UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

STRATIGRAPHIC AND DIAGENETIC CONTROLS ON PETROFACIES AND RESERVOIR-QUALITY VARIABILITY USING SEMI-SUPERVISED AND SUPERVISED MACHINE LEARNING METHODS: SYCAMORE FORMATION, SHO-VEL-TUM FIELD, OKLAHOMA, USA

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

DAVID ENRIQUE DUARTE CORONADO

Norman, Oklahoma

2021

STRATIGRAPHIC AND DIAGENETIC CONTROLS ON PETROFACIES AND RESERVOIR-QUALITY VARIABILITY USING SEMI-SUPERVISED AND SUPERVISED MACHINE LEARNING METHODS: SYCAMORE FORMATION, SHO-VEL-TUM FIELD, OKLAHOMA, USA

A DISSERTATION APPROVED FOR THE

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Matthew Pranter, Chair

Dr. Douglas Elmore

Dr. Deepak Devegowda

Dr. Heather Bedle

To my family and mentors

In memory of Dr. Roger Slatt

Acknowledgements

I would like to express my deepest gratitude to my advisor Dr. Matthew Pranter for his expertise, motivation, and continuous support. His guidance helped me throughout the research. To the committee members, Dr. Richard Elmore, Dr. Deepak Devegowda, and Dr. Heather Bedle. I am incredibly grateful to you for being part of my committee and for your valuable feedback. I wish to extend my gratitude to my coauthors Dr. Rafael Pires da Lima, Dr. Benmadi Milad, and Dr. Javier Tellez for their contribution and insights. Also, thanks to all my colleagues at the University of Oklahoma for the valuable discussions we had at Sarkeys. I am very thankful for my mentor, friend, and former advisor Dr. Roger Slatt (RIP), who gave me the opportunity to learn more about geology, and his work ethic, generosity, friendship, and enthusiasm towards knowledge. Besides, I would like to thank the OPIC and School of Geoscience staff for their valuable assistance and support, especially Rebecca for their constant support with all the forms and letters.

I would like to thank all the friends I made during the five and a half years I spent in Norman, for all the help and support they provided. Also, thanks to my friends in Colombia for always being there and making me laugh. I will not try to list their names, but I am sure they know this acknowledgment is for them. Finally, but not least, thanks to my Mom, Dad, and Sister, thank you for always being there for me. To Alejandra for being the best partner, anyone could ever have. This work would have never been possible without the help, constant support, guidance, and encouragement of you four.

Abstract

Diagenetic processes in sedimentary rocks have intrigued geologists, but they are poorly understood. In sedimentary fine-grained rocks, there is even less information, due to the complexity and lack of interest despite the current shale boom that has resulted in increased interest in mudrocks. This study aims to explain the geological processes that affect the reservoir quality on the Sycamore siltstones and present a model that shows the spatial distribution of the thin-section-derived petrofacies for the Sycamore Formation.

This study is divided into four chapters. In Chapter 1, I will introduce my motivation, the research questions, the data available, the contributions, and the limitations of the dissertation. Additionally, in this chapter, I provide a literature review on the geology of the Mississippian rocks in southern Oklahoma, our current understanding of the diagenesis on mudrocks, and a brief explanation of the most common machine learning techniques I will use in chapters 3 and 4.

In Chapter 2, I present a diagenetic study of a mixed-carbonate siliciclastic system. This study evaluates and explains the impact of diagenesis on reservoir quality of the Sycamore Formation and identifies other intrinsic geological factors that control diagenesis. Cementation and dissolution are the main diagenetic processes that impact the reservoir quality of the massive siltstones. Additionally, differences in the sediment supply between carbonate- and siliciclastic-dominated systems drive the calcite cementation and dissolution of feldspars in the Sycamore and consequently the reservoir quality.

In Chapter 3, I propose a workflow that uses machine learning techniques to bridge the resolution gap between thin sections and well logs. To that end, we tested two semi-supervised approaches that use XRF data as input in combination with dimensionality reduction techniques to reliably classify the thin-section-based petrofacies. Both the semi-supervised approaches we

use, a self-training approach and a labeled-clustering approach, achieve accuracies of more than 90% on this dataset. Therefore, implementing the proposed workflow we generate petrofacies logs for cored wells.

In Chapter 4, I present a machine learning-based workflow for assisting 3D models to reliably represent the petrofacies distribution in the Sycamore Formation. The workflow compares over 1,800 classification models and selects the best combination of well logs, algorithms, and hyperparameters to predict petrofacies in wells without cores. From 6 cored wells, the selected classification model predicted petrofacies logs on 75 wells. These predicted petrofacies logs were used to construct a 3D data-driven model.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction

Research questions and scope of the thesis

Diagenetic processes in sedimentary rocks have intrigued geologists, but they are poorly understood. In sedimentary fine-grained rocks, there is even less information, due to the complexity and lack of interest despite the current shale boom that has resulted in increased interest in mudrocks. This study addresses two main questions: What are the diagenetic processes that affect reservoir quality in the Sycamore Formation, especially in terms of the porosity and permeability? And what is the spatial distribution of the diagenetic products in the Sycamore Siltstones?

When addressing the first question, what are the diagenetic processes that affect reservoir quality in the Sycamore Formation? The first observation is that authigenic calcite decrease the porosity of siltstones. Hence, this study focuses on the calcite distribution in the siltstones, and it considers the hypothesis authigenic calcite is the main mechanism that controls the porosity and permeability of the Sycamore Formation. In order to test this hypothesis, we used conventional techniques such as petrographic analysis, elemental and mineralogical composition of the whole rock, and routine core analysis.

For the second question: what is the spatial distribution of the diagenetic products in the Sycamore Siltstones? We hypothesize that depositional processes drive the authigenic calcite distribution. However, the challenge remains in linking the information from the thin-sections scale with 3D models at a reservoir scale. For this, we propose to use machine learning techniques to bridge the resolution gap between thin-sections and 3D models to visualize the spatial variability of calcite-rich and -poor facies in the Sycamore.

The Sycamore Formation was selected for three main reasons. First, the Sycamore provides a unique example of clay and silt-rich mudstones. This study pays special attention to the siltstones because by studying them, one could fill the gap in knowledge between the diagenetic processes that are well studied on sandstones and the challenges brought by studying fine-grained rocks. Second, the Sycamore Formation is a mixed carbonate-siliciclastic system. This presents a unique opportunity to study the characteristic elements and processes of calcareous and siliciclastic systems that are coexisting. Thirdly, we selected the Sycamore because it has gained attention as an unconventional reservoir (Price et al., 2019; Milad et al., 2020; Duarte et al., 2021), as well as to reconstruct paleoenvironments (McGlannan et al., 2019).

This study is divided into two main portions, and it contributes to the understanding of diagenesis in the Sycamore Formation. The first portion, chapter 2, presents a unique diagenetic study of a fine-grained and mixed-carbonate siliciclastic system. This study presents the paragenetic sequence and a conceptual model to show the spatial distribution of calcite-poor and -rich siltstones in the Sycamore Formation. The second portion, divided into chapters 3 and 4, implements a workflow to build a robust 3D model of three thin-section-derived petrofacies. This workflow includes machine learning techniques to fill the resolution gap between scarce core data and the well log information. Therefore, the implementation of this workflow represents an advance on techniques to bridge the gap between information at a thin-section scale with 3D models at a reservoir scale.

Data available

More than 1000 ft. (300 m) of core distributed in six wells, and additionally one outcrop were used for this study; each core has well logs and at least one-foot resolution hand-held XRF:

- Myrtle Green 23 (115 ft) - API: 35137208380000
- Moore Est 2-A (210 ft) - API: 35137212590000
- Hunsucker 11-1 (44 ft) - API: 35019219380000
- Frensley E 1-X (278 ft) - API: 35137223440000
- Russell B-1 (292 ft) - API: 35137210400000
- Winters-Hardin (104 ft divided in two cores) - API: 35019211340000

Additionally, core plugs were taken from the cores for petrographic analysis (99 samples) including thin sections and secondary electron microscopy (SEM); routine core analysis (92 samples) including Porosity and permeability data; mineralogical composition (48 samples) using X-ray diffraction spectrometer (XRD), elemental composition (99 samples) using a hand-held X-ray fluorescence spectrometer (XRF) and organic geochemistry data (36 samples). A location map for the three chapter will be added to show the data used for each projects/publication.

Contributions and limitations

This study aims to contribute to the understanding of diagenesis in mixed carbonate-siliciclastic mudrocks and to bridge the resolution gap between thin-section-derived information and 3D models. The knowledge facilitated in this dissertation impacts the scientific community petroleum geologist working on the industry, and geoscientist in general that work with datasets at different scales. On one hand, the first part of the dissertation contributes to understand the

diagenetic and depositional processes for the Sycamore Formation. This is specially important for this Early Mississippian strata because eolian processes likely transported the Sycamore siltstones (McGlannan et al., 2019). Therefore, this eolian siltstones are important to reconstruct the paleoenvironment during the Early Mississippian time. On the other hand, the second part has two immediate impacts on the petroleum geology field: i) it provides a model of the authigenic calcite spatial distribution, and this contributes to predicting the reservoir quality in the Sycamore, and ii) it proposes a workflow that bridges the resolution gap between scarce information at thin-section scale and abundant well logs. This is important to geoscientists working with datasets at different scaled, and it contributes to generating robust 3D models at a reservoir scale from information at thin section scale. This, in the long term, will open the door for modeling more complex diagenetic processes observed only at a thin-section scale that help us to understand the subsurface strata to reveal geological processes from the past.

The first part of the dissertation presents a conceptual model to show the spatial distribution of calcite-poor and -rich siltstones. However, the question about the origin of the authigenic calcite is still open. The common techniques available to measure the chemical and isotopic compositions of cement in sandstones are not enough to measure the small (60 μm) calcite crystals in mudrocks. Therefore, new techniques are required to understand the origin of this material.

The second portion implements a workflow to bridge the resolution gap between thin sections and 3D models. The original goal of this project was to build 3D models to visualize the distribution of authigenic calcite, but instead, we built one model to show the distribution of three thin-section-defined petrofacies. This petrofacies are rock types with information from the lithology, mineralogy, and porosity/permeability data. Therefore, the workflow was implemented

using a discrete variable (petrofacies). However, other continuous geological attributes, such as percentage of authigenic calcite, TOC, and porosity could be estimated implementing this workflow. Nevertheless, we expect that more data is needed to predict continuous attributes at high accuracy.

1.2 Literature review

Geological Settings

Ardmore basin is a northwest-southeast depression of Pennsylvanian age located in southern Oklahoma. The Anadarko basin to the North, the Arbuckle uplift to the east, and Marietta basin to the southeast bound the Ardmore basin. Three major tectonic events explain the evolution of the Ardmore basin in the context of the Great Anadarko Basin: rifting, passive margin, and plate collision.

The first event is related to a three-arm rift or triple junction during the late Proterozoic to early Paleozoic. Two of these arms are associated with the opening of the Protoatlantic Ocean, the other arm failed and formed the Southern Oklahoma Aulacogen (SOA) (Ham et al., 1965; Burke and Dewey, 1973; Hoffman et al., 1974; Wickham, 1978; Allen, 2000). The second event is associated with subsidence and passive margin of the aulacogen, this controlled the deposition of a thick sedimentary sequence during early Cambrian to early Mississippian time, along an asymmetric foreland and the Anadarko shelf(Lowe, 1975; Evans, 1979; Lane and De Keyser, 1980; Gutschick and Sandberg, 1983; Keller et al., 1983; Perry, 1989). Finally, the closure of the Rheic sea during Pennsylvanian time due to the North America and Gondwana collision, triggered the Ouachita Orogeny. This formed major uplifts and basins that resulted in its present configuration (Ham et al., 1965; Feinstein, 1981; Perry, 1989; Ghosh et al., 2018). (Perry, 1989).

In Mississippian time three major structures existed: Anadarko-Ouchita basin, the northern Oklahoma shelf and the Ozark uplift. But late Mississippian-Pennsylvanian tectonism formed the Nemaha ridge and fragmented the Anadarko-Ouachita basin (Ham et al., 1965; Feinstein, 1981). During this collision several orogenic events formed major uplifts such as the Nemaha Ridge, Wichita and Arbuckles mountains, and major basins such as Marrieta, Ardmore and Anadarko. Ouchita uplift during Mississippian to Pennsylvanian time produced a major unconformity (Miss/Penn unconformity) (Siever, 1951; Stevenson et al., 1951; Sutherland, 1988; Nana Yobo and Yobo, 2014).

During Cambrian and Silurian time in the SOA, a thick sequences of carbonate facies were accumulated in the Ardmore basin. Then, during late Devonian a major transgression resulted in the deposition of the organic-rich Woodford shale on top of a major erosional unconformity. Throughout Early Mississippian time (359-340 My.) great part of North America was covered by an extensive, shallow and tropical sea (Ham, 1978; Gutschick and Sandberg, 1983) (Figure 1). But in order to understand Early Mississippian strata is necessary to study the entire Mississippian units. The Carboniferous system is divided from older to younger into the Mississippian and Pennsylvanian period. This study focuses in the Mississippian subsystem divided into Tournaisian, Visean and Serpukhovian stages based on the international chronostratigraphic chart (Cohen et al., 2019). But, in the North American system the Mississippian is divided into the Kinderhookian, Osagean, Meramecian, and Chesterian stages from older to younger. (Figure 2).

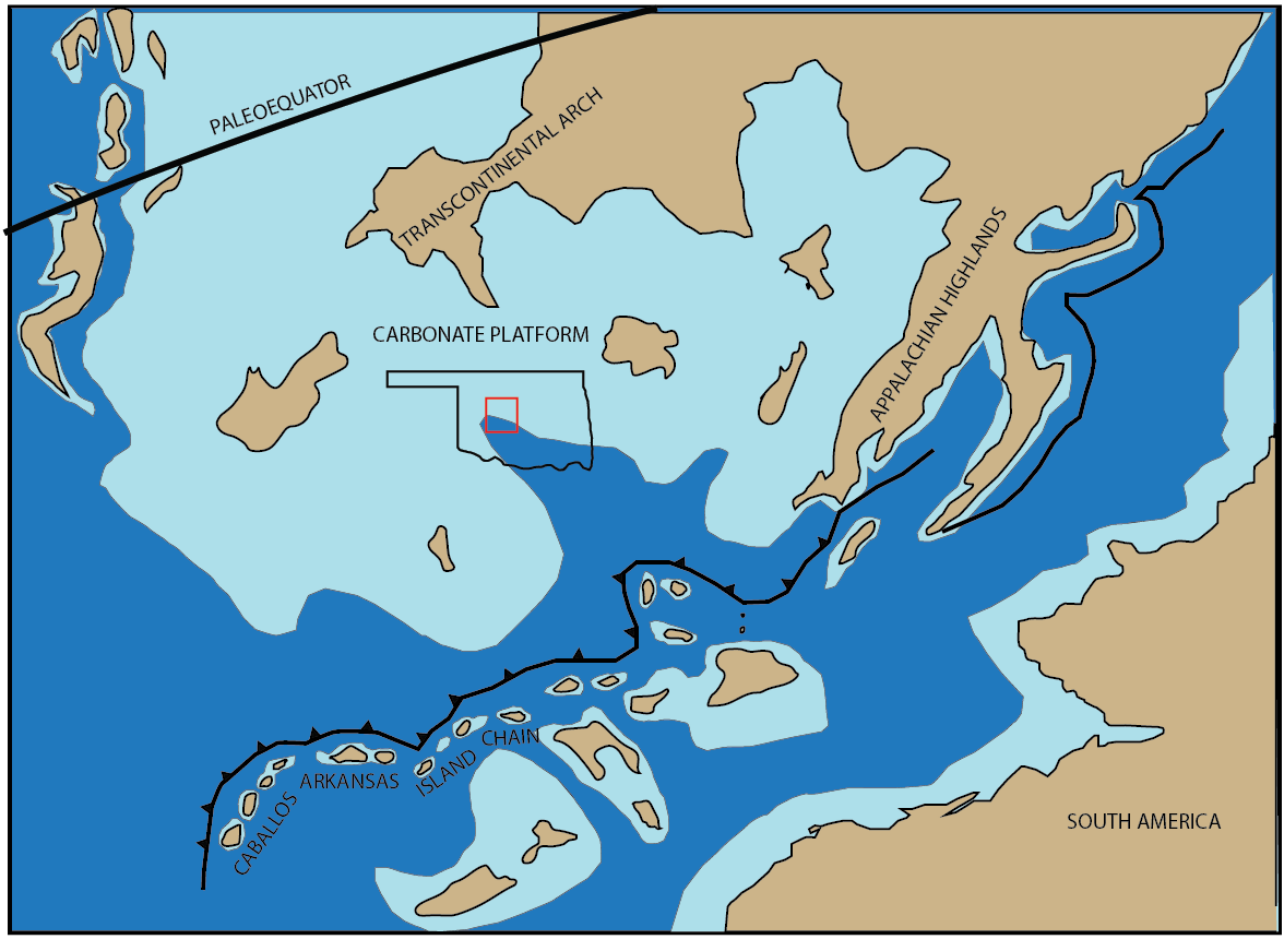

Figure 1: Paleography of the Early Mississippian strata. Taken from (Duarte, 2018) after (Gutschick and Sandberg, 1983; Blakey, 2013). Red square represents Anadarko basin.

During Kinderhookian time, shallow and protected seas were close to the Caballos-Arkansas Islands chain. Welden limestone (or "Pre-Sycamore") in the Arbuckles and Ardmore basin, and Kinderhook in Anadarko basin are assigned to Kinderhookian age (Curtis and Champlin, 1959). Some authors have placed the top of the Woodford in Kinderhookian time overlying immediately the Sycamore limestone (Braun, 1961; Bennison, 1956). Also, a lower green shale is reported in south Oklahoma which marks the base of the Welden formation (Huffman and Barker, 1950; Braun, 1959; Culp, 1961).

Osagean rocks are truncated into the south by Meramecian units, suggesting an uplift related to the Ouchita orogeny. Welden limestone in the Lawrence uplift and Osage in the northern part of the Anadarko basin are interpreted in part as Osagean in age (Curtis and Champlin, 1959; Bennison, 1956). In the southern portion of the Anadarko basin the thick Sycamore limestone is divided into an upper limestone and a lower cherty limestone with shale. The lower part is placed in the Osagean series and correlated with the upper portion of the reported Pre-Sycamore (Harlton, 1956).

Then, during late Mississippian time The Nemaha ridge experienced an uplift in related to the Ouachita orogeny, consequently Meramecian units thin toward the north. General marine transgression and subsidence in late Mississippian time resulted in the deposition of the Sycamore and Meramec formations in the northern and southern portions of the Great Anadarko Basin, respectively (Curtis and Champlin, 1959; Peace, 1994). Braun (1961) restricted the name Sycamore in the Ardmore and south Anadarko basin to the upper massive and competent silty limestone of the Sycamore and proposed Meramecian age deposition (Braun, 1961; Culp, 1961; Bennison, 1956; Harlton, 1956). Finally, the thickness of Chesterian rocks (Caney, Goddard and Chester Formations) decreases towards the Nemaha ridge (Curtis and Champlin, 1959) product of the Mississippian-Pennsylvanian tectonism.

Physicochemical conditions varied in the Mississippian system from the shelf in the north and West to basin in the south, resulting in a variety of lithofacies that are hard to correlate without a time framework. Also, during this time the tectonic activity increased resulting in three possible sources of material: Appalachian highlands in the North-east; transcontinental arch in the North-west; and the Caballos-Arkansas Island chain in the south. These two factors helped the facies to

vary laterally and thus the nomenclature of Mississippian units along Oklahoma is more complex than it should be.

This study focusses on the Sycamore formation, composed mainly of clay-rich and silt-rich gravity flow deposits. Sycamore in the Ardmore basin is composed of up to 150 m of clay- and silt-rich mudstones with rear fine sandstones. Late Mississippian to Early Pennsylvanian tectonism resulted in the exposure of Mississippian strata to the east in the Arbuckle mountains, but Mississippian strata also dips up to 6,000 m depth to the west in the subsurface.

Global Series/Stage			North American Series/Stages			Curtis and Champlin, 1959		Peace, 1994			Culp, 1961	Bennison, 1956	Braun, 1961
						Lawrence Uplift	Anadarko basin	Sooner Trend Area	T 8N - 13N	T 6N - 7N	Anadarko Basin	Anadarko	Anadarko Basin
Carboniferous	Middle	Moscovian	Carboniferous	Pennsylvanian	Morrowan			Morrow	Morrow	Morrow			
		Bashkirian								Springer			
	Lower	Serpukhovian		Mississippian	Chesterian	Canye	Chester	Chester	Chester/ Springer		Caney	Caney/Chester	Caney
		Visean			Meramecian		Meramec	Osage-Meramec	Caney	Caney	Sycamore	Sycamore	Sycamore
					Osagean	hiatus	Osage		Sycamore	Sycamore	Osage or "Pre-Sycamore"	?	No deposition
		Tournaisian										Welden	
					Kinderhookian	Welden	Kinderhookian				Welden		Welden
Devonain	Upper	famenian	Devonian		Chattanoogan	Woodford	Woodford	Woodford	Woodford	Woodford	Woodford	Woodford	Woodford
		frasnian											

Figure 2: Correlation of Mississippian units from different authors. (Bennison, 1956; Braun, 1959; Chenoweth et al., 1959; Curtis and Champlin, 1959; Culp, 1961; Peace, 1994)

Diagenesis of Mudrocks

Mudrocks are divided in claystones and siltstones depending on the most abundant grain size. In the following pages we used the term *siltstones* to the mudrocks that are formed mainly by silty size material, and *clay-rich mudstones* for the remaining mudrocks. In the petroleum geology, other common division for the mudrocks is based on the content of organic matter. Organic

mudstones or organic shales are used for the mudrocks with total organic content (TOC) above 1.5 %. These organic shales have been studies for their importance as self-source reservoirs or unconventional reservoir (Loucks and Ruppel, 2007; Slatt and O'Brien, 2011; Turner et al., 2016; Alnahwi and Loucks, 2019; Slatt et al., 2021). Several studies have been completed on organic shales to understand diagenetic processes that affects brittleness, porosity, and permeability on the organic-rich shales (Nadeau et al., 2005; Dennie et al., 2012; Jennings and Antia, 2013; Elmore et al., 2016; Milliken and Olson, 2017; Roberts and Elmore, 2018; Wickard et al., 2020). However, in their counterpart Siltstones, the lack of studies difficult their understanding with consequences on the petroleum geology, but also as a paleoenvironmental (Gerilyn S. Soreghan, 1992; Kessler et al., 2001; Soreghan et al., 2007, 2008, 2015; Giles et al., 2013). Paleographic reconstructions, grain size statistics and geochemistry from the Mississippian strata suggest eolian processes take part in transporting the silty sediments prior to deposition in a subaqueous environment.(McGlannan et al., 2019). Therefore, studying the siltstones from the Sycamore Formation has implications for the scientific community and the petroleum industry.

Diagenetic processes, such as cementation and dissolution, affect the reservoir quality of unconventional reservoirs in terms of porosity, permeability, and brittleness. This is especially important in unconventional reservoirs because how the rock response when is subjected to stress impacts the well plan (Wang and Gale, 2009; Slatt et al., 2018). The general idea is that rock with higher content of calcite, dolomite and quartz are more brittles, and rock with higher clay content are less brittle. Brittleness is the response of the rock when exposed to different loading conditions such as tension or compression (Zhang et al., 2014; Xia et al., 2019; Meng et al., 2021). Other important criterion to define the reservoir quality in mudstones is the capacity of the rock to store a fluid, we called this porosity. Porosity in mudrocks is variable, for example in organic mudstones,

organic porosity product of the maturity of organic matter is abundant (Slatt and O'Brien, 2011; Slatt et al., 2012, 2021; Mathia et al., 2016). However, siltstones are usually tight reservoir with different types of porous than the organic-rich shales. Finally, the other important criterion is permeability of mudstones that is highly controlled by the clay content, and cementation (Katsube and Williamson, 1994; Eichhubl and Aydin, 2005; Heij and Elmore, 2019; Wang et al., 2019; Duarte et al., 2021). Understanding the diagenetic processes that affect mudstones have implications in reconstructing the geological evolution of an area but also in the petroleum geology.

Because the lack of diagenetic studies on siltstones, the diagenetic processes are not well understood and this difficult the prediction of subsurface geology. Some of the most common diagenetic products linked with the Sycamore Formation are calcite cement, feldspar grains dissolution and authigenic clays.

Calcite cement

The calcite ($CaCO_3$) behavior is controlled by the following reversible reaction, where calcite is dissolved in the forward direction, and it is precipitated in the reverse direction:

$$CaCO_3\ (\text{solid}) + H_2CO_3 \rightarrow Ca^{+2} + 2HCO^{-3} \quad (1)$$

$$H_2O + CO_2 \rightarrow H_2CO_3 \quad (2)$$

$CaCO_3$ is the crystalline calcite; H_2CO_3 is the carbonic acid formed by the reaction between water (H_2O) and carbon dioxide (CO_2). Ca^{+2} is the calcium cation in solution and HCO^{-}_3 is the bicarbonate ion (anion). Changes in temperature, pressure, organic activity, or acidity are some of the controlling factors in the precipitation or dissolution of the calcium carbonates. However, CO_2 is the determining factor that promotes or inhibits the production of carbonic acid, H_2CO_3 in equation 1, shifting a system in equilibrium to the right or left respectively in equation 2.

The intergranular calcite cement in the Sycamore Formation was generated during early diagenesis supported by the absence of quartz cement, chlorite rims, and the lack of mechanical compaction evidenced by the floating texture in the massive siltstones. However, the origin of this cement is not clear and difficult to explain due to the deficiency of methodologies for separating calcareous cement from calcareous pellets, and the small size of the calcite crystals to measure the isotopic composition in situ. SEM observations indicate that the crystalline cement is replacing micritic matrix. Hence, we propose an internal source in the same formation as the source for that material needed to precipitate the crystalline cement (Kantorowicz et al., 1987; Saigal and Bjørlykke, 1987).

Internal sources, such as interbedded mudstones, dissolution of calcium-bearing feldspars, fragment fossils, and calcareous pellets may be responsible for the precipitation of crystalline calcite cement (Morad et al., 1990). Some authors (Freed and Peacor, 1989; Sun et al., 2019) have reported that mudstones and feldspars could provide the needed ions ($Na^{+} + Ca^{2} + Fe^{2} + Mg^{2}$) to precipitate carbonate cements. Nevertheless, this does not explain the great amount of calcite that have been precipitated in the massive siltstones. Therefore, it is more likely that micrite and carbonate pellets could account for the calcium needed for growing crystalline calcite. This idea is discussed in the chapter 2.

Feldspar grains dissolution

Silts-sized quartz and feldspar grains are abundant in the massive siltstones. Feldspars dissolution is the main mechanism that drives secondary porosity and it is usually attributed to the presence of organic acid fluids. In the Sycamore formation, these fluids could be sourced by clay-rich mudstones intervals within the Sycamore or most like originated from the underlying Woodford Shale.

The Woodford Shale is an organic rich shale, with TOC values up to 17% (Galvis et al., 2018), that have been considered as the sourced rock for excellence in Oklahoma, and it has generated hydrocarbons since Mississippian time (Comer, 1992; Cardott, 2012). CO_2 derived from this organic-rich shale during decarboxylation of organic matter dissolved in water (equation 2) may be responsible for the carbonic acid that contributes in the dissolution of feldspars in the Sycamore (equation 3) (Worden and Barclay, 2000; Rahman and Worden, 2016). Additionally, the dissolution of feldspars grains contributes to the formation of the authigenic clays in the sycamore formation by releasing K^+, Al^{3+}, and $SiO_2(aq)$ into the system (Sun et al., 2019).

Quartz cement is very rare in the Sycamore formation, and considering that mechanical compaction was minor, silica that resulted in quartz cement is likely to be sourced from dissolution of feldspars. The small presence of quartz cement could be explained by two factors: first, the lack of time for the grow of quartz, this is due that quartz cement develops at temperatures higher than 80-90 °C (Walderhaug and Bjørkum, 1992; Rahman and Worden, 2016; Li et al., 2017). Or the presence of clay minerals that prevents the quartz overgrowth (Ehrenberg, 1989; Rahman and Worden, 2016; Li et al., 2017). Additionally, some quartz grains exhibit corrosion textures associated with the calcite cement (Shaw and Primmer, 1991; Xiong et al., 2016).

$$3KAlSi_3O_8\,(K-\text{feldspar}) + 2H^+ \rightarrow KAl_3Si_3O_{10}(OH)_2\,(\text{Illite}) + 6SiO_2 + 2K^+ \qquad (3)$$

$$CaCO_3 + H^+ \rightarrow Ca^{+2} + HCO_3 \qquad (4)$$

$$H_2CO_3 \rightarrow HCO_3 + H^+ \qquad (5)$$

$$\text{Smectite} + K^+ \rightarrow \text{Illite} + SiO_2 + Na^+ + Ca^2 + Fe^2 + Mg^2 + H_2O \qquad (6)$$

Authigenic clays

Feldspar dissolution, responsible in great part for the secondary porosity, occurs under acidic conditions and contributes to form authigenic phyllosilicates (Bjorlykke and Aagaard, 1992; Worden and Morad, 2003; Rahman and Worden, 2016). Decarboxylation of organic matter in the Woodford releases CO_2 in which dissolves in water to produce carbonic acid with pH less than 7. The carbonic acid produces bicarbonate ions and hydrogen ions that are used in the K-feldspars and calcium carbonate reactions. When carbonic acid forms hydrogen ions and bicarbonate ions, the hydrogen ions are used for the dissolution of feldspars to generate clays such as illite. The remaining bicarbonate ions are used for the calcium carbonate precipitation.

Carbonic acid could be replaced by hydrogen and bicarbonate ions in the calcium carbonate reaction (equation 4), resulted in the simplified equation 5. Moreover, the reaction responsible for the feldspar dissolution take two hydrogen ions to form illite and release potassium ions and silica (aqueous) in the systems. Thus, reaction 5 provides the hydrogen ions needed in the dissolution of

feldspars but also provided the bicarbonate ions needed for the precipitation of calcium carbonate. Additionally, smectite-to-illite transformation during late diagenesis is driven by an increase on temperatures (> 100°C). This transformation releases ions (Na^{+} + Ca^{2} + Fe^{2} + Mg^{2}) used for the precipitation of calcium carbonate cements such as calcite and ferroan calcite, and dolomite (Sun et al., 2019).

Therefore, dissolution of feldspars, phyllosilicates growth, and calcite precipitation are paragentically related in some extend (Rahman and Worden, 2016). These diagenetic processes have been extensively studied on sandstones(Bjørkum, Per Arne et al., 1998; Worden et al., 2018) (Rahman and Worden, 2016; Li et al., 2017; Sun et al., 2019); however, this is not the case for siltstones (Vaisblat et al., 2021) specially for mixed carbonate siliciclastic system such as the Sycamore Formation. In the Sycamore Formation the small grains size and the two sources, carbonate factory and a siliciclastic feeder, are limitations that difficult the understanding of diagenetic processes, but we tried to understand the paragenetic sequence and propose the genesis of the diagenetic products. This study has implications in understanding the subsurface geology specially for petroleum geology.

Machine Learning techniques in geosciences

Rock cores provide unique information about the depositional settings of the sedimentary layers that underlie petroleum operations. Despite their importance, cores are scarce, likely because of the elevated extraction costs. Alternative approaches with limited resolution, such as well logs, are more commonly used to characterize petroleum reservoirs. The need of resolving subtle changes in the rock properties calls for the development of novel techniques that allow to predict core-

contained information from the already abundant well logs, while narrowing the resolution gap in the existing characterization techniques.

Geosciences has witnessed a flourishment in the application of machine learning (ML) techniques to address many of the challenges imposed by the lack of core data. The two main types of ML techniques are supervised and unsupervised methods. Supervised ML techniques have been used in geoscience to classify core-based geological attributes (e.g. lithofacies) from well log information (Dubois et al., 2007; G. Wang et al., 2014; Bhattacharya et al., 2016; Wood, 2019; Bressan et al., 2020). However, it has been clear that for the supervised model to perform well, big data sets containing numerous observations from both predictor and predicted variables are required, posing a stringent need for data that is not always feasible to fulfill. On the other hand, unsupervised techniques have been used in geosciences to define facies (e.g., chemofacies or Electrofacies) based on indirect tools such as well logs or hand-held X-Ray fluorescence spectrometer (XRF) (Gupta et al., 2017; Duarte, 2018; Duarte et al., 2019; Hardisty et al., 2021; Slatt et al., 2021).

The most common unsupervised methods are clustering analysis. Clustering is used when trying to understand the structure of the data, for this, groups or clusters are defined based on similarity of the data. The advantage of clustering is that labels or dependent variables are not needed so it has been used to define Electrofacies (Gupta et al., 2017; Davis, 2018; Hossain et al., 2020) from well log, or chemofacies from elemental composition. However, the lack of dependent variable implies that the validation process is ambiguous since there is not data to compare with the results. Therefore, supervised ML techniques are more popular in geosciences to predict some geological attributes.

For supervised ML techniques, a model is trained using labeled data (data with independent and dependent variables) to predict some geological attributes. For example, build a model to predict lithologies from well logs. Depending on the nature of the output data we use regression or classification models when using continuous or discrete data, respectively. For this dissertation, even though the goal is to understand the calcite and porosity distribution, both continuous variables, we combined this information in discrete data called petroafacies. One of the most important information derived from core data is the characterization of facies with similar mineralogical composition, pore types and distribution (i.e., Petrofacies). Determining petrofacies allows for the identification of prolific petroleum reservoirs (Watney et al., 1998) and the characterization of the diagenetic processes that help to predict petrophysical properties in the subsurface.

1.3 Dissertation outline

This dissertation is the result of the work I developed during my Ph.D. at the University of Oklahoma. This dissertation is organized into five chapters related to each other by the Sycamore Formation, but the body of the dissertation is from chapters 2 to 4. All the data for this dissertation came from six cores, one outcrop, and around 300 adjacent wells. All the chapters are the interest of petroleum geologists, however, chapters 3 and 4 use machine learning techniques that could be the interest of other fields. During my Ph.D., I had the privilege to work with several co-authors and they are recognized at the beginning of each chapter. Therefore, the chapters are written in first person plural because these chapters are organized based on journal papers published or submitted. However, I was responsible for the conception, design, analysis, interpretation of the data, figure creation, and the great majority of writing for these chapters.

Chapter 1 (this chapter) serves as an introduction to the dissertation, with the motivation, objectives, and implications of the project in general. Also, this chapter includes a literature review of the main topics. Then, Chapter 2, "Diagenetic Controls on Reservoir Quality of a Mixed Carbonate-Siliciclastic System: Sycamore Formation, Sho-Vel-Tum Field, Oklahoma, USA", is presented as it was published in Marine and Petroleum Geology (Duarte et al., 2021). This chapter presents a conceptual model to explain the effect of diagenesis on the reservoir quality in terms of porosity and permeability.

The second portion of the dissertation is divided into chapters 3 and 4. In this chapter, we used machine learning techniques that use XRF data and well logs that are incorporated in 3D models to validate the conceptual model from chapter 2nd. The objective for the 3D model is to give a visual distribution of the calcite-rich and -poor intervals in the Sycamore Formation. Chapter 3, "Semi-supervised Workflow to Generate Petrofacies Logs from Thin Sections and XRF Data", is presented as it was submitted to Marine and Petroleum Geology. This chapter aims to present a workflow to bridge the vertical resolution gap between core data and well logs. Chapter 4, "3D Petrofacies model assisted by supervised machine learning techniques to increase well control", is in preparation to be submitted in Marine and Petroleum Geology journal. The goal of this chapter is to build a workflow for increasing well control to build robust 3D models. In the end, the final remarks and future work are presented in chapter 5.

Chapter 2: Diagenetic Controls on Reservoir Quality of a Mixed Carbonate-Siliciclastic System: Sycamore Formation, Sho-Vel-Tum Field, Oklahoma, USA.

David Duarte, Benmadi Milad, R. Douglas Elmore, Matthew J. Pranter, and Roger Slatt

2.1 Preface

This chapter is presented here as it was published in the Marine and Petroleum Geology journal. A portion of the data for this project was provided by Energy 89 company through the STACK-Merge-SCOOP consortium led by Dr. Roger Slatt, the other portion was provided by the consortium. The results presented in this chapter show that reservoir quality in the Sycamore Formation is controlled by the diagenesis that is driven at the same time by the depositional facies. Some of the highlights for this chapter are:

- Presentation of the diagenetic evolution for the Sycamore Formation in the Ardmore Basin.
- Cementation and dissolution are the main diagenetic processes that impact reservoir quality of the massive siltstones.
- Sediment supply between carbonate and siliciclastic systems drives diagenetic processes that affect reservoir quality.
- Stratigraphic variation of cement between parasequences is linked with conditions during deposition and diagenesis.

2.2 Abstract

Reservoir quality (RQ), commonly related to diagenesis, is important to predict formation porosity and permeability. The lack of studies to understand the role of diagenesis on tight reservoirs is a limiting factor for predicting petroleum accumulation in unconventional reservoirs. This study

aims to determine the influence of diagenesis on RQ for the mixed carbonate-siliciclastic Sycamore Formation in southern Oklahoma. Additionally, this study assesses other geologic factors such as depositional facies control on diagenesis of the Sycamore.

Petrographic observation, geochemical data, and conventional plug analysis from six cores and one outcrop were used to accomplish three objectives: i) reconstruct the paragenetic sequence; ii) determine the role of diagenesis on RQ; and iii) elucidate geological factors that control the diagenetic evolution in the Sycamore strata.

Four facies were identified: Massive Calcite-cemented Siltstones, Massive Calcareous Siltstone, Bioturbated Mudstones, and Argillaceous Mudstones. However, this study focuses on the siltstones and their associated diagenetic processes and products. Porosity evolution in the siltstones is affected by cementation and dissolution processes. These diagenetic processes are related to each other; calcite cement decreases permeability during early diagenesis and controls the flow of later diagenetic fluids responsible for feldspar dissolution. Additionally, differences in the supply between a carbonate and siliciclastic material and a change from a carbonate-dominated system to a mixed system through time controls the distribution of calcite cement in the Sycamore.

This study allows geologists to understand the diagenetic evolution and to characterize RQ of the Sycamore and similar formations by considering the sediment supply between carbonate- and siliciclastic-dominated systems, changes in stratigraphy, and influx of acidic fluids. Identifying geological controls on diagenesis and defining the influence of diagenesis on RQ in tight formations is important for the exploration and development of unconventional reservoirs, to build predictable porosity and permeability models, and to identify petroleum accumulations.

2.3 Introduction

This study focuses on how diagenesis affected reservoir quality of the tight Mississippian Sycamore reservoir at Sho-Vel-Tum field (abbreviation of three oil fields combined, Sholem-Alechem, Velma, and Tatums) in the Ardmore Basin (**Figure *3***) and how depositional processes and associated facies impact diagenesis.

Reservoir quality in carbonate and siliciclastic rocks are affected by depositional processes, sea-level changes, diagenetic processes, tectonics, and input material either from a carbonate factory or a siliciclastic feeder (Worden et al., 2018). Diagenesis in coarse-grained siliciclastic systems has been extensively studied (McBride, 1984; Mansurbeg, De Ros, et al., 2012; Rahman and Worden, 2016; Xiong et al., 2016; Li et al., 2017; Yang et al., 2017; Sun et al., 2019; Okunuwadje et al., 2020). Original reservoir quality and consequent diagenesis is determined by depositional setting including the composition of the grains, grain size, and sorting (Walderhaug and Bjørkum, 1998; Worden et al., 2018). In the siliciclastic reservoirs, compaction, cementation and dissolution are the main processes controlling reservoir quality (Rahman and Worden, 2016; Li et al., 2017; Sun et al., 2019). However, reservoir quality of carbonate rocks is controlled by other factors including the carbonate factory, sea-level changes, paleoclimate, clastic input, and the composition of the seas (Worden et al., 2018). The aforementioned diagenetic factors that control reservoir quality in a mixed carbonate-siliciclastic system are present in the Sycamore Formation at Sho-Vel-Tum field.

This study focuses on the diagenesis of the Early Mississippian Sycamore Formation with the emphasis on how diagenetic alterations affected reservoir quality. The objectives of this study are i) to reconstruct the paragenetic sequence of the Sycamore Formation in the Sho-Vel-Tum field, ii) to determine the role of diagenesis on the reservoir quality, and iii) to elucidate the

geological factors that control the diagenetic evolution in the massive siltstones within the Sycamore Formation.

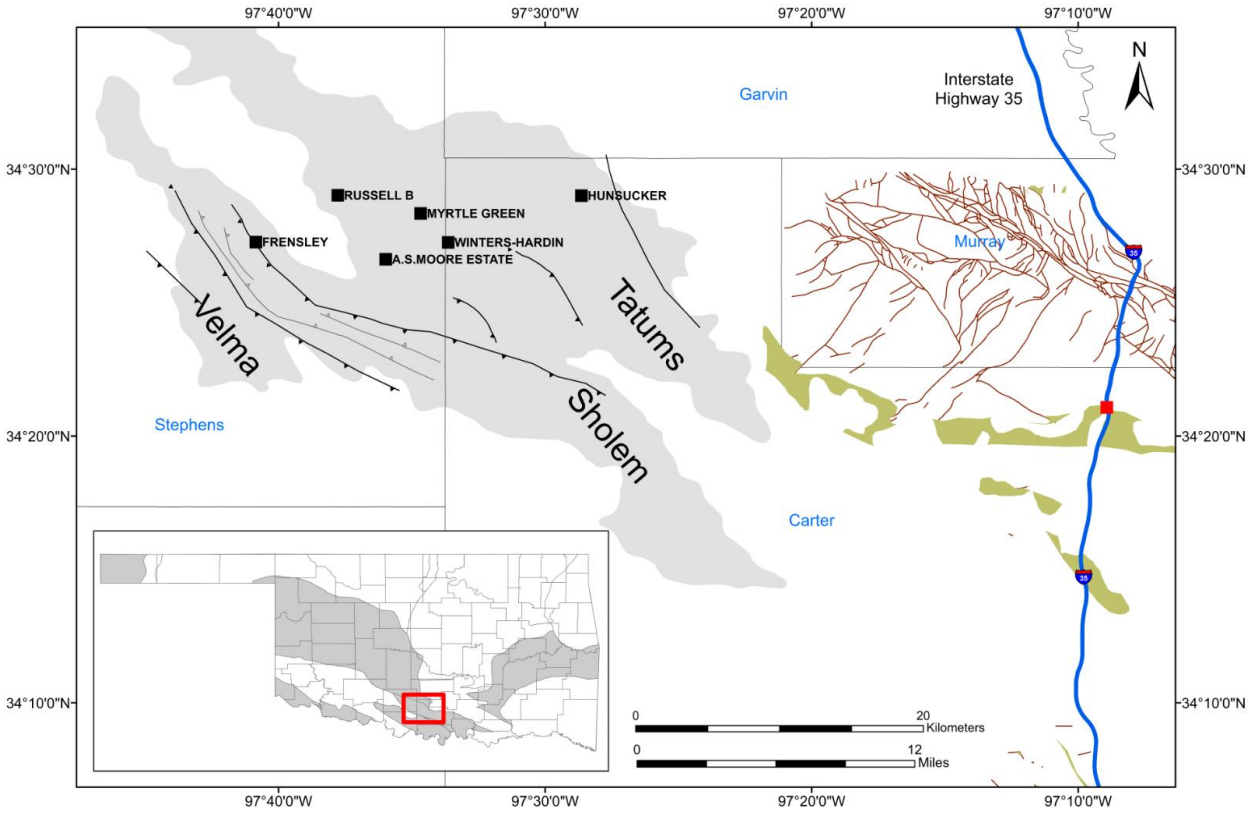

Figure 3: Location map of study area. Black squares are cored well locations in Sho-Vel-Tum field (gray area). Red square is the roadcut with Mississippian strata along the southern limb of the Arbuckle Mountains. Dark red lines represent the faults in Arbuckles and the green areas represent Mississippian rocks from the USGS geological map (Cederstrand, 1997; Boyd, 2002; Heran et al., 2003). Faults in black in Sho-Vel-Tum field are from (Carpenter and Tapp, 2014).

2.4 Geological setting

The Ardmore basin is a northwest-southeast-oriented basin located in Oklahoma, USA. It is bounded by the Anadarko basin to the north, the Arbuckle uplift to the east and north, and the Marietta basin to the southeast. Three major tectonic events explain the evolution of the Ardmore basin. First, during the late Proterozoic to early Paleozoic a three-arm rift resulted in the opening

of the Protoatlantic Ocean from two arms. The other arm failed and formed the Southern Oklahoma Aulacogen (SOA) (Ham et al., 1965; Burke and Dewey, 1973; Hoffman et al., 1974; Wickham, 1978; Allen, 2000). Then, subsidence and a passive margin setting in the aulacogen controlled the deposition of a thick sedimentary sequence (Lowe, 1975; Evans, 1979; Lane and De Keyser, 1980; Gutschick and Sandberg, 1983; Keller et al., 1983; Perry, 1989). Finally, during Pennsylvanian time, North America and Gondwana collided and triggered the Ouachita Orogeny. This formed major uplifts and basins that resulted in its present configuration (Feinstein, 1981; Ghosh et al., 2018).

During Cambrian and Silurian time in the SOA, a thick sequence of carbonate and clastic facies accumulated in the Ardmore basin. Then, during the late Devonian, a major transgression resulted in the deposition of the organic-rich Woodford Shale on top of a major erosional unconformity. Throughout Mississippian time, an extensive and tropical sea covered a great part of North America (Curtis and Champlin, 1959; Ham, 1978; Gutschick and Sandberg, 1983; Blakey, 2013). During this marine transgressive episode, limestones, clay- and silt-rich mudstones, such as the Sycamore Formation, were deposited. Through the Wichita Orogeny in late Mississippian-early Pennsylvanian time, erosion of Paleozoic sediments resulted in the deposition of marine shales, sandstones and conglomerates in the Ardmore and Marietta basins (Allen, 2000).

The Sycamore Formation was initially described by Taff (1903) as a limestone. Since then, multiple studies have been completed in outcrops (Cooper, 1926; Chenoweth et al., 1959; Milad and Slatt, 2018; Milad et al., 2020) and in the subsurface (Prestridge, 1957; Braun, 1959; Culp, 1961; Cole, 1989; Coffey, 2001) and defined the Sycamore Formation as a mixed carbonate-siliciclastic system. However, to the best of the authors' knowledge, only Coffey (2001) has addressed the Sycamore porosity and its relationship to diagenetic processes. Coffey (2001)

indicated that secondary porosity (fracture porosity) of the Sycamore in the Carter-Knox field, towards the north of the Sho-Vel-Tum field, was structurally developed due to uplift where higher fracture porosity occurred in structurally high areas.

The Sycamore Formation is bounded by two important source rocks, the Woodford and Caney shales (**Figure *4***), and it has two informal zones, a lower transition zone and an upper zone with intervals of massive siltstones separated by clay-rich mudstones (Fay, 1989; Donovan, 2001). The lower transition zone is separated from the underlying Woodford Shale by a glauconitic layer (Donovan, 2001; Duarte, 2018), and it is dominated by gray-greenish claystone (Galvis et al., 2018; Milad et al., 2020). The upper zone is dominated by gravity-flow deposits separated by shale-dominated intervals. Biostratigraphic studies suggest the Sycamore Formation was deposited during Meramecian time (Schwartzapfel, 1990; Schwartzapfel and Holdsworth, 1996) and some authors interpreted that it was formed by material derived from carbonate and siliciclastic systems in the north and west, and deposited as gravity flows (Schwartzapfel, 1990; Coffey, 2001; Donovan, 2001; Milad et al., 2020). Donovan (2001) mentioned a source for the carbonate portion to the west and north from a distal shelf and a fluvial system developed to the north and/or east.

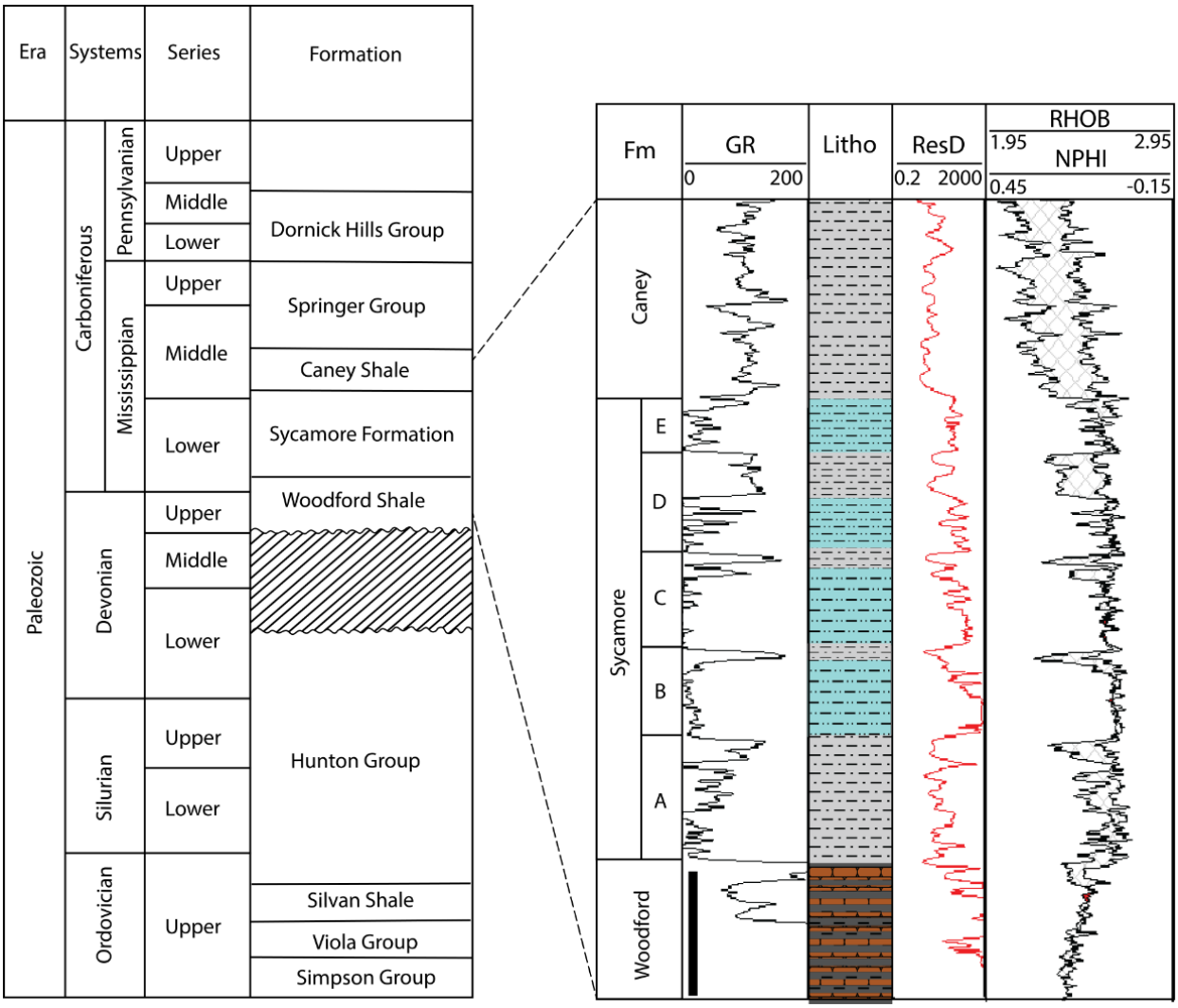

Figure 4: General stratigraphy of the study area. The Sycamore Formation is divided into 5 paraseqeunces (A-E). Parasequence A is a transition zone from the Woodford Shale to the massive siltstones in the Sycamore Formation. Parasequences B through E are massive siltstones (blue) divided by clay-rich mudstones (gray). Modified from Bebout et al. (1993) and Henry and Hester (1995).

2.5 Data and methods

Rock description and petrography

A total of 287 m (942 ft) of Sycamore core from six wells (Frensley, Russell B, Moore Est, Myrtle Green, Hunsucker, and Winters-Hardin) were described in this study (**Figure *3***) in terms of

lithology, grain size, bioturbation index, key surfaces, and fractures. Also, 137 m (450 ft) of Sycamore outcrop on interstate highway 35 (I-35)(Milad et al., 2020) was also used. Ninety-nine plugs from different cores were collected for: thin sections, X-ray diffraction (XRD), energy dispersive x-ray fluorescence (Ed-XRF), hand-held X-ray fluorescence (Hh-XRF), porosity, and permeability analyses. All thin sections were impregnated with blue epoxy and stained with Alizarin-Red S for porosity and calcite identification, respectively. A Zeiss AxioImager Z1TM petrographic microscope was employed to characterize the detrital siliciclastic grains, allochems, clays, and diagenetic phases. Later, this information was used to define the paragenetic sequence based on cross-cutting relationships. Point-counting (Dickinson, 1970) based on 300 points per thin section was used to quantify the amount of intergranular calcite and allochems in the bulk carbonate portion.

Scanning electron microscopy (SEM)

Ten selected thin sections were coated with gold prior to the analysis under the SEM. A Zeiss NEON 40 EsB SEM was used to study the diagenetic minerals with an emphasis on the intergranular calcite and types of porosity. The SEM was coupled with an energy dispersive spectroscopy analyzer (EDS), INCA Energy 250 Energy, to acquire observations for elemental composition in selected samples. Intergranular calcareous material could be either micritic matrix or crystalline since sometimes the limit for classification becomes diffuse when point counting. Thus, SEM was used to distinguish the calcite types in the massive siltstones.

X-ray diffraction spectroscopy (XRD)

Forty-eight samples were selected for bulk XRD mineralogical composition performed at Weatherford laboratories. Powdered samples were scanned with a Bruker AXS D4 Endeavor X-ray diffractometer using copper Kalpha radiation. The scanning parameters for a bulk scan are from 5° 2θ to 70° 2θ. The step size is 0.020°, and the dwell time at each step is 0.5 seconds. Mineral phases were identified with MDI JadeTM 9+ software and ICDD PDF 4+ 2018 database.

Porosity and permeability

Ninety-two samples were selected for conventional plug analysis where porosity, air permeability, and Klinkenberg permeability were measured at ambient temperature and 800 psi net confining stress. The samples were dried at 140 °F (60 °C) prior measurements.

X-ray fluorescence spectroscopy (XRF)

X-ray fluorescence spectroscopy from a Hh-XRF is a semiquantitative elemental analysis technique based on the principle that each element has a characteristic combination of fluorescent (or secondary) X-rays related to the energy lost. When an atom is excited by a primary X-ray source the electrons are ejected, this leaves the atom unstable. Then, electrons from higher shells drop to a shell closer to the nucleus; this release of energy produces the secondary X-ray. The energy of this X-ray is used to identify each element and consequently determine the bulk elemental concentration of the sample.

Bulk elemental composition was measured at one-foot resolution in the six cores, including the exact location of the 99 thin sections, to identify the described facies and to highlight zones with similar lithology but different elemental composition. We used a portable Bruker Traces IV-

SD hand-held X-ray fluorescence (Hh-XRF) at the University of Oklahoma to measure the elemental composition of the core plugs and the aforementioned cores at 0.3-m (1-ft) resolution. The major elements were scanned under vacuum at 15 kV accelerating voltage for 90 seconds, then, trace elements were analyzed under atmospheric pressure with a Ti-Al filter at 40 kV accelerating voltage for 60 seconds. The resulted data were converted to parts per million (ppm) using fundamental parameters (FP) for mudrocks in Rowe et al., 2012a.

2.6 Results

Facies classification

Mississippian strata are divided into two main lithologies (silt-rich mudstones and clay-rich mudstones) that are subdivided into four facies (**Figure *5***) based on grain size, mineralogical composition, and rock fabric. Siltstones have phyllosilicates content lower than 10% and mudstones have phyllosilicates content equal to or higher than 10% (**Figure *6***). The siltstones are structureless (or massive) with a low amount of clay, and the clay-rich mudstones are either massive or bioturbated and composed of clay- to silt-size grains. The massive siltstones are bounded at the base by a sharp contact and at the top by a transitional contact to clay-rich mudstones. Additionally, each of these lithologies are subdivided into two facies. The massive siltstones are divided into a) massive calcite-cemented siltstones (MCcSt) with intergranular calcite of > 40% and b) massive calcareous siltstones (MCSt) with intergranular calcite of 40% or less. The amount of detrital siliciclastic grains and clays in the MCSt are usually higher than the MCcSt. However, the clay-rich mudstones are divided into a) argillaceous mudstones (AMdst) with bioturbation index (BI) < 3 and b) bioturbated mudstones (BMdst) with BI equal to or > 4 following the Taylor and Goldring (1993) classification.

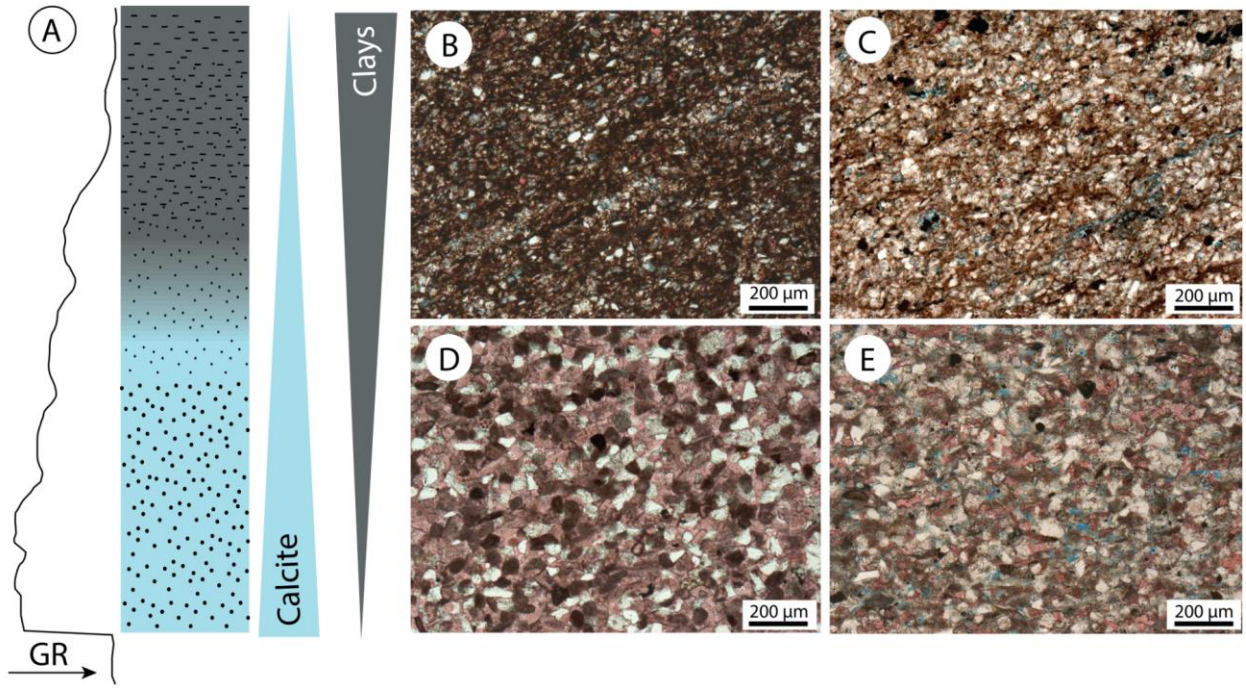

Figure 5: Facies in the Sycamore Formation. A) ideal facies succession for Sycamore deposits. Clay-rich mudstones are characterized by high GR response and low calcite content and include B) argillaceous mudstones and C) bioturbated mudstones. Silt-rich facies include D) massive calcite-cemented siltstones and E) calcareous siltstones.

The calcite content is variable in the Sycamore facies. The massive siltstones have higher calcite content than the mudstones; however, there are other differences between the massive siltstones and clay-rich mudstones. Importantly, the calcite component includes calcareous pellets, micrite, and crystalline calcite cement. **Figure *6***A shows a linear inverse relationship between the tectosilicates and the carbonates in the massive siltstones. Also, XRD data reveal that the massive siltstones can be differentiated in MCcSt and MCSt by the calcite content. However, it not possible to distinguish the bioturbated from the argillaceous mudstones based only on mineralogical composition (**Figure *6***).

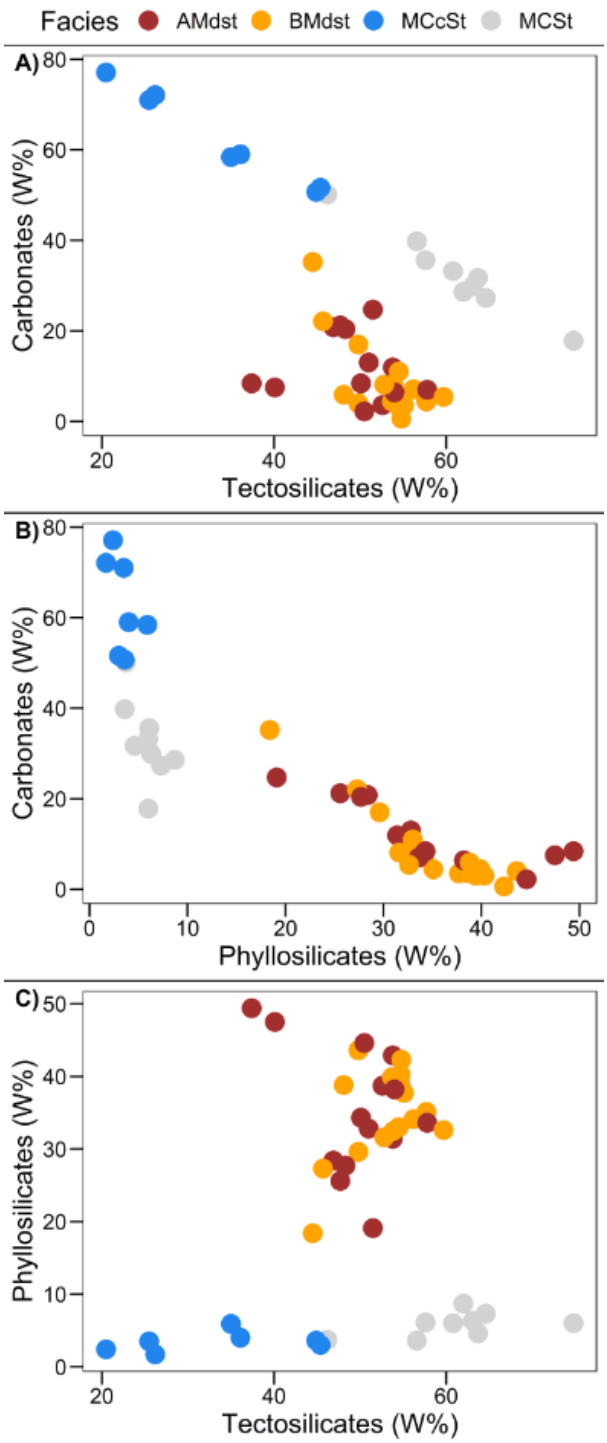

Figure 6: Cross plots from XRD data color coded by facies: Argillaceous mudstones (AMdst), bioturbated mudstones (BMdst), massive calcareous siltstones (MCcSt), and massive calcite-cemented siltstones (MCSt). A) Tectosilicates vs. carbonates. Silt-rich mudstones show a linear relationship between carbonates and tectosilicates. B) Phyllosilicates vs carbonates. Silt- and clay-rich mudstones are differentiated based on phyllosilicate content (silt-rich mudstones < 10 W% phyllosilicates). Massive calcite-cemented siltstones have higher carbonate content than the calcareous siltstones. C) Phyllosilicates vs. tectosilicates.

Stratigraphy

The contact between Sycamore Formation and the underlying Woodford Shale is unconformable and followed by a transition zone that consists of AMdst and BMdst facies. Similarly, the upper contact with the Caney Shale is unconformable. We divided the Sycamore into five parasequences (A, B, C, D, E) (**Figure *4***). These fining upward parasequences are bounded by flooding surfaces that are easily identified in the gamma-ray (GR) log; thus, we used them for correlations. Parasequence A corresponds to the lower transition zone from Donovan (2001) and Fay (1989), and it is characterized by clay-rich mudstones. Parasequences B, C, D, and E correspond with the upper Sycamore zone and they contain a higher proportion of massive siltstone facies (MCSt and MCcSt) at the base and a higher proportion of clay-rich mudstones at the top.

Detrital compositions

Massive siltstones are well sorted and mainly composed of tectosilicate grains (grains formed by single crystals of tectosilicates minerals) and ovoid calcareous pellets. Quartz grains, the most common tectosilicate grains, are subangular and range in size from silt to very fine sand and are more abundant in MCSt than in MCcSt. The pellets are well rounded and elongated in shape, and contrary to tectosilicate grains that are subangular. They are more abundant in the MCcSt than in the MCSt facies.

Clay-rich mudstones facies (AMdst and BMdst) are mainly composed of clays and quartz grains, but allochems and detrital organic matter are also present. Clay-rich mudstones have fewer pellets and calcareous rock fragments than the massive siltstones. However, fossil fragments longer than 1 cm (0.4 in) (mainly brachiopods) have not been observed in the massive siltstones. The most common ichnofacies is Phycosiphon (**Figure *7***A), but Nereites and Zoophycos also exist

in the Mississippian cores. From the I-35 Sycamore outcrop, Chondrites, Planolites, large Cosmoraphe, Thalassinoides, and Rhizocorallium are also documented in the clay-rich facies (Milad et al., 2020).

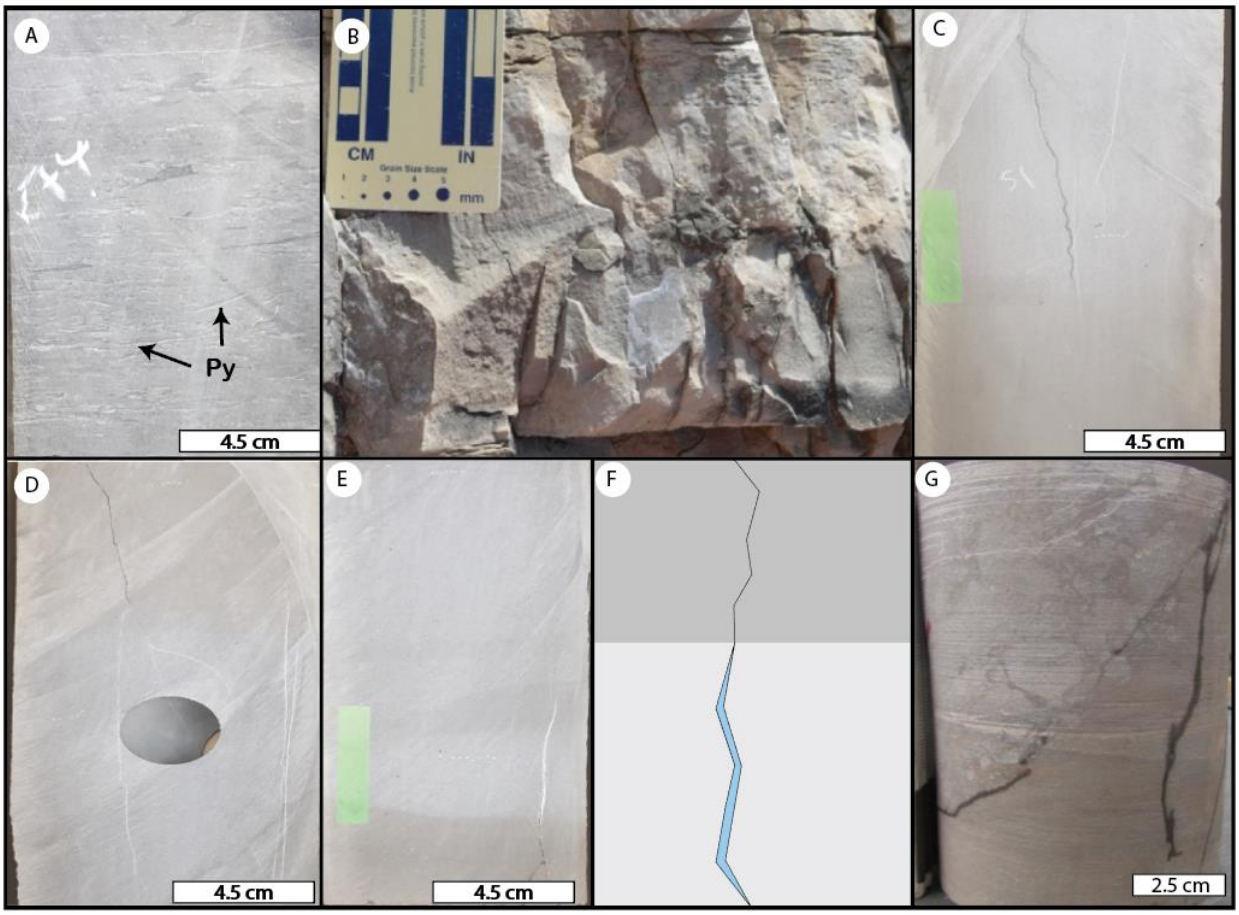

Figure 7: Sycamore Formation core and outcrop images. A) Bioturbation features caused by Phycosiphon (Py) in Moore East core. B) Average thickness of one complete parasequence in I-35 outcrop including massive siltstones at the base overlain by massive and bioturbated clay-rich mudstones. C-E) Fractures in the silt-rich siltstones in Myrtle Green core. Fractures are filled with calcite only when cutting massive calcite-cemented siltstone MCcSt. Note that the fractures filling material stop at the boundary between the MCcSt (lighter color) and MCSt (darker color). F) Schematic diagram shows how fractures are filled with calcite only when cutting through massive calcite-cemented siltstone MCcSt. The two siltstones types are easier to differentiate in core vs. outcrop. G) Breccia found in the Russell B core.

Diagenetic minerals and cements

The highest porosity and permeability observed in this study correspond to the massive calcareous siltstones (**Figure *8***A & **Figure *8***B). Consequently, the diagenetic history of these rocks is important. The next paragraphs present the main diagenetic minerals in these massive siltstones which are calcite and clay minerals, but we also identified pyrite, dolomite, and silica.

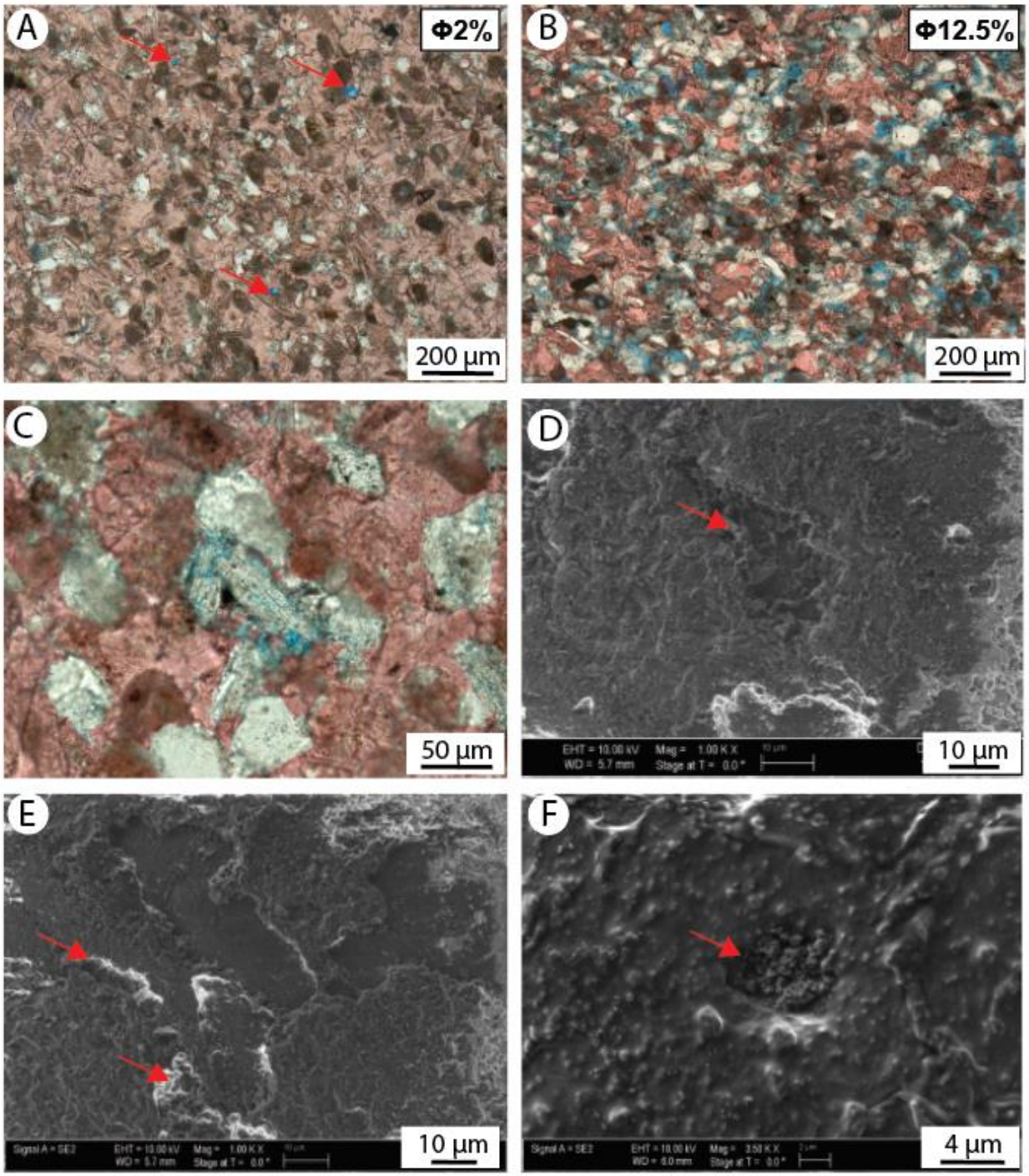

Figure 8: Sycamore Formation thin-section photomicrographs (A-C) and SEM images (D-F). A) Dissolution porosity in isolated feldspar grains in massive calcite-cemented siltstone (arrows) (appendix C53). B) Porous massive calcareous siltstone (appendix C8). C) Dissolution porosity in a feldspar grain (appendix C12). D-E) SEM images of dissolution porosity in massive calcareous siltstone (appendix C50). Note the rocks are not entirely cemented. F) Small-scale dissolution porosity in well-cemented massive calcite cemented siltstone (appendix C8).

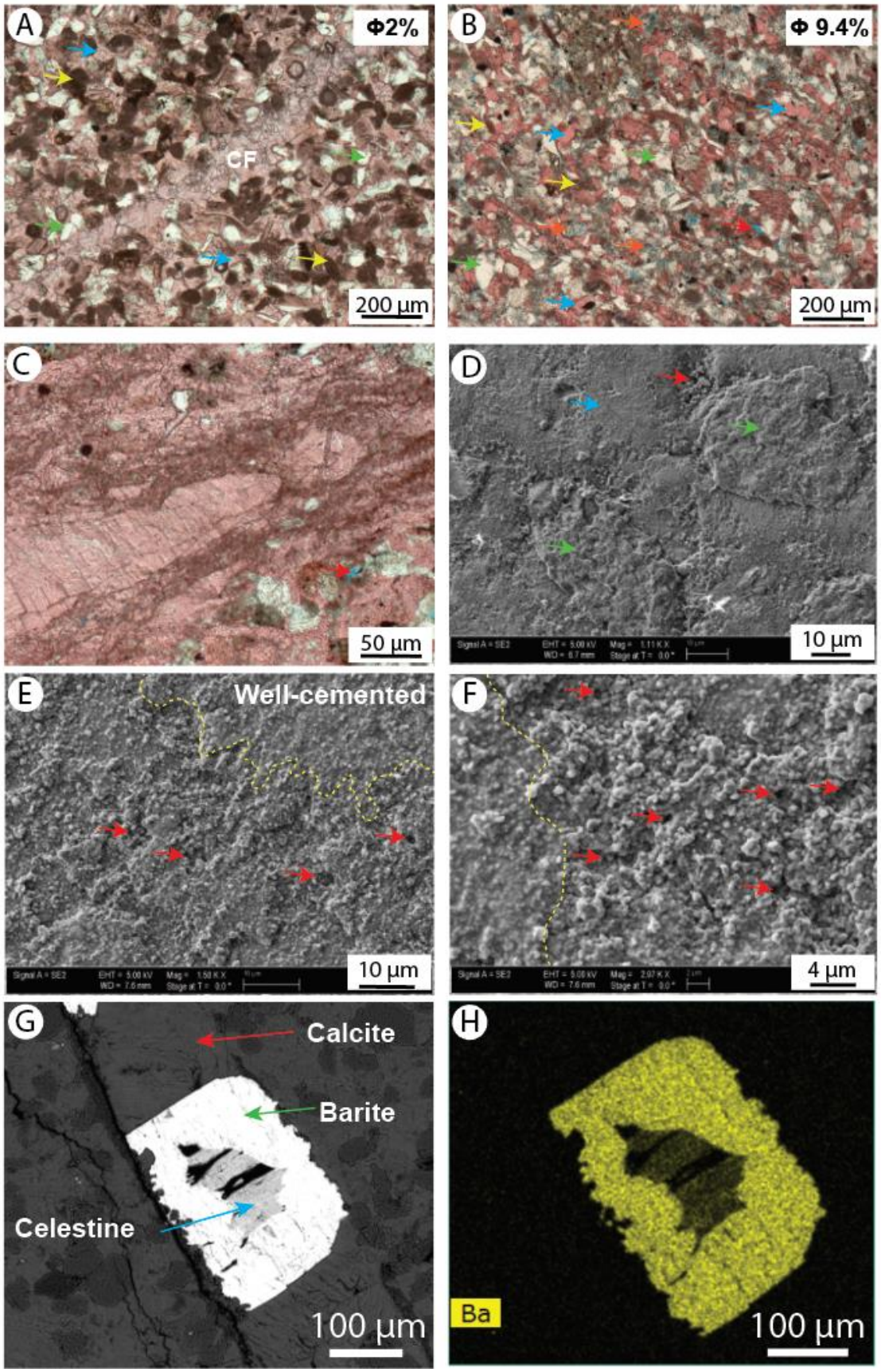
A
Φ2%
CF
200 µm
B
Φ 9.4%
200 µm
C
50 µm
D
10 µm
E
Well-cemented
10 µm
F
4 µm
G
Calcite
Barite
Celestine
100 µm
H
Ba
100 µm

Figure 9: Sycamore Formation thin-section photomicrographs (A-C) and SEM images (D-F). Various colored arrows highlight tectosilicate grains (green), pellets (yellow), intergranular calcite cement (blue), porosity (red), and dissolution porosity (orange). A) Massive calcite-cemented siltstone with abundant intergranular calcite and a calcite-filled fracture (CF) (appendix C24). Note that the cross-cutting relationship indicates that the calcite fill in the fracture is younger than the intergranular calcite cement. B) Massive calcareous siltstones. Note dissolution porosity is present in the sample and calcite cement is less abundant than in A (appendix C7). C) Micrite and calcite-filled fracture (appendix C20). D) Intergranular calcite cement, pellet, and quartz grain (appendix C50). E) Effect of calcite cement in micritic intervals (appendix C52). Note the division of the sample by the yellow line, the upper portion is well cemented without porosity and the lower portion is less cemented and with microporosity. F) Micritic interval with cemented portion to the left and visible micrite grains with microporosity to the right (appendix C52). G) SEM image showing a fracture filled with calcite, barite, and celestine (appendix C50). H) Barium map using an EDX coupled with the SEM to illustrate the variation in the elemental composition in the fracture (appendix C50). Based on cross-cutting relations calcite precipitated first, then celestine, and finally barite. These minerals have been reported in the underlying Woodford Shale (Roberts and Elmore, 2018) in a nearby location.

In general, calcite in the massive siltstone facies occurs as allochems, micrite, and calcite cement filling fractures (**Figure *9***A & **Figure *7***C-F) and the spaces between tectosilicate grains. This intergranular calcite, including micrite and cement, is higher in the MCcSt (avg. 44%) than in the MCSt (avg. 26%) facies. However, it is difficult to quantify these two types of calcite under the petrographic microscope due to the diffuse limit between them. The SEM is necessary to differentiate and qualitatively define the distribution of the micrite and intergranular calcite cement in the massive siltstones. The calcite cement is more common in the MCcSt than in the MCSt. In addition to the intergranular calcite cement, fracture-filling calcite is in sharp contact with other authigenic minerals including intergranular calcite cement (**Figure *9***A) and only precipitates in the fractures when cutting the MCcSt facies but not in the MCSt facies (**Figure *7***C-F). Also, this fracture-filling calcite is usually associated with celestine and barite (**Figure *9***G).

From XRD data, the main phyllosilicates in the Mississippian strata are illite/micas (up to 40%), mixed layers of illite-smectite (up to 4.9%), and chlorite (up to 0.9%). Phyllosilicates are less abundant than calcite cement but based on the XRD data they represent about 4.6% of the siltstone facies and 34% of the mudstone facies.

Porosity and permeability

Based on conventional core analysis, porosity ranges from 0.2 - 14.6% and permeability ranges from 0.0000044 - 0.27 mD. In general, calcareous siltstones have higher porosity followed by bioturbated mudstones whereas bioturbated mudstones have higher permeability followed by calcareous siltstones (**Figure *10***). Also, the distribution of the porosity shows that BMdst has a broad range of porosity and permeability.

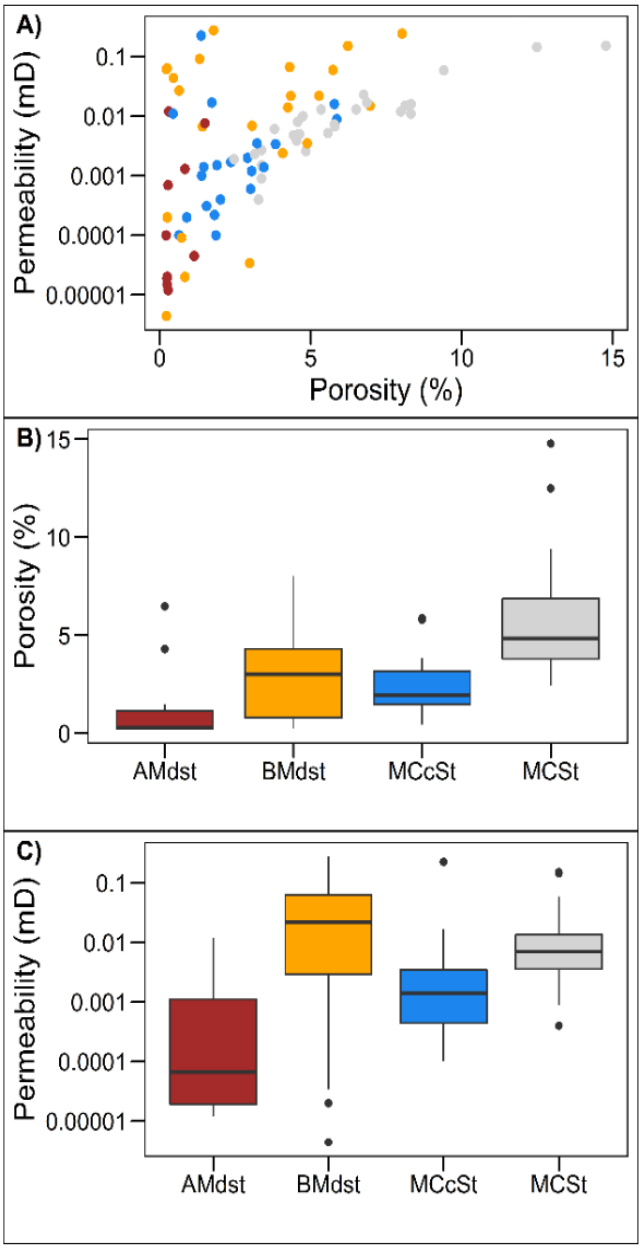

Figure 10: Porosity and permeability cross plot (A) and box plots (B, C). AMdst (n=16), BMdst (n=30), massive MCcSt (n=28), and MCSt (n=25) are illustrated. A) Cross plot of ambient porosity vs air permeability. The highest porosity values correspond to massive calcareous siltstones. A moderate positive relationship exists between porosity and permeability. However, some clay-rich mudstones and massive calcite-cemented siltstones with low porosity have permeability values higher than 0.01 mD. B, C) Box plots of porosity and permeability by facies. Massive calcareous siltstones have the highest porosity and Argillaceous mudstones have the lowest porosity values. Note the broad distribution of permeability in the Bioturbated mudstones.

Petrographic and SEM analyses show two main types of porosity in the massive siltstones: microporosity and dissolution porosity. Primary porosity in the massive siltstones is mainly related to microporosity between the mud-size calcareous grains. However, in some cases, integranular calcite cement occludes this microporosity. Microporosity is more common in the MCSt than MCcSt due to the limited abundance of calcite cement. Secondary porosity caused by the dissolution of detrital feldspar grains (dissolution porosity) is the other main type of porosity in the massive siltstones, MCcSt and MCSt. In addition, fracture porosity is present but is minor; therefore, it does not significantly impact RQ.

Elemental composition

average values in ppm of selected XRF elements, such as aluminum (Al), potassium (K), titanium (Ti), and silicon (Si) are substantially higher in clay-rich mudstones than in the Massive siltstones, the opposite happens with the carbonate proxies (Ca and Sr). By comparing the XRF values within clay-rich Mudstone facies, there is not a significant difference between BMdst and AMdst. However, by comparing the Massive Siltstones, the MCSt have a higher content of clay and continental input proxies (Ti and Zr), but less carbonate proxies than the MCcSt**.**

2.7 Discussion

The aforementioned data integration including petrophysical observations, SEM evaluation, geochemical analyses, core and outcrop investigation, and well-log analysis were used to: 1) reconstruct the paragenetic sequence, 2) determine the role of diagenesis on reservoir quality, and 3) investigate the geological factors that control the diagenetic evolution in the massive siltstone rocks. The two major Sycamore facies, which are massive siltstones and clay-rich

mudstones, are the result of differences in the conditions of deposition and sediment source. While massive siltstones are divided into calcareous siltstones and calcite-cemented siltstones based on the amount of calcite cement, the mudstones are divided into bioturbated and argillaceous based on the amount of clays and bioturbation.

Results show that the massive calcareous siltstones are more porous and permeable than the massive calcite-cemented siltstones. This is because MCSt was partially cemented, but not to the same degree as the MCcSt. Additionally, the highly cemented MCcSt restricted the movement of dissolving fluids inhibiting the generation of dissolution porosity. On the other hand, permeability in BMdst is relatively high compared to the other facies. Based on the petrographic analysis we found that BMdst are more prone to develop induced fractures during the sampling process, however more data needs to be analyzed to understand the relationship between bioturbation and fractures.

Paragenetic sequence of diagenesis

Petrographic and SEM analyses and XRD data were used to understand the paragenetic sequence of the Sycamore Formation. The main diagenetic processes that occurred in the Massive siltstones (**Figure *11***) are calcite cementation, feldspar and calcite dissolution, corrosion of quartz, smectite-to-illite transformation, calcite cement filling fractures, and pyrite formation.

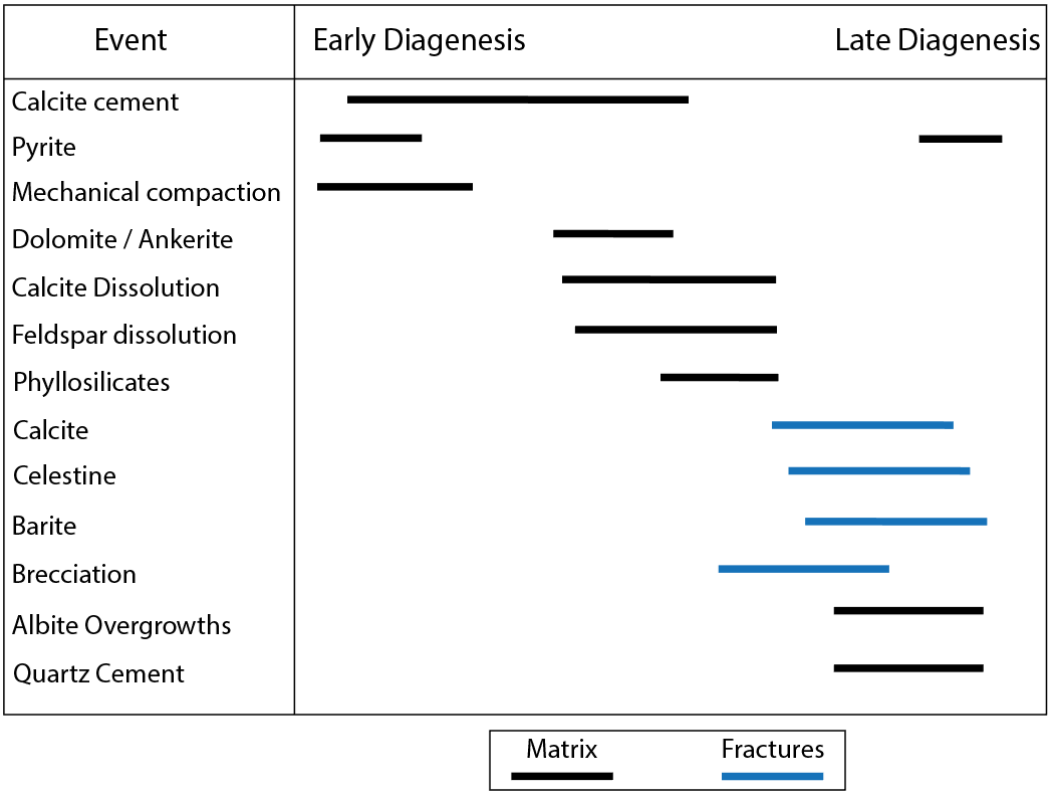

Figure 11: Paragenetic sequence of the Sycamore formation in the area of study. Note that calcite cement in the matrix started during early diagenesis and continue during middle diagenesis even when dissolution porosity started in some parts of the basin. Also, the calcite filled-fracture, and other digenetic minerals filling the fractures, are younger than the intergranular calcite cement.

Calcite cement

Two main types of authigenic calcite are present in the Sycamore Formation: intergranular calcite cement and calcite cement filling fractures. These types of calcite are not genetically related due to cross-cutting relationships that indicate the fracture fills are younger than the intergranular calcite cement (**Figure *9***A). There also could be differences in the source of the cementing fluids.

The intergranular calcite cement was precipitated during early diagenesis based on the absence of quartz cement and chlorite rims, and the floating texture in the massive siltstones (**Figure *9***A & **Figure *9***B), suggesting a pre-compaction event. Composition of early Mississippian seas were primary calcite over aragonite; thus, this support the idea of early cementation. However,

the origin of this cement using isotopic analysis is not clear and difficult to explain due to the deficiency of methodologies for separating calcareous cement from calcareous pellets, and the small size of the calcite crystals. Furthermore, SEM was used to distinguish between intergranular calcareous material that could be either micritic matrix or crystalline cement (**Figure *9***E-F). Based on SEM results (**Figure *8*** & **Figure *9***), micrite is more common in the MCSt and crystalline cement is more common in the MCcSt. SEM observations also indicate that the crystalline cement is replacing micritic matrix (**Figure *8***F, **Figure *9***E-F, & **Figure *12***). Hence, we propose an internal source in the same formation as the source for that material needed to precipitate the crystalline cement (Kantorowicz et al., 1987; Saigal and Bjørlykke, 1987).

Contrary to the intergranular calcite cement, fracture-filling calcite precipitated during late diagenesis (**Figure *9***A). Roberts and Elmore (2019) reported that the study area was affected by external fluids and that calcite, barite, and celestine fill fractures in the Woodford Shale. The presence of these minerals in the Sycamore suggests that, like the Woodford Shale, it was an open system late in its diagenetic evolution.

Internal sources, such as interbedded mudstones, dissolution of calcium-bearing feldspars, fossils, and calcareous pellets may be the source of the crystalline calcite cement (Morad et al., 1990). Some authors (Freed and Peacor, 1989; Sun et al., 2019) have reported that mudstones and feldspars could provide the needed ions ($Na^{+} + Ca^{2} + Fe^{2} + Mg^{2}$) to precipitate carbonate cement. Nevertheless, this does not explain the great amount of calcite that has been precipitated in the Massive siltstones. Therefore, it is more likely that micrite and allochems could account for most of the calcium needed for growing crystalline calcite. This idea is consistent with the observation that calcite cement is more abundant in samples with more calcareous material such as pellets.

This also might explain the floating texture given by the original abundance of micrite that has been replaced by the crystalline calcite cement.

Feldspar grain dissolution and quartz grain corrosion

Silt-sized tectosilicate grains are abundant in the massive siltstones. Feldspar dissolution is the main reason for secondary porosity followed by calcite dissolution, and both processes are usually attributed to the presence of organic acid-rich fluids (Harrison and Thyne, 1992; Worden and Barclay, 2000; Rahman and Worden, 2016). In the Sycamore Formation, these fluids could be sourced by clay-rich mudstones intervals within the Sycamore or most likely they originated from the underlying organic-rich Woodford Shale.

CO_2 derived from the organic-rich Woodford Shale during decarboxylation of organic matter may be responsible for the acidic fluid that is interpreted to have caused the dissolution of feldspars in the Sycamore (Worden and Barclay, 2000; Rahman and Worden, 2016). Additionally, the dissolution of feldspar grains contributes to the formation of the authigenic phyllosilicates in the Sycamore Formation by releasing K^+, Al^{3+}, and $SiO_2(aq)$ into the system (Sun et al., 2019).

Quartz cement in the Sycamore Formation is very rare and its minor presence might result from Si released during the feldspar dissolution.

Controls on reservoir quality

Reservoir quality in the Sycamore Formation, especially in the massive siltstones, is related to the diagenetic evolution. However, other geologic factors like depositional facies controlled the diagenesis in the Sycamore Formation.

Diagenetic controls on reservoir quality

Porosity evolution in the Sycamore Formation is affected by cementation and dissolution processes. Cementation, as opposed to compaction, is the main process that reduced original porosity and started during early diagenesis (**Figure *11***). The floating texture of massive siltstones supports the idea that calcite cement formed during early diagenesis, preventing the impact of mechanical compaction on the reservoir quality of the Sycamore. On the other hand, dissolution, occurred later than cementation in the paragenetic sequence. Our hypothesis is that initially, micrite fills intergranular spaces among the tectosilicate grains and generates microporosity between the mud-size calcareous grains (**Figure *12***). Then, cementation occurs, helped in part by the abundant calcareous material, and in part because microporosity facilitates the flow of the cementing fluids in the rock. This results in highly and poorly cemented facies. Finally, dissolving fluids move easier along permeable rocks than in highly cemented facies causing feldspar and calcite dissolution. Therefore, Massive Calcareous Siltstones-MCSt are less cemented and more porous than Massive Calcite-cemented Siltstones-MCcSt as a result of the cementation and dissolution processes.

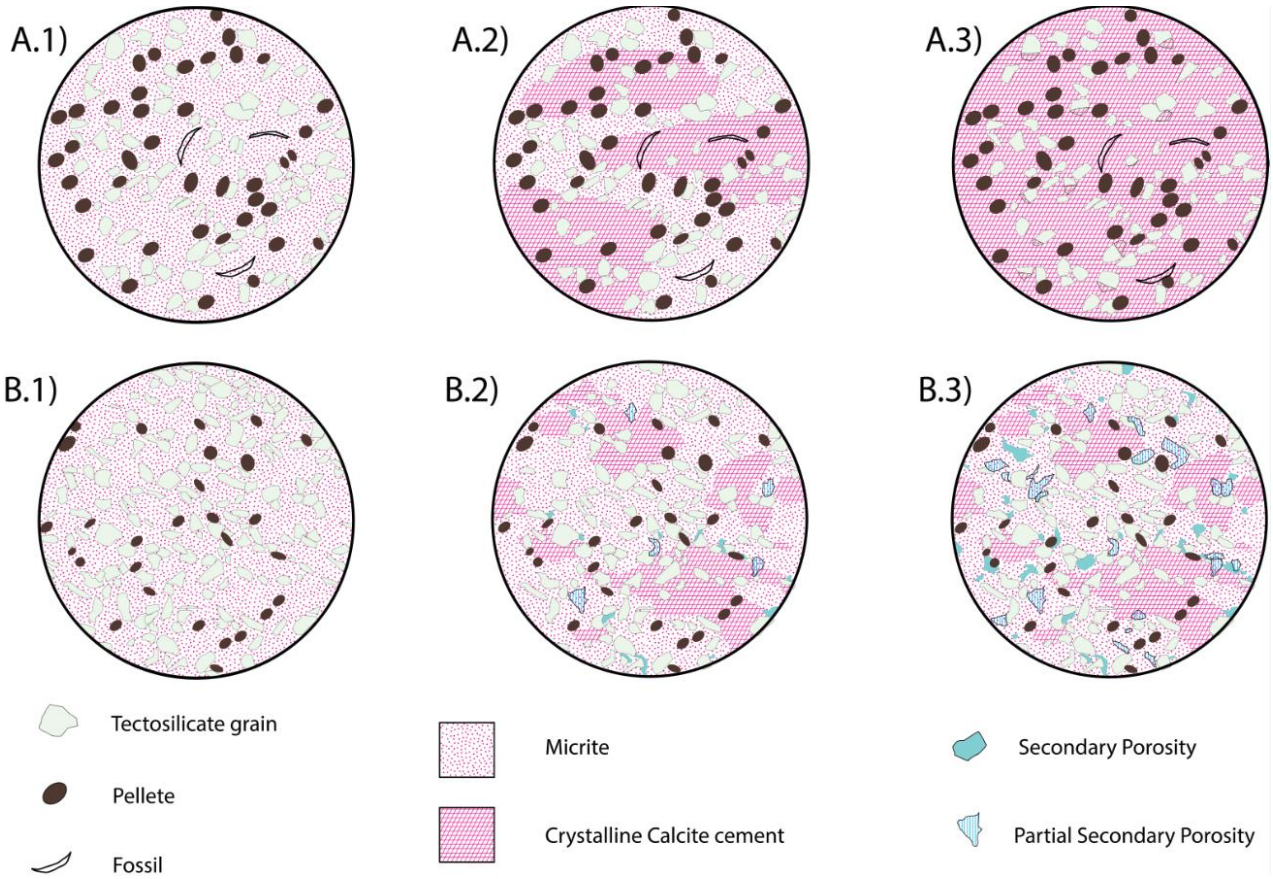

Figure 12: Schematic diagram to illustrate the simplified evolution of crystalline calcite cement and secondary porosity in A) massive calcite-cemented siltstones (MCcSt) and B) massive calcareous siltstones (MCSt) from relative time 1 to 3. Time 2 represent a transition point from depositional facies in time 1 to the facies described in this study in time 3. Sample B initially (time 1) has more tectosilicate grains and less micrite and pellets than sample B (in time 1). This resulted (time 3) in less calcite cement and more dissolution porosity in sample B than A.

Crystalline calcite cement occludes primary intergranular porosity and some microporosity (**Figure *12***). Therefore, micrite and the associated microporosity are important factors that control the presence of the calcite cement and consequently porosity. Point-count data (**Figure *13***) show an inverse relationship between undifferentiated intergranular calcite and porosity and permeability values. Point-count data and SEM observations strengthen the idea that calcite cement is a major (but not the only) factor that drives the evolution of the porosity in the Sycamore

Formation. Furthermore, the calcite cement decreases the permeability and consequently influences the movement of late diagenetic fluids responsible for the dissolution porosity.

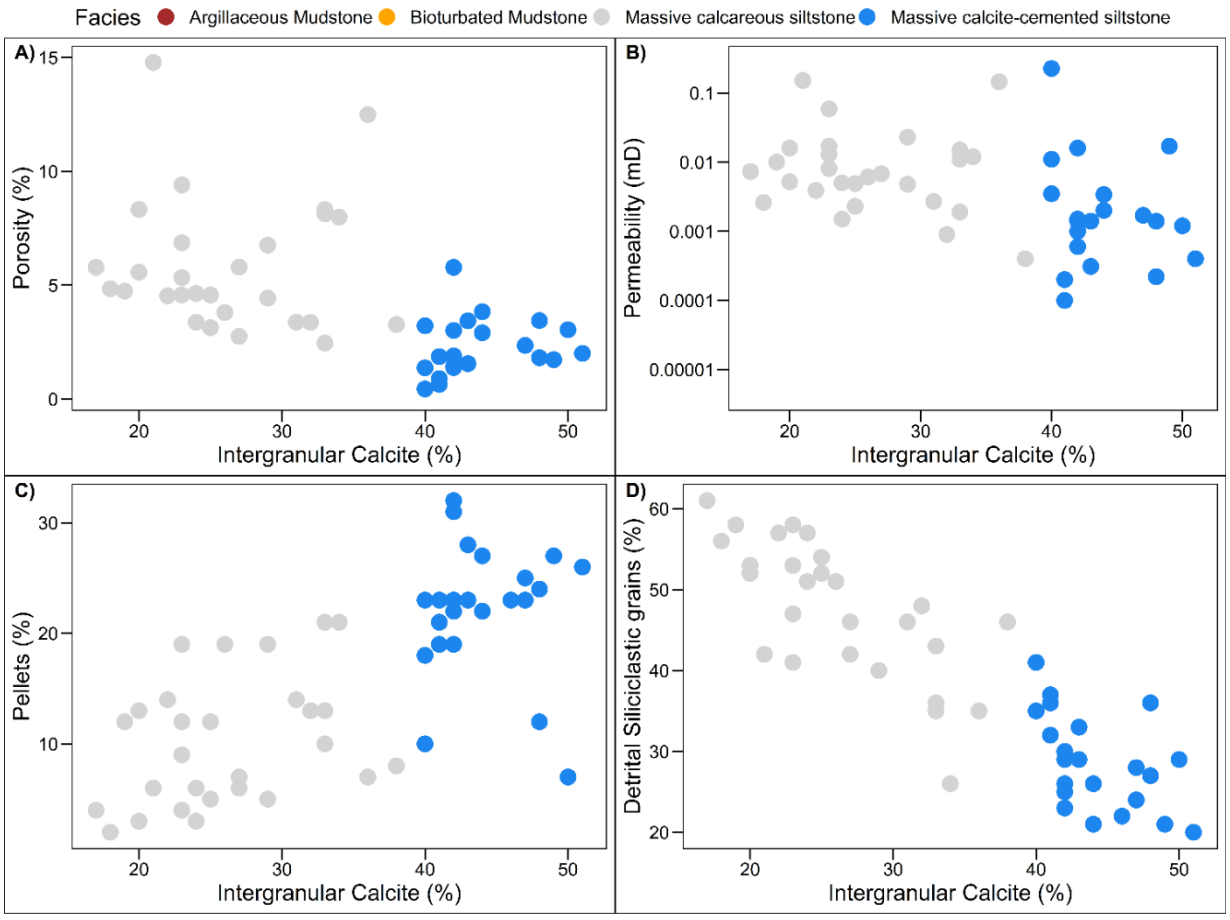

Figure 13: Cross plots using point counting data. They illustrate how the percentages of porosity (A), permeability (B), pellets (C), and detrital grains (E) vary with the percentage of intergranular calcite (undifferentiated) for massive calcite cemented siltstones (blue) and massive calcareous siltstones (gray). Pellet, detrital siliciclastic grain, and calcite cement percentages based on point counting (n=300). Porosity and permeability values are from conventional core plug analyses, and they present an inverse relationship with intergranular calcite. Also, note the direct relationship between intergranular calcite and pellets, but the inverse relationship between intergranular calcite and detrital siliciclastic grains.

Opposite to calcite cementation, feldspar and calcite dissolution increases the porosity of the rock. However, these two diagenetic processes are related to each other; calcite cement

decreases the permeability during early diagenesis and controls the flow of later diagenetic fluids. In other words, a decrease of permeability in zones with abundant calcite cement prevents late diagenetic fluids from permeating the rock and dissolving feldspar grains. Thus, comparing Sycamore massive facies, MCcSt facies has higher calcite cement and lower dissolution porosity than the MCSt. These resulted in low and high relative reservoir quality facies, respectively.

Depositional controls on diagenesis

Variation in the degree of diagenesis between the MCcSt and the MCSt is partially due to variations of conditions during deposition. For instance, the abundance of intergranular calcite is directly proportional with pellets abundance and inversely proportional with detrital siliciclastic grains (**Figure *13***C -D). This is due to greater input from the carbonate shelf in the north and west (Donovan, 2001) and results in abundant micrite filling spaces between tectosilicate grains. This implies that the abundance of calcite cement in the massive siltstones is driven by depositional facies. The greater the input of calcareous material in the depositional facies, the more intergranular micrite, and consequently the more calcite cement. Therefore, by understanding, sea-level change and position along the depositional profile or proximity to the sediment source can assist in predicting reservoir quality in the Sycamore.

MCSt facies are enriched in detrital siliciclastic grains and depleted/poor in calcareous material such as pellets, compared to the MCcSt that has less siliciclastic material, but more pellets (**Figure *13***C-D). These differences in the rock composition are the result of differences in sediment supply. **Figure *14***shows the relationship among paleoenvironmental proxies where MCcSt facies contains higher calcium (carbonate proxy) but less aluminum and titanium (clay and detrital input

proxies, respectively) than the MCSt facies. Therefore, the ratio of detrital grains over calcareous material, such as pellets and micrite, in the rock is given by the relative location between the feeder of the siliciclastic material and the source of the calcareous material. Consequently, depositional facies proximal from the carbonate source and distal from the siliciclastic source results in well cemented MCcSt. Also, depositional facies distal from the carbonate source but proximal to the siliciclastic input results in more porous MCSt facies. Therefore, diagenetic calcite cement is more abundant where the input from the carbonate shelf is more notable.

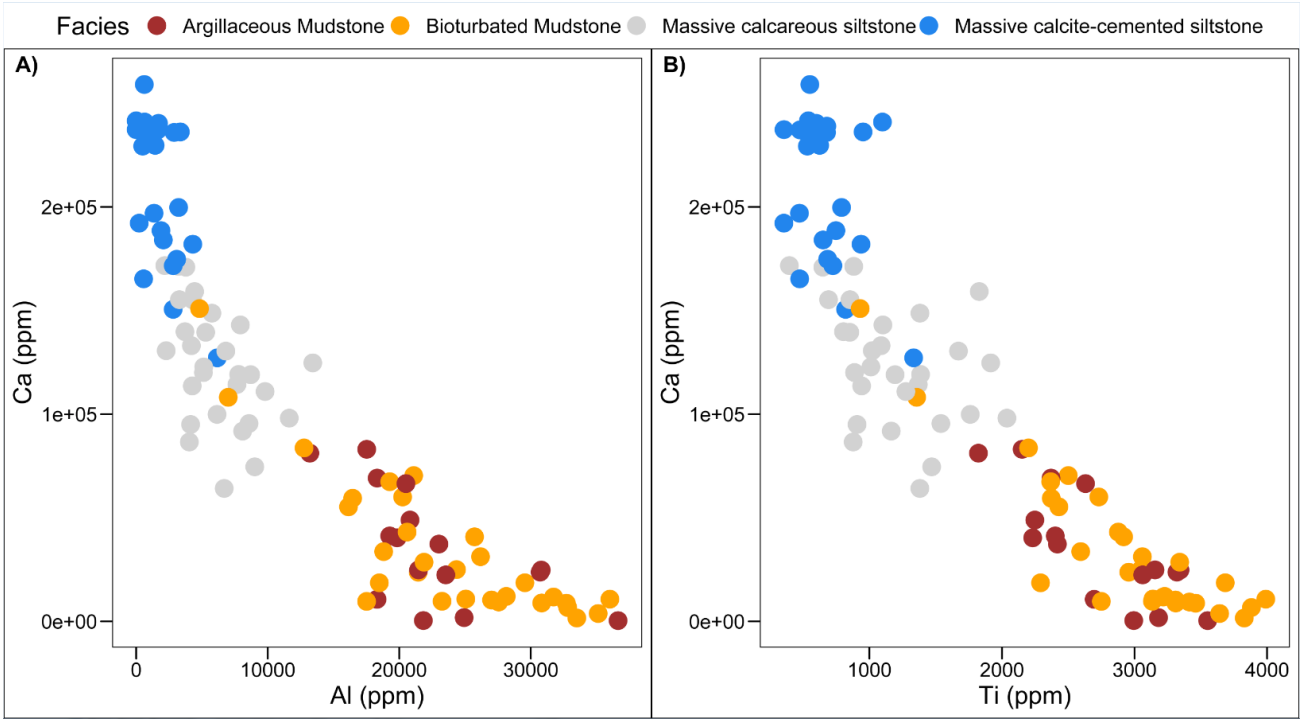

Figure 14: Cross plots of facies elemental abundance of calcium (Ca) as compared to aluminum (Al) and titanium (Ti) (in ppm based on HH-XRF). Al is a proxy for clays, Ca is a carbonate proxy, and Ti is a detrital siliciclastic input proxy. A) The Al vs Ca cross plot shows a negative relationship between the clay and carbonate content. B) The Ti vs Ca cross plot also shows a negative relationship between detrital siliciclastic input and carbonate content.

Additionally, more calcareous transported material such as micrite and pellets implies more material (e.g. Ca+) that could be the source to precipitate the crystalline calcite cement. This evidence presents at the fractures and the filling material (**Figure *7***C-F). The calcite cement only precipitated in the fractures when cutting the MCcSt facies but not in the MCSt. Therefore, the diagenetic evolution controlled in part by different depositional facies between the MCcSt and MCSt Sycamore facies has led to different reservoir quality.

Stratigraphic controls on diagenesis

The distribution of calcite cement in the Sycamore Formation is also temporally related to the paleoenvironmental conditions. Variable sediment input from the siliciclastic and carbonate systems through time resulted in differences in the distribution of calcite cement in the four parasequences that contain massive siltstone facies (B, C, D, E). Moreover, the consequence of this stratigraphic variation in the distribution of calcite cement is the evolution of different pathways or impermeable zones that affect reservoir quality and heterogeneity (Mansurbeg, Morad, et al., 2012).

Some authors have observed that the lowest porosity and permeability values are usually in the interface between mudstones and sandstones (Xiong et al., 2016; Yang et al., 2017; Sun et al., 2019). However, this is not the case in this study because, in the Sho-Vel-Tum field, the highest porosity values correspond with the base of parasequence C located immediately above the interface between clay-rich mudstones-massive siltstones; porosity decreases toward the top of parasequence C (**Figure *15***). Additionally, the lowest porosity and permeability values correspond

to parasequence B. The contact between these parasequences (B and C) is of special interest because the highest and lowest porosity values are located around this stratigraphic position.

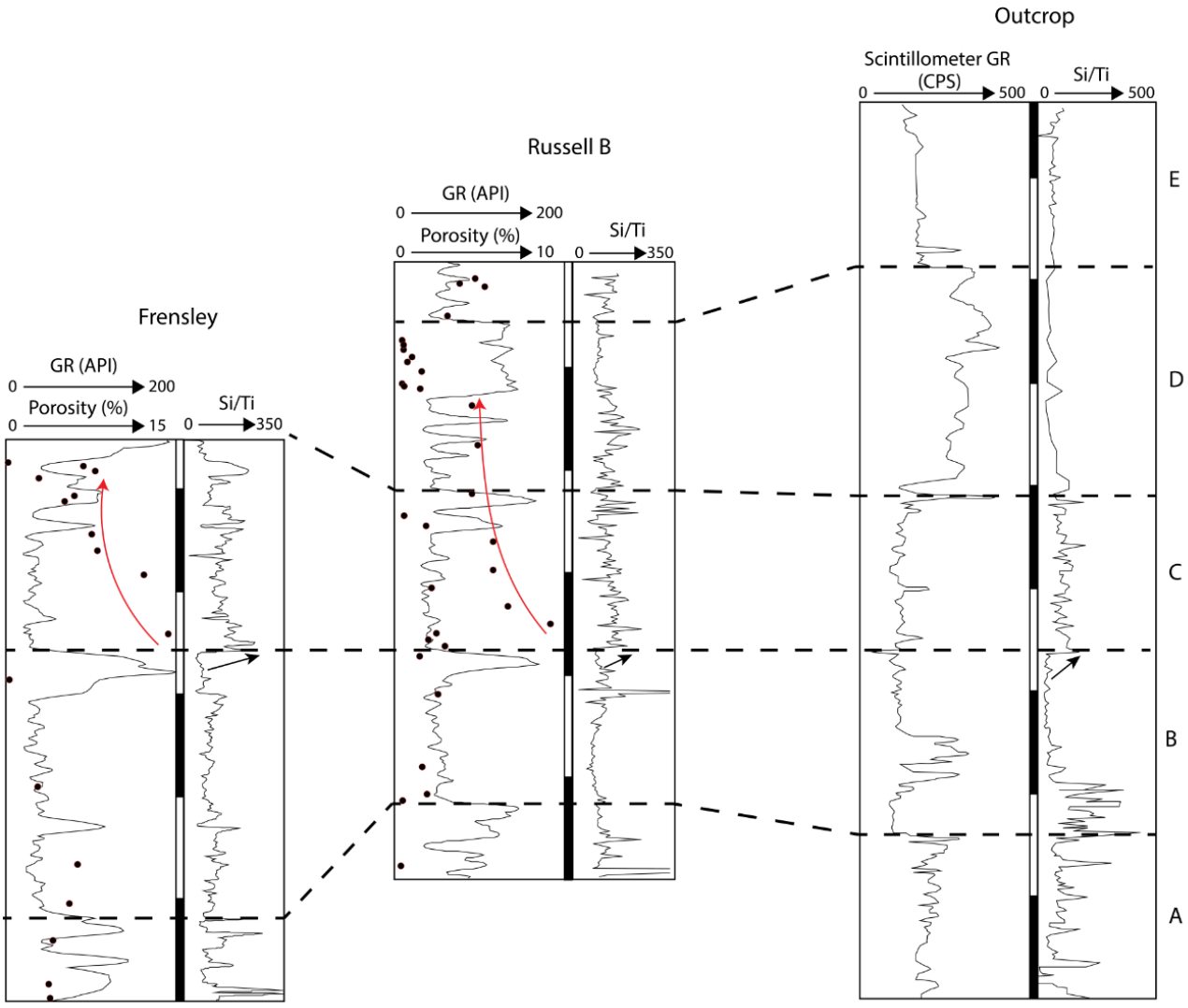

Figure 15: W-E stratigraphic cross section of Sycamore Formation Parasequences. Parasequence B is less porous than parasequences C, D and E (black dots represent porosity values base on core plugs). Additionally, the highest porosity values are at the base of parasequence C and decrease upward (red arrows). This is accompanied by a sharp increase in detrital siliciclastic input from parasequence B to C based on the Si/Ti ratio (black arrows). The top of parasequence B is the datum.

The siltstones from parasequence B have higher carbonate proxies (Ca) and lower clay proxies (Al and K) as compared to the other parasequence siltstones (C-E). Parasequence B is well cemented due to the replacement of abundant micritic material by crystalline calcite cement. The incursion in Si/Ti ratio from parasequence B to C (**Figure *15***) implies different compositions between these two sections due to changes in the detrital input composition. We observe that this boundary represents a sharp increase in clay content and a decrease in carbonate material from parasequence B to C. Therefore, this boundary could be related to a major sea-level fall and transition from a more carbonate-dominated system during early Mississippian, where a great part of North America was cover by shallow epicontinental seas (Gutschick and Sandberg, 1983), to a more siliciclastic-dominated system. This is also observed toward the north in the Anadarko basin, where the Osage formation is a pure carbonate system and the Meramec is a mixed carbonate-siliciclastic system (Price et al., 2019).

Model of depositional, stratigraphic, and diagenetic processes controlling the reservoir quality in the Sycamore Formation

Reservoir quality, especially porosity, in the Sycamore Formation is controlled by diagenetic processes such as calcite cementation and dissolution of feldspars and calcite; however, other intrinsic factors such as depositional environment and stratigraphy are also necessary to predict reservoir quality. The stratigraphic distribution of clay-rich mudstones and calcite cement distribution in massive siltstones is illustrated for the Sycamore parasequences (**Figure *16***).

Parasequence B overlies the transition zone (parasequence A), is well cemented, and exhibits the lowest porosity and permeability of the Sycamore parasequences that contains massive

siltstones. This might be caused by a major change in sea-level during Mississippian time that, analogous to the Anadarko basin in the north, resulted in a change depositional environment from a carbonate-dominated system (e.g. Osage), to a mixed carbonate-siliciclastic system (e.g. Meramec). The carbonate-dominated system resulted in a greater amount of crystalline calcite cement responsible for the low RQ in parasequence B.

Parasequence C has the highest porosity at its base and decreases in porosity upward (**Figure *15*** & **Figure *16***). This can be explained by diagenetic fluids that come from the organic-rich Woodford Shale and move on top of and across the well-cemented (impermeable) parasequence B. As a result, the diagenetic fluids move easily to the base of parasequence C, resulting in the generation of secondary porosity due to the dissolution of feldspars. This pattern has not only been observed from porosity and permeability data and petrographic analysis but also from well logs in Sho-Vel-Tum field where the highest porosity values are at the base of parasequence C. The distribution of calcite cement in the overlying parasequences (D and E) is variable depending on depositional facies and diagenetic history.

Although the aforesaid porosity trend is continuous in the Sho-Vel-Tum field, the absolute values vary depending on the area. For instance, Frensley and Russel B cores exhibit similar trends (**Figure *15***), but the Frensley core is generally more porous than the Russel B. This might be because the Frensley core is located next to a fault and fluids moving through this fault have a higher impact on the nearby formation by creating secondary porosity.

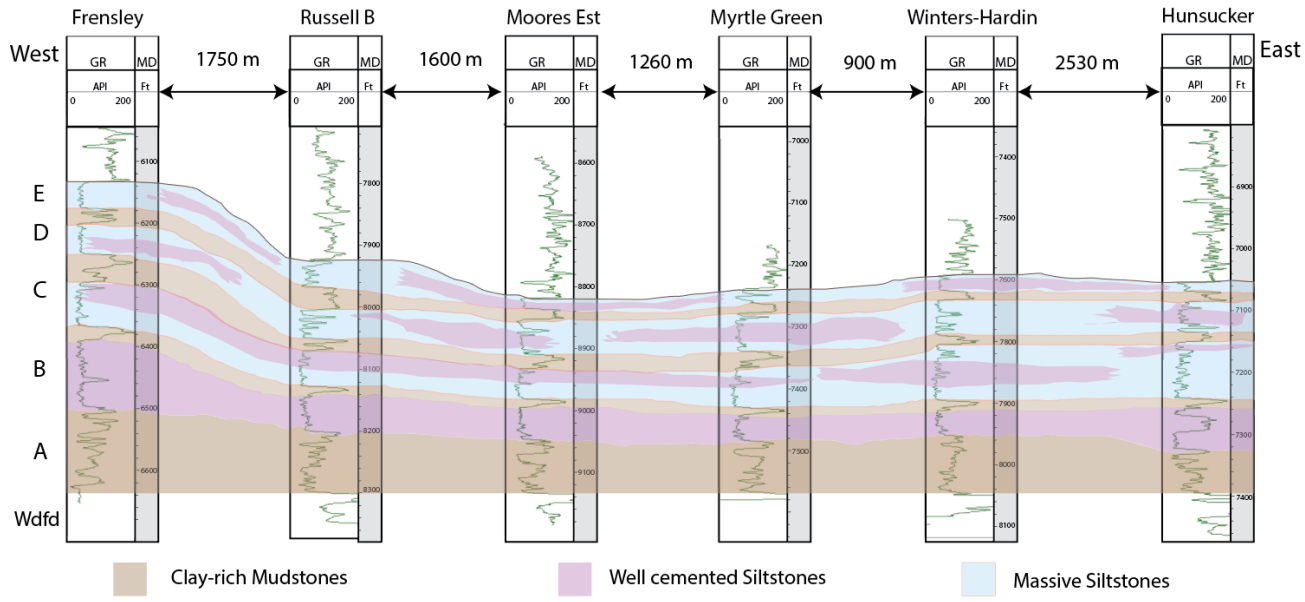

Figure 16: W-E stratigraphic cross section to illustrate the distribution of calcite cement in the Sycamore Formation siltstones at Sho-Vel-Tum Field. In general, the massive calcite-cemented siltstones are well cemented (pink) and the massive calcareous siltstones are the massive siltstones in blue. The massive siltstones from parasequence B are well cemented, then the lower portion of parasequence C is the most porous interval and decreases upward. The diagenetic fluids move easily to the base of parasequence C, resulting in the generation of secondary porosity due to the dissolution of feldspars, and as a consequence the rocks with better RQ are present in this interval.

2.8 Conclusions

- Clay-rich mudstones and massive siltstones are the two main facies of the Sycamore Formation at Sho-Vel-Tum field. The clay-rich mudstones are subdivided in argillaceous (AMdst) and bioturbated (BMdst) mudstones and the massive siltstones are subdivided in massive calcite cement (MCcSt) and massive calcareous (MCSt) siltstones.

- Cementation and dissolution are the main diagenetic processes that impact reservoir quality of the massive siltstones. These diagenetic processes are related to each other; calcite cement, that replaces micrite, decreases permeability during early diagenesis and controls the flow of later diagenetic fluids responsible for feldspar dissolution

- Differences in the sediment supply between carbonate- and siliciclastic-dominated systems controls the distribution of calcite cement in the Sycamore. Detrital siliciclastic input increase, accompanied by a decrease in pellets and micrite, leads to less calcite cement, thus better reservoir quality. The consequence of this variation in calcite cement is the evolution of different permeable and impermeable zones that affect how the acid fluids flow and generate secondary porosity.
- Stratigraphic variation of cement between the Sycamore parasequences is linked with conditions during deposition and diagenesis. The parasequence B is relatively more cemented than the other parasequences (C, D and E) that contains massive siltstones. Additionally, the highest porosity is at the base of the parasequence C and is associated with dissolution porosity of feldspars grains and carbonate material immediately on top of parasequence B.

2.9 Acknowledgements

This work was supported by the Woodford-STACK-Merge-SCOOP Consortium of the Institute for Reservoir Characterization at the University of Oklahoma directed by Dr. Roger Slatt. We thank our consortium sponsors for their support of this research and especially thank 89 Energy for sharing data. Also, we thank Dr. Abbas Seyedolali for his valuable comments and ideas that improved the manuscript. This work is in memory of Dr. Roger Slatt who was instrumental in developing and initiating this research.

Chapter 3: Semi-supervised Workflow to Generate Petrofacies Logs from Thin Sections and XRF Data

David Duarte, Rafael Pires de Lima, Deepak Devegowda, Matthew J. Pranter

3.1 Preface

This chapter is presented here as it was submitted to the Marine and Petroleum Geology journal. The data for this project was acquired through the STACK-Merge-SCOOP consortium led by Dr. Roger Slatt. The 2nd chapter resulted in a conceptual model to predict the reservoir quality in the Sycamore Formation. The 3 and 4 chapters combined try to validate this conceptual model by linking the thin-section information with 3D models. The results presented in this 3rd chapter concluded that machine learning techniques and XRF data could be used to increase the vertical resolution gap between thin-sections and well logs. Some of the highlights for this chapter are:

- UMAP dimensionality reduction technique demonstrated to be effective to transform high-dimensional XRF data to low-dimensional projections.
- A self-training approach combined with XRF-data enhances the vertical resolution of thin-section-defined petrofacies by creating pseudo-labels (predicted petrofacies) for cored intervals.
- Machine learning bridges the resolution gap between the scarce core data and the well log information.

3.2 Abstract

Core data provide valuable *in situ* information on the chemical and physical characteristics of subsurface formations. For example, from thin sections it is possible to define rock types with

similar mineral composition, lithologies, and pore types (i.e., petrofacies). Petrofacies logs illustrate the stratigraphic variability of reservoir rocks and are often used to constrain 3-D facies and petrophysical-property models. However, core data are commonly scarce, thus, alternative data, such as well logs, with a different vertical resolution are more commonly used to characterize the mineralogy, lithology, and porosity of subsurface formations. Given the differences in vertical resolution between thin sections and well logs, novel Machine Learning techniques and workflows are used to relate these data types and identify subtle changes in rock properties at an appropriate vertical resolution The data available are thin-section-based petrofacies and X-ray fluorescence (XRF) data on collocated cores. XRF data is relatively lower in resolution compared to thin-section images. Our aim is to use machine learning techniques to bridge the resolution gap between thin sections and well logs. To that end, we develop two semi-supervised approaches that use XRF data as input in combination with dimensionality reduction techniques to reliably classify the thin-section-based petrofacies. Both the semi-supervised approaches we use, a self-training approach and a labeled-clustering approach, achieve accuracies in excess of 90% on this dataset. Although we used different dimensionality reduction techniques, UMAP provided the best results for both semi-supervised approaches. By generating petrofacies logs, we bridge the resolution gap between core-based thin sections and well-log data. Additionally, we demonstrate that by including semi-supervised methods in routine core analysis, leads to enormous cost and time savings for stratigraphic correlation, identifying target zones, design horizontal wells, and constrain subsurface models.

3.3 Introduction

Table 1: Nomenclature for chapter 3

NOMENCLATURE			
Al	Aluminum	MCSt	Massive calcareous siltstones
Ca	Calcium	Mdst	Clay-rich mudstones
CGR	Core gamma-ray	ML	Machine learning
DBSCAN	Density-based spatial clustering analysis	PCA	Principal component analysis
ECS	Elemental Capture Spectroscopy	PCs	Principal components
GR	Gamma-ray log	SEM	Scanning electron microscope
HCA	Hierarchical clustering analysis	Si	Silicon
ICA	Independent component analysis	SVM	Support vector machine
ICs	Independent components	t-SNE	t- Distributed Stochastic Neighbor Embedding
K-means	K-means clustering analysis	UMAP	Uniform Manifold Approximation and Projection
KNN	K-nearest neighbor	XRF	X-ray fluoresce
MCcSt	Massive calcite-cemented siltstones	XRD	X-ray diffraction

From core observations and petrographic analysis of core-derived thin sections it is possible to define rock types with similar mineral composition, lithologies, and pore types (i.e., petrofacies). Core data and thin sections are generally scarce, therefore petrophysical properties from well logs are more commonly used to infer and characterize the mineralogy, lithology, and porosity of subsurface formations. Thin sections and associated core plugs provide direct mineralogical and porosity data for a subsurface formation, whereas mineralogy and porosity are not directly observed from well logs but can be estimate from other analysis such as X-ray diffraction (XRD) and routine plug analysis. Given these differences and the differences in vertical resolution and volume support between thin sections and well logs, novel Machine Learning (ML) techniques and workflows are used to relate these data types and to classify rock types or petrofacies at an appropriate vertical resolution so as to characterize subtle changes in rock properties.

There have been many applications of ML techniques to relate core and petrographic data to well logs. For example, supervised ML techniques have been used to classify core-based geological attributes (e.g. lithofacies) from well-log information (Dubois et al., 2007; G. Wang et al., 2014; Bhattacharya et al., 2016; Wood, 2019; Bressan et al., 2020). ML has fostered the

development of models that predict thin-section-defined information, such as mineralogical composition and diagenetic facies, from photographs and Elemental Capture Spectroscopy (ECS) logs (Cheng et al., 2018; Duarte-Coronado et al., 2019; Lai et al., 2020; Pires de Lima et al., 2020). However, it has been clear that for the supervised model to perform satisfactorily, big data sets containing numerous observations from both predictor and predicted variables are required, posing a stringent need for data that is not always feasible to fulfill.

Semi-supervised techniques, which stands somewhere between supervised and unsupervised learning, have emerged as an alternative tool. These techniques combine a small amount of labeled data (data with true labels) and a large amount of unlabeled data avoiding the challenges of finding a large amount of labeled data to train supervised models (Basu, 2009; Liu and Özsu, 2009). Semi-supervised approaches have been used to predict lithology and facies observations from well logs in a data set with abundant well logs but limited observations on the intended predicted variable (i.e. lithofacies) (Dunham, Malcolm, and Kim, 2020; Xie and Spikes, 2020). The challenge remains in obtaining unified and unambiguous workflows to enhance the resolution of the scarce core data at high accuracies and bridge the resolution gap between thin sections and well logs.

Determining petrofacies allows for the identification of prolific petroleum reservoirs (Watney et al., 1998) and the characterization of the diagenetic processes to predict subsurface petrophysical properties. We present a workflow that combines semi-supervised techniques and elemental composition of the rock to enhance the resolution of the thin-section-defined petrofacies in cores. By using the elemental composition and petrographic information from thin sections for the Mississippian Sycamore Formation at Sho-Vel-Tum Field in the Ardmore Basin of southern

Oklahoma (**Figure *17***), we develop a ML workflow to generate petrofacies logs (predicted petrofacies variations with depth) with high accuracy in cored wells.

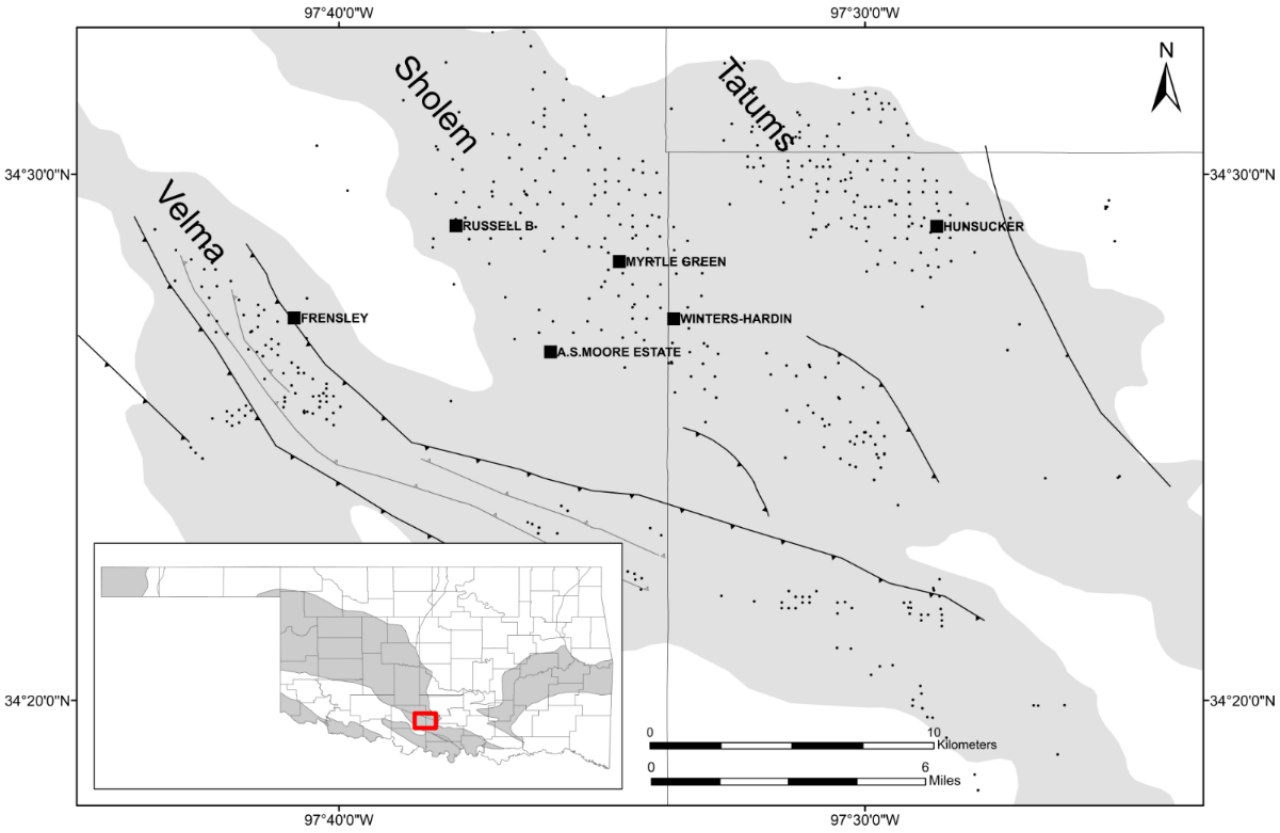

Figure 17: Location map of the study area. Black squares are cored well locations in Sho-Vel-Tum field (gray area) and black dots are non-cored wells. Faults in black in Sho-Vel-Tum field are from Carpenter and Tapp (2014).

3.4 Data & methods

We use core data for six wells from the Sycamore Formation in the prolific Sho-Vel-Tum Field (**Figure *18***). The core data include 96 thin sections that are used to define petrofacies and the elemental composition of the rock. The thin-section-defined petrofacies can be considered to be true labels that we attempt to predict in a ML workflow using elemental compositions derived

from the XRF. The elemental compositions for 29 elements was conducted at 0.3-m (1-ft) intervals in the cores (>1,000 samples) with a hand-held X-ray fluorescence analyzer (herein referred to as XRF data).

Figure *19* shows the general workflow with the main steps implemented in this study: i) define petrofacies (true labels) using thin sections, ii) preprocess XRF data, iii) use techniques to reduce the dimensionality of the XRF data, and iv) train two semi-supervised approaches use XRF data as input to classify petrofacies logs.

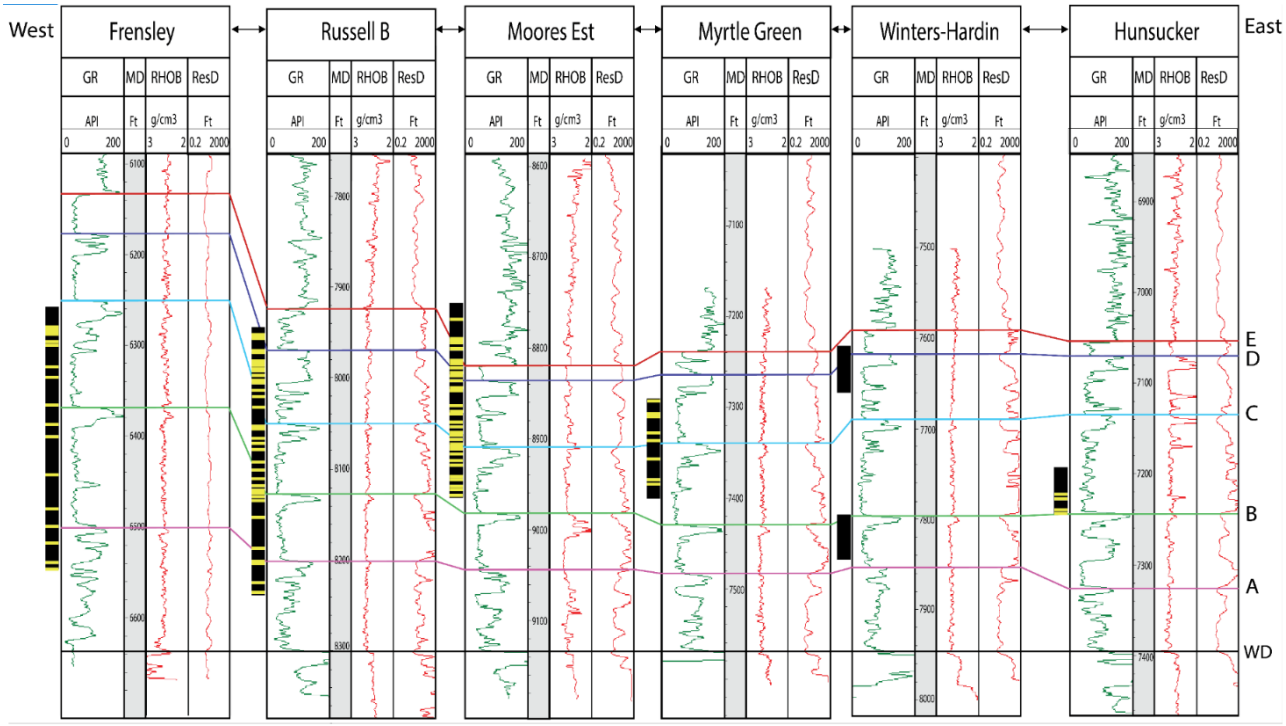

Figure 18: Cross section from west to east with the 6 cored wells and the thin-sections locations. The black rectangles next to the well logs (Gamma Ray-GR, density-RHOB and resistivity-RESD) represent the cored section where the X-ray fluorescence (XRF) data were measured at 0.3-m (1-ft) resolution. The yellow lines in the black rectangles represent the locations of the thin-sections that were used to define the petrofacies. The Sycamore Formation overlies the Woodford Shale (WD), and it is divided in parasequences from A to E.

Data processing

XRF data were normalized with min-max normalization method:

$$x\ s_i\ scaled = \frac{x_i - \min(\boldsymbol{x})}{\max(\boldsymbol{x}) - \min(\boldsymbol{x})} \quad (1)$$

where $x\ s_i$ is the scaled version of sample $i \in (1, 2, \ldots, N)$ of variable $\boldsymbol{x}$ containing N samples. This method rescales all variable values in the range of [0,1] to prevent one variable having an increased importance when compared to others only based on its range. Contrary to some other normalization methods, min-max scaler preserves the natural distribution of data. XRF data were normalized prior to being used in the following steps (**Figure *19***). Samples that lacked values for any variable were discarded.

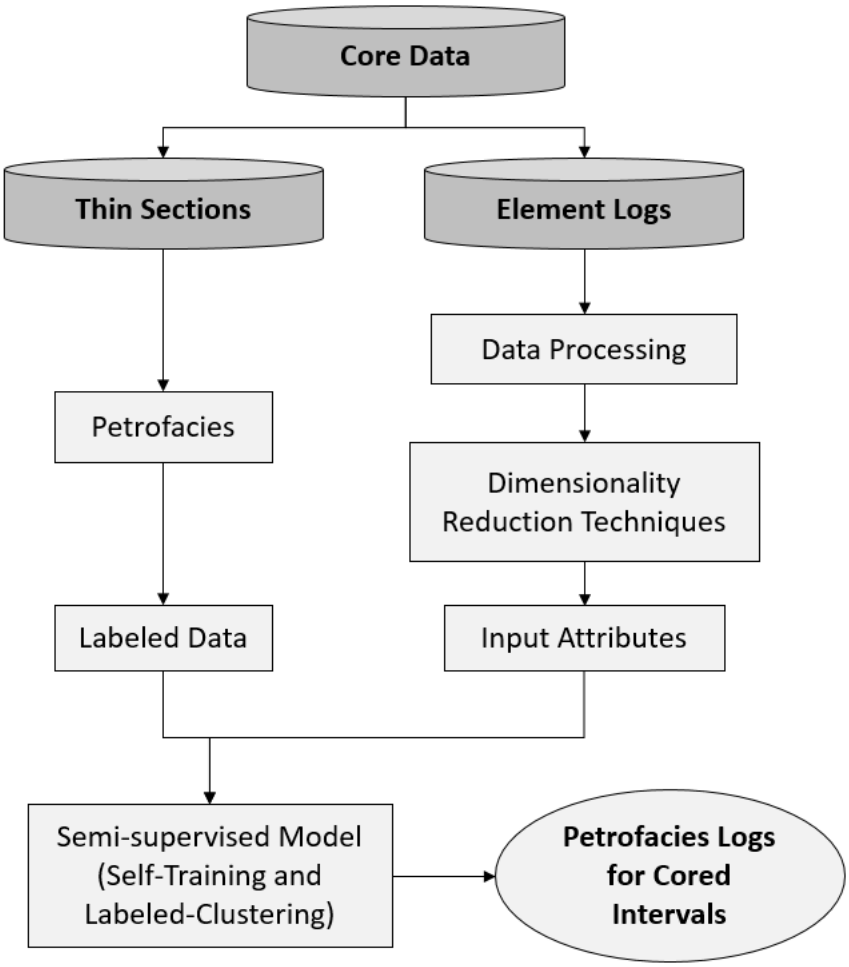

Figure 19: Generalized workflow that uses core thin-section-defined petrofacies and XRF data to create petrofacies log in cored intervals. Cylinders in dark gray represent data used and oval represent the output for each of the two main steps. The thin-sections (N=96) were used to define the three petrofacies and the X-ray fluorescence (XRF) provides the element logs (29 elements at 0.3-m resolution) of the rock. Two semi-supervised approaches were tested, self-training and labeled-clustering.

Dimensionality reduction techniques

ML methods have a tendency to perform poorly when using high-dimensional data (too many variables) due to a phenomenon defined as the curse of dimensionality, which causes data to appear equidistant in high-dimensions (Bellman, 1961). For example, clustering methods that measure

the distance between points produce poor results when dealing with too many variables. Thus, we use five techniques to transform 29-dimensional XRF data to a 3-dimensional representation. The first technique consisted of simply selecting the most important elements based on our geological knowledge of the rock. These were calcium (Ca), aluminum (Al), and silicon (Si). We called this set of elements "proxies". We also use four different algorithms including Principal Component Analysis (PCA), Independent Component Analysis (ICA), t- Distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP). We evaluated the dimensionality reduction algorithms based on how well these techniques cluster the true labels in 2D graphical representations (qualitatively), and in terms of the fidelity of the classification using the semi-supervised models. Thus, we used the accuracy of the final supervised model to define which dimensionality reduction technique resulted in models with higher accuracy.

PCA and ICA apply linear transformations to reduce high-dimensional data into fewer dimensions called principal components (PCs) and independent components (ICs), respectively (Pearson, 1901; Comon, 1994; Hyvärinen and Oja, 2000; Hyvärinen et al., 2001; Jolliffe, 2002). While PCA computes projections based on the variance of the data, ICA aims to separate superimposed information into different sources. On the other hand, t-SNE and UMAP are stochastics and nonlinear techniques, respectively, used mainly for visualizing high-dimensional data in a two or three-dimensional maps (van der Maaten and Hinton, 2008; McInnes et al., 2018). Because t-SNE is a stochastic method, it produces different results with different initializations. Thus, selecting the parameters, such as perplexity and learning rate, is important for the optimization of the method. Perplexity is related to the number of neighbors and the learning rate is related to the separation of each point is from its neighbors. In this study, we use values of 5 and

12, respectively. The main parameters to tune in UMAP are the number of nearest neighbors that balances local versus global structure, and the minimum distance between points in the low-dimensional space that controls how tightly the points are packed together. In this study, we use values of5 and 0.35, respectively.

Semi-supervised model

Semi-supervised ML methods use a small amount of data with true labels and a large amount of unlabeled data to generate pseudo labels (predicted petrofacies) (**Figure *20***). In this case, the data with true labels, or labeled data, are the thin-section-defined petrofacies. Collocated XRF data is used in two semi-supervised approaches to generate petrofacies logs at 0.3 m (1-ft) resolution. We implemented two using the XRF data as the input. The first approach, self-training, uses a supervised classification model to create the pseudo-labels (predicted petrofacies for unlabeled data). And the second approach, labeled-clustering, uses unsupervised algorithms to create the pseudo-labels.

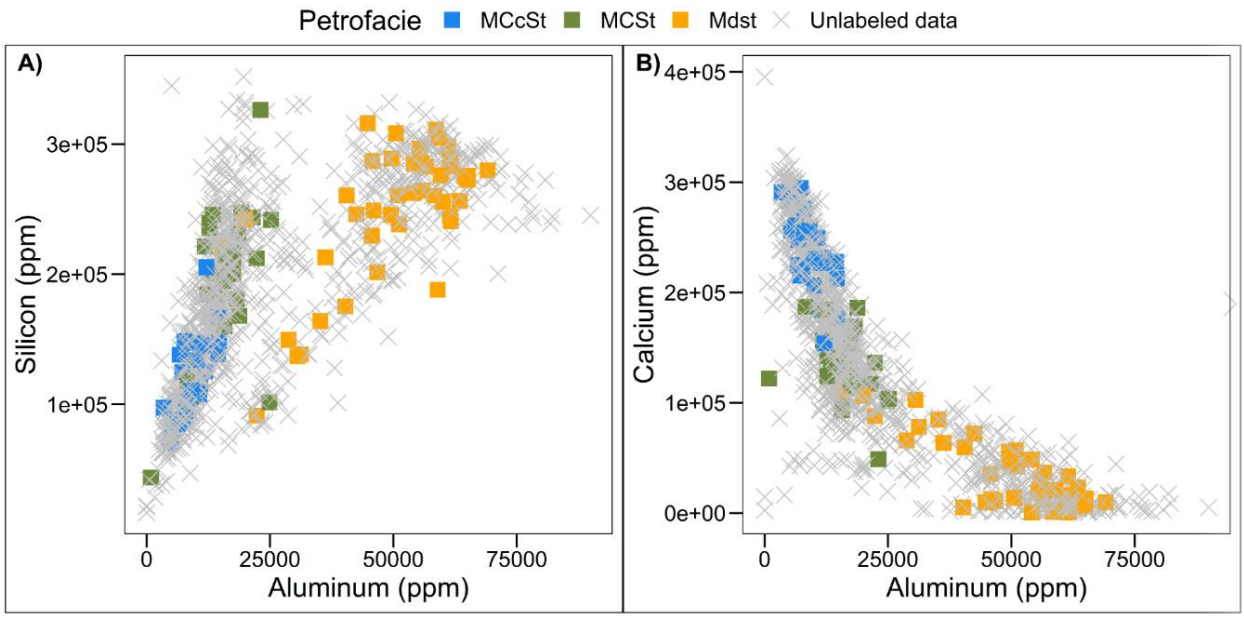

Figure 20: Cross plots of the elemental abundance in the cores. Aluminum (Al) as compared to silicon (Si) and aluminum (Al) as compared to calcium (Ca) (in ppm based on XRF data) show the distribution of unlabeled data (x). The three thin-section-defined petrofacies (squares) are referred in the text as true labels. Notice that labeled data represent a small portion (~8%) of the whole data and the three petrofacies are grouped based on the elemental composition (Al, Si, Ca). Also, in the Al vs Si cross plot clay-rich Mudstones (Mdst) are separated from massive siltstones (MCcSt and MCSt).

Self-training approach

With the self-training approach, we use three different input data: i) All 29 XRF-derived elements, ii) only the proxies (Al, Ca, and Si), and lastly, iii) the results from the dimensionality reduction techniques as the input data. First, we train a support vector machine (SVM) classification model (supervised) to classify petrofacies. (**Figure *21***). The classified petrofacies, or pseudo-labels, are associated with a probability of being correctly predicted. Pseudo-labels with higher probabilities than a set probability threshold of being correctly predicted are used to train a new model. This step is repeated until no more predicted petrofacies achieved the probability threshold (Steen, 2020; ***Figure 21***). We use three different probability thresholds (60%, 80%, and 90%). Finally, we

train a classification model using only the pseudo-labels, and we validate the model with the true labels. For the SVM (Boser et al., 1992) model parameters, we usegamma=0.01 and C=1 following hyperparameter tuning.

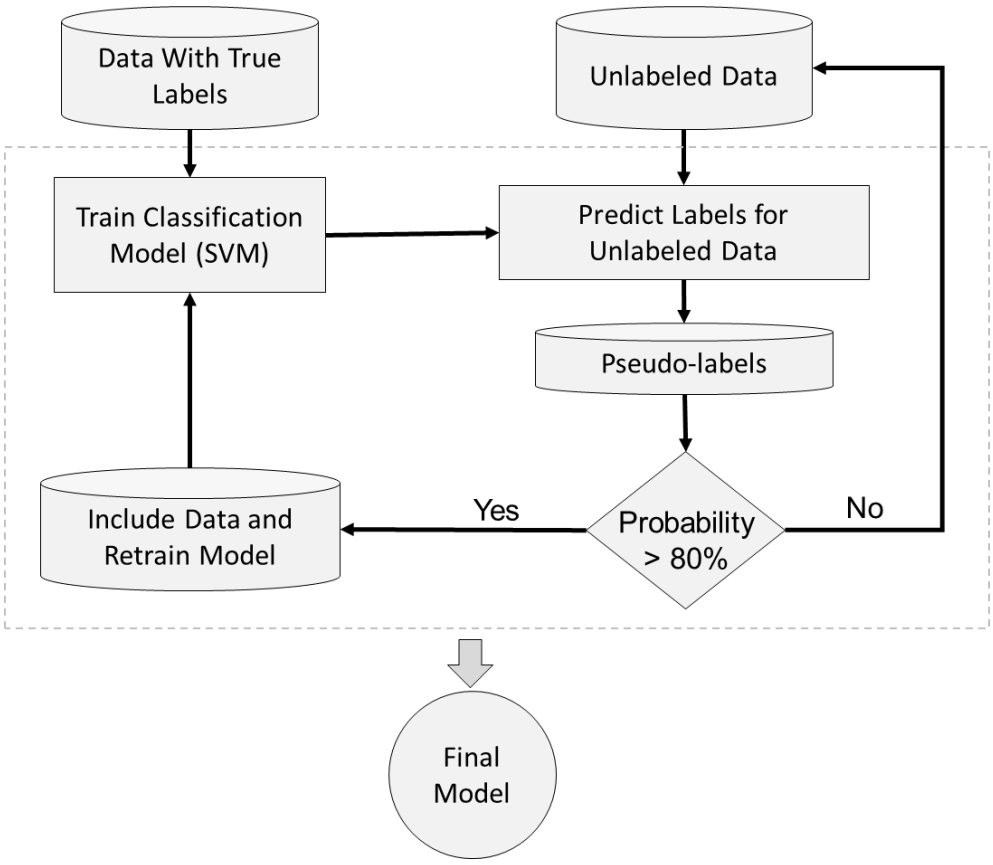

Figure 21: Self-training workflow. Self-training uses the scarce data with thin-section-defined petrofacies (true labels) to create pseudo-labels (classified petrofacies) for core intervals without thin-sections. First, we trained a support vector machine (SVM) classification model that uses XRF data to classify petrofacies in unlabeled data. These classified petrofacies, or pseudo-labels, were weighted by the probability of being correctly predicted. Then, pseudo-labels with probability higher than 80% of being correctly predicted were used to train a new model. This step was repeated until no more predicted petrofacies achieved the 80% probability threshold (Steen, 2020). This workflow was repeated with different input data and probability threshold (60%, 80%, and 90%). All 29 elements, the proxies (Al, Ca, Si), and the results of the dimensionality reduction techniques (PCA, ICA, t-SNE, UMAP) were used as the input data.

Labeled-Clustering approach

The second semi-supervised approach we use is based on unsupervised clustering and the resulting clusters are compared to the true labels obtained from thin sections. Again, we use three different sets of input data and compare the results. We use all 29 elements, only the three selected proxies and finally, the results of the best dimensionality reduction technique as determined by the results of the self-training approach.

We use three different clustering algorithms: K-means, hierarchical clustering analysis (HCA), and density-based spatial clustering (DBSCAN) (Xu and Tian, 2015). To validate the results, we assign labels to each cluster. These clusters are compared to thin-section-defined petrofacies to calculate the accuracy of the model. Accuracy is defined as the number of accurate predictions relative to the size of the dataset. Similar to accuracy in classification models, this helped to compare clustering techniques, that *sensu stricto* are unsupervised methods.

In all clustering methods, the number of clusters is defined as the number of petrofacies defined from thin sections. For k-means, we used 3 number of cluster, 100 iterations and the accelerated algorithm from Elkan (2003), while for HCA, we use Ward linkage (Ward, 1963) and Euclidean distance (Dokmanic et al., 2015). DBSCAN relies on identifying high density grouping of datapoints as clusters and relies on two important parameters: “eps”, the maximum distance between points for two points to be considered from the same cluster and “minSamples”, the minimum number of points in the neighborhood that the algorithm requires to form a cluster (Ester et al., 1996; Schubert et al., 2017; Campello et al., 2020). We used five for the *minSamples* parameter, but we varied the *eps* parameter based on the input data: we used 0.25 when using all 29 elements as the input data; 0.05 for the proxies; and 0.6 for UMAP projections. The advantage

of DBSCAN over other clustering algorithms is that it does not require the user to specify the number of clusters.

3.5 Results

Two main rock types are present in the Sycamore Formation: clay-rich mudstones (Mdst) characterized by high clay content, and siltstones characterized by abundant silt-size quartz grains, calcite cement, and calcareous allochems (**Figure *22***). Siltstones are divided into Massive calcareous siltstones (MCSt) and Massive calcite-cemented siltstones (MCcSt); MCSt being more porous than MCcSt. Calcium is higher in MCcSt than in MCSt due to calcite cement; however, silicon is higher in MCSt because it has higher quartz and clay content than MCcSt. Once the three petrofacies, Mdst, MCSt, and MCcSt were defined, we implemented the ML techniques.

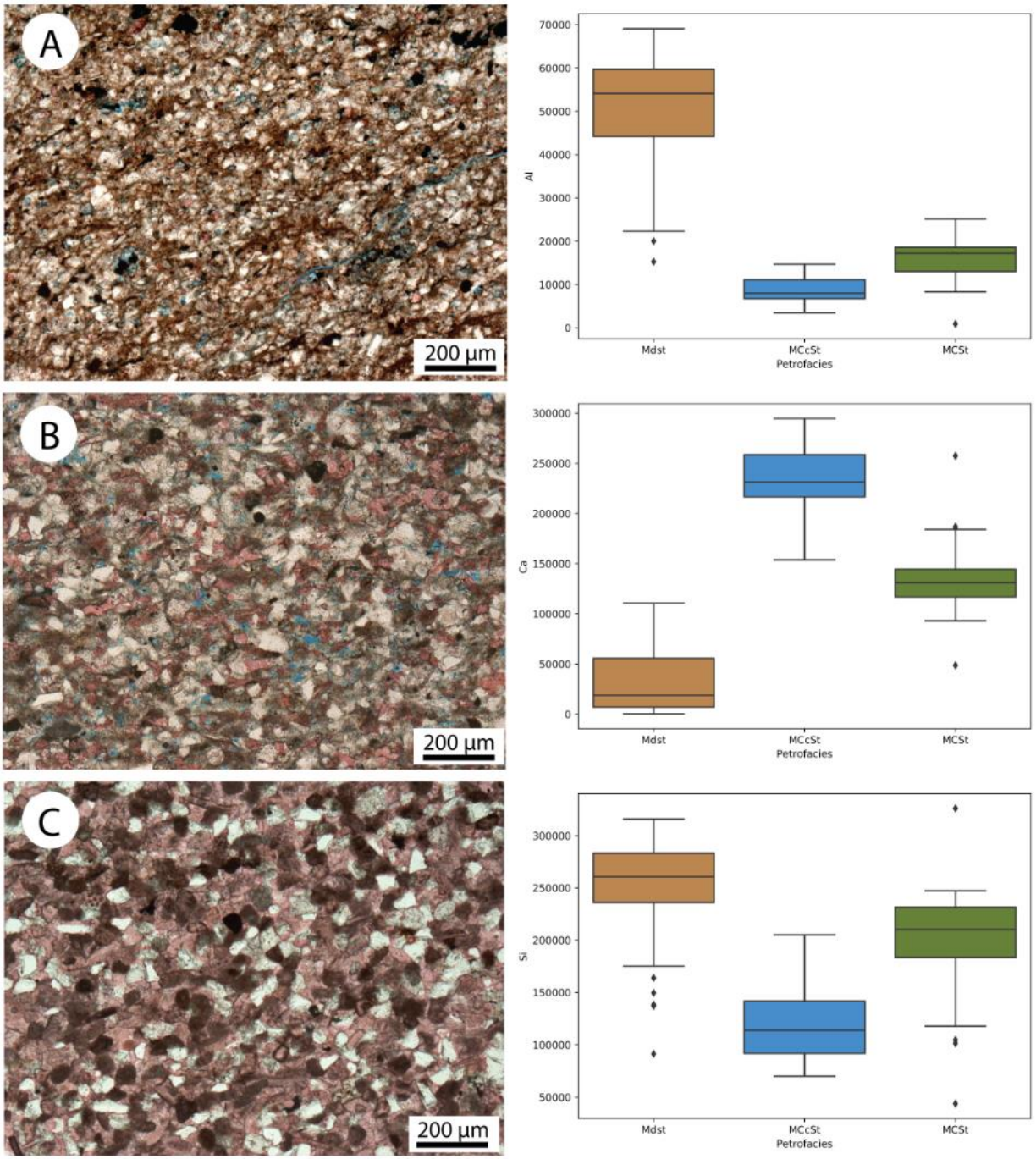

Figure 22: Petrofacies characterization. Thin-section images for the three thin-section-defined petrofacies A) Mudstones (Mdst), B) Massive calcareous siltstones (MCSt), and C) Massive calcite-cemented siltstones (MCcSt). The boxplots depict the Aluminum (Al), Calcium (Ca), and Silicon (Si) distribution for each petrofacies. Al is used as an indicator of the clay content, the Ca for carbonates and Si for Silicon. The clay-rich mudstones (Mdst) are characterized by high clays content, and massive siltstones by abundant silt-size quartz grains, calcite cement and calcareous allochems. These massive siltstones were divided into Massive calcareous siltstones (MCSt), and Massive calcite-cemented siltstones (MCcSt), being the former more porous than the latter MCcSt.

Also, calcium is higher in MCcSt than in MCSt due to calcite cement. On the other hand, Silicon is higher in MCSt because it has higher quartz content than in MCcSt.

Dimensionality reduction techniques

Visualizations are an effective tool to evaluate dimensionality reduction techniques. **Figure *23*** shows the results of the four transformations (PCA, ICA, t-SNE, and UMAP) including the unlabeled data, and the data with the true labels (thin-section-defined petrofacies). **Figure *23*** corroborates the observations from the petrographic analysis that rocks samples are divided into two main groups, mudstones (Mdst) and siltstones (MCSt and MCcSt).

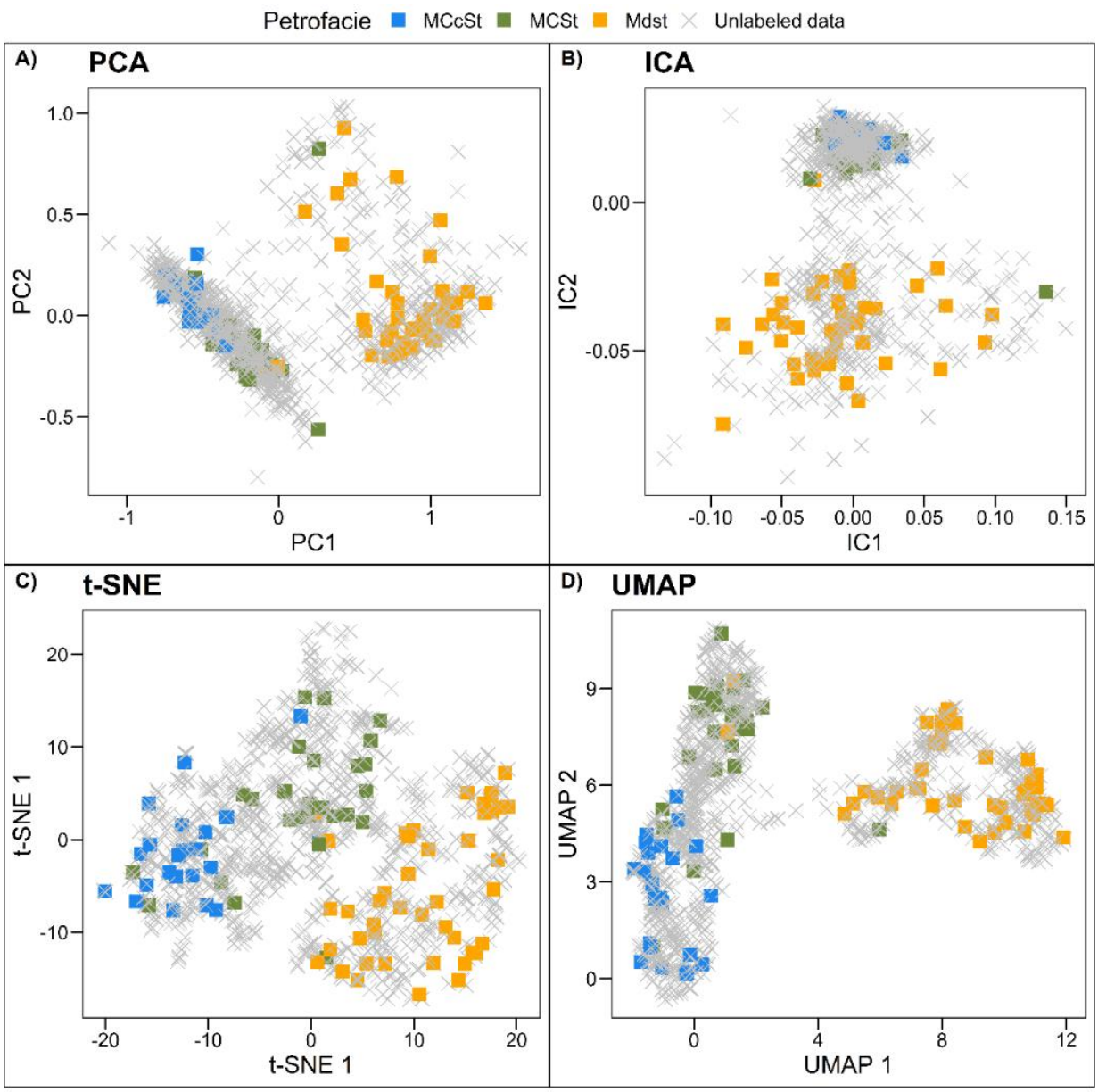

Figure 23: Cross plots of the resulting projections using four dimensionality reduction techniques A) principal component analysis (PCA), B) independent component analysis (ICA), C) t-Distributed Stochastic Neighbor Embedding (t-SNE) and D) Uniform Manifold Approximation and Projection (UMAP). These cross plots show how well the dimensionality reduction techniques group the three thin-section-defined petrofacies (squares) referred in the text as true labels, but also plot the distribution of unlabeled data (x) respect to the projections. Notice the separation in all four plots between the three petrofacies.

PCA and ICA separated data into two main groups; while both non-linear transformations (t-SNE and UMAP) separated MCSt from MCcSt better than linear transformations. Furthermore, UMAP separated siltstones from mudstones better than t-SNE. Also, the data points that were plotted in a group which they do not belong to were closer to their similar data points in the UMAP visualization than when using t-SNE.

Semi-supervised model

Two semi-supervised approaches were executed, self-training and clustering with true labels. In addition to the original 96 thin-section-defined petrofacies, the self-training approach created hundreds of pseudo-labels (**Figure *24***). We evaluated 18 different self-training methods with accuracies between 87% and 93%. For the methods that the self-training model did not classify at least one of each petrofacies, we did not calculate the accuracy (). We found that the higher the probability threshold (e.g., 90%) the more data remained unlabeled after the last iteration. In some cases, for example when using the proxies and PCA results, this led to higher accuracy values, but in other cases, such as when using ICA, not enough data were available to train and test the model accuracy. Also, corroborating the results that the UMAP was the best dimensionality reduction technique for this dataset, the models that used UMAP as the input data outperformed the models that used other dimensionality reduction techniques or the proxies.

Table 2: Self-training results using support vector machine (SVM). 18 cases were compared using different input data and probability thresholds. All the 29 elements, the selected proxies (Al, Ca, Si) and the results of four dimensionality reduction techniques (PCA, ICA, t-SNE, and UMAP) were selected as the input data. Additionally, for each input data, three probability thresholds were evaluated (60%, 80%, and 90%). The unlabeled data are the data points that remained unlabeled because the probability of these points to be correctly predicted was lower

than the probability threshold. Accuracy was calculated after training a model with the pseudo-labels and testing the model with the thin-section-defined petrofacies (true labels). The blank spaces mean that the accuracy was not calculated because the self-training model did not predict at least one sample for each (3) petrofacies.

Input Data	Probability Threshold	Unlabeled Data	Accuracy
All elements	60%	4	0.91
All elements	80%	58	0.91
All elements	90%	143	0.93
Proxies	60%	10	0.87
Proxies	80%	47	0.88
Proxies	90%	147	0.91
PCA	60%	14	0.89
PCA	80%	72	0.89
PCA	90%	195	0.91
ICA	60%	0	----
ICA	80%	0	----
ICA	90%	742	----
t-SNE	60%	5	0.92
t-SNE	80%	147	0.91
t-SNE	90%	763	---
UMAP	60%	14	0.93
UMAP	80%	64	0.93
UMAP	90%	190	0.93

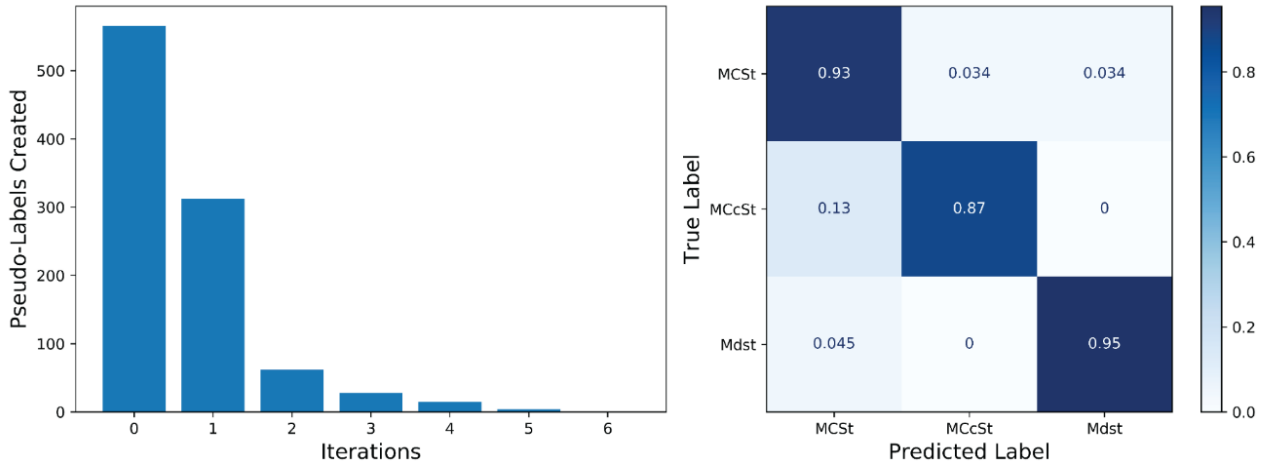

Figure 24: Self-training results using a support vector machine model (gamma=0.01, C=1), UMAP projections as the input data and 80% probability threshold. A) bar plot shows the number of data points that are labeled by each iteration (pseudo-labels). Notice that no pseudo-labels were created in the iteration number 6 because any predicted petrofacies achieved the 80% threshold of being correctly predicted, thus the self-labeling process stopped. B) confusion matrix of the model trained with all the pseudo-labels and tested with the true labels with accuracy of 93%.

In the clustering approach, K-means and HCA over performed the DBSCAN clustering technique (**Figure *25***). Additionally, similar than in self-training methods, K-means and HCA algorithms achieved higher accuracy when using UMAP than when using all the elements or the proxies (Al, Ca, and Si) as input data (**Table *2***). Contrary to K-means and HCA that include all points into one of clusters, DBSCAN considers noise for points that do not belong to a group; however, we interpreted the clusters and noise points to mimic the three petrofacies.

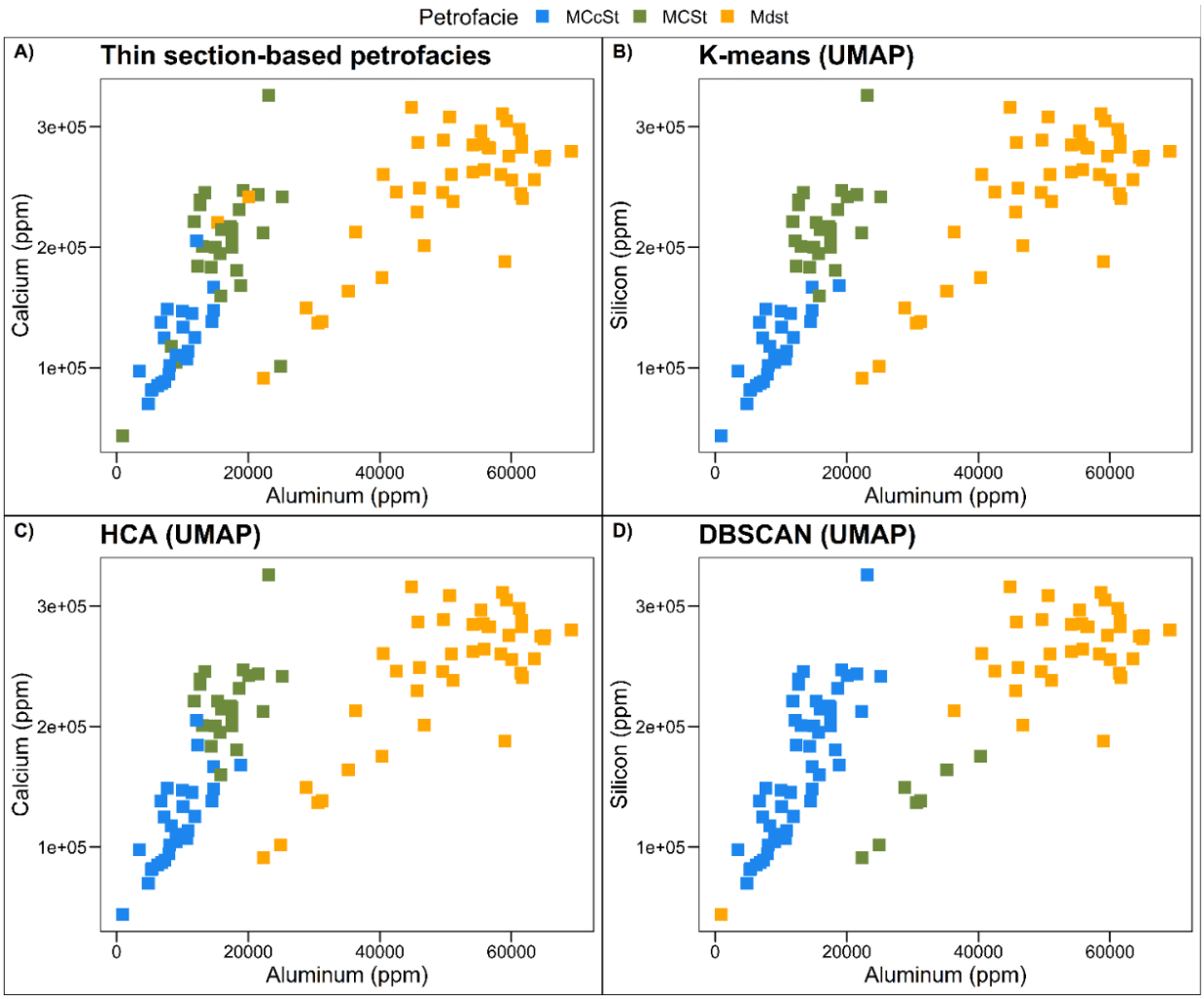

Figure 25: Cross plots of aluminum vs silicon (in parts per millon) colored with the thin-section-defined petrofacies in A) and with the clustering-based-petrofacies in B-C. UMAP was used as input data for the three clustering results depicted in the cross plots. Compare the results from K-means (B), HCA (C), and DBSCAN (D) with the true labels (A) and refer to Table 3 for metrics.

Table 3: Labeled-clustering results. 9 cases were compared using three clustering algorithms (K-means, HCA, and DBSCAN) and three different input data. All the 29 elements, the selected proxies (Al, Ca, Si) and the projections when using UMAP were selected as the input data. Accuracy for the clustering approach was calculated after comparing the interpreted clusters with the true labels.

Clustering Algorithm	Input Data	Accuracy
K-means	All elements	0.91
K-means	Proxies	0.85
K-means	UMAP	0.92
HCA	All elements	0.89
HCA	Proxies	0.79
HCA	UMAP	0.92
DBSCAN	All elements	0.71
DBSCAN	Proxies	0.69
DBSCAN	UMAP	0.67

Because the semi-supervised method that used UMAP and the self-training approach outperformed the models that used other input data and/or the clustering approach, we selected the self-training approach that used UMAP and 80% probability threshold as the semi-supervised method to predict petrofacies for unlabeled samples. This method labeled over 900 points and after the 6^{th} iteration, 64 data points remained unlabeled because the probability of these points to be correctly predicted was lower than 80%. This model was tested with the true labels, and it achieved 93% accuracy (**Table 3**).

3.6 Discussion

XRF data (element logs), thin-section-defined petrofacies, and semi-supervised techniques provided a successful strategy to predict petrofacies at a 0.3-m (1-ft) resolution with an accuracy

>90%. XRF is an inexpensive and nondestructive technique (Alnahwi and Loucks, 2019) that provides elemental information from the exact locations where the thin sections were sampled but also at the well-log resolution. XRF data provide enough information that, combined with dimensionality reduction techniques and semi-supervised methods, are an effective way to generate petrofacies logs for cored intervals that lack thin sections (**Figure *26***).

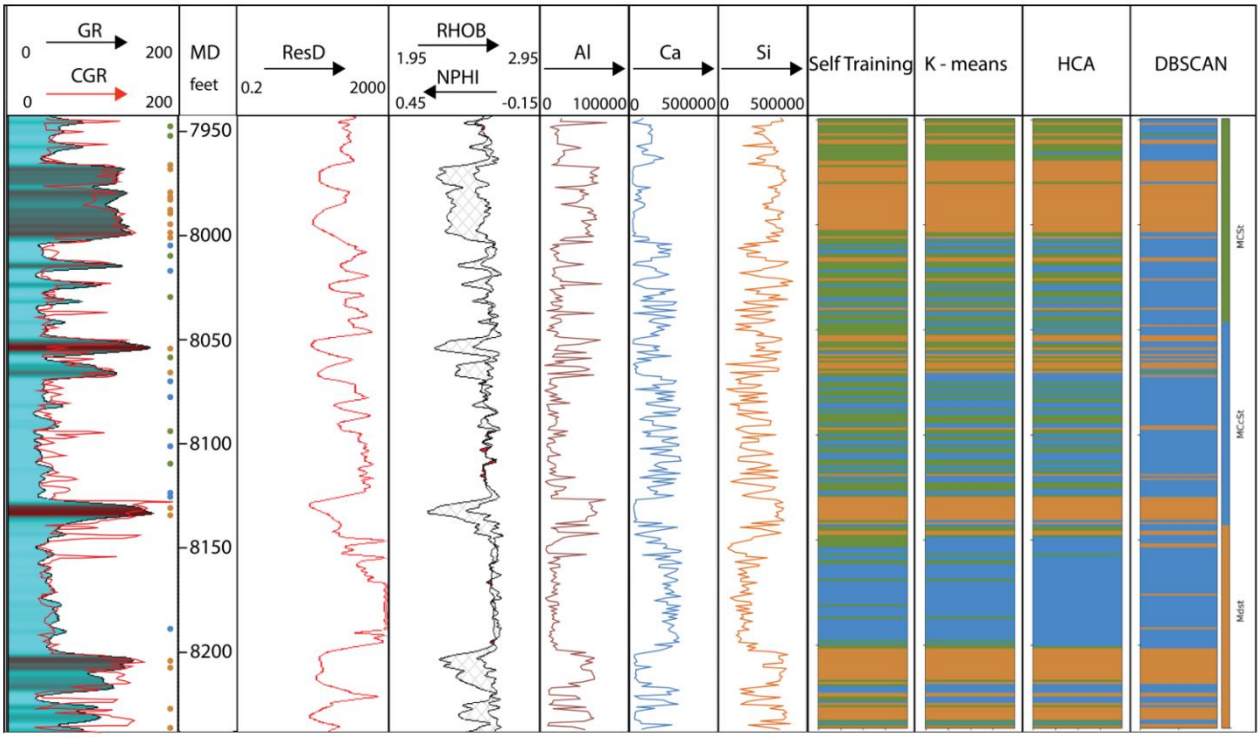

Figure 26: Russel B core profile with synthetic core gamma ray (CGR), well logs (GR, ResD, RHOB and NPHI), location of the thin sections (dots), XRF data (Al, Ca, and Si) and petrofacies logs creates using different semi-supervised methods (including self-training and clustering). The petrofacies log from the self-training approach uses Support vector machine as the classifier, UMAP projections as the input data and an 80% probability threshold. All the petrofacies log from K-means, HCA, and DBSCAN use UMAP projections as the input data. Notice the difference between in GR log and the core GR (CGR). Also, notice the difference between the petrofacies logs when using different semi-supervised methods.

We propose to always use a dimensionality reduction technique with XRF data because it is a good practice when dealing with high-dimensional data, and it avoids the issue of selecting the most important elements (proxies) for each specific dataset. Visualizations help to identify the best dimensionality reduction technique and to determine whether the relationship between the original variables is linear; an important criterion when selecting the best dimensionality reduction technique. **Figure *20***B shows a non-linear relationship between Al and Ca in this dataset which likely explains why PCA and ICA failed to separate the three petrofacies as these are linear transformations. On the other hand, t-SNE and UMAP, two non-linear transformations, more clearly separated the three petrofacies. We selected UMAP because the boundary between mudstones and siltstones was more clearly separated than in t-SNE. Every technique has limitations, for example, t-SNE requires more computation power than UMAP (McInnes et al., 2018) and varying the perplexity parameter leads to completely different results. Or for PCA, one selects the number of components to be used; therefore, important information might be left behind in the other principal components. In ICA, the first independent components might be different to the most important independent components. Therefore, one should consider the limitations of each dimensionality reduction technique at the early stages of the analysis.

We compare two effective semi-supervised approaches to predict petrofacies. We selected the self-training approach because, contrary to the clustering approach, the training is supervised in the initial model, then the model itself generates pseudo-labels based on a given probability of being correctly predicted. Another advantage of this approach is that only data points that fulfill the probability requirements of being predicted correctly were labeled; the residual data remained unlabeled at the end of the self-training process. Thus, we determine the probability threshold to find a balance between having too few pseudo-labels with high probability and many pseudo-

labels with a low probability of being correctly predicted. Like supervised methods, the amount of data used in the training process limits the performance of the self-training approach.

One of the main advantages of clustering methods is that petrofacies labels are not required to define clusters. However, because we used the clustering approach as a semi-supervised approach rather than an entirely unsupervised method, we needed labels to validate the cluster results. Additionally, contrary to self-training when using this approach, the results must be interpreted after clustering the data and this might lead to misinterpretations when dealing with more petrofacies. Like the self-training approach, cluster results were better when using UMAP than when using all the elements or proxies. Also, DBSCAN failed on generating pseudo-labels for this dataset because finding the best value for eps parameter is not intuitive because the two main types of rocks, clay-rich mudstones and massive siltstones, have different density (distance between points in the same petrofacies in space) (**Figure *20*** & **Figure *23***). The depositional setting affects the elemental composition of the rock. The massive siltstones were deposited in a shorter period of time and have a more homogenous composition than the clay-rich mudstones. DBSCAN also failed to generate pseudo-labels because DBSCAN detects outliers that is an advantage when used as an unsupervised method to understand the structure of the data, but it is a disadvantage when used as a semi-supervised method for trying to imitate the thin-section-defined petrofacies.

Creating an unambiguous methodology for clustering data as a semi-supervised method is the main disadvantage for this approach, thus it is necessary to try different methods and select the most feasible and accurate method for a specific dataset. Some authors have used semi-supervised ML techniques to improve well-log facies classification (e.g., Dunham et al., 2020b, 2020a; Xie and Spikes, 2020). However, none of those studies attempted to combine XRF data, dimensionality

reduction techniques, and semi-supervised ML techniques to enhance the resolution of thin-section-based information in cores.

Thin-section-defined petrofacies provide information on the texture, composition, and porosity of the rock, and they are important in characterizing formation reservoir quality. We demonstrate that including semi-supervised ML methods with petrographic and XRF data is an effective approach to predict petrofacies. By classifying petrofacies logs, we can use them for stratigraphic correlation, to identifying target zones, to design horizontal wells, and to constrain subsurface models.

3.7 Conclusions

In this paper, we propose an effective workflow that combines semi-supervised techniques with petrographic and XRF data to predict petrofacies in cored sections. The self-training approach outperformed the clustering semi-supervised approach, achieving accuracies of 93% in the Sycamore Formation. However, using clustering techniques with true labels as a semi-supervised approach achieved accuracies of 92% with K-means and HCA algorithms. Additionally, we obtained better results when using UMAP projections than when using all the elements or the proxies as the input data. Thus, we conclude that it is important to transform high-dimension XRF data to low-dimension projections regardless of the approach or algorithms used. For this dataset, the UMAP dimensionality reduction technique outperformed linear transformations such as PCA and ICA. We demonstrate that the implementation of the workflow we discuss herein has the potential for extracting high-quality information from cores and reducing the resolution gap between scarce core data and well logs.

Chapter 4: Spatial variability of petrofacies using supervised machine learning and geostatistical modeling: Sycamore Formation, Sho-Vel-Tum Field, Oklahoma, USA.

David Duarte, Rafael Pires de Lima, Javier Tellez, and Matthew Pranter

4.1 Preface

This chapter is in preparation to be submitted to the Marine and Petroleum Geology journal. This chapter is the second portion of the project that combines chapters 3 and 4 to validate the conceptual model from chapter 2 by linking the thin-section information with 3D models. After generating petrofacies logs from chapter 3, this 4th chapter presents a workflow that uses machine learning techniques to build a 3D petrofacies model. Some of the highlights for this chapter are:

- This workflow selects the best combination of well logs, classifiers, and hyperparameters over hundreds of models.
- Trained models use well-log data to classify thin-section-derived petrofacies with precision up to 77%.
- Machine-learning techniques are useful to generate petrofacies logs in non-cored wells to use to constrain 3D facies models.

4.2 Abstract

Core data provide unique information about subsurface geology such as lithology, mineralogy, and porosity. Such information is necessary to define petrofacies, which are often used for constraining 3D facies models and to illustrate the stratigraphic variability of reservoir rocks. This

study develops a machine learning-based workflow for assisting 3D models to reliably represent the petrofacies distribution in the Sycamore Formation. The workflow compares over 1,800 classification models and selects the best combination of well logs, algorithms, and hyperparameters to predict petrofacies in wells without cores. For each combination of well logs, we optimized four classification algorithms: Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF). To optimize each classifier, we used Grid-Search and 5-Fold cross-validation to find the best combination of three hyperparameters to tweak each algorithm. Then, out of 244 wells, 75 wells with predicted petrofacies logs were used to construct a 3D data-driven model. The presented workflow obtains more information from cores efficiently and at a low cost. At the same time, the 3D models assisted by the machine learning model aid in the understanding of the diagenetic processes that occurred in the subsurface and affected the reservoir quality.

4.3 Introduction

Core data provide unique information about subsurface geology such as lithology, mineralogy, and porosity. With this information it is possible to define petrofacies, which are often used for constraining 3D facies models to illustrate the stratigraphic variability of reservoir rocks. Despite their importance, cores are scarce because of their high cost. Well logs are more commonly available therefore; various techniques are used to predict core-derived information from well logs. Recently, machine learning (ML) techniques have been used to generate models for predicting core information.

Supervised machine learning techniques have been applied in the geosciences to address many issues related to core data. Supervised ML models have been used to classify lithofacies from well-log information (Dubois et al., 2007; G. Wang et al., 2014; Bhattacharya et al., 2016, 2016; Al-Mudhafar, 2017; Bestagini et al., 2017; Brcković et al., 2017; Bhattacharya and Mishra, 2018; Mandal and Rezaee, 2019; Wood, 2019; Halotel et al., 2019; Bressan et al., 2020; Ippolito et al., 2021). Additionally, some authors have optimized supervised models after finding the best combination of the options, called hyperparameters, that can be chosen to tailor each algorithm for a specific issue (Castro et al., 2017; Dunham, Malcolm, and Kim, 2020; Daviran et al., 2021). Also, some studies have addressed the importance of different well logs on model performance (Singh et al., 2020; Merembayev et al., 2021). In addition, cross-validation techniques have been used to compare and validate supervised models to avoid overfitting (G. Wang et al., 2014; Al-Mudhafar, 2016; Bhattacharya et al., 2016; Bhattacharya and Mishra, 2018; Handhal et al., 2020; Daviran et al., 2021; Pang et al., 2021). However, workflows on building robust supervised ML models to predict facies are inconsistent, and the challenge remains in obtaining unified and unambiguous workflows to select the best model for predicting thin-section-derived petrofacies using well logs.

The lack of well-established workflows may be attributed to the fact ML techniques in geosciences are at an early stage, the lack of standardized studies and standardized datasets, or the interest in testing new algorithms (classifiers) and techniques over consolidating workflows. Therefore, this study aims to develop a workflow, based on ML techniques, to generate rock-type logs to constrain 3D petrofacies models for the Sycamore Formation in the Sho-Vel-Tum Field (Figure 27 & Figure 28). The workflow compares hundreds of classification models and selects the best combination of well logs, algorithms, and hyperparameters for predicting core-derive

petrofacies (from Chapter 3) in wells without cores. This workflow has the versatility of including new algorithms and hyperparameters, and more well logs, providing the user with the ability to compare new ML models with previously published models. By selecting the best ML model out of hundreds of models, one can increase well control and construct robust 3D data-driven petrofacies models.

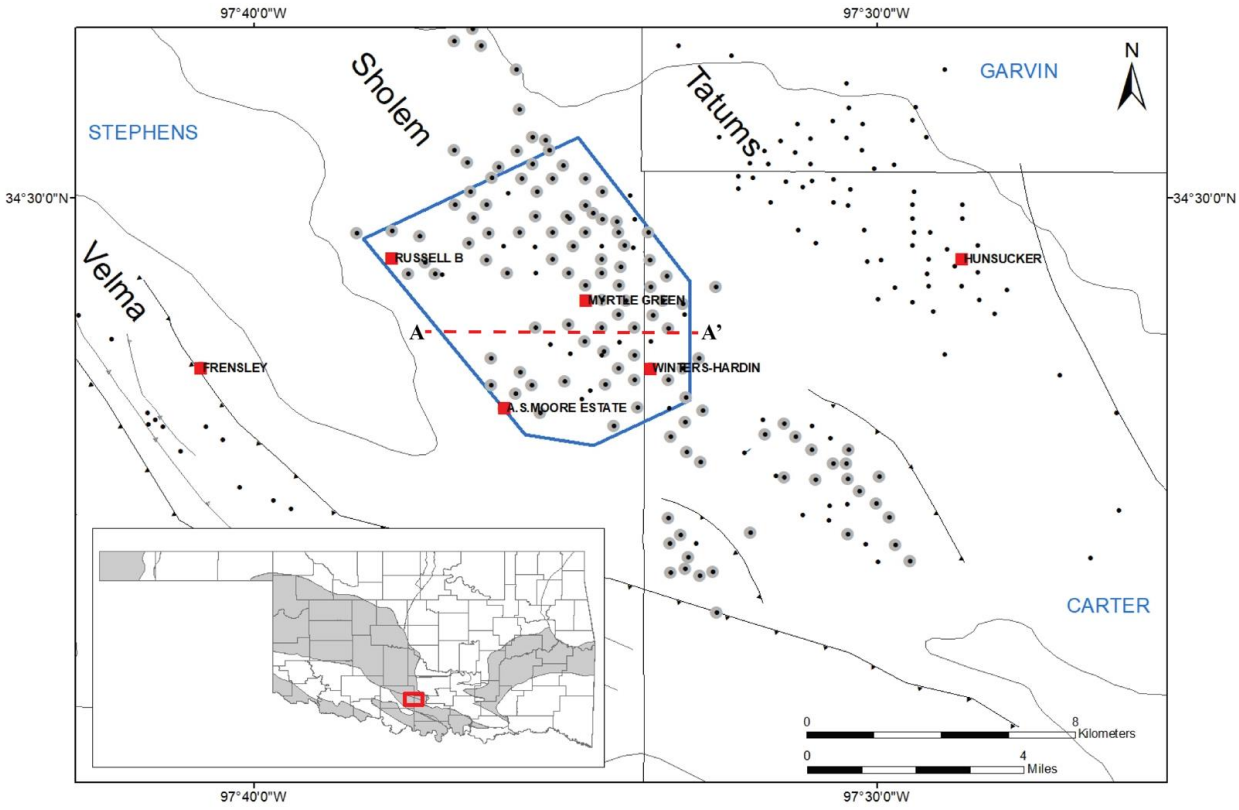

Figure 27: Location map of the study area. Red squares are wells with original/true petrofacies logs in the cored sections, black dots are non-cored wells, and black dots with gray areas correspond to uncored wells with predicted petrofacies logs. Names in black correspond to the three fields that combined forms the Sho-Vel-Tum field, 3D model boundary in blue. Faults in black are from (Carpenter and Tapp, 2014) and names in blue correspond to counties.

4.4 Data & methods

Six cored wells with thin section-derived petrofacies (petrofacies logs from Chapter 3) and additionally, 244 wells without cores in the Sho-Vel-Tum field were used for this study (Figure 27). All wells contain gamma-ray (GR), density (RHOB), and deep resistivity (RESD) logs. Seventy-five of the wells were used to construct the 3D model for the Sycamore Formation (Figure 28). Petrofacies logs is a term used to denote thin-section-derived petrofacies at 1-ft (x-m) resolution (from Chapter 3).

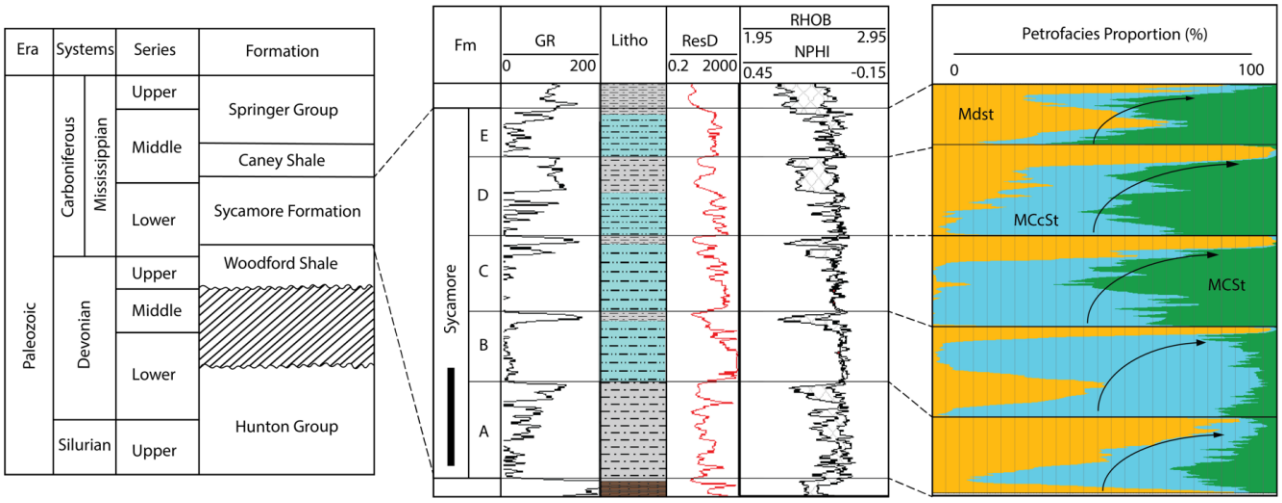

Figure 28: General stratigraphy of the study area on the left, Russell B well in the middle with the five finning-upward parasequences (A-E) from (Duarte et al., 2021), and vertical proportion curve (VPC) for the Sho-Vel-Tum field on the right. VPC shows the petrofacies proportion by zone/parasequence.

Figure 29 shows the general workflow with four main steps implemented in this study: i) first, we selected the best combination of well logs, algorithms (classifiers), and hyperparameters, ii) we built the classification model using the selected model, iii) we used the selected classification model for predicting petrofacies logs in uncored wells, and iv) we built a 3D model to evaluate the

petrofacies distribution in a portion of the Sho-Vel-Tum field. The following section explains the methods for the supervised classification model and the methods used to generate the 3D petrofacies model.

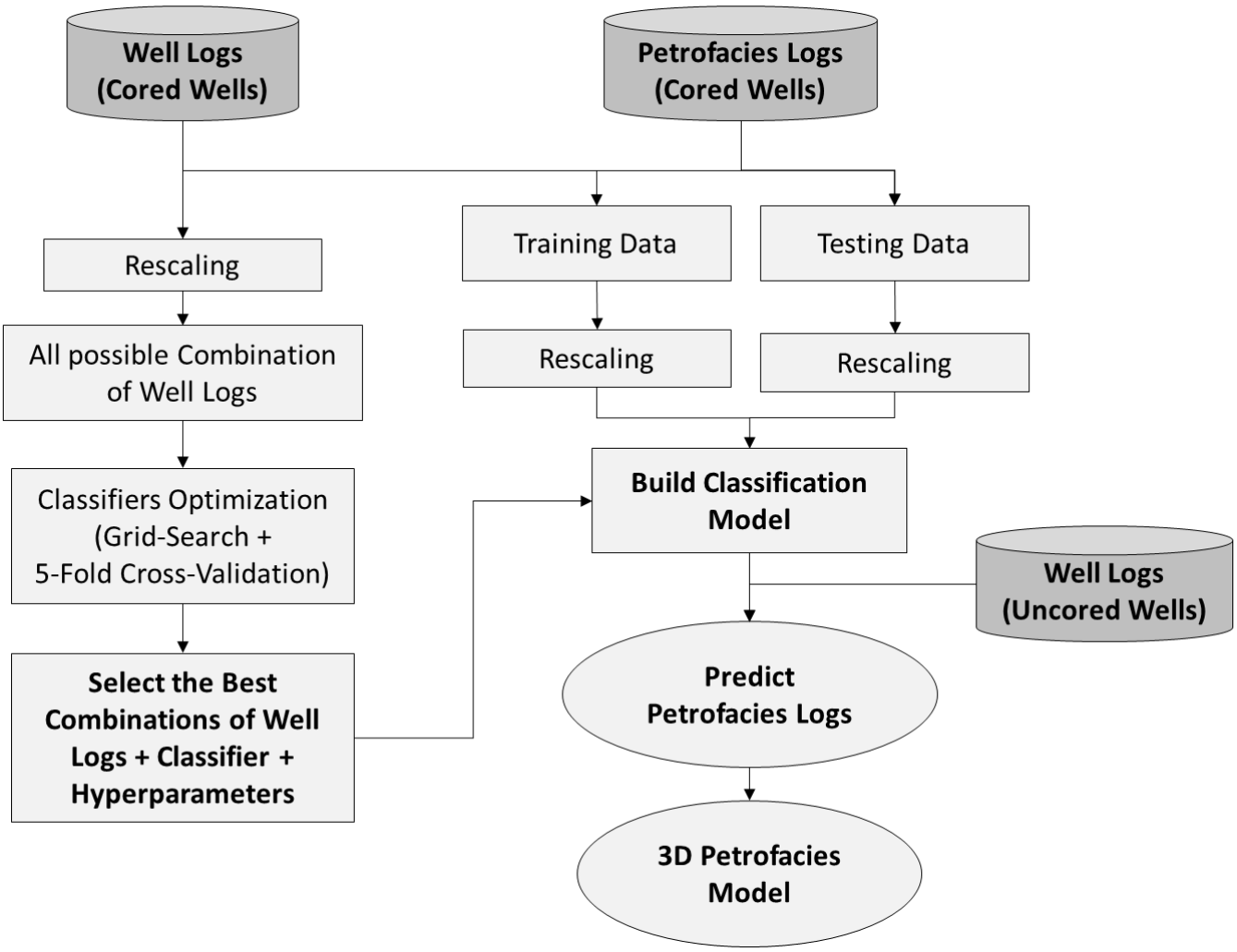

Figure 29: Generalized workflow that uses predicted petrofacies logs to build 3D petrofacies model. Cylinders in dark gray and ovals represent the input and output data respectively. First, we compared thousands of models to define the best combination of well logs, classifier, and hyperparameter using Grid-Search and k-fold cross-validation. Then, we used this information to create the final classification model that uses well logs to predict the petrofacies logs in uncored wells. Third, we used the model to predict petrofacies logs in uncored wells, and finally the predicted petrofacies logs were used to build the 3D petrofacies model.

Compare and select the best classification model for predicting petrofacies using well logs

We used Grid-Search (LaValle et al., 2004; Bergstra et al., 2012; W. Wang et al., 2014) and cross-validation (Lachenbruch and Mickey, 1968; Stone, 1974, 1977) to optimize and select the best classification model for predicting petrofacies in non-cored wells using well logs. However, first, we normalized GR, RHOB, and RESD logs using the min-max scaling method. This prevents one well log from having more importance than the other logs only based on its range. This method rescales all well logs in the range of [0,1]. Contrary to Z-score normalization, the min-max scaler preserves the natural distribution of data.

Model optimization

To improve the performance of the model, we tested over 1,800 models with different combinations of well logs, algorithms (classifiers), and hyperparameters. First, we identified all seven possible combinations of well logs using GR, RHOB, and RESD (e.g. GR + RHOB; GR + RHOB + RESD). Then, for each combination of well logs, we tested four classifiers using Scikit-learning libraries (Pedregosa et al., 2011): 1) Artificial Neural Network (ANN) (McCulloch and Pitts, 1943; Ivakhnenko and Lapa, 1968; Shannon and McCarthy, 2016), 2) K-Nearest Neighbor (KNN) (Cover and Hart, 1967), 3) Support Vector Machine (SVM) (Boser et al., 1992), and 4) Random Forest (RF) (Ho, 1995; Breiman, 2001). For optimizing each classifier, we used Grid-Search to find the best combination of hyperparameters that can be chosen to tweak each model (**Table *4***), this process is called hyperparameters tuning.

Table 4: Hyperparameter values used in the Grid-Search to optimize the classification model. We used four classifiers: Artificial Neural Network, Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest

Classifier	Hyperparameter 1	Hyperparameter 2	Hyperparameter 3	No Models
ANN	Hidden layer sizes: (50,50,50), (50,100,50), (100,)	Learning Rate: (Constant, Adaptive)	Solver: (sgd, adam)	168
KNN	Number of neighbors: (3,4,5,6,7,8,9,10,11,12,13,14,15)	p: (1,2)	Leaf size: (1,2,3,5)	1820
SVM	C: (0.1, 1, 10,50, 100)	Gamma: (0.001, 0.01, 0.1, 1, 10)	Kernel: (rbf, linear)	700
RF	Number of estimators: (1,5,10,50,100)	Minimum samples leaf: (1,2,3)	Maximum depth: (3,4,5,6,10)	1050

We selected three hyperparameters to optimize each of the four classifiers (**Table *4***), but more hyperparameters and corresponding options could be added as needed. ANN uses neurons organized into layers to predict the petrofacies. ANN has an input layer that receives the input data, an output layer to predict the petrofacies, and the hidden layer(s) in between the input and output layers. We varied the hidden layer size that is the number of neurons for each layer, the solver to specify the algorithm for weight optimization, and the learning rate (constant vs adaptive). The second algorithm, KNN, is a non-parametric model. KNN uses a K-number of nearest neighbors to predict the petrofacies. Similar to other ML algorithms, KNN uses distance metrics such as Euclidean distance to measure the distance between the point of interest and its nearest neighbors. For KNN, we tuned the number of neighbors the algorithm uses as queries, the p parameter to decide between using Manhattan or Euclidean distances, and the leaf size that affects the construction and query time. The third algorithm, SVM (Boser et al., 1992), uses a hyperplane in an n-dimensional space, given by the n-number of independent variables, to separate the data into the petrofacies. The closest points to the hyperplane are the support vectors and they

belong to different classes. We optimized the C parameter that penalizes misclassified points, the gamma parameter to determine the distance of influence of a training point, and the kernel type that could use radial- or linear-based functions. Finally, RF classifier assembles decision trees to predict the petrofacies (Ho, 1995; Breiman, 2001). We optimized the number of estimators which is the number of trees, the minimum samples leaf that is the minimum number of samples required to be at a leaf node, and the maximum depth of the three.

Cross-validation

Additionally to the Grid-Search method for evaluating different hyperparameters for each model, we included cross-validation (CV) to avoid overfitting and to better compare the models (Stone, 1977; Kohavi, 1995). With this method, we tested every model on multiple train-validation subsets to avoid differences on the model performance based on the way we divided the data in training and validating sets. We used a method called K-Fold cross-validation where the data is partitioned in K subsets. Then, one of subsets is used as the validating set and the remaining data ($K-1$) as the training set. We used K=5, thus for every combination of well logs, classifier, and hyperparameters we repeated the process five times using a different, randomly selected, validating subset every time. In other words, the first time, we randomly select 85% of the data as the training set and the remaining 20% as validating set, the second time we used another 20% as the new validating set, and we repeated this process five times until all the data were once, and only once, used for validating. Finally, the average precision and standard deviation of the five iterations were used to compare and select the best model. Precision was used over accuracy to compare the models because the petrofacies have different number of samples (Figure 30). This is called imbalanced

data. However, other metrics, such as recall and F1, could be easily added to the workflow. In fact, we used precision and recall metrics, but only precision is reported in the results of this paper.

Construct classification model and predict petrofacies logs

After we used Grid-Search and 5-Fold cross-validation to compare over 1,800 models and select the best combination of classifier, hyperparameters and well logs, we trained the selected model. For this model, we used 85% of the data for training and 15% of the data for validating the model. In this case, only 15% of the data was used for validation because we already validated the model using 5-Fold cross validation. Therefore, this step was used to corroborate the performance of the selected model. This selected model was used to predict petrofacies at well log-resolution when cores were unavailable. We predicted the petrofacies logs for 75 wells within the model boundary (Figure 27).

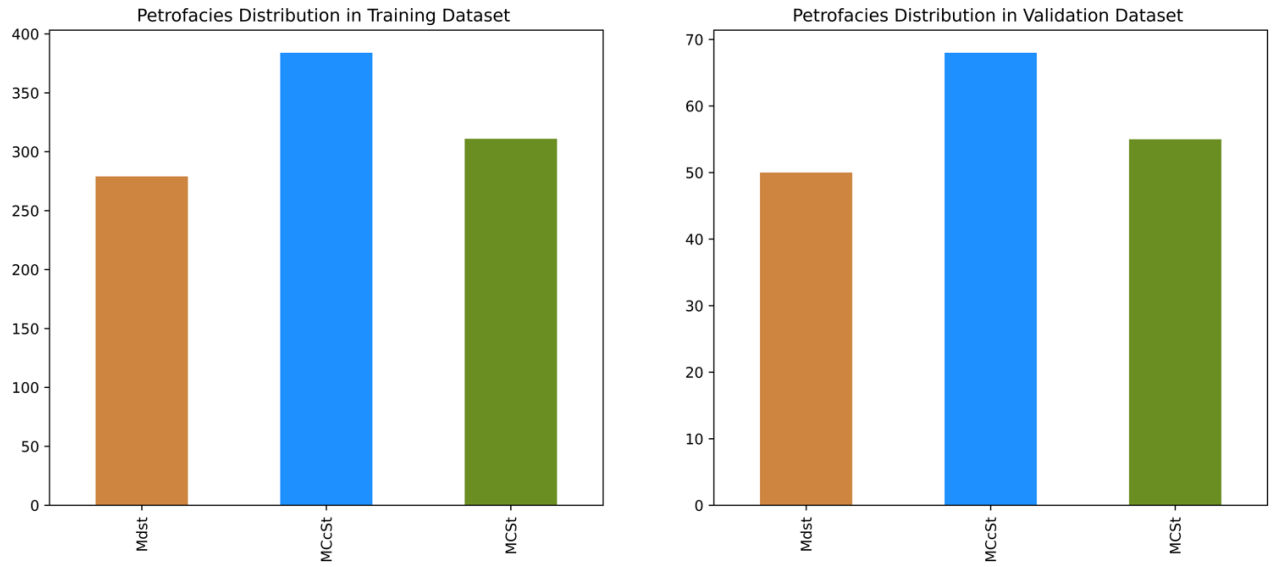

Figure 30: Petrofacies distribution before and after splitting for the selected model. The training dataset correspond to the 80% and the validation dataset to the remaining 20% of the entire data. Note the three petrofacies have different number of samples (x-axis), but we maintained the proportions equal in both sub-datasets.

3D petrofacies model

The 3D petrofacies model was created on a portion of the Sho-Vel-Tum Field (Figure 1) to understand the petrofacies distribution in the Sycamore Formation. This model was constructed with sequential-indicator simulation (SIS), the predicted petrofacies logs, and a 3D grid with areal cell dimensions of 100 × 100 ft (30 × 30 m) and an approximate layer thickness of 1 ft (0.3 m). The following constraints were used for the SIS model: i) the stratigraphic framework with five zones/parasequences (A-E) built from all 244 wells (i.e., the 3D grid with proportional layering scheme). ii) 75 upscaled petrofacies logs that were created from assigning the petrofacies that is most represented in the log to a cell from the 3D grid. iii) vertical proportion curves for the petrofacies in the Sho-Vel-Tum field, iv) petrofacies percentages by zone obtained from the upscaled logs, and v) vertical and horizontal variograms. The variograms ranges were estimated from variography analysis of the petrofacies logs (appendix).

4.5 Results

Three petrofacies were described in the Sycamore Formation: Clay-rich mudstones (Mdst) characterized by low porosity and CaO content, and a broad range of permeability (Figure 31). The variation in permeability is likely attributed to the effect of the bioturbation (Duarte et al., 2021); however, the acquisition of more data is needed for this evaluation. Massive Calcite-cemented Siltstones (MCcSt) are characterized by high CaO, and low porosity and permeability. And Massive Calcareous Siltstones (MCSt) are characterized by higher porosity and permeability than its counterpart, MCcSt.

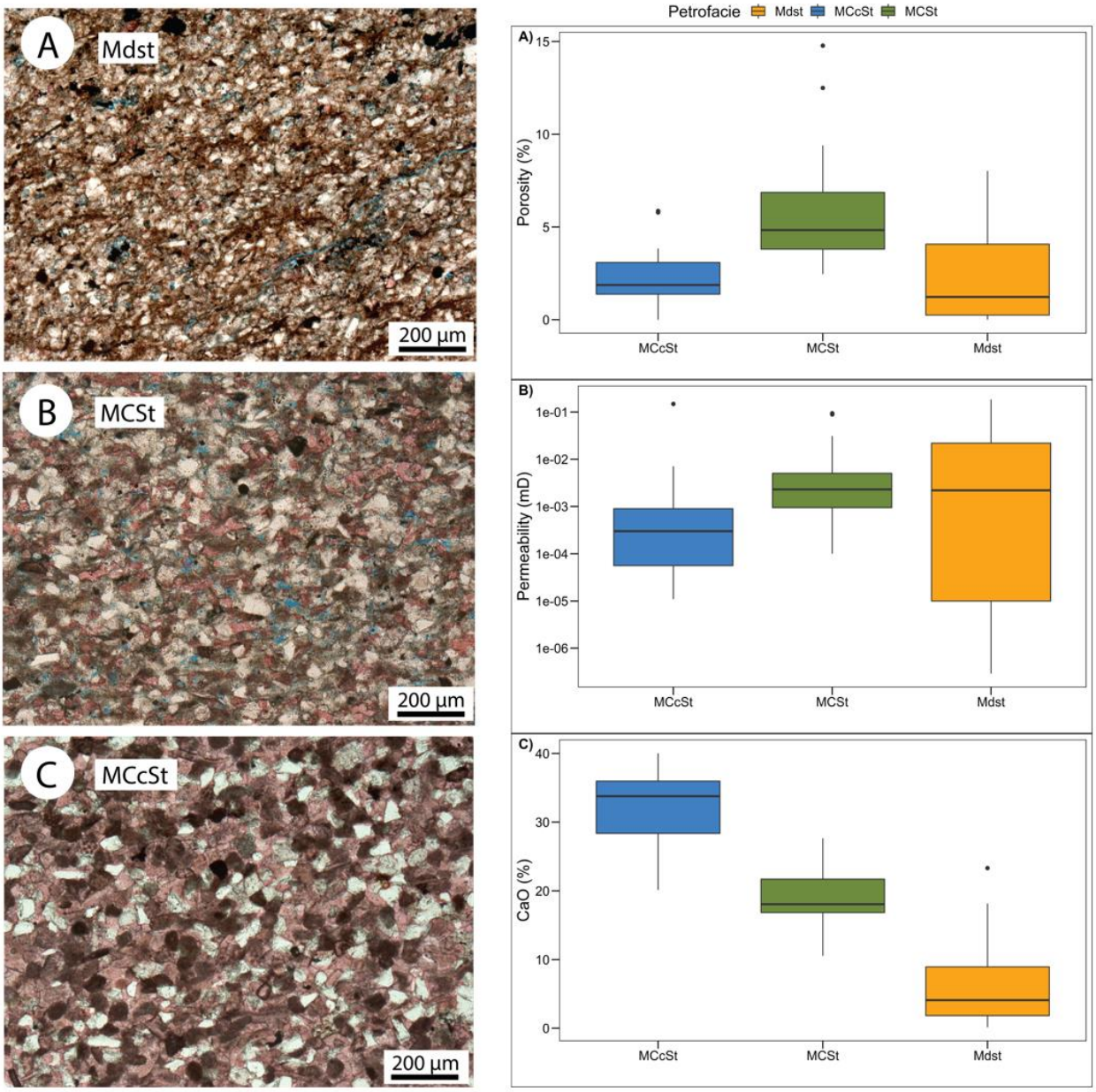

Figure 31: Petrofacies photomicrographs for the three thin-section-defined petrofacies A) Mudstones (Mdst), B) Massive calcareous siltstones (MCSt), and C) Massive calcite-cemented siltstones (MCcSt). The boxplots depict the porosity, permeability, and CaO distribution for each petrofacies. The clay-rich mudstones (Mdst) are characterized by a broad range of porosity and permeability. MCSt and MCcSt are massive siltstones that are differentiated by the amount of calcite cement (Duarte et al., 2021). MCSt is more porous and permeable but has less CaO than the MCcSt.

Classification model

First, we identify seven possible combinations of well logs using GR, RHOB, and RESD. Then, four classifiers were compared and their hyperparameters were optimized to predict the petrofacies. In total, 1,869 models (Figure 32) were created and distributed as following: 84 ANN models, 910 KNN models, 350 SVM models, and 525 RF models. Training all the models takes less than 30 minutes on a computer with CPU E5-2620 v4 @ 2.10GHz. Out of the first 100 models with the highest mean precision, only three models did not use all three well logs. Commonly, the models with the lowest mean precision used only one log. Models that only used GR performed better (up to 70% mean precision) than using RESD (up to 52%) or RHOB (up to 47%) logs. **Table 5** shows the top-three models for each classifier, its hyperparameters, and the selected well logs. RF and SVM achieved higher precision scores than ANN, and KNN. Only ten RF and three SVM models achieved mean precision higher than 75%. Furthermore, all well logs (GR, RHOB, and RESD) were used in all top-3 models. However, Figure 32 shows a narrower distribution and overall better performance for RF and KNN models than ANN and SVM models.

Table 5: Top-3 results for every classifier after using Grid-Search and cross-validation. We used Grid-Search to evaluate the best combination of well logs (GR, RESD, and RHOB), classifier (ANN, KNN, SVM, and RF), and hyperparameters. After evaluating over 1,800 models, these are mean precisions and standard deviation (std) of the top-3 results for every classifier. The mean precision came from 5-fold cross-validation to better evaluate and compare the models.

Classifier	Input (Well Logs)	Hyperparameters 1	Hyperparameters 2	Hyperparameters 3	Mean	std
ANN	(GR RESD RHOB)	hidden_layer_sizes: (50, 50, 50)	learning_rate: constant	solver: adam	0.74	0.06
	(GR RESD RHOB)	hidden_layer_sizes: (50, 50, 50)	learning_rate:adaptive	solver: adam	0.74	0.06
	(GR RESD RHOB)	hidden_layer_sizes: (100,)	learning_rate: constant	solver: adam	0.73	0.05
KNN	(GR RESD RHOB)	n_neighbors: 14	p: 2	leaf_size: 1	0.75	0.05
	(GR RESD RHOB)	n_neighbors: 14	p: 2	leaf_size: 3	0.75	0.05
	(GR RESD RHOB)	n_neighbors: 14	p: 2	leaf_size: 30	0.75	0.05
SVM	(GR RESD RHOB)	C: 50	gamma: 10	kernel: rbf	0.77	0.06
	(GR RESD RHOB)	C: 100	gamma: 10	kernel: rbf	0.76	0.06
	(GR RESD RHOB)	C: 10	gamma: 10	kernel: rbf	0.76	0.05
RF	(GR RESD RHOB)	No Estimators: 50	Max Depth: 5	Min Samples Leaf: 2	0.77	0.06
	(GR RESD RHOB)	No Estimators: 100	Max Depth:5	Min Samples Leaf:3	0.76	0.06
	(GR RESD RHOB)	No Estimators: 50	Max Depth:5	Min Samples Leaf: 3	0.76	0.07

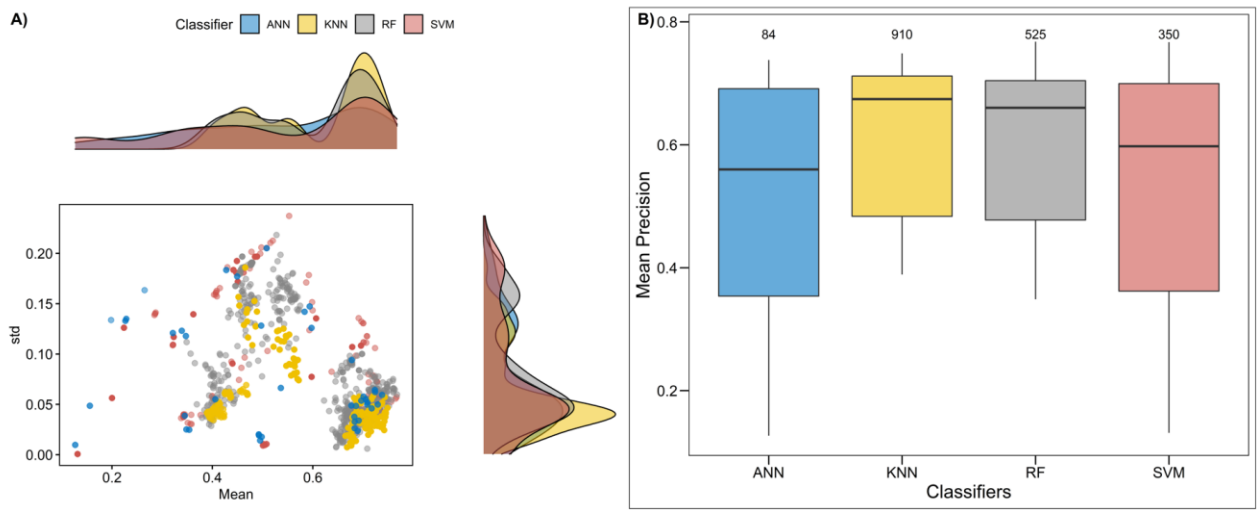

Figure 32: Cross plot between mean precision and standard deviation for all 1,869 models after cross-validation. Models achieved precision as low as 12% and as high as 77%. Thus, evaluating thousands of models is important to select the best model. B) Box plots show the variability in the mean precision results from every classifier. Note the number of models used for the graphs is indicated on top of the box plots.

The cross-validation results show that the best model obtained a mean precision of 77% using all well logs, RF classifier and the following hyperparameters: max depth of the tree=5, min samples required to be a leaf node=2, and number of estimators=50. Figure 33 shows the mean precision of all RF models and the best model that was selected to predict the petrofacies in uncored wells. However, the similarity of well-log response between the MCSt and the MCcSt blurs the differentiation of these petrofacies. The selected model increased the well control from six cored wells to around 120 wells with petrofacies logs in the Sho-Vel-Tum field.

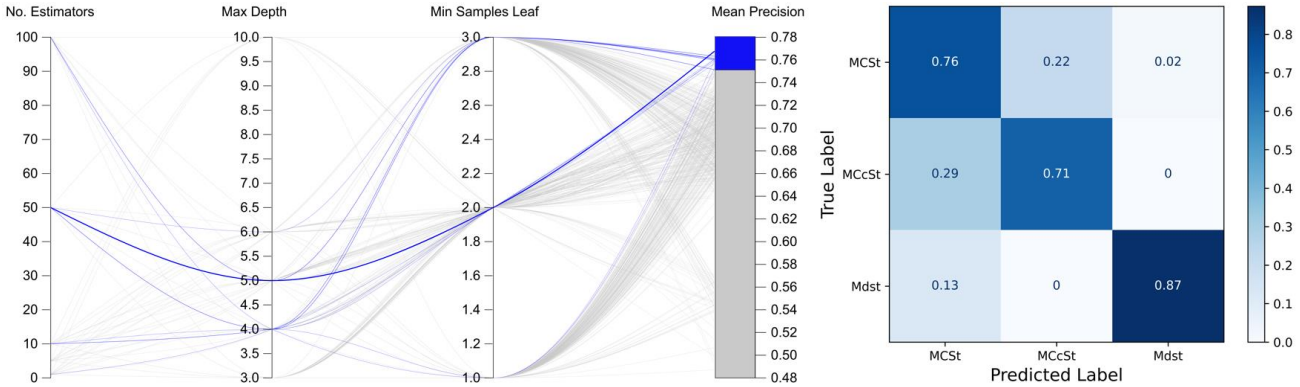

Figure 33: Comparison of all models generated in the hyperparameter tuning process of the Random Forest (RF) classifier. Left: parallel coordinates plot. The first three columns are the tuned hyperparameters from **Table 4** and the last column represents the mean precision of the model using 5-k fold cross-validation. The models with mean precision below 75% are colored in gray and the models with precision equal to or higher than 75% are colored in blue, additionally, the thick blue line represents the selected model based on the mean precision (77%) and standard deviation (0.06). Note most of the models with Max Depth equal to 10 have low precision. On the right, the confusion matrix shows the result of the selected classification model. Note the selected model struggle in differentiating between MCSt and MCcSt.

3D Petrofacies Model

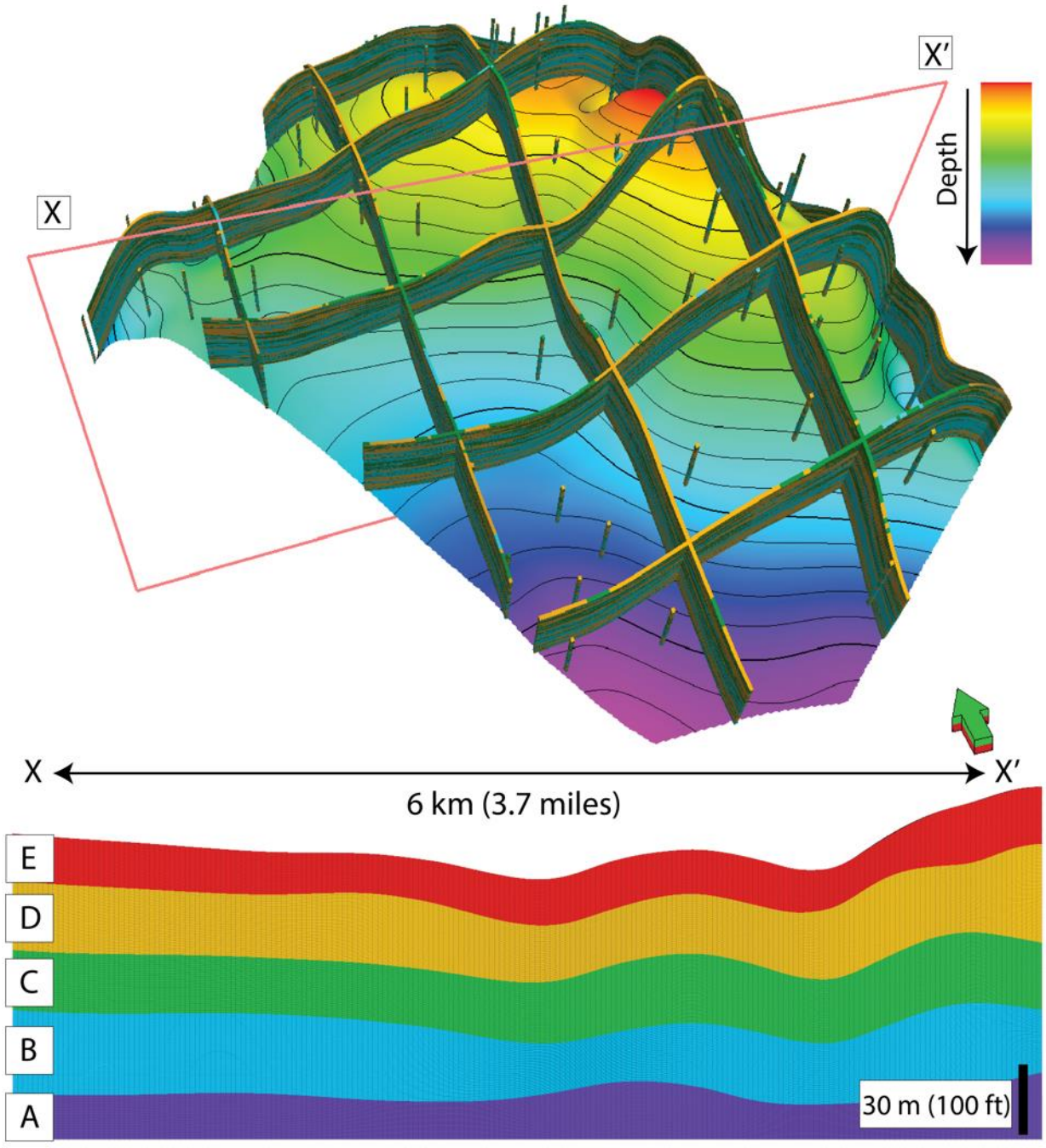

Figure 34: 3D data-driven model. On top, north-south, and east-west slices of the 3D petrofacies model, petrofacies logs, and top of the Woodford Shale horizon in the Sho-Vel-Tum field. On the bottom, cross-section from X to X' with the five zones that represent the finning upward parasequences in the Sycamore Formation.

The constructed 3D model has over 21 million cells and used 75 petrofacies logs (Figure 34) predicted from the previous classification model. The model maintains similar petrofacies percentages (Figure 35) as those of the original petrofacies logs and upscaled petrofacies logs. Also, Figure 36 shows the petrofacies in the cored wells at different steps: from the cores, the predicted petrofacies logs, and finally from the upscaled petrofacies logs. These three steps of the petrofacies logs show similarity between them; however, the upscaled petrofacies logs and the predicted petrofacies maintain similar vertical resolution (1-ft; x m). The vertical proportion curve (Figure 28) and the 3D petrofacies model show an increase of MCSt and a decrease of MCcSt from parasequence B to E. On top of that, Mdst are abundant in parasequence A, but also at the top of each parasequence.

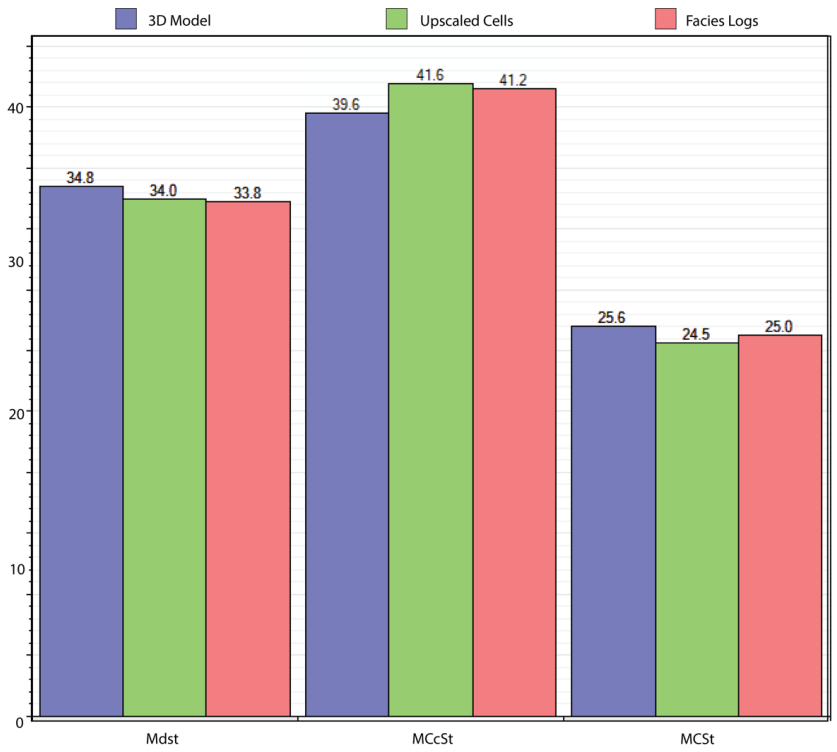

Figure 35: Histograms of the petrofacies proportions for the original petrofacies logs, upscaled petrofacies logs, and 3D petrofacies model

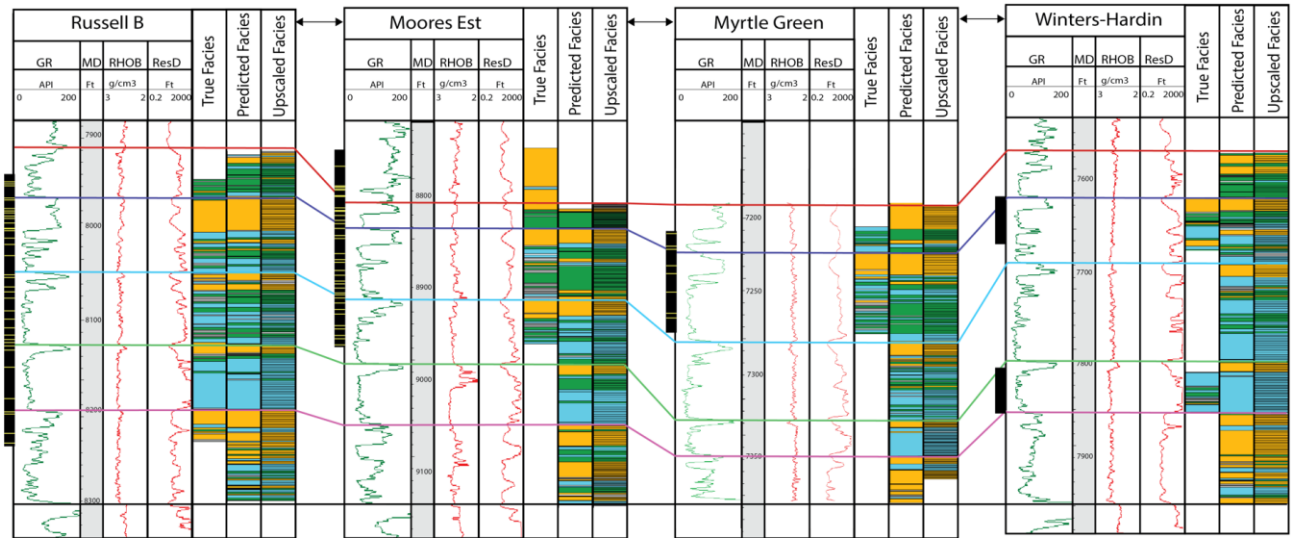

Figure 36: Cross-section from west to east with four wells with cores in the Sho-Vel-Tum field. The black rectangles next to the well logs (GR, RHOB, and RESD) represent the cored section with true/original petrofacies logs. True facies are the original petrofacies logs from (Chapter 3), the predicted facies are the petrofacies logs predicted using the selected classification model, and upscaled facies are the upscaled petrofacies logs that were used to constrain the 3D petrofacies model.

4.6 Discussion

We found that 3D facies models can be assisted by the proposed machine learning workflow to understand the petrofacies distribution of the Sycamore Formation in the Sho-Vel-Tum field. We used a supervised classification model with 77% precision that increased well control from 6 cored wells to 75 wells with predicted petrofacies logs. We also found that combining different classifiers, hyperparameters, and well logs may result in models with completely different performances.

Classification model

We found that by using Grid-Search and 5-Fold cross-validation for evaluating and comparing hundreds of models we are secure to select the best model, honoring major trends and avoiding overfitting. With over 1,800 trained models, selecting the best model is crucial, especially when precision varies from values as low as 13% and as high as 77%. In general, the models performed better when using all three well logs and selecting the best hyperparameters for each classifier made the difference between a useful model and a model that '*guesses'* the petrofacies.

5-fold cross-validation provides good information when selecting the best model over hundreds of models. However, our results show models with precision results lower than 80%. These low performances could be associated to differences between core data and well-log information. Correlating core depths with log data depth is still a limitation due to logging and coring techniques. Thus, we need better techniques to correct the depth shift between core data and well logs to achieve higher precision models. Also, including other well logs, such as neutron porosity (NPHI) and photoelectric factor (PE) may improve the model performances. Additionally, the wells used for this study were drilled in the 1970s, and service companies have improved and standardized the logging processes since then.

Previous studies have successfully used classification algorithms for predicting lithofacies from well logs (Dubois et al., 2007; G. Wang et al., 2014; Bhattacharya et al., 2016; Brcković et al., 2017; Bhattacharya and Mishra, 2018; Halotel et al., 2019; Mandal and Rezaee, 2019; Wood, 2019; Ippolito et al., 2021). Recently, in geosciences a significant effort on building classification models have been focused on using RF and SVM algorithms. However, using this workflow other algorithms such as, Probabilistic Neural Networks (Al-Mudhafar, 2017), Bayesian Network Theory (Bhattacharya and Mishra, 2018) and Gradient Boosting (Bestagini et al., 2017), can easily

be added to the workflow and compared with the models we evaluated. Some of these studies have addressed the superiority of RF and SVM models over other classifiers, but our data show that even though these classifiers achieved the higher precisions, in general the average cross-validation precision is higher for KNN than SVM (Figure 32***Figure 31***). On top of that, SVM and ANN results are more broadly distributed compared with the KNN and RF models. This indicates selecting the best hyperparameters could be as important as selecting the algorithm.

Some studies have compared different algorithms to select the best classifiers. Nonetheless, as previously observed, the performance of models is highly sensitive to the selected hyperparameters. That is the reason why some authors have optimized classification algorithms by tuning the hyperparameters (G. Wang et al., 2014; Bhattacharya and Mishra, 2018; Halotel et al., 2019; Mandal and Rezaee, 2019; Bressan et al., 2020; Daviran et al., 2021; Ippolito et al., 2021). Additionally, some authors have addressed the importance of well logs for lithofacies prediction (Singh et al., 2020; Ippolito et al., 2021; Merembayev et al., 2021), but we are not aware of other publications using well-log data to classify facies defined at the thin-section scale.

Contrary to previous studies, our workflow finds the best classifier (ANN, KNN, SVM, and RF), input data (well logs), and hyperparameters at the same time, this allows us to evaluate and select the best classification model for a specific dataset. Using this workflow, over 1,800 models were evaluated and compared, making the final model a robust one.

3D Petrofacies Model

We review a ML-based workflow to predict petrofacies logs. The predicted petrofacies logs in non-cored wells are used to increase well control and constrain 3D petrofacies models. We found

that the more detailed 3D data-driven petrofacies model is consistent with the conceptual model developed from Duarte et al. (2021) (Figure 37) that illustrates the distribution of three facies: well-cemented siltstones, massive siltstones, and clay-rich mudstones. These facies correspond with the three petrofacies described in this study: Massive Calcite-cemented Siltstones (MCcSt), Massive Calcareous Siltstones (MCSt), and Clay-rich Mudstones (Mdst), respectively (Figure 31).

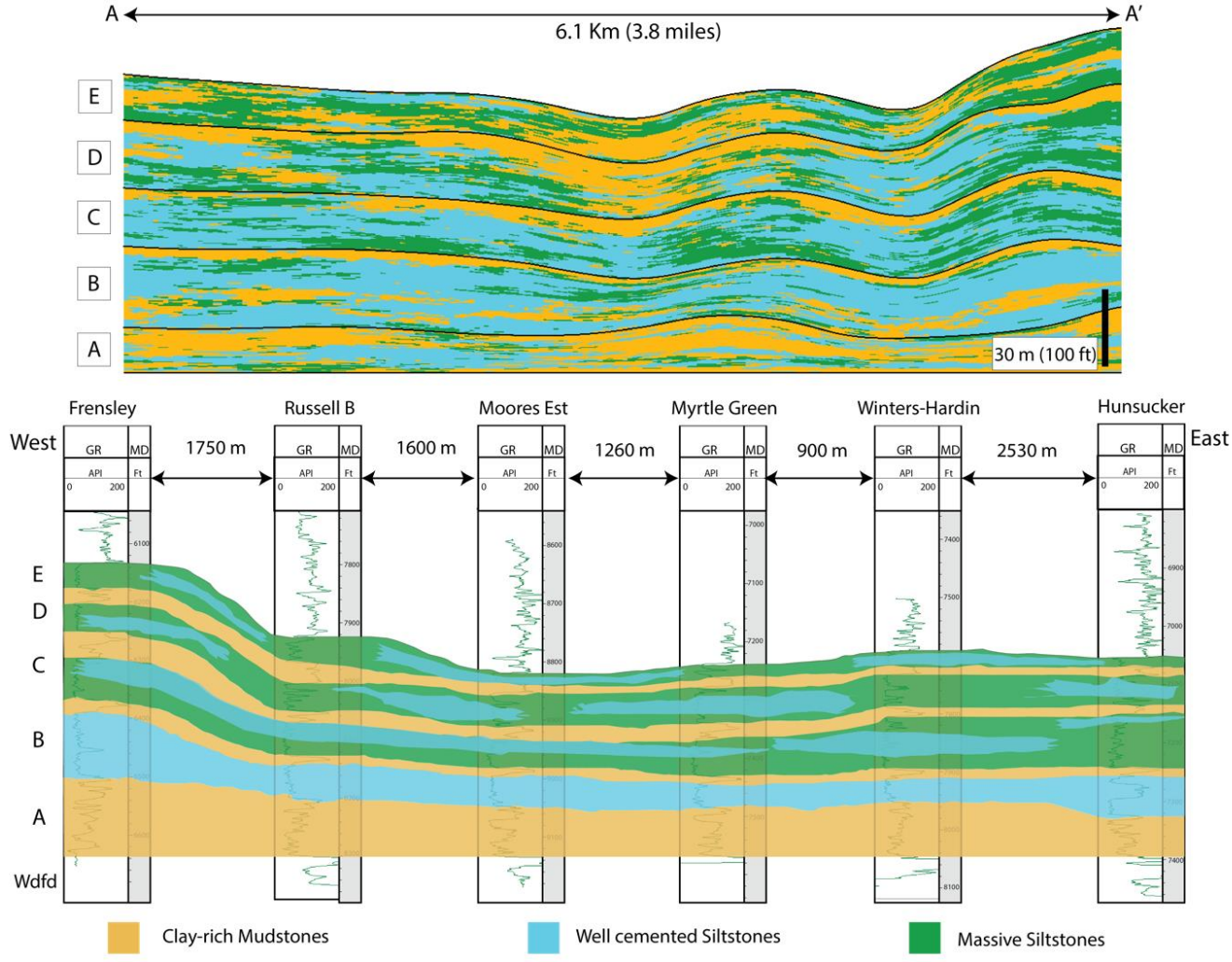

Figure 37: 3D data-driven model and conceptual model.. On top, a 3D model cross-section from A to A' in the location map. On the bottom, cross-section with the 6 cored wells and the conceptual model modified from (Duarte et al., 2021). Note the 3D model cross-section is consistent with the conceptual model in regard to petrofacies distribution. Parasequence A is a transition zone from

the Woodford Shale to the Sycamore Formation. Parasequence B is the most cemented interval in the Sycamore formation depicted by abundant MCcSt (blue), and the base of parasequence C shows the highest porosity values depicted by the presence of MCSt petrofacies. In general, the Sycamore formation transition from a carbonate-dominated system at the base to a siliciclastic-dominated system at the top.

The 3D model shows the distribution of three petrofacies for the five parasequences of the Sycamore Formation. Similar to the conceptual model of Duarte et al. (2021), Massive Calcite-cemented Siltstones (MCcSt) are more abundant in parasequence B (Figure 38), therefore, this sequence is characterized by a high abundance of calcite cement and low porosity. On the other hand, Massive Calcareous siltstones (MCSt) are more abundant at the base of parasequence C, where the highest porosity values have been reported. In general, both models show parasequence A is enriched in Clay-rich mudstones (Mdst). Parasequence B represents a carbonate-dominated environment with higher proportions of MCcSt petrofacies over MCSt. And parasequence C to E show an increase in a siliciclastic-dominated environment characterized by an increase in MCSt and Mdst petrofacies toward the top of the Sycamore Formation.

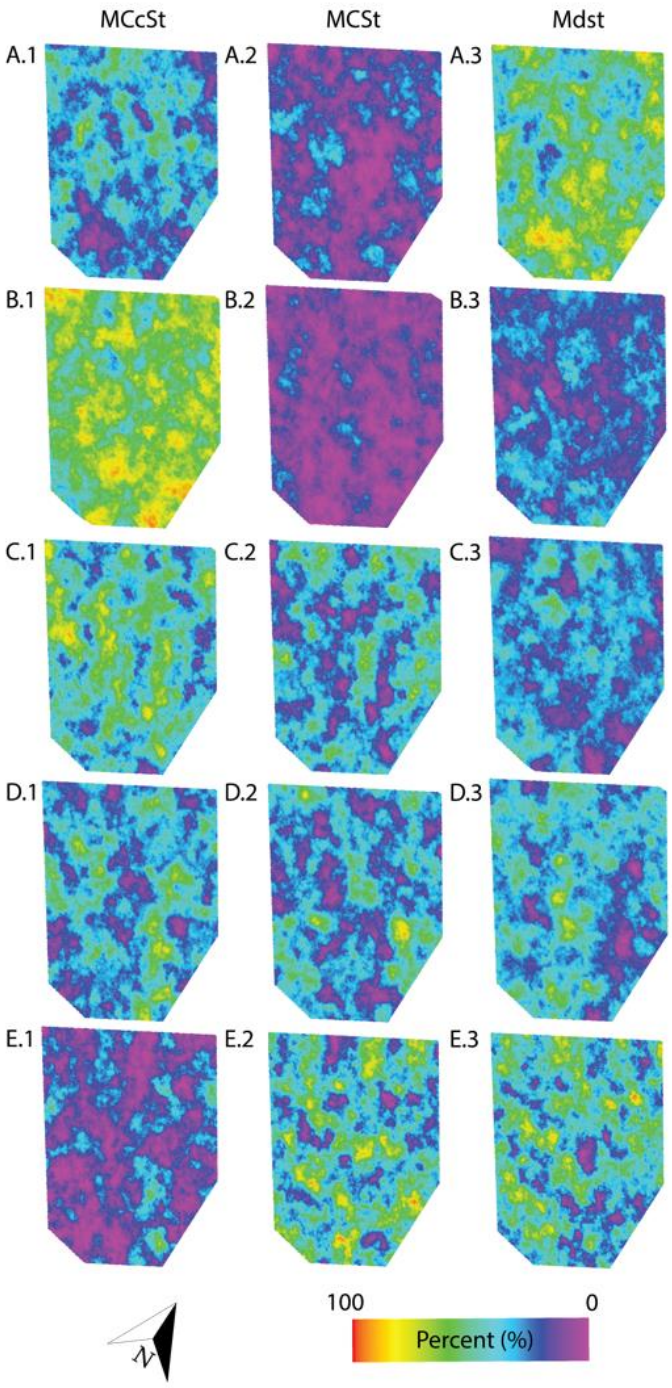

Figure 38: Petrofacies percentage maps for the MCSt (left) and MCcSt (right) by parasequence. Note the abundance of MCSt and Mdst petrofacies increase from parasequence B to E, contrary to petrofacies MCcSt that decreases from B to E. Parasequence B is the most cemented interval in the Sycamore Formation.

3D petrofacies models are important to quantifiy and illustrate the spatial variability of petrofacies that could control the distribution of geological attributes such as porosity,

permeability, and mineralogical composition implied by the petrofacies. Therefore, the presented workflow is important to obtain more information from cores efficiently and at a low cost. At the same time, the 3D data-driven models assisted by the proposed workflow aid the understanding of the diagenetic processes which occurred in the subsurface and affected reservoir quality. Future studies should be completed to compare 3D petrophysical models conditioned by different facies. For example, porosity models constrained to lithofacies models vs porosity models constrained to petrofacies models. This will highlight the value of the thin-section-based petrofacies.

4.7 Conclusions

In this paper, we propose a workflow that uses machine learning techniques to evaluate hundreds of models for selecting the best classification model that predicts petrofacies from well logs. The workflow consists of testing all combinations of the selected well logs, classifiers, and hyperparameters using a Grid-Search step followed by a 5-Fold cross-validation step. The first step demonstrated to be effective to generate over 1,800 models with different combinations of well logs, classifiers, and hyperparameters. The second step allowed us to validate and select the best model. The final model achieved 77% mean precision, but well-to-log calibrations are still a limitation to achieving higher performance. Furthermore, we demonstrated that the implementation of the workflow we discuss here has the potential for increasing control wells for constructing 3D data-driven facies models. More studies are needed to improve the proposed workflow, but in the current state, it has the capability to more effectively utilize core and well-log data at a low cost to help geoscientists understand the subsurface and to constrain geological models.

Chapter 5. Final Remarks and future work

This study was divided into two main portions, the first part, chapter 2, is a diagenetic study that presented a conceptual model to show the distribution of calcite-poor and -rich siltstones. The second part, chapters 3 and 4 validated the conceptual model with a robust 3D model. This part implemented a workflow that bridged the resolution gap between the scarce core data and the well log information to build robust 3D models based on thin-section information.

The knowledge facilitated in the first part impacts the scientific community and petroleum geologists in the industry. For the scientific community, this part contributed to the understanding of diagenesis in mixed carbonate-siliciclastic mudrocks, but also to reconstructing the paleoenvironment during the Early Mississippian time. From the 2nd chapter, we observed a major increase of silt-size quartz grains from parasequences B to C. This silt size material is so characteristic of eolian siltstones and it is important as a climatic archive (Soreghan et al., 2008). For the industry stakeholders, this first part provides a model of the authigenic calcite distribution, and this contributes to predicting the reservoir quality in the Sycamore. Understanding the areas and stratigraphic intervals that are more affected by diagenetic processes make the difference between good and bad reservoirs. The next step needs to be the understanding of the origin of the authigenic calcite, but the lack of techniques to measure the composition of the authigenic minerals in mudrocks difficult this task. However, we hope that these findings provide the foundation to begin studying the diagenetic processes in mixed-carbonated siliciclastic mudrocks.

The second portion contributes to generating robust 3D models at a reservoir scale from thin-section information. This also impacts both, the scientific community and the petroleum geologist working in the industry. The techniques implemented in the proposed workflow were

used to predict thin-section defined petrofacies. These petrofacies were used as proxies for the authigenic calcite distribution where the Massive Calcite-cemented Siltstones petrofacies are calcite-rich intervals, and the Massive Calcareous Siltstones petrofacies are calcite-poor intervals. However, the proposed workflow opened the door for modeling more complex diagenetic processes observed only at a thin-section scale. On top of that, the 3D petrofacies model has an immediate impact on the petroleum geologist working on the Sycamore Formation. This 3D model provides the visualization of the intervals with better reservoir quality in the Sho-Vel-Tum field. There are many research areas to be explored in the field of machine learning in geoscience, but first seismic data should be included to build a more robust model for the petrofacies distribution.

References

Al-Mudhafar, W. J., 2016, Incorporation of Bootstrapping and Cross-Validation for Efficient Multivariate Facies and Petrophysical Modeling: Society of Petroleum Engineers - SPE Low Perm Symposium, doi:10.2118/180277-MS.

Al-Mudhafar, W. J., 2017, Integrating well log interpretations for lithofacies classification and permeability modeling through advanced machine learning algorithms: J Petrol Explor Prod Technol, v. 7, p. 1023–1033, doi:10.1007/s13202-017-0360-0.

Allen, R. W., 2000, Complex Structural Features of the Ardmore Basin: Oklahoma City Geological Society in the Shale Shaker, v. 51.

Alnahwi, A., and R. G. Loucks, 2019, Mineralogical composition and total organic carbon quantification using x-ray fluorescence data from the Upper Cretaceous Eagle Ford Group in southern Texas: AAPG Bulletin, v. 103, no. 12, p. 2891–2907, doi:10.1306/04151918090.

Basu, S., 2009, Semi-Supervised Learning, *in* L. LIU, and M. T. ÖZSU, eds., Encyclopedia of Database Systems: Boston, MA, Springer US, p. 2613–2615, doi:10.1007/978-0-387-39940-9_609.

Bebout, D. G. ., M. Woodward, and J. M. Mendenhall, 1993, Atlas of Major Midcontinent Gas Reservoirs: Gas Research Institute.

Bellman, R., 1961, Adaptive Control Processes: A Guided Tour: Princeton University Press, 274 p., doi:10.2307/j.ctt183ph6v.

Bennison, A. P., 1956, Springer and Related Rocks of Oklahoma: Tulsa Geological Society Digest, v. 24, p. 111–115.

Bergstra, J., J. B. Ca, and Y. B. Ca, 2012, Random Search for Hyper-Parameter Optimization Yoshua Bengio: Journal of Machine Learning Research, v. 13, p. 281–305.

Bestagini, P., V. Lipari, and S. Tubaro, 2017, A machine learning approach to facies classification using well logs: no. August, p. 2137–2142, doi:10.1190/segam2017-17729805.1.

Bhattacharya, S., T. R. Carr, and M. Pal, 2016, Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA: Journal of Natural Gas Science and Engineering, v. 33, p. 1119–1133, doi:10.1016/j.jngse.2016.04.055.

Bhattacharya, S., and S. Mishra, 2018, Applications of machine learning for facies and fracture prediction using Bayesian Network Theory and Random Forest: Case studies from the Appalachian basin, USA: Journal of Petroleum Science and Engineering, v. 170, no. July, p. 1005–1017, doi:10.1016/j.petrol.2018.06.075.

Bjørkum, Per Arne, E. H. Oelkers, P. H. Nadeau, O. Walderhaug, and W. M. Murphy, 1998, Porosity Prediction in Quartzose Sandstones as a Function of Time, Temperature, Depth, Stylolite Frequency, and Hydrocarbon Saturation: AAPG Bulletin, v. 82, no. 4, p. 637–648.

Bjorlykke, K., and P. Aagaard, 1992, Clay Minerals in North Sea Sandstones.

Blakey, R., 2013, Early Mississippian paleogeography, Colorado Plateau geosystems, North America 345 Ma Mississippian (early to middle).

Boser, B. E., I. M. Guyon, and V. N. Vapnik, 1992, A Training Algorithm for Optimal Margin Classifiers, *in* Proceedings of the Fifth Annual Workshop on Computational Learning Theory: Association for Computing Machinery, COLT ’92, p. 144–152, doi:10.1145/130385.130401.

Boyd, D. T., 2002, GIS Files for Oklahoma oil and gas fields (distinguished by coalbed methane and field boundaries): Oklahoma Geological Survey Map GM-37.

Braun, J. C., 1959, A Stratigraphic Study of the Sycamore and Related Formations in the Southeastern Anadarko Basin: Shale Shaker Digest III.

Brcković, A., M. Kovačević, M. Cvetković, I. Kolenković Močilac, D. Rukavina, and B. Saftić, 2017, Application of artificial neural networks for lithofacies determination based on limited well data: Central European Geology, v. 60, no. 3, p. 299–315, doi:10.1556/24.60.2017.012.

Breiman, L., 2001, Random Forests: Machine Learning, v. 45, no. 1, p. 5–32, doi:10.1023/A:1010933404324.

Bressan, T. S., M. Kehl de Souza, T. J. Girelli, and F. C. Junior, 2020, Evaluation of machine learning methods for lithology classification using geophysical data: Computers and Geosciences, v. 139, p. 104475, doi:10.1016/j.cageo.2020.104475.

Burke, K., and J. F. Dewey, 1973, Plume-Generated Triple Junctions: Key Indicators in Applying Plate Tectonics to Old Rocks: The Journal of Geology, v. 81, no. 4, p. 406–433, doi:10.1086/627882.

Campello, R. J. G. B., P. Kröger, J. Sander, and A. Zimek, 2020, Density-based clustering: WIREs Data Mining and Knowledge Discovery , v. 10, no. 2, p. e1343, doi:https://doi.org/10.1002/widm.1343.

Cardott, B. J., 2012, Thermal maturity of Woodford Shale gas and oil plays, Oklahoma, USA: International Journal of Coal Geology, v. 103, p. 109–119, doi:10.1016/j.coal.2012.06.004.

Carpenter, M., and B. Tapp, 2014, Structural analysis of Sho-Vel-Tum using balanced cross sections, *in* N. Suneson, ed., Igneous and Tectonic History of the Southern Oklahoma Aulacogen, Guidebook 38: Oklahoma Geological Survey, p. 327–340.

Castro, W., J. Oblitas, R. Santa-Cruz, and H. Avila-George, 2017, Multilayer perceptron architecture optimization using parallel computing techniques: PLoS ONE, v. 12, no. 12, doi:10.1371/JOURNAL.PONE.0189369.

Cederstrand, J. R., 1997, Digital geologic map of Ardmore-Sherman quadrangles, south-central Oklahoma: doi:10.3133/OFR96370.

Cheng, G., Q. Yue, and X. Qiang, 2018, Research on Feasibility of Convolution Neural Networks for Rock Thin Sections Image Retrieval, *in* 2018 2nd IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference (IMCEC): p. 2539–2542, doi:10.1109/IMCEC.2018.8469642.

Chenoweth, P. A., J. C. Braun, S. C. Champlin, and J. D. Prestridge, 1959, Sycamore and related formations of Southern Oklahoma: Tulsa Geological Society Digest, v. 27.

Coffey, W., 2001, Lithostratigraphy and Porosity Characterization of the Sycamore Formation (Mississippian), and it's Relationship to Reservoir Performance, Carter-Knox Field, Grady and Stephens county, Oklahoma: no. August, p. 9–17.

Cohen, K. M., S. C. Finney, P. L. Gibbard, and J.-X. Fan, 2019, The ICS International Chronostratigraphic Chart: Episodes 36: 199-204. p.

Cole, T., 1989, A Surface to Subsurface Study of the Sycamore Limestone (Mississippian) Along the North Flank of the Arbuckle Anticline: Shale Shaker Digest XII.

Comer, J. B., 1992, Organic Geochemestry and Paleogeography of Upper Devonian Formation in Oklahoma and Wester Arkansas, *in* K. S. Johnsons, and B. J. Cardott, eds., Source Rocks in the Southern Midcontinent, 1990 Symposium: circular 93: Norman, OK, p. 70–93.

Comon, P., 1994, Independent component analysis, A new concept?*: Signal Processing, v. 36, p. 287–314.

Cooper, L. C., 1926, The Sycamore Limestone , Oklahoma Geological Survey Circular 9: 5 p.

Cover, T., and P. Hart, 1967, Nearest neighbor pattern classification: IEEE Transactions on Information Theory, v. 13, no. 1, p. 21–27, doi:10.1109/TIT.1967.1053964.

Culp, C. K., 1961, Stratigraphic Relations of the Sycamore Limestone (Mississippian) in Southern Oklahoma: Shale Shaker Digest III, p. 446–457.

Curtis, D. M., and S. C. Champlin, 1959, Depositional Environments of Mississippian Limestones of Oklahoma: Tulsa Geological Society Digest, v. 27, p. 90–103.

Daviran, M., A. Maghsoudi, R. Ghezelbash, and B. Pradhan, 2021, A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach: Computers & Geosciences, v. 148, p. 104688, doi:10.1016/J.CAGEO.2021.104688.

Davis, J. C., 2018, Electrofacies in Reservoir Characterization: Handbook of Mathematical Geosciences: Fifty Years of IAMG, p. 211–223, doi:10.1007/978-3-319-78999-6_11.

Dennie, D., R. D. Elmore, J. Deng, E. Manning, and J. Pannalal, 2012, Palaeomagnetism of the Mississippian Barnett Shale, Fort Worth Basin, Texas: Geological Society, London, Special Publications, v. 371, no. 1, p. 89–106, doi:10.1144/SP371.10.

Dickinson, W. R., 1970, Interpreting Detrital Modes of Graywacke and Arkose: SEPM Journal of Sedimentary Research, v. Vol. 40, no. 2, p. 695–707, doi:10.1306/74D72018-2B21-

11D7-8648000102C1865D.

Dokmanic, I., R. Parhizkar, J. Ranieri, and M. Vetterli, 2015, Euclidean Distance Matrices: Essential theory, algorithms, and applications: IEEE Signal Processing Magazine, v. 32, no. 6, p. 12–30, doi:10.1109/MSP.2015.2398954.

Donovan, N. R., 2001, Field study of the Sycamore Formation on Interstate Highway 35 in the Arbuckle Mountains Oklahoma, *in* K. Jhonson, ed., Silurian, Devonian, and Mississippian Geology and Petroleum in the Southern Midcontinent, 1999 Symposium: Oklahoma Geological Survey Circular 105: Norman, OK, Oklahoma Geological Survey and National Petroleum Technology Office, U.S. Department of Energy, p. 139–149.

Duarte-Coronado, D., J. Tellez-Rodriguez, R. Pires de Lima, K. Marfurt, and R. Slatt, 2019, Deep convolutional neural networks as an estimator of porosity in thin-section images for unconventional reservoirs : doi:10.1190/segam2019-3216898.1.

Duarte, D., 2018, Rock Characterization and Stratigraphy of the Mississippian Strata, Meramec/Sycamore. Merge play, Central Oklahoma: University of Oklahoma, 91 p.

Duarte, D., R. P. De Lima, R. Slatt, and K. Marfurt, 2019, Comparison of Clustering Techniques to Define Chemofacies in Mississippian Rocks in The STACK Play , Oklahoma, *in* Search and Discovery Article #42523 (2020). Adapted from poster presentation given at 2019 AAPG Annual Convention and Exhibition, San Antonio, Texas, May 19-22, 2019: doi:10.1306/42523Duarte2020.

Duarte, D., B. Milad, R. D. Elmore, M. J. Pranter, and R. Slatt, 2021, Diagenetic controls on reservoir quality of a mixed carbonate-siliciclastic system: Sycamore Formation, Sho-Vel-Tum Field, Oklahoma, USA: Marine and Petroleum Geology, v. 134, p. 105375, doi:https://doi.org/10.1016/j.marpetgeo.2021.105375.

Dubois, M. K., G. C. Bohling, and S. Chakrabarti, 2007, Comparison of four approaches to a rock facies classification problem: Computers and Geosciences, v. 33, no. 5, p. 599–617, doi:10.1016/j.cageo.2006.08.011.

Dunham, M. W., A. Malcolm, and W. J. Kim, 2020, Improved well-log classification using semisupervised label propagation and self-training, with comparisons to popular supervised

algorithms: GEOPHYSICS, v. 85, no. 1, p. O1–O15, doi:10.1190/geo2019-0238.1.

Dunham, M. W., A. Malcolm, and J. K. Welford, 2020, Improved well log classification using semi-supervised algorithms, *in* SEG International Exposition and Annual Meeting 2019: Society of Exploration Geophysicists, p. 2398–2402, doi:10.1190/segam2019-3208817.1.

Ehrenberg, S. N., 1989, Assessing the Relative Importance of Compaction Processes and Cementation to Reduction of Porosity in Sandstones: Discussion; Compaction and Porosity Evolution of Pliocene Sandstones, Ventura Basin, California: DISCUSSION: AAPG Bulletin, v. 73, no. 10, p. 1274–1276, doi:10.1306/44b4aa1e-170a-11d7-8645000102c1865d.

Eichhubl, P., and A. Aydin, 2005, Structure, petrophysics, and diagenesis of shale entrained along a normal fault at Black Diamond Mines, California.Implications for fault seal: AAPG Bulletin, v. 89, no. 9, p. 1113–1137, doi:10.1306/04220504099.

Elkan, C., 2003, Using the Triangle Inequality to Accelerate-Means, *in* Twentieth International Conference on Machine Learning (ICML-2003).

Elmore, R. D., G. W. Heij, and A. K. Wickard, 2016, Paragenesis of mineralized fractures and diagenesis of prominent North American shales: The Sedimentary Record, v. 14, no. 4, p. 4–10, doi:10.2110/SEDRED.2016.4.4.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu, 1996, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *in* E. Simoudis, J. Han, and U. Fayyad, eds., KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data: AAAI Press, p. 226–231.

Evans, J. L., 1979, Major structural features of the Anadarko basin: no. Figure 2, p. 1–13.

Fay, R. O., 1989, Geology of the Arbuckle Mountains along Interstate - 35, Carter and Murray Counties, Oklahoma: 50 p.

Feinstein, S., 1981, Subsidence and Thermal History of Southern Oklahoma Aulacogen: Implications for Petroleum Exploration: AAPG Bulletin, v. 65, no. 12, p. 2521–2533, doi:10.1306/03B599F9-16D1-11D7-8645000102C1865D.

Freed, R. L., and D. R. Peacor, 1989, Variability in temperature of the smectite/illite reaction in Gulf Coast sediments: 171–180 p.

Galvis, H., D. Becerra, and R. Slatt, 2018, Lithofacies and stratigraphy of a complete Woodford Shale outcrop section in South Central Oklahoma: Geologic considerations for the evaluation of unconventional shale reservoirs: Interpretation, v. 6, no. 1, p. SC15–SC27, doi:10.1190/INT-2017-0074.1.

Gerilyn S. Soreghan, 1992, Preservation and paleoclimatic significance of eolian dust in the Ancestral Rocky Mountains province: Geology, v. 20, no. 12, p. 1111–1114, doi:https://doi.org/10.1130/0091-7613(1992)020<1111:PAPSOE>2.3.CO;2.

Ghosh, S., H. A. Galvis-Portilla, C. M. Klockow, and R. M. Slatt, 2018, An application of outcrop analogues to understanding the origin and abundance of natural fractures in the Woodford Shale: Journal of Petroleum Science and Engineering, v. 164, p. 623–639, doi:10.1016/J.PETROL.2017.11.073.

Giles, J. M., M. J. Soreghan, K. C. Benison, G. S. Soreghan, and S. T. Hasiotis, 2013, Lakes, Loess, and Paleosols In the Permian Wellington Formation of Oklahoma, U.S.A.: Implications For Paleoclimate and Paleogeography of the MidcontinentLAKES, LOESS, AND PALEOSOLS IN THE PERMIAN WELLINGTON FORMATION OF OKLAHOMA: Journal of Sedimentary Research, v. 83, no. 10, p. 825–846, doi:10.2110/JSR.2013.59.

Gupta, I., C. Rai, C. Sondergeld, and D. Devegowda, 2017, Rock typing in Eagle Ford, Barnett, and Woodford formations, *in* SPE/AAPG/SEG Unconventional Resources Technology Conference 2017: doi:10.15530/urtec-2017-2669624.

Gutschick, R., and C. Sandberg, 1983, Mississippian continental margins on the conterminous United States: SEPM Special Publication, v. 33, no. 33, p. 79–96, doi:10.2110/pec.83.06.0079.

Halotel, J., V. Demyanov, and A. Gardiner, 2019, Value of Geologically Derived Features in Machine Learning Facies Classification: Mathematical Geosciences 2019 52:1, v. 52, no. 1, p. 5–29, doi:10.1007/S11004-019-09838-0.

Ham, W. E., 1978, Regional geology of the Arbuckle Mountains, Oklahoma: p. 1–61.

Ham, W. E., R. E. Denison, and A. C. Merritt, 1965, Basement rocks and structural evolution of southern Oklahoma - A Summary: Bulletin of the American Association of Petroleum Geologists, v. 49, no. 7, p. 927–934.

Handhal, A. M., A. M. Al-Abadi, H. E. Chafeet, and M. J. Ismail, 2020, Prediction of total organic carbon at Rumaila oil field, Southern Iraq using conventional well logs and machine learning algorithms: Marine and Petroleum Geology, v. 116, no. January, p. 104347, doi:10.1016/j.marpetgeo.2020.104347.

Hardisty, L. et al., 2021, Stratigraphic variability of Mississippian Meramec chemofacies and petrophysical properties using machine learning and geostatistical modeling, STACK trend, Anadarko Basin, Oklahoma: Interpretation, v. 9, no. 3, p. T987–T1007, doi:10.1190/INT-2020-0169.1.

Harlton, B., 1956, The Harrisburg Trough Stephens and Carter Counties, Oklahoma: p. 135-143.

Harrison, W. J., and G. D. Thyne, 1992, Predictions of Diagenetic Reactions in the Presente of Organic Acids: Geochemica et Cosmochimica Acta, v. 56, p. 565–586.

Heij, G. W., and R. D. Elmore, 2019, The magnetic fabric of the Wolfcamp shale, Midland Basin, west Texas: Understanding petrofabric variability, hydrocarbon associations, and iron enrichment: AAPG Bulletin, v. 103, no. 11, p. 2785–2806, doi:10.1306/04101918140.

Henry, M. E., and T. C. Hester, 1995, ANADARKO BASIN PROVINCE (058) With a section on the Woodford/Chattanooga/Arkansas/Novaculite of Midcontinent Play: 30, 51 p.

Heran, W. D., G. N. Green, and D. B. Stoeser, 2003, A Digital Geologic Map Database for the State of Oklahoma.

Ho, T. K., 1995, Random decision forests, *in* Proceedings of 3rd International Conference on Document Analysis and Recognition: p. 278–282 vol.1, doi:10.1109/ICDAR.1995.598994.

Hoffman, P., J. Dewey, and K. Burke, 1974, Aulacogens and their genetic relation to geosynclines, with a proterozoic example from great slave lake, Canada, *in* Modern and Ancient Geosynclinal Sedimentation: SEPM (Society for Sedimentary Geology), p. 38–55,

doi:10.2110/pec.74.19.0038.

Hossain, T. M., J. Watada, I. A. Aziz, and M. Hermana, 2020, Machine learning in electrofacies classification and subsurface lithology interpretation: A rough set theory approach: Applied Sciences (Switzerland), v. 10, no. 17, doi:10.3390/app10175940.

Huffman, G. G., and J. C. Barker, 1950, Mississippian Problems in the Lawrence Uplift , Pontotoc County , Oklahoma: no. 16, p. 22–24.

Hyvärinen, A., J. Karhunen, and E. Oja, 2001, What is Independent Component Analysis?, *in* Independent Component Analysis: John Wiley & Sons, Ltd, p. 145–164, doi:https://doi.org/10.1002/0471221317.ch7.

Hyvärinen, A., and E. Oja, 2000, Independent Component Analysis: Algorithms and Applications: 411–430 p.

Ippolito, M., J. Ferguson, and F. Jenson, 2021, Improving facies prediction by combining supervised and unsupervised learning methods: Journal of Petroleum Science and Engineering, v. 200, no. December 2019, p. 108300, doi:10.1016/j.petrol.2020.108300.

Ivakhnenko, A. G., and V. G. Lapa, 1968, Cybernetics and Forecasting Techniques: Nature, v. 219, no. 5150, p. 202–203, doi:10.1038/219202b0.

Jennings, D. S., and J. Antia, 2013, Petrographie Characterization of the Eagle Ford Shale, South Texas: Mineralogy, Common Constituents, and Distribution of Nanometer-scale Pore Types: AAPG Memoir, no. 102, p. 101–113, doi:10.1306/13391708M1023586.

Jolliffe, I. T., 2002, Principal Component Analysis: New York, Springer-Verlag New York, 488 p., doi:10.1007/b98835.

Kantorowicz, J. D., I. D. Bryant, and J. M. Dawans, 1987, Controls on the geometry and distribution of carbonate cements in Jurassic sandstones: Bridport Sands, southern England and Viking Group, Troll Field, Norway: Geological Society Special Publication, v. 36, no. 1, p. 103–118, doi:10.1144/GSL.SP.1987.036.01.09.

Katsube, T. J., and M. A. Williamson, 1994, Effects of diagenesis on shale nano-pore structure and implications for sealing capacity: Clay Minerals, v. 29, no. 4, p. 451–461,

doi:10.1180/CLAYMIN.1994.029.4.05.

Keller, G. R., E. G. Lidiak, W. J. Hinze, and L. W. Braile, 1983, The Role of Rifting in the Development of the Midctontinent U.S.A.: Tectonophysics, v. 94, p. 391–412.

Kessler, J. L. P., G. S. Soreghan, and H. J. Wacker, 2001, Equatorial Aridity in Western Pangea: Lower Permian Loessite and Dolomitic Paleosols in Northeastern New Mexico, U.S.A.: Journal of Sedimentary Research, v. 71, no. 5, p. 817–832, doi:10.1306/2DC4096B-0E47-11D7-8643000102C1865D.

Kohavi, R., 1995, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, *in* Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2: Morgan Kaufmann Publishers Inc., IJCAI'95, p. 1137–1143.

Lachenbruch, P. A., and M. R. Mickey, 1968, Estimation of Error Rates in Discriminant Analysis: Technometrics, v. 10, no. 1, p. 1–11, doi:10.1080/00401706.1968.10490530.

Lai, J., X. Fan, B. Liu, X. Pang, S. Zhu, W. Xie, and G. Wang, 2020, Qualitative and quantitative prediction of diagenetic facies via well logs: Marine and Petroleum Geology, v. 120, no. May, p. 104486, doi:10.1016/j.marpetgeo.2020.104486.

Lane, H. R., and T. L. De Keyser, 1980, Paleogeography of the Late Early Mississippian (Tournaisian 3) in the central and southwestern United States: Paleozoic Paleogeography of West-Central United States: Rocky Mountain Paleogeography Symposium 1, v. 6, p. 149–162.

LaValle, S. M., M. S. Branicky, and S. R. Lindemann, 2004, On the Relationship between Classical Grid Search and Probabilistic Roadmaps: The International Journal of Robotics Research, v. 23, no. 7–8, p. 673–692, doi:10.1177/0278364904045481.

Li, Y., X. Chang, W. Yin, T. Sun, and T. Song, 2017, Quantitative impact of diagenesis on reservoir quality of the Triassic Chang 6 tight oil sandstones, Zhenjing area, Ordos Basin, China: Marine and Petroleum Geology, v. 86, p. 1014–1028, doi:10.1016/j.marpetgeo.2017.07.005.

Liu, L., and M. T. Özsu (eds.), 2009, Semi-Supervised Clustering, *in* Encyclopedia of Database Systems: Boston, MA, Springer US, p. 2613, doi:10.1007/978-0-387-39940-9_3579.

Loucks, R. G., and S. C. Ruppel, 2007, Mississippian Barnett Shale: Lithofacies and depositional setting of a deep-water shale-gas succession in the Fort Worth Basin, Texas: AAPG Bulletin, v. 91, no. 4, p. 579–601, doi:10.1306/11020606059.

Lowe, D. R., 1975, Regional controls on silica sedimentation in the Ouachita system: Bulletin of the Geological Society of America, v. 86, no. 8, p. 1123–1127, doi:10.1130/0016-7606(1975)86<1123:RCOSSI>2.0.CO;2.

van der Maaten, L., and G. Hinton, 2008, Visualizing Data using t-SNE: 1–48 p.

Mandal, P. P., and R. Rezaee, 2019, Facies classification with different machine learning algorithm – An efficient artificial intelligence technique for improved classification: ASEG Extended Abstracts, v. 2019, no. 1, p. 1–6, doi:10.1080/22020586.2019.12072918.

Mansurbeg, H., S. Morad, P. Plink-Bjorklund, M. A. K. E-Ghali, M. A. Caja, and R. Marfil, 2012, Diagenetic alterations related to falling stage and lowstand systems tracts of shelf, slope and basin floor sandstones (Eocene Central Basin, Spitsbergen): no. January, p. 353–378.

Mansurbeg, H., L. F. De Ros, S. Morad, J. M. Ketzer, M. A. K. El-Ghali, M. A. Caja, and R. Othman, 2012, Meteoric-water diagenesis in late Cretaceous canyon-fill turbidite reservoirs from the Espírito Santo Basin, eastern Brazil: Marine and Petroleum Geology, v. 37, no. 1, p. 7–26, doi:10.1016/j.marpetgeo.2012.03.009.

Mathia, E. J., L. Bowen, K. M. Thomas, and A. C. Aplin, 2016, Evolution of porosity and pore types in organic-rich, calcareous, Lower Toarcian Posidonia Shale: Marine and Petroleum Geology, v. 75, p. 117–139, doi:10.1016/J.MARPETGEO.2016.04.009.

McBride, E. F., 1984, Compaction in sandstones: influence on reservoir quality: AAPG Bulletin, v. 68:4.

McCulloch, W. S., and W. Pitts, 1943, A logical calculus of the ideas immanent in nervous activity: The bulletin of mathematical biophysics, v. 5, no. 4, p. 115–133, doi:10.1007/BF02478259.

McGlannan, A. J., A. L. Bonar, L. S. Pfeifer, S. Adams, D. Duarte, C. Totten, A. Cullen, and G. S. Soreghan, 2019, A model for the origin of Devono-Mississippian mudrocks in greater

midcontinent North America, *in* GSA Annual Meeting: Geological Society of America, doi:10.1130/ABS/2019AM-337810.

McInnes, L., J. Healy, and J. Melville, 2018, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction: arXiv.

Meng, F., L. N. Y. Wong, and H. Zhou, 2021, Rock brittleness indices and their applications to different fields of rock engineering: A review: Journal of Rock Mechanics and Geotechnical Engineering, v. 13, no. 1, p. 221–247, doi:10.1016/J.JRMGE.2020.06.008.

Merembayev, T., D. Kurmangaliyev, B. Bekbauov, and Y. Amanbek, 2021, A Comparison of Machine Learning Algorithms in Predicting Lithofacies: Case Studies from Norway and Kazakhstan: Energies 2021, Vol. 14, Page 1896, v. 14, no. 7, p. 1896, doi:10.3390/EN14071896.

Milad B. and Slatt R., 2019, Outcrop to Subsurface Reservoir Characterization of the Mississippian Sycamore/Meramec Play in the SCOOP Area, Arbuckle Mountains, Oklahoma, USA: Unconventional Resources Technology Conference 991-MS, doi:10.15530/urtec-2019-991.

Milad, B., R. Slatt, and Z. Fuge, 2020, Lithology, stratigraphy, chemostratigraphy, and depositional environment of the Mississippian Sycamore rock in the SCOOP and STACK area, Oklahoma, USA: Field, lab, and machine learning studies on outcrops and subsurface wells: Marine and Petroleum Geology, p. 104278, doi:10.1016/j.marpetgeo.2020.104278.

Milliken, K. L., and T. Olson, 2017, Silica Diagenesis, Porosity Evolution, and Mechanical Behavior In Siliceous Mudstones, Mowry Shale (Cretaceous), Rocky Mountains, U.S.A.K.L. MILLIKEN AND T. OLSONCEMENT IN SILICEOUS MUDSTONE: Journal of Sedimentary Research, v. 87, no. 4, p. 366–387, doi:10.2110/JSR.2017.24.

Morad, S., I. Al-Aasm, K. Ramseyer, R. Marfil, and A. A. Aldahan, 1990, Diagenesis of carbonate cements in Permo-Triassic sandstones from the Iberian Range, Spain: evidence from chemical composition and stable isotopes: Sedimentary Geology, v. 67, no. 3–4, p. 281–295, doi:10.1016/0037-0738(90)90039-V.

Nadeau, P. H., P. A. Bjørkum, and O. Walderhaug, 2005, Petroleum system analysis: impact of

shale diagenesis on reservoir fluid pressure, hydrocarbon migration, and biodegradation risks: Geological Society, London, Petroleum Geology Conference series, v. 6, no. 1, p. 1267–1274, doi:10.1144/0061267.

Nana Yobo, L., and N. Yobo, 2014, Origin and distribution of the Mississippian-Pennsylvanian boundary unconformity in marine carbonate successions with a case study of the karst development atop the Madison formation in the Bighorn basin, Wyoming: p. 59.

Okunuwadje, S. E., S. A. Bowden, and D. I. M. Macdonald, 2020, Diagenesis and reservoir quality in high-resolution sandstone sequences: An example from the Middle Jurassic Ravenscar sandstones, Yorkshire CoastUK: Marine and Petroleum Geology, p. 104426, doi:10.1016/j.marpetgeo.2020.104426.

Pang, Y., B. Shi, X. Guo, X. Zhang, Z. Han, L. Cai, G. Xiao, and H. Liu, 2021, Source–reservoir relationships and hydrocarbon charging history in the central uplift of the south Yellow Sea basin (East Asia): Constrained by machine learning procedure and basin modeling: Marine and Petroleum Geology, v. 123, no. September 2020, p. 104731, doi:10.1016/j.marpetgeo.2020.104731.

Peace, H., 1994, Mississippian Facies Relationships Eastern Anadarko Basin, Oklahoma: Shale Shaker Digest XIII, p. 193–202.

Pearson, K., 1901, LIII. On lines and planes of closest fit to systems of points in space : The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, v. 2, no. 11, p. 559–572, doi:10.1080/14786440109462720.

Pedregosa, F. et al., 2011, Scikit-learn: Machine Learning in Python: Journal of Machine Learning Research, v. 12, no. 85, p. 2825–2830.

Perry, W. J., 1989, Tectonic Evolution of the Anadarko Basin Region, Oklahoma: Denver, United State Government Printing Office.

Pires de Lima, R., D. Duarte, C. Nicholson, R. Slatt, and K. J. Marfurt, 2020, Petrographic microfacies classification with deep convolutional neural networks: Computers and Geosciences, v. 142, no. June, p. 104481, doi:10.1016/j.cageo.2020.104481.

Prestridge, J. D., 1957, A Subsurface Stratigraphic Study of the Sycamore Formation in the

Ardmore Basin: Chenowith, P.A, 62 p.

Price, B., A. C. Pollack, A. P. Lamb, T. C. Peryam, J. R. Anderson, and B. J. Price, 2019, Depositional interpretation and sequence stratigraphic control on reservoir quality and distribution in the Meramecian STACK play, Anadarko Basin, Oklahoma, United States: AAPG Bulletin, v. 2, no. 2, p. 357–386, doi:10.1306/04301917411.

Rahman, M. J. J., and R. H. Worden, 2016, Diagenesis and its impact on the reservoir quality of Miocene sandstones (Surma Group) from the Bengal Basin, Bangladesh: Marine and Petroleum Geology, v. 77, p. 898–915, doi:10.1016/j.marpetgeo.2016.07.027.

Roberts, J. M., and R. D. Elmore, 2018, A diagenetic study of the Woodford Shale in the southeastern Anadarko Basin, Oklahoma, USA: Evidence for hydrothermal alteration in mineralized fractures: Interpretation, v. 6, no. 1, p. SC1–SC13, doi:10.1190/INT-2017-0071.1.

Rowe, H., N. Hughes, and K. Robinson, 2012, The quantification and application of handheld energy-dispersive x-ray fluorescence (ED-XRF) in mudrock chemostratigraphy and geochemistry: Chemical Geology, v. 324–325, p. 122–131, doi:10.1016/j.chemgeo.2011.12.023.

Saigal, G. C., and K. Bjørlykke, 1987, Carbonate cements in clastic reservoir rocks from offshore Norway—relationships between isotopic composition, textural development and burial depth: Geological Society, London, Special Publications, v. 36, no. 1, p. 313–324, doi:10.1144/GSL.SP.1987.036.01.22.

Schubert, E., J. Sander, M. Ester, H. P. Kriegel, and X. Xu, 2017, DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN: ACM Trans. Database Syst., v. 42, no. 3, doi:10.1145/3068335.

Schwartzapfel, J. A., 1990, Biostratigraphic investigations of Late Paleozoic (Upper Devonian to Mississippian) radiolaria within the Arbuckle Mountains and Ardmore Basin of south-central Oklahoma: University of Texas at Dallas, Richardson, Texas, 475 p.

Schwartzapfel, J. A., and B. K. Holdsworth, 1996, Upper Devonian and Mississippian radiolarian zonation and biostratigraphy of the Woodford, Sycamore, Caney, and Goddard

formations, Oklahoma (1996 edition) | Open Library, *in* Lawrence, Kansas, Cushman Foundation for Foraminiferal Research Special Publication 33: p. 275.

Shannon, C. E., and J. McCarthy, 2016, Automata Studies. (AM-34), Volume 34: Princeton, Princeton University Press, doi:https://doi.org/10.1515/9781400882618.

Shaw, H. F., and T. J. Primmer, 1991, Diagenesis of mudrocks from the Kimmeridge Clay Formation of the Brae Area, UK North Sea: Marine and Petroleum Geology, v. 8, no. 3, p. 270–277, doi:10.1016/0264-8172(91)90081-B.

Siever, R., 1951, The Mississippian-Pennsylvanian Unconformity in Southern Illinois: AAPG Bulletin, v. 35, no. 3, p. 542–581, doi:10.1306/3D9341D8-16B1-11D7-8645000102C1865D.

Singh, H., Y. Seol, and E. M. Myshakin, 2020, Automated Well-Log Processing and Lithology Classification by Identifying Optimal Features Through Unsupervised and Supervised Machine-Learning Algorithms: SPE Journal, v. 25, no. 05, p. 2778–2800, doi:10.2118/202477-PA.

Slatt, R. M. et al., 2012, Pore-to-regional-scale, integrated characterization workflow for unconventional gas shales: Shale reservoirs—Giant resources for the 21st century: AAPG Memoir 97, p. 127–150, doi:10.1306/13321461M97441.

Slatt, R. M., H. Galvis-Portillo, D. Becerra-Rondon, I. C. Ekwunife, R. Brito, J. Zhang, C. Molinares, E. Torres, D. Duarte, and B. Milad, 2018, Outcrop and Subsurface Geology Applied to Drilling, Sweet Spot and Target Zone Detection of Resource Shales: The Woodford Example: SPE/AAPG/SEG Unconventional Resources Technology Conference 2018, URTC 2018, doi:10.15530/URTEC-2018-2893838.

Slatt, R. M., and N. R. O'Brien, 2011, Pore types in the Barnett and Woodford gas shales: Contribution to understanding gas storage and migration pathways in fine-grained rocks: AAPG Bulletin, v. 95, no. 12, p. 2017–2030, doi:10.1306/03301110145.

Slatt, R. M., J. Zhang, E. Torres-Parada, R. Brito, D. Duarte, and B. Milad, 2021, Unconventional Petroleum Resources, *in* Encyclopedia of Geology: Academic Press, p. 783–807, doi:10.1016/B978-0-08-102908-4.00023-0.

Soreghan, G. S., N. G. Heavens, L. A. Hinnov, S. M. Aciego, and C. Simpson, 2015, Reconstructing the Dust Cycle in Deep Time: the Case of the Late Paleozoic Icehouse: The Paleontological Society Papers, v. 21, p. 83–120, doi:10.1017/S1089332600002977.

Soreghan, G. S., A. M. Moses, M. J. Soreghan, M. A. Hamilton, C. Mark Fanning, and P. K. Link, 2007, Palaeoclimatic inferences from upper Palaeozoic siltstone of the Earp Formation and equivalents, Arizona-New Mexico (USA): Sedimentology, v. 54, no. 3, p. 701–719, doi:10.1111/J.1365-3091.2007.00857.X.

Soreghan, G. S., M. J. Soreghan, and M. A. Hamilton, 2008, Origin and significance of loess in late Paleozoic western Pangaea: A record of tropical cold? Palaeogeography, Palaeoclimatology, Palaeoecology, v. 268, no. 3–4, p. 234–259, doi:10.1016/j.palaeo.2008.03.050.

Steen, D., 2020, A gentle introduction to self-training and semi-supervised learning: Medium: <https://towardsdatascience.com/a-gentle-introduction-to-self-training-and-semi-supervised-learning-ceee73178b38> (accessed December 23, 2020).

Stevenson, E., N. J. Puffer, M. M. Leighton, and R. Sieyer, 1951, The Mississippian Pennsylvanian unconformity in southern Illinois: The Bulletin of the American Association of Petroleum Geologists, v. 35, no. 3, p. 542–581.

Stone, M., 1977, Asymptotics for and against cross-validation: Biometrika, v. 64, no. 1, p. 29–35, doi:10.1093/biomet/64.1.29.

Stone, M., 1974, Cross-Validatory Choice and Assessment of Statistical Predictions on JSTOR: Journal of the Royal Statistical Society. Series B (Methodological), v. 36, no. 2, p. 111–47,.

Sun, N., J. Zhong, B. Hao, Y. Ge, and R. Swennen, 2019, Sedimentological and diagenetic control on the reservoir quality of deep-lacustrine sedimentary gravity flow sand reservoirs of the Upper Triassic Yanchang Formation in Southern Ordos Basin, China: Marine and Petroleum Geology, v. 112, no. September 2019, p. 104050, doi:10.1016/j.marpetgeo.2019.104050.

Sutherland, P. K., 1988, Late Mississippian and Pennsylvanian depositional history in the Arkoma basin area, Oklahoma and Arkansas: Bulletin of the Geological Society of

America, v. 100, no. 11, p. 1787–1802, doi:10.1130/0016-7606(1988)100<1787:LMAPDH>2.3.CO;2.

Taff, J. A., 1903, Preliminary Report on the Geology of the Arbuckle and Wichita Mountains in Indian Territory and Oklahoma: 81 p.

Taylor, A. M., and R. Goldring, 1993, Description and analysis of bioturbation and ichnofabric: Journal of the Geological Society, v. 150, no. 1, p. 141–148, doi:10.1144/gsjgs.150.1.0141.

Turner, B. W., J. A. Tréanton, and R. M. Slatt, 2016, The use of chemostratigraphy to refine ambiguous sequence stratigraphic correlations in marine mudrocks. An example from the Woodford Shale, Oklahoma, USA: Journal of the Geological Society, v. 173, no. 5, p. 854–868, doi:10.1144/jgs2015-125.

Vaisblat, N. et al., 2021, Compositional and Diagenetic Evolution of a Siltstone, with Implications for Reservoir Quality; an Example from the Lower Triassic Montney Formation in western Canada: Marine and Petroleum Geology, v. 129, p. 105066, doi:10.1016/J.MARPETGEO.2021.105066.

Walderhaug, O. and, and P. A. Bjørkum, 1992, Effect of meteoric water flow on calcite cementation in the Middle Jurassic Oseberg Formation, well 30/3-2, Veslefrikk Field, Norwegian North Sea: Marine and Petroleum Geology, v. 9, no. 3, p. 308–318, doi:10.1016/0264-8172(92)90079-T.

Wang, G., T. R. Carr, Y. Ju, and C. Li, 2014, Identifying organic-rich Marcellus Shale lithofacies by support vector machine classifier in the Appalachian basin: Computers and Geosciences, v. 64, p. 52–60, doi:10.1016/j.cageo.2013.12.002.

Wang, F. P., and J. F. W. Gale, 2009, Screening Criteria for Shale-Gas Systems: Gulf Coast Association of Geological Societies Transactions, *in* GCAGS 59th Annual Meeting: p. 779–793.

Wang, M., Z. Yang, C. Shui, Z. Yu, Z. Wang, and Y. Cheng, 2019, Diagenesis and its influence on reservoir quality and oil-water relative permeability: A case study in the Yanchang Formation Chang 8 tight sandstone oil reservoir, Ordos Basin, China: Open Geosciences, v. 11, no. 1, p. 37–47, doi:10.1515/GEO-2019-0004.

Wang, W., C. Zhou, C. Guan, X. Yuan, Z. Chen, and B. Wan, 2014, An integrated carbon, oxygen, and strontium isotopic studies of the Lantian Formation in South China with implications for the Shuram anomaly: Chemical Geology, v. 373, p. 10–26, doi:10.1016/J.CHEMGEO.2014.02.023.

Watney, W. L., W. J. Guy, J. H. Doveton, S. Bhattacharya, P. M. Gerlach, G. C. Bohling, and T. R. Carr, 1998, Petrofacies Analysis - A Petrophysical Tool for Geologic/Engineering Reservoir Characterization: AAPG Memoir, no. 71, p. 73–90.

Wickard, A. K., R. D. Elmore, and G. W. Heij, 2020, A Diagenetic Study of the Wolfcamp Shale in the Southeast Midland Basin, West Texas: A Petrographic and SEM Study of Two Cores: Memoir 120: Mudstone Diagenesis: Research Perspectives for Shale Hydrocarbon Reservoirs, Seals, and Source Rocks, p. 225–239, doi:10.1306/13672219M1213827.

Wickham, J., 1978, The southern Oklahoma Aulacogen-Structure and Stratigraphy of the Ouachita Mountains and the Arkoma Basin; Field Guide.

Wood, D. A., 2019, Lithofacies and stratigraphy prediction methodology exploiting an optimized nearest-neighbour algorithm to mine well-log data: Marine and Petroleum Geology, v. 110, no. May, p. 347–367, doi:10.1016/j.marpetgeo.2019.07.026.

Worden, R. H., P. J. Armitage, A. R. Butcher, J. M. Churchill, A. E. Csoma, C. Hollis, R. H. Lander, and J. E. Omma, 2018, Petroleum reservoir quality prediction: Overview and contrasting approaches from sandstone and carbonate communities, *in* Geological Society Special Publication: Geological Society of London, p. 1–31, doi:10.1144/SP435.21.

Worden, R. H., and S. A. Barclay, 2000, Internally-sourced quartz cement due to externally-derived CO2 in sub-arkosic sandstones, North Sea, *in* Journal of Geochemical Exploration: doi:10.1016/S0375-6742(00)00104-7.

Worden, R. H., and S. Morad, 2003, Clay minerals in sandstones: a review of the detrital and diagenetic sources and evolution during burial: Clay Cement in Sandstones: International Association of Sedimentologists, Special Publication, 3–41 p.

Xia, Y., H. Zhou, C. Zhang, S. He, Y. Gao, and P. Wang, 2019, The evaluation of rock brittleness and its application: a review study:

https://doi.org/10.1080/19648189.2019.1655485, doi:10.1080/19648189.2019.1655485.

Xie, W., and K. T. Spikes, 2020, Well-log facies classification using a semi-supervised algorithm, *in* SEG Technical Program Expanded Abstracts: Society of Exploration Geophysicists, p. 2151–2155, doi:10.1190/segam2020-3426030.1.

Xiong, D., K. Azmy, and N. J. F. Blamey, 2016, Diagenesis and origin of calcite cement in the Flemish Pass Basin sandstone reservoir (Upper Jurassic): Implications for porosity development: Marine and Petroleum Geology, v. 70, p. 93–118, doi:10.1016/j.marpetgeo.2015.11.013.

Xu, D., and Y. Tian, 2015, A Comprehensive Survey of Clustering Algorithms: Annals of Data Science, v. 2, no. 2, p. 165–193, doi:10.1007/s40745-015-0040-1.

Yang, T., Y. Cao, H. Friis, K. Liu, Y. Wang, L. Zhou, S. Zhang, and H. Zhang, 2017, Genesis and distribution pattern of carbonate cements in lacustrine deep-water gravity-flow sandstone reservoirs in the third member of the Shahejie Formation in the Dongying Sag, Jiyang Depression, Eastern China: Marine and Petroleum Geology, no. September 2016, p. 0–1, doi:10.1016/j.marpetgeo.2017.11.020.

Zhang, B., S. Guo, F. Li, P. Li, and K. J. Marfurt, 2014, Brittleness evaluation of resource plays by integrating petrophysics and seismic analysis: SEG Technical Program Expanded Abstracts, v. 33, p. 2667–2671, doi:10.1190/SEGAM2014-1679.1.

Appendix

Chapter 2

Well Name	Depth	Thin Section	Facie	XRD			
				TECTOSILICATES (Weight %)	CARBONATES (Weight %)	PHYLLOSILICATES (Weight %)	ADDITIONAL (Weight %)
Frensley E 10X	6275.3	C2	BMdst				
Frensley E 10X	6277.2	C3	BMdst				
Frensley E 10X	6281.1	C4	MCcSt				
Frensley E 10X	6289.7	C5	BMdst				
Frensley E 10X	6292.8	C6	BMdst				
Frensley E 10X	6316.8	C7	MCSt				
Frensley E 10X	6328.3	C8	MCSt	46.2	50.1	3.7	0.0
Frensley E 10X	6358.6	C9	MCSt	63.7	31.7	4.6	0.0
Frensley E 10X	6380.5	C10	BMdst				
Frensley E 10X	6393.6	C11	BMdst				
Frensley E 10X	6434.7	C12	MCcSt	20.5	77.1	2.4	0.0
Frensley E 10X	6472.7	C13	MCSt				
Frensley E 10X	6492.5	C14	MCcSt				
Frensley E 10X	6510.5	C15	AMdst	52.6	3.6	38.7	5.1
Frensley E 10X	6532.2	C16	BMdst	55.1	3.5	37.7	3.7
Frensley E 10X	6539.7	C17	BMdst	49.8	17.0	29.6	3.6
Hunsucker 1101	7215	C18	MCSt				
Hunsucker 1101	7219.4	C19	MCcSt				
Hunsucker 1101	7232.6	C20	MCSt				
Hunsucker 1101	7236	C21	MCSt				
Myrtle Green #23	7294.6	C22	MCcSt				
Myrtle Green #23	7308	C23	AMdst	37.4	8.4	49.4	4.8
Myrtle Green #23	7311.9	C24	MCcSt				
Myrtle Green #23	7328.9	C25	AMdst				
Myrtle Green #23	7337.6	C26	MCcSt				
Myrtle Green #23	7357.4	C27	MCSt				
Myrtle Green #23	7380.5	C28	MCSt				
Myrtle Green #23	7384.7	C29	MCcSt				
Russel B01	7945.4	C30	MCSt	64.6	27.3	7.3	0.8
Russel B01	7948.1	C31	MCSt				
Russel B01	7949.7	C32	MCSt	63.1	29.9	6.3	0.7
Russel B01	7963.8	C33	BMdst	54.6	3.0	39.4	3.0
Russel B01	7965.4	C34	AMdst	53.8	0.0	42.9	3.3
Russel B01	7975.6	C35	AMdst	53.8	11.9	31.4	2.9
Russel B01	7977.5	C36	AMdst	51.0	13.0	32.8	3.2
Russel B01	7979.5	C37	AMdst	47.7	21.2	25.6	5.5
Russel B01	7983.5	C38	BMdst	53.9	8.1	32.5	5.5
Russel B01	7986.4	C39	BMdst	53.7	4.4	39.9	2.0
Russel B01	7990.9	C40	BMdst	49.8	4.0	43.6	2.6
Russel B01	7996.6	C41	BMdst	57.7	4.4	35.1	2.8
Russel B01	7997.6	C42	AMdst	51.5	24.7	19.1	4.7
Russel B01	7999.7	C43	MCcSt				
Russel B01	8007.5	C44	MCSt	62.0	28.6	8.7	0.7
Russel B01	8014.6	C45	MCcSt				
Russel B01	8026.7	C46	MCSt	74.8	17.8	6.0	1.4
Russel B01	8047.2	C47	AMdst	50.5	2.2	44.6	2.7
Russel B01	8050.5	C48	MCSt				

Well Name	Depth	Thin Section	Facie	XRD			
				TECTOSILICATES (Weight %)	CARBONATES (Weight %)	PHYLLOSILICATES (Weight %)	ADDITIONAL (Weight %)
Russel B01	8060.8	C49	AMdst	46.9	20.8	28.4	3.9
Russel B01	8066.4	C50	MCcSt				
Russel B01	8073.5	C51	MCcSt	35.0	58.4	5.9	0.7
Russel B01	8087.9	C52	MCSt				
Russel B01	8096.8	C53	MCcSt	26.2	72.1	1.7	0.0
Russel B01	8105.8	C54	MCSt	56.6	39.8	3.6	0.0
Russel B01	8113.7	C55	MCSt	60.8	33.2	6.0	0.0
Russel B01	8118.7	C56	MCcSt				
Russel B01	8121.4	C57	MCcSt	44.9	50.7	3.6	0.8
Russel B01	8125.5	C58	BMdst	54.8	0.6	42.3	2.3
Russel B01	8129.8	C59	BMdst	54.8	3.6	38.4	3.2
Russel B01	8148.5	C60	MCSt				
Russel B01	8183.5	C61	MCcSt	25.5	71.0	3.5	0.0
Russel B01	8197.6	C62	BMdst	54.7	3.0	4.3	2.0
Russel B01	8200.3	C63	BMdst	54.5	11.0	33.0	1.5
Russel B01	8221.5	C64	BMdst	56.2	7.1	34.1	2.6
Russel B01	8232.3	C65	BMdst	44.5	35.2	18.4	1.9
Moores Est 20A	8761.7	C66	AMdst	40.1	7.5	47.5	4.9
Moores Est 20A	8783.2	C67	BMdst	45.7	22.1	27.3	4.9
Moores Est 20A	8784.8	C68	AMdst	54.0	6.4	38.2	1.4
Moores Est 20A	8785.4	C69	AMdst	57.8	7.0	33.6	1.6
Moores Est 20A	8788.2	C70	AMdst	48.3	20.4	27.7	3.6
Moores Est 20A	8793.7	C71	MCcSt				
Moores Est 20A	8795.4	C72	BMdst	59.7	5.4	32.6	2.3
Moores Est 20A	8807.2	C73	BMdst				
Moores Est 20A	8810.5	C74	BMdst				
Moores Est 20A	8812.1	C75	AMdst	50.1	8.4	34.3	7.2
Moores Est 20A	8815.2	C76	BMdst	48.1	5.9	38.8	7.2
Moores Est 20A	8821.2	C77	MCSt	57.6	35.6	6.1	0.7
Moores Est 20A	8824.1	C78	MCcSt				
Moores Est 20A	8829.5	C79	MCSt				
Moores Est 20A	8837.9	C80	BMdst	52.8	8.1	31.6	7.5
Moores Est 20A	8840.8	C81	BMdst				
Moores Est 20A	8857	C82	BMdst				
Moores Est 20A	8864.7	C83	MCcSt				
Moores Est 20A	8867.1	C84	MCSt				
Moores Est 20A	8876.4	C85	MCcSt				
Moores Est 20A	8881.2	C86	MCSt				
Moores Est 20A	8888.1	C87	MCSt				
Moores Est 20A	8890.2	C88	MCSt				
Moores Est 20A	8894.6	C89	MCSt				
Moores Est 20A	8900.7	C90	MCcSt	36.1	59.0	4.0	1.0
Moores Est 20A	8901.9	C91	MCSt				
Moores Est 20A	8905.2	C92	MCcSt	45.4	51.6	3.0	0.0
Moores Est 20A	8911.9	C93	BMdst				
Moores Est 20A	8918.1	C94	AMdst				
Moores Est 20A	8923.8	C95	BMdst				
Moores Est 20A	8936.3	C96	MCSt				
Moores Est 20A	8940.1	C97	MCSt				
Moores Est 20A	8946.7	C98	BMdst				
Moores Est 20A	8950.9	C99	MCcSt				
Moores Est 20A	8957.7	C100	MCcSt				

				Poro/Perm		Point Counting									
Well Name	Depth	Thin Section	Facie	Permeability Klinkenberg (millidarcys)	Ambient Porosity (%)	Detrital Grains	Pelletes	Other allochems	Other grains	Intergranular Calcite	Opaques	Other Cement	Secondary Porosity	Intergranular Porosity	TOTAL
Frensley E 10X	6275.3	C2	BMdst	0.0061	6.963856392										
Frensley E 10X	6277.2	C3	BMdst	0.161	8.024702562										
Frensley E 10X	6281.1	C4	MCcSt	0.0001	3.008654727	29	23	0	1	42	1	2	2	0	100
Frensley E 10X	6289.7	C5	BMdst	0.093	6.226215232										
Frensley E 10X	6292.8	C6	BMdst	0.0095	5.271394231										
Frensley E 10X	6316.8	C7	MCSt	0.0039	8.315129386	35	13	0	0	33	3	4	12	0	100
Frensley E 10X	6328.3	C8	MCSt	0.089	12.49087333	35	7	0	0	36	2	2	18	0	100
Frensley E 10X	6358.6	C9	MCSt	0.094	14.77766349	42	6	0	0	22	5	4	20	0	100
Frensley E 10X	6380.5	C10	BMdst	0	0										
Frensley E 10X	6393.6	C11	BMdst	0.036	4.30165501										
Frensley E 10X	6434.7	C12	MCcSt	0.0004	2.905550253	21	27	1	1	44	2	0	5	0	100
Frensley E 10X	6472.7	C13	MCSt	0.0047	6.506440088										
Frensley E 10X	6492.5	C14	MCcSt	0.0064	5.780453768	23	22	0	0	42	2	0	10	0	100
Frensley E 10X	6510.5	C15	AMdst	+	4.294352073										
Frensley E 10X	6532.2	C16	BMdst	+	3.910074079										
Frensley E 10X	6539.7	C17	BMdst	+	3.899812185										
Hunsucker 1101	7215	C18	MCSt	0.007	6.863879932	47	9	0	0	30	3	1	11	0	100
Hunsucker 1101	7219.4	C19	MCcSt	0	0	22	23	0	0	46	1	0	8	0	100
Hunsucker 1101	7232.6	C20	MCSt	0.0046	7.983602882	26	21	1	0	34	1	0	13	4	100
Hunsucker 1101	7236	C21	MCSt	0.0014	4.429990387	40	19	0	0	29	2	0	10	0	100
Myrtle Green #23	7294.6	C22	MCcSt	0.0003	3.429559847	29	28	0	0	43	0	0	0	0	100
Myrtle Green #23	7308	C23	AMdst	+	6.476993006										
Myrtle Green #23	7311.9	C24	MCcSt	0.0003	1.891872744	25	32	1	0	42	0	0	0	0	100
Myrtle Green #23	7328.9	C25	AMdst	0.0000079	1.141159415										
Myrtle Green #23	7337.6	C26	MCcSt	0.000017	0.892679886	36	21	1	0	41	2	0	0	0	100
Myrtle Green #23	7357.4	C27	MCSt	0.0064	8.325461055	52	3	0	0	29	3	4	7	0	100
Myrtle Green #23	7380.5	C28	MCSt	0.0058	8.130884037	36	10	0	0	39	2	5	8	0	100
Myrtle Green #23	7384.7	C29	MCcSt	0.0009	3.834731027	26	22	0	0	44	1	0	6	0	100
Russel B01	7945.4	C30	MCSt	0.0038	4.739591918	58	12	0	1	19	2	2	6	0	100
Russel B01	7948.1	C31	MCSt	0.0019	3.800586049	51	19	0	0	26	1	1	2	0	100
Russel B01	7949.7	C32	MCSt	0.0048	5.339576813	53	19	0	0	23	1	0	3	1	100
Russel B01	7963.8	C33	BMdst	0.0022	3.054960508										
Russel B01	7965.4	C34	AMdst	0	0										
Russel B01	7975.6	C35	AMdst	0.00001	0.209751265										
Russel B01	7977.5	C36	AMdst	0.0000025	0.246016861										
Russel B01	7979.5	C37	AMdst	0.0000023	0.238951433										
Russel B01	7983.5	C38	BMdst	0.0000027	0.828532405										
Russel B01	7986.4	C39	BMdst	0.012	0.639178496										
Russel B01	7990.9	C40	BMdst	0.0021	1.412850204										
Russel B01	7996.6	C41	BMdst	0.033	0.217514408										
Russel B01	7997.6	C42	AMdst	0.0000013	0.276967327										
Russel B01	7999.7	C43	MCcSt	0.148	1.3660816	35	23	0	0	40	1	0	0	0	100
Russel B01	8007.5	C44	MCSt	0.0011	4.530681939	57	14	0	0	22	3	0	4	0	100
Russel B01	8014.6	C45	MCcSt	0	0	28	23	1	0	47	1	0	0	0	100
Russel B01	8026.7	C46	MCSt	0.0006	4.835680693	56	2	0	0	22	8	9	3	0	100
Russel B01	8047.2	C47	AMdst	0	0										
Russel B01	8050.5	C48	MCSt	0.0027	4.567860765	58	12	0	0	23	2	1	4	0	100

				Poro/Perm		Point Counting									
Well Name	Depth	Thin Section	Facie	Permeability Klinkenberg (millidarcys)	Ambient Porosity (%)	Detrital Grains	Pelletes	Other allochems	Other grains	Intergranular Calcite	Opaques	Other Cement	Secondary Porosity	Intergranular Porosity	TOTAL
Russel B01	8060.8	C49	AMdst	0.0000017	0.24197096										
Russel B01	8066.4	C50	MCcSt	0.0071	1.725844407	21	27	0	0	49	1	0	1	0	100
Russel B01	8073.5	C51	MCcSt	0.0031	5.863501808										
Russel B01	8087.9	C52	MCSt	0.0022	5.796863332	42	6	0	1	33	3	3	12	1	100
Russel B01	8096.8	C53	MCcSt	0.000046	2.005068682	20	26	1	0	51	0	0	2	0	100
Russel B01	8105.8	C54	MCSt	0.0099	6.755173448	40	5	0	0	37	5	2	11	0	100
Russel B01	8113.7	C55	MCSt	0.031	9.400011465	41	4	0	0	35	3	2	15	0	100
Russel B01	8118.7	C56	MCcSt	0.0003	2.353193131	24	25	0	0	47	2	0	2	0	100
Russel B01	8121.4	C57	MCcSt	0.000015	1.858667961	32	23	0	0	41	2	1	1	0	100
Russel B01	8125.5	C58	BMdst	+	2.858940278										
Russel B01	8129.8	C59	BMdst	0.053	1.32077261										
Russel B01	8148.5	C60	MCSt	0.0004	2.456217293	43	21	0	0	33	3	0	0	0	100
Russel B01	8183.5	C61	MCcSt	0.0003	1.453248293	26	31	0	0	42	0	0	0	0	100
Russel B01	8197.6	C62	BMdst	0.185	1.78081495										
Russel B01	8200.3	C63	BMdst	0.035	0.239029826										
Russel B01	8221.5	C64	BMdst	0	0										
Russel B01	8232.3	C65	BMdst	0.00000029	0.220023896										
Moores Est 20A	8761.7	C66	AMdst	0.0025	1.493005218										
Moores Est 20A	8783.2	C67	BMdst	0.000019	0.248905814										
Moores Est 20A	8784.8	C68	AMdst	0.0043	0.297775564										
Moores Est 20A	8785.4	C69	AMdst	+	0.282738587										
Moores Est 20A	8788.2	C70	AMdst	0.0001	0.270712522										
Moores Est 20A	8793.7	C71	MCcSt	0.000011	0.637943301	37	19	0	0	41	3	0	0	0	100
Moores Est 20A	8795.4	C72	BMdst	0.032	5.734078414										
Moores Est 20A	8807.2	C73	BMdst	+	4.074100528										
Moores Est 20A	8810.5	C74	BMdst	0.0094	4.337669069										
Moores Est 20A	8812.1	C75	AMdst	0.0002	0.831903173										
Moores Est 20A	8815.2	C76	BMdst	+	1.608209093										
Moores Est 20A	8821.2	C77	MCSt	0.0006	3.37033398	46	14	0	0	31	9	0	0	0	100
Moores Est 20A	8824.1	C78	MCcSt	0.0009	3.218063493	41	10	0	0	40	10	0	0	0	100
Moores Est 20A	8829.5	C79	MCSt	0.0005	3.136591611	54	12	0	0	25	8	1	0	0	100
Moores Est 20A	8837.9	C80	BMdst	0.022	0.44841679										
Moores Est 20A	8840.8	C81	BMdst	0.0009	4.88071282										
Moores Est 20A	8857	C82	BMdst	0.0054	4.239403476										
Moores Est 20A	8864.7	C83	MCcSt	0.0002	1.381968609	30	19	7	0	42	2	0	0	0	100
Moores Est 20A	8867.1	C84	MCSt	0.0001	3.365365254	48	13	0	0	35	3	1	0	0	100
Moores Est 20A	8876.4	C85	MCcSt	0.00002	1.81191186	27	24	0	0	48	1	0	0	0	100
Moores Est 20A	8881.2	C86	MCSt	0.0015	5.565877853	53	13	0	0	20	4	5	4	0	100
Moores Est 20A	8888.1	C87	MCSt	0.0024	5.783587922	61	4	0	0	22	8	5	2	0	100
Moores Est 20A	8890.2	C88	MCSt	0.0014	4.565081567	52	5	0	0	31	6	5	1	0	100
Moores Est 20A	8894.6	C89	MCSt	0.0015	4.631162978	57	3	0	0	27	7	3	1	1	100
Moores Est 20A	8900.7	C90	MCcSt	0.000033	1.548742537	33	23	0	0	43	1	0	0	0	100
Moores Est 20A	8901.9	C91	MCSt	0.0003	3.368579482	51	6	0	0	29	9	5	0	0	100
Moores Est 20A	8905.2	C92	MCcSt	0.0039	0.439644558	35	18	0	0	44	2	1	0	0	100
Moores Est 20A	8911.9	C93	BMdst	0.00002	0.722575889										
Moores Est 20A	8918.1	C94	AMdst	0	0										
Moores Est 20A	8923.8	C95	BMdst	0.0000054	2.969815204										
Moores Est 20A	8936.3	C96	MCSt	+	2.752200292	46	7	0	1	34	6	5	1	0	100
Moores Est 20A	8940.1	C97	MCSt	0.0001	3.268542018	46	8	0	0	42	3	1	0	0	100
Moores Est 20A	8946.7	C98	BMdst	0.0006	4.068179894										
Moores Est 20A	8950.9	C99	MCcSt	0.0003	3.444067472	36	12	0	0	50	1	0	1	0	100
Moores Est 20A	8957.7	C100	MCcSt	0.0002	3.041449894	29	7	0	0	54	1	4	5	0	100

Chapter 4

Zone	Petrofacies	Major (ft)	Minor (ft)	Vertical (ft)	Azimuth (degrees from N)
	Mdst	2500	2000	15	315
Sycamore E	MCcSt	2500	2000	10	315
	MCSt	2500	2000	12	315
	Mdst	2500	2000	15	315
Sycamore D	MCcSt	2500	2000	13	315
	MCSt	2500	2000	12	315
	Mdst	2500	2000	25	315
Sycamore C	MCcSt	2500	2000	13	315
	MCSt	2500	2000	13	315
	Mdst	2500	2000	16	315
Sycamore B	MCcSt	2500	2000	12	315
	MCSt	2500	2000	8	315
	Mdst	2500	2000	8	315
Sycamore A	MCcSt	2500	2000	8	315
	MCSt	2500	2000	7	315