The Dissertation Committee for Han Jiang Certifies that this is the approved version of the following dissertation:

Characterization of Petrophysical Properties of Organic-rich Shales by Experiments, Lab Measurements and Machine Learning Analysis

Committee:

Hugh Daigle, Supervisor

Mukul Sharma

Maša Prodanović

Zoya Heidari

Farzam Javadpour

Characterization of Petrophysical Properties of Organic-rich Shales by Experiments, Lab Measurements and Machine Learning Analysis

by

Han Jiang

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2018

Dedication

This is dedicated to my beloved parents and fiancée.

Acknowledgements

I would like to express my sincere appreciation to my supervisor Dr. Hugh Daigle for his instruction, guidance, and invaluable help. He not only taught me how to conduct scientific research, but also gave me lots of freedom to choose the research topics, and try different techniques. Without his supervision and constant help this dissertation would not have been possible.

I would like to thank Dr. Mukul Sharma. He gave me many valuable advices on research. I also would like to acknowledge his Joint Industry Project of Hydraulic Fracturing and Sand Control for providing two-year funding for my research. I would also like to thank my other PhD committee members, Dr. Maša Prodanović, Dr. Zoya Heidari, and Dr. Farzam Javadpour. They provided many constructive comments and suggestions for my project and dissertation.

I would like to thank Dr. Nicolas W. Hayman, Dr. Michael Pyrcz, and Dr. Kitty Milliken for providing helps for my projects and paper. I would like to thank Glen Baum, Rod Russell, Gary Miscoe, and Daryl Nygaard for providing helps and advices for my lab work. I would like to thank Peter Polito and Dr. Athma Bhandari for helping me with the helium porosity and permeability measurements. I would like to thank Dr. David Medellín for providing the inversion program for NMR measurement. The NMR inversion code used in this study was provided by UT Austin's Research Consortium on Formation Evaluation. I would like to thank Dr. Jianshi Zhou, Dr. Zhiming Cui and Dr. Aijun Zhou for providing access and help for pyrolysis experiments. I would like to thank Donnie Brooks and Xavier Janson for providing access to the triaxial testing apparatus and performing tests.

I would like to thank my colleagues, Dr. Chunbi Jiang, Dr. Chunxiao Zhu, Dr. Boyang Zhang, and Dr. Weiwei Wu for helping my experiments. I would like to thank my colleagues, Xiao Tian, Christopher Griffith, and Abhishek Bihani for discussions of my research.

Finally, I would like to thank my fiancée Wendy Zhang. Thank you for your wisdom, patience, sacrifice. Thank you for letting me know who I am, what is love and the path for future.

Characterization of Petrophysical Properties of Organic-rich Shales by Experiments, Lab Measurements and Machine Learning Analysis

Han Jiang, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Hugh Daigle

The increasing significance of shale plays leads to the need for deeper understanding of shale behavior. Laboratory characterization of petrophysical properties is an important part of shale resource evaluation. The characterization, however, remains challenging due to the complicated nature of shale. This work aims at better characterization of shale using experiments, lab measurements, and machine leaning analysis.

During hydraulic fracturing, besides tensile failure, the adjacent shale matrix is subjected to massive shear deformation. The interaction of shale pore system and shear deformation, and impacts on production remains unknown. This work investigates the response of shale nanoscale pore system to shear deformation using gas sorption and scanning electron microscope (SEM) imaging. Shale samples are deformed by confined compressive strength tests. After failure, fractures in nanoscale are observed to follow coarser grain boundaries and laminae of OM and matrix materials. Most samples display increases in pore structural parameters. Results suggest that the hydrocarbon mobility may be enhanced by the interaction of the OM laminae and the shear fracturing.

Past studied show that the evolution of pore structure of shale is associated with thermal maturation. However, the evolution of shale transport propreties related to thermal

maturation is unclear due to the difficulty of conducting permeability measurement for shale.This work studies evolution of permeability and pore structure measurements using heat treatment. Samples are heated from 110°C to 650°C. Gas sorption and GRI (Gas Research Institute) permeability measurements are performed. Results show that those petrophysical parameters, especially permeability, are sensitive to drying temperature. Multiscale pore network features of shale are also revealed in this study.

Characterizing fluids in shale using nuclear magnetic resonance (NMR) T_1-T_2 maps is often done manually, which is difficult and subjected to human decisions. This work proposes a new approach based on Gaussian mixture model (GMM) clustering analysis. Six clustering algorithms are performed on T_1-T_2 maps. To select the optimal cluster number and best algorithm, two cluster validity indices are proposed. Results validate the two indices, and GMM is found to be the best algorithm. A general fluid partition pattern is obtained by GMM, which is less sensitive to rock lithology. In addition, the clustering performance can be enhanced by drying the sample.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Motivation and Objectives

In recent decades, shale gas has become a significant resource play in the USA. The combination of horizontal drilling and hydraulic fracturing allows the extraction of huge quantities of hydrocarbon from shale formations with extremely low permeability, which were previously thought to be either impractical or uneconomic (Wang et al., 2014; EIA 2018). According to the U.S. Energy Information Administration (EIA 2018), the continued development of shale gas and tight oil resources supports growth in natural gas plant liquids production, which is projected to reach 5.0 million bbl/d in 2023, a nearly 35% increase from the 2017 level.

The increasing significance of shale gas plays leads to the need for deeper understanding of shale behavior. Laboratory characterization of petrophysical properties is an important part of shale gas resource evaluation. There are many petrophysical properties which govern whether a particular shale will become a shale gas resource. Josh et al. (2012) summarized the key factors for shale plays: (1) organic matter abundance, type and thermal maturity, (2) porosity-permeability relationships and pore size distribution, and (3) mechanical properties (brittleness) and their relationship to mineralogy and rock fabric.

Shale is a fine-grained, organic-rich sedimentary rock with low porosity and low permeability, which behaves as both the source of and the reservoir for the hydrocarbons (Loucks and Ruppel 2007; Bhandari et al. 2015). Consisting of fine grains and organic matter, shale commonly contains a large proportion of micropores (< 2 nm) and mesopores (2-50 nm) (Loucks et al., 2009; Loucks et al., 2012). Gas sorption, especially N_2 and CO_2 sorption, is one of the most widely used techniques to quantitatively characterize micropore and mesopore structure in shale (e.g. Bustin et al., 2008; Ross and Bustin, 2009; Adesida

et al., 2011; Chalmers et al., 2012; Kuila and Prasad, 2013a, b; Clarkson et al., 2013; Wang et al., 2014). The main information that we can obtain from gas sorption is porosity, pore size distribution, and pore structure (Kruk and Jaroniec 2001; Kuila and Prasad, 2013a, b). These are important properties of a porous medium and affect most of its behavior, including elastic and mechanical behavior and flow of fluids (Kuila and Prasad, 2013b).

The economical production of hydrocarbon from shale depends on hydraulic fracturing. During hydraulic fracturing, a network of highly conductive fractures is created by tensile failure, which enhances hydrocarbon transport to the wellbore (Nolte, 2000; Arthur et al., 2009). The adjacent shale matrix is subjected to massive shear deformation, as shear mechanism is found active in the microseismicity induced by hydraulic fracturing (Maxwell and Cipolla, 2011; Williams-Stroud et al., 2013; Busetti et al., 2014; Roux, 2016). Whether shear deformation may lead to microfractures and enhance hydrocarbon transportation from nano-scale organic matter pores to the larger tensile fracture remain unknown. It is vitally important for laboratory measurements to characterize the pore system at nanoscale and quantify the response to the shear deformation.

Laboratory measurements of shale are challenging, and a lack of standardization in certain techniques makes comparing results difficult. Permeability measurements of shale typically employ unsteady-state methods, which include the pulse decay method (Brace et al., 1968; Dicker and Smits, 1988) and the GRI (Gas Research Institute) method (Luffel et al., 1993; Cui et al., 2009). For gas-based (e.g. helium, nitrogen) permeability methods, the shale sample often requires drying before the permeability measurement to obtain as accurate an estimate of the intrinsic permeability as possible (e.g. Cui et al., 2013; Alnoaimi et al., 2014; Heller et al., 2014; Ghanizadeh et al., 2015). However, in practice the drying temperature is often below 120°C, which may still cause the pore system to be only partially accessible to the probe gas, resulting in an incomplete characterization of shale

transport properties. In addition, past research has shown that when the temperature is above 300°C, the OM can be matured and create new pores and alter the structure of the pore system. All this leads to the need for a systematic investigation the evolution of shale permeability and pore structure after heat treatment.

Characterizing the distribution of water, hydrocarbon, and organic matter within organic shale is an important aspect of shale analysis, and developing more reliable laboratory and analysis methods to this is another challenge. Low-field nuclear magnetic resonance (NMR) has proved to be a powerful technique for characterizing shale rock and tight oil (e.g. Mullen, 2010; Odusina et al., 2011; Lewis et al., 2013; Tinni et al., 2014). Recently, NMR T_1-T_2 2-D maps are commonly used for fluid characterization in shale (Washburn and Birdwell, 2013; Daigle et al., 2014; Gips et al., 2014; Fleury and Romero-Sarmiento, 2016; Mehana and El-monier, 2016). However, the interpretation of the measurements result is often done manually, which is challenging and often subjected to human decisions. Furthermore, the complex nature of the shale pore/fluid system adds uncertainty for the interpretation. To overcome the challenge of the manual characterization method, it requires the exploration and development of advanced techniques such as machine learning.

The main objectives of this dissertation are 1) to systematically investigate the response of the shale pore system to shear deformation at the nanoscale; 2) to study the effects of removal of fluids and thermal maturation on shale permeability and pore structure measurements; and 3) to develop a new fluid characterization approach using NMR T_1-T_2 maps in shale based on clustering analysis. By the integration of experiments, lab measurements, and machine leaning analysis, this work allows a better characterization of shale petrophysical properties.

1.2 OUTLINE OF DISSERTATION

The dissertation consists of 7 chapters, and the content of each chapter is briefly summarized below:

Chapter 1 describes the problems and introduces the motivation and objectives of this project.

Chapter 2 reviews the properties of the shale and shale pore system, and techniques that have been applied to study the petrophysical properties of shale including gas sorption, permeability methods, and NMR. The theory of clustering and related algorithms are also reviewed.

Chapter 3 studies the response of shale pore system at nanoscale to shear deformation. Shale samples from the northern Rocky Mountains (NoRM) and the Eagle Ford (EF) Formations are experimentally deformed using confined compressive strength tests. N_2/ CO_2 sorption and scanning electron microscope (SEM) imaging are performed to characterize fracture effects on pore morphology at nanoscale. After failure, fractures with widths ranging from 10-100 nm up to 1-2 μm are observed to follow coarser grain boundaries and laminae of OM and matrix materials. Most samples display increases in pore structural parameters including Brunauer-Emmett-Teller (BET) surface area, N_2/ CO_2 porosity and surface fractal dimensions after failure. Compared to other parameters, surface fractal dimensions are less sensitive to shear failure. The interaction of the OM laminae and the shear fracturing may improve the connectivity of the OM laminae to the adjacent rock matrix, and thus enhance the hydrocarbon mobility.

Chapter 4 studies effects of removal of fluids and thermal maturation on permeability and pore structure measurements of shale. NoRM and EF shale samples are dried at four temperature levels (110°C, 250°C, 450°C, >= 600°C). N_2 gas sorption and GRI (Gas Research Institute) permeability measurements are performed after each heating

level. Results show that BET surface area, N_2 pore volume, GRI porosity and GRI permeabilities increase as drying temperature increases, due to a combination of progressive loss of volatiles with increasing temperature and artificially induced thermal maturation at high temperatures. GRI permeabilities are strongly affected by heating temperature, which suggest that caution is warranted in interpreting permeability measurements performed on shales. The results also indicate the multiscale pore network structure of the shale where a certain fraction of small pores could be partially or completely disconnected from the overall structure.

Chapter 5 proposes a new fluid characterization approach of NMR T_1-T_2 in shale based on clustering analysis. Six clustering algorithms including Gaussian mixture model (GMM) are performed on T_1-T_2 maps of NoRM and EF samples at as-received and dried at 110°C conditions. Two cluster validity indices are proposed for the selection of cluster number and best algorithm. GMM is shown to be the best algorithm in most of the cases. The drying procedure helps to reveal the NMR footprint from organic matter, allowing better performance of clustering compared to fluid condition at as-received conditions.

Chapter 6 extends the analysis of GMM clustering approach through the application to 4 organic-rich shale samples at as-received and dried at 110°C conditions. The two indices and GMM approach are further validated. Heating at 110°C can help to reveal the footprint of OM, generating better cluster performance. Fluid types identified at dried at 110°C conditions are comparable with previous studies. In addition, the fluid partitioning rule obtained by GMM show a general pattern that is less sensitive to rock lithology.

Chapter 7 summarizes this project and provides recommendations for future work.

Chapter 2: Background

2.1 Shale Pore System

Shale is a fine-grained, organic-rich sedimentary rock with low porosity and low permeability, which behaves as both the source of and the reservoir for the hydrocarbons (Loucks and Ruppel 2007; Bhandari et al. 2015). Consisting of fine grains and organic matter, shale commonly contains a large proportion of micropores (< 2 nm) and mesopores (2-50 nm) (Loucks et al., 2009; Loucks et al., 2012). These structures are critical for analyzing shale's sealing ability and mechanisms of hydrocarbon migration, as well as the storage state of gas and oil in shale (Wang and Ju, 2015). The characterization of the pore size distribution, specific surface area and total volume of micropores and mesopores is becoming the focus of a great deal of research (Bustin et al., 2008; Loucks et al., 2009; Ross and Bustin, 2009; Clarkson et al., 2013).

Matrix-related pore networks for shale are composed of nanometer- to micrometer-sized pores (Loucks et al., 2012). According to the classification system of Loucks et al. (2012), there are mainly three types of pores in shale: interparticle and intraparticle pores associated with the mineral matrix, and organic matter (OM) pores. The interparticles are pores found between particles and crystals (Figure 2.1). Intraparticle pores are pores located within particles, such as interplatelet pores with clay aggregates, and intercrystalline pores within pyrite framboids. OM pores are located within organic matter (Figure 2.2). Fracture pores are not part of the classification, as they are not controlled by the individual particles (Loucks et al., 2012)

The OM pores are created during hydrocarbon maturation (Jarvie et al., 2007; Loucks et al., 2009; Bernard et al., 2012), and exhibit irregular ellipsoidal shapes with a pore size range from about 1 to 500 nm in most gas shale systems. This nanoporosity has

been interpreted as resulting from the exsolution of gaseous hydrocarbons during the secondary thermal cracking of retained oil and has been suggested to greatly influence, if not control, gas storage capacity and permeability of gas shale systems (Ambrose et al., 2010; Curtis et al., 2010, 2012; Passey et al., 2010; Slatt and O'Brien, 2011).

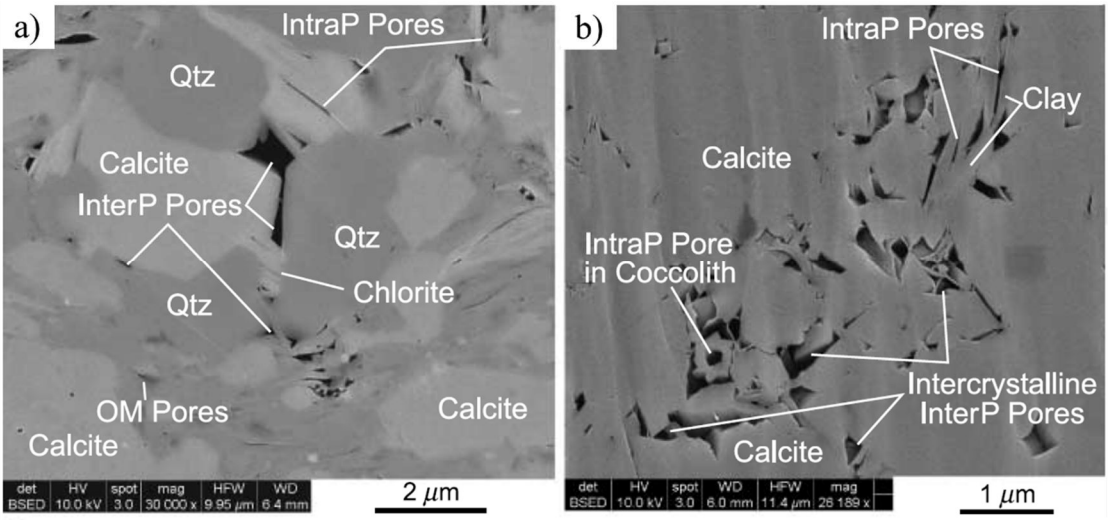

Figure 2.1 Example of interparticle (interP) and intraparticle (intraP) pores within shales. (a) Interparticle pores between quartz (Qtz) and calcite grains with cement overgrowths. The shale sample was from Lower Cretaceous Pearsall Formation, Maverick County, Texas. Vitrinite reflectance (Ro) = 1.5% (Loucks et al., 2012). (b) Sample contains intercrystalline-appearing interP pores. Intraparticle pores are also present in the center of the coccolith and along cleavage planes of distorted clay grains. The shale sample is from Upper Cretaceous Austin Chalk, La Salle County, Texas (Loucks et al., 2012). Vitrinite reflectance (Ro) = 0.9 %. Two plots are from Figure 6 in Loucks et al. (2012).

Vitrinite reflectance (Ro) is used to quantify the maturity of the shale samples. Shale with Ro less than 0.5~0.7% is considered immature; 0.5 to 0.7% < Ro <1.3% is referred to as the oil window; 1.3% < Ro < 2% is referred to as the zone of wet gas and condensate; and Ro > 2% is referred to as the dry gas zone (Tissot and Welte 2012). Besides thermal maturity, total organic carbon (TOC) content, mineral composition and OM type

can also affect the generation of OM-associated pores (Milliken et al. 2013; Mastalerz et al. 2013).

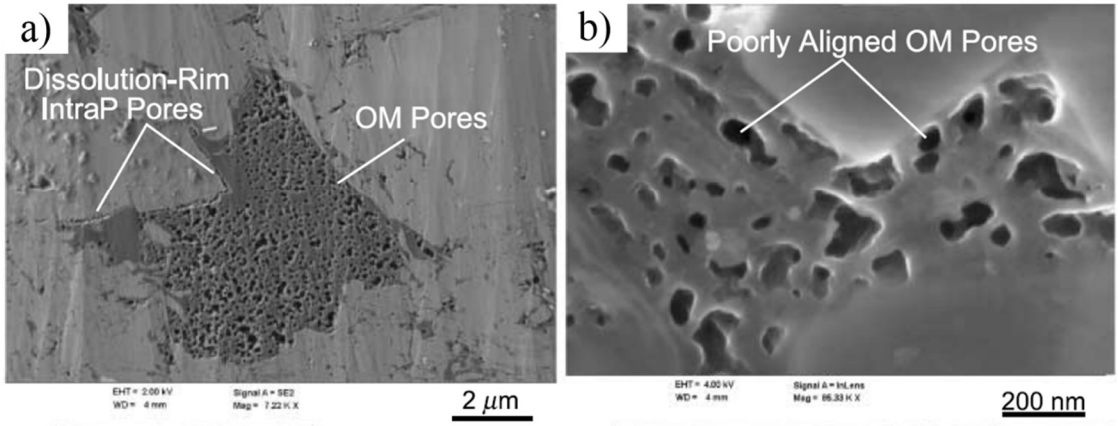

Figure 2.2 Example of organic matter (OM) pores within shales. (a) Large OM particle with OM pores. Vitrinite reflectance (Ro) = approximately 1.6%. Sample was from Mississippian Barnett Shale, Wise County, Texas (Loucks et al., 2012). (b) Organic-matter pores slightly aligned and showing complexity in third dimension. The shale sample had approximately 1.6% Ro. It was from Mississippian Barnett Shale, Wise County, Texas (Loucks et al., 2012). Two plots are from Figure 10 in Loucks et al. (2012).

2.2 Pore Connectivity and Multiple Scales

OM-associated pores can be developed in both depositional OM and migrated OM. Depositional OM retains its position and shape from the time of deposition, while migrated OM may change shape and location in response to temperature and pressure during burial (Loucks and Reed, 2014). Increasing maturation can convert kerogen into bitumen, which can stay in place or migrate into the adjacent interparticle pores. Depositional OM and migrated OM can reduce the connectivity with adjacent inorganic matrix, which was supported by measurements of shale samples before and after OM removal (Kuila et al., 2014).

The pore system in shales consists of disconnected networks of cracks and voids at multiple scales (Jiang et al. 2015; Daigle et al., 2017a, 2017b, 2018). The concept of multiscale pore network is depicted in Figure 2.3 (Daigle et al., 2017a). At the scale of about 10 microns, the individual organic matter clusters have a connected pore network. Those connected organic porosity clusters are connected through the intervening inorganic matter (Figure 2.3b). At higher scale (about 100 mirons), groups of organic matter clusters are connected through the inorganic matter (Figure 2.3c). An example of a pore network extracted from an x-ray microcomputed tomography (micro-CT) image is shown in Figure 2.3d, which indicating typical multiscale networks.

The importance of multiscale pore network features on shale permeability has been similarly demonstrated by other researchers (e.g. Ambrose et al., 2012; Mehmani et al., 2013; Mehmani and Prodanović, 2014; Jiang et al., 2015; Xu et al., 2016; Tian and Daigle, 2018b). For example, Tian and Daigle (2018b) constructed a shale network model based on nitrogen sorption isotherms of Barnett shale samples. By matching the measured permeability, the model showed that the pore spatial arrangement is related to the size of the pores, where pores with small size (< 8 nm) tend to develop on the walls of pores with larger size (> 8 nm).

Petrophysical measurements performed on crushed shale samples present challenges to interpretation due to multiple scales of the shale pore system (Jiang et al. 2015; Daigle et al., 2018). The disconnected network at multiple scales of shale pore systems causes sample fragments to have size-dependent response to mercury injection capillary pressure (MICP) and gas sorption measurements (Jiang et al. 2015). Daigle et al. (2018) computed the connectivity of the shale pore network using percolation theory based on MICP and gas sorption data. The results showed that many samples do not percolate at the scale of laboratory measurements, indicating a multiscale network.

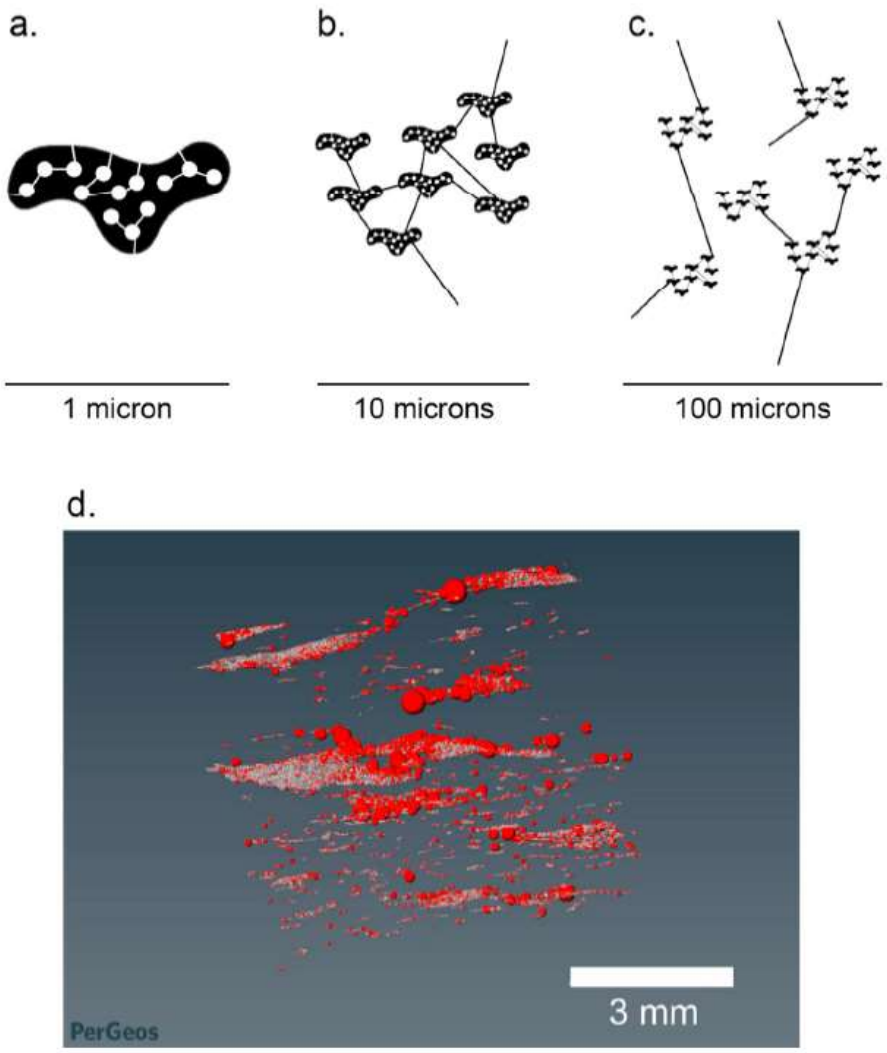

Figure 2.3 Multiscale pore system in shale. (a) Individual connected organic porosity cluster. White circles indicate pores, and white lines indicate connections between pores. (b) Connected organic clusters. They are connected through the intervening inorganic matrix. (c) Grouping of organic matter clusters, themselves connected through the inorganic matrix. (d) Example rendering of pore and fracture network in one of the intact siliceous shale samples based on a segmented x-ray microcomputed tomography (micro-CT) volume, showing typical multiscale features. Pores are represented by red spheres, and connections are represented by white lines/planes. Figure is based on Daigle et al. (2017a)

2.3 GAS SORPTION

Gas sorption, especially nitrogen (N_2) and carbon dioxide (CO_2) sorption, is one of the most widely used techniques to quantitatively characterize micropore and mesopore structure in shale (e.g. Bustin et al., 2008; Ross and Bustin, 2009; Adesida et al., 2011; Chalmers et al., 2012; Kuila and Prasad, 2013a, 2013b; Clarkson et al., 2013; Wang et al., 2014). The main information we can obtain from gas sorption is porosity, pore size distribution, and pore structure (Kruk and Jaroniec, 2001; Kuila and Prasad, 2013a, 2013b).

N_2 is the most widely used gas and can measure pores with pore size between 0.8 and 200 nm. CO_2 sorption is used to characterize micropores (< 2 nm). The drawback of the sorption technique is its inability to measure pores with sizes larger than 200 nm. For a full characterization of the entire pore range, combining of gas sorption with traditional MICP measurements is suggested (Bustin et al, 2008; Ross and Bustin, 2009). This should be done with caution as the inverted pore information from different measurements are based on different simplifying assumptions and the different techniques record different aspects of the pore structure.

During the sorption measurement (Figure 2.4), the adsorbate N_2 is dosed into the sample in controlled pressure increments. The pressure is allowed to equilibrate between doses. The adsorbate molecules physisorb on the surface of the solid through intermolecular forces. At the end of adsorption, the internal surface of the sample will be completely covered and all pores will be filled by condensed adsorbate. Desorption then proceeds by withdrawal of gas in prescribed pressure decrements. The result, a set of pressures and adsorbed gas quantities, is called the isotherm. Pressure is reported as relative pressure, which is the normalized pressure with respect to saturation vapor pressure (about 1 atm for N_2 at 77 K). Adsorbed quantity is reported in gas volume over unit sample mass at standard pressure and temperature condition.

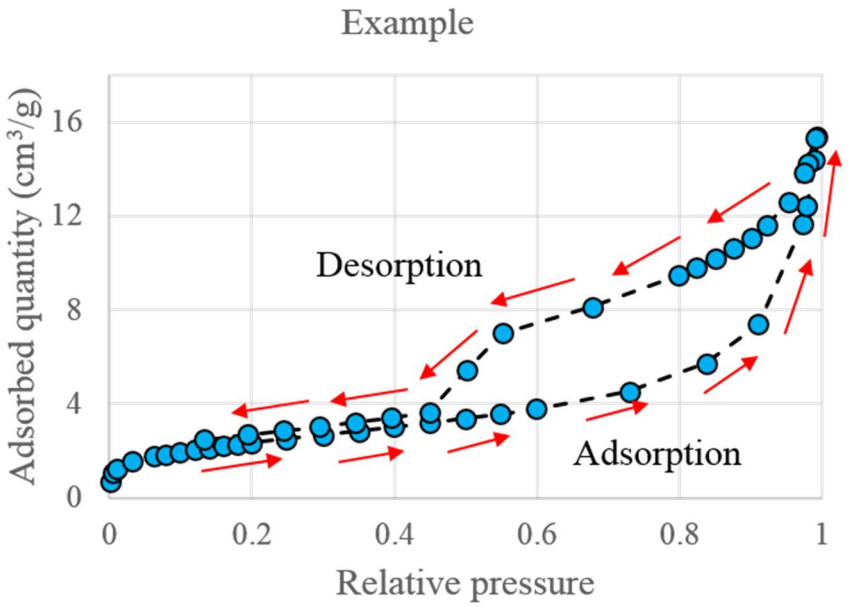

Figure 2.4 Example of gas adsorption and desorption isotherm of a shale sample using N_2. The isotherm is the change of adsorption (desorption) gas quantity in term of relative pressure. The lower branch is the adsorption branch and the upper branch is desorption branch. Blue circles are the isotherm points collected during the measurements. Arrows indicate the direction of the pressure. Pressure is reported as relative pressure, which is the normalized pressure with respect to saturation vapor pressure (1 atm for N_2 at 77K). The range of relative pressure is from 0 to 1. Adsorbed quantity is reported in gas volume per unit sample mass at standard pressure and temperature conditions.

Since shale is rich in micropores and mesopores (Loucks et al., 2009; Loucks et al., 2012), the adsorption process on shale pore walls starts with micropore filling, which occurs at very low relative pressure. As pressure increases, the adsorption for mesopores take places with two distinct stages: monolayer-multilayer adsorption and capillary condensation (Sing, 1985). In monolayer adsorption, all the adsorbed molecules are in contact with the surface of the adsorbent. In multilayer adsorption, the adsorption space accommodates more than one layer of molecules so that not all adsorbed molecules are in

direct contact with the surface of the adsorbent (Sing, 1985). In capillary condensation, the residual pore space which remains after multilayer adsorption has occurred is filled with condensed adsorbate separated from the gas phase by menisci.

2.3.1 Hysteresis Mechanisms

Gas adsorption-desorption isotherms exhibit hysteresis loops, where the desorption isotherm does not retrace the adsorption isotherm but rather lies above it over a range of relative pressures. The relative pressure point at which the loop is closed depends on the nature of adsorbate. For nitrogen at 77 K, the closure pressure point of the adsorption-desorption isotherm is around $P/P_0 = 0.42$ (Ravikovitch and Neimark, 2002).

There are mainly three factors contributing to the adsorption/desorption hysteresis, including capillary condensation effects (Gregg and Sing, 1982; Pinson et al., 2018), pore network effects (Mason, 1982; Seaton 1991; Tanev and Vlaev, 1993), and cavitation effects (Kadlec and Dubinin, 1969; Burgess and Everett, 1970; Ravikovitch and Neimark, 2002; Thommes et al., 2006).

The first effect is called the capillary condensation effect, which is due to differences in the shape of the liquid-vapor interface in an isolated pore during wetting and drying (Gregg and Sing, 1982; Pinson et al., 2018). Consider a simple pore consisting of a spherical cavity with a narrow cylindrical neck (Ravikovitch and Neimark, 2002). During adsorption, the pore filling process by condensation follows the formation of a liquid film on the cavity wall and thus is controlled by the radius of the curvature of the cavity r_c. During desorption, evaporation occurs after the formation of a hemispherical meniscus in the pore neck and is controlled by the radius of the neck r_n. The condensation pressure p_c is a function of the radius, according to the Kelvin-Laplace equation $\frac{p_c}{p_0} = \exp(-\frac{2\gamma v_1}{RTr})$, where p_0 is the saturation vapor pressure at the given temperature T, γ is

the liquid-vapor surface tension, v_1 is the liquid molar volume, and R is the universal gas constant. Since the cavity radius r_c is greater than the neck radius r_n, the condensation pressure during desorption is higher than that of adsorption.

The second effect is the pore network effect, which can act to broaden hysteresis in a system with a wide range of pore sizes and good pore connectivity (Pinson et al., 2018). The additional hysteresis is due to some pores remaining full below the relative pressure at which the empty state is thermodynamically favored, because they lack the connection with the vapor phase that is necessary to nucleate the liquid-to-vapor transition (Seaton 1991; Tanev and Vlaev, 1993).

The third effect that controls desorption at low relative pressure is called cavitation (Ravikovitch and Neimark, 2002; Groen et al., 2003; Thommes et al., 2006). This theory is developed to explain the lower closure point of sorption hysteresis loops (Kadlec and Dubinin, 1969; Burgess and Everett, 1970). The cause of the closure of desorption and adsorption branch is that the condensed phase becomes unstable at sufficiently low pressure. Upon further pressure decrease, the liquid-vapor meniscus ceases to exist and results in a forced closure of the hysteresis loop (Groen et al., 2003).

2.3.2 BET Surface Area

The Brunauer-Emmett-Teller (BET) method is used to calculate the specific surface area (Brunauer et al., 1938; Gregg and Sing, 1982; Sing, 1985; Roque-Malherbe, 2007). This method is derived in a similar fashion to the Langmuir isotherm, but with the allowance for multilayer adsorption. The relative pressure range used for calculation is between 0.05 to 0.3 (Webb and Orr, 1997). The relative pressure and the quantity of gas adsorbed may be transformed into a linear relationship (Brunauer et al., 1938):

$$\frac{1}{Q(\frac{p_0}{p}-1)} = \left(\frac{C-1}{Q_m C}\right)\left(\frac{p}{p_0}\right) + \frac{1}{Q_m C}, (2.1)$$

where Q is the gas adsorbed quantity, p/p_0 is the relative pressure, Q_m is the monolayer capacity, and C is a constant.

The BET method treats the expression on the left-hand side as a linear function in terms of p/p_0 on the right-hand side. Using a least-squares fit, one can obtain the slope and intercept. Two coefficients Q_m and C and be obtained from that, and specific surface area (hereinafter referred as BET surface area) can be calculated from the following equation (Sing, 1985):

$$A_{BET} = N_A \frac{Q_m}{Q_{mol}} \sigma, (2.2)$$

where N_A is the Avagadro constant, σ is the cross-sectional area effectively occupied by an adsorbed molecule (Roque-Malherbe, 2007), and Q_{mol} is the adsorption quantity per unit mass at standard temperature and pressure.

2.3.3 Pore Size Distribution

Geometric topology information like pore size distribution can be extracted from the adsorption isotherm. The Barrett-Joyner-Halenda (BJH) method and density functional theory (DFT) are two common methods applied to compute pore size distribution (Roque-Malherbe, 2007). BJH considers the adsorption process as monolayer-multilayer adsorption and condensation based on Kelvin's equation and a statistical adsorbed film thickness equation (Barrett et al. 1951). DFT, on the other hand, provides a modern statistical thermodynamic approach for calculating pore size distribution (Adesida et al., 2011). A detailed description of DFT can be found in Roque-Malherbe (2007). Because

the BJH model cannot account for interaction between adsorbed molecules and the opposing pore wall, which becomes important in pores < 7 nm in width, it fails to interpret micropores in a rigorous fashion (Lastoskie et al. 1993). DFT takes this interaction into consideration and works well in micropores, but it does not provide a direct relation between pressure and pore size which makes sorption modeling difficult. More details can be found from Appendix A.

2.4 Overview of Permeability Measurements

The low permeability of shale not only presents a challenge for commercial production, but also for experimental measurements of rock samples. Methods to determine permeability of low permeability rock cores and crushed rock samples directly can be divided into two categories: steady-state and unsteady-state.

2.4.1 Steady-state Method

The steady-state method is the standard to determine permeability for conventional rocks in the laboratory (e.g. Jones and Meredith,1998; Amann-Hildenbrand et al., 2012, 2013; Dong et al., 2012; Gensterblum et al., 2014). It is regarded as an accurate and reliable technique, with the advantage of a comparatively simple experimental set-up and a straightforward analytical solution (Sander et al., 2017). However, the time required to reach equilibrium is usually very long for rock samples with low permeability (Cui et al., 2009). For this reason, steady-state techniques are typically not used to measure the permeability of shale.

2.4.2 Pulse Decay Method

Due to the low permeability of shale, the most popular methods are unsteady-state methods, which include the pulse decay method and GRI method. These involve calculating the permeability based on pressure vs. time data.

The pulse decay method (Figure 2.5) appears to be the most commonly applied experimental method to determine permeability in low-permeability porous media (e.g. Brace et al., 1968; Kwon et al., 2001; Escoffier et al., 2005; Billiotte et al., 2008; Fedor et al., 2008; Chalmers et al., 2012; Firouzi et al., 2014; Mokhtari and Tutuncu, 2015; Zhang et al., 2015, 2016; Cao et al., 2016; Kumar et al., 2016). This method gains its popularity from its shorter experimental run times, and higher resolution for very low permeability measurements, compared with the steady-state method (Coyner et al., 1993; Cui et al., 2009).

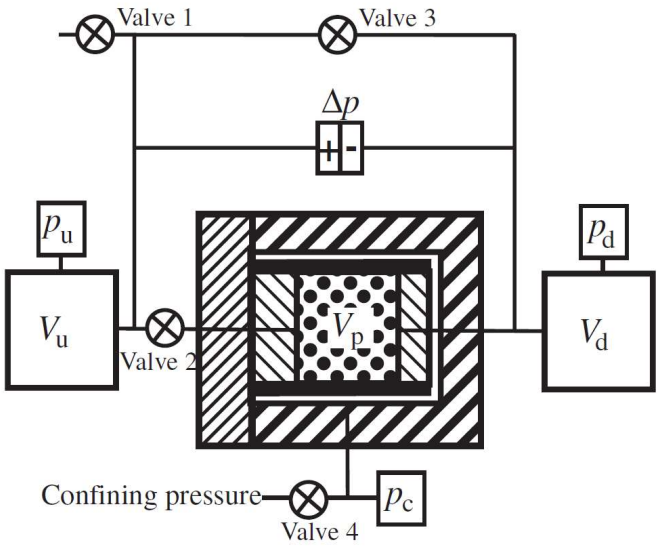

Figure 2.5 Example of the set-up of a pulse decay experiment. It consists of an upstream reservoir of volume V_u, a downstream reservoir of volume V_d, and a cell capable of applying hydrostatic confining pressure (P_c) and containing a cylindrical rock sample with a total pore volume V_p. The figure is from Cui et al. (2009).

Brace et al. (1968) proposed the pulse decay method to determine the nanodarcy permeability of granite using water and argon. The basic setup is presented in Figure 2.5. The pulse decay apparatus consists of an upstream reservoir, a downstream reservoir, and a cell capable of applying confining pressure. The sample, which is usually cylindrical, is held in the pressurized cell. Upstream and downstream pressures are measured by pressure transducers. The pressure difference between the upstream and downstream ends is measured by differential pressure transducers (Sander et al., 2017; Cui et al., 2009). Many set-ups also allow for an axial load to be applied to the sample during the experiment (Sander et al., 2017).

2.4.3 GRI Method

The GRI method, proposed by Luffel et al. (1993), is a variation of the pulse decay method (Sander et al., 2017). It is also commonly referred to as the pressure fall-off method (Cui et al., 2013). It provides an permeability estimation using a crushed sample. The crushed sample is analyzed in an apparatus consisting of a reference and a sample cell, valves and pressure transducers (Figure 2.6). The system should be temperature controlled, as gas properties like compressibility are assumed to be constant for permeability calculation (Cui et al., 2009). No confining pressure can be applied to the sample in this experimental set-up (Luffel et al., 1993).

The advantages of the GRI method applied to crushed samples (also known as the 'crushed method') lie in the speed of the experiments compared with core plug experiments (Luffel et al., 1993), as well as in the ability to use cuttings, which are typically easier to obtain than intact core plugs. The other advantage of the GRI method is that can provide a better estimation of shale matrix permeability, as crushing is believed to eliminate fractures

created by coring and handling (Luffel et al., 1993; Handwerger et al., 2011; Cui et al., 2013; Ghanizadeh et al., 2015).

The drawbacks of the method are: (1) performing measurements on crushed samples potentially eliminates the influence of larger scale features (such as microcracks), (2) the inability to perform measurements at reservoir conditions or measure how permeability evolves during production (Heller et al., 2014), and (3) no ability to control the direction of flow. In other words, only isotropic permeability can be measured.

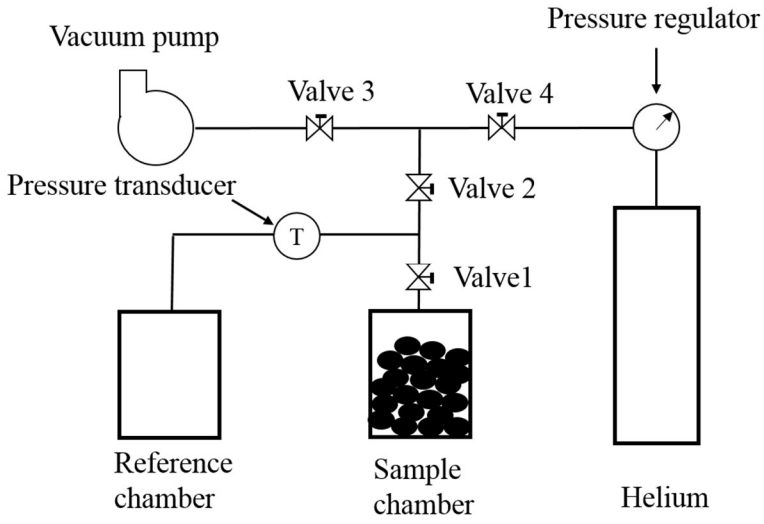

Figure 2.6 Example of the set-up of the GRI method using crushed sample. It consists of a reference chamber and a sample chamber. Black particles represent crushed shale samples. The reference chamber and sample chamber are kept in a temperature-controlled box (not shown in the figure) to keep a constant temperature.

2.4.4 Measurement and Calculation of GRI Method

The apparatus for the GRI measurement is shown in Figure 2.6. Initially, all valves are opened to the atmosphere through the outlet. A known weight of crushed sample is introduced to the sample chamber. Before the measurement, the sample is allowed to

stabilize for a few hours. The system is then evacuated using a vacuum pump (depicted as Stage 1 in Figure 2.7a). Then valves 1 and 3 are closed, and valves 2 and 4 are sequentially opened. Pressurized helium gas (around 220 psi) fills the reference chamber, and the system is allowed to equilibrate for a few minutes for stabilization (Stage 2 in Figure 2.7a). Then valve 1 is opened and the helium expands from the reference chamber into the sample chamber. The helium pressure immediately drops to a new value by filling the dead space in the sample chamber and subsequently decays as gas permeates the shale particles (Stage 3 in Figure 2.7).

Bulk porosity of the shale sample can be computed from the GRI measurement based on the Boyle's law. The vacuum pressure (P_1) at Stage 1 is calculated by the mean of pressure points with values less than 0.5 psi. The pressure at Stage 2 (P_2) is the mean of pressure points during that period. The final decay equilibrium pressure (P_3) at Stage 3 is the mean of the last several pressure points. The porosity is calculated using the following equation (Cui et al., 2009):

$$\Phi = \left[V_r\left(\frac{P_2}{z_2} - \frac{P_3}{z_3}\right) + (V_s - V_b)\left(\frac{P_1}{z_1} - \frac{P_3}{z_3}\right)\right] / (\frac{P_3}{z_3} - \frac{P_1}{z_1})V_b \,, (2.3)$$

where V_r, V_s are volumes of reference and sample chambers, V_b is the bulk volume of the sample, and z_1, z_2, and z_3 are the compressibility factors of helium at pressures P_1, P_2, and P_3.

Gas transport in tight and fine-grained porous rocks may be due to diffusion, advection, or a combination of both (Cui et al., 2009). In either case, gas transport can be described by diffusion-type equations with density (or gas pressure) as the primary unknown (Cui et al., 2009). Sample particles are assumed to be spheres, so the diffusivity equation in spherical coordinates is expressed as:

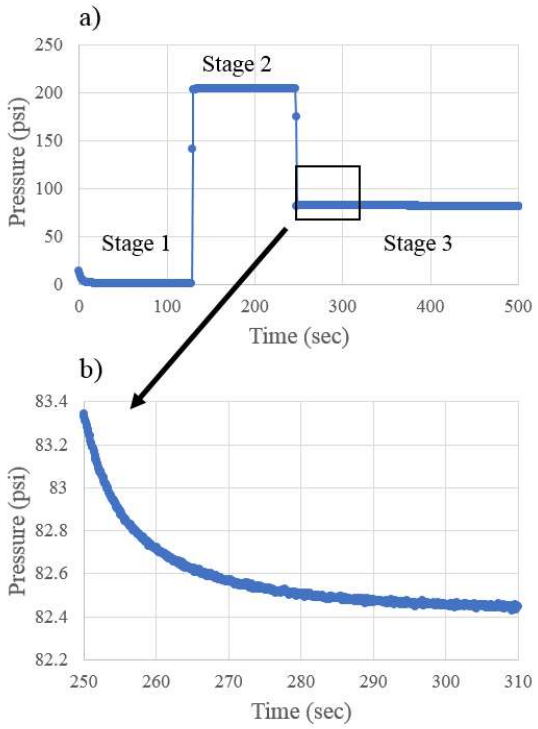

Figure 2.7 Example of recorded pressure data over time using a shale sample. (a) Pressure data includes Stage 1, Stage 2, and early time for Stage 3. (b) Early time for Stage 3 of the selected window in the plot of (a). Stage 1 is the period when the system is evacuated by a vacuum pump. Stage 2 is the period when pressurized helium gas fills the reference chamber. The Stage 3 is the period when helium expands from the reference chamber into the sample chamber. The entire measurement for Stage 3 is about 17-24 hours. The units of the time are second, and the units of pressure are psi.

$$\frac{\partial \rho}{\partial t} = \frac{K}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial \rho}{\partial r}\right) . (2.4)$$

The transport coefficient K is defined as:

$$K = \frac{k}{\mu c_g \Phi} , (2.5)$$

where the c_g is the gas compressibility, μ is the viscosity of the gas, k is the permeability and Φ is the porosity of the sample.

The boundary conditions are:

$$\frac{\partial \rho}{\partial r} = 0 \qquad at\ r = 0\ , (2.6a)$$

$$-4N\pi R_a^2 \frac{k}{\mu c_g}\frac{\partial \rho}{\partial r} = V_c \frac{\partial \rho}{\partial t} \qquad at\ r = R_a\ , (2.6b)$$

where $N = 3M/4\rho_b \pi R_a^3$, M is the sample mass, ρ_b is the sample bulk density, and V_c is the dead volume of both sample cell and reference cell.

The initial conditions are:

$$\rho = \rho_0 \qquad for\ 0 \le r < a \quad at\ t = 0\ , (2.7a)$$

and

$$\rho = \rho_{c0} \qquad for\ r = R_a \quad at\ t = 0\ , (2.7b)$$

where ρ_0 is the initial free gas density in the sample particles before gas expansion, ρ_{c0} is the average initial free gas density in the dead space in both sample and reference cells.

The analytical solution for gas density ρ in the void volume of the reference and sample cells given above conditions is expressed as (Carslaw and Jaeger, 1959; Cui et al., 2009):

$$\rho = \rho_{c0} - \frac{\rho_{c0} - \rho_0}{(K_c + 1)} + 6K_c(\rho_{c0} - \rho_0)\sum_{n=1}^{\infty} \frac{e^{-\frac{K\alpha_n^2}{R_a^2}t}}{K_c^2\alpha_n^2 + 9(K_c + 1)}\ , (2.8)$$

where α_n is the n^{th} root of

$$tan\ \alpha = \frac{3\alpha}{3 + K_c\alpha^2}\ . (2.9)$$

K_c is expressed as a function of reference volume V_r, sample volume V_s, sample bulk volume V_b and porosity Φ.

$$K_c = \frac{(V_r + V_s)}{V_b \, \Phi} \quad , (2.10)$$

Rearrange the above equation, we have

$$FR(t) = 6K_c(K_c + 1) \sum_{n=1}^{\infty} \frac{e^{-\frac{K\alpha_1^2}{R_a^2}t}}{K_c^2\alpha_n^2 + 9(K_c + 1)} \quad , (2.11)$$

where $FR(t)$ is defined as the gas fraction in the void volume of the reference and sample cells that will, but has not yet, penetrated into the sample particles. In addition, the adsorption effect of helium on the pore space can be ignored.

If K_c is large (e.g. K_c > 50), which implies that the gas reservoir or total void volume of the reference and sample cells are much larger than the gas storage capacity of the sample particles, the logarithmical value of FR becomes a linear function of time (Cui et al., 2009). The logarithm form of Equation (2.11) can be expressed as

$$ln(FR) = b - \frac{K\alpha_1^2}{R_a^2}t \, . (2.12)$$

Through linear regression, the slope $\frac{K\alpha_1^2}{R_a^2}$ of Equation (2.12) can be obtained. Based upon Equation (2.5), the permeability k can thus be estimated.

2.5 NUCLEAR MAGNETIC RESONANCE

Nuclear magnetic resonance (NMR) refers to the response of atomic nuclei to magnetic fields (Callaghan, 1993, Coates et al., 1999). It is a non-destructive method for characterization of porous materials. The two fundamental parameters that are investigated

are the longitudinal relaxation time T_1 and the transverse relaxation time T_2. The values of T_1 and T_2 depend on the fluid types, pore sizes and materials that make up the pore walls (Coates et al., 1999; Daigle et al., 2014). Based on those features, NMR measurements give insight of the pore structures and fluids of the rock. It is routinely applied in downhole logging (Mullen, 2010; Lewis et al., 2013) as well as lab investigation (Odusina et al., 2011; Tinni et al., 2014).

2.5.1 NMR Theory

NMR measurements can be made on nuclei with an odd number of protons or neutrons or both, such as hydrogen (^{1}H), carbon (^{13}C), and sodium (^{23}Na). Hydrogen, which has only one proton and no neutrons, is abundant in both water and hydrocarbons, has a relatively large magnetic moment, and produces a strong signal. Almost all NMR logging and NMR rock studies of porous media are based on responses of the nucleus of the hydrogen atom (Coates et al., 1999).

The first step of making an NMR measurement is to align magnetic nuclei with an external magnetic field B_0. When subjected to B_0, the hydrogen nuclei tend to precess and align with the direction of external magnetic field. The precessional frequency f is called Larmor frequency, and is given by:

$$f = \frac{\gamma B_0}{2\pi}, (2.13)$$

where γ is the gyromagnetic ratio, a measure of the strength of the nuclear magnetism.

Different nuclei have different γ values. For ^{1}H, $\gamma/2\pi$ = 42.58 MHz/T (Coates et al., 1999). For a given nuclear species, the gyromagnetic ratio has a fixed value, and the Larmor frequency is a function of the strength of the static magnetic field (Coates et al., 1999).

A measurable net magnetization in the longitudinal direction (z axis; Figure 2.8) is established through the alignment of protons. Applying a 90° oscillating B_1 in the transverse plane perpendicular to B_0 causes the magnetization to tip 90° to transverse plane (x-y plane; Figure 2.9). When the B_1 field is turned off, the proton population begins to dephase, or lose phase coherency—that is, the precessions of the protons will no longer be in phase with one another. Therefore, as dephasing progresses, the net magnetization decreases (Figure 2.9; Coates et al., 1999).

The longitudinal relaxation T_1 is measured by the inversion recovery method. In the inversion recovery method, the first 180° pulse inverts the magnetization 180° relative to the static magnetic field. After a specific wait time (the inversion time), a 90° pulse rotates the magnetization into the transverse plane, and the degree of recovery of the initial magnetization is measured (Coates et al., 1999).

The transverse relaxation T_2 is measured by CPMG (Carr-Purcell-Meiboom-Gill) sequence (Carr and Purcell, 1954; Meiboom and Gill, 1958). CPMG is a series of 180° pulses following the 90° oscillating pulse B_1. This pulse sequence can partially rephase the protons and generate a series of magnetization signals called a spin-echo train. Due to irreversible dephasing of molecular interactions and diffusion, the magnitude of the spin-echo train decays, and is characterized by a time constant T_2 (Coates et al., 1999).

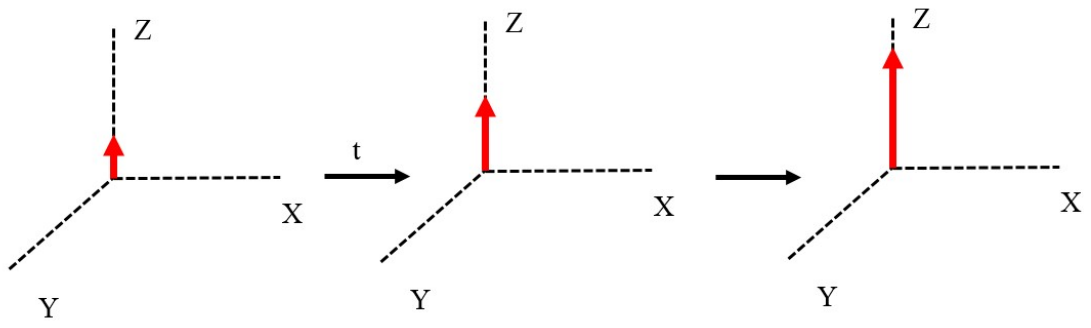

Figure 2.8 Buildup of magnetization in z direction. The magnetization is denoted by the red arrow.

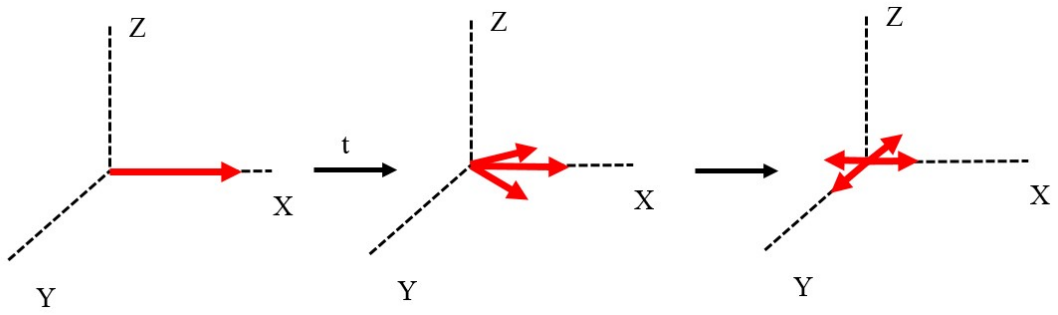

Figure 2.9 Decay of magnetization in x-y plane. The magnetization is denoted by the red arrow.

The expressions for T_1 and T_2 are:

$$M_z(t) = M_{0z}\left(1 - e^{-\frac{t}{T_1}}\right), (2.14)$$

$$M_{xy}(t) = M_{0xy}e^{-\frac{t}{T_2}}, (2.15)$$

where $M_z(t)$ is the magnitude along the z axis. M_{0z} is the final and maximum magnetization. $M_{xy}(t)$ is the magnitude along the x-y plane. M_{0xy} is the initial magnetization for the transverse relaxation.

2.5.2 Relaxation Mechanisms

For fluids in porous media, three independent relaxation mechanism are involved: bulk fluid processes, which affect both T_1 and T_2 relaxation; surface relaxation, which affects both T_1 and T_2 relaxation; and diffusion in the presence of magnetic field gradients, which only affects T_2 relaxation (Coates et al., 1999).

T_1 and T_2 of pore fluids may be expressed as:

$$\frac{1}{T_2} = \frac{1}{T_{2B}} + \frac{1}{T_{2S}} + \frac{1}{T_{2D}}, (2.16)$$

$$\frac{1}{T_1} = \frac{1}{T_{1B}} + \frac{1}{T_{2S}} . (2.17)$$

$T_{1\mathrm{B}}$ and T_{2B} are the bulk relaxation the pore fluid as it would be measured in a container so large that container effects would be negligible. $T_{2\mathrm{S}}$ and T_{2s} are the surface relaxation times of the pore fluid resulting from the pore surface. T_{2D} is the T_2 diffusion relaxation time of the pore fluid as induced by diffusion in the magnetic field gradient.

Furthermore, T_2 can be written as:

$$\frac{1}{T_2} = \frac{1}{T_{2B}} + \rho_2 \left(\frac{S}{V}\right) + \frac{D(\gamma G\, TE)^2}{12} , (2.18)$$

where ρ_2 is T_2 surface relaxivity (T_2 relaxing strength of the grain surfaces), $\frac{S}{V}$ is ratio of pore surface area to volume, D is molecular diffusion coefficient, γ is gyromagnetic ratio of a proton, G is field-strength gradient, and TE is inter-echo spacing used in the CPMG sequence.

T_1 can be written as:

$$\frac{1}{T_1} = \frac{1}{T_{1B}} + \rho_1 \left(\frac{S}{V}\right) , (2.19)$$

where ρ_1 is T_1 surface relaxivity (T_1 relaxing strength of the grain surfaces) and $\frac{S}{V}$ is ratio of pore surface area to volume.

2.5.3 NMR Inversion

2.5.3.1 Multi-exponential Decay

Equation (2.15) can be expressed in units of porosity (or fluid volume) by assuming 100% water saturation, which can be done by calibration sample made of a known volume of water:

$$m(t) = \phi e^{-\frac{t}{T_2}}, (2.20)$$

where ϕ is the rock porosity.

A natural rock commonly exhibits a distribution of pore sizes and frequently contains more than one type of fluid. Thus, instead of a single-exponential decay, Equation (2.20) should be expressed as a summation of multiple exponential components:

$$m(t) = \sum \phi_j e^{-\frac{t}{T_{2j}}}, (2.21)$$

where ϕ_j is the porosity coefficient which contributes to the total porosity from pores associated with the j^{th} component and T_{2j} is the decay constant of the j^{th} component of transverse relaxation.

In the measurement data, the decay function $m(t)$ is a function of amplitude over n points of timestamp. Furthermore, the continuous T_2 distribution is discretized into q points. Expressing Equation (2.21) as a system of equations yields

$$\begin{pmatrix} m_1 \\ m_2 \\ \cdots \\ m_i \\ \cdots \\ m_n \end{pmatrix} = \begin{pmatrix} K_{1,1} * \phi_1 + K_{1,2} * \phi_2 + \cdots + K_{1,j} * \phi_j + \cdots + K_{1,q} * \phi_q \\ K_{2,1} * \phi_1 + K_{2,2} * \phi_2 + \cdots + K_{2,j} * \phi_j + \cdots + K_{2,q} * \phi_q \\ \cdots \\ K_{i,1} * \phi_1 + K_{i,2} * \phi_2 + \cdots + K_{i,j} * \phi_j + \cdots + K_{i,q} * \phi_q \\ \cdots \\ K_{n,1} * \phi_1 + K_{n,2} * \phi_2 + \cdots + K_{n,j} * \phi_j + \cdots + K_{n,q} * \phi_q \end{pmatrix}, (2.22)$$

where $K_{i,j} = e^{-\frac{t_i}{T_{2j}}}$, i is the i^{th} time index, j is the j^{th} pore component index.

The matrix form of Equation (2.22) is:

$$m = K\phi, \qquad K \in \mathbb{R}^{n*q}, \quad m \in \mathbb{R}^{n}, \quad \phi \in \mathbb{R}^{q} \ , (2.23)$$

2.5.3.2 Inversion Technique

One common method is using direct matrix inversion to find out the best porosity components $\phi_i (i = 1, \ldots, n)$ (Hansen, 2010; Medellín et al., 2015; Medellín et al., 2016). The residual e_i is defined as:

$$e_i = m_i - \sum K_{i,j}\phi_j \ , (2.24)$$

It represents the difference between the i^{th} observed decay and i^{th} response value that is predicted by the linear model. The residual sum of squares (RSS) is defined as:

$$RSS = e_1^2 + e_2^2 + \cdots + e_i^2 + \cdots + e_n^2 \ , (2.25)$$

or:

$$RSS = \sum_{i=1}^{n} (m_i - \sum K_{i,j}\phi_j)^2 \ , (2.26)$$

RSS is called loss function. The goal is to find out the best porosity components $\phi_i (i = 1, \ldots, n)$, so that the loss function can be minimized (Hansen, 2010). However, due

to the nature of inversion problem, a small random perturbation of m can lead to very large perturbation of ϕ. In practice, an additional term called a regularization term is added to the RSS (Hansen, 2010). The final expression for the loss function with regularization is:

$$Loss = \sum_{i=1}^{n} (m_i - \sum K_{i,j}\phi_j)^2 + \lambda \sum_{j=1}^{p} \phi_j^2 = RSS + \lambda \sum_{j=1}^{p} \phi_j^2 \,, (2.27)$$

where λ is a tuning parameter for the regularization term.

A small λ poses a small regularization effect on the cost function, so the solution might still be affected by noise in the measurement. On the other hand, a large λ can lead to artificial smoothing of the solution. The choice of λ needs to be made carefully. A detailed discussion of choosing the tuning parameter can be found in Bauer and Lukas (2011).

Since ϕ cannot below zero, non-negative least square (NNLS) or linear programming (LP) can be used to solve for the loss function. The algorithm of NNLS follows the work of Lawson and Hanson (1995), which is available in MATLAB as LSQNONNEG.

2.5.4 NMR 2-D T_1-T_2 Measurement

For conventional reservoirs, the standard T_2 NMR measurement is sufficient to detect the quantity of water or gas in the porous media. However, for complex porous media such as shales, a single type T_2 or T_1 is not enough (Fleury and Romero-Sarmiento, 2016). The pore sizes are typically in the nanometer range, leading to very short T_2 relaxation times. In addition, a significant amount of the porosity in shales resides in organic phases (Washburn and Birdwell, 2013), giving rise to the possibility of

homonuclear dipolar coupling between fluids present and the organic matrix. This makes the interpretation of the shale system challenging. Although T_1 is easier to measure compared with T_2, the amount of signal measured from the T_1 experiments is often significantly less than what is determined by T_2 experiments. The low signal quality of T_1 leads to difficulty in characterizing the sample, and new NMR techniques are required to better understand the describe shale samples (Washburn and Birdwell, 2013)

Because of the above problems, using T_1/T_2 2D measurement on shale has become popular in recent years (Washburn and Birdwell, 2013; Daigle et al., 2014; Gips et al., 2014; Fleury and Romero-Sarmiento, 2016). The T_1-T_2 method provides better differentiation between the different fluids responses. T_1/T_2 ratio is a function of fluid viscosity and pore size (Daigle et al., 2014; Fleury and Romero-Sarmiento, 2016). For a small-molecular size and low-viscosity fluid like water, the T_1/T_2 ratio is close to 1. The ratio will increase with an increase in molecular size and viscosity.

Furthermore, Daigle et al., (2014) introduced a method to map T_1 and T_2 to two new variables: T_1/T_2 ratio (denoted as R) and secular relaxation time (denoted as Ts):

$$R = \frac{T_1}{T_2}, (2.28)$$

$$Ts = \frac{1}{\frac{1}{T_2} - \frac{1}{T_1}} = \frac{T_2 T_1}{T_1 - T_2}, (2.29)$$

R and Ts are functions of pore size and viscosity. R is about 1 for viscosity less than 1000 poise and increases at higher viscosities. On the contrary, Ts is large at low viscosity and decreases with increasing viscosity (Daigle et al., 2014). What's more, Ts is sensitive to the variation of the pore size. In smaller pores, interactions with paramagnetic ions on pore walls become important cause of relaxation, causing Ts to

decrease (Daigle et al., 2014). Based on characteristics of R and Ts, seven regions are defined and shown in Figure 2.10. Linear decision boundaries are chosen to cut the new 2-D space into non-overlapping sub-spaces where each sub-space represents one fluid type (Daigle et al., 2014; Jiang et al., 2018b).

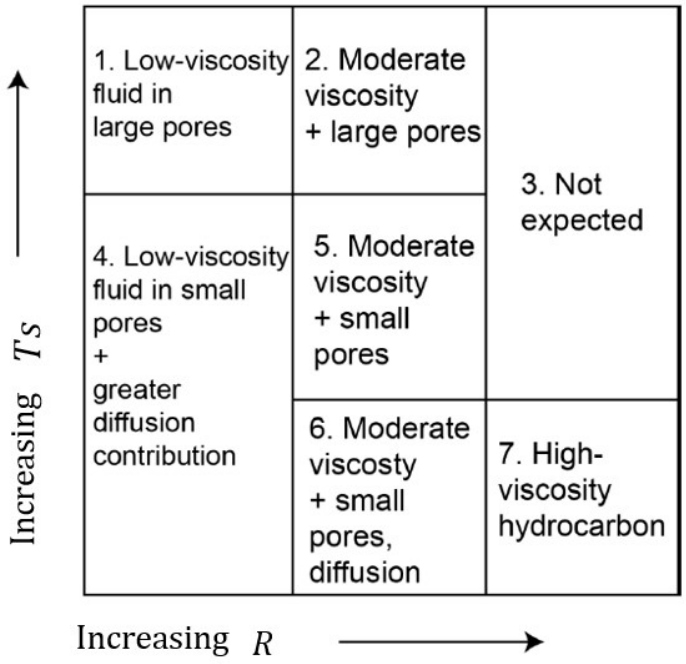

Figure 2.10 Relaxation regimes on R and Ts. Seven regions are defined. Figure is modified from Daigle et al. (2014).

2.6 CLUSTER ANALYSIS

Machine learning is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Cluster analysis is a type of machine learning approach to discover the natural groups of a set of observations (Jain et al., 1999; Gan et al., 2007; Jain, 2010; Aggarwal and Reddy, 2013). Cluster analysis has been widely applied in numerous fields such as image segmentation (Shi and Malik, 2000; Comaniciu and Meer, 2002), document retrieval

(Sahami and Koller, 1998; Bhatia and Deogun, 1998), biology (Baldi and Hatfield, 2002; Yeung et al., 2003), and geochemistry (Templ et al., 2008; Grunsky, 2010). It can be an alternative method for fluid characterization of T_1-T_2 maps in shale.

2.6.1 Basic Notions

A dataset used for clustering is a set of n observations denoted as $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots \boldsymbol{x_n}\}$. An observation $\boldsymbol{x_i}$ is a single data item, which consists of a vector of m elements: $\boldsymbol{x_i} = (x_{i,1}, x_{i,2}, \ldots x_{i,m})$. The individual elements are called features or dimensions. The dataset can be viewed as a $n \times m$ matrix.

A distance measure is a metric of the n-dimensional feature space used to quantify the similarity of observations (Jain et al., 1999). The most popular distance measure is the Euclidean distance. A detailed discussion of distances can be found in Jain et al. (1999). The expression for Euclidean distance between 2 observations $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ is below:

$$d(\boldsymbol{x_i}, \boldsymbol{x_j}) = \sqrt{\sum_{k=1}^{m} (x_{i,k} - x_{j,k})^2}, (2.30)$$

where $x_{i,k} - x_{j,k}$ is the difference of the two observations i and j in the k^{th} dimension.

2.6.2 Algorithms

The process of clustering is to assign observations to different groups, so that observations in the same group are as similar as possible, and observations in different groups are as dissimilar as possible (Jain, 2010; Aggarwal and Reddy, 2013). There are thousands of clustering algorithms proposed in the literature (Jain, 2010), and the popular

algorithms can be described with the help of the taxonomy of clustering (Figure 2.11; Jain et al., 1999; Jain, 2010).

Clustering algorithms can be broadly divided into two branches at the top level: hierarchical and partitional (Jain et al., 1999; Jain, 2010). In the hierarchical method, each observation starts with itself as a cluster, and clusters are successively merged together to form larger clusters. The algorithm recursively produces a nested series of partitions. The partitional method, on the other hand, produces all the partitions at the same time without imposing the hierarchical structures. There are 5 major approaches including squared error, density-based, model-based, graph theoretic, and mode seeking (Jain et al., 1999; Jain, 2010).

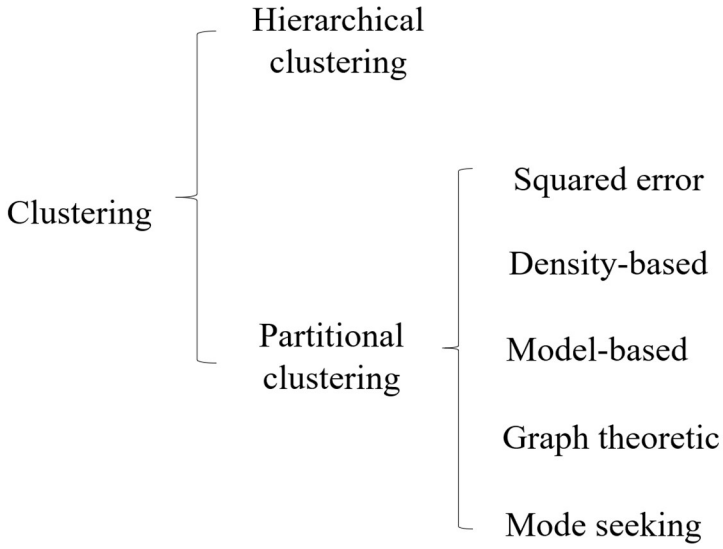

Figure 2.11 Taxonomy of clustering approaches.

Most hierarchical algorithms are variants of the single linkage (Sneath and Sokal, 1973), complete linkage (King, 1967) and minimum-variance or Ward linkage (Ward, 1963; Murtagh, 1983). The difference in between the three methods are the definition of cluster distance. In simple linkage, the distance between the two clusters is the minimum

of the distances between all pairs of observations drawn from the two clusters. In complete linkage, the distances are the maximum distances of all pairwise observation distances. In Ward linkage, the distance is the increase in the sum of cluster distances when the new observation is added to the cluster.

For partitional algorithms, the most intuitive and frequently used criterion function is the squared error criterion (Jain et al., 1999). The expression of the squared error criterion for a clustering α of a data set A is:

$$e^2(\alpha, A) = \sum_{j=1}^{K} \sum_{i=1}^{n_j} \left\| \boldsymbol{x}_i^{(j)} - \boldsymbol{c}_j \right\|^2, (2.31)$$

Where $\boldsymbol{x}_i^{(j)}$ is the i^{th} data point belonging to the j^{th} cluster, $\boldsymbol{c}_j$ is the centroid of the j^{th} cluster. K is the total cluster number. n_j is the total number of points for j^{th} cluster.

The most popular and simplest algorithm for squared error criterion is k-means (Steinhaus, 1956; Ball and Hall, 1965; MacQueen, 1967). The k-means algorithm starts by randomly choosing k observations as cluster centers. Observations are assigned to the closest cluster centers. Then, the mean of observations within each cluster becomes the new cluster center. This process is repeated until some convergence criteria are met (e.g. no new reassignment of cluster labels).

One type of partitional methods is the density-based clustering. In density-based clustering, clusters are defined as the high-density regions in the feature space separated by low-density regions. It can be used for detecting clustering with an arbitrary geometry. One widely used method is density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996).

Another type of clustering is called model-based clustering. It assumes that data is generated from a mixture of components, where each component is described by one probability distribution. Often the distribution is assumed to be Gaussian (Day, 1969; McLachlan et al., 1999; Fraley and Raftery, 2006), so the method is called the Gaussian mixture model (GMM). Model parameters are solved iteratively by the expectation-maximization algorithm (Dempster et al., 1977; Neal and Hinton, 1998; McLachlan and Krishnan, 2007).

Graph theoretic clustering is one type of clustering method. One popular such algorithm is called spectral clustering (Shi and Malik, 2000; Stella and Shi, 2003). It builds a weighted graph in which nodes correspond to observations and edges are related to the distance between the observations (Jain 1999, Ng et al., 2002). It separates the graph into exactly two parts and recursively finds k clusters (Ng et al., 2002).

The last branch that is covered in this article is mode seeking. The mode is the local maxima of probability density functions (Sasaki et al., 2017). One famous method, mean shift, makes use of modes of the estimated density function for clustering (Fukunaga and Hostetler, 1975; Cheng, 1995). The mean shift method starts by regarding all data samples as potential cluster centers. It interactively updates them toward the nearest modes. The observations that converge to the same mode are assigned the cluster labels (Sasaki et al., 2017).

Chapter 3: Porosity-deformation Relationships in Organic-rich Shale[1]

3.1 INTRODUCTION

Shales exhibit a wide range of textures, compositions, and mechanical properties (Loucks et al., 2009; Loucks et al., 2012; Pommer and Milliken, 2015; Eliyahu et al., 2015; Emmanuel et al., 2016). Diagenesis controls much of this variation through burial, compaction, cementation, and thermal maturation. As the original porosity in the sediment collapses with burial (e.g. Velde, 1996), some pores can be preserved within both detrital and diagenetically produced or altered grains (e.g. Desbois et al., 2009). Intragranular porosity is particularly important in organic matter (OM) (Loucks et al., 2009; Loucks et al., 2012; Pommer and Milliken, 2015), as opposed to intergranular pores between grains in the surrounding matrix of clay, cement, and other materials (e.g. Schneider et al., 2011). These pore systems can be related to one another, because the fine pore network and connectivity can be dependent not just on the abundance of OM, but also on its distribution (Loucks and Reed, 2014). Here, we explore how deformation that occurs during unconventional production via hydraulic fracturing can have a varying effect on porosity due to these diagenetically produced textural variations.

Diagenetic processes directly lead to the properties of shale, in turn impacting production efforts at the field scale. During hydraulic fracturing, a network of highly conductive fractures enhances hydrocarbon transport to the wellbore (Nolte, 2000; Arthur et al., 2009). The fluid injection brings the rock volume to shear failure, causing micro-

---

1 This chapter is based on: Jiang, H., H. Daigle, N. W. Hayman, and K. L. Milliken. “Porosity-deformation relationships in organic-rich shale”, AAPG memoir (2018a). (in press)

Daigle, H., N. W. Hayman, E. D. Kelly, K. L. Milliken, and H. Jiang. "Fracture capture of organic pores in shales." *Geophysical Research Letters* 44, no. 5 (2017b): 2167-2176.

seismicity and enhanced permeability (Dusseault, 2011; Maxwell and Cipolla, 2011; Williams-Stroud et al., 2013; Busetti et al., 2014; Roux, 2016). Modeling of field production data suggests that the permeability of the rock lying between the main, induced, meter-spaced fractures is enhanced by a factor of 10-100 (Patzek et al., 2013). It stands to reason that as the main fractures are reactivated, additional damage in the matrix between the fractures could further expand the zone responsible for production, and numerical simulations have indeed shown that the shear deformation may be able to reactivate networks of pre-existing fractures and faults (Johri and Zoback, 2013).

The induced shear deformation may cause microfractures and enhance hydrocarbon transportation from nano-scale organic matter pores to the larger tensile fracture. Characterization the shale pore system at nanoscale, however, remains challenging. Shale commonly contains a large proportion of micropores (< 2 nm) and mesopores (2-50 nm) (Loucks et al., 2009; Loucks et al., 2012). Some clay pores are even less than 1 nm (Kuila and Prasad, 2013b), which is below the resolution of imaging techniques. Gas sorption, especially N_2 and CO_2 sorption, can quantitatively characterize micropore and mesopore structure in shale (e.g. Bustin et al., 2008; Ross and Bustin, 2009; Adesida et al. 2011; Chalmers et al., 2012; Kuila and Prasad 2013a, 2013b; Clarkson et al., 2013; Wang et al., 2014). The main information we can obtain from gas sorption is porosity, pore size distribution, and pore structure (Kruk and Jaroniec 2001; Kuila and Prasad 2013a, 2013b).

In this study, we experimentally introduced shear deformation on shale samples using confined compressive strength tests. Gas sorption was used to characterize pore structure before and after failure at the nanometer (nm) scale as well as scanning electron microscope (SEM) imaging. We used data from multiple aliquots of each sample to reduce bias of results due to sample heterogeneity. Key parameters of pore morphology including pore size distribution (PSD), BET surface area, and surface fractal dimension were

calculated from sorption measurements. When complemented by (SEM) imaging, this allows the more comprehensive study of deformation changes associated with shale nano-scale pore structure. Furthermore, effects including bedding, total clay and cementation, shape and distribution of organic matter were discussed.

3.2 DIAGENETIC HISTORY

The focus of this study is on the mechanisms for nanoscale porosity change due to field-scale hydraulically induced fracture by considering the micromechanics of failure. The distinct depositional setting, characteristic grain assemblages, and systematic contrast in organic matter content of the two shale units we focus on here impact the evolution of porosity and mechanical rock properties (Milliken, 2014). By contrasting two very different shale types (see Table 3.1 for X-ray diffraction (XRD) data), we aim to resolve the role of diagenetic components in the deformation.

The Eagle Ford Formation, in southern Texas, is a coccolithic, organic-rich shale that contains a mixture of mineral- and OM-hosted pores of both primary and secondary origins (Pommer and Milliken, 2015). Destruction of primary porosity and generation of secondary porosity in high-maturity samples is controlled by the relative physical and chemical stabilities of the grain assemblage and early diagenetic components as they undergo later diagenesis burial diagenetic processes, including abundant microquartz cement (Milliken et al., 2016). Pommer and Milliken (2015) reported the maturity of the samples by vitrinite reflectance ranges from low maturity (0.5% Ro) to high maturity (1.3% Ro). The total organic carbon (TOC) has a wide range from 1.2-13.6 wt.%.

In contrast, the samples from the northern Rocky Mountains are considerably more quartz-rich (40-60 wt.%). In general, the quartz in siliceous shales takes several forms: extrabasinal detrital silt (some with an earlier diagenetic history as exemplified by

transported pre-transport overgrowths), replacement of skeletal debris, minor overgrowths on detrital quartz and replaced radiolaria, pore-filling in the intragranular pores of allochems, and authigenic microquartz that is dispersed through the clay-size matrix (Milliken and Olson, 2017). The overall porosity decline is compaction-dominated, despite the exceptional abundance of cement. Milliken and Olson (2017) proposed that the presence of significant volumes of cement would lead to brittle behavior where cements are most prominently developed. The maturity of shale samples by vitrinite reflectance ranges from 0.72% - 1.1% Ro, and the total organic carbon (TOC) ranges from 1.55 to 3.75 wt.% (Milliken and Olson, 2017). We return to this topic of cementation and embrittlement in the Discussion section.

3.3 METHODS

The sampling and experimental workflow is shown in Figure 3.1. Samples were from an organic shale in the northern Rocky Mountains (NoRM) and the Eagle Ford shale (EF). We drilled core plugs parallel and perpendicular to the bedding planes. Samples were deformed by confined compressive strength tests. After the tests, failed samples drilled parallel to bedding are referred to as horizontally failed (HFail), while failed samples drilled perpendicular to bedding are referred to as vertically failed (VFail). Imaging and sorption measurements were performed on intact and shear failed samples. Detailed procedures are described below.

3.3.1 Samples

A total of 8 organic-rich shale cores were used in this study, with 3 samples (EF 1_223, EF 2_50, EF 2_93) from the Eagle Ford unit (hereinafter referred as EF shale), and 5 samples (NoRM 3_14, NoRM 3_42, NoRM 3_53, NoRM 4_14, NoRM 4_34) from the

northern Rocky Mountains, referred to as NoRM shale. The location of the NoRM shale is withheld by the donor though the samples are similar to those described by Milliken and Olson (2017). All samples were preserved in mineral oil until experimentation. We found no sign of mineral oil imbibing into the samples through nuclear magnetic resonance measurements (Daigle et al., 2017b).

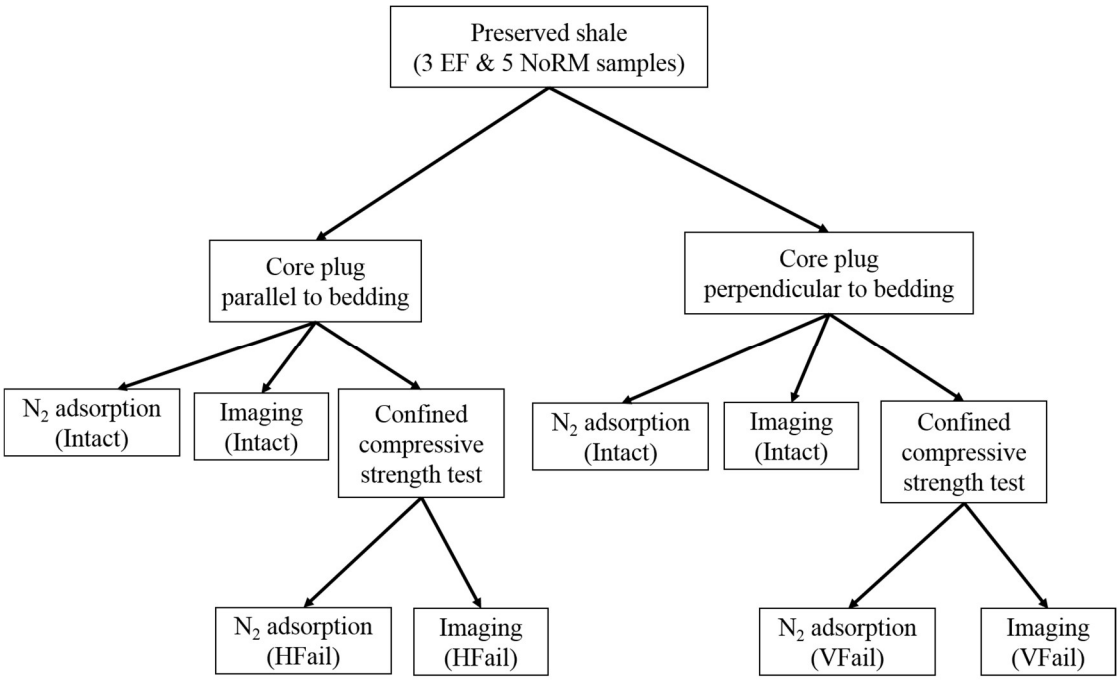

Figure 3.1 Sampling and experimental workflow. NoRM refers to the shales from the northern Rocky Mountains. EF refers to the shales from Eagle Ford. Failed samples drilled parallel to bedding are referred to as horizontally failed ('HFail'). Failed samples drilled perpendicular to bedding are referred to as vertically failed ('VFail'). Samples without confined compressive strength tests are referred to as 'Intact'.

3.3.2 XRD and Rock-eval

The mineralogy of the shale was analyzed using XRD by Weatherford Laboratories in Houston, Texas. The data are listed in Table 3.1. Following Milliken (2014), the total clay content and total cement content are estimated by the following equations:

$$Total\ clay\ =\ Illite/Mica\ +\ Mx\ I/S\ , (3.1)$$

$$Total\ cement =\ calcite\ *\ 0.5\ +\ quart *\ 0.85, (3.2)$$

where Mx I/S is mixed-layer illite/smectite.

Assumptions for the above equations are that 50% of calcite takes the form of cement, which is a common assumption in carbonate-rich rocks (Bathurst, 1972); and authigenic quartz has recently been reported to be 85% (or more) of total quartz in both the EF and NoRM shales (see Milliken et al., 2016; Milliken and Olson, 2017).

Table 3.1 Summary of XRD results in weight percent. Mx I/S is mixed-layer illite/smectite. The total clay = Illite/Mica + Mx I/S. The total cement = calcite * 0.5 + quartz * 0.85. Samples from the Eagle Ford formation have names starting with 'EF'. Samples from the northern Rocky Mountains formation have names starting with 'NoRM'.

Sample	Illite/ Mica	Mx I/S	Calcite	Quartz	Total clay	Total cement
EF 1_223	11	8	62	13	19	42
EF 2_50	19	21	40	13	40	31
EF 2_93	6	7	70	10	13	44
NoRM 3_14	19	19	3	41	38	36
NoRM 3_42	13	10	1	55	23	47
NoRM 3_53	9	11	1	54	20	46
NoRM 4_14	10	9	0	63	19	54
NoRM 4_34	19	13	0	52	32	44

XRD shows that calcite dominates the EF shale samples, whereas quartz dominates the NoRM samples. Clay minerals (illite/smectite and illite/mica) are present in significant amounts in both lithologies, whereas the quartz and calcite contents differ quite dramatically between the two. The bulk mineralogy is generally consistent with that found from previous research on EF shale (Pommer and Milliken, 2015) and NoRM shale (Milliken and Olson, 2017).

Rock-eval analysis was conducted by Weatherford Laboratories in Houston, Texas. According to the results (Table 3.2), the total organic carbon (TOC) content of the Eagle Ford samples varies from 3.12 to 4.73 wt.%, while the TOC of NoRM samples has a wider variation, from 2.82 to 4.56 wt.%.

Table 3.2 Summary of rock-eval results. TOC = Total Organic Content wt.%; S_1 = volatile hydrocarbon content, mg/g; S_2 = remaining hydrocarbon generative potential, mg/g; S_3 = carbon dioxide content, mg/g; HI = hydrogen index; OI = oxygen index.

Sample	TOC (%)	S_1	S_2	S_3	Tmax (°C)	HI	OI
EF 1_223	3.12	8.69	4.83	0.46	445	155	15
EF 2_50	3.79	7.26	5.44	0.45	444	144	12
EF 2_93	4.73	7.83	4.92	0.41	448	104	9
NoRM 3_14	4.40	4.54	2.6	0.37	454	59	8
NoRM 3_42	4.56	5.02	2.72	0.43	460	60	9
NoRM 3_53	3.15	2.42	1.56	0.25	452	50	8
NoRM 4_14	3.64	5.18	2.36	0.21	459	65	6
NoRM 4_34	2.82	2.58	1.57	0.24	457	56	9

3.3.3 Confined Compressive Strength Test

The received shale cores were subsampled using a low-rate coring machine with mineral oil as a lubricating fluid. One core plug was drilled parallel to the bedding planes, and another one was drilled perpendicular to the bedding planes (Figure 3.2a). Each plug

was 2.54 cm (1”) in diameter and 5-7 cm in length. They were preserved in light mineral oil until experimentation.

Shear failure was induced by subjecting samples to confined compressive strength tests (Figure 3.2b). The core plug was wrapped in a thermo-shrinkable sleeve before being loaded into the testing cell. During the test, the confining stress increased to 10 MPa over the course of 1 min. The axial stress was increased by displacing the axial ram at a rate of 0.01" per minute. The test was completed when sample failure occurred.

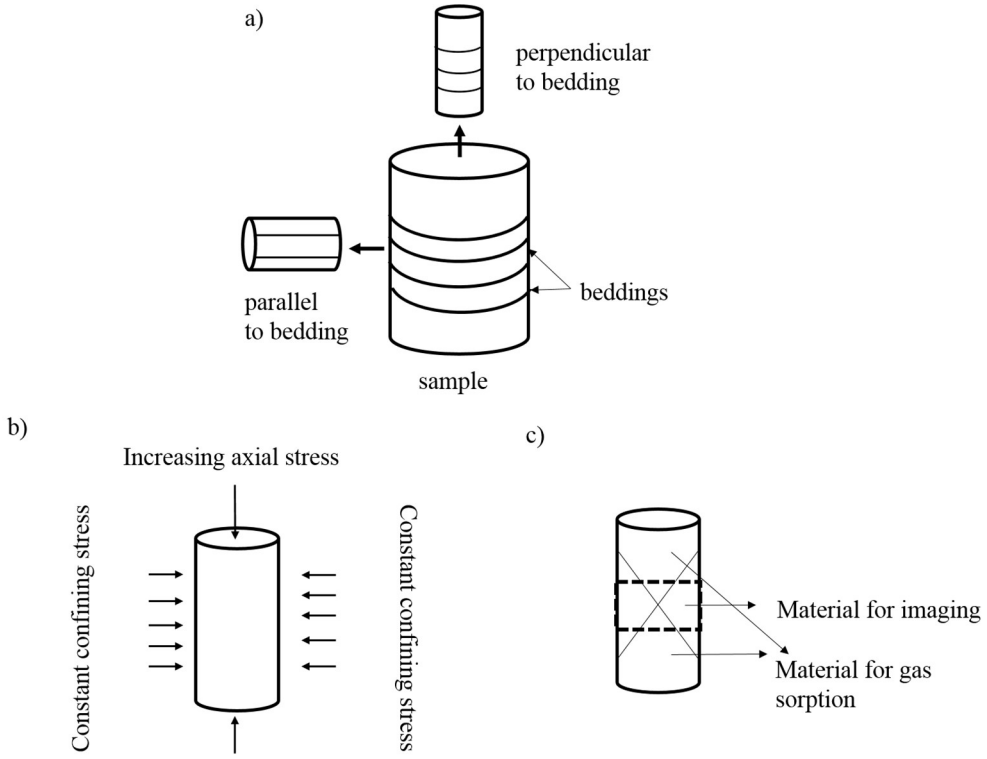

Figure 3.2 (a) Illustraion of coring. One core plug is drilled parallel to the bedding planes, and another one is drilled perpendicular to the bedding planes. (b) Stress conditions during confined compressive strength tests. (c) Illustration of where imaging and gas sorption material was taken from the failed samples. We took the material from the main fractures plane for images and gas sorption.

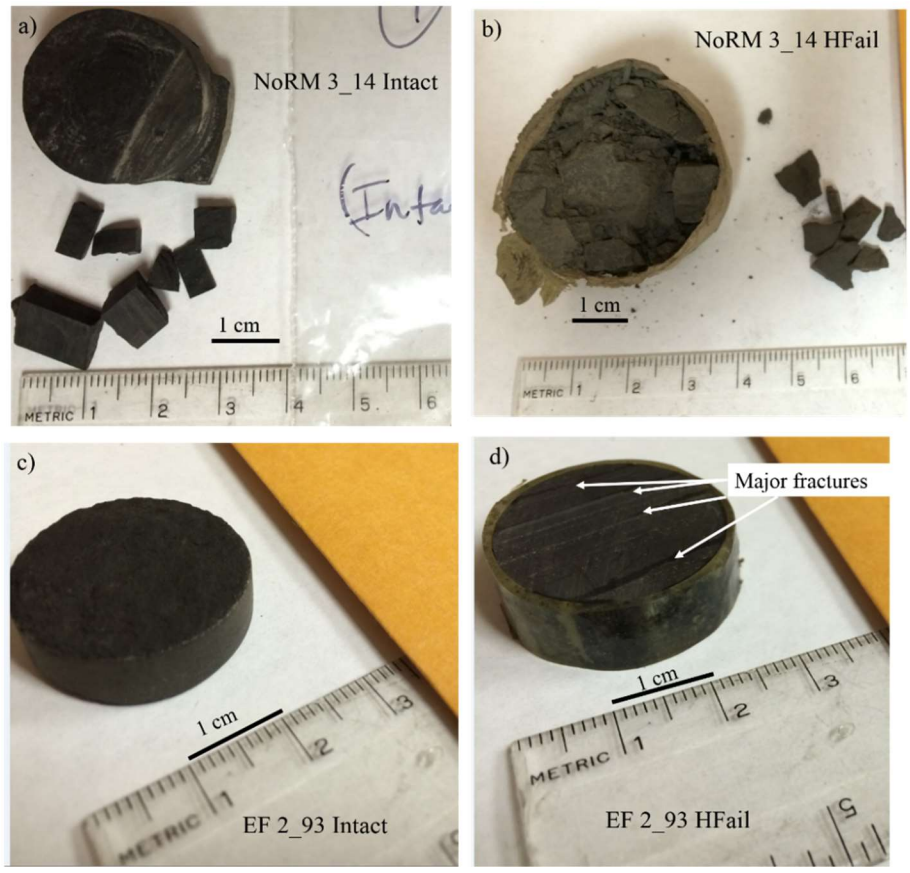

Figure 3.3 Illustraion of SEM material of (a) intact NoRM sample, (b) failed NoRM sample, (c) intact EF sample, (d) failed EF sample. The subsamples are chosen to be on or near the major fractures. The orientation of subsample preparation provided views of the bedding planes ('Bed') or cross-sectional views ('X section').

3.3.4 SEM Imaging

For imaging, 46 subsamples (Figure 3.2c, Figure 3.3) were selected from 8 samples of the NoRM and Eagle Ford shales (Daigle et al., 2017b). These samples reflect deformed and intact counterparts. The orientation of subsample preparation provided views of the bedding planes ('Bed') or cross-sectional views ('X section'), from cores either 'horizontally failed' or 'vertically failed' with respect to bedding. The SEM images for failed samples were chosen to be on or close to the major fractures as shown in Figure 3.3.

Imaging was performed using a FEI Nova Nano SEM 430 FE-SEM (field-emission scanning electron microscope). Prior to the confined compressive strength tests, about 1 cm of material was removed from the bottom of the core plugs for imaging. After the test, a 1 cm slice of material was removed from the middle of each core plug for imaging. Subsamples were prepared in-house via ion milling (Milliken et al., 2013) and Ir coating (4-5 nm thickness). Both backscattered electron and x-ray elemental maps via EDS (energy dispersive x-ray spectroscopy) were collected.

3.3.5 Gas Sorption

Low pressure N_2 and CO_2 sorption measurements were conducted using a Micrometitics 3Flex surface analyzer. We performed measurements on multiple aliquots of materials of each sample to ensure repeatability and mitigate the effects of cm-scale heterogeneity. Sample material was collected from the core plugs before and after the confined compressive strength tests. A total of 8 HFail samples and 2 VFail samples (EF 1_223 and NoRM 3_53) were used for the measurements. Samples were oven dried at 110°C for 24 hours and hand crushed to less than 40 US mesh (0.42 mm). Approximately 1-1.5 g of crushed sample was used for N_2 gas sorption at 77 K and CO_2 sorption at 273.15 K.

Given the area covered by each adsorbed molecule, the surface area of the solid surface thus can be calculated. The method used to calculate the surface area is called Brunauer-Emmett-Teller (BET) method, which incorporates multilayer coverage during adsorption (Brunauer et al., 1938; Yang et al., 2014). The pore size distribution (PSD) represents the pore volume abundance of each pore size in the sample. Here, we computed PSDs by nonlocal density functional theory (NLDFT; Roque-Malherbe, 2007; Adesida et

al., 2011) with slit-shaped carbon pores (Tarazona, 1985; Tarazona and Vicente, 1985). More details on gas sorption theory can be referred to Chapter 2.3.

Integrating the N_2 and CO_2 pore size distributions, we calculated the meso-/macropore volume (pore diameter > 2 nm) and the micropore volume (pore diameter < 2 nm), according to the classification of International Union of Pure and Applied Chemistry (IUPAC; Sing, 1985).

3.3.6 Surface Fractal Dimension

Shale is a multi-scale, heterogeneous material with a complex pore structure. As such, it is difficult to describe the geometry of the solid surface. Fractal theory provides a powerful tool to characterize heterogeneous media like shale. We computed two fractal dimensions based on N_2 adsorption data (D_1 and D_2). D_1 is the fractal dimension calculated from the N_2 adsorption isotherm with relative pressure < 0.45, and D_2 is the dimension from the isotherm with relative pressure > 0.45. Detailed description about fractal theory and surface fractal dimension are in the following section.

3.3.6.1 Overview of Fractal Dimension

To describe the fractal dimension, we first introduce the concept in a simple fashion by using an example of a regular 1-D line (Figure 3.4). The line initially has unit length. If we magnify the line by a factor of 2 (the magnification factor r), the line becomes two units long. The total number (N) of unit length lines is also 2.

The dimension D for a regular l-D line is:

$$D = \frac{log(N)}{log(r)} = \frac{log(2)}{log(2)} = 1 \quad (3.3)$$

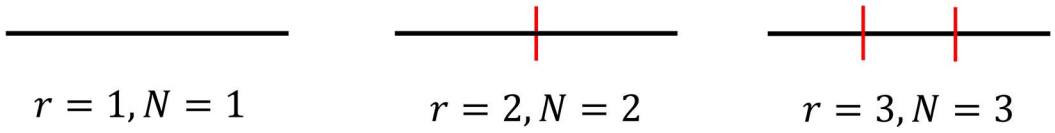

Figure 3.4 Example of a regular 1-D line. A line may be broken into N self-similar sub-lines, each with magnification factor r. For a normal 1-D line, r and N are the same.

Using this formula, we can calculate the fractal dimension of a Koch curve, a mathematical fractal curve (Addison, 1997). As shown in Figure 3.5, when the line length increases from unit length to 3 unit lengths (r = 3), the total number of unit lines becomes 4 instead of 3.

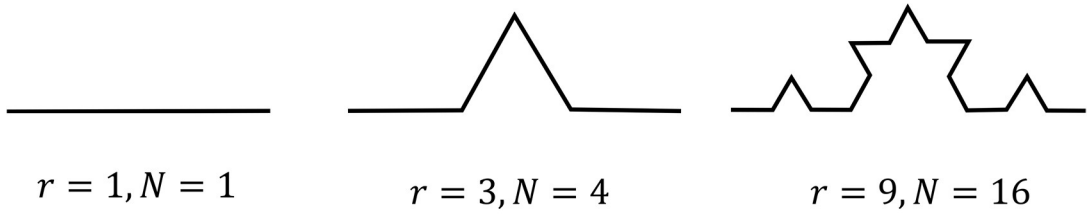

Figure 3.5 Illustration of fractal theory using a 1-D Koch curve. r is the magnification factor. N is the total number of unit length lines. Note: The fractal dimension of the line given r = 9 is also equal to 1.26.

The fractal dimension D is:

$$D = \frac{log(N)}{log(r)} = \frac{log(4)}{log(3)} = \frac{log(16)}{log(9)} = 1.26\,. \quad (3.4)$$

3.3.6.2 Surface Fractal Dimension from N_2 Sorption

One widely used fractal model is the surface fractal, where surfaces or boundaries separating mass and pore spaces are fractal. Surface fractal are measured by surface fractal dimension D. For simplicity, we use the notation D for surface fractal dimension. The value of the surface fractal dimension varies from 2 to 3. A value of 2 indicates a smooth surface, whereas 3 indicates an extremely rough surface.

N_2 adsorption data can be used to compute the surface fractal dimension. The method used for this purpose is called Frenkel-Halsey-Hill (FHH) method (Frenkel, 1946; Hill, 1946; Halsey, 1948), and it is the most effective and widely used model for evaluating surface fractal dimension from gas adsorption data (e.g. Yang et al., 2014; Liu et al., 2015; Jiang et al., 2016).

The surface fractal dimension D can be determined as

$$\frac{V}{V_m} \propto [\, RT\, ln\left(\frac{P_0}{P}\right)]^{k}\,, (3.5)$$

where V is the volume of adsorbed gas molecules at equilibrium pressure P, V_m is the volume of gas molecules in a monolayer, R is the universal gas constant, T is the absolute temperature when the isotherm is obtained, and P_0 is the saturated vapor pressure of nitrogen at temperature T (Sokołowska et al., 2001; Yang et al., 2014). k is a coefficient related to the surface fractal dimension D as $D = k + 3$ (Jaroniec et al., 1997).

Incorporating this relationship between D and k, Equation 3.5 can be written in log-log form:

$$ln(V) = (D - 3)\, ln(ln(P_0/P)) + C\,, (3.6)$$

where C is an additional constant that accounts for the amount of adsorbed volume when the fractal regime is first reached (Jiang et al., 2016). The value of D is obtained by applying a linear regression for regression for $ln(V)$ versus $ln(ln(P_0/P))$.

In this work, we computed two fractal dimensions based on N_2 adsorption data (D_1 and D_2). D_1 is the fractal dimension calculated from N_2 adsorption isotherm with relative pressure < 0.45, and D_2 is the dimension from the isotherm with relative pressure > 0.45. Previous studies showed that the two fractal dimensions have different values, which is

probably due to different gas sorption mechanisms in these two regions (Wang et al., 2015; Jiang et al., 2016).

3.3.7 About Experiments and Measurements

The limitation for SEM images is that only a local region is chosen for imaging, which is ususally around tens of micrometers in scale. It therefore may not be representative of the entire core plug. There is also a lack of information about the third dimension of the pore structure. The limitation of pore sizes determined fromgas sorption is that it is an indirect method, and all parameters are calculated based on physical models, which may not represent natural rocks like shale with chemically heterogeneous pore surfaces. The pore size distribution is computed using an inversion technique (Appendix A), and may not represent the actual pore size distribution due to errors from inversion. In addition, there might still be disconnected pore networks which were not accessed by the probing gas, although we crushed the rock.

In any laboratory study of fracturing behavior, there is always the question of which fractures are induced experimentally, which are generated during core retrieval and handling, and which are present in situ in the subsurface. Specifically, the hand crushing procedure for the gas sorption measurements may open new fractures. Though we cannot rule out that some of the fractures were caused by the sampling process, we note that there were quantifiable differences between pre- and post-failure porosity that was correlated with lithology, and hence diagenetic history. Since we applied the same handling procedures for porosity measurements of intact and failed rock, the uncertainty due to artifacts from that particular measure is assumed to be minimized.

3.4 RESULTS

3.4.1 SEM Imaging

SEM images provide the in-situ documentation of the microfracture and pore system in NoRM vs. EF samples, as well as information on the amount of cement and detrital clay minerals, and textural information about porosity distribution (Figures 3.6-3.8). The microstructure of NoRM samples are a mixture of laminated and particulate OM between predominantly quartz grains. On the other hand, the EF samples are dominated by coarser calcite grains, and most organic matter is dispersed throughout the samples.

After failure (Figure 3.7), fractures with widths ranging from 10-100 nm up to 1-2 μm are observed to follow coarser grain boundaries and laminae of OM and matrix materials. In more laminated materials, fracture lengths are up to hundreds of micrometers, which are likely continuous across entire sample volume. Some fractures initiate along grain contacts and primarily propagate through OM.

Though the textural and diagenetic controls on porosity distribution are undoubtedly more complex, our limited observations find that the siliceous NoRM samples in particular have instances where the least cemented sample (e.g. Figure 3.8a, b) have highly porous OM and enhanced porosity within deformed clay aggregates. More cemented samples (e.g. Figure 3.8c, d) have less porous OM and less 'distributed' porosity within the cemented matrix.

Above analysis provides a qualitative assessment of the characteristics of fracture induced by shear deformation. In addition, Tian and Daigle (2018a) proposed an automated fracture detection technique on image based on machine learning, and analyzed 100 SEM images obtained from these same intact and failed samples. Twenty-four of the images were measured from EF samples, and 76 were from NoRM samples. Their results showed

that the failed NoRM samples had slightly more detectable microfractures (2.2 fractures/image) than the intact samples (1.7 fractures/image). The lengths of the fractures in failed NoRM samples tended to be longer than those in intact samples. On the other hand, failed EF samples had a similar number fractures on average compared to intact samples (1.6 fractures/image), and the lengths of the fractures in failed EF samples were smaller than those in intact EF samples. They concluded that the observed fractures were evidence of interaction of preexisting fabric of the samples with the experimentally induced deformation (Tian and Daigle, 2018a)

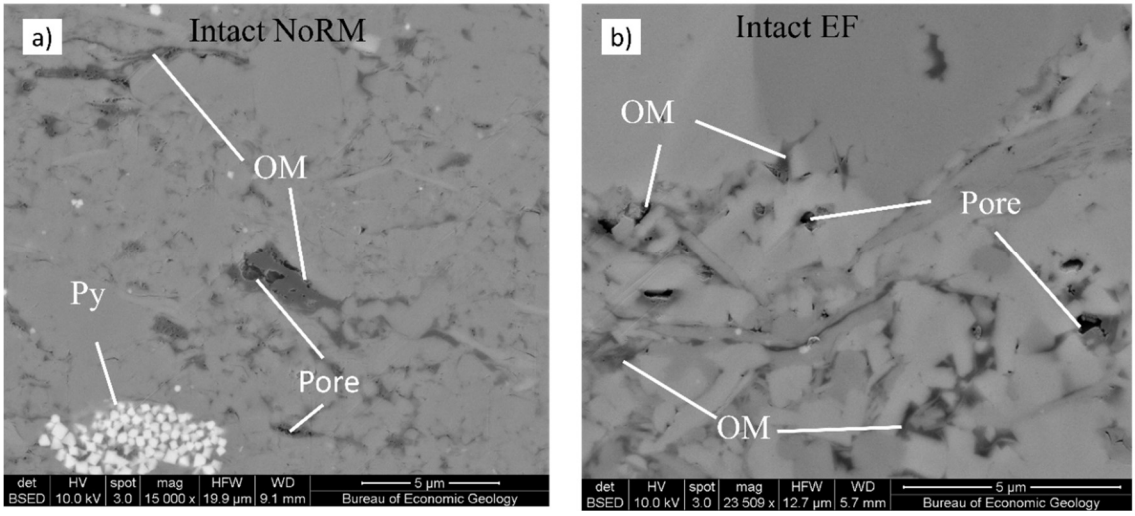

Figure 3.6 SEM images of the intact samples from two shale formations. (a) SEM image of an intact NoRM 3_42 sample. The brightest regions are pyrite framboids. The OM particulates contain a complex pore structure. The darkest regions of the images are mostly organic matter. The brightest regions are pyrite (Py), while pores appear black. (b) SEM image of an intact EF sample (EF 1_223). Pores (black) occur within organic matter (OM) dispersed within calcareous matrix of predominately coocolith fragments (white). NoRM refers to the shales from the northern Rocky Mountains. EF refers to the shales from Eagle Ford. Image IDs are (a) 12-3_42_BP_Xsec_Sample12, (b) 40-1_223_BN_Xsec_40_area5a.

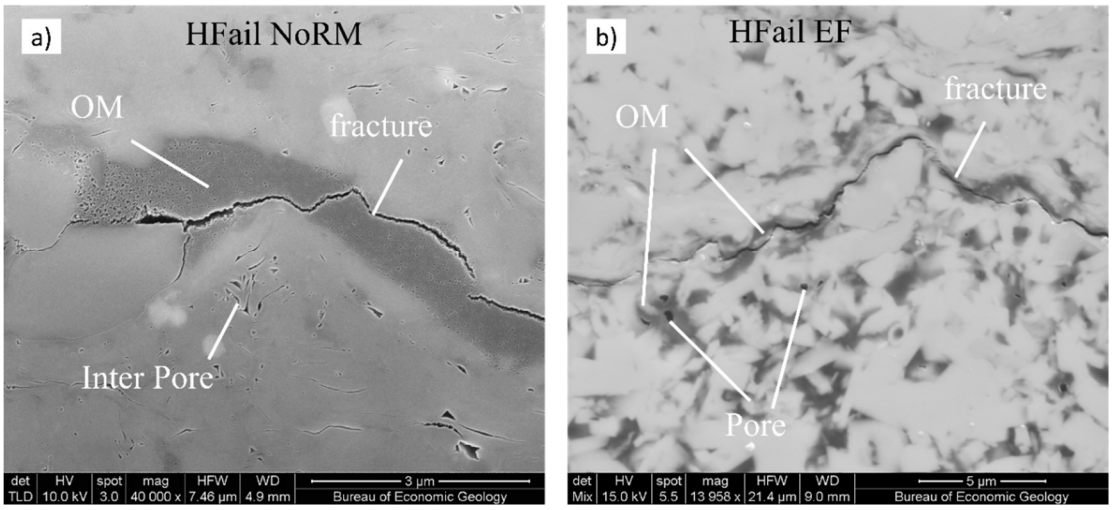

Figure 3.7 SEM images of horizontally failed shale samples. Axial stress is applied normal to image. (a) SEM image of a horizontally failed NoRM 3_14 sample. The fracture intersected pores in OM, but bypassed intergranular pores. (b) SEM image of a horizontally failed EF 2_93 sample. NoRM refers to the shale form the northern Rocky Mountains. EF refers to shale from Eagle Ford. Image IDs are (a) 08-3_14_BP_Xsec_BSE_FEI_8Area1, (b) 34-2_93_BP_Xsec_34_area_2b.

3.4.2 Gas Sorption

Eight horizontally failed samples and two vertically failed (EF 1_223 and NoRM 3_53) samples were characterized using N_2 and CO_2 gas sorption measurements. The calculated BET specific surface areas, N_2 total pore volume, CO_2 total pore volume and two surface fractal dimensions D_1 and D_2 for all samples are presented in Table 3.3.

3.4.2.1 Isotherms and Pore Size Distributions

Gas sorption isotherms are shown in Figure 3.9 and Figure 3.10 for N_2 and CO_2, respectively. Failed and intact samples from the same core are plotted in the same sub-figure. After shear failure, most samples from the two formations show an increase of their sorption quantity for both N_2 and CO_2. Sample NoRM 3_14 has the largest increase in the sorption quantity for N_2, while NoRM 4_14 has the largest increase in the sorption quantity

for CO_2. The vertically failed samples, including samples EF 1_223 and NoRM 3_53, exhibit greater sorption quantities compared with the horizontally failed samples.

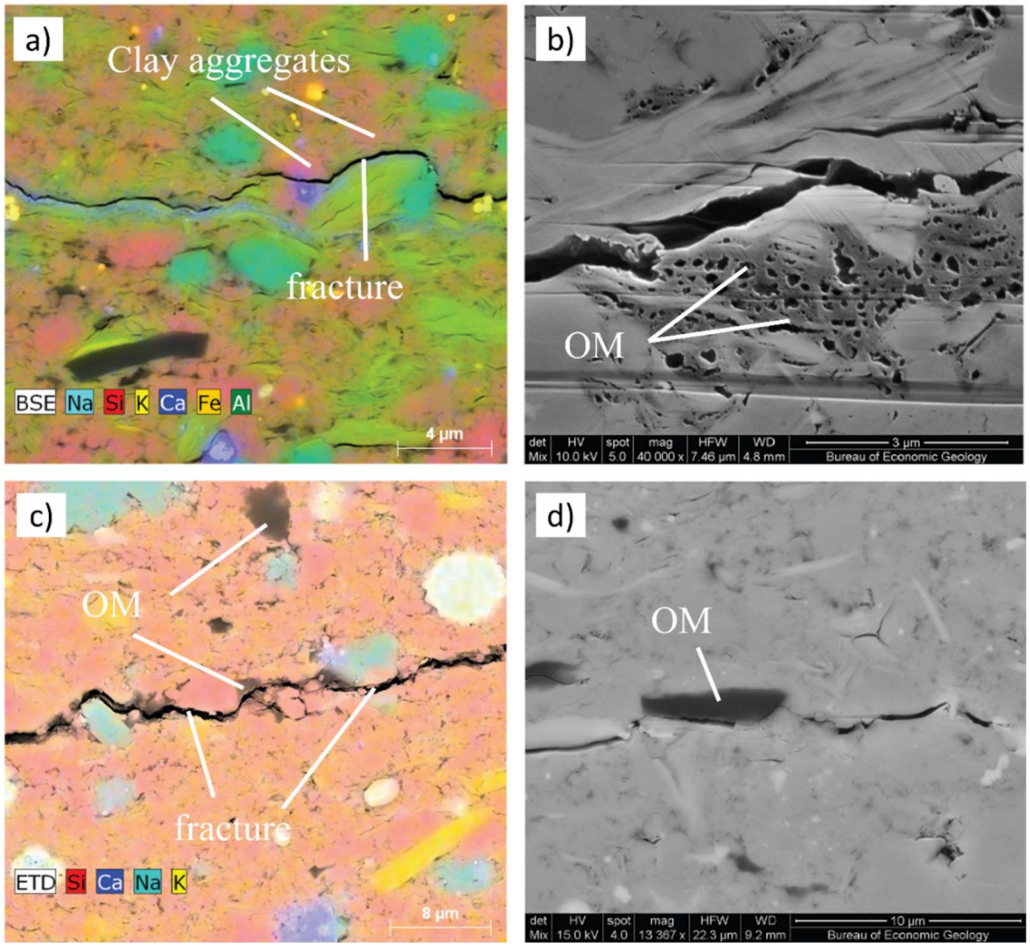

Figure 3.8 More SEM images of horizontally failed siliceous NoRM shale samples. (a) EDS-BSE compositional map of the horizontally failed NoRM 3_14 illustrating the composition (relatively cement poor and clay rich), and cracks both along the central fracture as well as within clay aggregates. (b) SEM image of the vertically failed NoRM 3_14 showing OM pores next to cracked clay aggregates, (c) SEM image of the vertically failed NoRM 3_53, which was relatively cement rich and clay poor, (d) SEM image of the horizontally failed NoRM 4_34 with relatively nonporous OM. Image IDs are (a) 08-3_14_BP_Xsec_EDS_8area3, (b) 10-3_14_BN_Xsec_10, (c) 26-3_53_BN_Xsec_26_Area6, and (d) 38-4_34_BP_Xsec_BSE_38_area2a.

Table 3.3 Summary of pore structural parameter results. 'HF' refers to horizontally failed samples, 'VF' refers to vertically failed samples, and 'In' refers to intact samples. NoRM refers to the shales from the northern Rocky Mountains. EF refers to the shales from Eagle Ford.

Sample	Label	BET surface area (m^2/g)	N_2 pore volume (cm^3/g)	CO_2 pore volume (10^{-3} cm^3/g)	Fractal dimension D_1	Fractal dimension D_2
EF 1_223	HF 1	4.80	0.015	0.72	2.42	2.64
	HF 2	4.80	0.016	0.77	2.41	2.62
	VF 1	5.27	0.015	0.87	2.43	2.63
	VF 2	5.56	0.016	-	2.45	2.64
	In 1	4.32	0.014	0.65	2.40	2.63
	In 2	4.94	0.015	0.85	2.41	2.63
EF 2_50	HF 1	6.19	0.019	0.75	2.41	2.64
	HF 2	5.38	0.017	0.82	2.40	2.64
	In 1	6.22	0.019	0.95	2.40	2.63
	In 1	6.29	0.019	1	2.41	2.63
EF 2_93	HF 1	7.34	0.022	0.86	2.43	2.65
	HF 2	7.98	0.023	0.75	2.41	2.65
	In 1	6.90	0.021	0.81	2.42	2.64
	In 2	6.82	0.021	0.77	2.41	2.64
NoRM 3_14	HF 1	11.25	0.021	1.93	2.51	2.65
	HF 2	11.77	0.021	2.03	2.54	2.64
	In 1	8.06	0.014	1.54	2.53	2.65
	In 2	7.70	0.013	1.57	2.54	2.65
	In 3	8.02	0.013	-	2.55	2.66
	In 4	8.44	0.014	-	2.53	2.65
NoRM 3_42	HF 1	10.44	0.010	1.48	2.65	2.71
	HF 2	8.95	0.010	1.45	2.62	2.70
	In 1	9.80	0.013	1.78	2.61	2.67
	In 2	10.22	0.011	1.77	2.62	2.69
	In 3	9.23	0.011	1.69	2.63	2.68
	In 4	9.02	0.010	1.75	2.64	2.68
NoRM 3_53	HF 1	7.36	0.017	1.56	2.55	2.64
	HF 2	7.24	0.016	-	2.55	2.63
	VF 1	7.79	0.017	1.56	2.54	2.63
	In 1	6.78	0.015	1.4	2.56	2.64
	In 2	6.76	0.015	1.35	2.57	2.64
	In 3	6.69	0.016	-	2.55	2.64

Table 3.3 Continued.

Sample	Label	BET surface area (m^2/g)	N_2 pore volume (cm^3/g)	CO_2 pore volume (10^{-3} cm^3/g)	Fractal dimension D_1	Fractal dimension D_2
NoRM 4_14	HF 1	8.77	0.009	1.68	2.64	2.69
	HF 2	9.39	0.009	1.75	2.63	2.70
	HF 3	8.54	0.010	1.38	2.62	2.71
	In 1	6.91	0.009	1.14	2.59	2.67
	In 2	6.85	0.009	1.08	2.58	2.67
	In 3	6.29	0.009	-	2.56	2.67
NoRM 4_34	HF 1	7.13	0.013	1.73	2.68	2.65
	HF 2	6.91	0.014	1.88	2.66	2.68
	In 1	8.62	0.015	1.95	2.64	2.69
	In 2	8.18	0.015	1.87	2.65	2.68
	In 3	7.26	0.013	1.96	2.66	2.67

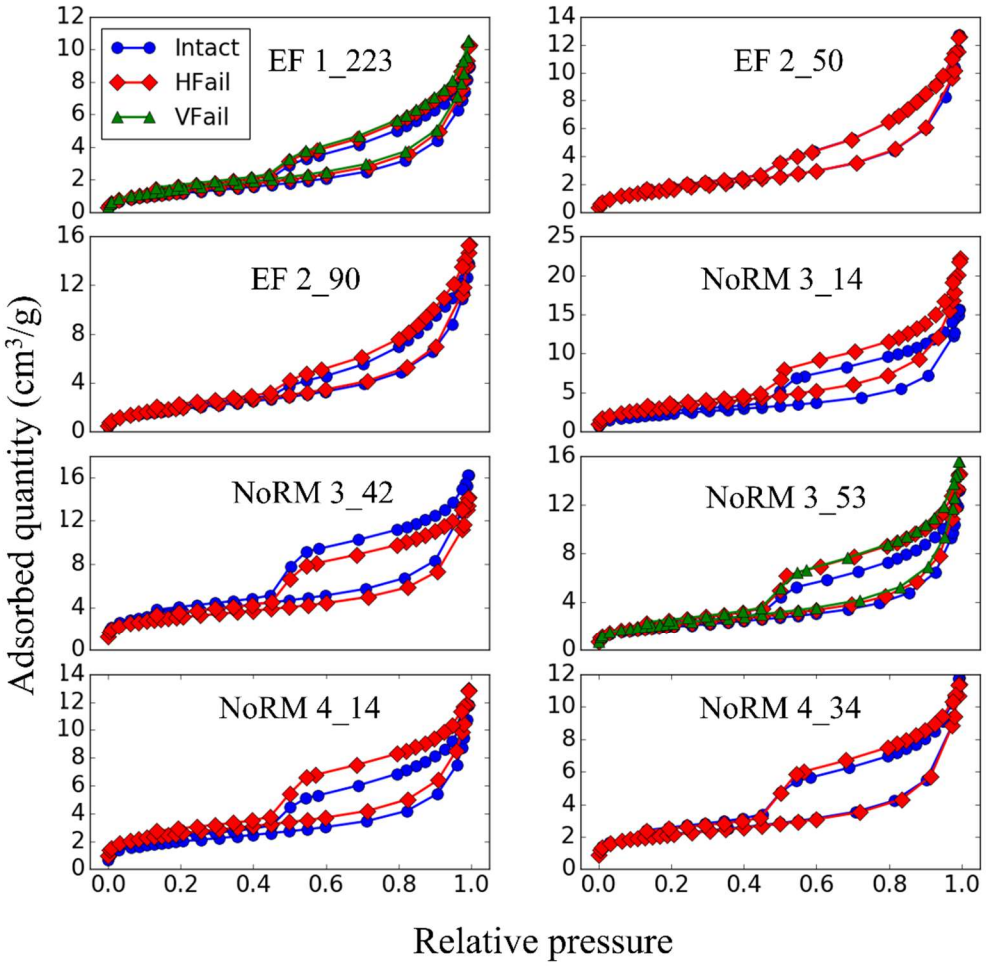

Figure 3.9 Comparisons of N_2 isotherms for failed and intact samples of the calcareous EF and siliceous NoRM shale. EF1_223 and NoRM 3_53 had vertically failed samples. Intact samples are represented by blue circles. Horizontally failed samples are represented by red diamonds. Vertically failed samples are represented by green triangles. Note that only one measurement from intact and failed samples are plotted in the graph.

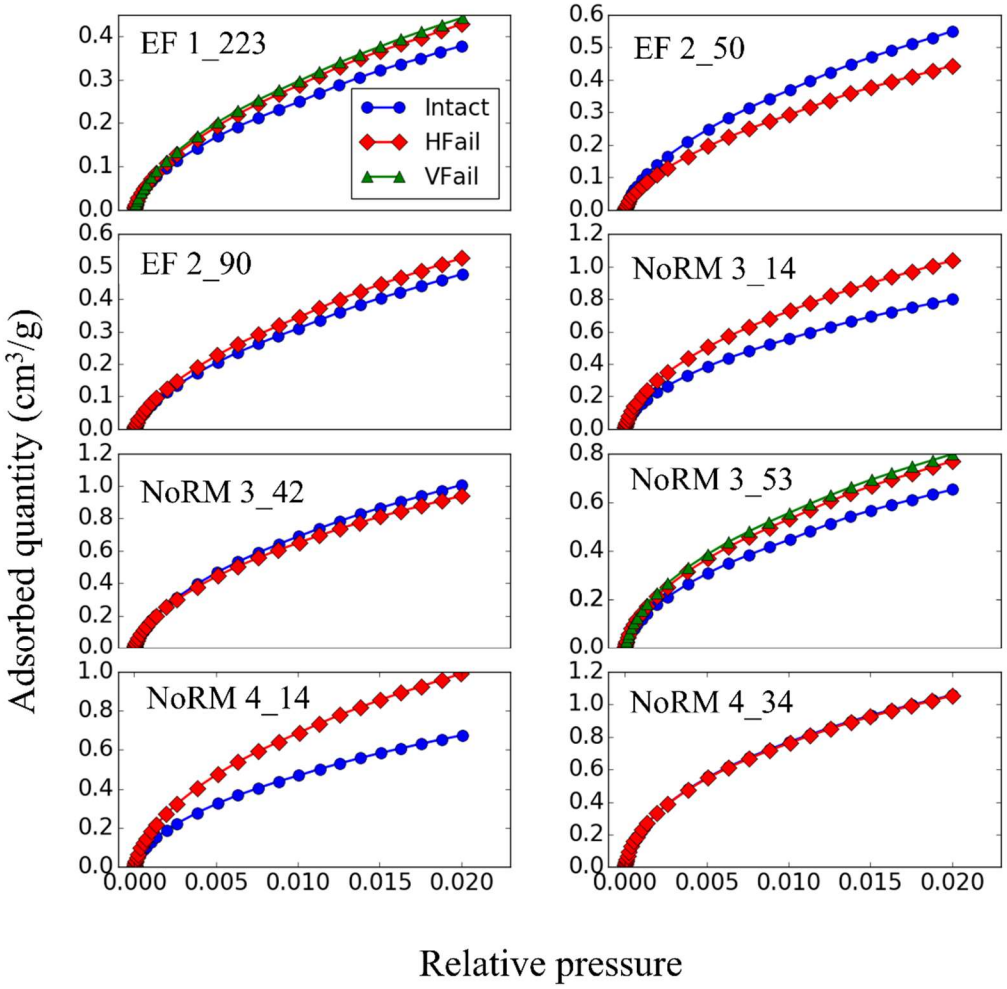

Figure 3.10 Comparisons of CO_2 isotherms for failed and intact samples of the calcareous EF and siliceous NoRM shale. The CO_2 adsorption isotherms are Type I, indicating microporous solids. Intact samples are represented by blue circles. Horizontally failed samples are represented by red diamonds. Vertically failed samples are represented by green triangles. Note that only one measurement from intact and failed samples are plotted in the graph.

Pore size distributions of N_2 and CO_2 are calculated, and N_2 results are shown in Figure 3.11. The N_2 pore size ranges from 1.8 nm to 100 nm. After failure, most of the samples display an increase in pore volume in both the micro- and meso-/macropore range. NoRM samples show a more significant change. The shape of NoRM 3_14 changes after failure. Similar to the isotherm and BET surface area, the vertically failed samples EF 1_223 and NoRM 3_53 have larger increases in pore volume compared to the horizontally failed samples.

3.4.2.2 BET Surface Area and Pore Volume

BET surface area and pore volume are shown in Figure 3.12. Surface area (Figure 3.12a) of intact of EF samples vary from 4.32 to 6.98 m^2/g. The surface areas of the intact NoRM samples have higher values, varying from 6.78 to 12.08 m^2/g. After shear failure, most of the failed rocks display an increase in surface area, especially for 3_14 and 4_14. The vertically failed samples have a larger surface area compared to both horizontally failed and intact samples.

N_2 and CO_2 pore volume are shown in Figure 3.12b and Figure 3.12c. The intact EF shales have higher N_2 pore volumes than the intact NoRM samples, whereas smaller CO_2 pore volumes for those EF shales. After shear failure, most of the failed samples show an increase in both N_2 and CO_2. Sample NoRM 3_14 displays large increases for both volumes.

Furthermore, we calculated the ratio of failed samples and intact samples in terms of meso-/macropore ($\geq$ 2 nm) volume and micropore ($<$ 2 nm) volume (Figure 3.13). The meso-/macropore is the total N_2 pore volume with pore size greater than 2nm. The micropore volume is the sum of both CO_2 pore size and N_2 pore volume with pore size less than 2nm.

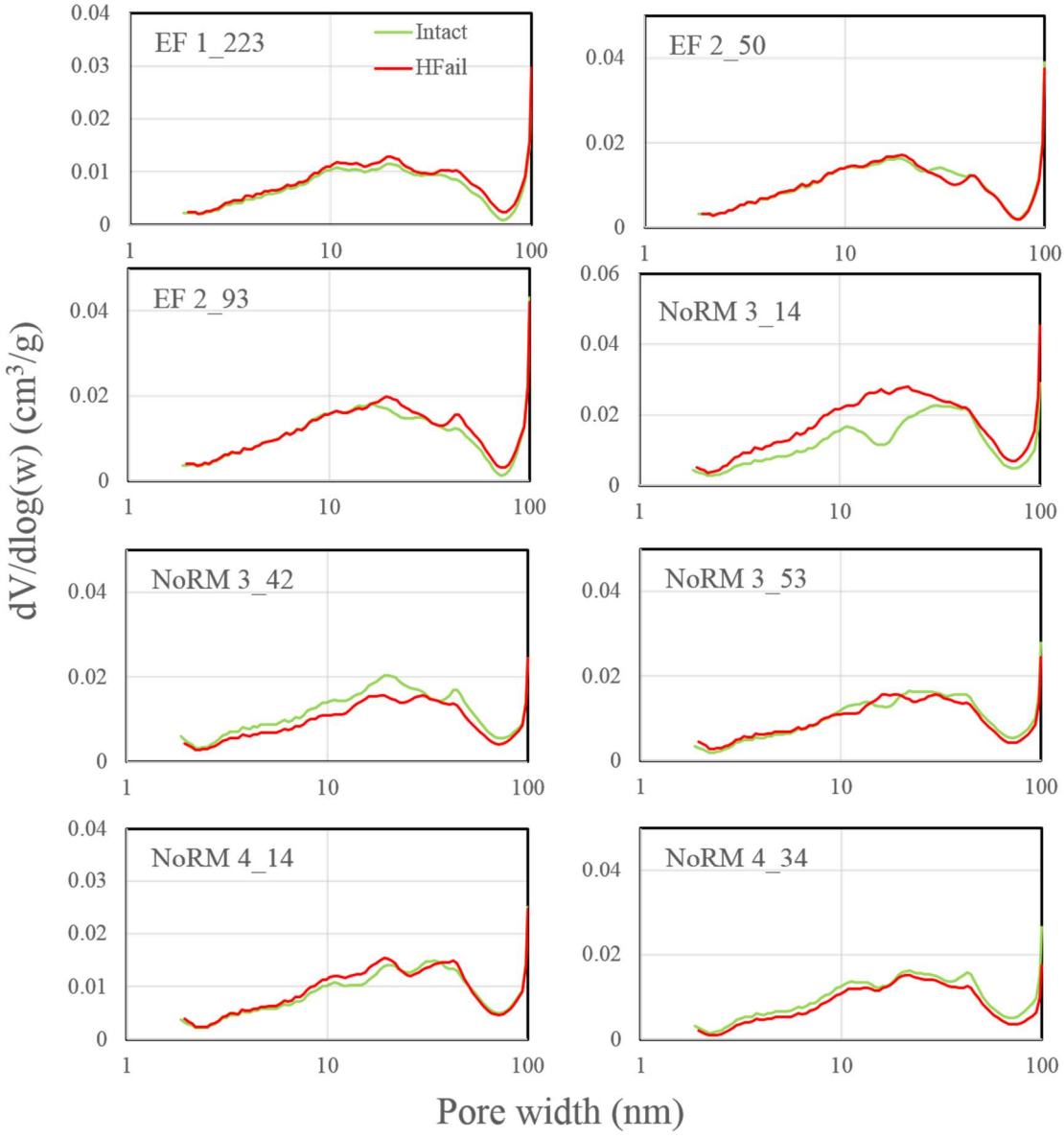

Figure 3.11 N_2 pore size distributions of intact (green) and horizontally failed (red) samples. N_2 Pore size distributions are based on non-local functional theory model using slit-shape pores. The regularization parameter is 1.0. The range of the pore width is from 1.8 to 100 nm. The pore volume is reported in dV/dlog(w), which is the derivative pore volume (V) normalized to natural logarithm of pore width (w).

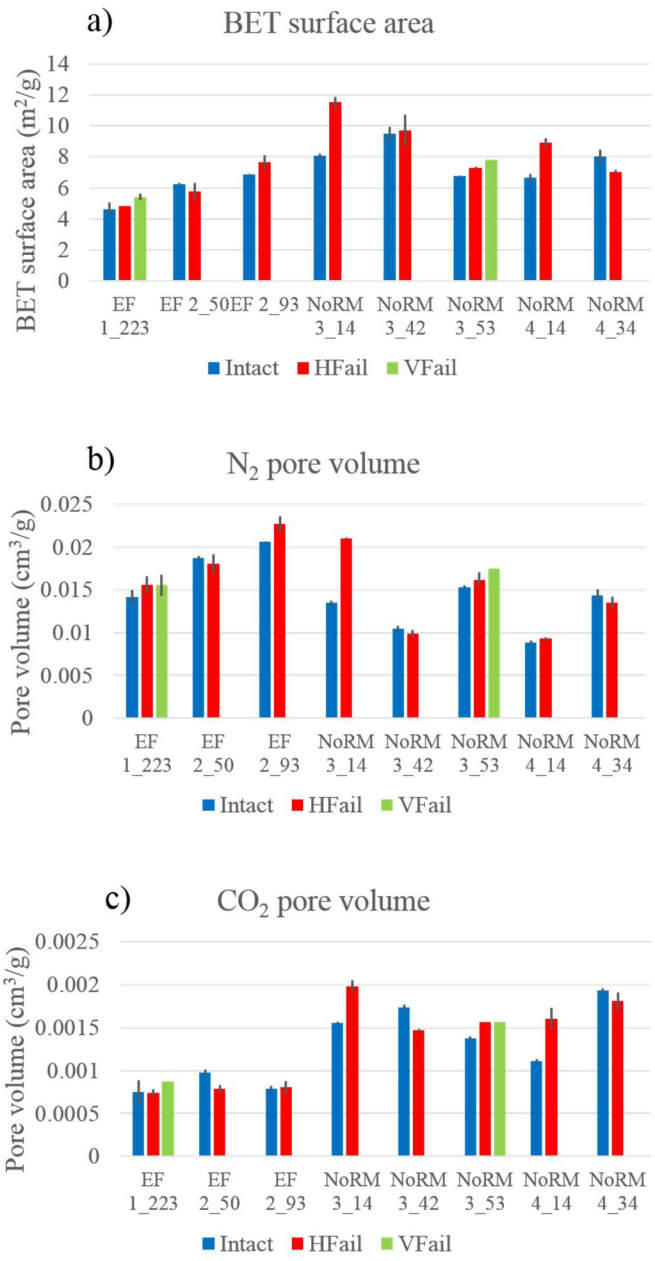

Figure 3.12 (a) BET surface area of intact and failed samples. (b) N_2 pore volume of intact and failed samples. (c) CO_2 pore volume of intact and failed samples. The error bar is shown in the plot as the black line.

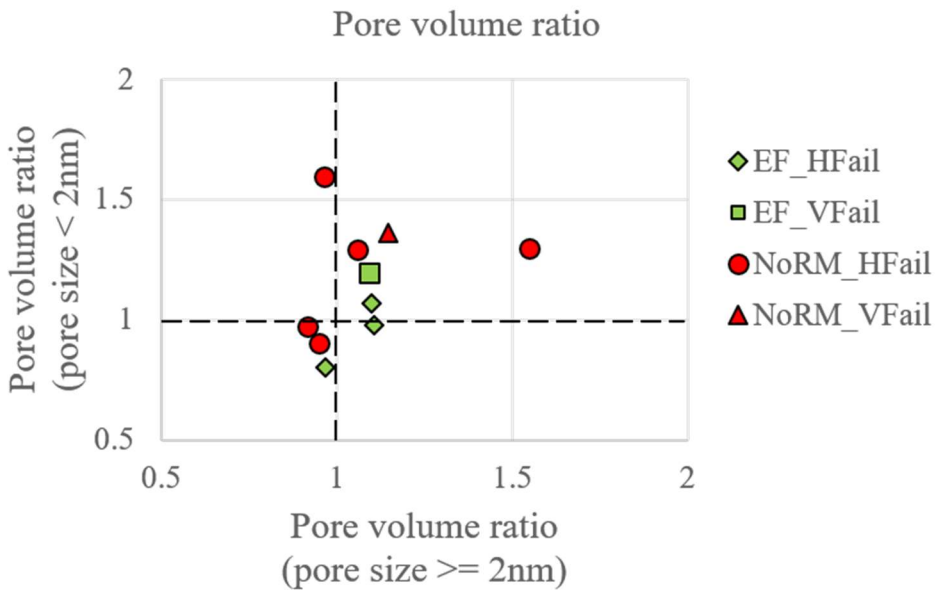

Figure 3.13 Ratio of meso-/macropore (≥ 2 nm) volume and ratio of micropore (< 2 nm) volume of the shale samples. The ratio is the pore volume of failed samples over the pore volume of intact samples. The micropore volume is the sum of both CO_2 pore size and N_2 pore volume with pore size less than 2nm. EF samples are marked in green diamonds (HFail) and a green square (VFail). NoRM samples are marked in red circles (HFail) and a red triangle (VFail).

Most samples from both formations show an increase in pore volume in both the meso-/macropore and micropore size range after failure, although a few failed samples have a decrease in porosity. Some samples have their pore volume increase about 1.5 fold. One EF sample and two NoRM samples, however, have a roughly 5%-10% reduction in pore volume after failure.

3.4.2.3 Fractal Dimension

The surface fractal dimensions D_1 and D_2 are shown in Figure 3.14. All intact EF samples had similar D_1 as well as D_2, indicating a more consistent rock fabric structure. Compared to BET surface area and pore volume, the surface fractal dimensions

are less sensitive to shear failure. However, NoRM 4_14 and NoRM 3_42 increase their fractal dimension to some degree. This might be the sign that the shear failure cause systematic changes in shale nanoscale pore surface, producing more complex surface textures. However, the magnitudes of such changes are likely small.

In short, BET surface area, N_2/ CO_2 porosity, pore size distribution and surface fractal dimensions indicate the impact of shear deformation on the nanoscale pores. The pore volume ratio data show that the porosity change is greater in the NoRM samples than in the EF samples during deformation. Similarly, the fractal surface dimension appear to be more greatly impacted by deformation in the NoRM samples than the EF samples.

3.5 DISCUSSIONS

3.5.1 Effect of Bedding

According to Figures 3.9-3.14, the absolute change in pore volume for any given sample depend primarily on the direction of loading. The vertically failed samples tend to have a larger increase in the total pore volume than the horizontally failed ones. For sample NoRM 3_53, the vertically failed sample has a similar increase of pore volume for meso-/macropore range and micropore range compared to the horizontally failed sample. The vertically failed EF sample (EF_1_223), on the other hand, shows a larger micropore volume increase and a smaller meso-/macropore volume increase. In this context, the primary fabric anisotropy (bedding) in combination with variable cementation states may be imparting heterogeneity in mechanical response.

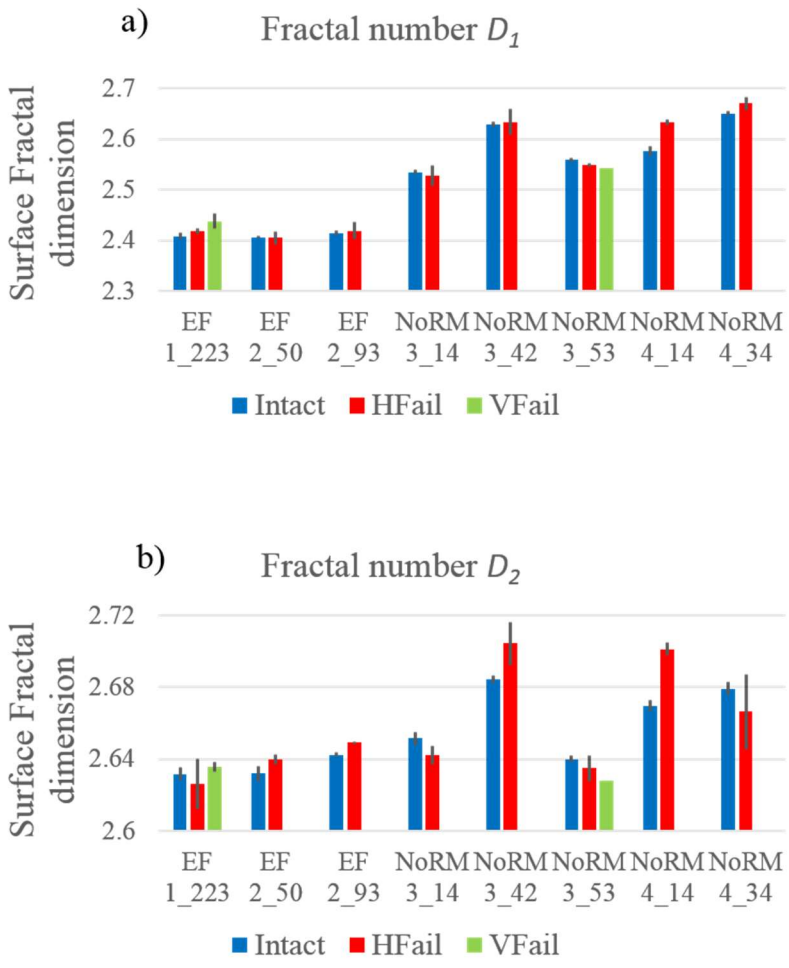

Figure 3.14 (a) Fractal dimension D_1 and (b) fractal dimension D_2 of intact and failed shale samples. The error bar is shown in the plot as the black line.

3.5.2 Effects of Total Clay and Cementation

We offer the following hypothesis for a further role for diagenesis in governing the mechanics of pore-volume and surface-area change with deformation. This hypothesis stems from the ratio of surface area and total pore volume (N_2 pore volume plus CO_2 pore volume) in failed samples relative to intact samples. We plot the surface area ratio and total pore volume ratio with total clay content and total cement content in Figures 3.15 and 3.16.

The plots show that the EF and NoRM shales have similar linear tends. However, an outlier (sample NoRM 3_14) has higher BET surface area ratio and total pore volume ratio. Notwithstanding the outlier, the analyses from the two formations have similar tends, indicating the general impact of minerology on porosity changes with deformation. Note that the total clay content has a negative impact on surface area ratio and total pore volume ratio. We suggest that this is because, in general, clay-rich samples also have less cement and have pores that are more prone to collapse during the deformation. On the other hand, the total cement content has a positive impact on surface area ratio and total pore volume ratio, because cement favors fracturing surface increase.

The outlier (NoRM 3_14) potentially highlights two important aspects of our hypothesized relationship between mechanics and diagenetic history. Firstly, this sample is anomalously cement-poor and clay rich for the samples we investigated and therefore the opening of pores within deforming clay aggregates is enhanced relative to fracturing (Figure 3.8a). Secondly, the OM in this sample appears to be especially particulate and pore-rich (Figure 3.8b). Thus, there is an intrinsic high porosity at the sub-micron scale that is not widely observed in the other NoRM samples. Though clearly limited by the few numbers of samples we observed, this hypothesis highlights how thermal maturation leading to OM porosity is one control, while cementation and the heterogeneous distribution of clay mineral content is another, and the two together can lead to anomalous porosity increases during hydraulic fracturing.

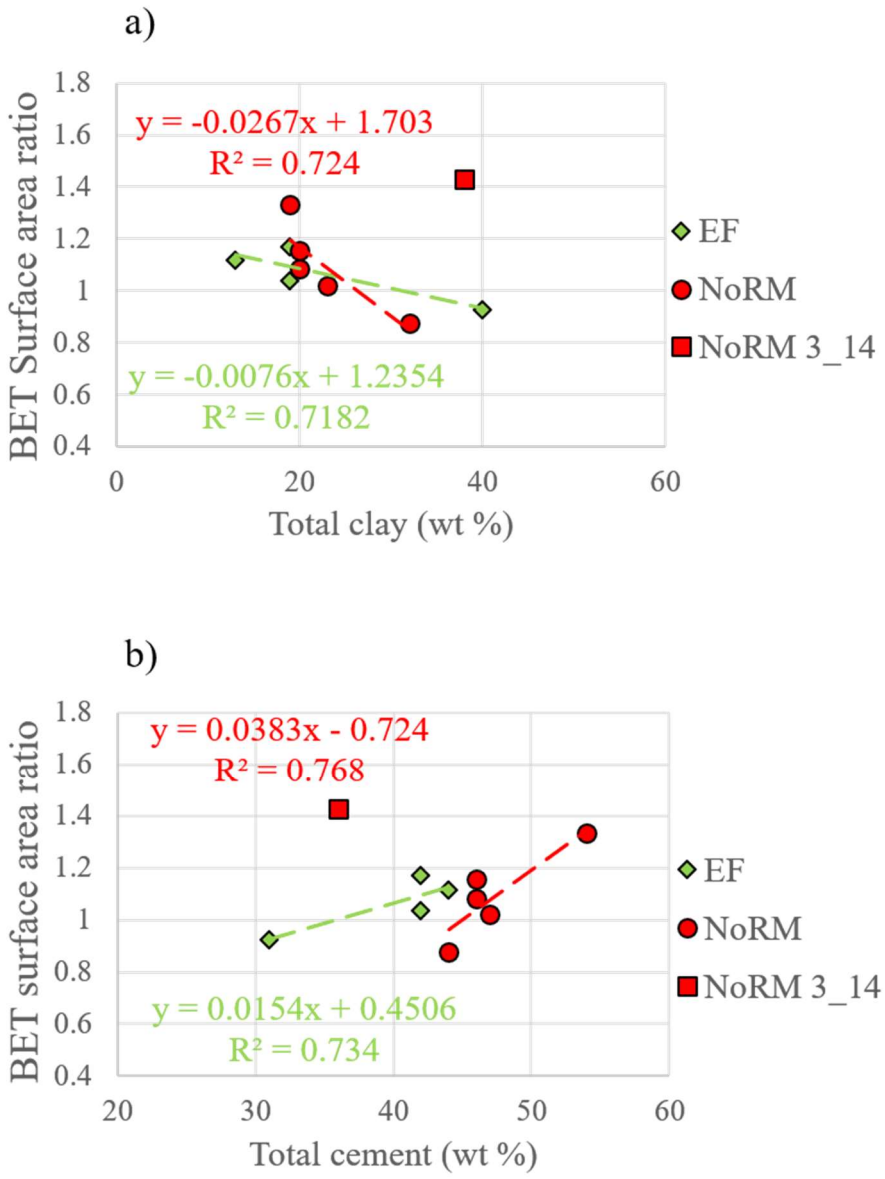

Figure 3.15 (a) Ratio of BET and total clay content. (b) Ratio of BET and total cement content. The total clay = Illite/Mica + Mx I/S. The total cement = calcite * 0.5 + quartz * 0.85. The EF samples are marked in green diamonds and NoRM samples except NoRM 3_14 are marked in red circles. The outlier NoRM 3_14 is marked with a red square. Linear equations for two formations are obtained by removing NoRM 3_14.

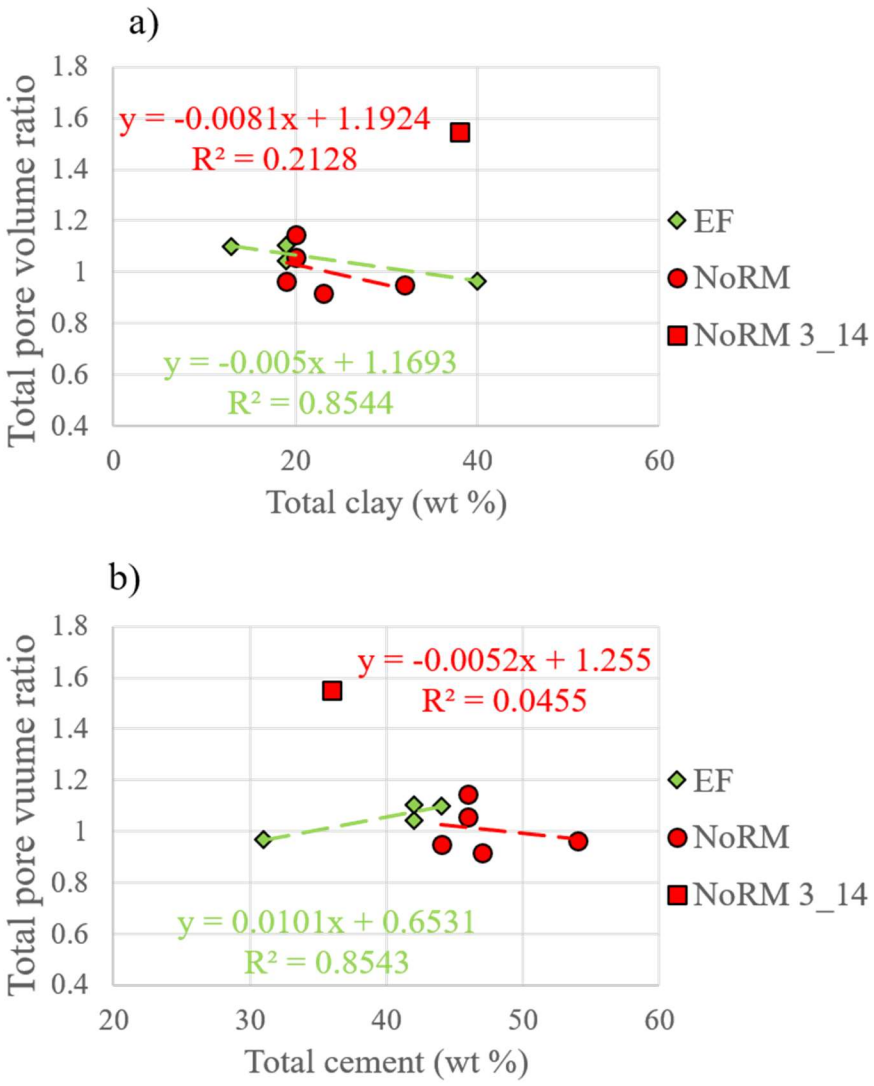

Figure 3.16 (a) Ratio of total pore volume and total clay content. (b) Ratio of total pore volume and total cement content. The total clay = Illite/Mica + Mx I/S. The total cement = calcite * 0.5 + quartz * 0.85. The EF samples are marked in green diamonds and NoRM samples except NoRM 3_14 are marked in red circles. The outlier NoRM 3_14 is marked with a red square. Linear equations for two formations are obtained by removing NoRM 3_14.

3.5.3 Effect of OM

The SEM images of intact samples (e.g. Figure 3.6) suggest that the NoRM samples had more fine-scale organic matter laminations than the EF samples. The mechanically soft (OM and clay) and stiff minerals (quartz and carbonate) are more discretely partitioned in laminae for NoRM samples. Under shear deformation, fractures propagate into the OM-rich soft zone at the mechanical contrast of the two layers. SEMs of failed samples (Figure 3.7 and Figure 3.8) indicate that the fracture further propagate along the OM boundary and into the OM pores. Previous research has shown that OM, especially kerogen, has a lower modulus than surrounding carbonate and silicate grains (Eliyahu et al., 2015; Emmanuel et al., 2016), and that some kinds of OM can fracture under certain circumstances (Daigle et al., 2017b). In contrast, the clay in the shale matrix exhibits ductile (distributed) deformational textures that formed through grain rearrangements and porosity closure during deformation (Dehandschutter et al., 2004; Laurich et al., 2014).

OM-associated pores can be developed in both depositional OM and migrated OM. Depositional OM retains its position and shape from the time of deposition, while migrated OM may change shape and location in response to temperature and pressure during burial (Loucks and Reed, 2014). Loucks and Reed (2014) suggest that the connectivity decreases in the laminated OM compared with the dispersed OM due to OM isolation. Our study indicates that, in contrast, the laminated OM is more sensitive to the shear deformation. The poorer connectivity of the laminated OM will receive relatively greater improvements with deformation. Through the newly formed pathway the hydrocarbons are connected with the main flow channels. Compared to the dispersed OM, the shear failure is more effective to capture laminated OM pores, enhancing the production of shale with the laminated OM. In turn, because many of these microstructural and mechanical properties are developed early in the diagenetic history, the pore evolution of the shale over geologic

time will also be impacted by this contrast between shear and fracture deformation across contrasting distributions and types of OM and surrounding matrix. Future research should consider the timing and role of cementation in establishing these OM relationships, and also the relative importance of cementation to these "granular" controls on syn-deformational porosity change.

3.6 CONCLUSIONS

The effects of shear deformation on shale pore structure at nanoscale was investigated using SEM analysis and gas sorption measurements. Confined compressive strength tests were conducted on preserved shale core plugs that were drilled parallel and perpendicular to the bedding planes.

According to the SEM analysis, fractures with widths ranging from 10-100 nm up to 1-2 μm are observed to follow coarser grain boundaries and laminae of OM and matrix materials. In more laminated materials, fracture lengths are up to hundreds of micrometers, which are likely continuous across entire sample volume. Some fractures initiate along grain contacts and primarily propagated through OM.

N_2/ CO_2 sorption measurements were performed on intact and failed samples The BET surface area, N_2/ CO_2 porosity, N_2/ CO_2 pore size distribution and surface fractal dimensions indicate the impact of shear deformation on the nanoscale pores. Most samples show an increase in their sorption quantity, pore volume, and BET surface area following failure.

Diagenetic differences between calcareous EF and siliceous NoRM samples may lead to different responses to deformation. The differences in rock fabric created by different diagenetic histories cause different nanoscale fracture patterns, including anomalous porosity increases due to pore distributions within OM, heterogeneous

distribution of cement between samples, and enhanced porosity within deformed clay aggregates.

Fractures tend to propagate along the OM laminae and get access to the OM pores. The interaction of the OM laminae and the shear fracturing may improve the connectivity of the OM laminae to the adjacent rock matrix, and thus enhance the hydrocarbon mobility.

Chapter 4: Effects of fluids removal and thermal maturation on permeability and pore structure of organic-rich shale: Results from heat treatment experiments[2]

4.1 INTRODUCTION

Oil and gas present in unconventional shale systems account for a significant portion of the world's hydrocarbon resources (Wang et al., 2014; EIA 2018). The increasing significance of shale gas plays has led to the need for deeper understanding of shale reservoir properties. Using laboratory measurements to determine porosity, pore size distribution and permeability of unconventional reservoirs is critical for reservoir characterization, forecasting production, determination of well spacing, and designing hydraulic fracture treatments (Cui et al., 2013).

Permeability is an important parameter for characterizing shale transport properties. Permeability measurements typically employ unsteady-state methods, which include the pulse decay method (e.g. Brace et al., 1968; Dicker and Smits 1988; Alnoaimi et al., 2014; Heller et al., 2014; Bhandari et al., 2015; Ghanizadeh et al., 2015; Bhandari et al., 2017), and the GRI (Gas Research Institute) method (e.g. Luffel et al., 1993; Egermann et al., 2005; Cui et al., 2009; Tinni et al., 2012; Cui and Brezovski, 2013; Cui et al., 2013).

The general approach for the pulse decay method is to establish a differential pressure between the upstream and downstream ends of the core sample, usually cylindrical in shape. The record of differential pressure versus time is used to calculate the sample's axial permeability (Brace et al., 1968; Dicker and Smits 1988; Cui et al., 2009). This method gains its popularity from its shorter experimental run times, and higher resolution

2 This chapter is based on Jiang, H., Daigle, H. Effects of drying temperature on permeability and pore structure measurements of organic-rich shale. *Journal of petroleum science and engineering*. (under review)

for very low permeability measurements, compared with the steady-state method (Coyner et al., 1993; Cui et al., 2009).

The GRI method, on the other hand, provides a fast estimation of permeability using a crushed sample. In this method, the sample is placed in an evacuated chamber, and then a probing gas is allowed to expand into the chamber. The gas pressure falls off due to the probe gas filling the interior pore space of the sample, and this pressure decay can be used to estimate permeability by numerical modeling (Luffel et al., 1993) or analytical solution (Cui et al, 2009). A detailed review on permeability methods can be found in Chapter 2.4.

For permeability methods, especially those based on gas (e.g. helium, nitrogen), the moisture of the sample can significantly affect the measured permeability (Ghanizadeh et al., 2014; Gao and Li, 2018). Gao and Li (2018) measured the permeability of shale at different water saturations, and showed that decreasing water saturation can exponentially increase the permeability coefficient (Gao and Li, 2018). For this reason, the sample often requires drying before the permeability measurement to obtain as accurate an estimate of the intrinsic permeability as possible (e.g. Cui et al., 2013; Heller et al., 2014; Alnoaimi et al., 2014; Ghanizadeh et al., 2015). By removing the moisture, the probe gas is able to access pores that would otherwise have been blocked by in-situ fluids such as water. Ghanizadeh et al. (2014) reported that the permeability increased by a factor of 6 compared to the as-received state after drying at 105°C.

However, in practice the drying temperature is often below 120°C, which may still cause the pore system to be only partially accessible to the probe gas, resulting in an incomplete characterization of shale transport properties. Past studies have used experiments and theoretical models to investigate the evaporation behavior of confined fluids from nanopores (Fisher et al., 1981; Narayanan et al., 2011; Lu et al., 2015). Their results showed that, for a given temperature (e.g. 90°C), the thickness of adsorbed water

film on pore walls increases as the pore geometry decreases. It can be even more difficult for fluids evaporation from shale matrix due to the abundance of nanopores, poor network connectivity, and multiscale pore system (Jiang et al., 2015; Daigle et al., 2017a). In addition, pores associated with clay aggregates are a fundamental textural feature of shale nanostructure (Kuila et al., 2014), which affect the pore structure and flow properties. The temperature for removing water from clay interlayers needs to be higher than 200°C (Al-Harahsheh et al., 2011; Alnoaimi et al., 2014).

In addition, when the temperature is above 300°C (Al-Harahsheh et al., 2011), the OM can be matured and create new pores. Past experiments used hydrous pyrolysis to study the evolution of pores during the thermal maturation of OM (Ko et al., 2014; Hu et al., 2015; Ko et al., 2016; Ko et al., 2018). SEM imaging results and gas sorption results showed that OM pores were found to be associated with stages of OM maturation, and formation of new pores was related to gas generation and structural rearrangement of OM (Hu et al., 2015; Ko et al., 2016).

To understand the effects of fluids removal and thermal maturation by heat treatment on shale permeability and pore structure, we investigated the evolution of permeability and pore structure of shale by heating the samples at multiple temperature stages. Shale samples were pyrolyzed at 110°C, 250°C, 450°C and $\geq$ 600°C. By using the same sample, the effect of shale heterogeneity at the core scale is greatly reduced. For each heating level, the GRI method was chosen to determine the apparent gas permeability and bulk porosity (hereinafter referred as GRI permeability and GRI porosity). The GRI method is believed to eliminate the microfractures introduced by coring and handling through the crushing process, which provides a better estimation of shale matrix permeability (Luffel et al., 1993; Handwerger et al., 2011; Cui et al., 2013; Ghanizadeh et al., 2015). In addition, the pore characteristics at the nanoscale (<100 nm) such as pore size

distribution and surface area were determined by N_2 gas sorption. All those measurements provide useful information about the effects of fluids removal and thermal maturation for sample preparation procedures on laboratory measurements.

4.2 METHODS

The workflow of this study is shown in Figure 4.1. Four preserved shale samples were used for this work. They were crushed and sieved using 20-35 US mesh trays. They were then dried in an oven at 110°C for 2-4 days, and the same samples were pyrolyzed in a tubing furnace at 250°C,450°C, and $\geq$ 600°C for 24 hours under the protection of argon. For every heating level, samples were cooled down, and GRI measurements were conducted on about 40 g of sample. GRI porosity and GRI permeabilities were then computed. In addition, N_2 gas sorption measurements were conducted on 1-1.5 g of sample with fragment size < 35 mesh (<500 μm). N_2 pore volume, N_2 pore size distribution and Brunauer-Emmett-Teller (BET) surface area were obtained. Detailed procedures are described below.

4.2.1 Samples

In this study, four shale samples from two formations were used for lab measurements. Two samples (EF 1_223, EF 2_93) were from the Eagle Ford shale (Karnes County, TX) and the other two samples (NoRM 3_14, NoRM 4_34) were from a siliceous shale in the northern Rocky Mountains, USA (the location and identity of the siliceous shale has been withheld at the donor's request). All samples were preserved in mineral oil until experimentation. We found no sign of mineral oil imbibing into the samples based on nuclear magnetic resonance measurements (Daigle et al., 2017b).

The X-ray diffraction (XRD) results (Table 3.1 of Chapter 3) provide concentrations of minerals, indicating a wide mineral variation among samples from the same formation. Calcite dominates samples from the Eagle Ford formation, whereas quartz dominates the NoRM samples. Clay minerals also make up a significant fraction of both formations, ranging from 13 to 38 wt.%. According the rock-eval report (Table 3.2 of Chapter 3), the Total Organic Content (TOC) varies from 2.82 to 4.73 wt.%.

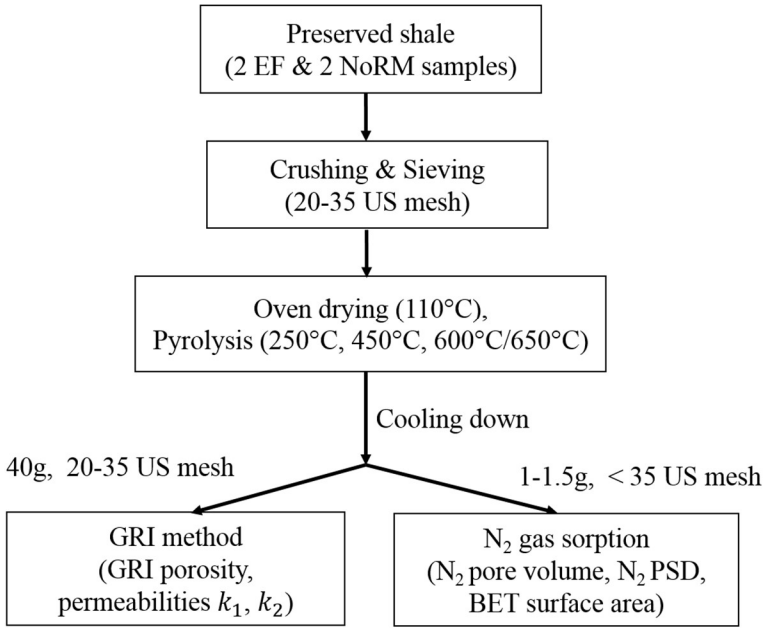

Figure 4.1 Workflow of this study. Two samples were from the Eagle Ford formation (denoted as 'EF'). Two samples were from the northern Rocky Mountains formation (denoted as 'NoRM'). Pyrolysis experiments were conducted in a tubing furnace under an argon atmosphere. GRI stands for Gas Research Institute. PSD stands for pore size distribution. BET stands for Brunauer-Emmett-Teller.

The sample bulk density at as-received conditions is determined using the oil-immersion method. A sample mass is first measured in air. It is then fully immersed in a light mineral oil (with density of 0.808 cm^3/g) and its apparent mass upon immersion is

recorded. According to Archimedes' principle, the bulk volume of the sample equals to the volume of oil that the sample displaces. The oil volume is obtained by dividing the mass difference of sample before and after immersion in oil by the oil density.

Samples were crushed and sieved using 20 and 35 US mesh size (0.841 mm to 0.5 mm) trays. About 40 g of samples between 20 to 35 mesh size was collected for GRI measurements, and 1-1.5 g of samples with size under 35 mesh were collected for N_2 sorption measurements. Sizes between 20 to 35 mesh size are the recommended values for GRI method (Luffel et al., 1993; Cui et al., 2009), whereas smaller particle sizes are preferable for the gas sorption measurement (Kulia and Prasad, 2013a).

4.2.2 Heating

After crushing, samples were dried in an oven for 2-4 days at a temperature of 110°C. The mass before and after drying were recorded. Pyrolysis was conducted on samples in a tubing furnace under an argon atmosphere at 250°C, 450°C, and $\geq$ 600°C (600°C or 650°C) for 24 hours.

The setup of the tubing furnace is shown in Figure 4.2. The inlet of the tubing is connected to the high pressure argon gas cylinder. The tubing outlet is connected to an Erlenmeyer flask (a laboratory flask with a flat bottom) filled with water. The sample is first weighted and placed in a crucible boat. The boat is moved to the center of the tubing. The tubing is then sealed by tightening the screws of the flanges on both ends. During the pyrolysis, the sample is protected by the argon gas. After the pyrolysis, the weight of the sample is recorded again.

The sample mass after the heating is reduced due to the fluid loss. By assuming that the bulk volume V_b is the same before and after heating, bulk density ρ_2 after heating can be estimated:

$$V_b = \frac{m_2}{\rho_2} = \frac{m_1}{\rho_1}, (4.1a)$$

$$\rho_2 = \frac{m_2}{m_1}\rho_1, (4.1b)$$

where ρ_1 and ρ_2 are sample bulk density before and after heating, and m_1 and m_2 are sample mass before and after heating. Bulk densities for 4 samples at all heating levels as well as as-received conditions are listed in Table 4.1.

After each temperature level, both GRI and N_2 sorption measurements were conducted. As the pyrolysis and measurements were conducted on the same sample, the effect of shale heterogeneity at the core scale is anticipated to be greatly reduced.

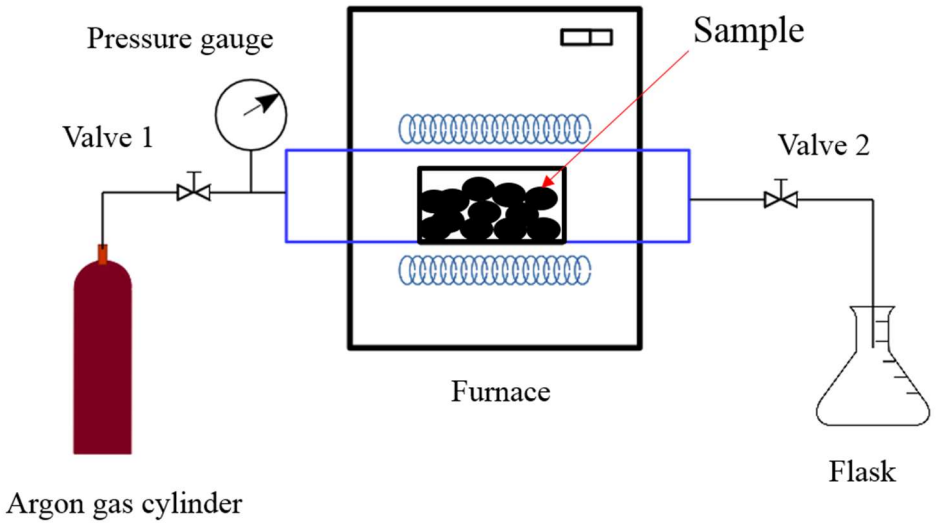

Figure 4.2 Example of the set-up of the tubing furnace for shale sample pyrolysis. The pyrolysis are conducted at 250°C, 450°C and ≥ 600°C for 24 hours. The sample particles are placed in a crucible boat and moved to the center of the tubing (in blue color). During pyrolysis, the sample is protected by the argon gas (red cylinder). Valve 1 controls the flow of argon gas from the gas cylinder. Valve 2 controls the outlet of the tubing. The gas outlet was connected to an Erlenmeyer flask (a laboratory flask with a flat bottom) filled with water.

Table 4.1 Bulk densities for 4 shale samples after heating different heating levels as well as as-received conditions. Samples from the Eagle Ford formation have names starting with 'EF'. Samples from the northern Rocky Mountains formation have names starting with 'NoRM'.

Sample	Bulk Density (cm^3/g)				
	As-received	110°C	250°C	450°C	≥ 600°C
EF 1_223	2.491	2.473	2.434	2.425	2.394
EF 2_93	2.461	2.405	2.400	2.367	2.338
NoRM 3_14	2.437	2.400	2.395	2.355	2.294
NoRM 4_34	2.524	2.505	2.486	2.451	2.394

4.2.3 N_2 Sorption

N_2 sorption measurements were conducted at 77K using a Micromeritics 3Flex surface analyzer. Samples were first dried in the oven at 110°C for 24 hours to remove hygroscopic moisture. Then, 1-1.5 g of sample was placed in a sample tube and degassed under a N_2 stream at 150°C for 4 hours to remove any remaining hygroscopic moisture (Daigle et al., 2017b; Jiang et al., 2018a). This preparation step was applied for all samples regardless of heating level.

Based on gas sorption measurements, surface area was calculated based the Brunauer-Emmett-Teller (BET) method (Brunauer et al., 1938), and pore size distributions were interpreted by nonlocal density functional theory (NLDFT, Roque-Malherbe, 2007; Adesida et al., 2011) using the model for slit-shaped carbon pores. In the pore size distribution, the pore volume is reported in dV/dlog(w), which is the derivative pore volume (V) normalized to natural logarithm of pore width (w). This is the most commonly graphical representation of pore size distribution, which normalizes the effect of irregular experimental point spacing (Kuila and Prasad, 2013a). More details on gas sorption can be found in Chapter 2.3.

4.2.4 GRI Method

4.2.3.1 Experiment Setup

GRI permeability of the sample was conducted in a pycnometer using helium (He) gas (Figure 4.3). The setup consists of two chambers: the reference chamber and sample chamber. To reduce the influence of the ambient temperature, the setup is kept in an isothermal cabinet with a constant temperature of 30°C, which is slightly above the room temperature.

Initially, all valves are opened to the atmosphere through the outlet. A known weight of crushed sample is introduced to the sample chamber. Before the measurement, the sample is allowed to stabilize for a few hours. The system is then evacuated using a vacuum pump (depicted as Stage 1 in Figure 4.4a).

In Stage 2, valves 1 and 3 are closed, and valves 2 and 4 are sequentially opened. Pressurized helium gas (around 220 psi) fills the reference chamber, and the system is allowed to equilibrate for a few minutes for stabilization (Figure 4.4a).

In Stage 3, valve 1 is opened and the helium gas expands from the reference chamber into the sample chamber. The helium pressure immediately drops to a level by filling the dead space in the sample chamber and subsequently decays as gas permeates the shale particles (Figure 4.4a/ Figure 4.4b).

4.2.3.2 GRI Porosity

According to Boyle's law, Bulk porosity Φ of the shale sample can be computed . The vacuum pressure (P_1) at Stage 1 is the mean of pressure points that are less than 0.5 psi. The pressure at Stage 2 (P_2) is the mean of pressure points during that period. The final decay equilibrium pressure (P_3) at Stage 3 is the mean of the last several pressure points. The porosity is calculated using the following equation (Cui et al., 2009):

$$\Phi = \left[V_r\left(\frac{P_2}{z_2} - \frac{P_3}{z_3}\right) + (V_s - V_b)\left(\frac{P_1}{z_1} - \frac{P_3}{z_3}\right)\right] / (\frac{P_3}{z_3} - \frac{P_1}{z_1})V_b \,, (4.2)$$

where V_r, V_s are volumes of reference and sample chambers, V_b is the bulk volume, and z_1, z_2, and z_3 are the compressibility factors of helium at pressures P_1, P_2, and P_3. The bulk volume is based on Table 4.1.

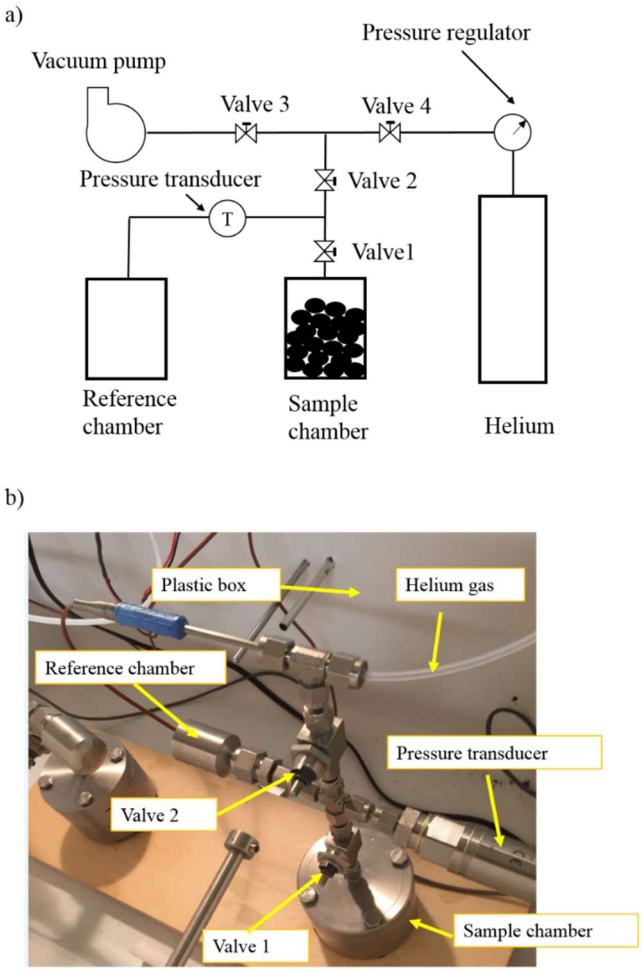

Figure 4.3 (a) Schematic and (b) actual set-up of the GRI method using crushed sample. It consists of a reference chamber and a sample chamber. The reference chamber and sample chamber are kept in a temperature-controlled plastic box to keep a constant temperature. Valve 1 controls the gas passage between the reference chamber and the sample chamber. Note: for (a), black particles represent crushed shale samples.

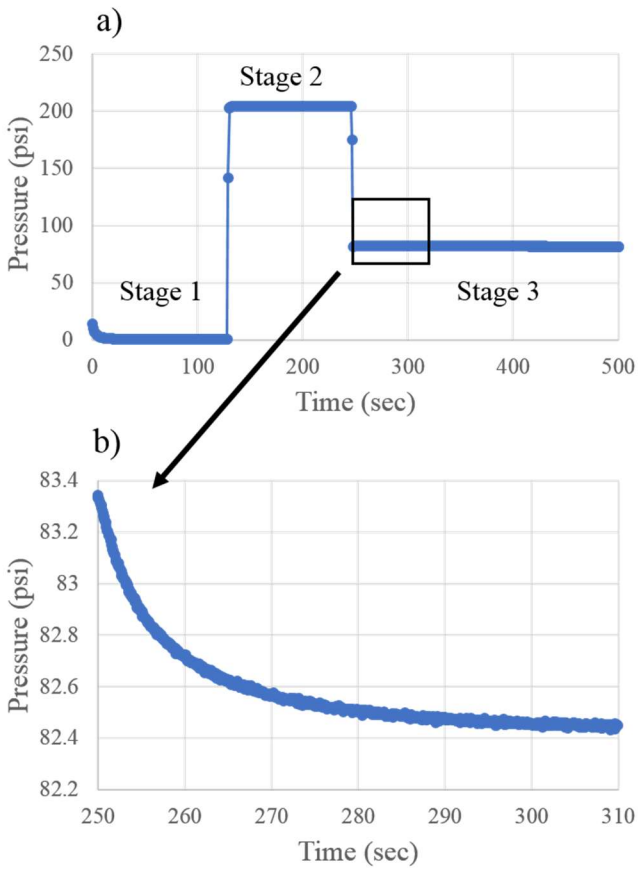

Figure 4.4 Example of recorded pressure data over time. (a) Pressure data includes Stage 1, Stage 2, and early time for Stage 3. (b) Early time for Stage 3 of the selected window in (a). Stage 1 is the period when the system is evacuated by a vacuum pump. Stage 2 is the period when pressurized helium gas fills the reference chamber. The Stage 3 is the period when helium expands from the reference chamber into the sample chamber. The entire measurement for Stage 3 is about 17-24 hours. The time is in the unit of second, and the pressure is in the unit of psi.

4.2.3.3 GRI Permeability

Furthermore, permeability can be calculated based on the pressure decay data in Stage 3 (Cui et al., 2009). Since the gas first fills up the pores of the high permeability zones and then slowly migrates into the low permeability zone, there can be more than one permeability value from the measurement (Cui et al., 2013). Here we computed two

permeabilities k_1 and k_2: k_1 is the early time permeability for $ln(FR)$ between -1.5 and -2.5, and k_2 is the late time permeability with $ln(FR)$ between -3.5 and -4. FR is defined below.

FR is the mass fraction of potential gas relative to the total gas that is taken by the sample at the end of Stage 3 (Cui et al., 2009). The potential gas is defined as the gas mass in the void volume of the reference and sample chambers that will eventually be taken up by sample particles. The expression for FR at a given time t is:

$$FR(t) = 1 - \frac{(K_c + 1)\big(\rho_c - \rho(t)\big)}{\rho_c - \rho_1} \text{ , (4.3)}$$

where K_c is the ratio of gas storage capacity of the total void volume of the reference and sample chambers, $\rho(t)$ is the gas density at time t. ρ_1 is the initial gas density in the sample pore space (same as the average gas density for Stage 1), and ρ_c is the average initial gas density in the sample and reference chambers at the beginning of Stage 3. K_c and ρ_c are expressed as:

$$K_c = \frac{(V_r + V_s)}{V_b\, \Phi} \text{ , (4.4)}$$

$$\rho_c = \frac{\rho_2 V_r + \rho_1 (V_s - V_b)}{V_r + V_s - V_b} \text{ , (4.5)}$$

where V_r, V_s are volumes of reference and sample chambers, V_b is the bulk volume, andρ_1, ρ_2 are the density at pressures P_1 and P_2 given in Equation (4.2).

By assuming the size particles are spheres with a relatively uniform radius (R_a), FR has an analytical form (Cui et al., 2009):

$$FR(t) = 6K_c(K_c + 1) \sum_{n=1}^{\infty} \frac{e^{-\frac{K\alpha_n^2}{R_a^2}t}}{K_c^2\alpha_n^2 + 9(K_c + 1)} \text{ , (4.6)}$$

where α_n is the n^{th} root of

$$tan\ \alpha = \frac{3\alpha}{3 + K_c\alpha^2}\ , (4.7)$$

and K is:

$$K\ = \frac{k}{\mu c_g \Phi}\ , (4.8)$$

where k is the permeability, μ is the gas viscosity, c_g is the gas compressibility, and Φ is the porosity.

If K_c is large (e.g. K_c > 50), the logarithmical value of FR becomes a linear function of time (Cui et al., 2009). The natural logarithm form of Equation (4.6) can be expressed as

$$ln(FR) = b\ - \frac{K\alpha_1^2}{R_a^2} t\ , (4.9)$$

Through linear regression, the slope of Equation (4.9) can be combined with Equation 4.9 to compute the permeability k. Details of permeability calculation using GRI method can be found in Chapter 2.4.4.

The reproducibility of the GRI measurements was tested using two samples. We performed GRI measurements twice for sample NoRM 3_14 after drying at 110°C. The porosities were 8.55% and 8.50%, and permeabilities k_1 were 19.12 and 22.50 nanodarcies (nD) while permeabiliies k_2 were 2.80 and 0.58 nD. We performed measurements sample NoRM 4_34 after drying at 40°C. The porosities from the two measurements were 3.36% and 2.83%. Permeabilities k_1 were 0.3 and 0.36 nD, and permeabilities k_2 were 0.22 and 0.86 nD. The tests show that the variability of porosity

and k_1 is small, whereas the variation is higher for k_2. GRI data shown in the Results section have a greater variation than those due to measurement error.

4.3 RESULTS

4.3.1 Gas Sorption

Nitrogen adsorption-desorption isotherms for EF and NoRM samples are shown in Figures 4.5 and 4.6. All samples exhibit H3 hysteresis loops, indicating that the material has both mesopores (2-50 nm) and macropores (>50 nm) (Sing, 1985). As heating temperature increases, the adsorption and desorption quantity of the sample increase, implying an increase tend of pore volume. The most significant quantity increase occurs after heating at temperatures $\geq$ 600°C. On the other hand, heating from 110°C to 250°C is associated with less significant change to the isotherms, suggesting that the change of pore volume between 110°C and 250°C is not significant.

N_2 pore size distributions are calculated from the isotherms, and depicted in Figure 4.7 and Figure 4.8. The pore size ranges from 1.8 to 100 nm. Similar to the isotherms, the most significant increase occurs when heating temperatures $\geq$ 600°C. The variation of pore volume is not strong when the temperature increases from 110°C to 250°C. In addition, pore volume in pores larger than 10 nm display a more significant increase compared to those smaller than 10 nm. Besides the increase in adsorbed quantity, the shape of the isotherm, however, remains similar through all temperature levels, indicating little change in pore network connectivity (e.g. Seaton, 1991). Note that the exception is sample EF 2_93 at 650°C, which could be the demineralization of carbonates in some degree when the temperature ranges from 650 and 850°C (Al-Harahsheh et al. 2011)

BET surface areas and N_2 pore volumes are also obtained. The results are shown in Figure 4.9 and summarized in Table 4.2. Most shale samples generally show a increasing

trend in surface areas as the heating temperature increases. For samples EF 2_93 and NoRM 4_34, the maximum surface area occurs after heating at 450°C.

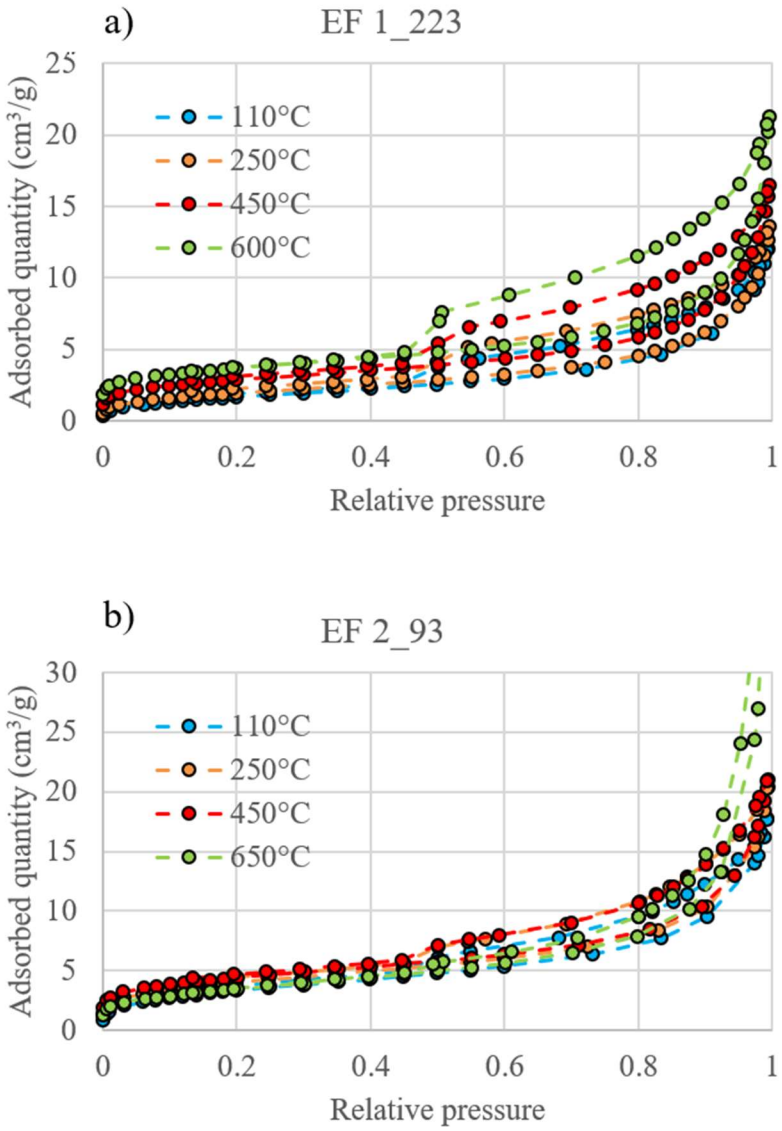

Figure 4.5 N_2 gas sorption measurements for samples (a) EF 1_223 and (b) EF 2_93. The two samples were from Eagle Ford formation (Karnes County, Texas, USA). Isotherms (marked by different colors) were collected after heating at 4 different temperature levels. Pressure is reported as relative pressure, and adsorbed quantity is reported in gas volume per unit sample mass at standard pressure and temperature conditions.

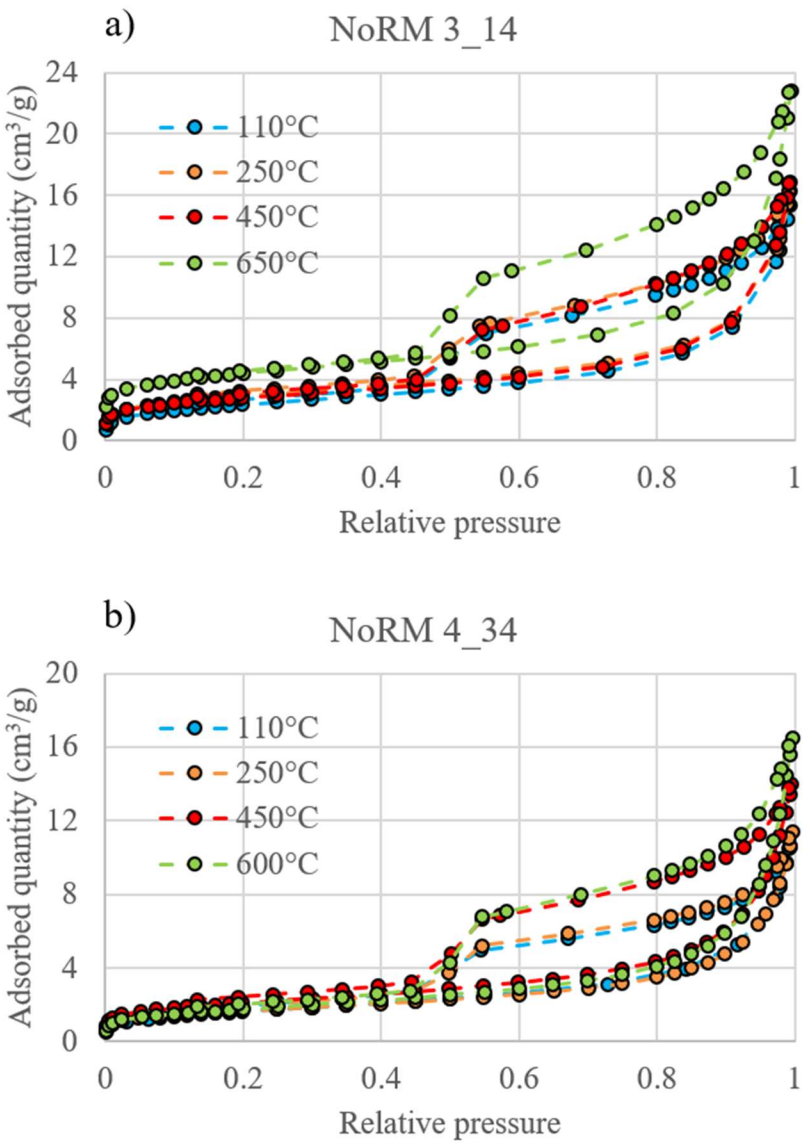

Figure 4.6 N_2 gas sorption measurements for samples (a) NoRM 3_14 and (b) NoRM 4_34. The two samples were from the northern Rocky Mountains formation (USA). Isotherms (marked by different colors) were collected after heating at 4 different temperature levels. Pressure is reported as relative pressure, and adsorbed quantity is reported in gas volume per unit sample mass at standard pressure and temperature conditions.

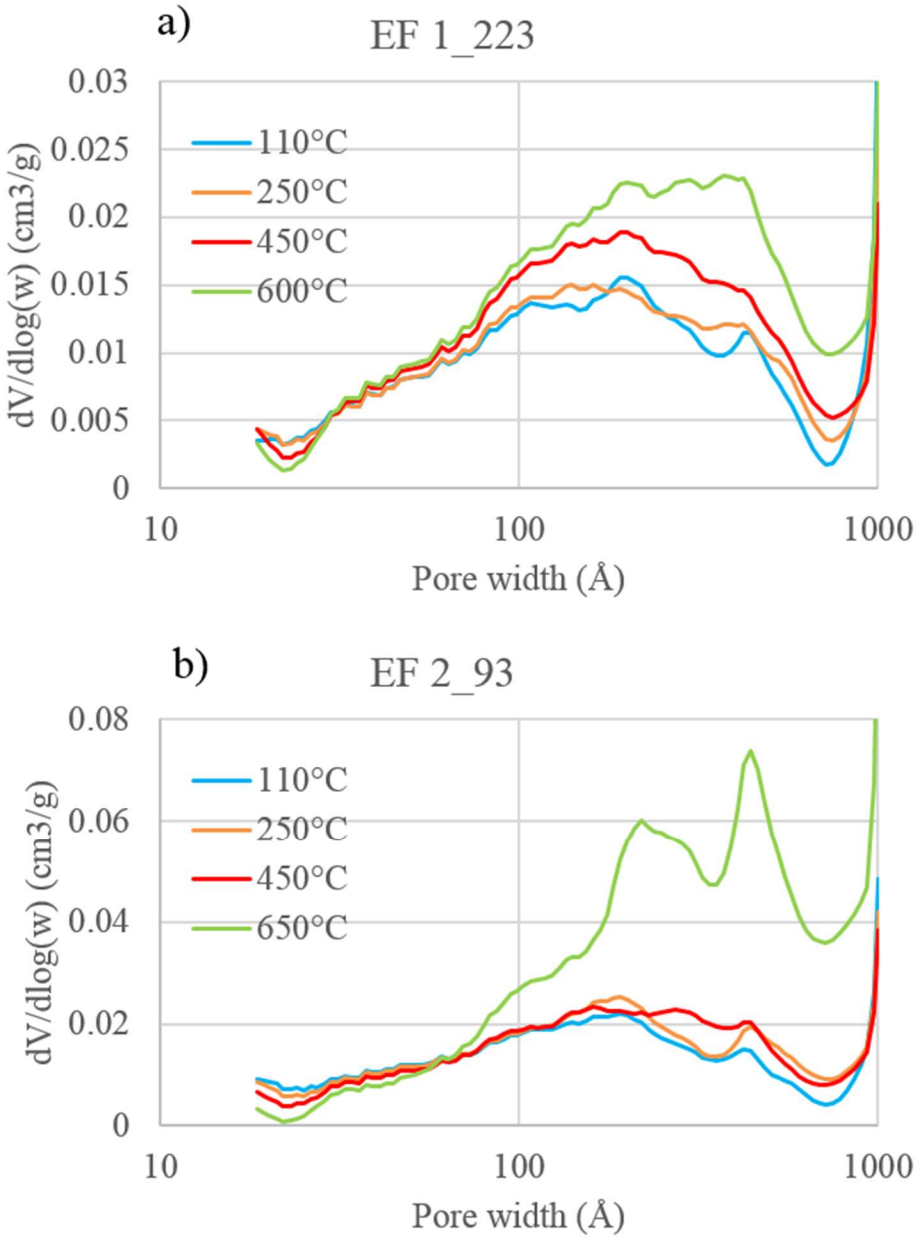

Figure 4.7 N_2 pore size distributions of four samples of (a) EF 1_223 and (b) EF 2_93. Pore size distributions after different heating levels are in different colors. The pore size distribution was based on nonlocal density functional theory using the model for slit-shaped pores. The regularization value is 1.0. The range of the pore width is from 1.8 to 100 nm. The pore volume is reported in dV/dlog(w), which is the derivative pore volume (V) normalized to natural logarithm of pore width (w).

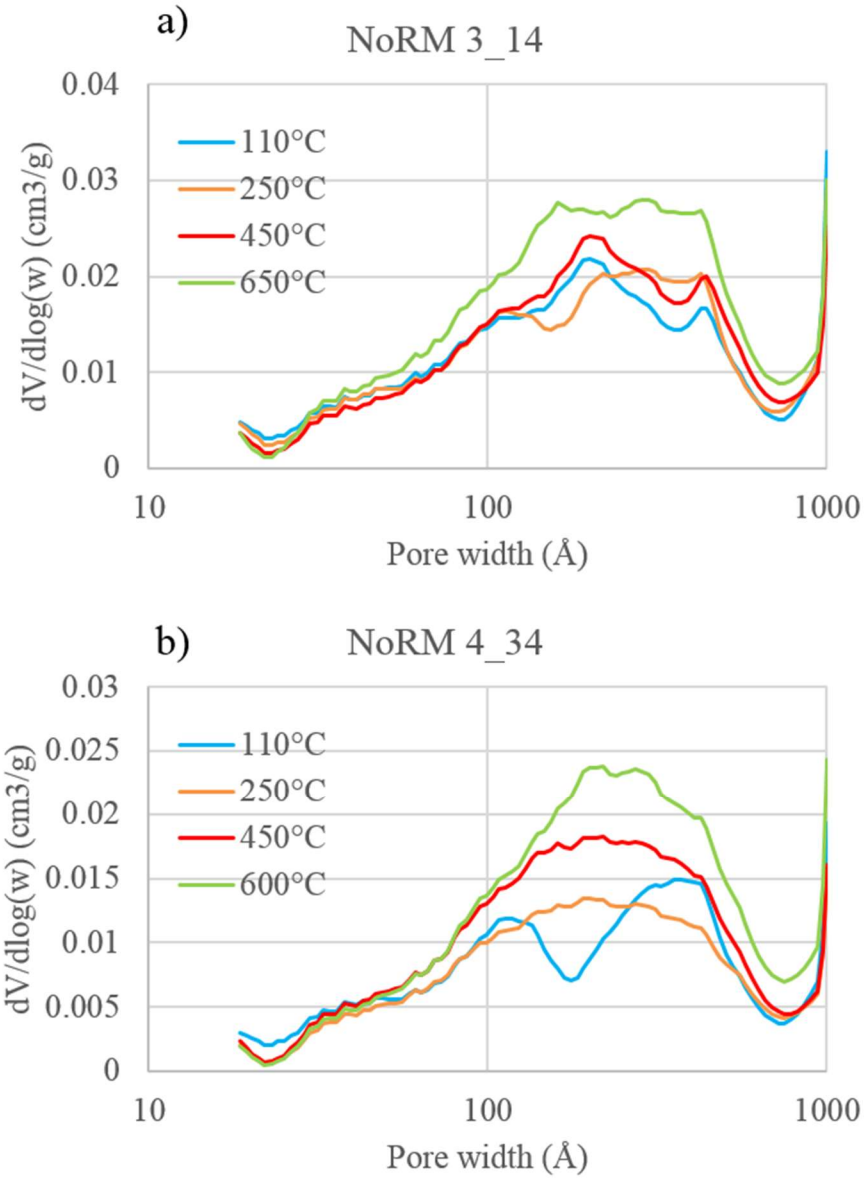

Figure 4.8 N_2 pore size distributions of four samples of (a) NoRM 3_14 and (b) NoRM 4_34. Pore size distributions after different heating levels are in different colors. The pore size distribution was based on nonlocal density functional theory using the model for slit-shaped pores. The regularization value is 1.0. The range of the pore width is from 1.8 to 100 nm. The pore volume is reported in dV/dlog(w), which is the derivative pore volume (V) normalized to natural logarithm of pore width (w).

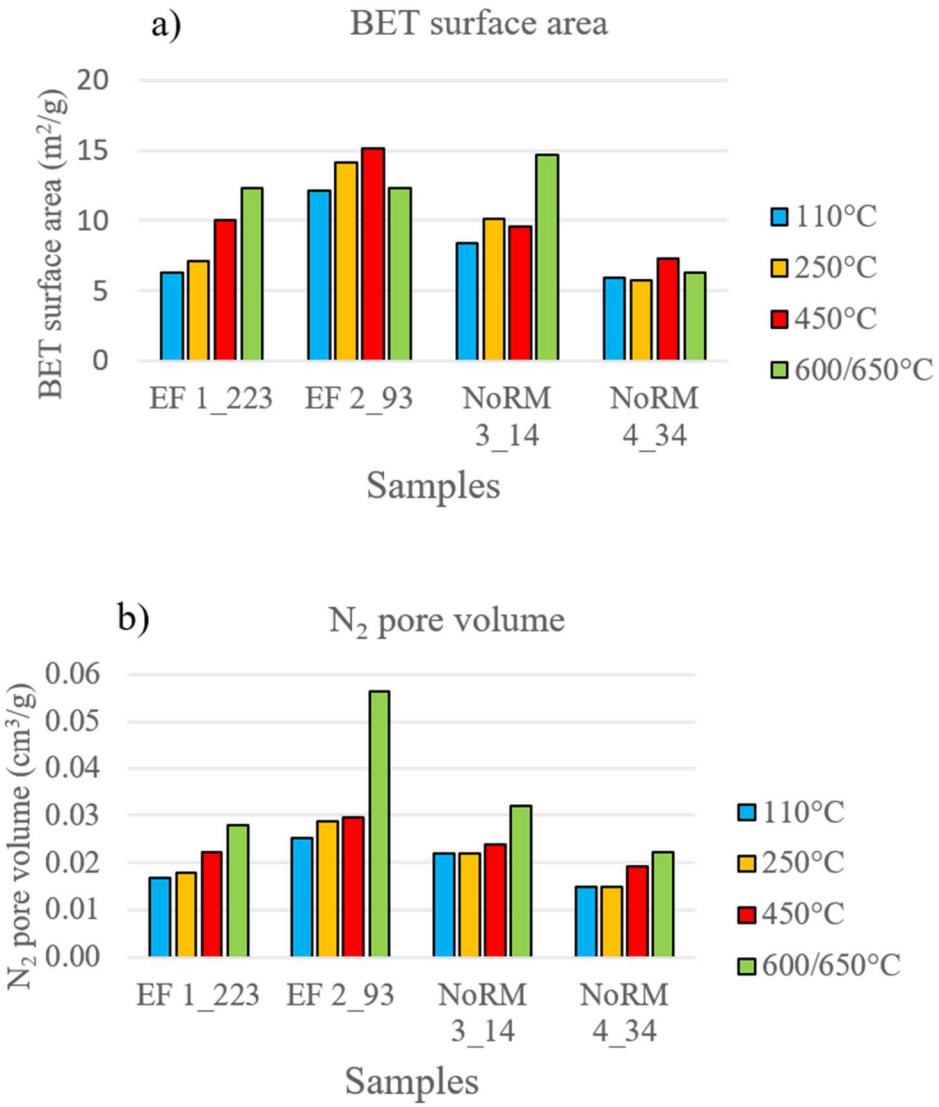

Figure 4.9 (a) BET surface areas and (b) N_2 pore volumes of four samples after heating at 4 levels.

Table 4.2 BET surface areas and N_2 pore volumes of 4 shale samples based on gas sorption measurements. The BET surface area (N_2 pore volume) ratio is the value at a given temperature level with respect to the value at 110°C. The units of BET surface area are m^2/g, and the units of N_2 pore volume are cm^3/g.

Sample	Heating level	BET surface area		N_2 pore volume	
		Value	Ratio	Value	Ratio
EF 1_223	110°C	6.32	1	0.017	1
	250°C	7.15	1.13	0.018	1.06
	450°C	10.04	1.59	0.022	1.31
	600°C	12.34	1.95	0.028	1.65
EF 2_93	110°C	12.17	1	0.025	1
	250°C	14.11	1.16	0.029	1.14
	450°C	15.12	1.24	0.030	1.17
	650°C	12.28	1.00	0.057	2.24
NoRM 3_14	110°C	8.44	1	0.022	1
	250°C	10.09	1.20	0.022	1.00
	450°C	9.63	1.24	0.024	1.10
	650°C	14.65	1.74	0.032	1.46
NoRM 4_34	110°C	5.89	1	0.015	1
	250°C	5.76	0.98	0.015	1.00
	450°C	7.28	1.24	0.019	1.29
	600°C	6.27	1.07	0.022	1.49

N_2 (total) pore volume (Figure 4.9b) is obtained from the pore size distribution. It displays an increasing trend as temperature increases, although the increase for individual pore volume in the pore size distribution plots (e.g. Figure 4.7) is not that visually significant. After heating at 450°C, all samples increase their pore volume by 10% - 31% with respect to 110°C, and have a further increase after heating $\geq$ 600°C. This indicates that pore volumes at the nanoscale depend highly on the temperature level. In addition, EF and NoRM shales show different pore volume responses after heating from 110°C to 250°C. The EF samples have 6% - 14% of increase in pore volume, whereas the NoRM samples show almost no change.

4.3.2 GRI Method

GRI porosity and two permeabilities k_1 and k_2 are obtained from the GRI measurements. The permeability is the apparent permeability to gas and no gas slip correction is applied. Adding corrections like the Klinkenberg effect provides little benefit, since there are no measurement or data analysis standards for GRI method (Heller et al., 2014). The data are plotted in Figure 4.10 and listed in Table 4.3.

The GRI porosity ranges from 4.7 to 8.6% after heating at 110°C (Figure 4.10a). Similar to N_2 pore volume, GRI porosity shows an increasing trend against temperature . After heating $\geq$ 600°C, the porosity has the highest increase, ranging from 8.4 to 12.8%. The maximum increase is about 1.8-fold compared to the value at 110°C conditions.

Most of the GRI permeabilities are within the range of 1 to 30 nanodarcy (nD). Similar to porosity, there is a general increase for both permeabilities k_1 and k_2 against the heating temperature (Figures 4.10b, c). k_1 at the 110°C level shows a wider variation than k_2. They both have a significant increase after heating at 450°C. A few samples, however, have a reduction of permeability after heating $\geq$ 600°C. The correlation plot (Figure 4.11) shows that k_1 has a positive correlation with k_2, with values higher than k_2.

Furthermore, the GRI porosity shows a good correlation with k_1, while there is a more scattered relation between the porosity and k_2 (Figure 4.12). The linear regression between $\log(k_1)$ and the porosity is performed by removing the outlier point (indicated by the red arrow in Figure 4.12a).

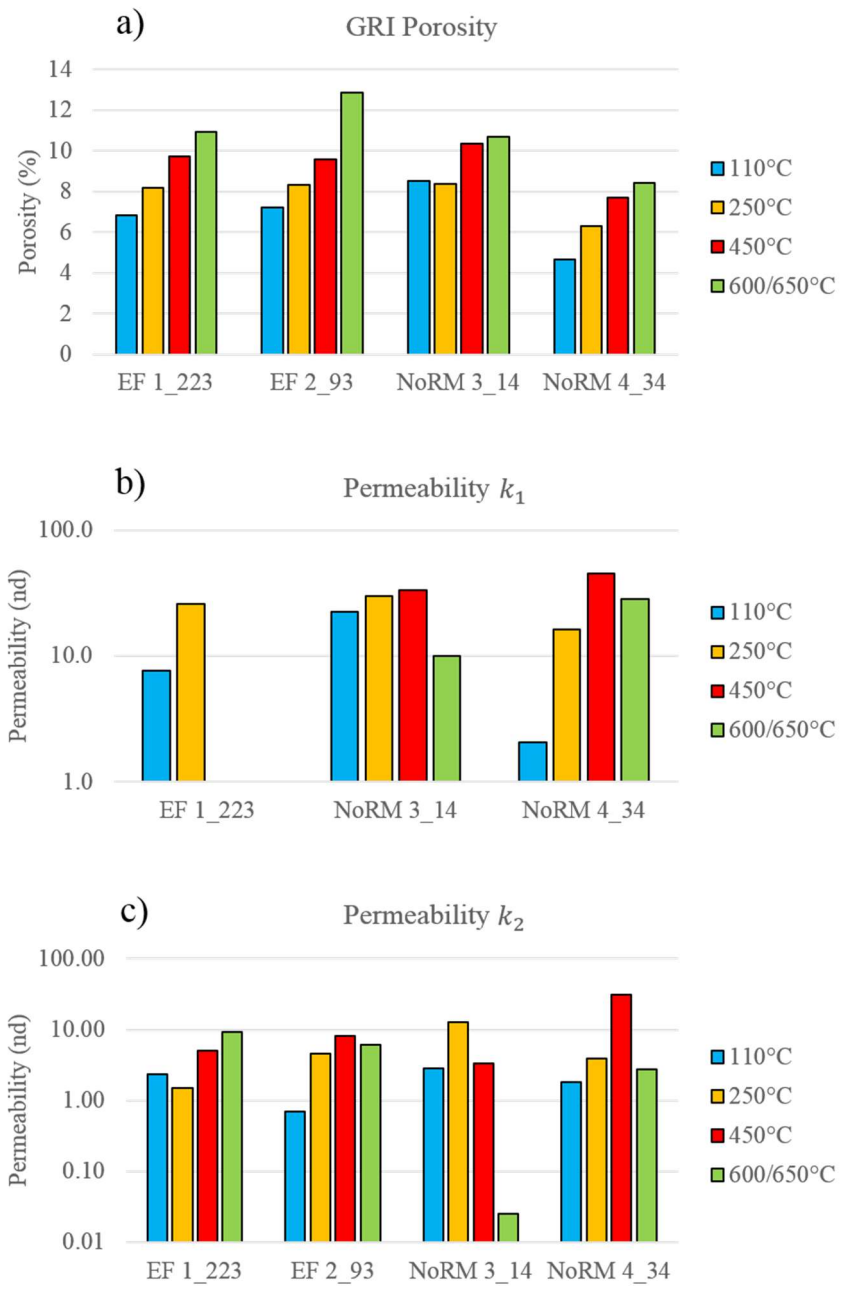

Figure 4.10 Results from GRI method for 4 shale samples after heating at 4 different levels: (a) GRI porosity, (b) GRI permeability k_1, and (c) GRI permeability k_2. The units of GRI porosity are %, and the units of permeability are nano-Darcies. Permeabilities of (b) and (c) are in log scale.

Table 4.3 GRI porosities, and permeabilities of 4 shale samples. The units of GRI porosity are %, and the units of permeability are nano-Darcies. k_1 is the permeability obtained with $ln(FR)$ of 1.5 - 2.5. k_2 is the permeability obtained with $ln(FR)$ of 3.5 - 4.0. The ratio is the permeability value with respect to 110°C. R^2 is the R squared value, a measure of goodness of fit of the linear regression. Note that a few EF samples only have k_2 reported, as the k_1 was not able to be computed due to the fast drop of $ln(FR)$ at the early period.

Sample	Heating level	GRI porosity		GRI permeability k_1			GRI permeability k_2		
		Value	Ratio	Value	Ratio	R^2	Value	Ratio	R^2
EF 1_223	110°C	6.80	1	7.60	1	0.99	2.34	1	0.76
	250°C	8.16	1.20	25.86	3.40	0.98	1.48	0.63	0.94
	450°C	9.72	1.43	-	-	-	5.06	2.16	0.65
	600°C	10.93	1.61	-	-	-	9.25	3.95	0.69
EF 2_93	110°C	7.98	1	-	-	-	0.70	1	0.62
	250°C	8.33	1.15	-	-	-	4.56	6.52	0.14
	450°C	9.56	1.32	-	-	-	8.16	11.67	0.97
	650°C	12.84	1.78	-	-	-	6.11	8.75	0.81
NoRM 3_14	110°C	8.55	1	22.50	1	0.99	2.8	1	0.99
	250°C	8.50	0.98	29.74	1.32	0.99	12.60	4.50	0.67
	450°C	8.36	1.22	33.16	1.47	0.98	3.29	1.17	0.96
	650°C	10.67	1.26	10.04	0.45	0.89	0.02	0.01	0.71
NoRM 4_34	110°C	4.67	1	2.07	1	0.99	1.83	1	0.56
	250°C	6.31	1.35	16.19	7.81	0.99	3.95	2.16	0.67
	450°C	7.70	1.65	45.36	21.88	0.99	31.21	17.09	0.83
	600°C	8.43	1.81	28.15	13.58	0.89	2.78	1.5	0.85

To sum up, the above results show that heating has a significant impact on pore structure parameters and permeability measurements of shale. Heating at 110°C produces measurable pore space for the probing gas by removing the fluid to some degree. Heating at 250°C introduces more accessible space by further fluids removal including bound water of clay aggregates. Heating at 450°C and $\geq$ 600°C can further expose the pore space by the continuous fluid removal. They may also create new pores through thermal maturation of organic matter, leading to additional increase of porosity and permeability. The pore structure parameters such as BET surface area, N_2 pore volume, and GRI porosity have

good correlations with the heating temperatures. The GRI permeabilities are significantly affected by the temperature, which show an increasing trend with a more complicated scattered pattern.

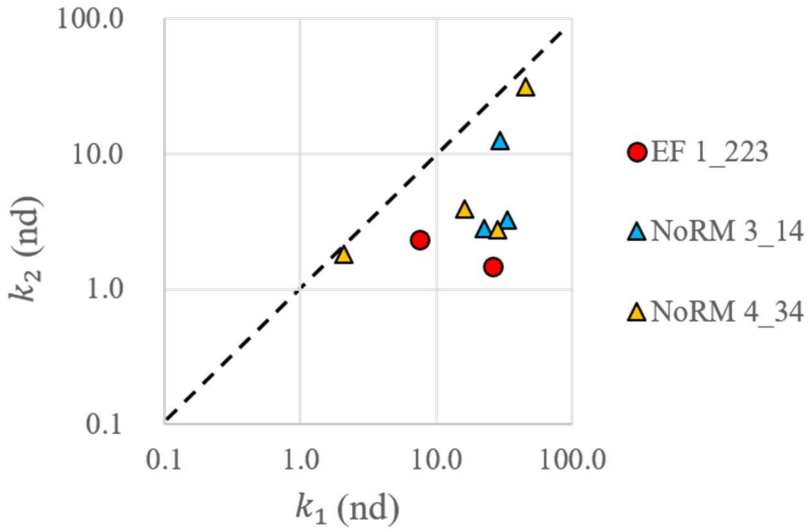

Figure 4.11 Correlation between GRI permeability k_1 and GRI permeability k_2. Samples (circles) from the Eagle Ford formation have names starting with 'EF'. Samples (triangles) from the northern Rocky Mountains formation have names starting with 'NoRM'. The units of permeability are nanodarcies. The dashed line represents 1:1 equivalence.

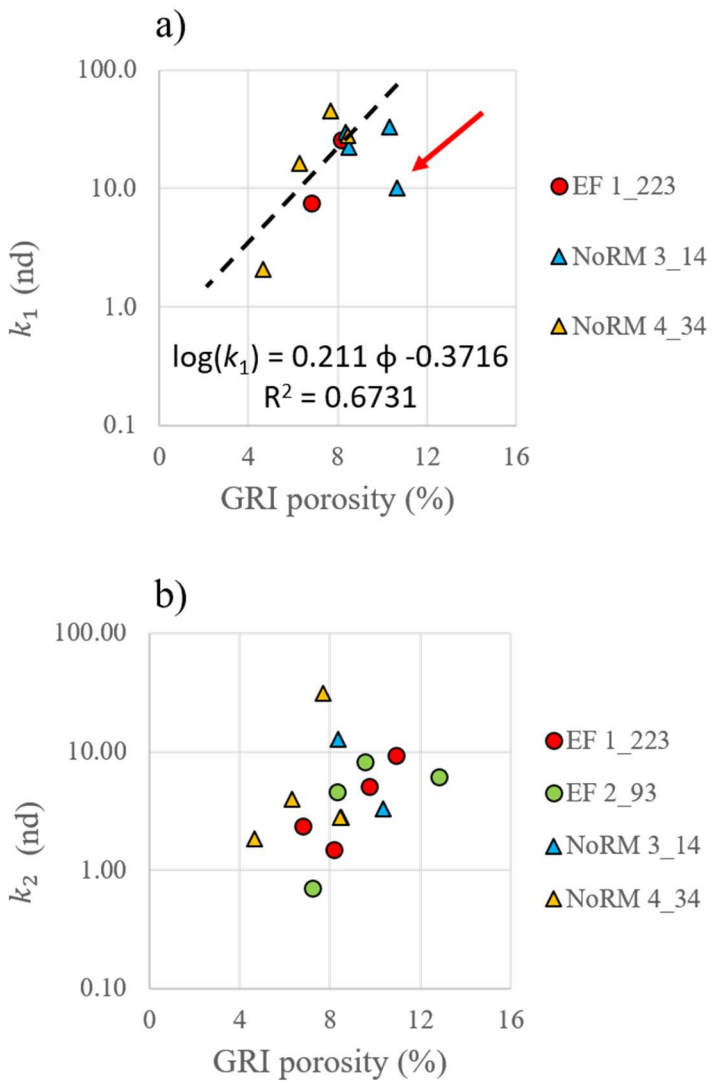

Figure 4.12 (a) Correlation between GRI porosity and permeability k_1. The linear regression coefficients by removing the outlier point (marked by the red arrow) is shown on the plot. The dashed line represents the line based on linear regression. (b) Correlation between GRI porosity and permeability k_2. Samples (circles) from the Eagle Ford formation have names starting with 'EF'. Samples (triangles) from the northern Rocky Mountains formation have names starting with 'NoRM'. The units of permeability are nanodarcies.

4.4 DISCUSSION

4.4.1 Effects of heat treatments

The pore parameters I consider include BET surface area, N_2 pore volume and GRI porosity measurements. These parameters all increase as heating temperature increases. These increases can be first ascribed to continuous loss of fluids. After heating at 110°C, free fluids like water and hydrocarbons in large pores are mostly likely to be removed. This creates accessible pore space for the probe gas. After heating at 250°C, capillary bound fluids in small pores and clay bound water can be further removed, resulting in an increase in measurable pore volume and surface area. After heating at 450°C and above 600°C, besides removal of clay bound water, OM can be matured. The thermal maturation can cause organic matter to degrade and produce new pores space, contributing to additional increases in pore parameters.

The increase of pore volume can explain the increase of permeability from heating temperature of 110°C to 450°C. Permeability, however, decreases after heating above 600°C. One reason may be hydrocarbon production due to thermal maturation: the produced hydrocarbons remain in the organic pores and block the pore system, causing a decrease in overall permeability.

Although almost all isotherms remain similar at all thermal stages, sample EF 2_93 experienced a significant change to its shape after heating at 650°C. The hysteresis of the isotherm decreased, indicating an increase of pore connectivity (Seaton, 1991). EF 2_93 also showed a 2.24-fold increase in its N_2 pore volume relative to that after heating at 110°C. This may be due to the decomposition of carbonate in the sample (e.g., Al-Harahsheh et al. 2011), as it has a significantly high calcite content (70 wt. %). The degradation of carbonate can create new mineral-hosted pores that increase the pore

volume and pore connectivity, even though the permeability measurement doesn't show strong increase evidence after 650°C.

Note that our heat treatment to mature shale is conducted at atmospheric pressure and without the presence of water. Another way to mature shale is by hydrous pyrolysis, which is conducted in a high pressure stainless-steel reactor with a 5 wt. % NaCl solution (Lewan, 1993; Hu et al., 2015). Hydrous pyrolysis is believed to better simulate natural petroleum formation. The presence and absence of liquid water can affect the thermal decomposition of generated bitumen (Lewan, 1997). In the absence of liquid water, the formation of an insoluble bitumen (e.g. pyrobitumen) is the dominant reaction pathway, whereas the formation of saturated-enriched oil is the dominant reaction pathway in the presence of liquid water. The different OM products due to different heat conditions might affect the formation of pore system and pathway. Caustion is warranted when compared with results from hydrous pyrolysis.

4.4.2 Correlation between Gas Sorption and GRI Measurements

Cross-plots of N_2 pore volume and GRI porosity as well as permeabilities are shown in Figure 4.13. The N_2 pore volume shows good correlations with both GRI porosity and permeability k_1. Since N_2 pore volume is a measure of shale matrix pore space at the nanoscale (pores smaller than 100 nm), this suggests that the shale matrix at the nanoscale is positively correlated the porosity and permeability properties. Note that the relationship between N_2 pore volume and k_2 is more complicated (Figure 4.13c).

Those results imply that the larger pores that contribute to k_1 are well connected, while the smaller ones that contribute more to k_2 appear to be more poorly connected. It should be noted here that these "larger" pores are smaller than about 100 nm in width, as this is the largest pore that can be seen by nitrogen sorption. This result is consistent with

modeling performed by Jiang et al. (2015) and Tian and Daigle (2018b) on Barnett shale samples, wherein a certain fraction of small pores were assumed to be partially or completely disconnected from the overall structure. The importance of multiscale pore network features on shale permeability has been similarly demonstrated by other researchers (Ambrose et al., 2012; Mehmani et al., 2013; Mehmani and Prodanović, 2014; Tahmasebi et al., 2015). This result is therefore consistent with other literature and serves as a further indication of pore size-dependent connectivity within the pore network of organic shales.

4.4.3 Permeability Measurement Interpretation

An interesting related issue is raised by recalling that the variations in pore volume observed in our samples are brought about by heating at different temperatures. Permeability ratios of k_1 and k_2 at higher levels with respect to values at 110°C are shown in Figure 4.14.

Accordingly, k_1 is strongly affected by heating temperature, with individual samples exhibiting a nearly tenfold increase in permeability over the range of temperatures. k_2 is also affected, though the magnitude of the effect is somewhat smaller and considerably more varied. These effects occur even at temperatures low enough that diagenetic reactions (e.g. maturation) are limited. Caution is therefore warranted in interpreting permeability measurements performed on shales, as the preparation technique appears to have a significant influence on the results.

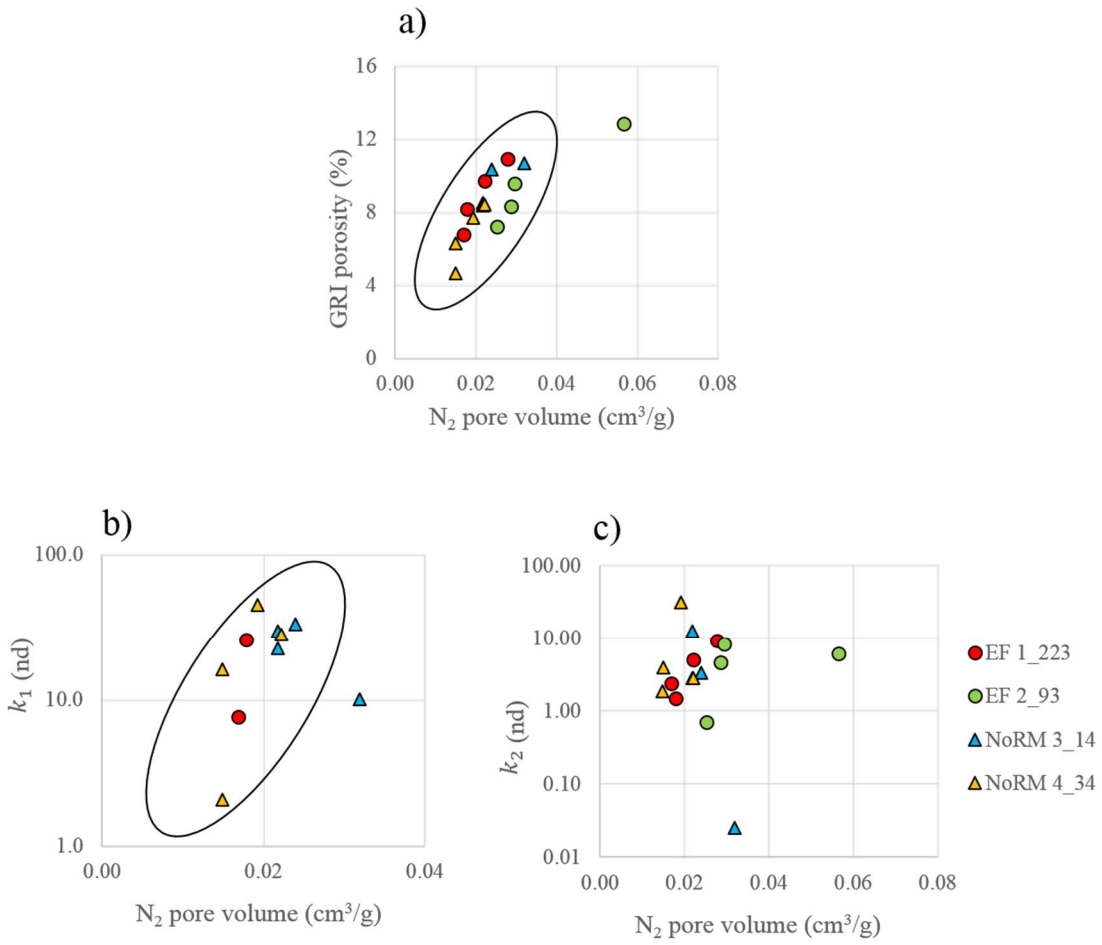

Figure 4.13 Correlation between N_2 pore volume and (a) GRI porosity, (b) permeability k_1, and (c) permeability k_2. The ellipse shows data points with a good correlation. Samples (circles) from the Eagle Ford formation have names starting with 'EF'. Samples (triangles) from the northern Rocky Mountains formation have names starting with 'NoRM'.

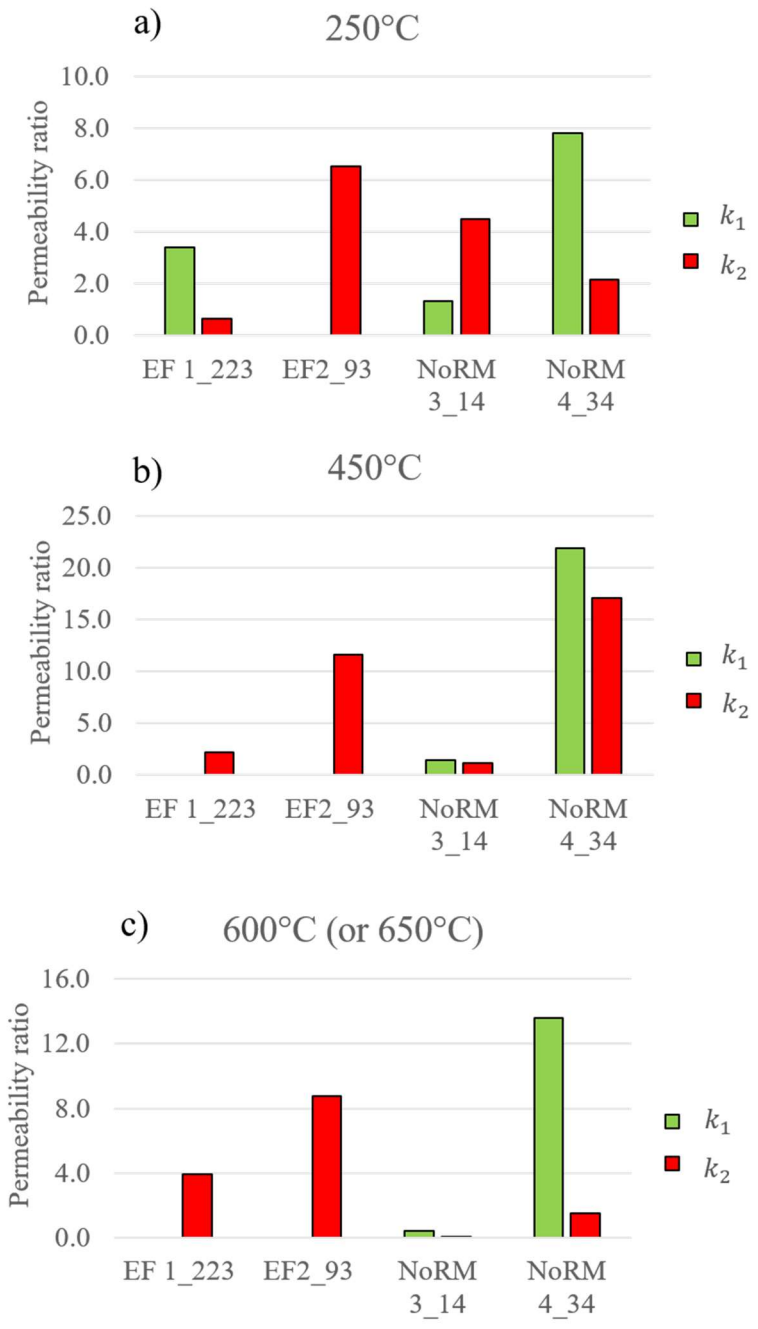

Figure 4.14 Permeability ratios of k_2 (green) and k_2 (red) with respect to 110°C, at (a) 250°C (b) 450°C and (c) 600°C (or 650°C).

4.4.4 Effect of TOC

The plot for N_2 pore volume against TOC (measured on unheated sample; Table 3.2) for each sample is shown in Figure 4.15.

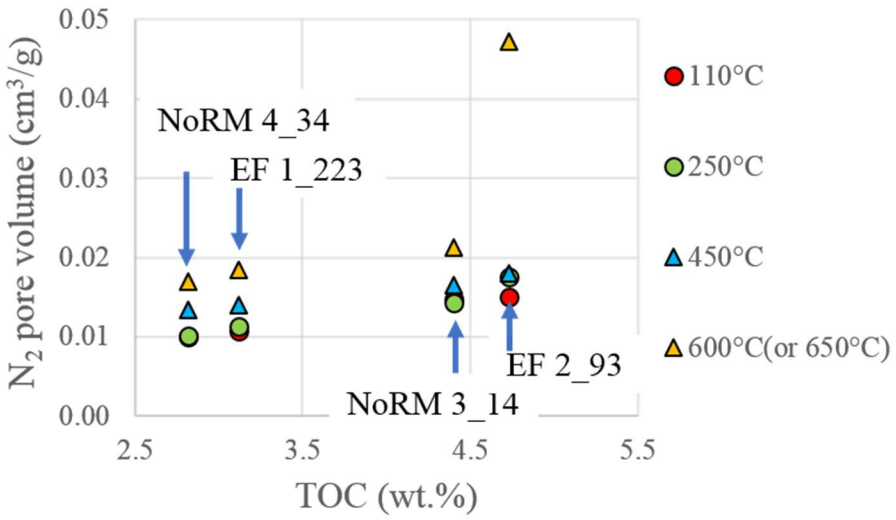

Figure 4.15 Plot of N_2 pore volume after heating at 4 different levels against TOC (Total Carbon Content).

The TOC is positively correlated with the N_2 pore volume, indicating a good correlation between organic matter and pore volume at nanoscale. Impact of organic matter maturation due to heating on the nanoscale pore volume is much higher for higher TOC. Please notice that the NORM 3_14 and EF 2_93 were heated at 650°C instead of 600°C. The higher temperature may also contribute to some amount of increase. Overall, it indicates that organic matter maturation plays an important role in pore volume evolution.

4.5 CONCLUSIONS

Through heating samples at 4 different temperature levels, we observed changes in petrophysical parameters include BET surface area, N_2 pore volume, pore size distribution,

GRI porosity, and permeabilities k_1 and k_2. Those petrophysical properties are sensitive to the heating temperature for the sample preparation. Surface area, N_2 pore volume, GRI porosity are positively correlated with heating temperatures, whereas permeabilities k_1 and k_2 show general increase trends.

Compared to other parameters, GRI permeabilities are strongly affected by heating temperature. The high sensitivity of permeability measurement to drying temperature suggest that caution is warranted in interpreting permeability measurements performed on shales, as the preparation technique appears to have a significant influence on the results.

Two factors contributing to the increase of those petrophyscial properties. The fluids are continuously removed for multiple-stage heating, resulting in an increase of accessible pore space for probe gas. The thermal maturation can also be a source for the increase of pore structure parameters and permeability, especially for temperature above 600°C.

The cross-plots between N_2 pore volume and GRI porosity and permeabilities indicate that the shale matrix at the nanoscale (< 100 nm) is correlated with these larger-scale properties. And the larger pores that contribute to k_1 are well connected, while the smaller ones that contribute more to k_2 appear to be more poorly connected. This indicates the multiscale network features, where a certain fraction of small pores are assumed to be partially or completely disconnected from the overall structure.

Chapter 5: Investigation of Clustering Algorithms for Fluid Characterization Using NMR T_1-T_2 Maps of Organic-rich Shale[3]

5.1 INTRODUCTION

Nuclear magnetic resonance (NMR) has proven to be a powerful technique for characterizing unconventional oil and gas resources. It has been used to estimate important petrophysical quantities such as total porosity, movable-fluid porosity, fluid type, and saturation (e.g. Mullen, 2010; Odusina et al., 2011; Lewis et al., 2013; Tinni et al., 2014). In addition, the advanced 2-D T_1-T_2 measurement has become popular for shale characterization (Washburn and Birdwell, 2013; Daigle et al., 2014; Gips et al., 2014; Fleury and Romero-Sarmiento, 2016; Mehana and El-monier, 2016). The result of the T_1-T_2 measurement is shown as a 2-D T_1-T_2 map, where the fluid volume is a function of T_2 (x-axis) and T_1 (y-axis).

In the T_1-T_2 map, fluid characterization is performed based on the T_1/T_2 ratio (e.g. Daigle et al., 2014; Fleury and Romero-Sarmiento, 2016; Singer et al., 2016; Jiang et al., 2018b). According to Bloembergen-Purcell-Pound (BPP) theory, T_1 and T_2 can be expressed as a function of correlation time (Bloembergen et al., 1947). The correlation time describes the average time for a molecule to rotate one radian, and is controlled by the molecule size and viscosity. For mobile liquids, T_1 and T_2 have a similar value, so their ratio is close to one. For larger molecules and more viscous fluid, the ratio is greater than 1. Under the framework of BPP theory, the characterizing of different fluid populations from the T_1-T_2 maps are done by specific T_1/T_2 ratios. For water in large pores, T_1/T_2 ratio is around 1. T_1/T_2 ratio of hydrocarbon increases due to the increase of the molecular size

[3] This chapter is based on Jiang, H., Daigle, H., Tian, X., Pyrcz, M., Griffith, C., Zhang, B. A Comparison of Clustering Algorithms applied to Fluid Characterization using NMR T1-T2 Maps of Shale. *Computer & Geosciences*. (under review)

and viscosity. The ratio for organic matter (OM; bitumen and kerogen) is even higher (Fleury and Romero-Sarmiento, 2016).

Most fluid characterization is done manually. For example, Fleury and Romero-Sarmiento (2016) defined several elliptical-shaped fluid regions on the T_1-T_2 maps to identify components like water, methane and kerogen (Figure 5.1a). Daigle et al. (2014) introduced a partitioning approach (Figure 5.1b) by projecting the NMR distribution from T_1-T_2 space to T_1/T_2 ratio (denoted as R) versus secular relaxation time (denoted as Ts). They are defined in Equation (2.29) and (2.30). Linear decision boundaries are chosen to cut the new 2-D space into several non-overlapping sub-spaces where each sub-space represents one fluid type (Daigle et al., 2014; Jiang et al., 2018b).

Those manual approaches, however, are often empirical and subjected to human decisions. In the method of Fleury and Romero-Sarmiento (2016), the elliptical-shaped fluid regions are empirically defined based on a few sample measurements, and regions outside the ellipses remain undefined. In the method of Daigle et al. (2014), boundary locations are determined empirically, and the linearity assumption of the boundary is not sufficiently validated. Furthermore, the complex nature of the shale pore/fluid system adds uncertainties for interpretation using manual partitioning methods. Most fluids are seen to have a connected fingerprint in the T_1-T_2 domain, and regularization used in the NMR inversion technique can cause further smoothing of the distribution (Venkataramanan et al., 2018). In a low signal-to-noise scenario (e.g. downhole measurements), differentiation of multiple fluids is even more difficult (Xie and Xiao, 2011). As a result, manual partitioning of the T_1-T_2 distribution may be subjected to great uncertainties, and is not practical in those cases.

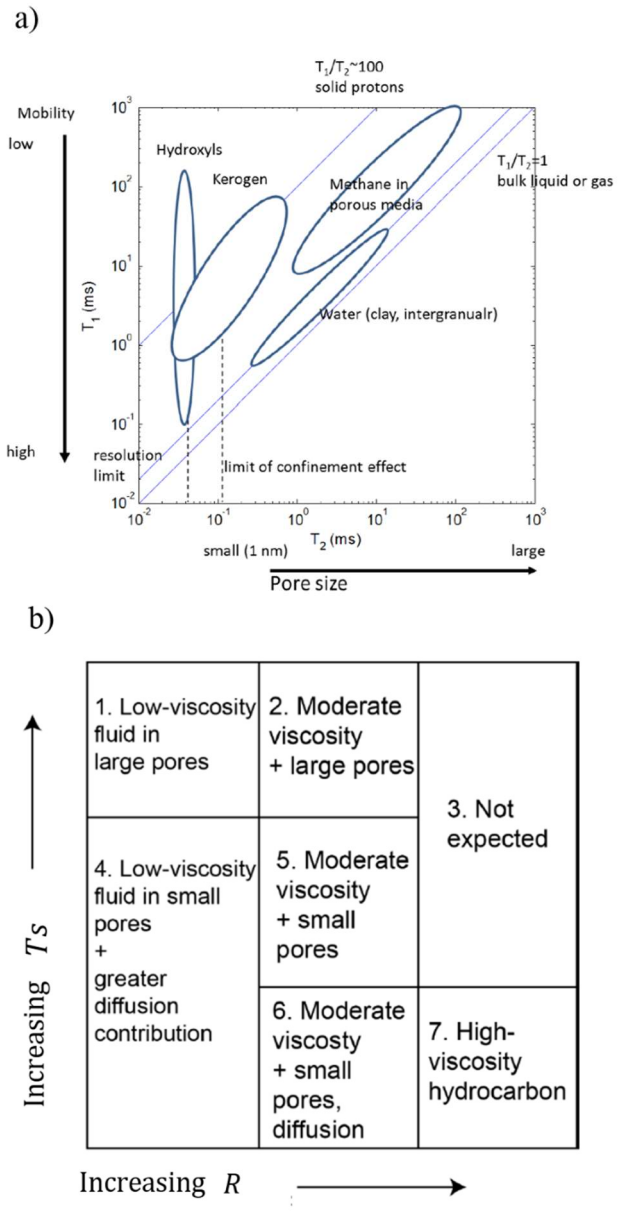

Figure 5.1 Two manual approaches for fluid characterization of NMR T_1-T_2 maps. (a) First approach introduced by Fleury and Romero-Sarmiento (2016). Four expected fluid component regions are defined. The diagonal lines correspond to T_1/T_2 ratios = 1, 2 and 100. Note that the hydroxyl component refers to hydroxyl groups in clay. The subplot is from Fleury and Romero-Sarmiento (2016). (b) Second approach based on R versus Ts. The vertical boundaries correspond to T_1/T_2= 10 and 100. The subplot is modified from Daigle et al. (2014).

Recently, machine learning approaches have been presented for NMR interpretation (e.g. Jain et al., 2013; Birdwell and Washburn, 2015; Venkataramanan et al., 2018). Jain et al. (2013) applied exploratory factor analysis for NMR T_2 logging measurements to identify fluid components from the T_2 distribution. Venkataramanan et al. (2018) proposed an unsupervised method based on blind source separation (BSS) for the T_1-T_2 distribution. The method identified different fluid components for a continuous T_1-T_2-depth log by combining non-negative matrix factorization and a hierarchical clustering method (Venkataramanan et al., 2018). In addition, quantitative correlations between T_1-T_2 maps and organic geochemical properties can be achieved by a partial least-squares regression approach (Birdwell and Washburn, 2015).

In this study, we investigated the performances of different clustering algorithms for fluid characterization using T_1-T_2 maps. Two cluster validity indices were proposed to evaluate the clustering quality, providing quantitative guidance for both choosing the cluster number and selecting the best algorithm. Rather than using continuous logging data, we performed clustering analysis on individual T_1-T_2 maps of shale samples in the as-received state and after drying at 110°C. The drying procedure helps to reveal the footprint from organic matter, allowing identification of low-signal fluids in OM pores (Note: OM is called 'fluid component' in this study for simplicity; Fleury and Romero-Sarmiento, 2016). The best algorithm was selected based on the two validity indices. The results of the best algorithm were further qualitatively validated by comparing the identified fluid centers to those documented in literature. Our work provides a practical guide for applying cluster analysis in fluid characterization in NMR T_1-T_2 core analysis.

5.2 METHODS

The workflow of this study is shown in Figure 5.2. We collected NMR T_1-T_2 data from two organic-rich shales at as-received and dried at 110°C conditions. T_1-T_2 data were processed and fed into 6 different clustering algorithms. Two cluster validity indices were calculated and the best algorithm was selected based on the evaluation. Detailed procedures are described below.

5.2.1 Samples

Samples used in this study are organic-rich shales from the Eagle Ford formation (denoted as EF) and a siliceous formation from the northern Rocky Mountains (denoted as NoRM). Samples include EF 1_223 and NoRM 3_14. Details of the samples can be found in Chapter 3.3. Samples were hand crushed and sieved through multiple trays. About 60 g of particles with mesh size between 20-35 US mesh were collected and dried in an oven at 110°C for 2-4 days.

NMR measurements were performed for both samples immediately after crushing (marked as as-received) and after oven drying at 110°C. All measurements were made at ambient pressure and temperature. We assume that a negligible amount signal was lost due to crushing. Drying in the oven can reveal the signal from organic matter (Fleury and Romero-Sarmiento, 2016), thus helping detecting signals from OM and fluids in OM-hosted pores. The measurements were conducted by an Oxford GeoSpec 2 low-field NMR instrument with an operating frequency of 2.15 MHz. T_1-T_2 data were acquired with inversion recovery steps followed by a Carr-Purcell-Meiboom-Gill (CPMG) sequence with an echo spacing $TE = 0.1$ ms (Carr and Purcell, 1954; Meiboom and Gill, 1958; Singer et al., 2016; Nicot et al., 2016). CPMG is a series of 180° pulses following the 90° oscillating

pulse B_1 in the x-y plane. The echo spacing is the time between the excitation 180° pulses. More details can be found in Chapter 2.5.

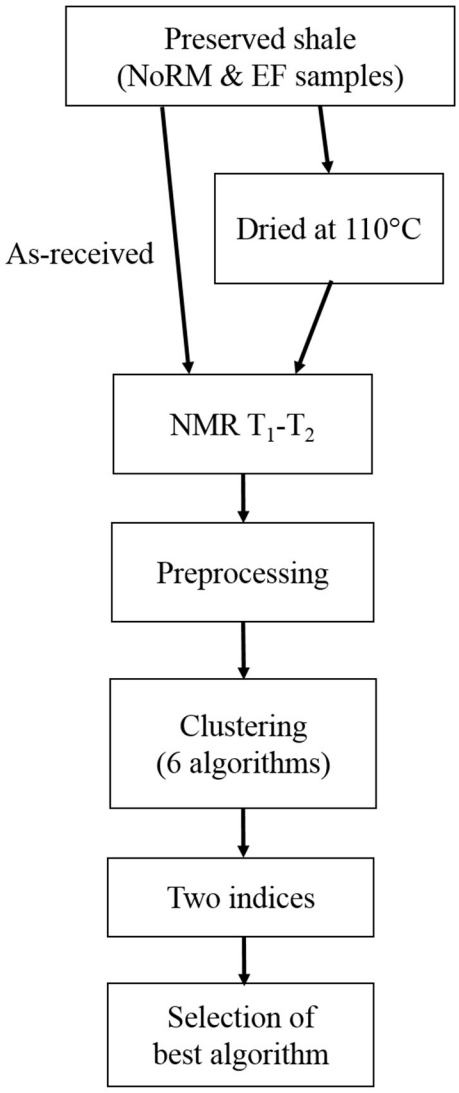

Figure 5.2 Workflow of clustering analysis. NoRM refers to the shales from the northern Rocky Mountains. EF refers to the shales from Eagle Ford.

T_1-T_2 data were processed by an in-house NMR inversion package (Medellín et al., 2015; Medellín et al., 2016). The computed T_1-T_2 distribution is a set of data points. Each

point contains three elements: T_2, T_1, and the fluid volume. The range of T_1 and T_2 are from 0.01 ms to 1000 ms.

5.2.2 Data Preprocessing

A manifold for the T_1-T_2 map needs to be defined, on which the clustering algorithm is performed (Venkataramanan et al., 2018). A sub-space of the T_1-T_2 domain is obtained by ignoring the region where T_1/T_2 ratio is less than 0.5, as the theoretical T_1/T_2 value is anticipated to be equal or larger than 1 according to BPP theory (Bloembergen et al., 1947). Since T_1 and T_2 values span several orders of magnitude, a logarithmic transform is applied, which converts heavily skewed data to a more symmetric distribution (Templ et al., 2008).

Fluid volume f is the third dimension of the distribution data (T_1, T_2, f). It is the density function of T_1-T_2, which represents the abundance (frequency) for a grid point (T_1, T_2) in terms of volume. The clustering requires converting the 3-D data to a 2-D T_1-T_2 distribution, where the frequency is represented by the number of points on the grid (T_1, T_2). To do that, f is normalized by the following equation:

$$f_{norm} = round\left(\frac{f(T_1, T_2)}{f_c}\right), \quad (5.1)$$

where f_{norm} is the normalized fluid volume and f_c is a threshold that represents the minimum fluid volume to be considered for the manifold.

The threshold f_c is similar to the normalization threshold coefficient τ_1 defined in Venkataramanan et al. (2018), and the choice of its value reflects the trade-off between noise variance and bias. When the threshold is large, gird points with information of the fluids may be erroneously omitted. When the threshold is small, grid points close to or less

than the noise level may be erroneously included. Through trial and error, 10% of the maximum fluid volume is chosen as f_c in this study.

Through this normalization, a fluid volume greater than f_c is converted to an integer that is greater or equal to 1. Fluid volumes smaller than f_c are ignored. Each grid point is then duplicated f_{norm} times to produce the 2-D dataset for clustering.

5.2.3 Clustering

Clustering algorithms can be broadly divided into hierarchical and partitional method at the top level (Figure 5.3; Jain et al., 1999; Jain, 2010). In the hierarchical method, each observation starts with itself as a cluster, and clusters are successively merged together to form larger clusters. The algorithm recursively produces a nested series of partitions. The partitional method, on the other hand, produces all the partitions at the same time without imposing the hierarchical structures.

Six commonly used algorithms are selected (Figure 5.3). Three algorithms are from the family of hierarchical methods (Ward linkage, complete linkage and balanced iterative reducing and clustering using hierarchies or BIRCH). Three algorithms are from partitional methods (k-means, Gaussian mixture model or GMM, and spectral clustering). More details about clustering algorithms can be found in Chapter 2.6.

Clustering is performed on the T_1-T_2 manifold using scikit-learn, an open source python machine learning library (Pedregosa et al., 2011). The T_1-T_2 dataset is fed into the selected clustering model. The model starts the training process by iteratively learning the cluster model parameters until some convergence criteria is met (Jain et al., 1999). After completing the training process, the trained model generates the cluster label for each T_1-T_2 pair. More details of clustering using scikit-learn can be found in its documentation

(Pedregosa et al., 2011). Details of clustering based on python can be found in Appendix B.

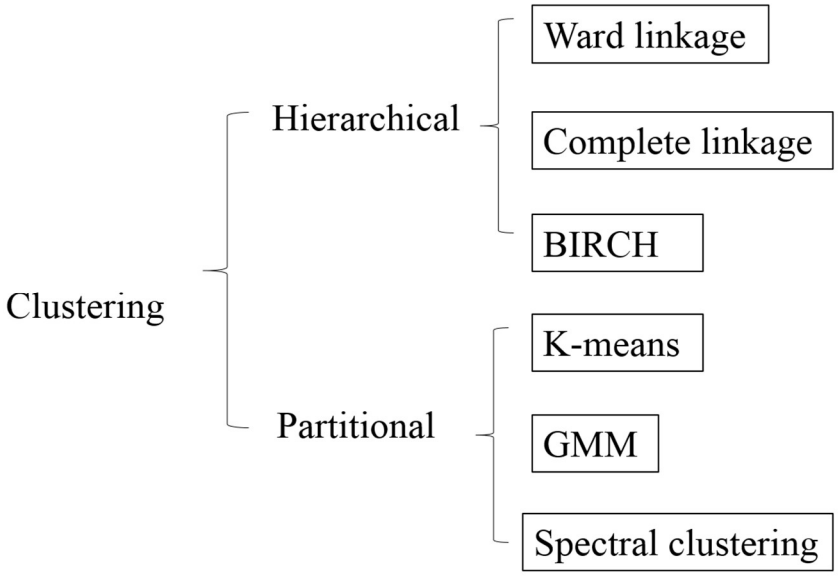

Figure 5.3 Taxonomy of clustering methods. Clustering names in boxes are methods used in this study. DBSCAN is short for density-based spatial clustering of applications with noise. GMM is short for Gaussian mixture model. BIRCH is short for balanced iterative reducing and clustering using hierarchies.

5.2.4 Evaluation of Clustering

Evaluating cluster quality is essential since any clustering algorithm will produce several different results for every dataset (Templ et al., 2018). The validity measures should support the decision for selecting the optimal cluster number. More importantly, they should also provide a value for judging the quality of the clustering result, so that the best clustering algorithm can be selected.

Here, we propose two validity indices for the T_1-T_2 fluid characterization problem. These indices are based on the assumption from BPP theory that individual fluid populations will be characterized by a single value or narrow range of T_1/T_2 ratio. One

index measures the quality of the single cluster, and the other measures the quality between different clusters.

5.2.4.1 Range Ratio

The first index is the T_1/T_2 ratio range (RR). The expression is defined below:

$$RR = \frac{max\left(\frac{T_1}{T_2}\right)}{min\left(\frac{T_1}{T_2}\right)}, (5.2)$$

where $max(T_1/T_2)$ is the maximum T_1/T_2 ratio for grid points in one cluster. $min(T_1/T_2)$ is the mimimum T_1/T_2 ratio for the same cluster. Since the ratio spans several orders of magnitude, the maximum and minimum T_1/T_2 ratios are divided rather than subtracted.

RR tries to measure the variation of T_1/T_2 ratio within the same cluster. A fluid with similar compositions should exhibit similar properties like viscosity and molecular size, resulting in a similar T_1/T_2 ratios on the T_1-T_2 map. Thus, a good cluster, which can represent a single fluid population, should have a small variation of T_1/T_2 ratios over its grid point population (Figure 5.4a). This means that a small RR is preferable, and a cluster with low quality will have a large RR (Figure 5.4b).

5.2.4.2 Angle Difference

The second index is the angle difference (AD), and is expressed below:

$$AD = |\alpha - 45°|, (5.3)$$

where α is expressed in degrees, and is the slope of the two-fluid partitioning boundary by linear regression in $\log(T_2)$ - $\log(T_1)$ space.

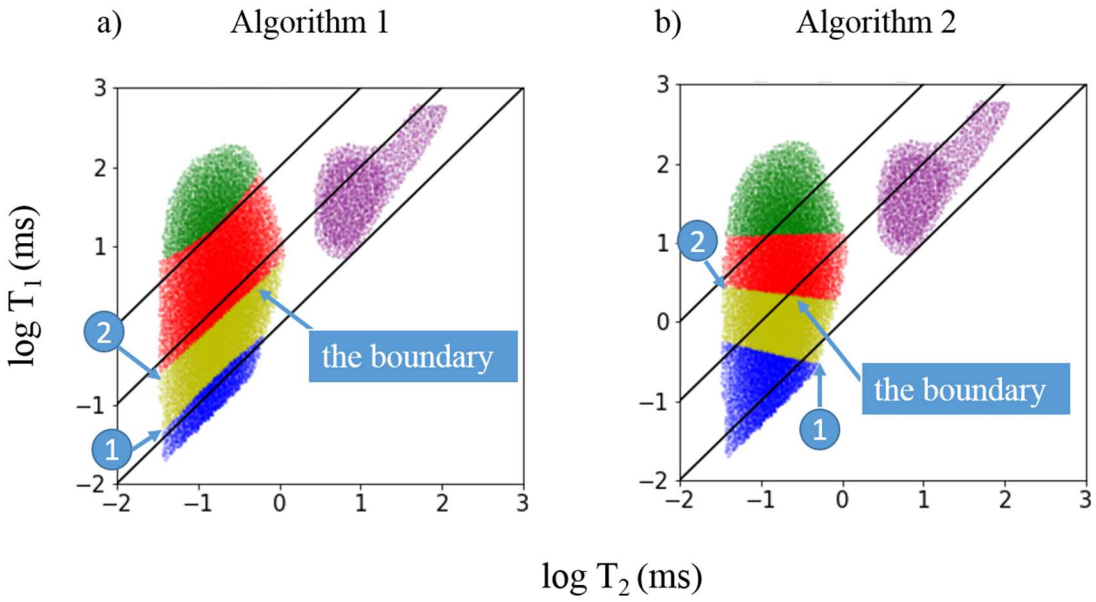

Figure 5.4 Illustration for (a) good clustering quality and (b) low clustering quality. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100. The circle labeled number 1 represents the smallest T_1/T_2 ratio for the yellow cluster. The circle labeled number 2 represents the largest T_1/T_2 ratio for the same yellow cluster.

We use grid points that are close to or on the boundary to represent the cluster boundary. Since all our algorithms apply hard assignment, this means each grid point can only belong to one cluster. In addition, if a grid point α_i from Cluster α is close to the boundary of Cluster α and Cluster β, its neighbors within a critical distance should have a least one point from Cluster β. If α_i doesn't contains points from β within the critical distance, α_i is not on the boundary. If, one the other hand, there is more than one point from β within the critical distance, the point with the minimum distance to α_i (denoted as β_i) , as well as α_i are considered as points on the boundary. The critical distance will have effects on the width of the boundary. A large critical distance will generate a wide boundary, and a small critical distance will generate a narrow boundary. In this work, we chose 0.05 as the critical distance, which yields a reasonable result (Figure 5.5). The

outcome is a set of points that are at the close or on the boundary as shown in Figure 5.5. Applying linear regression for the boundary points, we can obtain its slope. Based on this, we developed an algorithm to extract the boundary grid points. The pseudocode is shown below:

Algorithm: extract grid points for two-fluid boundary

Input: cluster α, and cluster β, critical distance d_c
Output: boundary points B
compute the number of grid points (n_α, n_β) of α and β
initiate B
for $i = 0$ to n_α **do**
for $j = 0$ to n_β **do**
$d_j = \sqrt{(\alpha_{i,0} - \beta_{j,0})^2 + (\alpha_{i,1} - \beta_{j,1})^2}$
$d_{min} \leftarrow$ minimum of d_j for all j
if $d_{min} < d_c$ **then**
$k \leftarrow$ the index for minimum d_j
add α_i and β_k to B
remove duplicate points of α_i or β_k in B

AD tries to measure the angle deviation of the fluid partitioning boundary to a theoretical fluid boundary line with a constant T_1/T_2 ratio. When partitioning fluids with different characteristic T_1-T_2 ratios on the maps, a good fluid partitioning boundary should have a slope which is roughly close to a constant T_1/T_2 ratio (Figure 5.4). This results in a small difference between the slopes of the partitioning boundary and 45°, and thus a smaller AD indicates a better partitioning boundary.

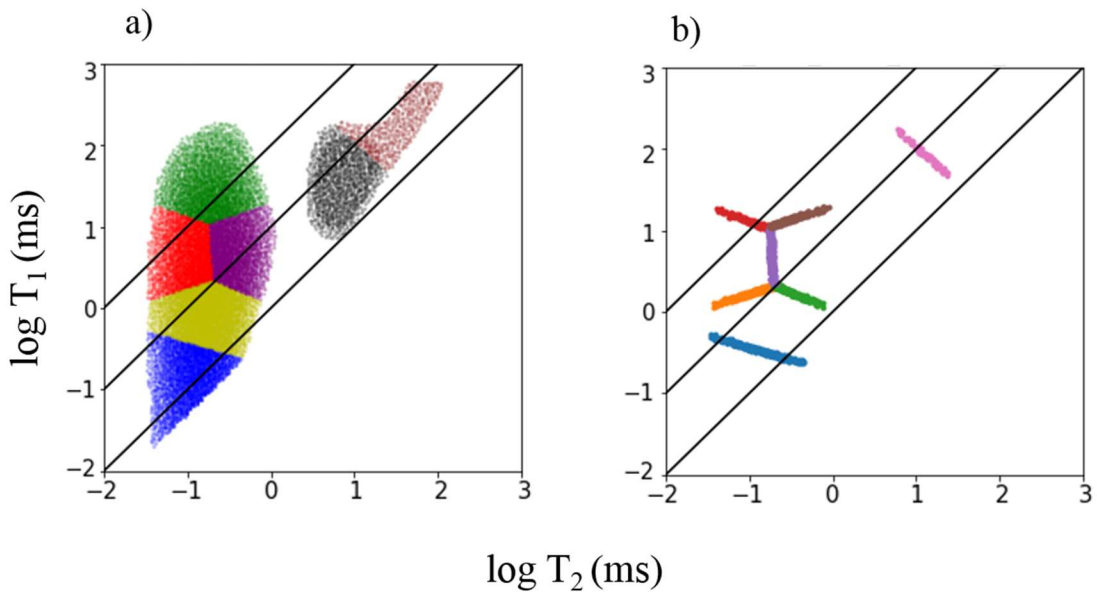

Figure 5.5 (a) Clustering result given 7 cluster numbers. Cluster labels are denoted by different colors. (b) Boundary points extracted by the algorithm. Different color denote points belong to different boundaries.

5.3 RESULTS

Clustering were performed on T_1-T_2 maps of shale sample NoRM 3_14 and EF 1_223 at as-received and dried at 110°C conditions.

5.3.1 T_1-T_2 Maps

The T_1-T_2 maps for sample NoRM 3_14 at as-received and dried at 110°C conditions are depicted in Figure 5.6. The sample at as-received conditions (Figure 5.6a) has a total fluid volume of 2.87 cm^3. It displays a large volume of signal from $T_2 < 1$ ms region, and expands along the T_1 axis. The large variation of T_1/T_2 ratios indicates a mixture of multiple fluid components. In addition, a distinct distribution is found in the $T_2 > 1$ ms region. After drying at 110°C (Figure 5.6b), the sample reduces its total fluid volume to 1.31 cm^3. Most signal from the region of $T_2 > 1$ ms disappears.

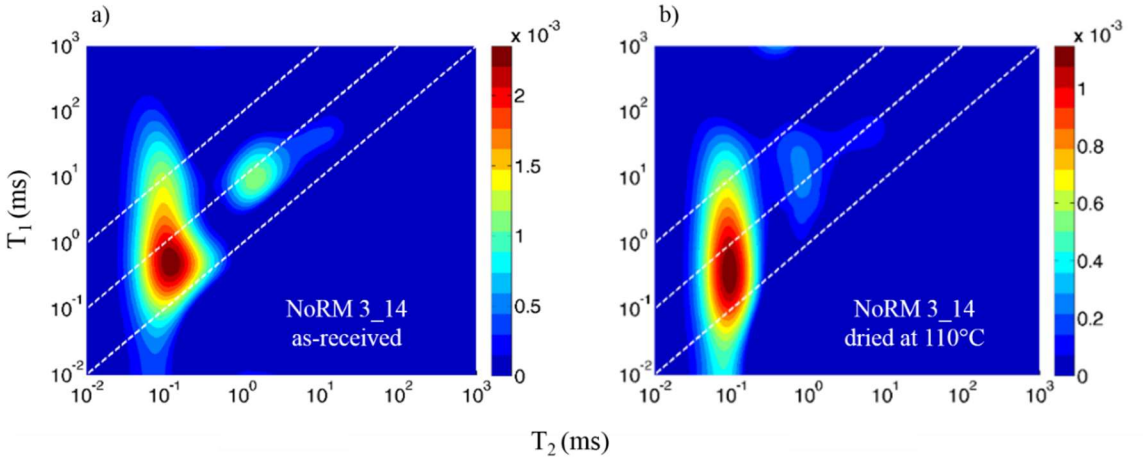

Figure 5.6 T_1-T_2 map for siliceous shale sample NoRM 3_14 (a) at as-received and (b) dried at 110°C conditions. The warmer colors indicate greater pore volume. The units of pore volume are cm^3. T_1 and T_2 both range from 10^{-2} ms to 10^3 ms. Three dashed lines correspond to T_1/T_2 ratio = 1, 10, 100. NoRM refers to the shales from the northern Rocky Mountains.

The T_1-T_2 maps for sample EF 1_223 at as-received and dried at 110°C conditions are depicted in Figure 5.7. Sample EF 1_223 has a total fluid volume of 2.86 cm^3 for its as-received conditions (Figure 5.7a). Most of the fluid volume is located in the region of $T_2 > 1$ ms, forming an elliptical shape. The center of the ellipse has a T_1/T_2 ratio close to 10. The distribution extends to the $T_2 < 1$ ms region with a tail-like shape. The peak volume of the tail has a T_1/T_2 ratio around 8.

After drying at 110°C, the total fluid volume of EF 1_223 is greatly reduced to 0.51 cm^3 due to fluid evaporation. As depicted in Figure 5.7b, most of the signal from the region of $T_2 > 1$ ms disappears. The peak fluid volume shifts to smaller T_2 region, with the shape widely extending along T_1 axis. The signal of organic matter, suppressed at as-received conditions, is anticipated to be more prevalent.

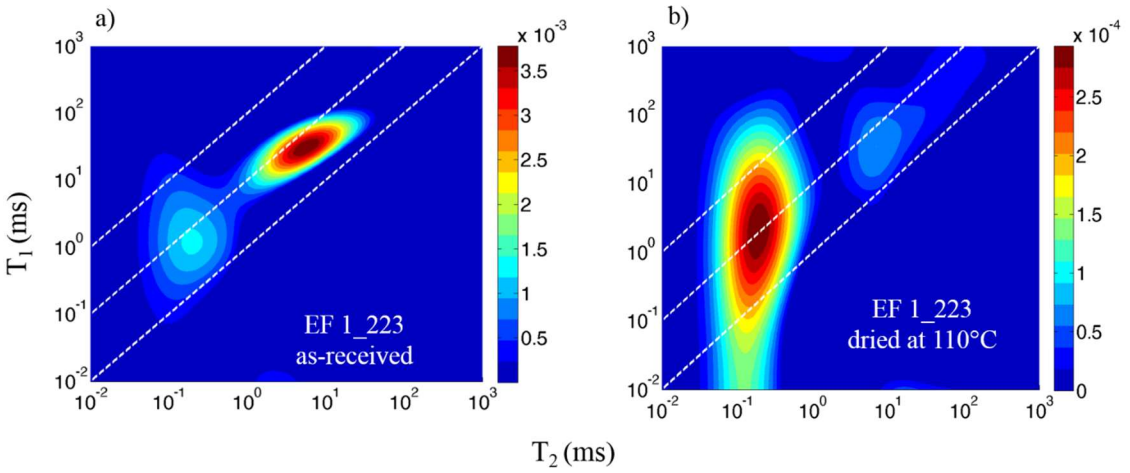

Figure 5.7 T_1-T_2 map of calcareous shale sample EF 1_223 (a) at as-received and (b) dried at 110°C conditions. The warmer colors indicate greater pore volume. The units of pore volume are cm^3. T_1 and T_2 both range from 10^{-2} ms to 10^3 ms. Three dashed lines correspond to T_1/T_2 ratio = 1, 10, 100. EF refers to the shales from Eagle Ford.

5.3.2 Clustering Results

Six clustering algorithms were applied to the 2 shale samples at both as-received and dried at 110°C conditions. We started with a small value of cluster number and successively increased the cluster number until 6-7. A large cluster number should be avoided as it can cause difficulty in interpretation of individual clusters.

We plotted the two validity indices RR and AD against the number of clusters for each algorithm. For each cluster number, there are multiple cluster groups and thus multiple values of the index, so the highest index value is selected to represent the cluster quality. Small values of both indices are preferable. The optimal cluster number can be selected according to the graph. In addition, the best clustering algorithm can also be selected by comparing the different plots.

5.3.2.1 Clustering for NoRM 3_14 at As-received Conditions

The results of the two validity indices of NoRM 3_14 at as-received conditions are shown in Figure 5.8 and Figure 5.9. Figure 5.8 shows plots of the two validity indices against the number of clusters of algorithms including BIRCH, complete linkage and GMM. Figure 5.9 shows plots of the other 3 algorithms (k-means, spectral clustering and Ward linkage). GMM displays a clear minimum of RR at a cluster number of 4. The small RR value indicates that the partitioning clusters given that cluster number have narrow ranges of T_1/T_2 ratios. One the other hand, the other 5 methods show a decreasing trend against the cluster number, implying that a larger cluster number is preferable.

For the second validity index AD, GMM shows a minimum AD value for a cluster number of 4. The small AD value indicates that those boundaries have good alignments with the T_1/T_2 ratio line. The other methods, however, have at least one large AD value when cluster number is greater than 3, implying that the boundaries are not well aligned to the T_1/T_2 ratio line at large cluster number.

The analysis of RR and AD concludes that GMM with cluster number of 4 yields the best partitioning result, whereas other algorithms fail to provide a good quality of partitioning.

Figure 5.10 shows clustering results of the 6 algorithms with cluster number of 4. Partitioning methods yield smooth boundaries whereas hierarchical methods generate non-smooth boundaries. On the other hand, the hierarchical methods start the process by treating the individual point as one cluster, then progressively reduce cluster number by merging two small clusters into one cluster. This iterative merging of clusters generates non-linear boundaries. Figure 5.10 show that GMM gives the best partitioning results, which visually validate the capability of cluster validity indices. The two decision

boundaries of GMM almost parallel to the reference T_1/T_2 ratio lines. Decision boundaries of other algorithms are not consistent to the reference ratio lines.

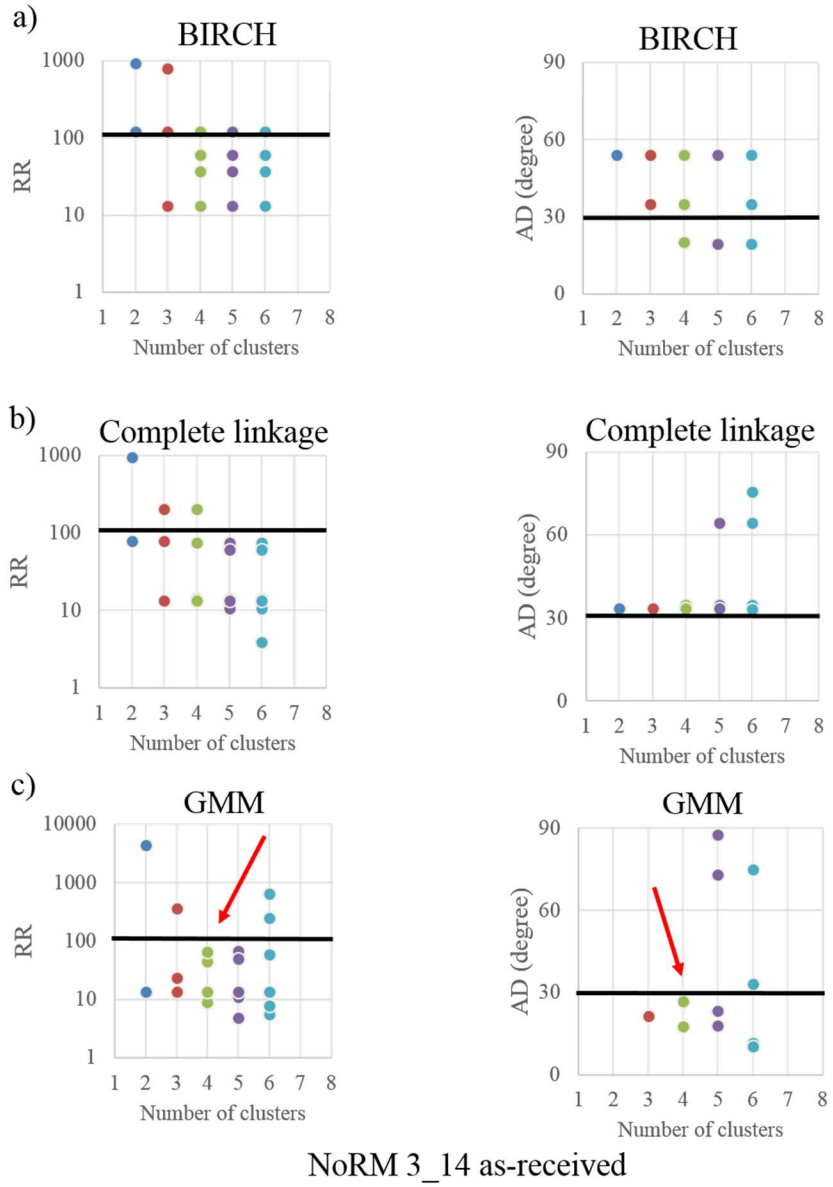

Figure 5.8 Two cluster validity indices RR and AD given different number of clusters for sample NoRM 3_14 at as-received conditions: Algorithms include (a) BIRCH, (b) complete linkage, and (c) GMM. The number of clusters is from 2 to 6. Small values of RR and AD indicate better clustering performance. The optimal clustering number is 4 for GMM as indicated by red arrows.

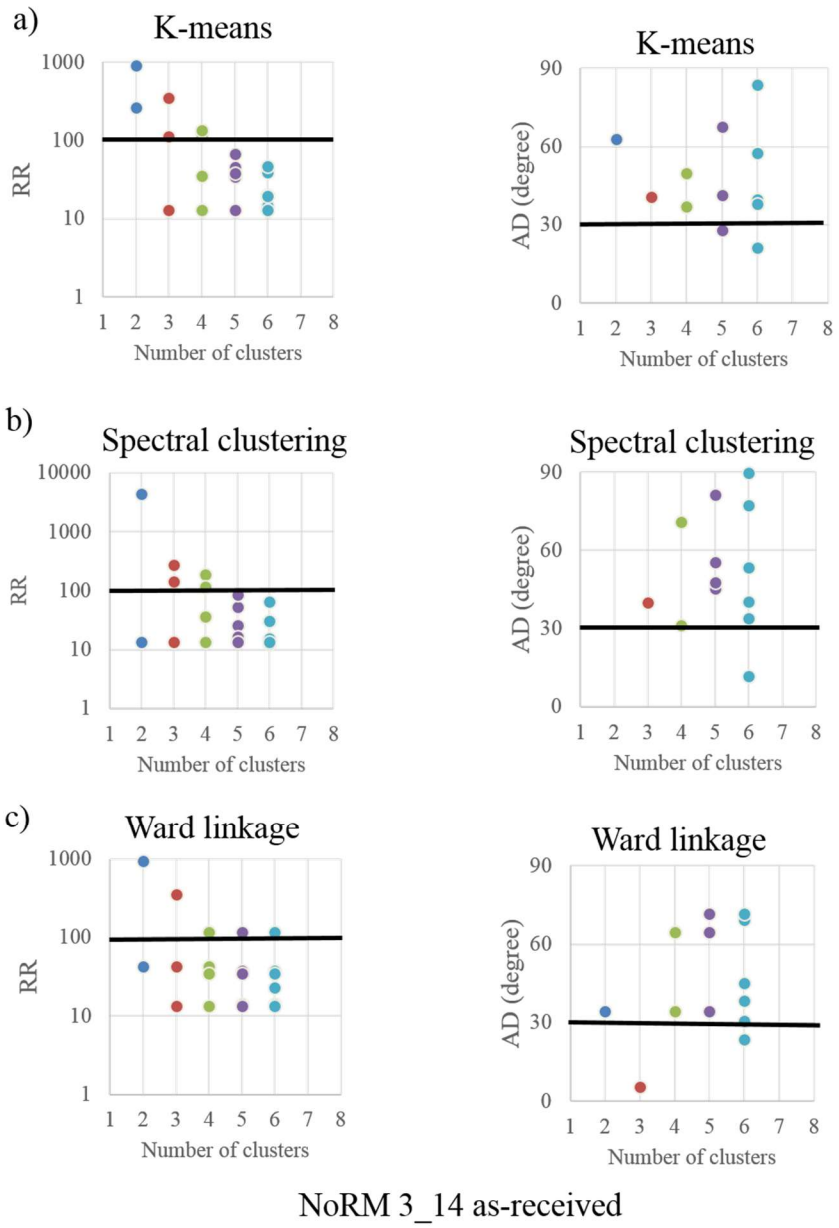

Figure 5.9 Two cluster validity indices RR and AD given different number of clusters for sample NoRM 3_14 at as-received conditions: Algorithms include (a) k-means, (b) spectral clustering, and (c) Ward linkage. The number of clusters is from 2 to 6. Small values of RR and AD indicate better clustering performance.

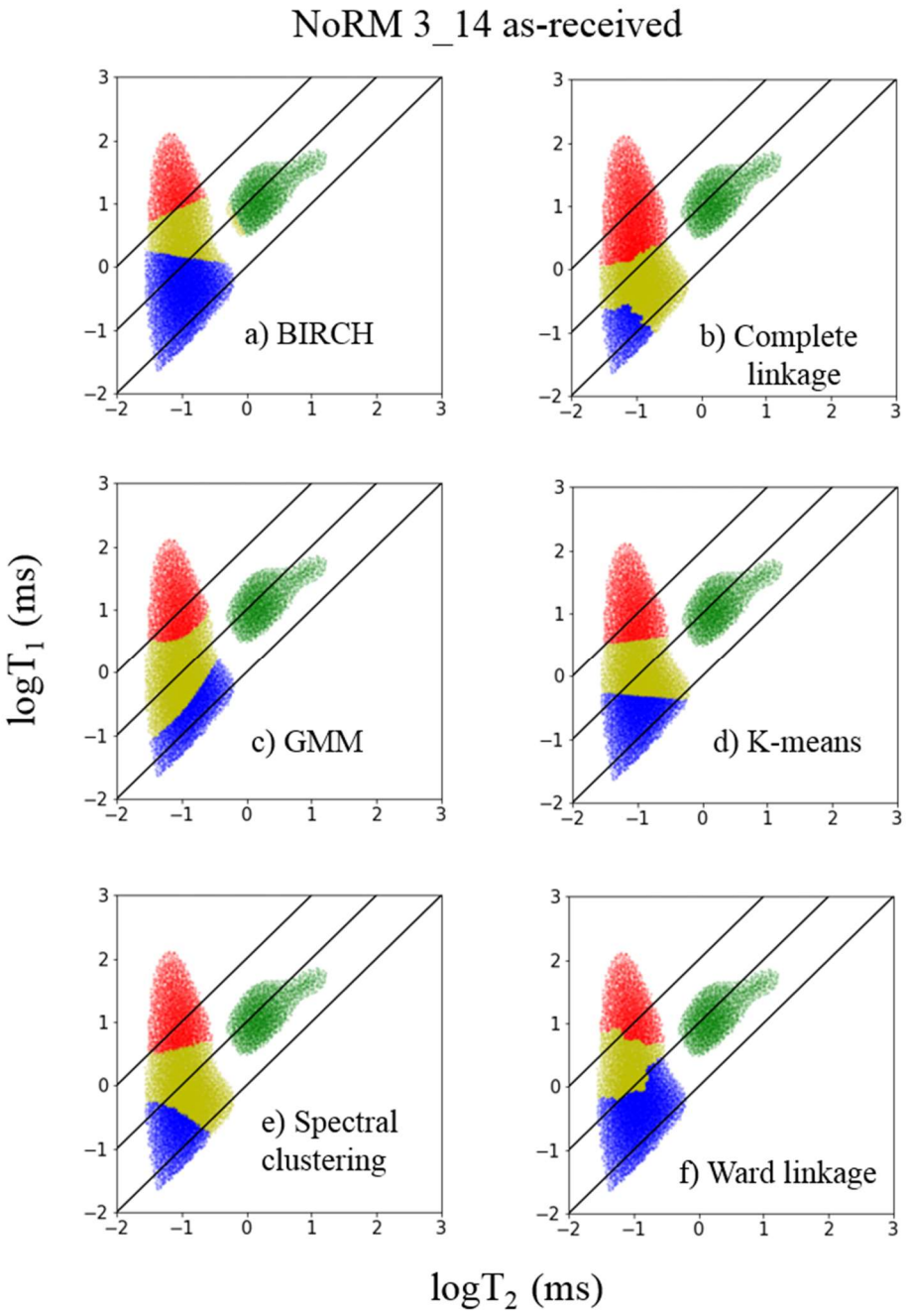

Figure 5.10 Clustering results of 6 algorithms for NoRM 3_14 at as-received conditions. The cluster number is 4. The results are plotted on log-log scales. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100.

5.3.2.2 Clustering for NoRM 3_14 and EF 1_223 at Dried at 110°C Conditions

The two indices are computed for samples NoRM 3_14 and EF 1_223 dried at 110°C conditions against the number of clusters. Based on the same evaluation procedure, GMM with an optimal cluster number of 5 produces the most suitable partitioning results, as predicted by two indices (Figure 5.11 and Figure 5.12). Other algorithms fail to generate good partitioning results. Clustering results on the T_1-T_2 map with 5 clusters are shown in Figures 5.13 and 5.14 for NoRM 3_14 and EF 1_223, which visually validates the conclusion from the two indices.

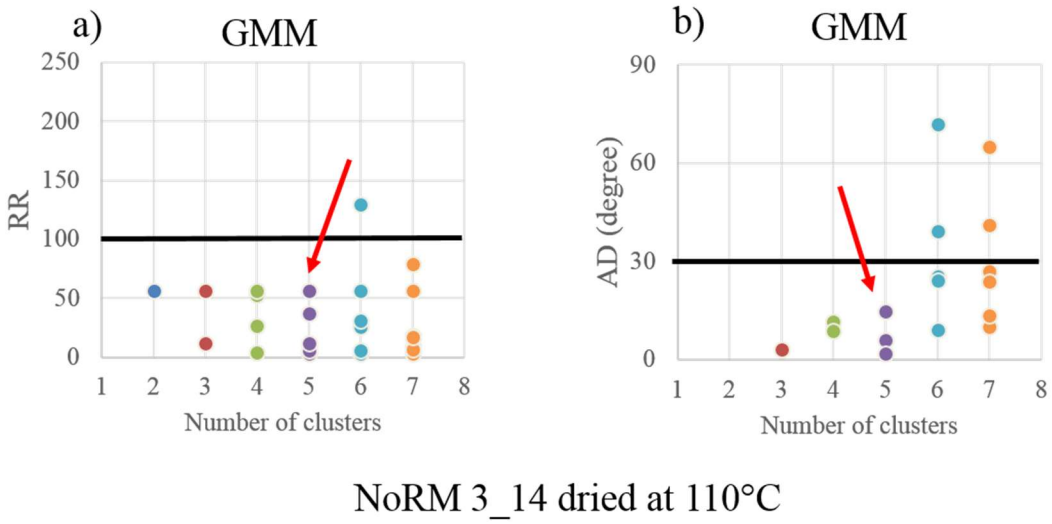

Figure 5.11 Two cluster validity indices (a) ratio range (RR) and (b) angle difference (AD) of GMM for NoRM 3_14 dried at 110°C. The number of clusters is from 2 to 7. Small values of RR and AD indicate better clustering performance. The optimal clustering number is 5 for GMM method as indicated by red arrows.

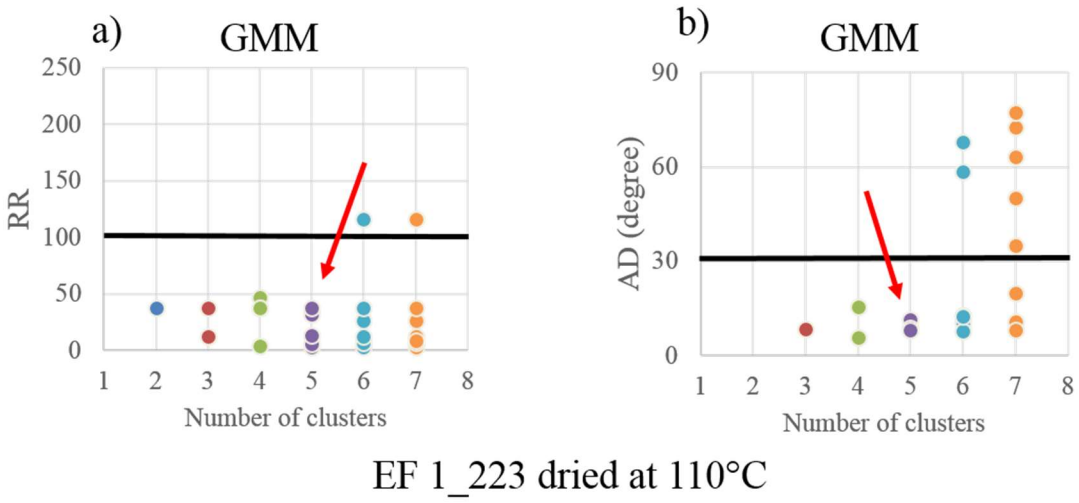

Figure 5.12 Two cluster validity indices (a) ratio range (RR) and (b) angle difference (AD) of GMM for EF 1_223 dried at 110°C conditions. The number of clusters is from 2 to 7. Small values of RR and AD indicate better clustering performance. The optimal clustering number is 5 for GMM method as indicated by red arrows.

5.3.2.3 Clustering for EF 1_223 at As-received Conditions

Clustering is applied for EF 1_223 at as-received conditions. The results, however, don't yield good performance even with GMM, according the two indices in Figure 5.15. The partitioning on the T_1-T_2 map (Figure 5.16) suggests that the GMM cluster boundaries don't well align with the reference T_1/T_2 ratio lines. Other methods fail as well. All 6 algorithms try to create sub-partitions in the large T_2 region, where one fluid component is anticipated due to its narrow T_1/T_2 ratio range.

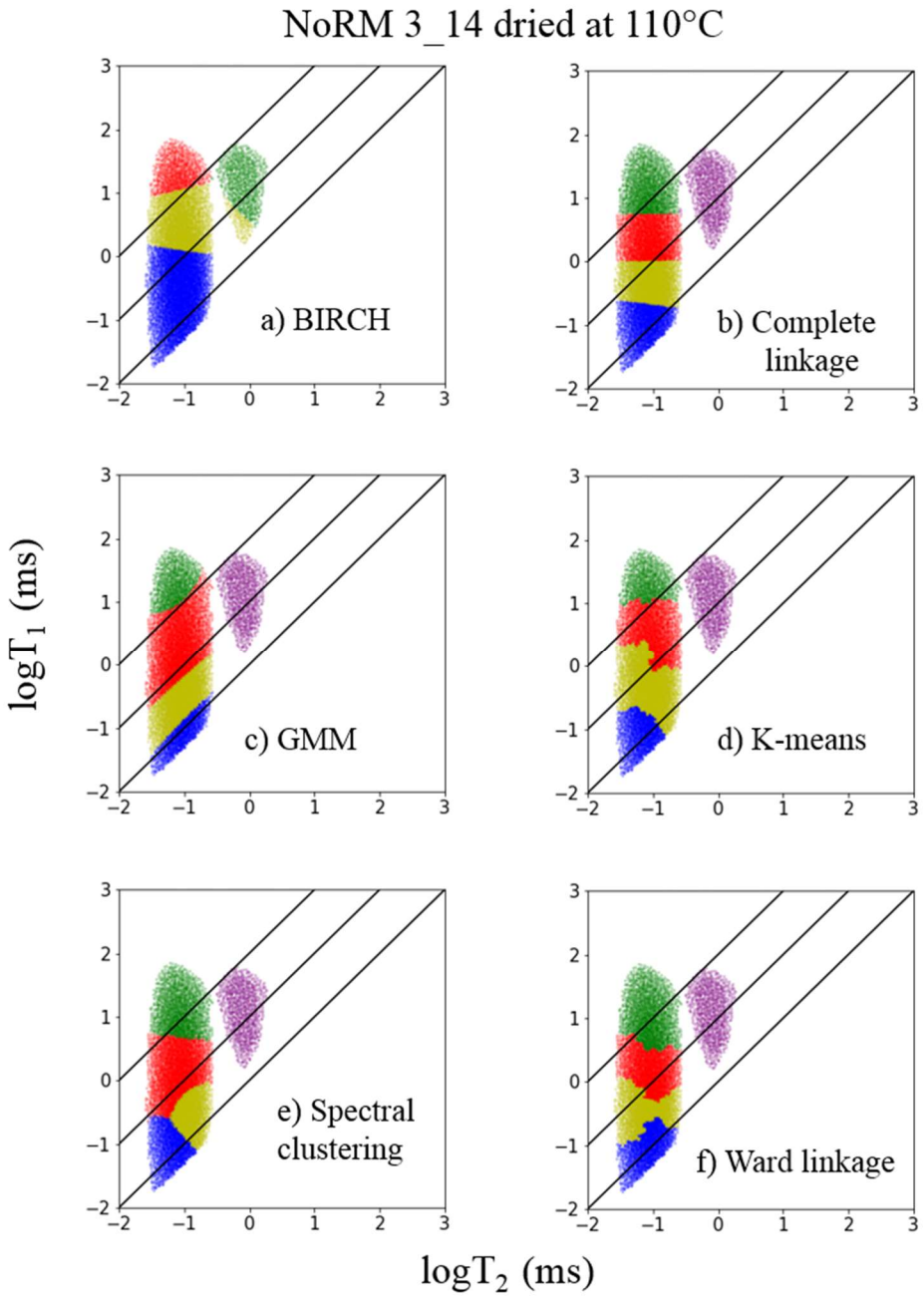

Figure 5.13 Clustering results of 6 algorithms for NoRM 3_14 dried at 110°C conditions. The cluster number is 5. The results are plotted on log-log scales. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100.

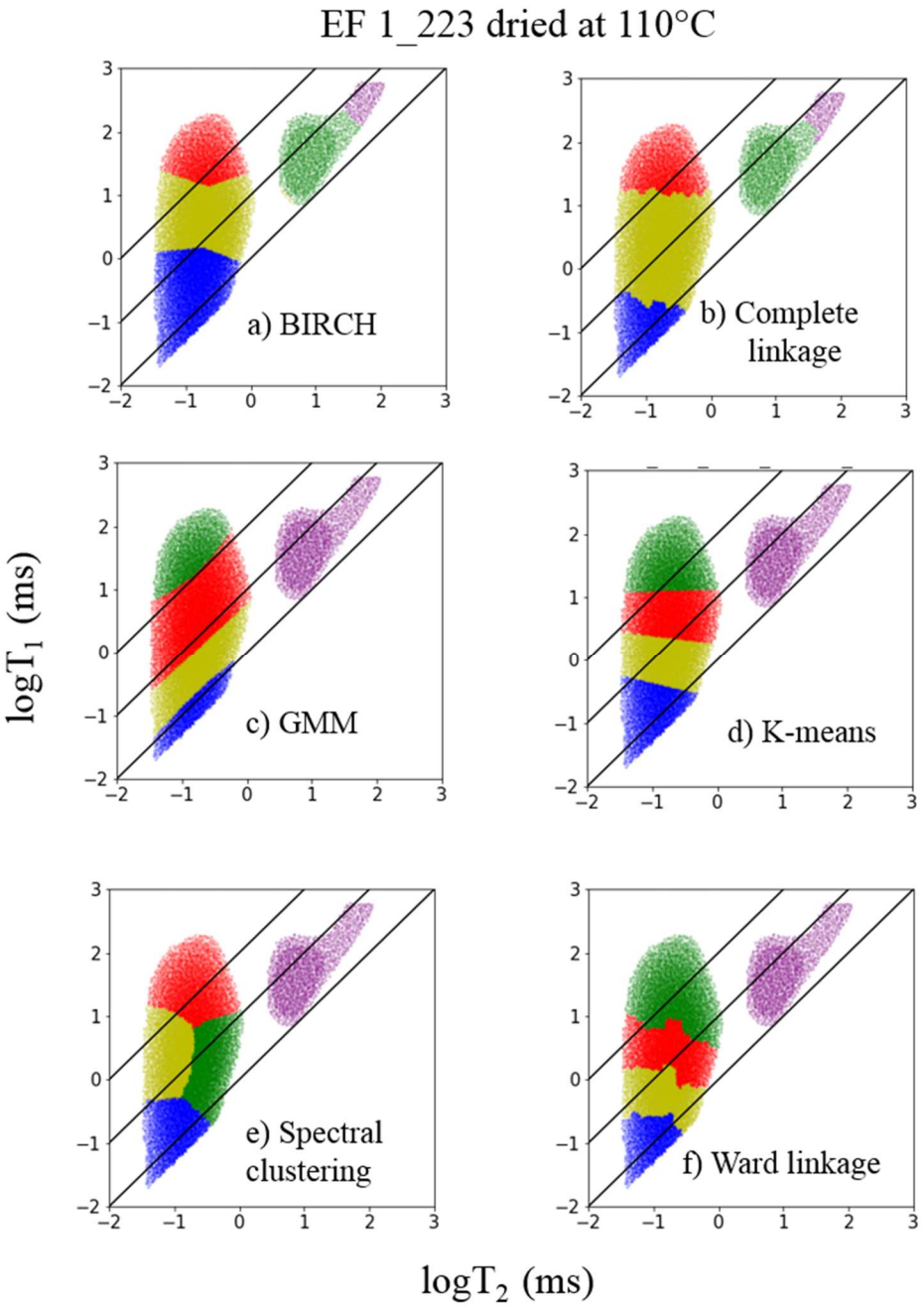

Figure 5.14 Clustering results of 6 algorithms for EF 2_93 dried at 110°C conditions. The cluster number is 5. The results are plotted on log-log scales. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100.

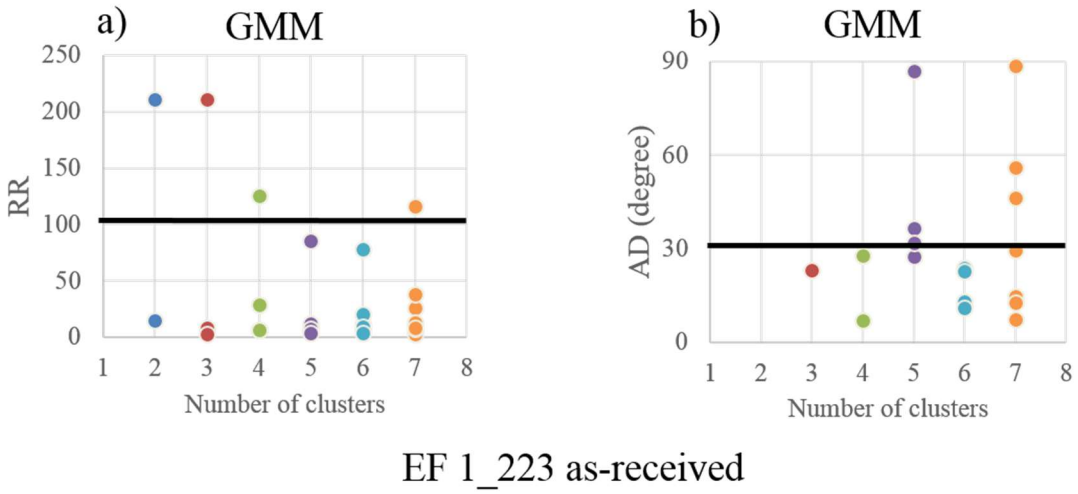

Figure 5.15 Two cluster validity indices (a) ratio range (RR) and (b) angle difference (AD) for GMM given different number of clusters for sample EF 1_223 at as-received conditions. The number of clusters is from 2 to 7. Small values of ratio range and angle difference indicate better clustering performance.

Unlike NoRM 3_14, sample EF 1_223 at as-received conditions is dominated by fluid in the larger T_2 region (Figure 5.7a). The footprint of kerogen and bitumen in the small T_2 region is not clearly observable, which may be suppressed by the dominant fluid signal. During the clustering training process, the optimization algorithm may be sensitive to the large signal intensity in the large T_2 region, causing sub-partitions in that region. This results in a relatively overall low quality of GMM partitioning.

In summary, the proposed two validity indices based on BPP theory yield good quantitative estimation of cluster quality, providing a practical guide for selecting cluster number and best algorithm. In most cases, GMM with cluster number to be 4-5 is the best algorithm, which shows the highest consistence respect to the reference fluid boundaries. Other algorithms, however, fail to provide good clustering based on the two indices and partitioning plot on T_1-T_2 maps.

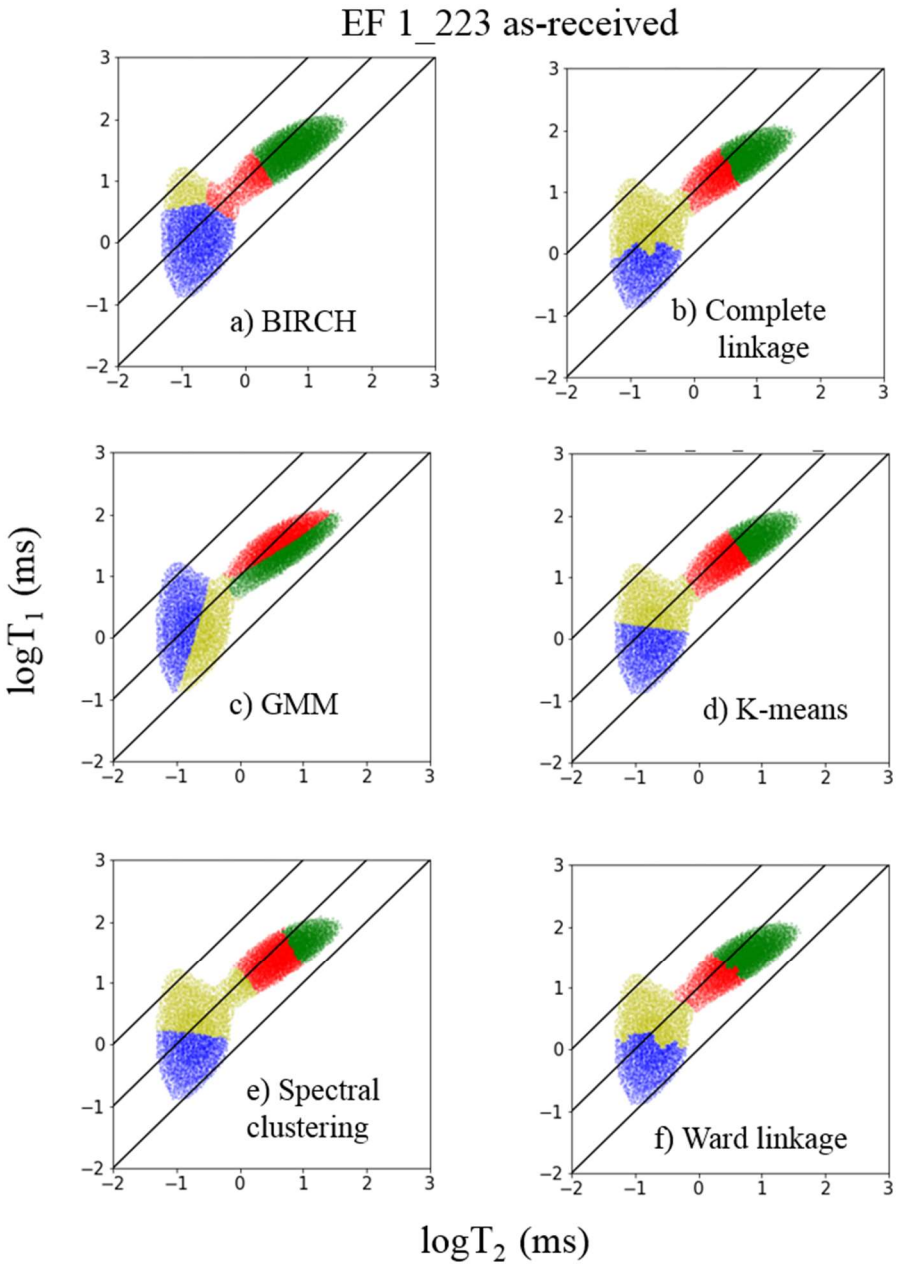

Figure 5.16 Clustering results of 6 algorithms for EF 1_223 at as-received conditions. The cluster number is 4. The results are plotted on log-log scales. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100.

5.4 Interpretation of Fluid Populations

5.4.1 Theory of GMM

We first introduce the theory of GMM. GMM belongs to one type of clustering approach called model-based clustering (Jain et al., 1999; Aggarwal and Reddy, 2013). The model-based clustering assumes that data is generated from a mixture of components, where each component is described by a probability model. In GMM, the Gaussian probability distribution is used (Day, 1969; McLachlan et al., 1999; Fraley and Raftery, 2006).

For a 2-D case, the Gaussian distribution function is expressed as:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{V}|}} exp\left[-\frac{1}{2}(\boldsymbol{x}-\mu)^T\boldsymbol{V}^{-1}(\boldsymbol{x}-\mu)\right], (5.4)$$

where $\boldsymbol{V}$ is the 2-D covariance matrix. $|\boldsymbol{V}|$ is its determinant, $\boldsymbol{x}$ is the two-dimensional variable, which contains two features T_2, T_1, and μ is the 2-D mean vector $(\overline{T_2}, \overline{T_1})$.

The 2-D covariance matrix $\boldsymbol{V}$ is defined:

$$\boldsymbol{V}(T_2, T_1) = \begin{bmatrix} VAR(T_2) & COV(T_2, T_1) \\ COV(T_1, T_2) & VAR(T_1) \end{bmatrix}, (5.5)$$

where the variance VAR and covariance COV is defined as:

$$VAR(X) = \sum_{i=1}^{N} \frac{(X_i - \bar{X})(X_i - \bar{X})}{N}, (5.6)$$

$$COV(X, Y) = \sum_{i=1}^{N} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{N}, (5.7)$$

where N is the total number of observations, and $\bar{X}$ (or $\bar{Y}$) is the mean of the variable.

The 2-D GMM is a probability density function expressed in a weighted summation of individual 2-D Gaussian distributions:

$$g(\boldsymbol{x};\boldsymbol{\pi}) = \sum_{k=1}^{K} \pi_k f_k(\boldsymbol{x}) \ , (5.8)$$

where K represents the total number of components and π_k is the k^{th} component density (weight), which is the prior probability that an observation is drawn from the k^{th} Gaussian model represented by f_k (Jain et al., 1999; Aggarwal and Reddy, 2013).

The GMM parameters, including component density and each set of Gaussian distribution parameters (mean and covariance matrix), are estimated from the given dataset with the use of the expectation maximization algorithm (EM; Dempster et al., 1977; Neal and Hinton, 1998; McLachlan and Krishnan, 2007). From the trained model, observations are assigned to the components with the highest probabilities.

5.4.2 Fluid Typing

Clustering results at dried at 110°C conditions are generally better than as-received conditions, and both 5 cluster are identified. Results from dried conditions are used for fluid typing. Means μ (interpreted as the fluid population centers in the T_1-T_2 map) and covariances $\boldsymbol{V}$ of dried condition are plotted in Figure 5.17 for the two samples. We use T_2 of 0.5 ms to separate fluids into large pores ($T_2 > 0.5$ ms) and small pores ($T_2 < 0.5$ ms), as a larger T_2 value roughly corresponds to a larger pore size.

GMM fluid types are qualitatively confirmed by comparing the identified fluid centers to those documented in literature. The mean values are summarized in Table 5.1, along with expected fluid population centers from the literature (Fleury and Romero-

Sarmiento, 2016; Nicot et al., 2016; Washburn and Cheng, 2017; Romero-Sarmiento et al., 2017).

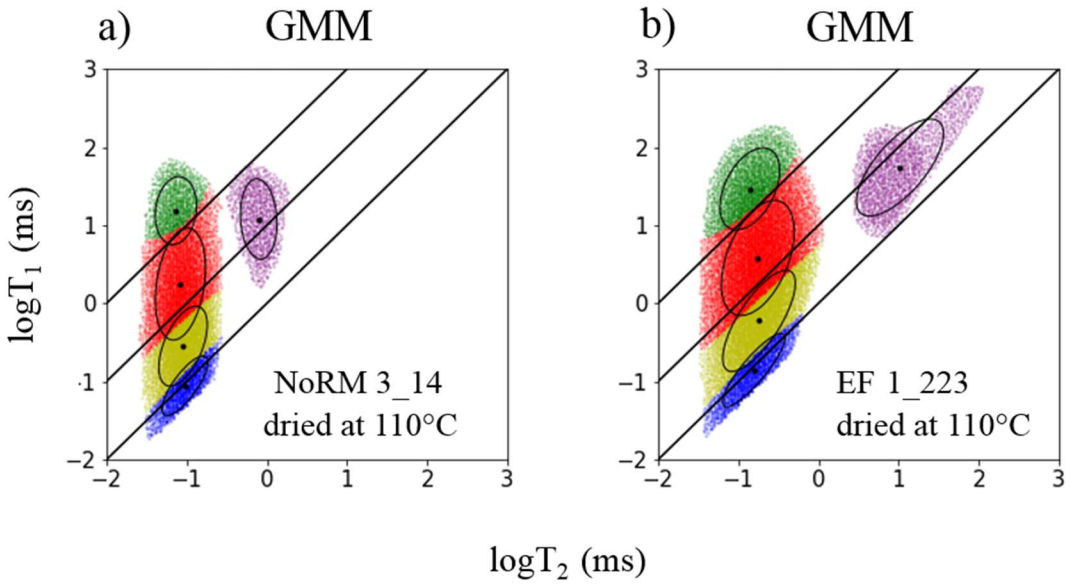

Figure 5.17 GMM results for (a) NoRM 3_14 and (b) EF 1_223 dried at 110°C condition. Cluster labels are denoted by different colors, which including bound water (blue), bound hydrocarbon (yellow), bitumen (red), kerogen (green) and free fluid (purple). The covariance matrix of each Gaussian distribution is represented by the black ellipse, and the mean is represented by a black dot in the center of ellipse. Three lines represent T_1/T_2 ratio = 1, 10, 100.

According to Table 5.1, fluid types predicted by GMM are comparable with previous studies. The GMM cluster component of NoRM 3_14 in the large T_2 region has an average T_2 around 0.8 ms, and an average T_1/T_2 ratio around 16. The GMM cluster of EF 1_223 has an average T_2 around 10 ms, and an average T_1/T_2 ratio round 5.3. The fluid is likely to be hydrocarbon in large pores (Nicot et al., 2016; Romero-Sarmiento et al., 2017).

Table 5.1 Comparison of fluid components identified from this study and previous studies. Samples were dried at 110°C conditions. NoRM refers to the shales from the northern Rocky Mountains. EF refers to the shales from Eagle Ford. Notes: [1] is methane. [2] is free hydrocarbon is light oil (isopar L).

Sample	Small pores								Large pores	
	Bound water		Bound hydrocarbon		Bitumen		Kerogen		Free fluid	
	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2
EF 1_223	0.16	0.9	0.18	3.3	0.17	21.8	0.14	208.7	10.38	5.3
NoRM 3_14	0.10	0.9	0.09	3.2	0.08	21.7	0.07	210.1	0.79	15.2
Fleury and Romero-Sarmiento (2016)	-	1-2	-	-	-	-	0.01-0.1	50, 180, 250	-	15-20[1]
Nicot et al. (2016)	0.1-1	1-2	-	> 3[2]	-	-	-	-	-	> 3[2]
Washburn and Cheng (2017)	-	-	-	4-10	-	20-30	-	1000	2	4-10
Romero-Sarmiento et al. (2017)	-	-	0.25	4	0.07	14.2	0.5	160	-	-

Furthermore, the other 4 components with small T_2 (< 0.5 ms) of both two samples suggest fluids in small pores, which is also confirmed by the previous studies listed in Table 5.1. The cluster component in blue with an average T_1/T_2 ratio close to unity (0.9) tends to be bound water (e.g. clay-bound water; Fleury and Romero-Sarmiento, 2016; Nicot et al., 2016). The yellow component with T_1/T_2 of 3 is likely to be bound hydrocarbon, as a ratio around 4 is a typical signature of oil relaxing in the pore network (Nicot et al., 2016; Romero-Sarmiento et al., 2017). The red component with higher T_1/T_2 around 22 is classified as bitumen, as suggested by Washburn et al. (2013), and Washburn and Cheng (2017). The green component with T_1/T_2 ratio above 200 is classified as kerogen

accordingly (Washburn et al., 2013; Romero-Sarmiento et al. 2017; Washburn and Cheng, 2017).

5.4.3 Computation of Fluid Volumes

After the clustering is performed on the dried at 110°C conditions, the T_1-T_2 space is partitioned, where each grid point in the space is assigned with a cluster label. By summation of fluid volumes for each cluster, the fluid volume of each fluid component can be obtained.

Since the clustering performed relatively poorly for samples at as-received conditions, we use the clustering results from dried at 110°C conditions to inform the clustering at as received conditions to mitigate the effects of the dominant fluid. For the sample at as-received conditions, T_1-T_2 grid points are assigned with the corresponding cluster labels from dried at 110°C conditions (Figure 5.18). The assumption of this 'label-transferring' is that the same number of fluid components is present and the composition and identity of those fluids have not changed after heating at 110°C.

Fluid volumes for 5 fluid components of both two samples are listed in Table 5.2. NoRM 3_14 shows a similar fluid portion for both as-received condition and dried at 110°C conditions. For EF 1_223 at the as-received conditions, the hydrocarbon at large T_2 region dominates the NMR signal. After heating at 110°C, the hydrocarbon in the large pore significantly is reduced.

In this work, we validated the cluster centers using data from literature. The computed volumeshowever, need experimental validation. This could be done by comparing fluid volumes extracted by other methods such as the Dean-Stark method. Other ways for validation can be the injection of known fluids with known fluid volume or performing clustering on conventional rocks like sandstones with a simpler pore geometry.

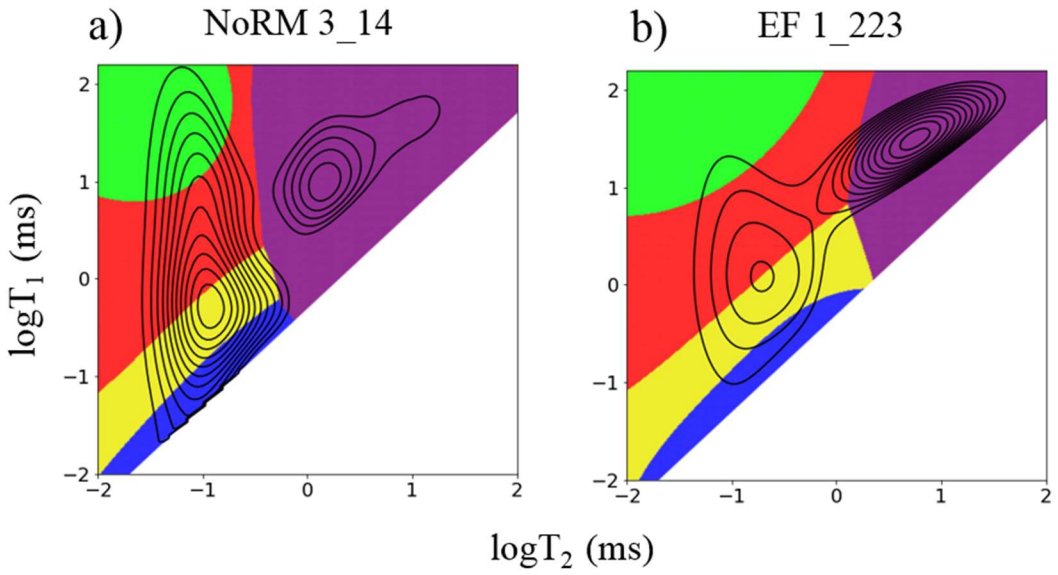

Figure 5.18 Fluid volume computation for as-received conditions of (a) NoRM 3_14 and (b) EF 1_223 using clustering from samples dried at 110°C conditions. The T_1-T_2 distributions at as-received conditions are shown using contour plots for better visualization. GMM fluid regions at dried at 110°C conditions are denoted by different colors, including bound water (blue), bound hydrocarbon (yellow), bitumen (red), kerogen (green) and free fluid (purple). The results are plotted on log-log scales.

Table 5.2 Fluid volumes in cm^3 of 5 fluid components of 2 samples at as-received and dried at 110°C conditions. EF refers to the shales from Eagle Ford. NoRM refers to the shales from the northern Rocky Mountains.

Sample	Small pores				Large pores
	Bound water	Bound hydrocarbon	Bitumen	Kerogen	Free fluid
EF 1_223, as-received	0.098	0.462	0.781	0.073	1.442
EF 1_223, dried at 110°C	0.048	0.115	0.218	0.067	0.066
NoRM 3_14, as-received	0.287	0.775	0.984	0.328	0.494
NoRM 3_14, dried at 110°C	0.157	0.342	0.503	0.126	0.183

5.5 DISCUSSIONS

5.5.1 Physical Meaning of GMM

GMM clustering matches the fluid partitioning predicted by BPP theory. As Jain (2010) pointed out: When there is a good match between the clustering algorithm and the data, good partitions are obtained. On one hand, GMM is a generative process, which tries to understand the underlying process through which clustering is generated, instead of simply learning the decision boundaries (Aggarwal and Reddy, 2013). The GMM assumes that the observations are drawn from a mixture of Gaussian probability distributions. On the other hand, fluid properties like the viscosity of individual components can be modeled with a Gaussian probability distribution. On the T_1-T_2 map, the mean of the signal represents the average property value, and the spread of the property can be modeled by the covariance matrix in two-dimensions. The physical mixture of multiple fluid components in the porous medium yields the mixture of multiple fluid responses in the NMR T_1-T_2 distribution. As a result, the GMM assumption is physically consistent with the fluid distribution in shale.

5.5.2 Comparison of GMM and Manual Methods

The fluid partitioning obtained by GMM can be linked to partitioning results obtained from Fleury and Romero-Sarmiento (2016). The covariance matrix of GMM (Figure 5.17) is consistent with the elliptical partitioning regions based on the manual method (Figure 5.1a), since the ellipse is the geometrical representation of the covariance matrix. This can be proved by the eigen-decomposition of the covariance matrix, where the eigenvectors represent the directions of the major and minor axis and the magnitudes of the two axes are the eigenvalues.

On the other hand, compared to manual partitioning, GMM provides a robust data-driven way to automatically obtain the fluid regions and fluid centers. Through statistically learning the pattern from the data, it overcomes the subjective nature of manually choosing the shape and boundaries.

5.6 CONCLUSIONS

We investigated 6 clustering algorithms for fluid characterization in NMR T_1-T_2 measurements from 2 organic-rich shale samples at as-received and dried at 110°C conditions. Two cluster validity indices including T_1/T_2 ratio range (RR) and angle difference (AD) were proposed. These indices are based on the assumption from BPP theory that individual fluid populations will be characterized by a single value or narrow range of T_1/T_2 ratio.

For most cases, GMM is the best algorithm, as its partitions shows the highest consistency with the theoretical fluid boundary lines predicted by BPP theory. Other 5 algorithms fail to generate good partitioning results.

Clustering algorithms are sensitive to the fluid distribution. For sample EF 1_223 at as-received conditions, none of the algorithms (include GMM) produce good clustering results, which could be due to the relatively large volume in the large T_2 region. Compared to as-received conditions, drying the sample producing better clustering results by revealing the footprint of organic matter.

We further evaluated GMM approach using its cluster centers from dried at 110°C conditions. Five fluid components were identified using the cluster centers of GMM, including free fluid in large pores ($T_2 > 0.5$ ms) and 4 fluid components in small pores ($T_2 < 0.5$ ms). The 4 components are bound water, bound hydrocarbon, bitumen and kerogen. Their values are consistent with the expected fluid responses from literature. In addition,

we computed the fluid volume of each cluster and transferred the labels to as-received conditions.

The GMM-based clustering approach is suitable for fluid characterization NMR T_1-T_2 measurements of shale where multiple fluid components are present in different pore systems. Compared to the manual partitioning methods, it overcomes the subjective nature of human decisions and provides a robust machine learning approach for fluid partitioning in NMR T_1-T_2 for shales.

Chapter 6: Characterization of Fluid Distributions in NMR T_1-T_2 Maps of Shales based on Gaussian Mixture Model Clustering Analysis

6.1 INTRODUCTION

In Chapter 5, we introduced a new fluid partitioning approach for NMR T_1-T_2 maps in shale using Gaussian mixture model (GMM) clustering analysis. We proposed two indices including T_1/T_2 ratio range (RR) and angle difference (AD) based on Bloembergen-Purcell-Pound (BPP) theory. Six clustering algorithms were performed on T_1-T_2 maps of 2 organic-rich shale samples for both as-received conditions and dried at 110°C conditions. The comparison of 6 clustering algorithms indicate that GMM is the best algorithm for most of the cases, as its partitions show the highest consistency with the theoretical fluid boundary lines predicted by BPP theory. In addition, clustering is found to be sensitive to the fluid distribution. Heating can help the performance of clustering, whereas clustering tends to perform relatively poorly at as-received conditions.

To further validate the GMM clustering approach, in this chapter, we present a more thorough analysis using multiple EF and NoRM samples. Four organic-rich shales (two EF and two NoRM samples) were used in this work, and their T_1-T_2 maps were collected at both as-received and dried at 110°C conditions. We applied GMM on the T_1-T_2 maps and computed the two indices against cluster number for selecting optimal cluster number. In addition, we focus on discussing fluid patterns of shales from different formations.

6.2 METHODS

Shale samples used in this study are from the Eagle Ford formation (denoted as EF) and from the northern Rocky Mountains (denoted as NoRM). A total of 4 samples were used (EF 2_50, EF 2_93, NoRM 4_14, NoRM 4_34). Details of the sample properties can be found in Chapter 3.3. About 60 g of sample particles with mesh size between 20-35 US mesh were collected and dried in an oven at 110°C for 2-4 days. NMR T_1-T_2 measurements were performed for both samples immediately after crushing and after oven drying at 110°C. T_1-T_2 data were acquired with inversion recovery steps followed by a Carr-Purcell-Meiboom-Gill (CPMG) sequence (Carr and Purcell, 1954; Meiboom and Gill, 1958). The T_1-T_2 maps were computed by the Matlab-based inversion program (Medellín et al., 2015; Medellín et al., 2016). More details on sample preparation and measurement procedures can be found in Chapter 5.2.

We applied the GMM clustering analysis described in Chapter 5. The workflow is briefly summarized here: First, a sub-space of T_1-T_2 domain is obtained by ignoring the region where T_1/T_2 ratio is less than 0.5. Second, a logarithmic transform is applied to the T_1 and T_2 values to convert the heavily skewed data to a more symmetric distribution. Fluid volume f is normalized using Equation (5.1), and the threshold f_c is chosen as 10% of the maximum fluid volume. The normalization converts f to f_{norm}, an positive integer. Volume values smaller than f_c are ignored. The clustering requires converting the 3-D data (T_1, T_2, f) to a 2-D T_1-T_2 distribution, where the frequency is represented by the number of points on the grid point (T_1, T_2). As a result, each grid point is duplicated f_{norm} times to produce the 2-D clustering dataset.

GMM belongs to one type of clustering approach called model-based clustering (Jain et al., 1999). GMM assumes that data is generated from a mixture of components, and each component is described by a Gaussian probability distribution (Day, 1969;

McLachlan et al., 1999; Fraley and Raftery, 2006). The cluster number needs to be pre-defined before performing clustering. One starts with a small value of cluster number and successively increases the cluster number. The maximum cluster number is restricted to 7 to avoid the difficulty of fluid typing interpretation. For a given cluster number, GMM is trained using the python machine learning library scikit-learn (Pedregosa et al., 2011). After training the model, the cluster label for each grid point is generated. This is achieved by assigning the grid point to the cluster with the highest probabilities.

After clustering, two cluster validity indices, ratio range (RR) and angle difference (AD), are calculated. Their expressions are defined in Equation (5.2) and (5.3). They provide guidance for selecting the optimal cluster number. RR measures the variation of T_1/T_2 ratio within the same cluster, whereas AD measures the angle deviation of the fluid partitioning boundary to a theoretical fluid boundary line with a constant T_1/T_2 ratio. Small values of RR and AD are preferable. More details can be found in Chapter 5.2.2 – 5.2.4.

6.3 Results and Discussions

6.3.1 T_1-T_2 Maps

The T_1-T_2 maps for 2 EF samples and 2 NoRM at as-received conditions are shown in Figure 6.1. The T_1-T_2 maps for samples dried at 110°C conditions are in Figure 6.2. The total NMR fluid volumes are listed in Table 6.1.

For as-received conditions, samples from the same formation have a similar fluid distribution pattern, whereas samples from different formations have distinct distribution characteristics. The two EF samples are dominated by the fluid in the region of $T_2 > 1$ ms. The dominant fluid forms an elliptical shape. Some signal extends to the smaller T_2 region ($T_2 < 1$ ms), forming a tail-like shape. On the other hand, NoRM samples display a

more complicated geometry: there is a distinct separation of signal in $T_2 > 1$ ms and $T_2 < 1$ ms regions. Yet, the peak volume locations of the two samples are different: The peak volume is in $T_2 > 1$ ms for NoRM 4_14, whereas the peak fluid volume for NoRM 4_34 is in the $T_2 < 1$ ms region.

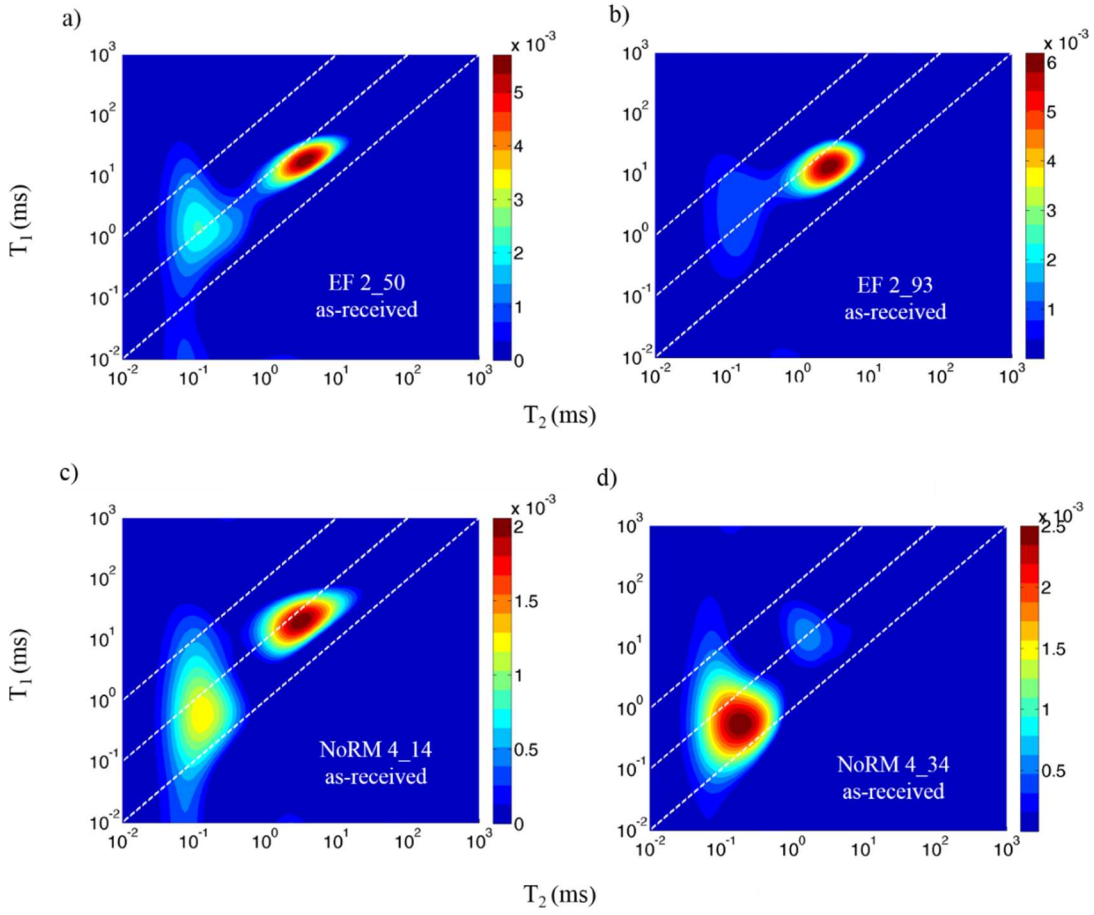

Figure 6.1 T_1-T_2 maps for samples at as-received conditions: (a) EF 2_50, (b) EF 2_93, (c) NoRM 4_14, and (d) NoRM 4_34. The warmer colors indicate greater pore volume. The units of pore volume are cm^3. T_1 and T_2 both range from 10^{-2} ms to 10^3 ms. Three dashed lines correspond to T_1/T_2 ratio = 1, 10, 100. EF refers to the shales from Eagle Ford. NoRM refers to the shales from the northern Rocky Mountains.

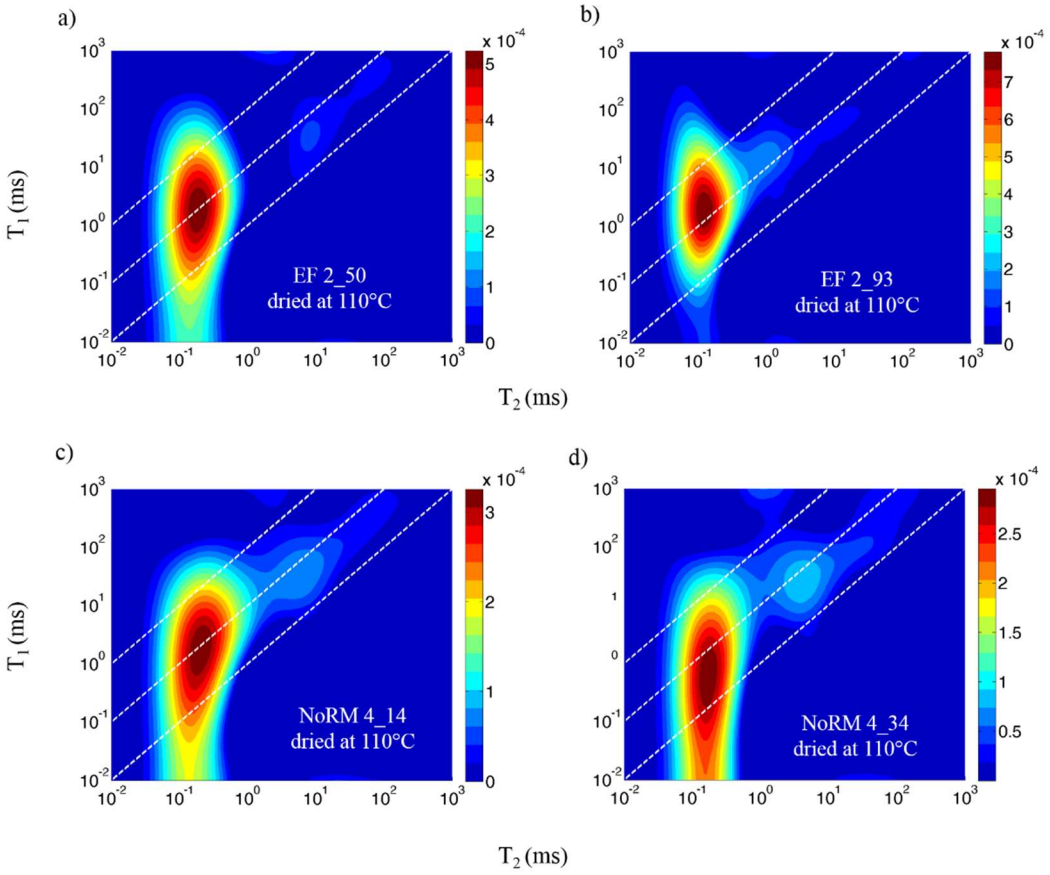

Figure 6.2 T_1-T_2 maps for samples dried at 110°C conditions: (a) EF 2_50, (b) EF 2_93, (c) NoRM 4_14, and (d) NoRM 4_34. The warmer colors indicate greater pore volume. The units of pore volume are cm^3. T_1 and T_2 both range from 10^{-2} ms to 10^3 ms. Three dashed lines correspond to T_1/T_2 ratio = 1, 10, 100. EF refers to the shales from Eagle Ford. NoRM refers to the shales from the northern Rocky Mountains.

After heating the samples, the fluid volume is reduced due to evaporation (Table 6.1). The NMR signal shows a significant change compared to as-received conditions. The fluid volume, especially in the large T_2 region, is greatly reduced, and the peak fluid volume shifts to smaller T_2 region. The signal of organic matter (OM), which was suppressed at as-received conditions, is anticipated to be more prevalent in these measurements.

In addition, both EF and NoRM samples show a similar distribution pattern after heating. The fluid distribution difference at as-received conditions may stem from different rock diagenetic histories, and this difference is reduced by the heating effect.

Table 6.1 Total NMR fluid volumes of 3 EF samples and 3 NoRM samples at as-received and dried at 110°C conditions. EF refers to the shales from Eagle Ford. NoRM refers to the shales from the northern Rocky Mountains.

Sample	Total fluid volume (cm^3)	
	As-received conditions	Dried at 110°C conditions
EF 1_223	2.86	0.51
EF 2_50	3.57	0.82
EF 2_93	2.93	1.00
NoRM 3_14	2.87	1.31
NoRM 4_14	2.12	0.62
NoRM 4_34	2.38	0.51

6.3.2 Clustering Results

6.3.2.1 Dried at 110°C Conditions

GMM clustering was first performed on T_1-T_2 data of samples dried at 110°C conditions, since heating helps to reveal the footprint of OM, generating better cluster performance. The two cluster validity indices RR and AD are plotted against the cluster number (Figure 6.3, 6.4), which can analyze the clustering performance and select the optimal cluster number.

For each cluster number, the highest value point is chosen to represent the cluster quality. Most of the indices show a V-shape against the cluster number, where the optimal cluster number can be easily found on the bottom. According to the analysis, the optimal number is 5, as there is often a minimum value for the two indices when cluster number is 5. In a few cases (e.g. Figure 6.3a), values for a cluster number of 4 are similar

or even lower than those for a cluster number of 5. However, we choose a fixed cluster number, which allows the comparison of fluid types among samples.

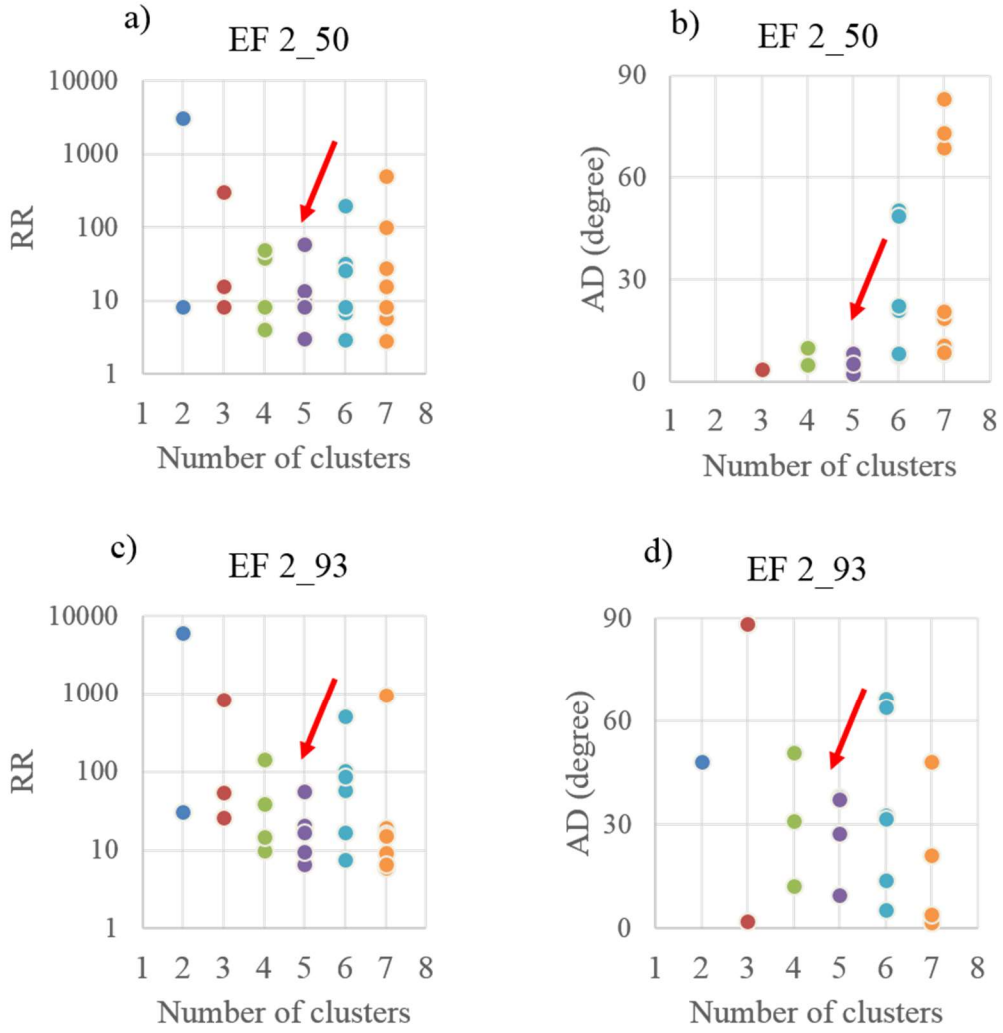

Figure 6.3 Cluster validity indices of GMM clustering against different number of clusters for T_1-T_2 maps dried at 110°C conditions: (a) RR of EF 2_50, (b) AD of EF 2_50, (c) RR of EF 2_93, and (d) AD of EF 2_93. RR is ratio range and AD is angle difference. EF refers to the shales from Eagle Ford. The optimal clustering number is indicated by the red arrow.

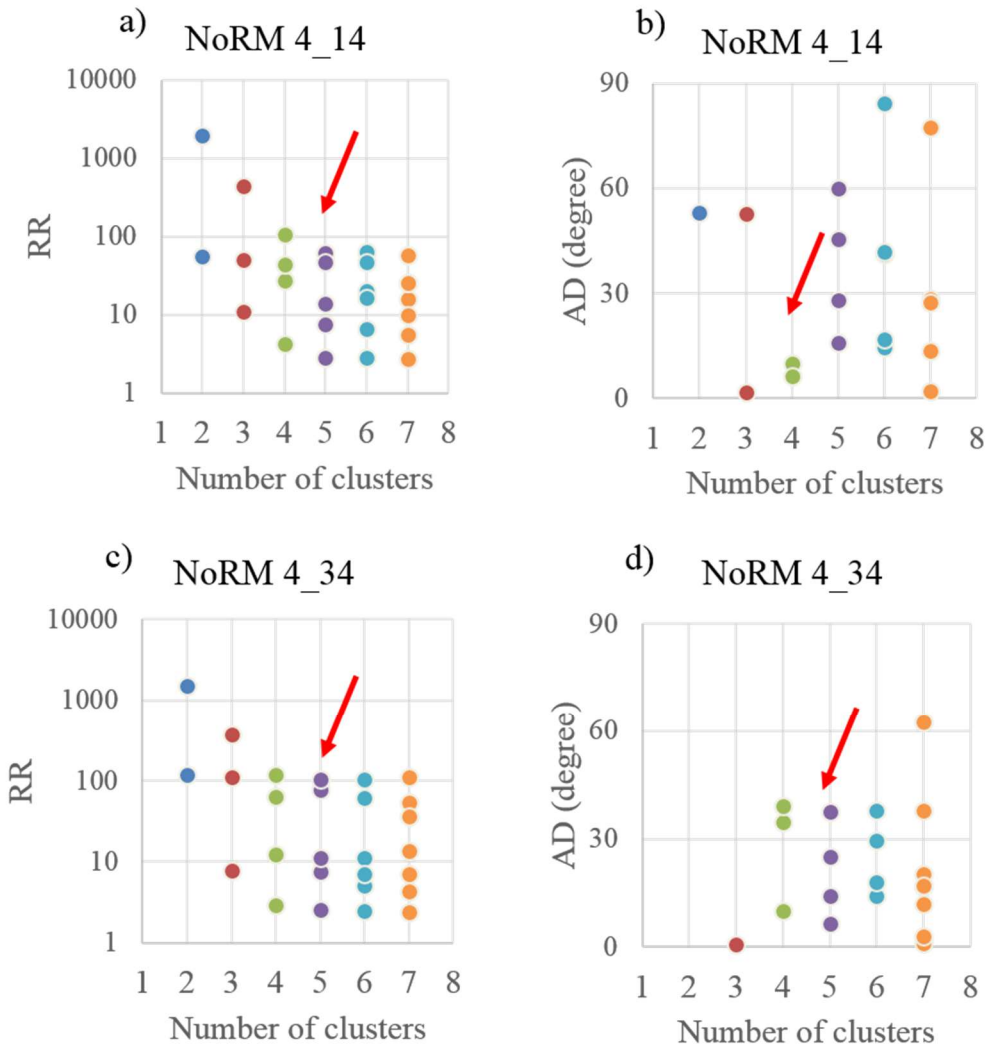

Figure 6.4 Cluster validity indices of GMM clustering against different number of clusters for T_1-T_2 maps dried at 110°C conditions: (a) RR of NoRM 4_14, (b) AD of NoRM 4_14, (c) RR of NoRM 4_34, and (d) AD of NoRM 4_34. RR is ratio range and AD is angle difference. The optimal clustering number is indicated by the red arrow.

The clustering results for 4 samples are plotted in Figure 6.5. Partitioning results show good consistency with the reference T_1/T_2 ratio lines, especially for EF samples. The decision boundaries parallel the reference lines. On the other hand, NoRM fluid boundaries show a small deviation against the reference lines. One possible explanation

is because the presence of the purple cluster in the large T_2 region, whose footprint is mixed with clusters in the small T_2 region.

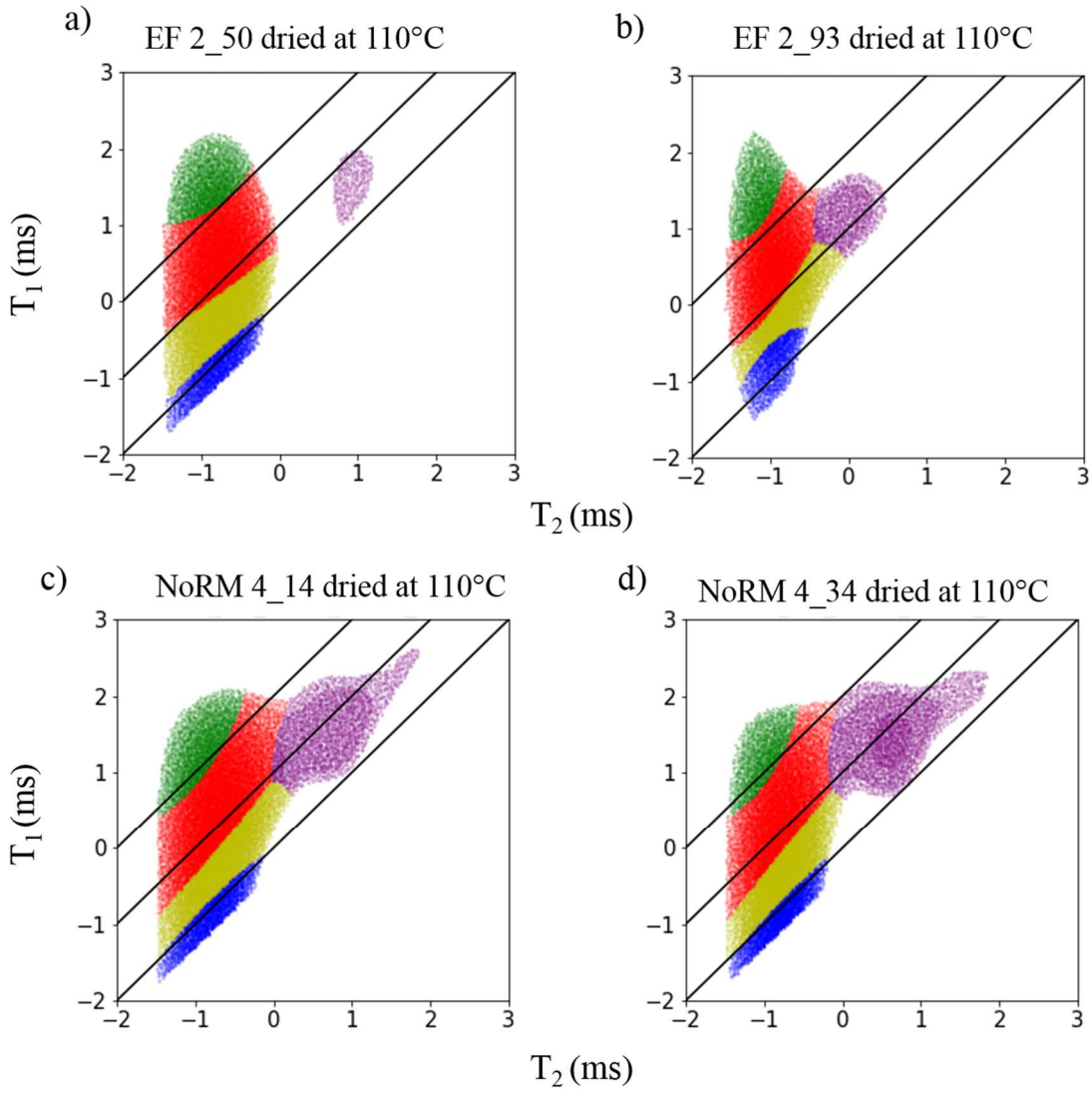

Figure 6.5 Clustering results for 4 samples dried at 110°C conditions. The cluster number is 5. The results are plotted on log-log scales. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100.

6.3.2.2 As-received Conditions

We applied GMM with cluster number of 5 to samples at as-received conditions (Figure 6.6). Most of the clustering results for as-received conditions have relatively poor performance. The cluster boundaries don't align well with the reference lines. For EF samples, GMM tries to make sub-partitions in the large T_2 region. For sample NoRM 4_14 (Figure 6.6c), however, it shows good clustering results. Similar to NoRM 3_14 in Chapter 5 (Figure 5.9c), both samples have a clear separation of the cluster (purple) in larger T_2 and rest clusters in smaller T_2 regions. This separation might allow the model to choose a better partitioning pattern.

6.3.3 Fluid Typing

Compared to as-received conditions, clustering results for dried at 110°C conditions show better partitioning. As a result, fluid typing is conducted using means (interpreted as fluid population centers) for GMM results from dried at 110°C conditions. Centers for the five identified fluids are summarized in Table 6.2 as well as those documented in literature. The results include EF 1_223 and NoRM 3_14 from Chapter 5.

Again, fluid types predicted by GMM are comparable with previous studies. We use T_2 of 0.5 ms to separate fluids in large pores ($T_2 > 0.5$ ms) and small pores ($T_2 < 0.5$ ms) since a larger T_2 value generally corresponds to a larger pore size. The free fluid in large pores (purple) shows a wide range of T_2 and T_1/T_2 ratio (Table 6.2). Its composition is likely to be hydrocarbon.

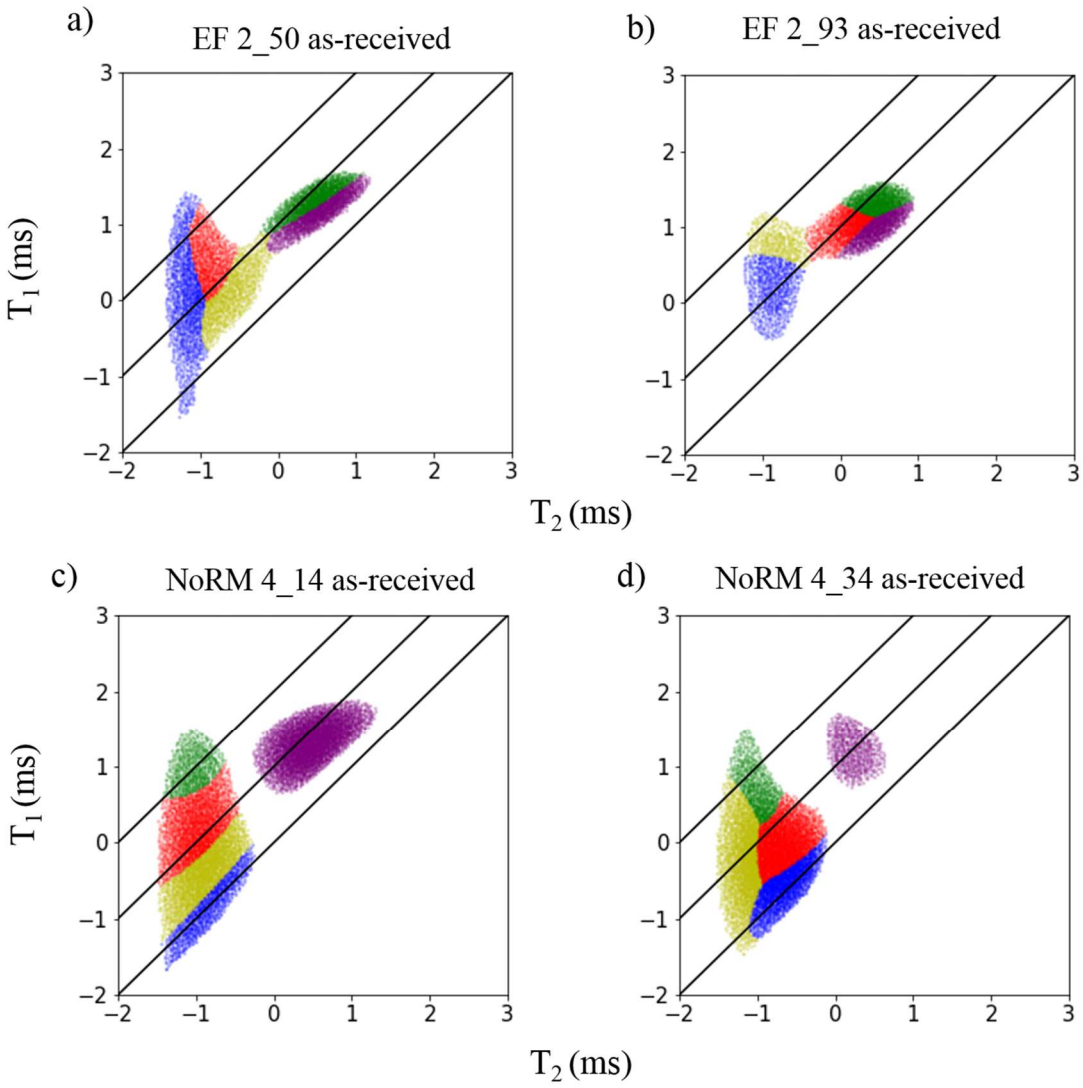

Figure 6.6 Clustering results for 4 samples at as-received conditions. The cluster number is 5. The results are plotted on log-log scales. Cluster labels are denoted by different colors. Three lines correspond to T_1/T_2 ratio = 1, 10, 100.

Table 6.2 Comparison of fluid components identified from this study and previous studies. fluid components were identify using GMM clustering in samples dried at 110°C conditions. EF refers to the shales from Eagle Ford. NoRM refers to the shales from the northern Rocky Mountains. Note: [1] is methane. [2] is light oil (isopar L).

Sample	Small pores								Large pores	
	Bound water		Bound hydrocarbon		Bitumen		Kerogen		Free fluid	
	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2	T_2 (ms)	T_1/T_2
EF 1_223	0.16	0.9	0.18	3.3	0.17	21.8	0.14	208.7	10.38	5.3
EF 2_50	0.16	0.9	0.17	3.6	0.16	23.8	0.13	222.0	8.2	4.2
EF 2_93	0.12	1.6	0.17	6.2	0.11	30.5	0.07	309.4	0.83	16.9
NoRM 3_14	0.10	0.9	0.09	3.2	0.08	21.7	0.07	210.1	0.79	15.2
NoRM 4_14	0.15	1.0	0.13	3.2	0.10	14.8	0.09	87.0	2.99	7.0
NoRM 4_34	0.17	0.8	0.18	2.7	0.16	16.8	0.10	130.1	3.30	8.7
Fleury and Romero-Sarmiento (2016)	-	1-2	-	-	-	-	0.01-0.1	50, 180, 250	-	15-20[1]
Nicot et al. (2016)	0.1-1	1-2	-	> 3[2]	-	-	-	-	-	> 3[2]
Washburn and Cheng (2017)	-	-	-	4-10	-	20-30	-	1000	2	4-10
Romero-Sarmiento et al. (2017)	-	-	0.25	4	0.07	14.2	0.5	160	-	-

Four fluids in small pores include bound water, bound hydrocarbon, bitumen, and kerogen. They have very similar T_2 and T_1/T_2 ratios across all 6 samples, which validate the GMM clustering performance. This also reveals that samples from the two formations have a similar fluid partitioning pattern after heating. It implies that GMM obtains a general fluid partitioning rule that is less sensitive to rock lithology.

Furthermore, the fluid volume of each fluid at dried at 110°C conditions can be obtained by summation of the fluid volume of each cluster label. In addition, the 'label-

transferring' approach in Chapter 5.4.3 is applied for T_1-T_2 maps for as-received conditions. Grid points on the map are assigned the corresponding cluster labels from dried at 110°C conditions (Figure 6.7) and their fluid volume are thus obtained.

Fluid volumes for 5 fluid components are listed in Table 6.3. The relative proportions are computed and shown in stacked bar-plots in Figure 6.8. At as-received conditions, the free fluid at large T_2 dominates the NMR signal. For all EF and NoRM 4_14 samples, the dominant fluid makes up 50%-60% of the total fluid volume. The dominant fluid in large pores may suppress the NMR signal of other components such as bound water and kerogen. After heating at 110°C, the fluid in large pores significantly reduces to less than 20%. The contribution of kerogen, bitumen to the signal increase. The change of relative fluid saturation helps the clustering algorithm to generate a better partitioning result.

6.4 CONCLUSIONS

We extended the GMM clustering approach through the application to 4 organic-rich shale samples at as-received and dried at 110°C conditions. Through the clustering analysis, we validated the two proposed indices for their ability of selecting optimal cluster number. Most of the indices show a V-shape against the cluster number, where the optimal cluster number corresponds to the minimum index value.

Clustering is sensitive to the fluid distribution. For as-received conditions, clustering may perform poorly if the distribution is dominated by signal at large T_2 values, and if the dominated signal is mixed with signal from small T_2 values. Heating at 110°C can help to reveal the footprint of OM, generating better cluster performance. The identified fluid clusters at dried conditions are comparable with fluid types from previous

studies. In addition, the fluid partitioning rule obtained by GMM show a general pattern that is less sensitive to rock lithology.

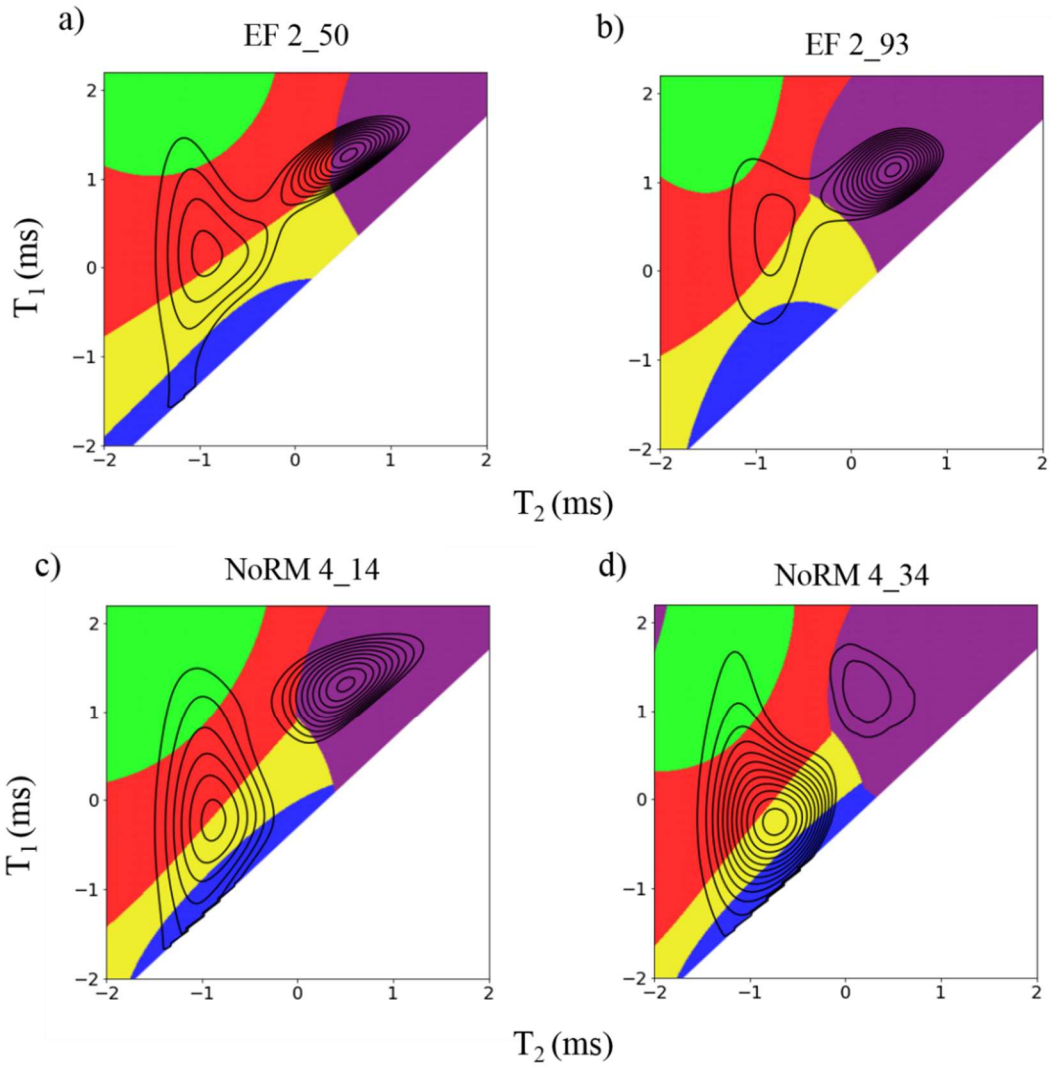

Figure 6.7 Fluid volume computation for as-received conditions of (a) EF 2_50, (b) EF 2_93, (c) NoRM 4_14, and (d) NoRM 4_34 using clustering from samples dried at 110°C conditions. The T_1-T_2 distributions at as-received conditions are shown using contour plots for better visualization. GMM fluid regions at dried at 110°C conditions are denoted by different colors, including bound water (blue), bound hydrocarbon (yellow), bitumen (red), kerogen (green) and free fluid (purple). The results are plotted on log-log scales.

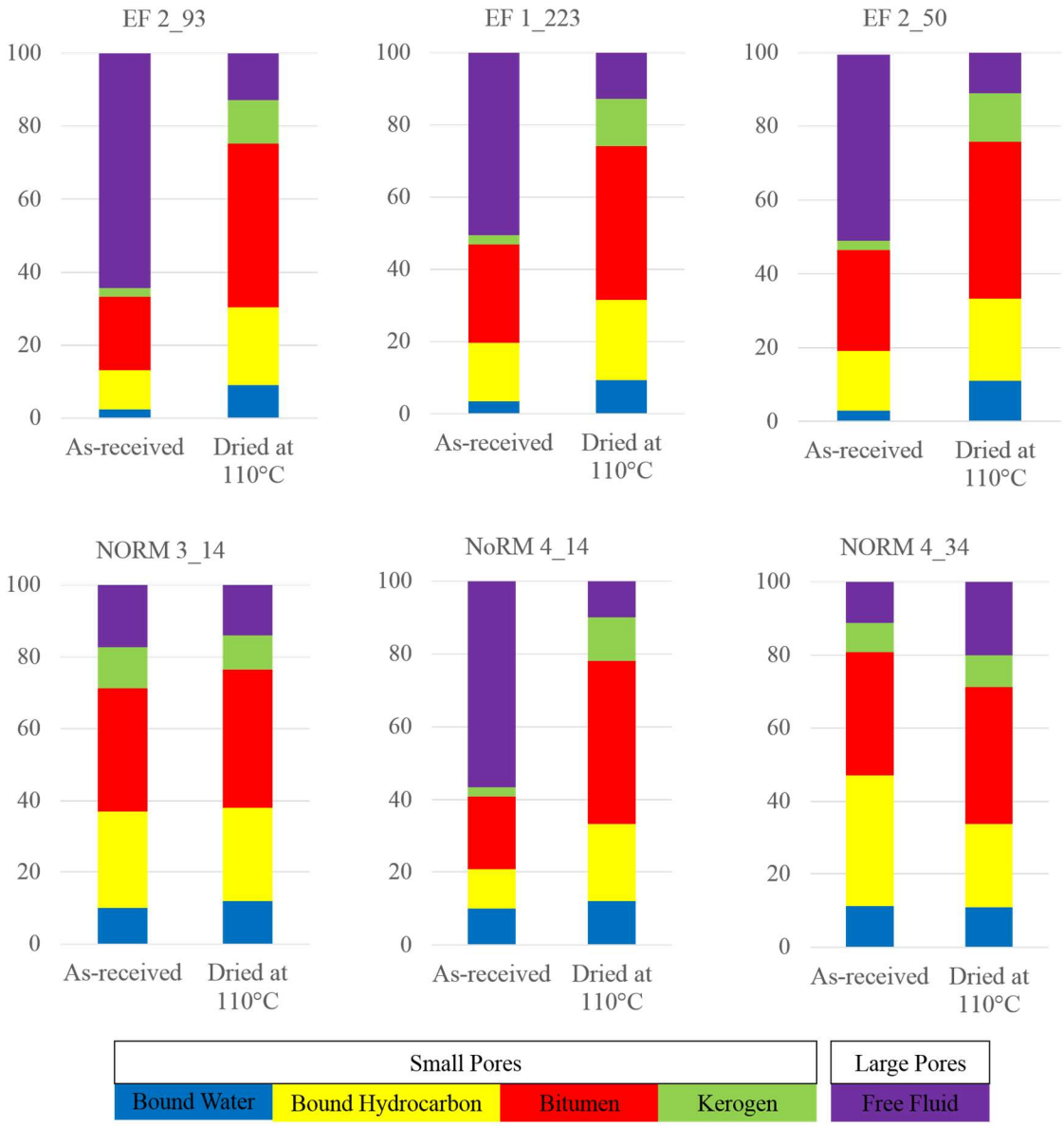

Figure 6.8 Relative portions of 5 fluid components for 4 samples at both the as-received and dried at 110°C conditions. The units of the relative portion is percentage (%), and 5 portions are summed to be 1 (100%). Fluid components include bound water (blue), bound hydrocarbon (yellow), bitumen (red), kerogen (green) and free fluid (purple).

Table 6.3 Fluid volumes in cm^3 of 5 fluid components at as-received and dried at 110°C conditions. Six samples are included. EF refers to the shales from Eagle Ford. NoRM refers to the shales from the northern Rocky Mountains.

Sample	Small pores				Large pores
	Bound water	Bound hydrocarbon	Bitumen	Kerogen	Free fluid
EF 1_223, as-received	0.098	0.462	0.781	0.073	1.442
EF 1_223, dried at 110°C	0.048	0.115	0.218	0.067	0.066
EF 2_50, as-received	0.104	0.752	1.616	0.133	0.971
EF 2_50, dried at 110°C	0.090	0.217	0.362	0.097	0.054
EF 2_93, as-received	0.069	0.315	0.592	0.073	1.886
EF 2_93, dried at 110°C	0.091	0.213	0.448	0.119	0.128
NoRM 3_14, as-received	0.287	0.775	0.984	0.328	0.494
NoRM 3_14, dried at 110°C	0.157	0.342	0.503	0.126	0.183
NoRM 4_14, as-received	0.167	0.408	0.618	0.138	0.791
NoRM 4_14, dried at 110°C	0.060	0.138	0.259	0.072	0.090
NoRM 4_34, as-received	0.268	0.852	0.800	0.191	0.265
NoRM 4_34, dried at 110°C	0.056	0.117	0.191	0.044	0.103

Samples from the same formation have a similar fluid distribution for as-received conditions, whereas samples from different formations have distinct fluid distribution characteristics. The fluid distribution difference at as-received conditions may stem from different rock diagenetic histories. The heating, one the other hand, reduces the difference through fluid evaporation.

Chapter 7: Conclusions and Future Work

7.1 CONCLUSIONS

The goal of this PhD work was to determine petrophysical properties of organic-rich shale using experiments, lab measurements, and machine learning analysis. To achieve this, we used mature organic-rich shale samples from Eagle Ford (EF) and the northern Rocky Mountains (NoRM) formations. X-ray diffraction (XRD) and rock-eval analysis were performed on the samples to characterize minerologies and organic geochemical properties. Calcite dominated the EF shale samples, whereas quartz dominated the NoRM samples. Effects of shear deformation on shale pore structure at the nanoscale was investigated using N_2/ CO_2 gas sorption and scanning electron microscope (SEM) analysis. In addition, effects of drying temperature on shale permeability and pore structure measurement were studied using N_2 gas sorption and Gas Research Institute (GRI) method. Furthermore, to overcome the limitation of manual methods for interpreting nuclear magnetic resonance (NMR) T_1-T_2 maps of shale, a new fluid characterization approach based on Gaussian mixture model (GMM) clustering analysis was developed.

N_2/ CO_2 sorption measurements were combined with SEM imaging to characterize the response of shale pore system to the shear deformation at nanoscale. Confined compressive strength tests were conducted on preserved NoRM and EF shale core plugs. After failure, most samples displayed increases in pore structure parameters including Brunauer-Emmett-Teller (BET) surface area, N_2/ CO_2 porosity, pore size distribution, and surface fractal dimensions. Compared to other parameters, surface fractal dimensions were less sensitive to shear failure. After failure, fractures with widths ranging from 10-100 nm up to 1-2 μm were observed to follow coarser grain boundaries and laminae of OM and matrix materials. The interaction of the OM laminae and the shear fracturing may improve

the connectivity of the OM laminae to the adjacent rock matrix, and thus enhance the hydrocarbon mobility. In addition, different responses to deformation of samples between the two formations shales may be due to diagenetic differences. The differences in rock fabric created by different diagenetic histories cause different nanoscale fracture patterns, including anomalous porosity increases due to pore distributions within OM, heterogeneous distribution of cement between samples, and enhanced porosity within deformed clay aggregates.

N_2 sorption and GRI measurements were conducted on NoRM and EF samples dried at 4 different temperatures to characterize the effects of fluid removal and thermal maturation under heat treatment. BET surface area, N_2 pore volume, N_2 pore size distribution, GRI porosity, and GRI permeabilities showed increasing trends against heating temperature, indicating those petrophysical properties are sensitive to the drying temperature for the sample preparation. GRI permeabilities were strongly affected by heating temperature, which suggest that caution is warranted in interpreting permeability measurements performed on shales, as the preparation technique appears to have a significant influence on the results. Two factors contributing to the increase of those petrophysical properties include continuous removal of fluids, and thermal maturation at high temperatures (especially above 600°C). The multiscale pore network features were reflected by the cross-plots between N_2 pore volume and GRI permeabilities, which indicates that larger pores are well connected, while the smaller ones appear to be more poorly connected. This is consistent with modeling performed by Jiang et al. (2015) and Tian and Daigle (2018b) on Barnett shale samples, wherein a certain fraction of small pores were assumed to be partially or completely disconnected from the overall structure.

A new fluid characterization approach was developed for NMR T_1-T_2 maps in shale based on GMM clustering analysis. Three hierarchical methods and three partitional

clustering methods, including GMM, were applied on T_1-T_2 maps in NoRM and EF shales at as-received and dried at 110°C conditions. We proposed two cluster validity indices including T_1/T_2 ratio range (RR) and angle difference (AD) to evaluate the cluster quality. These indices are based on the assumption from Bloembergen-Purcell-Pound (BPP) theory that individual fluid populations will be characterized by a single value or narrow range of T_1/T_2 ratio. Results show that the two indices provide guidance for selecting the optimal cluster number and best algorithm. For most cases, GMM was the best algorithm, as its partitions showed the highest consistency with the theoretical fluid boundary lines predicted by BPP theory. Furthermore, clustering algorithms were sensitive to the fluid distribution on the T_1-T_2 map. For sample EF 1_223 at as-received conditions, none of the algorithms (including GMM) produced good clustering results, which could be due to the relatively large volume in the large T_2 region. One the other hand, drying the sample helped to reveal the footprint of organic matter, producing better clustering results. Compared to manual partitioning methods, GMM-based clustering overcomes the subjective nature of human decisions and provides a robust machine learning approach for fluid characterization in NMR T_1-T_2 for shales.

We extended the analysis of GMM clustering approach by applying to 4 organic-rich shale samples at as-received and dried at 110°C conditions. The two indices and GMM approach were further validated by the results. Heating at 110°C can help to reveal the footprint of OM, generating better cluster performance. Fluid types identified at dried at 110°C conditions were comparable with previous studies. In addition, the fluid partitioning rule obtained by GMM show a general pattern that is less sensitive to rock lithology.

7.2 FUTURE WORK

Our work on shale nanoscale pore response to shear deformation shows that the shear failure is more effective to capture laminated OM pores, compared to the dispersed OM. Future research should consider the timing and role of cementation in establishing these OM relationships, and also the relative importance of cementation to these "granular" controls on syn-deformational porosity change. In addition, while the properties of the induced microfractures have been characterized (Tian and Daigle, 2018a), the resulting permeability enhancement has yet to be quantified. Future research should focus on such quantification, along with identification of the rock properties and operational parameters that yield optimal results.

Permeabilities measured by GRI method were strongly affected by heating temperature. Future work can extend this study from GRI method to pulse decay method using uncrushed samples. In addition, the data collected from N_2 sorption and permeability measurements can be used to build a pore network model to better understand the shale pore structure and its transport properties. Finally, a detailed analysis should be conducted to determine the preparation procedures that yield the permeability value most relevant for completion and production considerations.

The GMM-based fluid characterization approach was developed using data collected from core samples. For practical implementation, the information from core analysis should be combined and applied to logging measurements. Future work can be done by testing and extending our clustering workflow using logging measurements. In additional, other fundamental laboratory measurements can be done for further investigation and validation of our clustering approach, including re-saturation of shale samples with known fluid, extracting fluid compositions from shale samples by Dean-Stark method.

Appendix A: Pore Size Distribution from Gas Sorption

The pore size distribution can be computed from the adsorption isotherm of the sample. Assuming that adsorption in each pore acts independently, the pore size contribution to the total adsorption isotherm is proportional to the fraction of the total pore volume of the sample. The total amount adsorbed quantity $Q(p)$ is expressed as an integration of the product of individual isotherm kernel function q and pore size distribution function f in respect to pore size H:

$$Q(p) = \int q(p,H) * f(H)dH, (A-1)$$

where $Q(p)$ is the quantity adsorbed at pressure p. $q(p,H)$ is the isotherm kernel function, which describes the quantity adsorbed under pressure p and pore size H per pore volume. $f(H)$ is the pore volume distribution function, with $f(H)dH$ being the pore volume having pore size between H and $H+dH$. Numerical values of the kernel function can be derived using modern statistical mechanics like density functional theory or molecular simulations (Adesida et al., 2011; Roque-Malherbe, 2007). They also can be calculated from classical thermodynamic theories such as BJH method (Barrett et al. 1951).

Equation (A-1) should be discretized into as a summation of multiple pore volume components:

$$Q(p) = \sum F_j q\left(p, H_j\right), (A-2)$$

$$F_j = f\left(H_j\right)\Delta H, (A-3)$$

where F_j is the pore volume for pore size H_j.

In the measured data, the adsorbed quantity $Q(p)$ is a function of pressure over n points. Expressing Equation (A-1) as a system of equations yields:

$$\begin{pmatrix} Q_1 \\ Q_2 \\ \cdots \\ Q_i \\ \cdots \\ Q_n \end{pmatrix} = \begin{pmatrix} q_{1,1} * F_1 + q_{1,2} * F_2 + \cdots + q_{1,j} * F_j + \cdots + q_{1,m} * F_m \\ q_{2,1} * F_1 + q_{2,2} * F_2 + \cdots + q_{2,j} * F_j + \cdots + q_{2,m} * F_m \\ \cdots \quad \cdots \quad \cdots \\ q_{i,1} * F_1 + q_{i,2} * F_2 + \cdots + q_{i,j} * F_j + \cdots + q_{i,m} * F_m \\ \cdots \quad \cdots \quad \cdots \\ q_{n,1} * F_1 + q_{n,2} * F_2 + \cdots + q_{n,j} * F_j + \cdots + q_{n,m} * F_m \end{pmatrix}, \quad (A-4)$$

Similar to NMR T_2 distribution in Chapter 2.5.3, one can solve for F using direct matrix inversion technique. The loss function is expressed as:

$$Loss = \sum_{i=1}^{n} (Q_i - \sum q_{i,j} F_j)^2 + \lambda \sum_{j=1}^{p} F_j^2 = RSS + \lambda \sum_{j=1}^{p} F_j^2, (A-5)$$

where λ is a tuning parameter for the regularization term. RSS is the residual sum of squares.

The error of fit or root mean squared (RMS) error is defined as:

$$RMS = \sqrt{\frac{RSS}{n-1}}, (A-6)$$

As shown from Figure A-1a, the 3Flex software allows the user to decide the value for the regularization parameter λ. According to Figure A-1b, different values yield similar reconstructed isotherms. The shape of the pore size distribution, however, is sensitive to λ (Figure A-1c). A smaller λ (e.g. 0.0001) poses a small regularization effect on the cost function, causing a significant variation of the pore size distribution amplitude. On the other hand, a larger λ (e.g. 1) can produce a smooth pore size

distribution. Here, we chose 1.0 for the value, as a smoother pore size distribution is expected for shale. We applied the same λ through all samples for consistency.

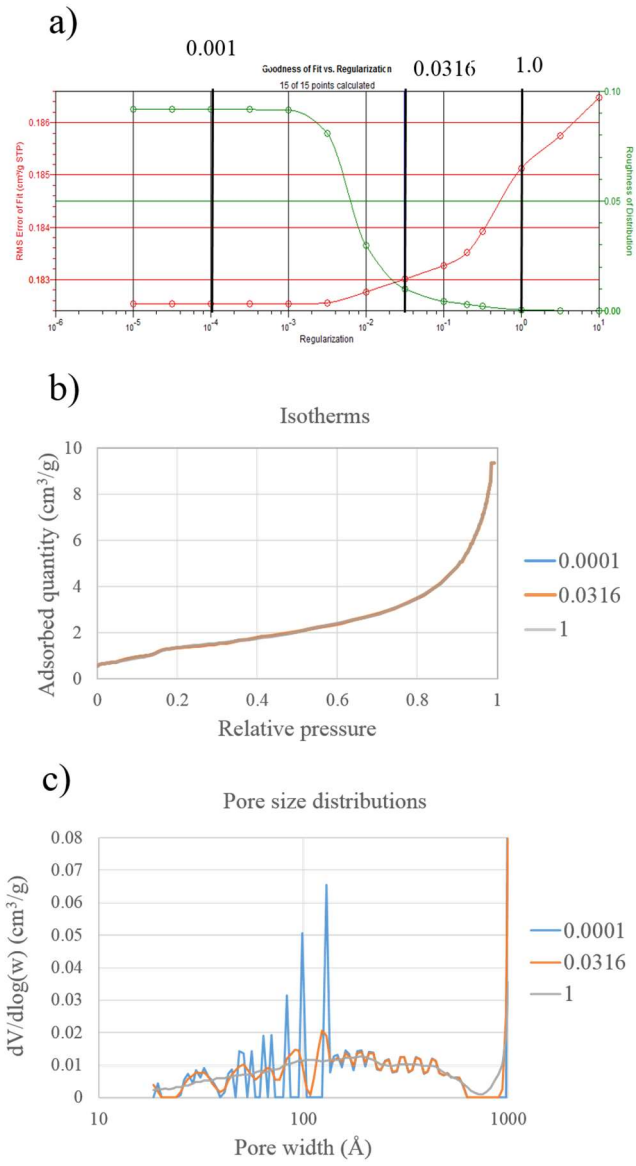

Figure A-1 (a) The error of fit (green) and roughness of the distribution (red) against regularization parameter λ. (b) Reconstructed isotherms given different regularization λ values (0.0001, 0.0316, 1). (c) Pore size distributions different regularization λ values (0.0001, 0.0316, 1).

Appendix B: Documentation for NMR Clustering Source Code

The source code is open sourced and can be found on Github[4]. Details of the documentation can also be found on GitHub. Feel free to email the author (jianghan2013@gmail.com) for any question.

Before you run this code, you need to have python (3.6 recommended) installed. You also need to install packages including numpy, matplotplib, pandas, and sklearn. You may also want to install jupyter notebook to run the notebook file. Those packages can be installed simultaneously if you install using anaconda (https://www.anaconda.com/).

Click 'clone and download' to clone using git in command window or download the zip file.

yourcomputer> git clone https://GitHub.com/jianghan2013/NMR_clustering_paper.git

Once you clone the NMR_clustering_paper folder from the GitHub repo, go to into the folder and you will find 4 folders.

Figure B-1 Files in the GitHub repo.

The test data is stored in the data folder, the processed data is stored in the data_process folder, and core python functions are stored in the utilis folder.

To run your first testing example, open python in command window or create a new jupyter notebook, and follow the step shown below (Figure B-2). A demo notebook can be found in the demo.ipynb file.

4 https://github.com/jianghan2013/NMR_clustering_paper.

First, download this repo.

- You need to have 'python' installed.
- You also need to install 'numpy', 'matplotplib', 'pandas', and 'sklearn'.
- You may also want to install jupyter notebook to run notebook file.

Second, import those libararies in your python environment

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Also import the Model_runner class

```
from utilis.model_runner import Model_runner
```

Creat a instance

```
nmr_cluster = Model_runner()
```

Load data

```
nmr_cluster.load_data()
```

Do preprocessing

```
nmr_cluster.preprocess()
```

Perform clustering

```
nmr_cluster.fit()
```

Figure B-2 Steps to run your first test case.

The ‘Model runner’ class is the top class for this work. It contains functions including ‘__init_’, ‘load_data’, ‘preprocess’, ‘fit’, and ‘evaluate’. In ‘__init__’ function, the user can set parameters for all the rest functions and file locations. You can set the parameters for data preprocessing, clustering model used for clustering, total number of clustering. In ‘load_data’ function, it load input data T_1,T_2, and fluid volume from external files. In ‘preprocess’ function, it converts the input data to data used for clustering. It starts from creating a sub-space of T_1-T_2 domain to the last step where each grid point is

duplicated f_{norm} times to produce the 2-D clustering dataset. In 'fit' function, it initiates cluster model based on sklearn package and performs clustering. In 'evaluate' function, it computes the two evaluation indices.

References[5]

Adesida, A. G., I. Akkutlu, D. E. Resasco, and C. S. Rai. "Characterization of barnett shale kerogen pore size distribution using DFT analysis and grand canonical monte carlo simulations." In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2011.

Addison, Paul S. *Fractals and chaos: an illustrated course*. CRC Press, 1997.

Aggarwal, C. C., and C. K. Reddy, eds. *Data clustering: algorithms and applications*. CRC press, 2013.

Al-Harahsheh, M., O. Al-Ayed, J. Robinson, S. Kingman, A. Al-Harahsheh, K. Tarawneh, A. Saeid, and R. Barranco. "Effect of demineralization and heating rate on the pyrolysis kinetics of Jordanian oil shales." *Fuel Processing Technology* 92, no. 9 (2011): 1805-1811.

Alnoaimi, K. R., C. Duchateau, and A. Kovscek. "Characterization and measurement of multi-scale gas transport in shale core samples." In *Unconventional Resources Technology Conference*, Denver, Colorado, 25-27 August 2014, pp. 1140-1158. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2014.

Amann-Hildenbrand, A., P. Bertier, A. Busch, and B. M. Krooss. "Experimental investigation of the sealing capacity of generic clay-rich caprocks." *International Journal of Greenhouse Gas Control* 19 (2013): 620-641.

Amann-Hildenbrand, A., A. Ghanizadeh, and B. M. Krooss. "Transport properties of unconventional gas systems." *Marine and Petroleum Geology* 31, no. 1 (2012): 90-99.

Ambrose, R. J., R. C. Hartman, M. D. Campos, I. Y. Akkutlu, and C. Sondergeld. "New pore-scale considerations for shale gas in place calculations." In *SPE Unconventional Gas Conference*. Society of Petroleum Engineers, 2010.

Ambrose, R. J., R. C. Hartman, M. D. Campos, I. Y. Akkutlu, and C. H. Sondergeld. "Shale gas-in-place calculations part I: New pore-scale considerations." *SPE Journal* 17(1) (2012): 219-229.

Arthur, J. D., B. Bohm, and M. Layne. "Hydraulic fracturing considerations for natural gas wells of the Marcellus Shale." *Journal of Gulf Coast Association of Geological Societies* 59 (2009): 49-59.

5 The references follow the format of Chicago referencing style.

Baldi, P., and G. W. Hatfield. *DNA microarrays and gene expression: from experiments to data analysis and modeling*. Cambridge university press, 2002.

Ball, G. H., and D. J. Hall. *ISODATA, a novel method of data analysis and pattern classification*. Stanford research inst Menlo Park CA, 1965.

Barrett, E. P., L. G. Joyner, and P. P. Halenda. "The determination of pore volume and area distributions in porous substances. I. Computations from nitrogen isotherms." *Journal of the American Chemical society* 73, no. 1 (1951): 373-380.

Bathurst, R. G. *Carbonate sediments and their diagenesis*. Vol. 12. Elsevier, 1972.

Bernard, S., R. Wirth, A. Schreiber, H. Schulz, and B. Horsfield. "Formation of nanoporous pyrobitumen residues during maturation of the Barnett Shale (Fort Worth Basin)." *International Journal of Coal Geology* 103 (2012): 3-11.

Bhandari, A. R., P. B. Flemings, and P. J. Polito. "Permeability Behavior and the Effective Stress Law for a Partially Fractured Eagle Ford Shale Sample." In *Poromechanics* VI, (2017): 1892-1899.

Bhandari, A. R., P. B. Flemings, P. J. Polito, M. B. Cronin, and S. L. Bryant. "Anisotropy and stress dependence of permeability in the Barnett shale." *Transport in porous media* 108, no. 2 (2015): 393-411.

Bhatia, S. K., and J. S. Deogun. "Conceptual clustering in information retrieval." *IEEE Transactions on Systems, Man, and Cybernetics*, Part B (Cybernetics) 28, no. 3 (1998): 427-436.

Billiotte, J., D. Yang, and K. Su. "Experimental study on gas permeability of mudstones." *Physics and Chemistry of the Earth*, Parts A/B/C 33 (2008): S231-S236.

Birdwell, J. E., and K. E. Washburn. "Multivariate analysis relating oil shale geochemical properties to NMR relaxometry." *Energy & Fuels* 29, no. 4 (2015): 2234-2243.

Bloembergen, N., E. M. Purcell, and R. V. Pound. "Relaxation effects in nuclear magnetic resonance absorption." In *Resonances*, pp. 411-444. 1990.

Brace, W. F., J. B. Walsh, and W. T. Frangos. "Permeability of granite under high pressure." *Journal of Geophysical research* 73, no. 6 (1968): 2225-2236.

Burgess, C. G. V., and D. H. Everett. "The lower closure point in adsorption hysteresis of the capillary condensation type." *Journal of Colloid and Interface Science* 33, no. 4 (1970): 611-614.

Busetti, S., W. Jiao, and Z. Reches. "Geomechanics of hydraulic fracturing microseismicity: Part 1. Shear, hybrid, and tensile events." *AAPG Bulletin* 98, no. 11 (2014): 2439-2457.

Bustin, R. M., A. M. M. Bustin, A. Cui, D. Ross, and V. M. Pathi. "Impact of shale properties on pore structure and storage characteristics." In *SPE shale gas production conference*. Society of Petroleum Engineers, 2008.

Callaghan, P. T. *Principles of nuclear magnetic resonance microscopy*. Oxford University Press on Demand, 1993.

Cao, C., T. Li, J. Shi, L. Zhang, S. Fu, B. Wang, and H. Wang. "A new approach for measuring the permeability of shale featuring adsorption and ultra-low permeability." *Journal of Natural Gas Science and Engineering* 30 (2016): 548-556.

Carr, H. Y., and E. M. Purcell. "Effects of diffusion on free precession in nuclear magnetic resonance experiments." *Physical review* 94, no. 3 (1954): 630.

Carslaw, H. S., and J. C. Jaeger. *Conduction of heat in solids*. Oxford Science Publications. Oxford, England, 1959.

Chalmers, G. RL, D. J. Ross, and R. M. Bustin. "Geological controls on matrix permeability of Devonian Gas Shales in the Horn River and Liard basins, northeastern British Columbia, Canada." *International Journal of Coal Geology* 103 (2012): 120-131.

Cheng, Y. "Mean shift, mode seeking, and clustering." *IEEE transactions on pattern analysis and machine intelligence* 17, no. 8 (1995): 790-799.

Clarkson, C. R., N. Solano, R. M. Bustin, A. M. M. Bustin, G. R. L. Chalmers, L. He, Y. B. Melnichenko, A. P. Radliński, and T. P. Blach. "Pore structure characterization of North American shale gas reservoirs using USANS/SANS, gas adsorption, and mercury intrusion." *Fuel* 103 (2013): 606-616.

Coates, G. R., L. Xiao, and M. G. Prammer. *NMR logging: principles and applications*. Houston: Haliburton Energy Services, 1999.

Comaniciu, D., and P. Meer. "Mean shift: A robust approach toward feature space analysis." *IEEE Transactions on pattern analysis and machine intelligence* 24, no. 5 (2002): 603-619.

Coyner, K., T. J. Katsube, M. E. Best, and M. Williamson. "Gas and water permeability of tight shales from the venture gas field, offshore Nova Scotia." *Papers-Geological Survey of Canada* (1993): 129-129.

Cui, A., and R. Brezovski. "Laboratory permeability and diffusivity measurements of unconventional reservoirs: useless or full of Information? A Montney example from the western Canada sedimentary basin." In *Proceedings of the SPE Unconventional Resources Conference and Exhibition-Asia Pacific*, Brisbane, Australia, pp. 11-13. 2013.

Cui, A., R. Wust, B. Nassichuk, K. Glover, R. Brezovski, and C. Twemlow. "A nearly complete characterization of permeability to hydrocarbon gas and liquid for

unconventional reservoirs: a challenge to conventional thinking." In *Unconventional resources technology conference*, pp. 1716-1732. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2013.

Cui, X., A. M. M. Bustin, and R. M. Bustin. "Measurements of gas permeability and diffusivity of tight reservoir rocks: different approaches and their applications." *Geofluids* 9, no. 3 (2009): 208-223.

Cui, X. A., R. M. Bustin, R. Brezovski, B. Nassichuk, K. Glover, and V. Pathi. "A new method to simultaneously measure in-situ permeability and porosity under reservoir conditions: Implications for characterization of unconventional gas reservoirs." In *Canadian Unconventional Resources and International Petroleum Conference*. Society of Petroleum Engineers, 2010.

Curtis, M. E., R. J. Ambrose, and C. H. Sondergeld. "Structural characterization of gas shales on the micro-and nano-scales." In *Canadian unconventional resources and international petroleum conference*. Society of Petroleum Engineers, 2010.

Daigle, H., N. W. Hayman, H. Jiang, X. Tian, and C. Jiang. "Multiscale pore networks and their effect on deformation and transport property alteration associated with hydraulic fracturing." *Energy Procedia* 125 (2017a): 71-79.

Daigle, H., N. W. Hayman, E. D. Kelly, K. L. Milliken, and H. Jiang. "Fracture capture of organic pores in shales." *Geophysical Research Letters* 44, no. 5 (2017b): 2167-2176.

Daigle, H., C. Jiang, X. Tian, and H. Jiang. "Multiscale Pore Systems in Shales and Their Effect on Laboratory Measurements and Transport Properties." *Unconventional Resources Technology Conference (URTEC)*, 2018.

Daigle, H., A. Johnson, J. P. Gips, and M. Sharma. "Porosity evaluation of shales using NMR secular relaxation." In *Unconventional Resources Technology Conference, Denver, Colorado*, 25-27 August 2014, pp. 1205-1216. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2014.

Day, N. E. "Estimating the components of a mixture of normal distributions." *Biometrika* 56, no. 3 (1969): 463-474.

Dehandschutter, B., S. Vandycke, M. Sintubin, N. Vandenberghe, P. Gaviglio, J-P. Sizun, and L. Wouters. "Microfabric of fractured Boom Clay at depth: a case study of brittle–ductile transitional clay behaviour." *Applied Clay Science* 26, no. 1 (2004): 389-401.

Dempster, A. P., N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the royal statistical society*. Series B (methodological) (1977): 1-38.

Desbois, G., J. L. Urai, and P. A. Kukla. "Morphology of the pore space in claystones–evidence from BIB/FIB ion beam sectioning and cryo-SEM observations." *eEarth Discussions* 4, no. 1 (2009): 1-19.

Dicker, A. I., and R. M. Smits. "A practical approach for determining permeability from laboratory pressure-pulse decay measurements." In *International meeting on petroleum engineering*. Society of Petroleum Engineers, 1988.

Dong, M., Z. Li, S. Li, and J. Yao. "Permeabilities of tight reservoir cores determined for gaseous and liquid CO2 and C2H6 using minimum backpressure method." *Journal of Natural Gas Science and Engineering* 5 (2012): 1-5.

Dusseault, M. B. "Geomechanical challenges in petroleum reservoir exploitation." *KSCE Journal of Civil Engineering* 15, no. 4 (2011): 669-678.

Egermann, P., R. Lenormand, D. G. Longeron, and C. Zarcone. "A fast and direct method of permeability measurements on drill cuttings." *SPE Reservoir Evaluation & Engineering* 8, no. 04 (2005): 269-275.

EIA. *Annual Energy Outlook 2018, With Projections to 2050*. Washington, DC: Energy Information Administration; 2018.

Eliyahu, M., S. Emmanuel, R. J. Day-Stirrat, and C. I. Macaulay. "Mechanical properties of organic matter in shales mapped at the nanometer scale." *Marine and Petroleum Geology* 59 (2015): 294-304.

Emmanuel, S., M. Eliyahu, R. J. Day-Stirrat, R. Hofmann, and C. I. Macaulay. "Impact of thermal maturation on nano-scale elastic properties of organic matter in shales." *Marine and Petroleum Geology* 70 (2016): 175-184.

Escoffier, S., F. Homand, A. Giraud, N. Hoteit, and K. Su. "Under stress permeability determination of the Meuse/Haute-Marne mudstone." *Engineering Geology* 81, no. 3 (2005): 329-340.

Ester, M., H. Kriegel, J. Sander, and X. Xu. "A density-based algorithm for discovering clusters in large spatial databases with noise." In *Kdd*, vol. 96, no. 34, pp. 226-231. 1996.

Fedor, F., G. Hámos, A. Jobbik, Z. Máthé, G. Somodi, and I. Szűcs. "Laboratory pressure pulse decay permeability measurement of Boda Claystone, Mecsek Mts., SW Hungary." *Physics and Chemistry of the Earth*, Parts A/B/C 33 (2008): S45-S53.

Firouzi, M., K. Alnoaimi, A. Kovscek, and J. Wilcox. "Klinkenberg effect on predicting and measuring helium permeability in gas shales." *International Journal of Coal Geology* 123 (2014): 62-68.

Fisher, L. R., R. A. Gamble, and J. Middlehurst. "The Kelvin equation and the capillary condensation of water." *Nature* 290 (1981): 575-576.

Fleury, M., and M. Romero-Sarmiento. "Characterization of shales using T1–T2 NMR maps." *Journal of Petroleum Science and Engineering* 137 (2016): 55-62.

Fraley, C., and A. E. Raftery. MCLUST version 3: *an R package for normal mixture modeling and model-based clustering*. WASHINGTON UNIV SEATTLE DEPT OF STATISTICS, 2006.

Frenkel, J. *Kinetic Theory of Liquids*. Oxford University Press, 1946.

Fukunaga, K., and L. Hostetler. "The estimation of the gradient of a density function, with applications in pattern recognition." *IEEE Transactions on information theory* 21, no. 1 (1975): 32-40.

Gan, G., C.n Ma, and J. Wu. *Data clustering: theory, algorithms, and applications*. Vol. 20. Siam, 2007.

Gao, J., and Z. Li. "Water saturation-driven evolution of helium permeability in Carboniferous shale from Qaidam Basin, China: An experimental study." *Marine and Petroleum Geology* (2018).

Gensterblum, Y., A. Ghanizadeh, and B. M. Krooss. "Gas permeability measurements on Australian subbituminous coals: fluid dynamic and poroelastic aspects." *Journal of Natural Gas Science and Engineering* 19 (2014): 202-214.

Ghanizadeh, A., S. Bhowmik, O. Haeri-Ardakani, H. Sanei, and C. R. Clarkson. "A comparison of shale permeability coefficients derived using multiple non-steady-state measurement techniques: Examples from the Duvernay Formation, Alberta (Canada)." *Fuel* 140 (2015): 371-387.

Gips, J. P., H. Daigle, and M. Sharma. "Characterization of Free and Bound Fluids in Hydrocarbon Bearing Shales Using NMR and Py GC-MS." In *Unconventional Resources Technology Conference*, Denver, Colorado, 25-27 August 2014, pp. 1217-1225. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2014.

Gregg, S. J., and K. S. W. Sing. *Adsorption, surface area and porosity*, Acad. Press, London 3 (1982).

Groen, J. C., L. A. Peffer, and J. Pérez-Ramírez. "Pore size determination in modified micro-and mesoporous materials. Pitfalls and limitations in gas adsorption data analysis." *Microporous and Mesoporous Materials* 60, no. 1-3 (2003): 1-17.

Grunsky, E. C. "The interpretation of geochemical survey data." *Geochemistry: Exploration, Environment, Analysis* 10, no. 1 (2010): 27-74.

Halsey, G. D. “Physical adsorption on non-uniform surfaces.” *The Journal of Chemical Physics* 16 (1948): 931-937.

Handwerger, D. A., J. Keller, and K. Vaughn. "Improved petrophysical core measurements on tight shale reservoirs using retort and crushed samples." In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2011.

Hansen, P. C. *Discrete inverse problems: insight and algorithms.* Vol. 7. Siam, 2010.

Heller, R., J. "Statistical mechanics of multimolecular adsorption. I." *The Journal of Chemical Physics* 14 (1946): 263-267.

Hu, H., T. Zhang, J. D. Wiggins-Camacho, G. S. Ellis, M. D. Lewan, and X. Zhang. "Experimental investigation of changes in methane adsorption of bitumen-free Woodford Shale with thermal maturation induced by hydrous pyrolysis." *Marine and Petroleum Geology* 59 (2015): 114-128.

Jain, A. K. "Data clustering: 50 years beyond K-means." *Pattern recognition letters* 31, no. 8 (2010): 651-666.

Jain, A. K., M. N. Murty, and P. J. Flynn. "Data clustering: a review." *ACM computing surveys (CSUR)* 31, no. 3 (1999): 264-323.

Jain, V., C. C. Minh, N. Heaton, P. Ferraris, L. Ortenzi, and M. T. Ribeiro. "Characterization of underlying pore and fluid structure using factor analysis on NMR data." In *SPWLA 54th Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts, 2013.

Jaroniec, M., M. Kruk, and J. Olivier. "Fractal analysis of composite adsorption isotherms obtained by using density functional theory data for argon in slitlike pores." *Langmuir* 13, no. 5 (1997): 1031-1035.

Jarvie, D. M., R. J. Hill, T. E. Ruble, and R. M. Pollastro. "Unconventional shale-gas systems: The Mississippian Barnett Shale of north-central Texas as one model for thermogenic shale-gas assessment." *AAPG bulletin* 91, no. 4 (2007): 475-499.

Jiang, C., H. Daigle, and S. Bryant. "A bundle of short conduits model of the pore structure of gas shale." In *Unconventional Resources Technology Conference*, San Antonio, Texas, 20-22 July 2015, pp. 1530-1543. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2015.

Jiang, F., D. Chen, J. Chen, Q. Li, Y. Liu, X. Shao, T. Hu, and J. Dai. "Fractal analysis of shale pore structure of continental gas shale reservoir in the Ordos Basin, NW China." *Energy & Fuels* 30, no. 6 (2016): 4676-4689.

Jiang, H., H. Daigle, N. W. Hayman, and K. L. Milliken. "Porosity-deformation relationships in organic-rich shale", AAPG memoir (2018a). (in press)

Jiang, H., H. Daigle, B. Zhang, and X. Tian. "Characterization of thermal evolution of pores and fluids in shales using NMR 2D measurement." In *Unconventional Resources Technology Conference*, Houston, Texas, 23-25 July 2018, pp. 1961-

1970. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2018b.

Johri, M., and M. D. Zoback. "The evolution of stimulated reservoir volume during hydraulic stimulation of shale gas formations." In *Unconventional Resources Technology Conference*, pp. 1661-1671. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2013.

Jones, C., and P. Meredith. "Experimental study of elastic wave propagation anisotropy and permeability anisotropy in an illitic shale." In *Proceedings of the SPE/ISRM Rock Mechanics in Petroleum Engineering Conference*, vol. 2, pp. 307-313. 1998.

Josh, M., L. Esteban, C. D. Piane, J. Sarout, D. N. Dewhurst, and M. B. Clennell. "Laboratory characterisation of shale properties." *Journal of Petroleum Science and Engineering* 88 (2012): 107-124.

Kadlec, O., and M. M. Dubinin. "Comments on the limits of applicability of the mechanism of capillary condensation." *Journal of Colloid and Interface Science* 31, no. 4 (1969): 479-489.

King, B. "Step-wise clustering procedures." *Journal of the American Statistical Association* 62, no. 317 (1967): 86-101.

Ko, L. T., R. G. Loucks, T. Zhang, S. C. Ruppel, and D. Shao. "Pore and pore network evolution of Upper Cretaceous Boquillas (Eagle Ford–equivalent) mudrocks: Results from gold tube pyrolysis experiments." *AAPG Bulletin* 100, no. 11 (2016): 1693-1722.

Ko, L. T., S. C. Ruppel, R. G. Loucks, P. C. Hackley, T. Zhang, and D. Shao. "Pore-types and pore-network evolution in Upper Devonian-Lower Mississippian Woodford and Mississippian Barnett mudstones: Insights from laboratory thermal maturation and organic petrology." *International Journal of Coal Geology* 190 (2018): 3-28.

Ko, L., T. Zhang, R. G. Loucks, and S. C. Ruppel. "Boquillas (Eagle Ford) Formation pore evolution results from laboratory heating experiments." In *Unconventional Resources Technology Conference*, Denver, Colorado, 25-27 August 2014, pp. 1006-1012. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2014.

Kruk, M., and M. Jaroniec. "Gas adsorption characterization of ordered organic− inorganic nanocomposite materials." *Chemistry of Materials* 13, no. 10 (2001): 3169-3183.

Kuila, U., D. K. McCarty, A. Derkowski, T. B. Fischer, T. Topór, and M. Prasad. "Nano-scale texture and porosity of organic matter and clay minerals in organic-rich mudrocks." *Fuel* 135 (2014): 359-373.

Kuila, U., and M. Prasad. "Application of nitrogen gas-adsorption technique for characterization of pore structure of mudrocks." *The Leading Edge* 32, no. 12 (2013a): 1478-1485.

Kuila, U., and M. Prasad. "Specific surface area and pore-size distribution in clays and shales." *Geophysical Prospecting* 61, no. 2 (2013b): 341-362.

Kumar, H., D. Elsworth, J. P. Mathews, and C. Marone. "Permeability evolution in sorbing media: analogies between organic-rich shale and coal." *Geofluids* 16, no. 1 (2016): 43-55.

Kwon, O., A. K. Kronenberg, A. F. Gangi, and B. Johnson. "Permeability of Wilcox shale and its effective pressure law." *Journal of Geophysical Research: Solid Earth* 106, no. B9 (2001): 19339-19353.

Lastoskie, C., Keith E. Gubbins, and Nicholas Quirke. "Pore size distribution analysis of microporous carbons: a density functional theory approach." The journal of physical chemistry 97, no. 18 (1993): 4786-4796.

Laurich, Ben, J. L. Urai, G. Desbois, C. Vollmer, and C. Nussbaum. "Microstructural evolution of an incipient fault zone in Opalinus Clay: Insights from an optical and electron microscopic study of ion-beam polished samples from the Main Fault in the Mt-Terri Underground Research Laboratory." *Journal of structural geology* 67 (2014): 107-128.

Lawson, C. L., and R. J. Hanson. *Solving least squares problems*. Vol. 15. Siam, 1995.

Lewan, M. D. "Laboratory simulation of petroleum formation." In Organic geochemistry, pp. 419-442. Springer, Boston, MA, 1993.

Lewan, M. D. "Experiments on the role of water in petroleum formation." Geochimica et Cosmochimica Acta 61, no. 17 (1997): 3691-3723.

Liu, X., J. Xiong, and L. Liang. "Investigation of pore structure and fractal characteristics of organic-rich Yanchang formation shale in central China by nitrogen adsorption/desorption analysis." *Journal of Natural Gas Science and Engineering* 22 (2015): 62-72.

Loucks, R. G., and R. M. Reed. "Scanning-electron-microscope petrographic evidence for distinguishing organic-matter pores associated with depositional organic matter versus migrated organic matter in mudrock." *GCAGS Journal* (2014): 51-60.

Loucks, R. G., R. M. Reed, S. C. Ruppel, and U. Hammes. "Spectrum of pore types and networks in mudrocks and a descriptive classification for matrix-related mudrock pores." *AAPG bulletin* 96, no. 6 (2012): 1071-1098.

Loucks, R. G., R. M. Reed, S. C. Ruppel, and D. M. Jarvie. "Morphology, genesis, and distribution of nanometer-scale pores in siliceous mudstones of the Mississippian Barnett Shale." *Journal of sedimentary research* 79, no. 12 (2009): 848-861.

Loucks, R. G., and S. C. Ruppel. "Mississippian Barnett Shale: Lithofacies and depositional setting of a deep-water shale-gas succession in the Fort Worth Basin, Texas." *AAPG bulletin* 91, no. 4 (2007): 579-601.

Lu, Z., S. Narayanan, and E. N. Wang. "Modeling of evaporation from nanopores with nonequilibrium and nonlocal effects." *Langmuir* 31, no. 36 (2015): 9817-9824.

Luffel, D. L., C. W. Hopkins, and P. D. Schettler Jr. "Matrix permeability measurement of gas productive shales." In *SPE annual technical conference and exhibition*. Society of Petroleum Engineers, 1993.

MacQueen, J. "Some methods for classification and analysis of multivariate observations." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281-297. 1967.

Mason, G.. "The effect of pore space connectivity on the hysteresis of capillary condensation in adsorption—desorption isotherms." *Journal of Colloid and Interface Science* 88, no. 1 (1982): 36-46.

Mastalerz, M., A. Schimmelmann, A. Drobniak, and Y. Chen. "Porosity of Devonian and Mississippian New Albany Shale across a maturation gradient: Insights from organic petrology, gas adsorption, and mercury intrusion." *AAPG bulletin* 97, no. 10 (2013): 1621-1643.

Maxwell, S. C., and C. L. Cipolla. "What does microseismicity tell us about hydraulic fracturing?." In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2011.

McLachlan, G., and T. Krishnan. *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons, 2007.

McLachlan, G. J., D. Peel, K. E. Basford, and P. Adams. "The EMMIX software for the fitting of mixtures of normal and t-components." *Journal of Statistical Software* 4, no. 2 (1999): 1-14.

Medellín, D., V.R. Ravi, and C. Torres-Verdín. "New Methods for Interpretation of NMR Measurements Acquired in Spatially Heterogeneous Rocks." In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2015.

Medellín, D., V. R. Ravi, and C. Torres-Verdín. "Multidimensional NMR inversion without Kronecker products: Multilinear inversion." *Journal of Magnetic Resonance* 269 (2016): 24-35.

Mehana, M., and I. El-monier. "Shale characteristics impact on Nuclear Magnetic Resonance (NMR) fluid typing methods and correlations." *Petroleum* 2, no. 2 (2016): 138-147.

Mehmani, A., and M. Prodanović. "The application of sorption hysteresis in nano-petrophysics using multiscale multiphysics network models." *International Journal of Coal Geology* 128-129 (2014): 96-108.

Mehmani, A., M. Prodanović, and F. Javadpour. "Multiscale, multiphysics network modeling of shale matrix gas flows." *Transport in porous media* 99, no. 2 (2013): 377-390.

Meiboom, S., and D. Gill. "Modified spin-echo method for measuring nuclear relaxation times." *Review of scientific instruments* 29, no. 8 (1958): 688-691

Milliken, K. "A compositional classification for grain assemblages in fine-grained sediments and sedimentary rocks." *Journal of Sedimentary Research* 84, no. 12 (2014): 1185-1199.

Milliken, K. L., S. M. Ergene, and A. Ozkan. "Quartz types, authigenic and detrital, in the Upper Cretaceous Eagle Ford Formation, south Texas, USA." *Sedimentary Geology* 339 (2016): 273-288.

Milliken, K. L., and T. Olson. "Silica Diagenesis, Porosity Evolution, and Mechanical Behavior In Siliceous Mudstones, Mowry Shale (Cretaceous), Rocky Mountains, USAKL MILLIKEN AND T. OLSONCEMENT IN SILICEOUS MUDSTONE." *Journal of Sedimentary Research* 87, no. 4 (2017): 366-387.

Milliken, K. L., M. Rudnicki, D. N. Awwiller, and T. Zhang. "Organic matter–hosted pore system, Marcellus formation (Devonian), Pennsylvania." *AAPG bulletin* 97, no. 2 (2013): 177-200.

Mokhtari, M., and A. N. Tutuncu. "Characterization of anisotropy in the permeability of organic-rich shales." *Journal of Petroleum Science and Engineering* 133 (2015): 496-506.

Nicot, B., N. Vorapalawut, B. Rousseau, L. F. Madariaga, G. Hamon, and J. Korb. "Estimating saturations in organic shales using 2D NMR." *Petrophysics* 57, no. 01 (2016): 19-29.

Nolte, K. G. *Reservoir stimulation*. Edited by Michael J. Economides. Vol. 18. New York: Wiley, 2000.

Mullen, J. "Petrophysical characterization of the Eagle Ford Shale in south Texas." In *Canadian Unconventional Resources and International Petroleum Conference*. Society of Petroleum Engineers, 2010.

Murtagh, F. "A survey of recent advances in hierarchical clustering algorithms." *The Computer Journal* 26, no. 4 (1983): 354-359.

Narayanan, S., A. G. Federov, and Y. K. Joshi. “Interfacial transport of evaporating water contained in nanopores.” *Langmuir* 27, no. 17 (2011): 10666-10676.

Neal, R. M., and G. E. Hinton. "A view of the EM algorithm that justifies incremental, sparse, and other variants." In *Learning in graphical models*, pp. 355-368. Springer, Dordrecht, 1998.

Ng, A. Y., M. I. Jordan, and Y. Weiss. "On spectral clustering: Analysis and an algorithm." In *Advances in neural information processing systems*, pp. 849-856. 2002.

Odusina, E. O., C. H. Sondergeld, and C. S. Rai. "NMR study of shale wettability." In *Canadian Unconventional Resources Conference*. Society of Petroleum Engineers, 2011.

Passey, Q. R., K. Bohacs, W. L. Esch, R. Klimentidis, and S. Sinha. "From oil-prone source rock to gas-producing shale reservoir-geologic and petrophysical characterization of unconventional shale gas reservoirs." In *International oil and gas conference and exhibition in China*. Society of Petroleum Engineers, 2010.

Patzek, T. W., F. Male, and M. Marder. "Gas production in the Barnett Shale obeys a simple scaling theory." *Proceedings of the National Academy of Sciences* 110, no. 49 (2013): 19731-19736.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas. "Scikit-learn: Machine learning in Python." *Journal of machine learning research* 12, no. Oct (2011): 2825-2830.

Pinson, M. B., T. Zhou, H. M. Jennings, and M. Z. Bazant. "Inferring pore connectivity from sorption hysteresis in multiscale porous media." *Journal of colloid and interface science* 532 (2018): 118-127.

Pommer, M., and K. Milliken. "Pore types and pore-size distributions across thermal maturity, Eagle Ford Formation, southern Texas." *AAPG Bulletin* 99, no. 9 (2015): 1713-1744.

Ravikovitch, P. I., and A. V. Neimark. "Experimental confirmation of different mechanisms of evaporation from ink-bottle type pores: equilibrium, pore blocking, and cavitation." *Langmuir* 18, no. 25 (2002): 9830-9837.

Romero-Sarmiento, M., S. Ramiro-Ramirez, G. Berthe, M. Fleury, and R. Littke. "Geochemical and petrophysical source rock characterization of the Vaca Muerta Formation, Argentina: Implications for unconventional petroleum resource estimations." *International Journal of Coal Geology* 184 (2017): 27-41.

Roque-Malherbe, R. M. A. *Adsorption and diffusion in nanoporous materials*. CRC press, 2007.

Ross, D. J., and R. M. Bustin. "The importance of shale composition and pore structure upon gas storage potential of shale gas reservoirs." *Marine and Petroleum Geology* 26, no. 6 (2009): 916-927.

Roux, P. F. "Microseismic activity: What does it really say? Insights from Coulomb failure function analysis." In *Unconventional Resources Technology Conference, San Antonio, Texas, 1-3 August 2016*, pp. 1978-1991. Society of Exploration

Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2016.

Sahami, M., and D. Koller. "Using machine learning to improve information access." PhD diss., Stanford University, Department of Computer Science, 1998.

Sander, R., Z. Pan, and L. D. Connell. "Laboratory measurement of low permeability unconventional gas reservoir rocks: A review of experimental methods." *Journal of Natural Gas Science and Engineering* 37 (2017): 248-279.

Sasaki, H., T. Kanamori, A. Hyvärinen, G. Niu, and M. Sugiyama. "Mode-Seeking Clustering and Density Ridge Estimation via Direct Estimation of Density-Derivative-Ratios." *arXiv preprint* arXiv:1707.01711 (2017).

Schneider, J., P. B. Flemings, R. J. Day-Stirrat, and J. T. Germaine. "Insights into pore-scale controls on mudstone permeability through resedimentation experiments." *Geology* 39, no. 11 (2011): 1011-1014.

Seaton, N. A. "Determination of the connectivity of porous solids from nitrogen sorption measurements." *Chemical Engineering Science* 46, no. 8 (1991): 1895-1909.

Shi, J., and J. Malik. "Normalized cuts and image segmentation." *IEEE Transactions on pattern analysis and machine intelligence* 22, no. 8 (2000): 888-905.

Sing, K. S. "Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (Recommendations 1984)." *Pure and applied chemistry* 57, no. 4 (1985): 603-619.

Singer, P. M., Z. Chen, and G. J. Hirasaki. "Fluid typing and pore size in organic shale using 2D NMR in saturated kerogen isolates." Petrophysics 57, no. 06 (2016): 604-619.

Slatt, R. M., and N. R. O'Brien. "Pore types in the Barnett and Woodford gas shales: Contribution to understanding gas storage and migration pathways in fine-grained rocks." *AAPG bulletin* 95, no. 12 (2011): 2017-2030.

Sneath, P., and R. Sokal. *Numerical Taxonomy*. Freeman, London. (1973).

Sokołowska, Z., M. Hajnos, C. Hoffmann, M. Renger, and S. Sokołowski. "Comparison of fractal dimensions of soils estimated from adsorption isotherms, mercury intrusion, and particle size distribution." *Journal of Plant Nutrition and Soil Science* 164, no. 5 (2001): 591-599.

Steinhaus, H. "Sur la division des corp materiels en parties." *Bull. Acad. Polon. Sci* 1, no. 804 (1956): 801.

Stella, X. Yu, and J. Shi. "Multiclass spectral clustering." In null, p. 313. IEEE, 2003.

Tahmasebi, P., F. Javadpour, and M. Sahimi. “Multiscale and multiresolution modeling of shales and their flow and morphological properties.” *Nature Scientific Reports* 5 (2015): 16373.

Tanev, P. T., and L. T. Vlaev. "An attempt at a more precise evaluation of the approach to mesopore size distribution calculations depending on the degree of pore blocking." Journal of colloid and interface science 160, no. 1 (1993): 110-116.

Tarazona, P. "Free-energy density functional for hard spheres." *Physical Review* A 31, no. 4 (1985): 2672.

Tarazona, P., and L. Vicente. "A model for density oscillations in liquids between solid walls." *Molecular Physics* 56, no. 3 (1985): 557-572.

Templ, M., P. Filzmoser, and C. Reimann. "Cluster analysis applied to regional geochemical data: problems and possibilities." *Applied Geochemistry* 23, no. 8 (2008): 2198-2213.

Thommes, M., B. Smarsly, M. Groenewolt, P. I. Ravikovitch, and A. V. Neimark. "Adsorption hysteresis of nitrogen and argon in pore networks and characterization of novel micro-and mesoporous silicas." *Langmuir* 22, no. 2 (2006): 756-764.

Tian, X., and H. Daigle. "Machine-learning-based object detection in images for reservoir characterization: A case study of fracture detection in shales." *The Leading Edge* 37, no. 6 (2018a): 435-442.

Tian, X., and H. Daigle. "Permeability prediction from a pore-scale network model constrained by low-pressure nitrogen sorption isotherms." *Journal of Petroleum Science and Engineering* 162 (2018b): 554-566.

Tinni, A., E. Fathi, R. Agarwal, C. H. Sondergeld, I. Y. Akkutlu, and C. S. Rai. "Shale permeability measurements on plugs and crushed samples." In *SPE Canadian Unconventional Resources Conference*. Society of Petroleum Engineers, 2012.

Tinni, A., E. Odusina, I. Sulucarnain, C. Sondergeld, and C. Rai. "NMR response of brine, oil and methane in organic rich shales." In *SPE Unconventional Resources Conference. Society of Petroleum Engineers*, 2014.

Tissot, B., and D. Welte. *Petroleum formation and occurrence: a new approach to oil and gas exploration.* Springer Science & Business Media, 2012.

Velde, B. "Compaction trends of clay-rich deep sea sediments." *Marine Geology* 133, no. 3-4 (1996): 193-201.

Venkataramanan, L., N. Evirgen, D. F. Allen, A. Mutina, Q. Cai, A. C. Johnson, A. Y. Green, and T. Jiang. "An Unsupervised Learning Algorithm To Compute Fluid Volumes From NMR T 1-T 2 Logs in Unconventional Reservoirs." In *SPWLA 59th Annual Logging Symposium*. Society of Petrophysicists and Well-Log Analysts, 2018.

Wang, G., and Y. Ju. "Organic shale micropore and mesopore structure characterization by ultra-low pressure N2 physisorption: Experimental procedure and interpretation model." *Journal of Natural Gas Science and Engineering* 27 (2015): 452-465.

Wang, M., H. Xue, S. Tian, R. W. Wilkins, and Z. Wang. "Fractal characteristics of Upper Cretaceous lacustrine shale from the Songliao Basin, NE China." *Marine and Petroleum Geology* 67 (2015): 144-153.

Wang, Q., X. Chen, A. N. Jha, and H. Rogers. "Natural gas from shale formation–the evolution, evidences and challenges of shale gas revolution in United States." *Renewable and Sustainable Energy Reviews* 30 (2014): 1-28.

Ward Jr, J. H. "Hierarchical grouping to optimize an objective function." *Journal of the American statistical association* 58, no. 301 (1963): 236-244.

Washburn, K. E., and J. E. Birdwell. "Updated methodology for nuclear magnetic resonance characterization of shales." *Journal of Magnetic Resonance* 233 (2013): 17-28.

Washburn, K. E., and Y. Cheng. "Detection of intermolecular homonuclear dipolar coupling in organic rich shale by transverse relaxation exchange." *Journal of Magnetic Resonance* 278 (2017): 18-24.

Webb, P. A., and C. Orr. "Surface area and pore structure by gas adsorption." *Analytical methods in fine particle technology* (1997): 53-153.

Williams-Stroud, S., C. Ozgen, and R. L. Billingsley. "Microseismicity-constrained discrete fracture network models for stimulated reservoir simulation." *Geophysics* 78, no. 1 (2013): B37-B47.

Xie, Ranhong, and Lizhi Xiao. "Advanced fluid-typing methods for NMR logging." Petroleum Science 8, no. 2 (2011): 163-169.

Xu, R., M. Prodanović, and C. J. Landry. "Simulation of Gas Adsorption and Capillary Condensation in Shale Nanopores Using Lattice Boltzmann Modeling." In *Unconventional Resources Technology Conference*, Houston, Texas, 23-25 July 2018, pp. 1986-2000. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, 2018.

Yang, F., Z. Ning, and H. Liu. "Fractal characteristics of shales from a shale gas reservoir in the Sichuan Basin, China." *Fuel* 115 (2014): 378-384.

Yeung, K. Y., M. Medvedovic, and R. E. Bumgarner. "Clustering gene-expression data with repeated measurements." *Genome biology* 4, no. 5 (2003): R34.

Zhang, R., Z. Ning, F. Yang, X. Wang, H. Zhao, and Q. Wang. "Impacts of nanopore structure and elastic properties on stress-dependent permeability of gas shales." *Journal of Natural Gas Science and Engineering* 26 (2015): 1663-1672.

Zhang, R., Z. Ning, F. Yang, H. Zhao, and Q. Wang. "A laboratory study of the porosity-permeability relationships of shale and sandstone under effective stress." *International Journal of Rock Mechanics and Mining Sciences* 81 (2016): 19-27