Liquid Flow and Gas Diffusion in Natural Rocks: Experimental and Simulation Studies

By

Xuewei Ning

DISSERTATION

Submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy at

the University of Texas at Arlington

August 2023

Abstract

Fluid Flow and Gas Diffusion in Natural Rocks: Experimental and Simulation Studies

Xuewei Ning

The University of Texas at Arlington. 2023

Supervising Professor: Qinhong Hu

Natural rocks, as key components in energy and environmental geosciences, play crucial roles in groundwater extraction, geological storage of high-level nuclear waste, geothermal energy mining, and petroleum exploration in both conventional and unconventional reservoirs. Unconventional reservoirs, specifically the tight shale formations, have seen a significant usage in petroleum production in recent decades, largely due to advances in fracturing stimulation techniques which counter the inherently low permeability and connectivity of shale reservoirs by expanding natural fractures and creating artificial ones for an enhanced petroleum production. However, the low mobility (liquid flow and gas diffusion) of stored fluids in naturally fractured formations with tight matrix blocks presents a challenge; consequently, understanding liquid flow and gas diffusion in fractured low-permeability media has become a paramount issue in porous media studies.

This research combines integrated methodologies of experimental and simulation studies (numerical and machine learning regression models) to delve into the processes of spontaneous imbibition and gas diffusion, respectively, within various porous natural rocks such as shales, sandstones and carbonate rocks.

For the imbibition analysis, several Barnett Shale samples with fractures were examined with flow direction oriented either parallel (P) or transverse (T) to the bedding plane. A newly developed model, integrating concepts of percolation theory implemented via MATLAB, allows for capturing 3D porous media imbibed with 2D fractures, an innovation over previous piston-like or multiply-sized pore models. The model incorporates the complex interplay of factors, such as porosity, fracture distribution, and pore connectivity. Results demonstrate a sensitivity of imbibed water mass to the number of fractures directly connected to the water source, importing a novel parameter for understanding the wetting-front progression in fractured tight shale.

Subsequently, we utilize machine learning regression techniques, specifically linear regression, Gradient Boosting regression, Decision Tree regression, and Random Forest regression, to predict gas diffusion across 15 different types of rocks with both heterogenous and homogeneous structures. We prioritize attributes such as rock type, sample radius, sample height, porosity, permeability, and tracer concentration difference for their impacts on diffusion velocity and the feature importance. Notably, the Random Forest model identifies the tested gas concentration difference as the most significant factor in affecting gas diffusion.

Our dual-method (experimental and simulation) approach offers valuable insights into liquid flow and gas diffusion behavior in natural rocks, providing a platform for their targeted usage in energy and environmental geosciences. Machine learning models can expedite and economize the testing process, highlighting their potential to enhance operational efficiency in the studies of porous media.

This dissertation is structured into five chapters. The first chapter (Chapter I) provides an overview and succinct introduction to the contents of the subsequent chapters. Chapter II features a paper that has already been published, showcasing the liquid imbibition into fractured Barnett

Shale. In Chapter II, the research elucidates the intricate imbibition mechanism prevalent in fractured shale formations. Additionally, it offers a versatile model suitable for Barnett Shale lithology, delineating the spatial-temporal dynamics during fluid imbibition within porous substrates. Chapter III scrutinizes the impact of pore throat distribution and mineralogical constitution on the velocity and amplitude of spontaneous fluid imbibition across a diverse range of natural lithologies. Chapter IV elucidates the proficiency and precision of machine learning models in predicting the gas diffusion process within natural rock matrices. The findings underscore the Random Forest model's aptitude in delineating feature significance and quantifying the contributing attributes for gas diffusion. Chapter V draws conclusions by synthesizing the findings and implications of this research during the duration of this PhD program.

Acknowledgements

I would like to extend my deepest gratitude to my advisor, Dr. Hu, whose guidance, support and mentorship have been invaluable throughout my doctoral journey. Dr. Qinhong Hu, your wisdom and dedication have made a significant impact on my academic career. Your teachings have not only shown me the practical aspects of conducting experimental work but also educated me with the patience and resilience necessary for scientific exploration. I am deeply grateful for your confidence in me, and for providing me the opportunity to pursue a doctoral degree. Your mentorship has been instrumental in shaping my academic journey and for that, I am eternally grateful.

In addition, I would like to express my heartfelt gratitude to my friends, Yujing Yang, Dr. Yifei Yu, Liuting Shang, Lei Di, Xiao Shi, Binqian Yin, Zehao Ye, Siyu Zhou, Bangde Liu, Haiqi Zhang, Zhengyuan Zhu, Chenxi Ma, Caifang Wu, Boen Su, Jie Han. Thank you, Jessica, Marshell, Hasnain, Dr. Lingyiqian, Luo. Your camaraderie and friendship have been a tremendous source of joy and relaxation, providing a perfect balance to my academic pursuits. The matches we played and the moments we shared together have been invaluable in maintaining my mental and physical well-being throughout this intense journey. I also want to extend thanks to the many other friends whose names may not be mentioned here but have significantly contributed to making my time as a doctoral student more enjoyable and fulfilling. Your collective support and encouragement have been integral to my academic achievements.

Furthermore, I will thank a thousand times to my husband, Dr. Yunxiang Bai. Your technical assistance and emotional support have been pillars of strength for me. You have filled my life with happiness and cherished memories. I am truly grateful to have you and our cat, Baozi, by my side.

Last, I owe a profound debt of gratitude to my parents, Qinghua Zhang and Xinliang Ning. Your unwavering love and patience have been my guiding lights. Your faith in me and your encouragement have propelled me forward in times of doubt. Indeed, I could not have asked for more caring and supportive parents. Your love and sacrifices have played an instrumental role in shaping the person I am today.

Dedication

To Dr. Yunxiang Bai

To my parents, family members and friends

Table of Contents

List of Tables

Chapter II

Chapter III

Chapter IV

List of Figures

Chapter II

Chapter III

Chapter IV

Chapter I: Introduction

Studying fluid liquid flow and gas diffusion in natural rocks holds significant importance across various scientific and industrial domains. This research area plays a crucial role in understanding the behavior of fluids, such as water, oil, and gas, within subsurface rock formations. The primary motivations behind studying these phenomena through experiments and simulations are to unravel the intricate processes that occur in subsurface reservoirs, enhance resource extraction techniques, and improve our comprehension of geological systems.

One key reason for investigating fluid flow and gas diffusion in natural rocks is related to hydrocarbon exploration and production. Oil and gas reservoirs are often found within porous rock formations, and the movement of fluids within these rocks greatly influences the efficiency of resource recovery. By comprehending the intricate mechanisms of fluid flow and diffusion, researchers can optimize drilling, extraction, and reservoir management strategies. This knowledge is pivotal in maximizing the recovery of valuable energy resources, reducing operational costs, and prolonging the lifespan of reservoirs.

Additionally, understanding fluid behavior in natural rocks is essential for groundwater management and environmental monitoring. Many regions rely on underground aquifers as primary sources of freshwater. The study of fluid movement within rock formations helps predict the movement of contaminants and pollutants, enabling more effective strategies for groundwater protection and sustainable resource management. This knowledge is crucial for safeguarding ecosystems and ensuring the availability of clean water supplies for both current and future generations.

Experiments and simulations serve as invaluable tools in this realm of study. Experiments conducted in controlled laboratory settings provide insights into fundamental fluid-rock interactions under various conditions. These experiments allow researchers to isolate specific variables, test hypotheses, and validate theoretical models. Simulations, on the other hand, utilize advanced computer models to recreate complex subsurface environments virtually. This approach enables researchers to explore scenarios that might be impractical or impossible to replicate in a lab, such as investigating the long-term effects of fluid injection or migration within rock formations.

Incorporating both experiments and simulations in research on fluid flow and gas diffusion in natural rocks leads to a synergistic approach. Experiments provide empirical data that validate and refine simulation models, while simulations offer insights into macroscopic behaviors and long-term effects that are difficult to capture in experiments alone. Together, these methods contribute to a deeper understanding of geological systems, aid in optimizing resource extraction techniques, and inform strategies for sustainable resource management and environmental protection.

This research presents the investigation of fluid flow behaviors in a range of natural rocks with different physical and structural characteristics. The test samples include shale, carbonates, sandstones, limestones, and dolomite which are located across the United States. The fluid flow behavior in imbibition tests and simulation results for Barnett Shale is presented in Chapter II. The investigation of imbibition test for various kinds of rocks and related simulation models is presented in Chapter III, and the gas diffusion and the model simulation are presented in Chapter IV.

Chapter II presents the simulation model of liquid imbibition tests in Barnett Shale. Imbibition experiments have been instrumental in examining petrophysical properties of porous media, but current empirical, analytical, and numerical models are limited due to their simplification of fluid flow paths and the neglect of irregular fractures. To address this, this research constructed a numerical model replicating laboratory-scale experiments, simulating the imbibition process in fractured Barnett Shale samples with MATLAB. The model revealed a steady increase of imbibition rates in samples with fractures parallel to the bedding plane (P-direction), while shale samples with transverse fractures (T-direction) exhibited a decrease in imbibition rate upon initial saturation due to low pore connectivity. The wetting phase movement (WPM) model, accounting for 3D porous media with 2D fractures, also highlighted the importance of fracture parameters in influencing imbibition rates. Ultimately, the study sheds light on the complex imbibition process in fractured shale and provides a model adaptable for other rock types to depict spatial-temporal behavior during a dynamic liquid imbibition in porous media.

Chapter III expands the investigation by conducting imbibition tests in a diverse set of 15 natural rocks, each with unique physical attributes, including porosity, permeability, heterogeneity, and wettability (quantified by the contact angle). This chapter also explores the mineral composition via X-Ray Diffraction (XRD) and assesses the pore size distribution utilizing Mercury Intrusion Porosimetry (MIP). In terms of the computational techniques, the pore size distribution data are integrated into the machine learning models, which assists in quantifying the influence of pore size on imbibition rates, thus highlighting its critical role in fluid flow dynamics.

Chapter IV discussed the gas diffusion in above 15 natural rocks and applied machine learning models to predict the gas diffusion behaviors. Gas diffusion plays a pivotal role in petroleum exploration and production. This study harnesses machine learning techniques, specifically linear

regression, Gradient Boosting regression, Decision Tree regression, and Random Forest regression, to predict gas diffusion across 15 distinct rock types encompassing both heterogeneous and homogeneous structures. The comparative application of these regression models has enabled the ranking of critical attributes influencing gas diffusion rates, including rock type, sample size (radius and height), porosity, permeability, and tracer-gas concentration difference. Our findings underscore the efficacy of the Random Forest model in identifying key determinants of gas diffusion, with gas concentration difference and sample size emerging as the most influential factors. Consequently, these predictive machine learning methodologies offer valuable tools for minimizing experimental tests and enhancing reservoir management such as efficient oil and gas production, by illuminating the intricacies of gas diffusion within diverse rock structures.

Chapter II:

A New Model for Simulating the Imbibition of a Wetting-Phase Fluid in a Matrix-Fracture Dual Connectivity System

Xuewei Ning[a], Robert Ewing[b], Qinhong Hu[a]*, Qiming Wang[a], and Xiaoming Zhang [a,c,d]

[a]Department of Earth and Environmental Sciences, The University of Texas at Arlington, TX 76019, USA

[b]The Climate Corporation, 710 2nd Ave., Seattle, WA 98104, USA

[c]Key Laboratory of Tectonics and Petroleum Resources, Ministry of Education, China University of Geosciences,

Wuhan 430074, China

[d]Faculty of Earth Resources, China University of Geosciences, Wuhan 430074, China

Published in

Geofluids

Corresponding author: maxhu@uta.edu

Keywords: Spontaneous imbibition, fluid flow, Barnett Shale, heterogeneity, percolation theory, probability

Abstract

Imbibition experiments are a well-established method for evaluating the petrophysical properties of porous media, having seen extensive use in the past decade. Numerous empirical, analytical, and numerical models have been constructed to simulate the spontaneous imbibition of the wetting phase fluid into these porous media, yet they exhibit inherent limitations. Traditional studies have often oversimplified the imbibition process, either portraying it as a piston-like displacement or modeling the porous medium as interconnected multi-sized pores. These approaches overlook the impact of irregular fractures and nonuniform flow paths within the matrix, rendering the results incomplete. In our study, we utilized a numerical model to replicate laboratory-scale data and conducted imbibition tests on several fractured Barnett Shale samples with fractures either parallel (P) or transverse (T) to the bedding plane. A MATLAB-built model was then implemented, integrating the imbibition process in fractures and the matrix using percolation theory principles. Our findings indicate a more consistent increase in imbibition rates for rocks with P-direction fractures compared to those with T-direction fractures. This is attributed to the low pore connectivity within the shale matrix, which restricts the upward movement of water, leading to a sudden decrease in the imbibition rate for shales with T-direction fractures once the bottom layer becomes saturated. Our Wetting Phase Movement (WPM) model successfully simulates 3D porous media with 2D fractures and reveals a direct relationship between the rate of imbibition and the physical parameters of the fractured porous media such as porosity and fracture distribution. By employing Monte Carlo methods, we examined the influence of these parameters on imbibition, predicting elapsed time and cumulative water imbibition in the Barnett Shale samples. It was observed that the rate of water absorption is particularly sensitive to the number of fractures directly connected to the water

source and the connectivity between neighboring grid cells, both of which are crucial for wetting-front progression. While our findings are based solely on Barnett Shale samples, the developed model can be adapted to other fractured rocks, offering insights into the spatial and temporal behavior during dynamic imbibition processes, often elusive in experimental setups.

1. **Introduction**

Many reservoirs have been utilized in various fields of energy and environmental geosciences, such as groundwater exploitation, safe storage of high-level nuclear waste in geological repository, geothermal mining, and petroleum exploration in both conventional and unconventional rock formations. As these reservoirs form, regularly and irregularly sized fractures at different scales and orientations could be developed, providing preferential flow pathways for transmitting the petroleum and water (Roshan et al., 2015; Zhang et al., 2017). Unconventional reservoirs show the significance in petroleum production in recent decades after stimulated fracturing of tight shales (Akbarabadi et al., 2017; Gombia et al., 2008). Because of the extremely low permeability and connectivity of shale reservoirs, some field-scale stimulation techniques are necessary to expand the natural, and initiate artificial, fractures, in order to enhance the petroleum production from the shale (Chalmers et al., 2012; Cipolla et al., 2009). As naturally fractured formation with the tight matrix block can cause the low mobility of stored fluids (Dejam et al., 2018), a better understanding of how the fluid flows in fractured low-permeability media has been a challenging problem in the porous media studies. Spontaneous imbibition is a critical process during the hydraulic stimulation and wells shut-in for production preparation, which cause over 50% of the injected fluids to flow through the matrix and fractures by capillary force (Alkouh et al., 2014). This displacement process affects many reservoir properties, for example, the relative permeability and consequent petroleum movement into the

producing well (Gombia et al., 2008). Thus, it is vital to investigate the performance of shales during imbibition process and simulate the fluid flow in shale formations (Zeng et al., 2020).

It is important to clarify the invasion percolation-like motion evolution of the wetting fluid during imbibition into fractured porous media, where the viscous, gravitational, and capillary forces are the main drivers for fluid flow. Dominated by interfacial phenomena, the capillary-driven flow is extremely important in nm-sized pore system, influenced by the flow pathways of connected pore space (Gao et al., 2016). The first imbibition model in the petroleum field was developed to predict oil recovery (Naar and Henderson, 1961). For consolidated sand, relative permeability curves showed that, at any given saturation, the permeability of oil during imbibition is smaller than during drainage stages (Naar and Henderson, 1961; Gao et al., 2018). The theoretical relationship between drainage and imbibition was applied to show relative permeability characteristics but failed to consider heterogeneity of fractures and the corresponding relationship with the matrix.

An analytical solution, including the well-known Lucas Washburn equation (Lucas, 1918; Washburn 1921), was proposed for capillary rise in tubes when studying imbibition phenomena. Eden's growth model (Eden, 1961) was proposed to describe the specific types of two-dimensional cluster growth, in which clusters grow stochastically based on accumulation of materials on the boundary instead of the inside. The growth process along with the surface, especially in some natural processes, has inspired many researchers to focus on the growth site, and which particles are the new ones to be added into the whole system. The study of diffusion limited aggregation reflected the stochastic nature of particle walk (Witten and Sander, 1981). A third growth model, invasion percolation, was motivated by fluid displacement under capillary forces in porous media (Wilkinson and Willemsen, 1983). Besides two-phase flow, three-phase

flow in porous media was investigated; a model was built for nonaqueous phase liquid flow instead of allowing all phases to flow, and it only can be applied to two-dimensional rock with uniform pore throat and body sizes (Fenwick and Blunt, 1998).

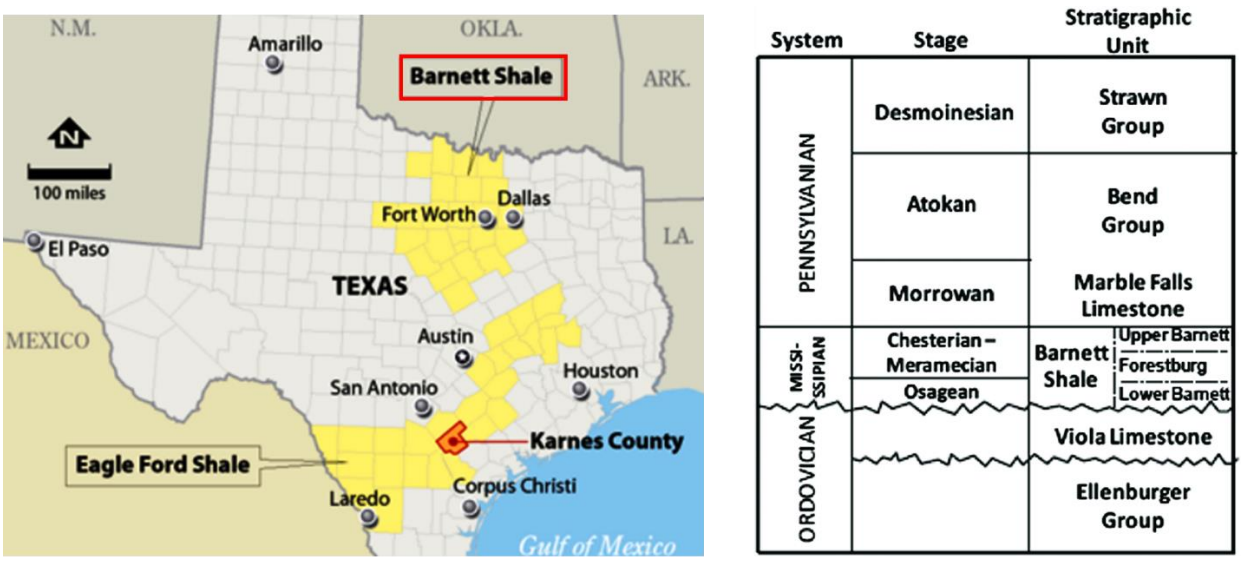
In response to the shale revolution since the first decade of the 21st century, spontaneous imbibition into low permeability shale has been attracting much attention (Cai et al., 2014; Dehghanpour et al., 2012; Hu et al., 2012; Stavropoulou et al., 2020). A two-dimensional model was presented to simulate the imbibition process in heterogeneous porous media, by tracking the moving wetting front with a new interface integral method (Suo et al., 2019). Based on Hagen-Poiseuille flow, spontaneous imbibition in tortuous capillaries with variably shaped apertures was simulated by Cai (Cai et al., 2014), and their model also considered different geometrical factors as the limitations while those variabilities may cause the changes of imbibed water. Based on the model of Cai (Cai et al., 2014), the model of Wang et al. (2018) for porous media in tight reservoirs calculated the imbibition length, which is hard to measure experimentally; however, as the model took more geometrical parameters into account than other imbibition models, even a small changeable reason may lead to differences in results. However, over-parametrization and complex governing equations in a model create a computational burden, especially at large scales of investigation. The rock properties may vary widely even over short distances, such that a model having uniform properties will be unable to fit field data properly, and an overparametrized model will lack flexibility and predictability. An analytical model focused on the imbibition process for confined nano-fractures in shales, as Zeng et al. (2020) modified the Navier Stokes equation and considered the whole sample with same and uniform physical properties instead of discriminating the difference between the matrix and microfractures. In microscale studies, the multicomponent flow in heterogeneous porous media

was simulated by the modified lattice Boltzmann method (Spaid and Phelan, 1998). Extending the continuum numerical framework to complex domains is necessary, especially among low-permeability porous media (Liu and Gan, 2019).

Several papers reported that a disconnection of the pore network usually occurs only in media with low porosity (Bernabé et al., 2010; Madden, 1983; Mavko and Nur, 1997), but some occasional reasons may cause the successive probability of water to pass through. Thus, fluid flow and chemical transport in porous media with low connectivity can be described using percolation theory, which contains lattice and bonds stochastically (Ewing and Horton, 2002; Ewing et al., 2010). Furthermore, neutron and X-ray tomography experiments were used to study the fracture development of Callovo Oxfordian claystone and provided a feasible approach to visualizing the development of water-induced fractures during imbibition (Stavropoulou et al., 2019), but recent simulators still lack the visualization of imbibition process in three-dimensional (3D) scales (Zeng et al., 2020, Meng et al., 2019; Wang et al.,2019; Wang and Zhao, 2021; Wang et al., 2020).

2 Samples

The WPM model was used to describe and interpret the laboratory imbibition tests of Barnett Shale. Located in the Bend Arch of the Fort Worth Basin, the Barnett Shale is known as a low-permeability shale gas reservoir. This field covers 7.08×10^{10} m^3 with abundance of natural gas and oil as a commercial exploration place (Wang et al., 2020). Barnett Shale samples were taken from an outcrop located in the Fort Worth Basin, a regional syncline of north-central Texas.



(A) (B)

Figure 1: (A) Barnett Shale location in Texas (map credit: Paul Horn/Inside Climate News, modified from David Hasemyer); (B) Simplified stratigraphic column showing the Barnett Shale and adjacent strata in the northeastern part of the Fort Worth Basin. The upper Barnett, Forestburg Limestone, and lower Barnett are all informal members of the Barnett Shale Formation (modified from Pollastro et al., 2007).

As presented in Figure 1, this work will start with samples and methods for the Barnett Shale and imbibition experiments as well as basic theory and assumptions of modeling concepts and implementation, continue to mathematical modeling for two- and three-dimensions and simulations with fractured Barnett Shale with either P- or T -direction fractures, proceed to model predictions with cubic and cylindrical sample shapes of the Barnett Shale, and end with summary and conclusions. The description of Barnett Shale core samples and its properties have been provided by Loucks and Ruppel (Loucks and Ruppel, 2007) and Gao and Hu (2016). The samples used for imbibition tests were extracted from those layers parallel (P) or transverse (T) to the bedding plane, which are termed as P- or T-fractured samples. Firstly, we measured the

physical properties of Barnett Shale, such as porosity, permeability, fracture numbers and aperture widths, and air-liquid contact angle. Secondly, we took several pairs of P- or T-fractured samples to carry out the imbibition tests. The procedures and corresponding data processing of the imbibition test were presented in previous papers (Hu et al., 2001; Hu et al., 2005; Hu et al., 2012; Gao and Hu, 2012; Wang et al., 2021). The cylindrical samples are around 15 mm in diameter and 20 mm in height. Every imbibition test was not ended until the additional mass of imbibed water can be ignored. Thirdly, after testing several pairs of fractured shale samples, a numerical model, combining the spontaneous water movement in fractures and in the matrix by capillary force, was built as below.

3. Hypothesis

1) The cumulative water imbibed relates to the size of rock samples and its porosity.
2) The rate of water imbibition relates to the fracture distribution, width, numbers, and the level of connectivity. Larger, denser, well-connected fractures may accelerate the imbibition.
3) The P-fractures accelerate the water imbibition while T-fractures hamper the same process.
4) The probability has an impact on imbibition. The probability of water going through a matrix grid cell with extremely low porosity and permeability can be considered as 0.

4. Basic Theory and Assumptions

The imbibition process, in which a wetting phase fluid displaces a non-wetting phase fluid, is driven by capillary forces (Alkouh, 2014; Gao and Hu, 2016; Shi et al., 2019). Fig. 2A illustrates the distribution of pores and the matrix. For better simulating the water imbibition into rock samples, re-construction and simplification of the connectivity relationship between pores and

fractures are significant to imbibition simulation. Fig. 2B reflects the diversity of pore sizes and pore throats in 3D fashion for a shale sample. But the simulated results cannot be guaranteed if the computational burden is too heavy for calculation processes. Thus, finding the balance point between reality and efficiency is significant for simulation builders.

At the start of the imbibition test, the rock sample dried under 60^{o}C for two days and no water has yet imbibed. With time, water penetrates into the samples though the matrix and fractures, as shown by the growing dark regions. The hydraulic conductivity of fractures is much (~5-6 orders of magnitudes) larger than that of the tight matrix, so the rate of water imbibition in fractures is much faster. Thus, the pore size diversity of the rock can become a problem for the model. The major complex problem is how to simulate the imbibition process which may be affected by fracture aperture and distribution. This research will investigate the influence of fractures and pore properties on the imbibition process, especially in terms of rate of imbibed water over time.

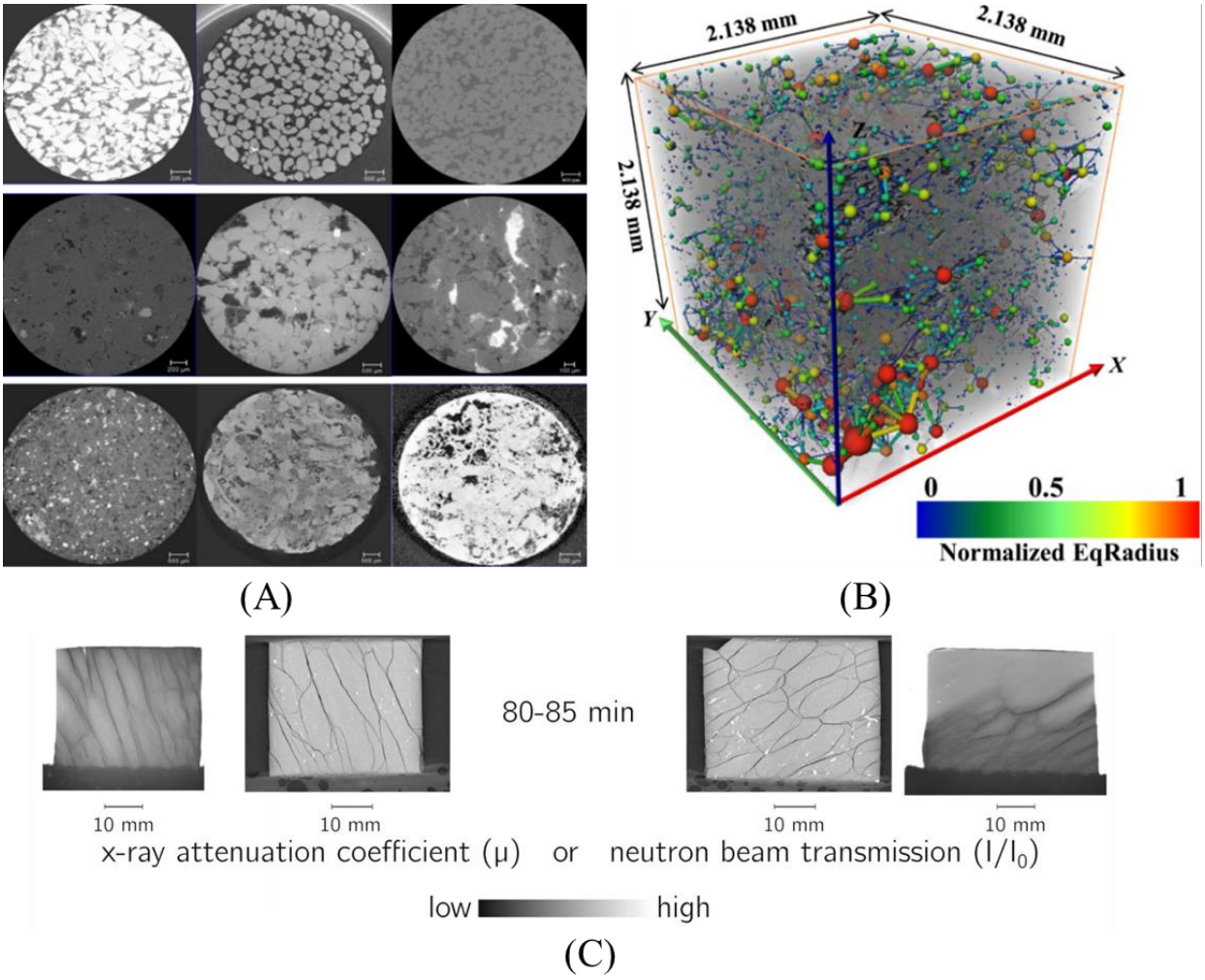


Figure 2: (A) Cross-section of original CT (computed tomography) image (Song et al., 2019); (B) Extracted pore network model from shale rock sample S9 (Song et al., 2019); (C) X-ray and neutron tomography images of imbibition process over 80-85 mins (Stavropoulou et al., 2019); P or T: parallel or transverse to the bedding plane.

In this study, the model combines the capillary force and spontaneous imbibition process in both the matrix and fractures. The virtual view of rock sample construction with grid cells is shown in Fig. 3. At the start of the imbibition tests, rock samples had an initial water saturation of "zero" (or residual saturation). The rock sample was connected to the electronic balance, with only its bottom face contacting the water inside a closed chamber (Wang et al., 2020). The sample was

coated on the wall and loosely covered on the top face to minimize vapor absorption inside the chamber full of humid air with water reservoir The experimental data used in this study were corrected for submergence depth decline and associated buoyancy gain by the method of Hu et al. (2012).

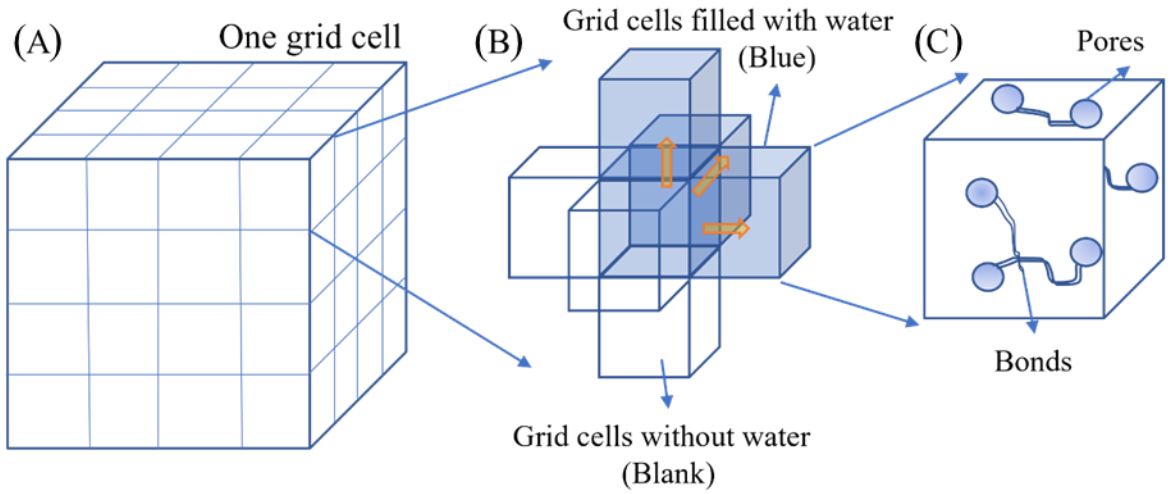


Figure 3: Construction of rock samples with grid cells: (A) Virtual cubic rock sample with grid cells; (B) Water moves from one grid cell to three stochastic directions in one step (between grid cells); (C) One grid cell with pores and bonds (void space connected with other pores).

The assumptions for the WPM model include the following:

(1) For the matrix of the Barnett Shale, the average porosity is around 13% and permeability ranges from 7 to 50 nano-Darcy. The fracture permeability, obtained from the parallel plate model (Ning et al., 2020), is six orders of magnitude larger than the matrix. Because the cylindrical samples are small (around 2.0 cm in diameter and 1.5 cm in height) the WPM model considers the matrix to be homogeneous, especially under the condition of much disparate permeability between fractures and the matrix.

(2) The rock sample can be divided arbitrarily into equal-sized cubic elements (grid cells), with each grid cell assumed to contain the same volume of pores.

(3) The volume of a fracture depends upon its aperture, width and length.

(4) The bonds need to exist to make pore bodies connect with each other. Bond volume can be considered to be part of the pore's volume, so it needn't be considered separately.

(5) The aperture widths can be drawn at random from a distribution that is independently derived from thin section petrography.

(6) The main driving mechanism for wetting-phase imbibition is capillary force. Other driving mechanisms, such as water adsorption by clay minerals (Bernard, 1967), electrical double layer expansion (Nasralla and Nasr-El-Din, 2014), and osmotic effect (Neuzil, 2000), can be ignored.

(7) During one model time step, the wetting front advances simultaneously into all neighboring connected grid cells. In other words, the wetting front advances one unit of chemical distance (Havlin et al., 1985) in one time step.

(8) The permeability reflects the rate of imbibition steps in the model; a larger permeability means one step represents longer time.

(9) Air can be trapped in dangling ends of the pore space during imbibition. Consequently, an individual grid cell may not attain 100% saturation even if water has been flowing through it.

(10) The wetting front stops moving once it reaches the upper boundary of the sample.

5. Mathematical models

The wetting phase movement (WPM) model was derived from the percolation theory which has been applied to water absorption into soils, geological, and other porous media samples (Hunt et

al., 2014). A rock was viewed as a composition of some same-sized grid cells. Every two neighboring grid cells have a connection probability to make fluid go through. Once the pore size or permeability is too small, the corresponding connection probability will be less than the critical value and the fluid movement will stop. Previous models, such as Hunt et al. (2014) and Cai et al. (2018), focus on the complex relationships of pores and pore throats; once there is a slight change on the initial values of basic parameters, the simulation results could be very different, thus, the results may be unpredictable and unreliable. In WPM model, the rock sample was divided into multiple grid cells having identical properties (for example, pore numbers and pore size) to ensure that the moving path has the same probability. The model could embed various-sized or directional fractures to mimic real samples. Each of the two neighboring grid cells are either connected or not, according to the stochastic probability value. The probability that two neighboring grid cells are connected, termed as connection probability, provides a reliable possibility to simulate the paths that water follows during the imbibition process. In this work, the WPM model was implemented by MATLAB software. First, the model was built on 2D and 3D systems with fractures. Then, the model simulates the experimental data for the Barnett Shale samples. The samples have two kinds of fracture orientation with respect to the upward imbibition direction, parallel and transverse to the bedding plane; these P- and T-direction fractures are shown as lines in 2D models and parallelograms in 3D models (see Appendix A).

5.1 2D model for quantifying water percolation and capillary forces during imbibition process

Fig. 4 lays out the structure of the 2D simulation models implemented in this work. The parametrization step sets the material properties (water surface tension and contact angle) and

decorates the lattice to set fracture locations and apertures, and inter-grid cell connections. Since the sample bottom is in contact with water, the grid cells of lowest layer have 100% water saturation, barring the small contribution of trapped air. The second step is to begin the percolation process, as the water imbibes from the grid cells of first layer up into the second layer. Each square-shaped grid cell has four directions (up, down, left, and right) for the water molecules to move with an equal probability. But grid cells in the bottom layer have only one direction, which is upward. If one grid cell is crossed by a fracture, the pores in this cell become water saturated. Then the water will quickly fill that fracture until it reaches a certain level which is calculated according to Jurin's law (Jurin, 1719). The neighboring grid cells near the fracture will be filled with water.

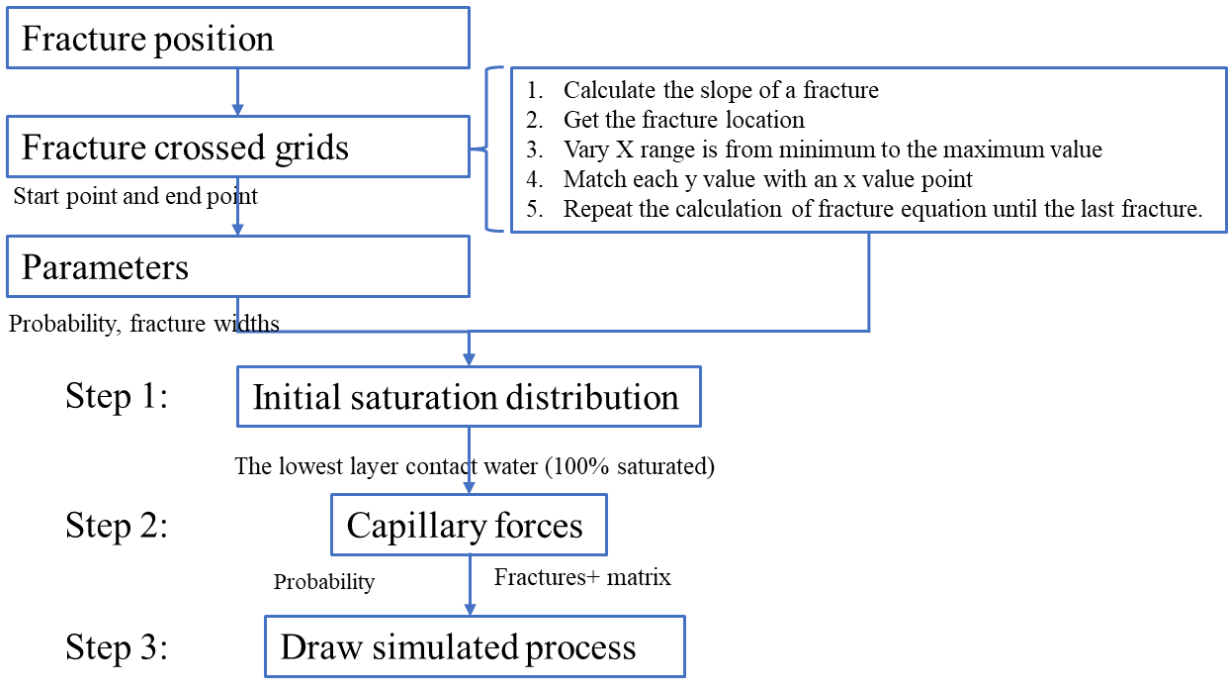


Figure 4: Workflow of 2D simulation model

In the first step, the sample size and the real fracture position(s) are noted, and all grid cells containing the fractures are identified. Since the bottom layer of rock sample is in contact with

water, in the first step, the water saturation of the first grid layer is set to 1 (100% water saturation). Thus, all grid cells in the first layer have no chance for more imbibition, while only the grid cells of second layer can be saturated in the next step. The rate of imbibition is related to the permeability and capillarity of rock samples. Each step represents a period of real time; if the permeability is higher, each step represents longer time. As all models are based on the Barnett Shale, the permeability and corresponding imbibition rates are the same in the simulations, with one step representing one hour. There are two domain theories in this model: the theory of invasion percolation is used within the porous matrix, and Jurin's law in the fractures as shown as Eq. (1) (Jurin, 1719),

$$h = \frac{2\sigma cos\theta}{\rho w g r} \tag{1}$$

As the pore surface of natural rock could be rough, the model adopts a friction coefficient to adjust the capillary rise in the model. Thus, when calculating the real capillary rise height, a parameter, named as friction coefficient (FC), is used in the WPM model; the modified capillary rise is shown as Eq. (2).

$$h = \frac{2\sigma cos\theta}{\mathrm{FC} \times \rho w g r} \tag{2}$$

where σ is the liquid-air surface tension (force/unit length), θ is the contact angle of water with the solid, ρ_w is the water density (mass/volume), g is the local acceleration due to gravity (length/time^2), and r is the radius of the pore throat. We used σ = 0.0728 N/m at 20 °C, ρ_w = 1000 kg/m^3, and g = 9.81 m/s^2 (Wang et al. 2021).

Although the fracture volume can be ignored in grid cells the fracture width is significant in the calculation of capillary rise in fractures. The first layer contacts the water source directly, so the first layer is seen as the free water surface. The model can calculate the capillary rise according

to the Jurin's law. Thus, we can know how high the water will rise in the fracture and in which grid cell it will stop.

Table 1 shows the basic information of a 2D water movement model, which includes the location of fractures and the parameters used for calculating the capillary force. The fracture volume can be neglected when calculating the imbibed volume in the model because the aperture size is much smaller than a grid cell. But in the capillary rise, the aperture width is significant. Thus, the rate of water imbibition will match the observed rate in the experiments.

Table 1 Parameters used in the 2D simulation model.

Parameters	Value
Water surface tension at 20^{o}C (N/m)	0.0728
Contact angle (^{o})	60
Fracture aperture (m)	5×10^{-6}
Gravity (m/s^2)	9.81
Density of wet phase water (kg/m^3)	1000
Density of non-wet phase air (kg/m^3)	0
Ideal capillary rise in one fracture (m) (Eq. 1)	1.48
Friction Coefficient	10000
Adjusted capillary rise in one fracture (m) (Eq. 1)	0.000148

Fig. 5 shows several example steps of the imbibition process in a 2D model, with the blue cells showing the advancement of wetting front. Once a grid cell containing a part of fracture is saturated, the capillary force will make water rapidly fill other parts of the fracture. Water in a matrix cell advances stochastically from any interface cells, but the water can only invade a neighboring cell if the connection between two cells is active. We have assumed that all matrix

cells are identical, so the weight of water imbibition at any time can be obtained simply from the number of saturated cells, volume of a cell, and water density.

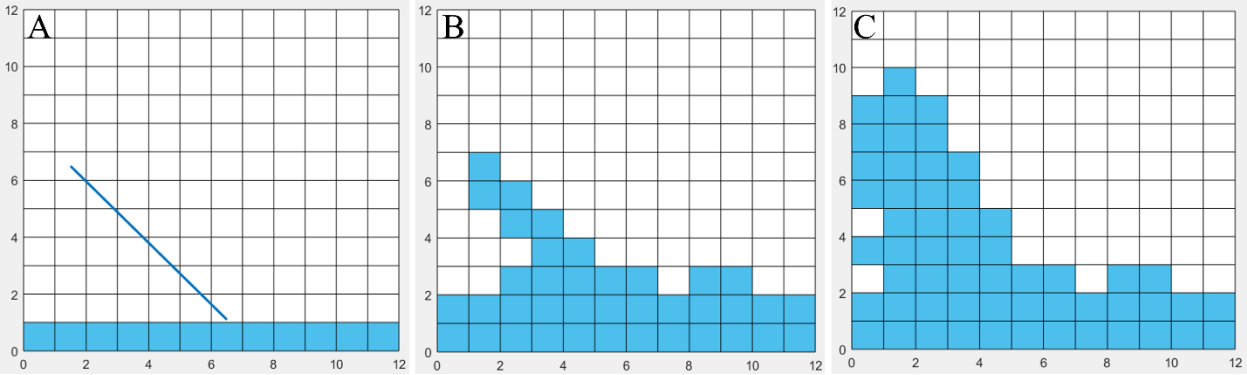


Figure 5: Imbibition process in 2D grid cells by percolation theory: (A) Step 1; (B) Step 3; and (C) Step 6. White and blue cells represent these without and with water; the blue line represents a fracture.

5.2 3D simulation model of water advancement under capillary forces

Flow in the 3D simulation model, shown in Fig. 6, is much more complex than that in the 2D model, starting with counting how many imbibed cells are crossed by fractures over time. Imbibition in the 3D model will also need to include vapor transport and absorption.

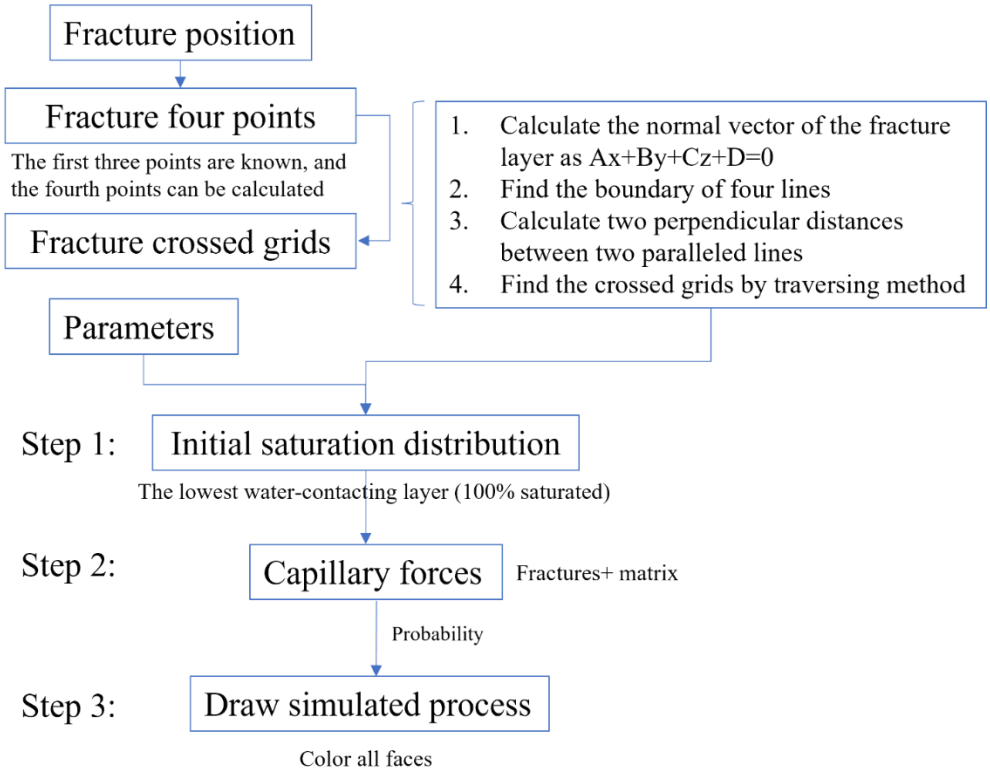


Figure 6: Workflow of 3D simulation model

The total volume of imbibed grid cells is the simulation result of mass of water imbibed over time on the current simulation step, which is shown as Eq. 3 and Eq. 4. The ratio of accessible porosity to total porosity (RAPTP) of connected pore volume is a new definition to describe how much of the grid can be filled with imbibing fluid. If a grid cell has several pores and paths linked to neighboring cells, it doesn't mean that all the connected void space in this grid could be filled with fluid during the imbibition process. RAPTP is an effective parameter to consider the percentage of a grid cell being imbibed by water.

$$V_{onegrid} = \frac{\pi R^2 H}{N_{grids}} \quad (3)$$

$$M = V_{onegrid} \times \phi \times N_{imbibed} \times RAPTP \times \rho_w \quad (4)$$

R and H are the radius and height of the rock sample; N_{grids} is the number of grid cells used in the simulation; $V_{onegrid}$ is the volume of a single grid cell; ϕ is the average porosity of the sample; $N_{imbibed}$ is the number of grid cells filled with the wetting-phase fluid; M is the mass of cumulative imbibed water.

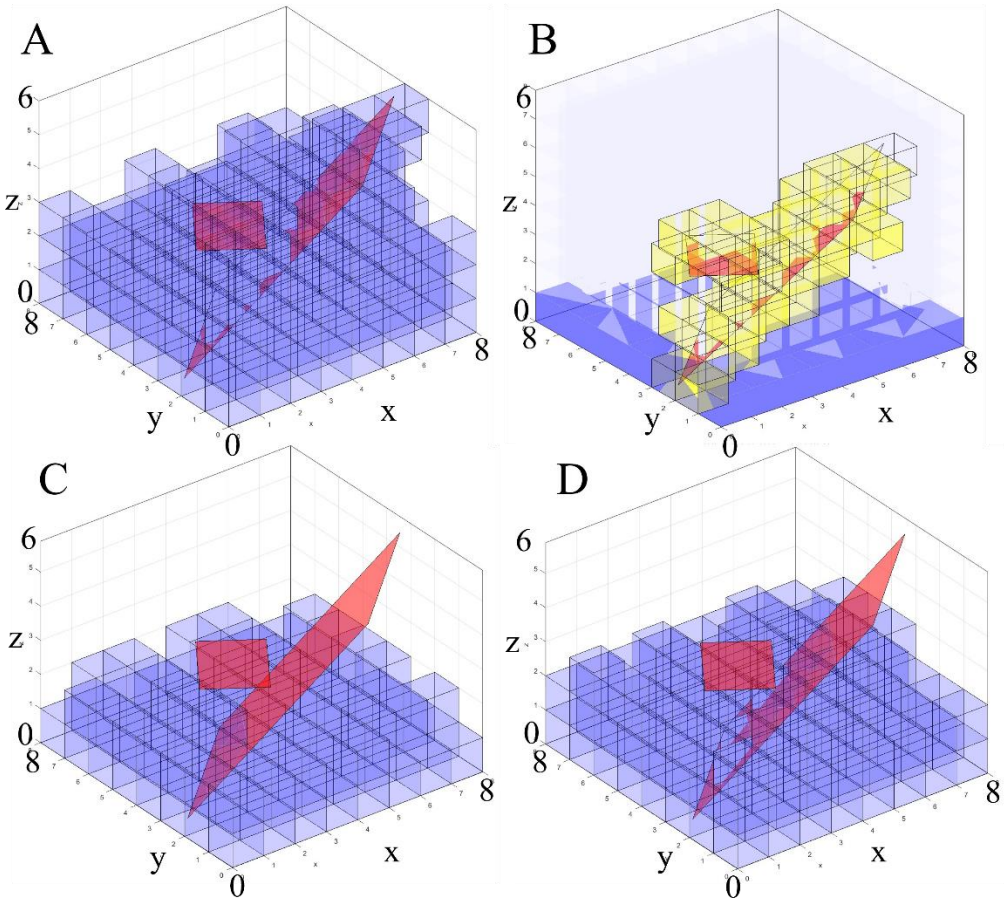


Figure 7: Simulation of the imbibition process in a square-shaped rock with fractures, with fracture-containing grid cells shown in yellow, at different steps of simulation: (A) First; (B) Second; (C) Fourth; and (D) Six.

Fig. 7 shows the process of imbibition with the red planes representing the fractures and yellow cells crossed by the fractures. The fracture shape is a parallelogram to avoid errors in settling the

complex boundaries of fractures and the computational burden in counting the crossed cells of fractures. The model could generate stochastically virtual fractures instead of calculating the positions of fractures.

The dark blue cells on the bottom layer are in direct contact with a full saturation in step one (Fig. 7 (A)). The light blue cells on the surface of other five directions contact air directly. These air-contact cells can also absorb water vapor due to the wettability of rock sample, as a thin water layer on the surface. For conveniently observing the invaded cells by vapor absorption, the color of cells invaded by vapor is chosen as 0% transparency. Figs. 7 (B)-(D) show the saturated grid distribution of simulation results at Steps 2, 4, and 6, respectively.

As shown in Fig. 7, all grid cells belonging to the bottom layer are therefore filled with water and counted into “movement tracing space” (a storage space in the model to track the positions of all imbibed cells), which stores the location of invaded grid cells at each simulation step. Based on chemical distance, the wetting front at the next step advances from the cells which are being filled with water at the current step. If water fails to move to a neighboring cell because the connection probability between those two cells was less than the threshold probability, the model will consider that those two cells are not directly connected. Grid cells that are completely unconnected are impermeable and never attain saturation greater than zero.

If some grid cells are invaded by vapor absorption before imbibition, the imbibed water will fill the rest of void space. The water vapor from the air has an impact on the imbibed mass, with the extent depending upon the rock properties and relative humidity. For example, the rock may absorb more water than normal when the humidity is high, or when the rock contains swelling clays (Zhang et al., 2017). From Fig. 7, we can see that the water moves along the fractures. Then the cells crossed by fractures are filled with water faster than the matrix cells whom near

the fracture cells will be filled with water in the next step by the connection probability by imbibition.

6 Simulations of fractured Barnett Shale

The WPM model was then used to describe and interpret the laboratory imbibition tests of Barnett Shale. The sample bottom was in contact with water for the imbibition tests of Barnett Shale samples, either with T (transverse to the bedding plane) direction fractures or with P (parallel to the bedding plane) direction fractures, and the imbibition results are shown in Fig. 8. The initial contact of imbibing deionized water causes the steep and noisy phase which should be not considered in the total imbibition process (Hu et al., 2015); the WPM model only uses the stable experimental data when the imbibition tests for T-and P-direction fractures are at 30 and 50 seconds, respectively. The sample mass obtained by the electronic balance over time was greater than actually imbibed mass obtained from the independent weighing. The submergence depth decrease of rock sample in a fluid reservoir from imbibition and evaporation will lead to the reduction of buoyant force, and the corrected data after buoyant change following the approach of Hu et al. (2001) was used for the WPM model.

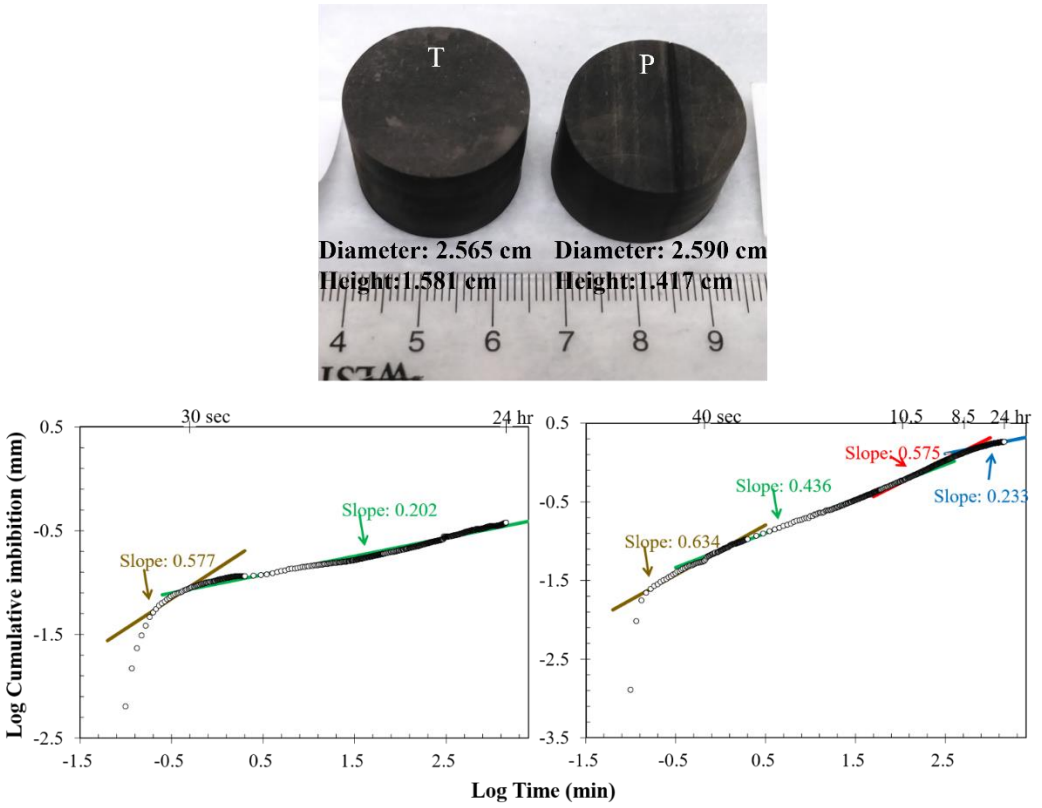


Figure 8: (Top) Picture of Barnett Shale sample (porosity at 13%) with T (left) and P (right) directions. (Bottom) Water imbibition results for the Barnett Shale sample under initially dry condition for T and P directions.

Fig. 9 shows four images of fractures with their apertures noted. Most fracture apertures are in the range of 5 to 80 μm, and these values are used in the WPM model.

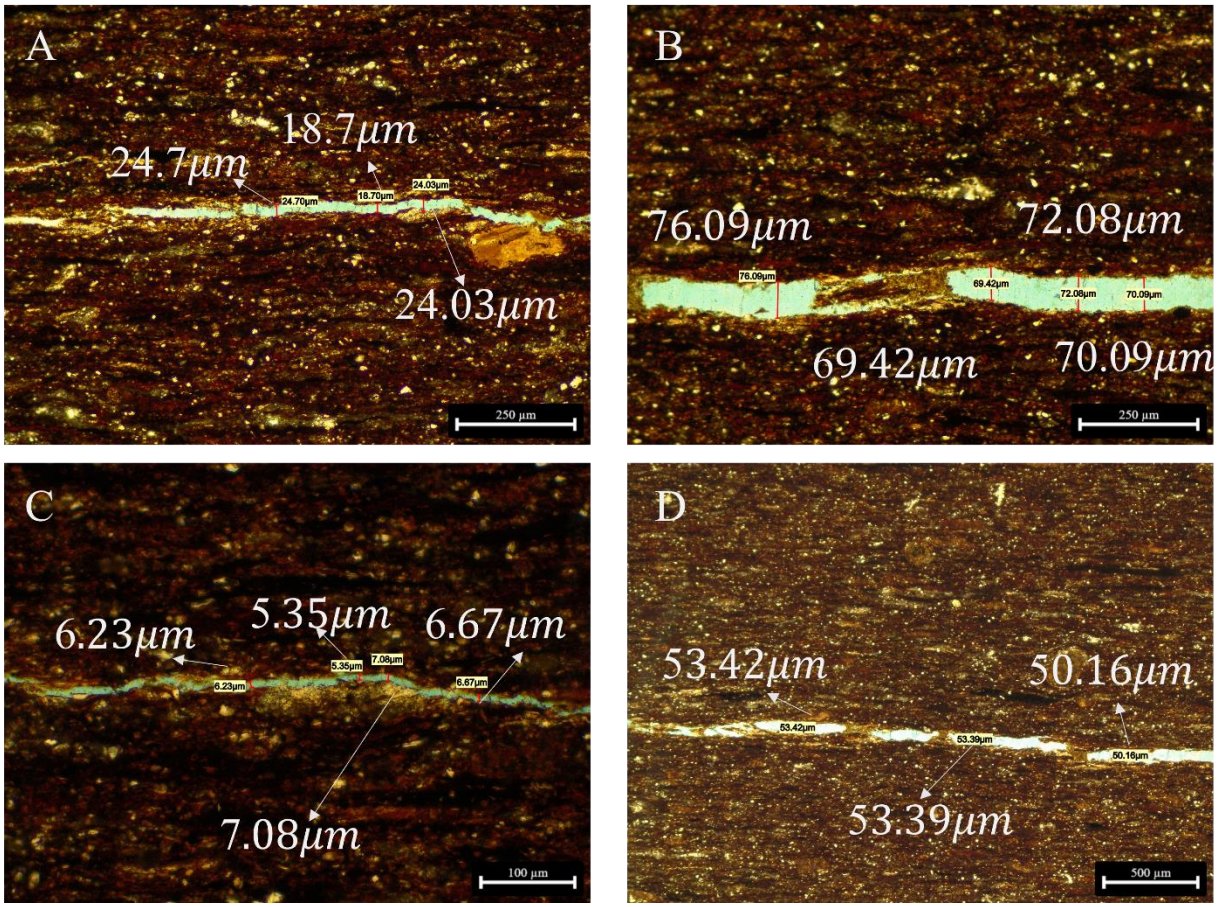


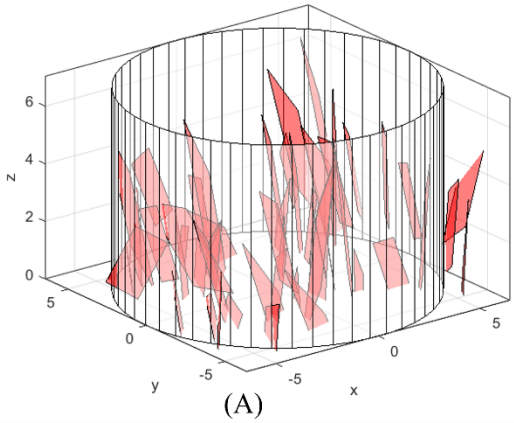
Figure 9: Fracture images with different apertures as observed from the thin section petrography by Leica DM 750P: (A) between 18.7–24.7 μm; (B) between 69.4 - 76.1 μm; (C) between 5.35 – 7.08 μm; and (D) between 50.2-53.4 μm.

There are two main fracture layouts in this study on the Barnett Shale, P and T directions. Though the shale samples have heterogeneous properties due to many factors, for example, the directions and the distribution of fractures, the connection probabilities in the same layer can be considered to be similar. The probability is larger through vertical axis than horizontal axis in P direction sample, while the probability has the opposite performance in T direction sample. According to percolation theory (Ewing et al., 2002), there are two kinds of probability in the imbibition process, one between every two neighboring cells, and the other is the description of connectivity of a sample by Monte Carlo methods. When the 1st probability is given, the 2nd

probability can be obtained from statistical tests. In other word, the 1st probability is the one for each cell to have active connections (site percolation) or for each connection between neighboring cells to be active (bond percolation), and the 2nd probability is a value that represents the connectivity of the whole sample. Therefore, the 2nd probability is comprised of the value of 1st probability.

6.1 Dual matrix-fracture system with P-direction fractures

Fig. 10A shows the sketch of rock surface and virtual fractures. To avoid the uncertainty bias during simulation, the average simulation result is obtained by running the Monte Carlo experiments 1000 times to obtain statistically reliable results, but the running time will be longer than a single run in the same simulation. The laboratory imbibition data after evaporation and buoyancy correction and the average simulated results are compared in Fig. 8B which shows a close match by the WPM model; the error difference between data and simulation is less than 3% in the long run.



(A)

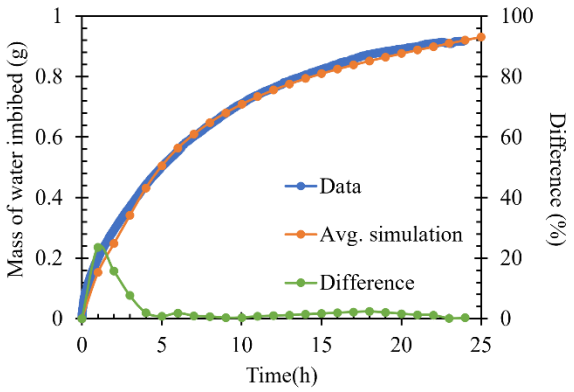


Figure 10: (A) Sketch of virtual P-direction fractures generated by the WPM model for a cylindrical Barnett sample; (B) Comparison of imbibition data vs. simulation results for P-direction fractures.

To understand the importance of influencing factors, the model was evaluated by various sensitivity analyses (Fig. 11). Every simulation line is obtained from the average of 1000 Monte Carlo runs. Fig. 11A shows the sensitivity analysis of the ratio of accessible porosity to total porosity (RAPTP). Porosity was determined by water immersion porosimetry after vacuum pulling and Archimedes' method (Vennard and Street, 1975) which could measure the volume of all pores connected to the sample's surface (Hu et al., 2012). The accessible pore volume which could be filled with wetting-phase-fluid in a dynamic imbibition test is less than the measured one after full saturation. If there are some pores connected to other large pores, but the pore throat is too small to let water move in, the small pore volume will be considered as non-connective pores instead of being a part of effective pore volume. It's shown that when the RAPTP is 0.89, the simulation matches the best, with 89% surface-accessible pore space participating in the dynamic imbibition process. Though the rates of water imbibed over time with various RAPTP are almost equal at each step, the minor impact of RAPTP becomes large with the increasing imbibition time. The connection probability has the same impact on the mass of water imbibed (shown in Fig. 9B). The greater value of the probability of imbibition, the more cells can be filled with wetting-phase fluid, and the greater value of the mass of water imbibed over time.

There is another important influencing factor during imbibition, which is the water vapor absorption into the sample. The shale sample has been coated with epoxy, a kind of water-proof polymer, on the exterior side. The vapor still could enter the tight shale sample through the top

side though it was loosely covered with aluminum foil. Then, the vapor movement is the same as imbibition movement. If a grid cell is permeable, it allows vapor to move through the pores. The probability of vapor movement accessibility also has an impact on the mass of water imbibed over time though the influence is not as prominent as the imbibition probability (Fig. 11C). There is an essential parameter to represent the ability of a sample to let water vapor attach on surface, vapor maximum value, shown as Eq. 7. If the vapor enters the interior of a rock, the wetting phase will become a thin layer to attach onto the pore surface. Though the volume of such a layer is small, it still takes up a part of pore space. The maximum volume of vapor that could attach onto the pore surfaces depends on the rock wettability, and its influence becomes more prominent over time.

$$Vapor_{max} = \frac{themaximumvolumeofvapor \in agrid}{thevoidvolumeofagrid \in 60°Ccondition} \quad (7)$$

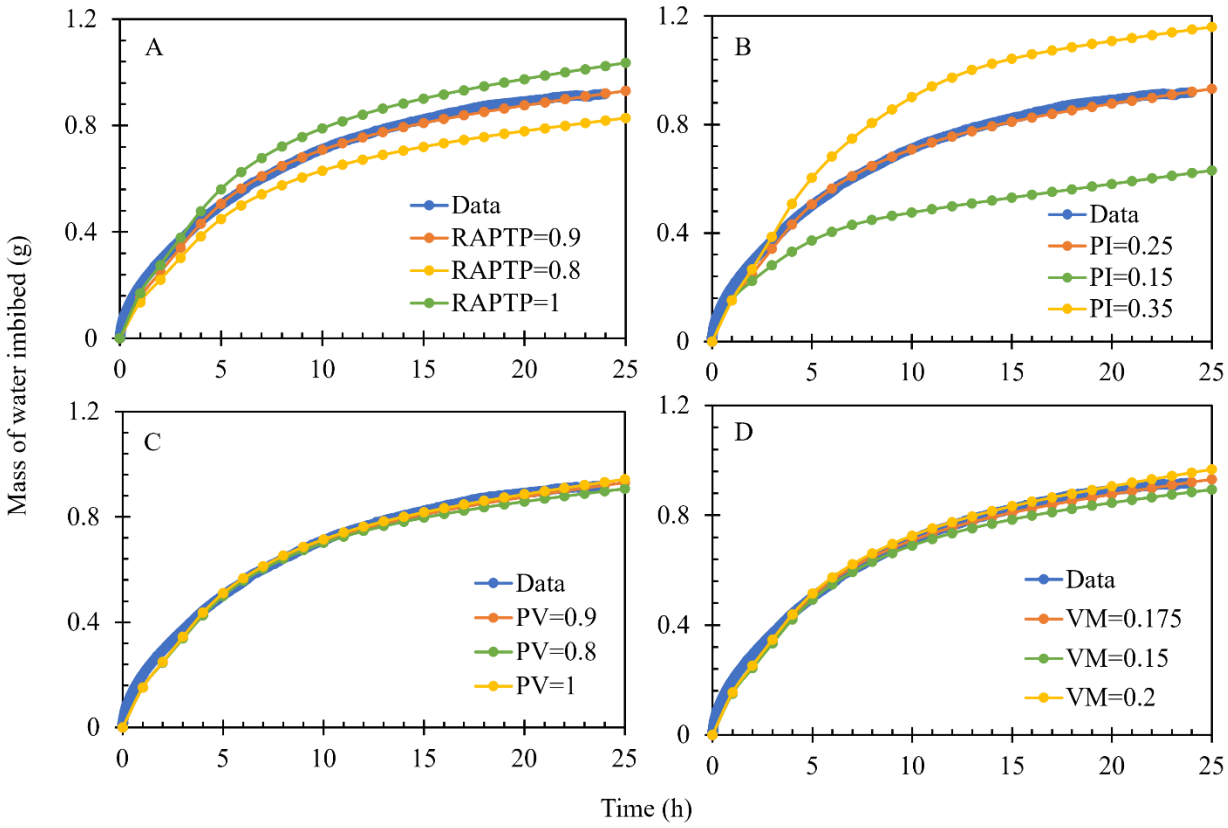


Figure 11: Comparison of the imbibed water mass and simulation of the Barnett Shale sample with P direction, as influenced by: (A) RAPTP; (B) probability of imbibition (PI) in matrix cells; (C) probability of vapor (PV) absorption in matrix cells; and (D) maximum of vapor (MV) saturation in matrix cells.

6.2 Dual matrix-fracture system with T-direction fractures

Fig. 12A shows the sketch of rock surface and virtual fractures, and Fig. 12B shows the comparison of the laboratory data and simulation results of water imbibition over time. As shown in Fig. 12B, the mass of imbibed water over time has a sudden increase in the first hour. If the fractures are mostly in P direction, as settled to be vertical, the water movement may face less restriction than the sample with T-direction fractures. When the wetting-phase liquid enters the cells crossed by T-direction fractures, the fracture (all the cells at the same level) will be

filled with liquid over a short time. In the meanwhile, the upper matrix cells have a relatively lower probability than the fracture cells at the same height level to be filled with liquid because of the gravity. As a result, if the T-direction fractures are exposed to the exterior at the bottom layer, those grid cells will be filled with water suddenly because they directly contact the surrounding water phase. During the imbibition process, once the water moves from matrix cells to proceed to fracture cells, the water will fill all cells crossed by fractures and cause a sudden increase of water imbibition.

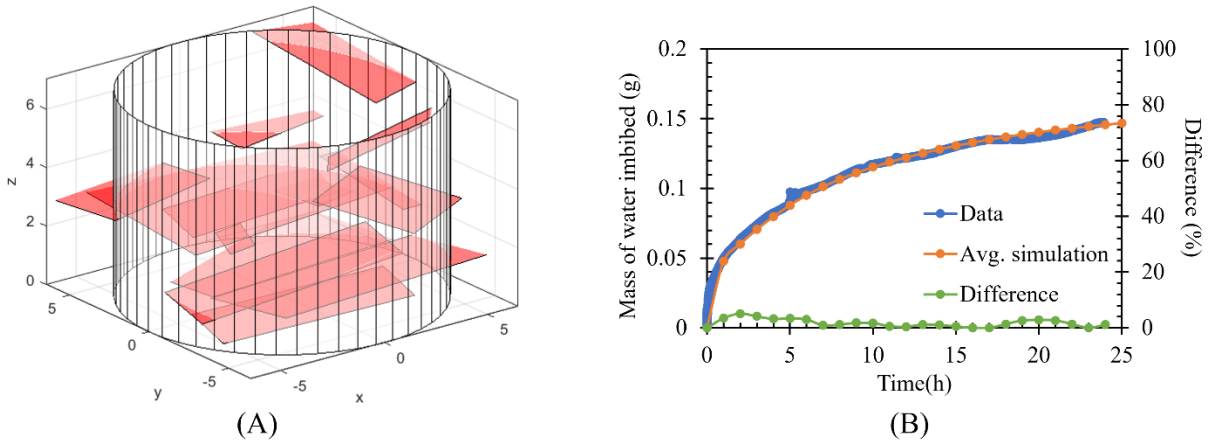


Figure 12: (A) Sketch of virtual T-direction fractures generated by the WPM model within a cylindrical Barnett sample; (B) Comparison of imbibition data vs. simulation results with T-direction fractures.

Fig. 13 (A)A shows the sensitivity analysis of ratio of RAPTP in Barnett Shale sample with T-direction fractures. The initial stage is unstable as the spontaneous imbibition is controlled by capillary force (Gao and Hu, 2016). The capillary force is determined by the number of fractures inserted into the cells of the bottom layer. To mimic several fractures located at the bottom layer, some fractures shown in Fig. 12 (A) may be established to contact water at the first step even

though their positions are not in the bottom layer. Fig. 13 (B) shows the influence of RAPTP; the higher ratio of saturation, the more water being imbibed into the rock. Fig. 13 (C) indicates that fractures are sensitive to direct contact of water. The probability of imbibition in Barnett matrix cells is extremely low. Since the vapor could enter small pores easier than liquids, the probability of vapor absorption is much higher than the probability of imbibition. Fig. 13 (D) shows that the probability of vapor absorption has little effect on the mass of imbibed water at the beginning, but it tends to have an increasing influence over time.

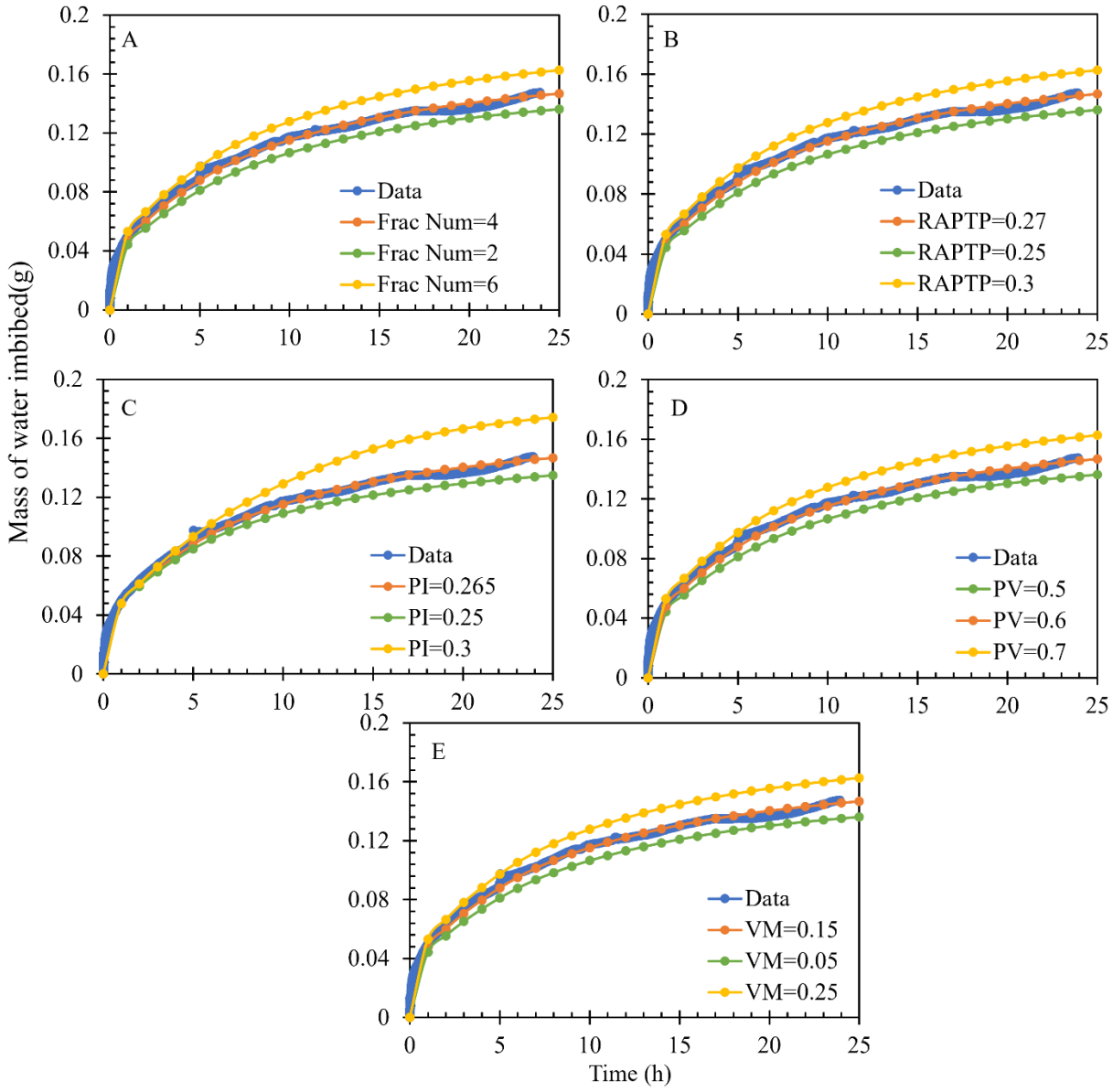


Figure 13: Comparison of laboratory data with simulation results for the T direction sample for the imbibed water mass with respect to; (A) numbers of fully saturated fractures (FN) in the first simulation step; (B) RAPTP; (C) PI in matrix cells; (D) PV in matrix cells; and (E) MV in matrix cells.

This phenomenon demonstrates that vapor absorption is not the main factor on mass of water imbibed rate at first but exists during the whole experiment. The maximum value of vapor

saturation sensitivity analysis shows the same tendency of simulation results which are shown in Fig.13 (E). Though it is hard to know the distribution of fractures in the rock sample without X-ray or neutron tomography, the WPM model could well match the mass of water imbibed over time with the laboratory data.

7. Model prediction

7.1 Predictions for cubic Barnett Shale samples

The WPM model is able to predict the imbibition process in the same rock with different geometries. For the simulation accuracy, all the parameters used in the previous benchmark simulation tests (Figs.12-13) of the Barnett Shale are applied in the prediction (Fig. 14). Though the bulk volumes of the samples are different, the size of the virtual simulation grid cells are set to the same value. For example, if the height of the cylindrical sample is 15 mm while the height of the cubic sample is 10 mm, the simulation grid cells are set as the same value (0.1 mm). Fig. 14 shows four cubic samples, which have P-direction fractures, T-direction fractures, P-direction matrix and T-direction matrix. Due to the smaller scale of cubic samples, there are only two visible fractures in P-fractured and T-fractured rocks (Figs.14 (A)-(B)) and there is no visible fracture in the matrix-only samples (Figs.14 (C) & (D)).

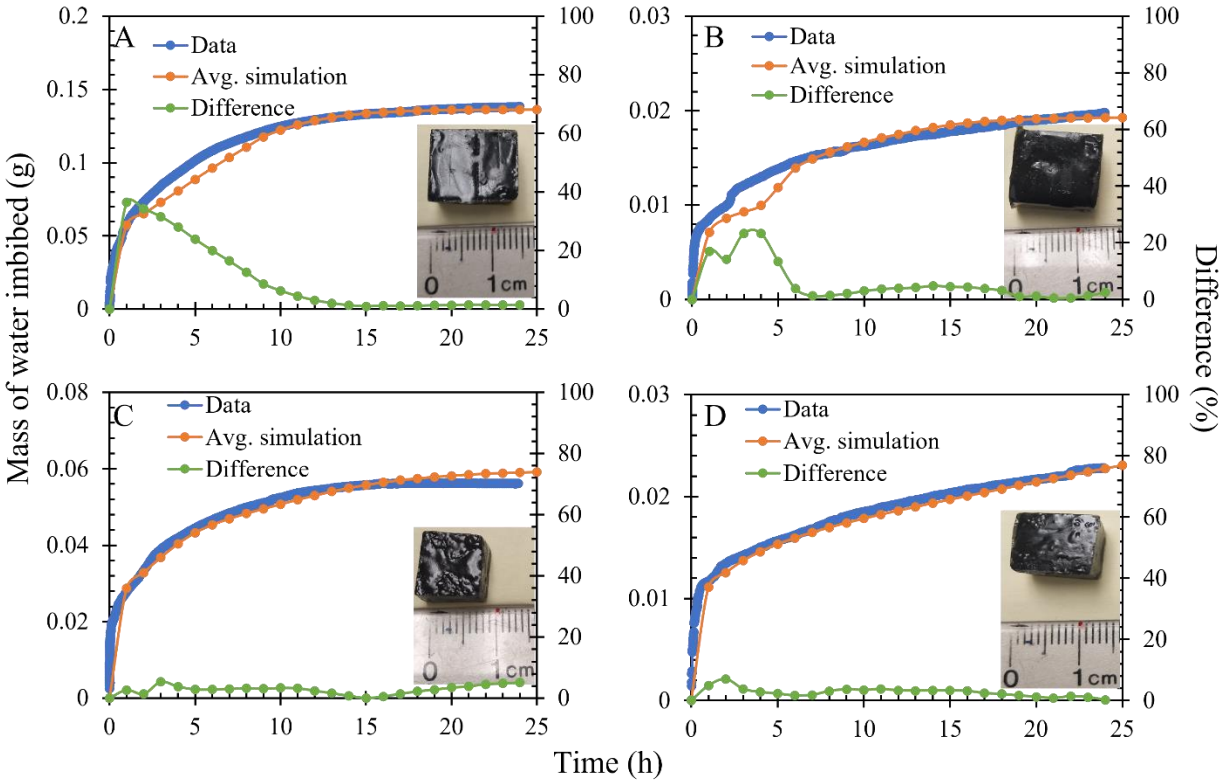


Figure 14: Comparison of imbibed water mass over time in Barnett Shale samples: (A) P-direction fractures; (B) T-direction fractures; (C) P-direction matrix; and (D) T-direction matrix.

Fig. 14 shows the comparisons of laboratory data and simulation results. All parameters are almost the same as the parameters used in the simulations of cylindrical samples, which are presented in the sensitivity analyses (Figs. 11 and 13).

The predicted results have the same trends as the cylindrical runs. At the beginning of these experiments, the rate of water imbibition is related to the number and position of fractures. After the rapid imbibition stage, the rate of water imbibition tends to rely on the matrix properties, likely the pore size and pore throat in controlling the capillary pressure and permeability of shale matrix.

7.2 Prediction of cylindrical Barnett Shale

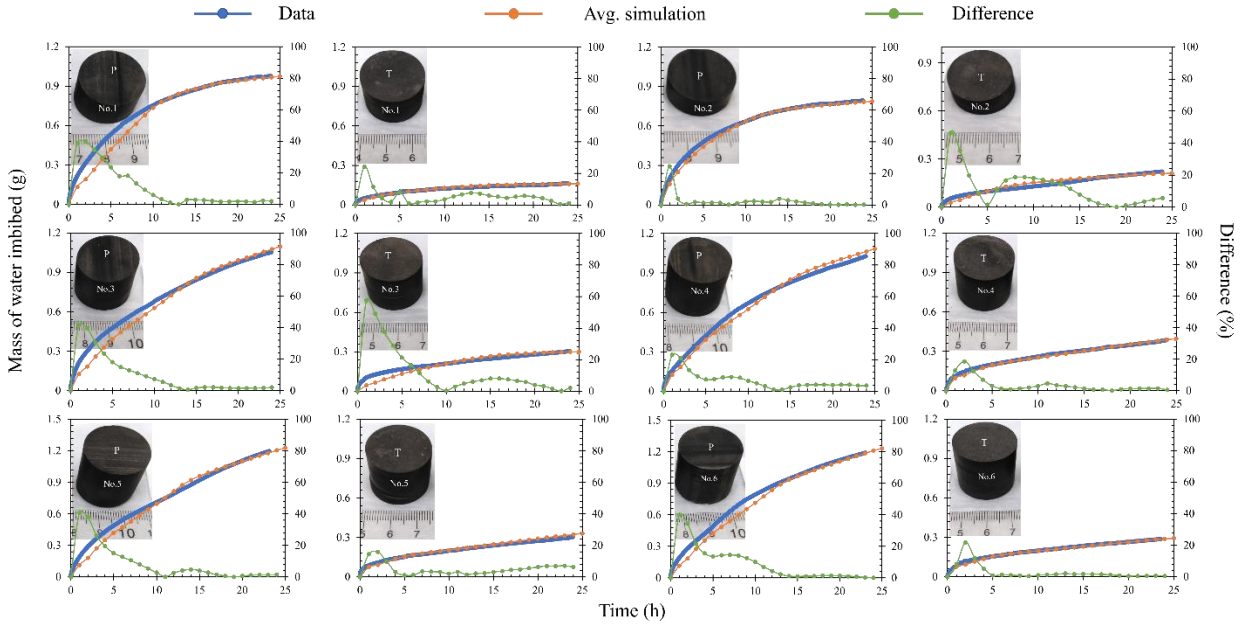


Figure 15: Comparison of mass of imbibed water over time in Barnett Shale samples

To further assess the accuracy and viability of the WPM model, we tested six pairs of cylindrical Barnett rock samples cored from the same large block, while the dimensions of these samples are shown in Table 2.

Fig. 15 shows the comparisons of laboratory data and simulation results for cylindrical Barnett samples. In these figures, sample size is a significant factor to influence the ultimate mass of imbibed water in 24 hours. The sizes of Barnett Shale samples are listed in Table 2. For a given sample size and sampling location, samples with P-fractures have a better ability to imbibe water. Even though the P-fractured samples are prone to take more water compared to T-fractured samples, the imbibition rate in P-fractured samples will slow down over time while the rate of T-fractured samples tend to be steady.

Table 2 The sizes of Barnett Shale samples

	Diameter (mm)	Height (mm)	Porosity (mm^3/mm^3)
1P	25.90	14.17	0.131
1T	25.65	15.81	0.138
2P	25.58	11.47	0.131
2T	25.63	9.94	0.136
3P	25.09	21.60	0.131
3T	25.86	20.95	0.139
4P	26.64	21.94	0.134
4T	26.26	20.90	0.133
5P	25.45	22.90	0.130
5T	25.21	19.97	0.138
6P	25.92	21.59	0.127
6T	25.45	22.00	0.124

8 Conclusions

In this work, we introduce a new model for wetting phase imbibition in the fractured porous media. The significant advantages over previous model are embedding simplified multiple fractures into matrix, calculating the water mass imbibed and visualizing the water movement in every step. The significance between the previous models and current one is that the percolation theory was introduced to simulate the stochastic movement by Monte Carlo approach. Therefore, this new model can test the simulation cases thousands of times, in order to obtain the most possible results in the simulation. The size, position, and distribution of fractures are based on the laboratory observations and experiments which make the simulation results more reliable.

The creative method of emplacing virtual fractures also helps the tracing of water movement paths statistically; thus, the model can have a more reliable prediction of the cumulative imbibed mass. However, as the Barnett Shale samples are reasonably considered in this work as homogeneous in the matrix, other scenarios, such as heterogeneous and irregularly shaped rocks with microfractures and large difference in pore size distribution, may not be predicted by this current model. Overall, from integrated experimental and modeling approaches, this work concludes with the following main advantages and contributions of the new model:

The model adopted the aperture width frequency from thin sections, which means the capillary rise depends on each fracture itself. The WPM model is developed to be able to change the ratio of accessible porosity to total porosity, and to simulate the water movement and mass of water imbibed in every step according to the heterogeneous fractures and homogeneous matrix. The new model reflects the performance of percolation process in rocks. The governing equations, e.g., capillary rise, are easily applied into the model to relieve the computational burden. This advantage satisfies the need of studying complex fractures in rock to ensure the simulation accuracy of imbibed mass.

The distribution of fractures has a significant impact on the imbibition process. The multiple crossed fractures could provide effective channels to enlarge the conductivity in the matrix. The water moves easier along with fractures, so the mass of imbibed water over time is larger in P- than T-direction rocks. Horizontal fractures prevent water from moving forward, even though the matrix is homogeneous, the upper grid cells are unable to contact water percolated from the bottom grid cells because of the heterogeneous fractures.

The sensitivity analysis shows that the percolation probability determines the rate of water imbibition. The vapor saturation and probability in matrix grid cells affect the value of mass of

water being imbibed in every step, especially at the beginning of the imbibition process. The values of vapor saturation and corresponding probability have more influence on the rock samples with T-direction fractures than P-directions.

The reason for the different rates of imbibed water may be due to the fracture distribution. If most fractures are in contact with the bottom layer, like P-fractures, where the grid cells have 100% water saturation, the matrix cells have a higher probability to be saturated. With T-fractures, the fractures in the bottom layer are saturated, but the water has less probability to move upwards in other directions or to encounter the next fracture. Thus, the steady rate of imbibition may be due to the sample lacking fractures which may cause the acceleration of water imbibition. Overall, the WPM provides capabilities of describing the experimental results, and predicting the spatial and temporal distribution of water mass during an imbibition process, which are not easily shown experimentally, for geological samples.

Acknowledgment

This project was completed with funding provided by the Nuclear Energy University Program, Office of Nuclear Energy, U.S. Department of Energy, award number DE-NE0008797.

References

Alkouh, A., 2014. New Advances in shale gas reservoir analysis using water flowback data, Physical Review Letters, 49, 479–488.

Akbarabadi, M., Saraji, S., Piri, M., G., D., Delshad, M., 2017. Nanoscale experimental investigation of in-situ wettability and spontaneous imbibition in ultra-tight reservoir rocks, Advances in Water Resources, 107, 160–179

Bernard, G., 1967. Effect of floodwater salinity on recovery of oil from cores containing clays. SPE Paper Number 1725, 10.2523/1725-MS.

Bernabé, Y., Li, M., Maineult, A., 2010. Permeability and pore connectivity: A new model based on network simulations. Journal of Geophysical Research, 115 B10, B10203.

Cai, J., Perfect, E., Cheng, C., Hu, X., 2014. Generalized modeling of spontaneous imbibition based on Hagen-Poiseuille flow in tortuous capillaries with variably shaped apertures. Langmuir, 30, 5142-5151.

Chalmers, G.R., Bustin, R.M., Power, I.M., 2012. Characterization of gas shale pore systems by porosimetry, pycnometry, surface area, and field emission scanning electron microscopy/transmission electron microscopy image analyses: examples from the Barnett, Woodford, Haynesville, Marcellus, and Doig unit. AAPG Bulletin, 96, 1099–1119.

Cipolla, C.L., 2009. Modeling production and evaluating fracture performance in unconventional gas reservoirs. Journal of Petroleum Technology, 61, 84–90.

Dehghanpour, H., Zubair, H. A., Chhabra, A., Ullah, A., 2012. Liquid intake of organic shale. Energy & Fuels, 26, 5750-5758.

Dejam, M., Hassanzadeh, H., Chen, Z., 2018. Semi-analytical solution for pressure transient analysis of a hydraulically fractured vertical well in a bounded dual-porosity reservoir. Journal of Hydrology, 565, 289–301.

Eden, M. 1961. A two-dimensional growth process. Proceedings of Fourth Berkeley Symposium on Mathematics, Statistics, and Probability. University of California Press, Berkeley, pp. 223–239.

Ewing, R.P., Horton, R., 2002. Diffusion in sparsely connected pore spaces: temporal and spatial scaling. Water Resources Research, 38(12), 1285.

Ewing, R.P., Hu, Q.H., Liu, C.X., 2010. Scale dependence of intra-granular porosity, diffusivity, and tortuosity. Water Resources Research, 46, W06513.

Fenwick, D.H., Blunt, M.J., 1998. Three-dimensional modeling of three phase imbibition and drainage. Advances in Water Resources, 21(2), 121–143.

Gao, Z.Y., Hu, Q.H., 2012. Using spontaneous water imbibition to measure the effective permeability of building materials. Special Topics and Reviews in Porous Media - An International Journal, 3(3): 209–213.

Gao, Z., Hu, Q.H., 2016. Initial water saturation and imbibition fluid affect spontaneous imbibition into Barnett Shale samples. Journal of Natural Gas Science and Engineering, 34, 541–551.

Gombia, M., Bortolotti, Villiam., Brown, R., Camaiti, M., Fantazzini, P., 2008. Models of water imbibition in untreated and treated porous media validated by quantitative magnetic resonance imaging. Journal of Applied Physics. 103, 9.

Havlin, S., Trus, B., Weiss, G. H., & Ben-Avraham, D., 1985. The chemical distance distribution in percolation clusters. Journal of Physics A: Mathematical and General, 18(5), L247–L249.

Hu, Q., Persoff, P., Wang, J.S.Y., 2001. Laboratory measurement of water imbibition into low permeability welded tuff. Journal of Hydrology, 242, 64-78.

Hu, Q., Ewing, R.P., Dultz, S., 2012. Low pore connectivity in natural rock. Journal of Contaminant Hydrology, 133, 76-83.

Hu, Q.H., Ewing, R.P., Rowe, H.D., 2015. Low nanopore connectivity limits gas production in Barnett Formation. Journal of Geophysical Research – Solid Earth, 120(12), 8073–8087.

Hunt, A., Ewing, R., Ghanbarian, B., 2014. Percolation Theory for Flow in Porous Media. Lecture Notes in Physics, vol 674. Springer, Berlin, Heidelberg.

Li, J., Gao, Y., Jiang, H., Liu, Y., Dong, H., 2018. Pore-scale imaging of the oil cluster dynamic during drainage and imbibition using in situ X-Ray microtomography. Geofluids, 1-13.

Jurin, J. 1719. II. An account of some experiments shown before the Royal Society; with an enquiry into the cause of the ascent and suspension of water in capillary tubes. Philosophical Transactions of the Royal Society of London. 30 (355), 739–747.

Loucks, R.G., Ruppel, S.C., 2007. Mississippian Barnett Shale: lithofacies and depositional setting of a deep-water shale-gas succession in the Fort Worth basin, Texas. AAPG Bulletin, 91(4), 579e601

Lucas, R., 1918. Ueber das Zeitgesetz des kapillaren Aufstiegs von Flüssigkeiten. Kolloid-Zeitschrift 23(1), 15–22.

Madden, T.R., 1983. Microcrack connectivity in rocks: a renormalization group approach to the critical phenomena of conduction and failure in crystalline rocks. Journal of Geophysical Research, 88B1, 585–592.

Mavko, G., Nur, A., 1997. The effect of a percolation threshold in the Kozeny-Carman relation. Geophysics, 62(5), 1480–1482.

Meng, Q., Cai, Z., Cai, J., Yang, F., 2018. Oil recovery by spontaneous imbibition from partially water-covered matrix blocks with different boundary conditions. Journal of Petroleum Science and Engineering, 172, 454-464.

Morris, J. 1979. Traversing binary trees simply and cheaply. Information Processing Letters, 9, 197-200.

Naar, J., Henderson, J. H., 1961. An imbibition model - Its application to flow behavior and the prediction of oil recovery. Society of Petroleum Engineers Journal, 1(02), 61–70.

Nasralla, R., Nasr-El-Din, H., 2014. Double-layer expansion: Is it a primary mechanism of improved oil recovery by low-salinity waterflooding? SPE Reservoir Evaluation & Engineering. 17. 10.2118/154334-MS.

Neuzil, C., 2000. Osmotic generation of ‘anomalous’ fluid pressures in geologic environments. Nature, 403, 182-4.

Ning, X., Feng, Y., Wang, B., 2020. Numerical simulation of channel fracturing technology in developing shale gas reservoirs. Journal of Natural Gas Science and Engineering, 83, 103515.

Roshan, H., Ehsani, S., Marjo, C.E., Andersen, M.S., Acworth, R.I., 2015. Mechanisms of water adsorption into partially-saturated fractured shales: an experimental study. Fuel, 159, 628–637.

Song, R., Wang, Y., Liu, J., Cui, M., Lei, Y., 2019. Comparative analysis on pore-scale permeability prediction on micro-CT images of rock using numerical and empirical approaches. Energy Science Engineering, 7(6), 2842-2854.

Stavropoulou, E., Ando, E., Tengattini, A., Briffaut, M., Dufour, F., Atkins, D., Armand, G., 2019. Liquid water uptake in unconfined Callovo Oxfordian clay-rock studied with neutron and X-ray imaging. Acta Geotechnica, 14, 19-33.

Stavropoulou, E., Andò, E., Roubin, E., Lenoir N., Tengattini, A., Briffaut, M., Besuelle, P., 2020. Dynamics of water absorption on Callovo-Oxfordian claystone revealed with multimodal x-ray and neutron tomography. Frontiers in Earth Science. 8,6.

Suo, Si., Liu, M., Gan, Y., 2018. Modelling imbibition processes in heterogeneous porous media. Transport in Porous Media. 126, 615–631.

Shi, Y., Yassin, M.R., Yuan, L., Dehghanpour, H., 2019. Modelling imbibition data for determining size distribution of organic and inorganic pores in unconventional rocks. International Journal of Goal Geology, 201, 26-43.

Spaid M.A.A., Phelan F.R. Jr., 1998. Modeling void formation dynamics in fibrous porous media with the lattice Boltzmann method. Composites Part A: Applied Science and Manufacturing, 29(7), 749–755.

Vennard, J.K., Street, R.L., 1975. Elementary Fluid Mechanics. John Wiley and Sons, Inc, New York.

Wang, F., Zhao, J., 2021. Mathematical model of liquid spontaneous imbibition into gas-saturated porous media with dynamic contact angle and gravity. Chemical Engineering Science, 229, 116139.

Wang, Q., Wen, Z., Hu, Q., Xu, H., Meendsen, F., Shu, Y., Qiao, H., 2021. Pore geometry characteristics and fluid-rock interaction in the Haynesville Shale, East Texas, United States. Energy & Fuels, 35(1), 237-250.

Wang, Q., Zhang, X., Hu, Q., Lin, X., 2020. Spontaneous imbibition of water in the Barnett Shale: Quantitate the influence of artificial fracture and bedding plane. AAPG Search and Discovery Article #11353, 3.

Wang, Z., Yang, Z., Ding, Y., Lin, W., He, Y, Duan, X., 2018. A generalized capillary imbibition model for porous media in tight reservoirs. Advances in Civil Engineering, 4148734.

Washburn, E.W., 1921. The dynamics of capillary flow. Physical Review, 17(3), 273–283.

Witten, T., Sander, L., 1981. Diffusion-limited aggregation, a kinetic critical phenomenon. Physical Research Letters, 47(19). 1400.

Wilkinson, D.J., Willemsen, J., 1983. Invasion percolation: A new form of percolation theory. Journal of Physics A: Mathematical and General, 16 (14), 3365.

Zeng, F., Zhang, Q., Guo, J., Zhang, Y., Ren, W., Jiang, Q., Xiang, J., 2020. Analytical spontaneous imbibition model for confined nanofractures. Journal of Geophysics and Engineering. 17 (4), 635–642.

Zhang, C., Ranjith, P. G., Perera, M., Zhao, J., 2017. Characteristics of clay-abundant shale formations: Use of CO_2 for production enhancement. Energies, 10(11), 1887.

Chapter III

Liquid imbibition in non-fractured porous media with various pore-size-distribution: Experimental and simulation results

Xuewei Ning, Khawaja Hasnain Iltaf, Qinhong Hu

Department of Earth and Environmental Sciences, the University of Texas at Arlington, TX 76019, USA

Keywords: imbibition test, heterogeneity, pore size distribution, petrophysical characters, fluid flow pathways, machine learning, simulation models

Abstract

This chapter examines the influence of pore structure and mineral composition on the rate and magnitude of spontaneous liquid imbibition into a wide variety of natural rocks. In the contemporary age, characterized by heightened exploitation of subsurface porous reservoirs for purposes ranging from petroleum extraction to environmental initiatives like CO2 sequestration and water storage, the study of fluid liquid flow and gas diffusion in natural rocks gains utmost relevance.

The study involves an extensive analysis of 15 distinctive natural rock samples, each with unique characteristics, employing a series of experimental methodologies such as water spontaneous imbibition, contact angle measurements, X-ray diffraction (XRD), and Mercury Intrusion Porosimetry (MIP). The rock samples, systematically segmented into smaller cylindrical units, have their heterogeneity tested in two directions at 90 degrees from each other. The contact angle test, instrumental in characterizing material surface properties and interfacial interactions, is integrated into the machine learning approaches. XRD analyses prove particularly effective in examining the phase and crystal structure of various mineral phases in these natural rocks. The standard MIP technique aids in the measurement of pore throat size distribution, porosity, and the comparison of injection pressure versus mercury saturation across diverse rock types. Experimental data from these methods is incorporated into the dataset to augment the predictive accuracy of select machine learning techniques, which includes gas concentration difference between outside and inside monitored chamber, contact angle, pore-throat distribution, porosity, permeability, the size of the sample (height and radius).The results show the accuracy and efficiency of machine learning in predicting spontaneous imbibition in various complex natural rocks.

1. Introduction

Conventional reservoirs have been utilized for hundreds of years while unconventional reservoirs have been explored for decades, both reservoirs share a significant status in petroleum exploitation. The pore structure and mineral composition are the two important geophysical features in tight sandstone.(Cai et al., 2020; Zhu et al., 2020; Sheng et al., 2020). The pore structure includes the geometry, connectivity of void space and pore size distribution (Anovitz and Cole, 2015; Wang et al., 2020; Liu et al., 2020). Based on the current advanced hydraulic fracturing techniques, spontaneous liquid imbibition is considered as an effective way to improve oil recovery, especially in unconventional reservoirs (Cai et al., 2020; Liu et al., 2021).

The pore structure and mineral composition are viewed as two essential factors in influencing the rate and ultimate fluid recovery from spontaneous imbibition (Austad et al., 2015; Liu et al., 2021; Mehana et al., 2018). Furthermore, the rock matrix and fluid properties are considered as two main influencing aspects to affect tight sandstone's spontaneous (Cai et al., 2020; Guo et al., 2020). In recent decades, underground porous reservoirs are not only being exploited for petroleum, but also for making use of protecting the environment, for example, CO_2 sequestrations and water storage (Georgiadis et al., 2012; Granados et al., 2021).

Many studies for investigating imbibition process have been carried out. Contact angle test is a method to characterize material surface properties and interfacial interactions between gas, liquid and solid phases (Alhammadi et al., 2017; Rezaei et al., 2020). X-ray diffraction (XRD) analysis is one of the most direct methods to study the phase and crystal structure of various mineral phases in porous media. The material composition, molecular configuration, conformation crystal form, and intramolecular bonding mode determines the unique diffraction pattern of the material (Bernstein, 2007). XRD, as a rapid routine quantitative analyses method, scans both the

mineral contents of organic-rich rocks (Mandile et al., 1995; Al-Ghamdi, 2019). Mercury Intrusion Porosimetry (MIP) is a commonly used technique for measuring pore throat size distribution, porosity, and injection pressure vs. mercury saturation for various types of rocks (Lohr et al., 2018; Peng et al., 2017; Mitchell et al., 2008).

In this work, 15 kinds of natural rock samples with various characteristics were tested by spontaneous water imbibition, contact angle, XRD, and MIP approaches. Fig. 1 displays the geographical positions of the rock samples subjected to testing, while Fig. 2 provides comprehensive geological details about these samples. All rock samples are cut into small cylinder samples at different sizes. Heterogeneous rocks were tested in two perpendicular directions which are vertical from each other; for the rock samples with clearly identifiable bedding planes, they are referred as parallel and transverse to the bedding plane, P- and T-direction for short. The contact angle measurement, XRD analysis, and MIP tests will be conducted to obtain the characterization of rock properties, such as the pore structure, mineral composition.

2. Samples

A total of 15 types of rock samples in U.S. were studied in this research (Fig. 1).

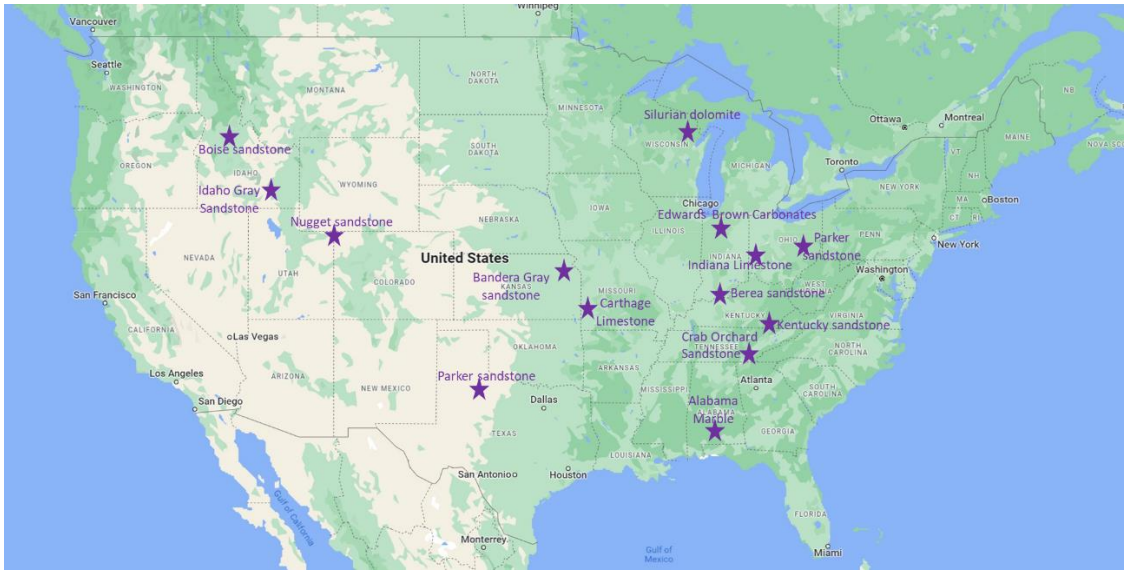


Figure 1: Locations of 15 natural rock samples in U.S.

Table 1 Sample information of 15 kinds of natural rocks

	Sample name	Location	Age	Lithology	Porosity	Permeability	Perm by	Homogeneous
1	Alabama Marble	Alabama	Early Paleozoic	Marble	1-2%	.001-.003	KCl	Yes
2	Bandera Gray Sandstone	Kansas	Paleozoic	Sandstone	19-21%	27-47 mD	KCl/N2	No
3	Berea Sandstone	Eastern Ohio, western Pennsylvania, western West Virginia, and eastern Kentucky	Late Devonian period	Sandstone	18-21%	80-120mD KCL, 200-230mD N2	KCl/N2	Yes
4	Boise Sandstone	Idaho	Late Miocene	Sandstone	28%	700-3000mD	KCl/N2	No
5	Carthage Marble (Burlington limestone)	Southwestern portion of the Missouri	Palaeozoic	Limestone	2-5%	.004-.007mD	KCl	NO
6	Crab Orchard Sandstone	Kentucky	Silurian	Sandstone	7-9%	0.075-0.14mD	KCl	Yes
7	Edwards Brown Limestone	Texas	Early Cretaceous	Limestone	40%	0.075-0.14mD	KCl	Yes
8	Kentucky Sandstone	Michigan, Ohio, Pennsylvania, West Virginia, Kentucky	Mississippian	Sandstone	14-19%	0.1-4mD	KCl	Yes
9	Idaho Gray Sandstone	Idaho	Neoproterozoic to Paleoproterozoic	Sandstone	29-30%	5-7 Darcy	N2	Yes
10	Indiana 2-4 Limestone	Indiana	Mississippian	Limestone	12-14%	2-4mD	KCl	Yes
11	Indiana 200 Limestone	Indiana	Mississippian	Limestone	17-19%	135-220mD	KCl/N2	Yes
12	Nugget Sandstone	Colorado, Idaho and Utah	Early Jurassic	Sandstone	10-12%	5-15mD	KCl	No
13	Parker Sandstone	Vermont	Cambrian	Sandstone	17-18%	15-30mD	KCl	No
14	Scioto Sandstone	Ohio	Silurian	Sandstone	16-18%	.001-.01mD	KCl	Yes
15	Silurian Dolomite	Wisconsin	Silurian	Dolomite	16-17%	35-100mD	KCl/N2	No

3. Hypothesis

1) The cumulative water imbibed is related to the volume of rock and the value of porosity.

2) The imbibition rate may be affected by the rock type, including characteristics of the contact angle, mineral composition, pore throat distribution. The different petrophysical may cause various imbibition rates.
3) Homogeneity rocks have a stable imbibition rate. Water imbibition rate heterogeneity rocks affected by the pore structure and pore throat in different directions.
4) The water imbibition rate has higher speed in higher-permeability direction, vice versa.

4. Methods

4.1 Liquid immersion porosimetry

Liquid immersion porosimetry, a porosity measurement by using liquid saturation and immersion technique with deionized water, was applied to measure the total porosity of rocks. The sample container for porosimetry tests is 16.0 cm in height and 11.5 cm in diameter which allows multiple rock samples to be tested in the same round. One round of vacuum saturation only takes one or two days at an extremely low price.

The principles of liquid immersion porosimetry are presented in the following equations with known parameters, such as the size of the rock samples, fluid density, and various weights obtained during the testing process.

$$V_g + V_o = V_b \tag{1}$$

$$\emptyset = \frac{V_o}{V_b} = \frac{\frac{W_s - W_d}{\rho_f}}{\frac{W_f}{\rho_f}} \tag{2}$$

$$\rho_b = \frac{W_d}{V_b} \tag{3}$$

$$\rho_g = \frac{W_d}{V_g} = \frac{\rho_b \cdot V_b}{(1 - \emptyset) \cdot V_b} \tag{4}$$

where $\emptyset$ is the porosity of sample; V_o is the void volume of sample in cm^3; V_b is the bulk volume of sample in cm^3; V_g is the grain volume in g/cm^3; W_s is the sample weight with liquid saturation after vacuum-pulling in g; W_d is over-dry weight of sample in g; W_f is submerged weight of sample in liquid in g; ρ_f is the liquid density in g/cm^3; ρ_b is the bulk density of the sample in g/cm^3; ρ_g is the grain density of the sample in g/cm^3.

(A) (B)

Figure 2: (A) Vacuum saturation setup; (B) Balance and Archimedes' bucket.

Before measuring samples with liquid immersion porosimetry, samples were recorded for the weights and sizes in an air-dry condition. Then, samples were dried in the oven at a constant temperature as 60°C for 48 hours. The long-time drying step aims at removing the moisture in the connected pores. After evaporating the possible liquid in the void space, samples were placed into a desiccator for 30 minutes. Once the samples cooled down, the stable oven-dry weight was recorded. The irregular or cylindrical samples can be placed in the vacuum chamber directly while the cubic samples need to be put into the custom-designed tray-holders first to keep track

of each individual sample. After all samples were placed in positions and the vacuum chamber closed, the pump started the evacuation process of air in connected pore space (Fig. 2A). In the first round of evacuation process, the pump evacuating air lasted 6 to 8 hours for small samples, or 12-16 hours for larger core plugs. The pressure in the chamber dropped to 0.06-0.1 torr (7.999-13.332 Pascal, or 99.99% vacuum) to make sure all the possible air in the connected pore space of samples was removed. In the next step, since the solubility of CO_2 in water is larger than the air's, CO_2 was injected into the vacuum chamber to replace residual air inside the samples. Then, the second evacuation process started to flush CO_2 for 12 to 18 hours. After completing the second round of evacuation, the liquid, for example, deionized water (after being boiled and cooled down), was injected into the vacuum chamber. The liquid invaded the edge-accessible pores of samples while 30 psi CO_2 further pressed deionized water into the samples. The immersion process lasted 3 to 4 hours. After completing the vacuum saturation, the water-saturated samples were weighted in the air and in the same fluid by the custom-designed Archimedes' bucket (Fig. 2B). Before weighing samples in the air, slightly moistened Kimwipes with DI water were used to wipe off the excess liquid attached to the sample surface. The custom-designed Archimedes' bucket was designed for weighing the submerged weight in the DI water. In the final step, Eqs. (1) to (4) were used to calculate the porosity, bulk volume, and grain density of these natural rocks.

4.2 X-Ray Diffraction

X-Ray Diffraction (XRD) analysis is a method for mineral identification. XRD is able to identify the crystal structure (the physical arrangement of elemental atoms) and quantitatively investigate the mineral composition of tested samples. The ideal powder size prepared for the normal XRD is 10~50 μm (Pecharsky et al., 2003), while the Rietveld analysis is better to decrease the powder

size to 1~5 μm (Snellings et al., 2014). Powder samples were measured at 40 KV and 30 mA, with the angle ranging from 8-70 degrees (Fig. 3).

Figure 3: Shimadzu MaximaX XRD-7000

4.3 Contact angle

Contact angle measurement can characterize the average wettability of natural rocks which has significant potential utilities (Decker et al., 1999). The mechanical equilibrium of the drop under the action of three interfacial tensions determines the value of the contact angle of a liquid drop on a solid surface (Kwok et al., 1999). The solid surface tensions in an air-liquid-rock system has been recognized by Young (1805).

$$\gamma_{la}cos\theta = \gamma_{sa} - \gamma_{sl} \quad (5)$$

where θ is the Young contact angle, γ_{la} is the liquid-air interfacial tension; γ_{sa} is the solid-air interfacial tension; γ_{sl} is the solid-liquid interfacial tension.

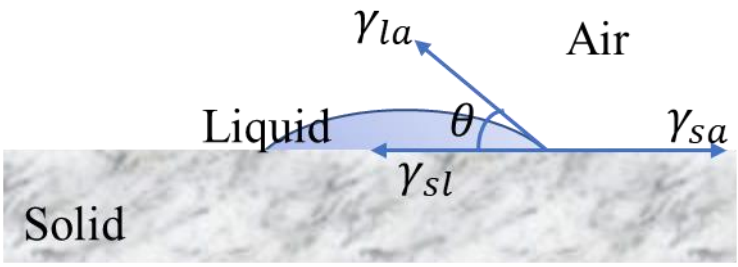


Figure 4: Schematic diagram of a sessile- drop contact angle system.

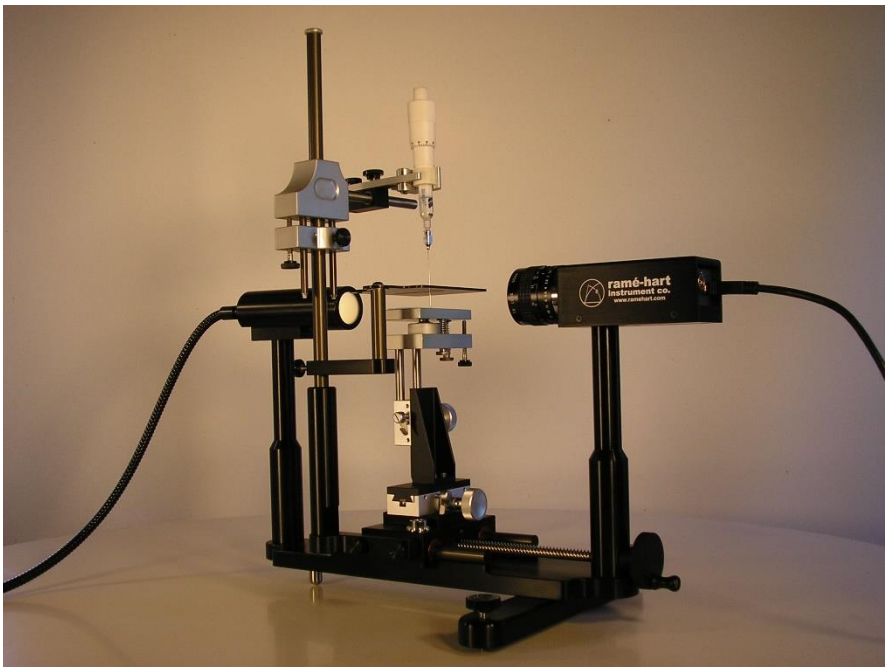


Figure 5: Ramé-hart Model 250 Tensiometer

In this study, Ramé-hart Model 250 Tensiometer was applied in measuring the contact angle of 15 natural rock samples. DI water represents hydrophilic liquid. The rock surface was polished flat and finely by sandpaper to reduce the impact of surface roughness. Samples were dried in the

oven before each test of contact angle. During measurements, each rock piece was required to be adjusted to be horizontal with the platform of Tensiometer. The related software was used for capturing the spreading process of droplets contacting the solids. After capturing the spreading process, the data of contact angle were obtained from these captured pictures.

However, when the observed solid surface is rough or contaminated, there will be contact angle hysteresis. Thus, the contact angle may vary from place to place and the Young's equation can only be applied as a local equilibrium (de Gennes, 1985). Since the surface roughness has a strong effect on the contact angle and wettability of a surface, the roughness ratio was introduced to modify the Young contact angle, which is shown as Eq. (6).

$$\cos(\theta_m) = r\cos(\theta_Y) \tag{6}$$

where θ_m is the Young's contact angle, and r is the roughness ratio being defined as the ratio between the actual and projected solid surface area.

Therefore, in this study, the contact angle of 15 rock samples was tested in different roughness conditions. The angle of contact that a liquid form with a solid surface holds a pivotal role in defining the capillary force that comes into play within porous materials or capillary tubes. This force emerges from the intermolecular interactions at the interface between the liquid, the solid, and the surrounding atmosphere. Furthermore, the capillary force exerts a substantial influence on the speed at which a liquid is absorbed into a porous substance, a phenomenon known as imbibition. Imbibition denotes the process through which a liquid is drawn into porous materials such as soil, paper, or sponges due to the effects of capillary forces. The pace at which imbibition unfolds is guided by multiple factors encompassing capillary force, porosity, permeability, and the inherent characteristics of both the liquid and the porous material. In the simulation model,

the contact angle in fractures determines the velocity of waterfront movement and the rate of imbibed water. Thus, the contact angle in various situations needs to be considered thoroughly.

4.4 Mercury Intrusion porosimetry (MIP)

Mercury intrusion porosimetry (MIP) is a potent technique employed to assess a diverse range of solid and powdered materials, providing insights into parameters such as porosity, pore size distribution, and pore volume. This characterization method utilizes a porosimeter, an instrument equipped with a pressurized chamber, to compel mercury to infiltrate the voids within a porous substrate. As pressure is applied, larger pores are filled initially, followed by a progression into progressively smaller pores. This technique enables the characterization of both inter-particle pores (between individual particles) and intra-particle pores (within particles themselves).

The Washburn Equation establishes a connection between applied pressure and pore diameter by utilizing physical properties of the non-wetting liquid (in this context, mercury). These physical properties encompass the contact angle between the material and the mercury, as well as the surface tension of mercury. The instruments employed at PTL facilitate pressure application spanning from approximately 1 psi up to 60,000 psi, corresponding to the measurement of pores ranging from about 250 µm down to 0.003 µm (3 nm). The contact angle of mercury on the material being tested holds significance in obtaining optimal outcomes. This angle can either be provided, measured, or default values can be utilized during the analysis. The sample is situated within a segment of the penetrometer cell, available in various volumes to accommodate powders or intact solid pieces. The dimensions of the sample are typically limited to approximately 2.5 cm in length and 1.5 cm in width. Mercury porosimetry finds extensive application in industries like catalyst and petrochemical, aiding in determining pore size and volume of catalyst substrates such as silica and alumina zeolites. Furthermore, the relatively high

quantity of mercury entrapped in the highly porous rocks creates artefacts that obliterate the rock structure in the micro-CT images (Zhang et al., 2016; Njiekak et al., 2018). The Lucas Washburn equation, in its broadest formulation, characterizes the extent of liquid penetration L into a capillary tube or pore over time, which is shown in Eq. (7). The value of the diffusion coefficient D is influenced by both the structure of the capillary and the characteristics of the infiltrating fluid.

$$L = (Dt)^{\frac{1}{2}} \tag{7}$$

Where, L is the penetration length of a liquid into a capillary pore or tube. D and t are simplified diffusion coefficient and time, respectively.

The relationship between a dynamic viscosity η, surface tension γ, pore radius r and penetration distance L are shown in Eq. (8).

$$L = \sqrt{\frac{\gamma r t cos(\theta)}{2\eta}} \tag{8}$$

Where, L, γ, r, θ, η are penetration distance, surface tension, pore radius, contact angle between the penetrating liquid and the solid (tube wall), and dynamic viscosity, respectively.

This equation also extends with some issues to imbibition into porous materials. Eq. (9) describes the relationship between capillary pressure, radius of the capillary and the contact angle.

$$P_c = \frac{2\sigma \cos\theta}{r} \tag{9}$$

where P_c is the capillary forces; r is the radius of the capillary; ρ is the density of the liquid; σ its surface tension; θ is same contact angle which is shown in Eq. (8).

The MIP method can be applied in analyzing the distinct pore-throat distribution frequency. For fine grained rock types such as mudstone and siltstone, their pore spaces are usually dominated by nanopores with pore-throats less than 0.05 μm while for coarse grained rock types such as sandstone and conglomerate, the high percentage of void spaces are occupied by large-sized megapores with pore-throats which are larger than 10 μm (Zhang et al., 2016, Yin et al., 2020). Mercury porosimetry data is generally analyzed by the Washburn equations, which mentioned in Eq. (7) and Eq. (8) (Watt-Smith et al., 2007; Khan et al., 2007). The Washburn equation, also known as Lucas–Washburn equation, describes capillary flow in parallel cylindrical tubes (Lucas, 1918; Washburn, 1921). The Washburn equation is able to be extended with some issues, for example, the imbibition into porous materials.

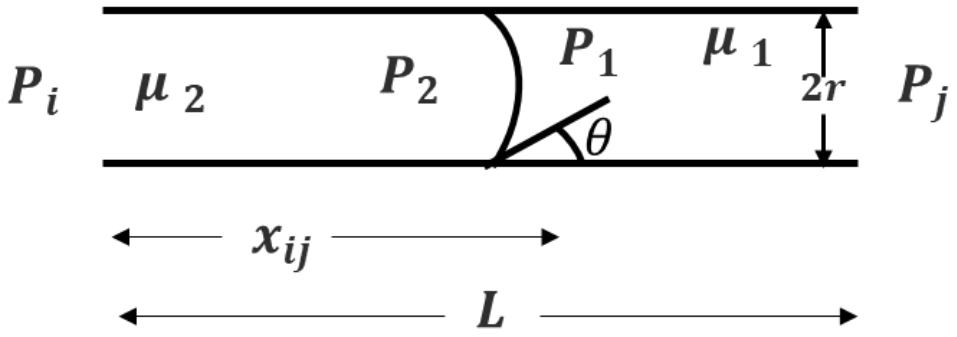


Figure 6: Capillary flow in a single tube

The volume flux $q_{ij}$from node i to node j in the tube are found from the Washburn equation (Washburn, 1921) for capillary flow with two fluids present as Eq. (8)

$$q_{ij} = -\frac{\pi r_{ij}^2 K_{ij}}{\mu_{eff}} \frac{\left(\Delta P_{ij} - P_c\right)}{L} \tag{8}$$

where $K_{ij} = \frac{r_{ij}^2}{8} = permeability\ of\ the\ rock$

$\mu_{eff} = \mu_2 x_{ij} + \mu_1(1 - x_{ij})$=weighted effective viscosity due to the two fluids

$\Delta P_{ij} = P_j - P_i$=pressure difference between nodes i and j.

$P_c = P_1 - P_2$=capillary pressure in the tube ij

x_{ij}=position of the pore-interface in the tube

r_{ij}=radius of the tube

Furthermore, the hysteresis represents the characterizations of raw porosimetry data (Rigby & Edler, 2002). The proposed semi-empirical alternatives to the Washburn equation for the analysis of raw mercury porosimetry data allow hysteresis to be characterized for the pore structure. The pore radius (nm) for mercury intrusion and mercury retraction are given by Eq. (9) and Eq. (10), respectively.

$$r = \frac{302.533 + \sqrt{91526.216 + 1.478p}}{p} \tag{9}$$

$$r = \frac{68.366 + \sqrt{4673.91 + 471.122p}}{p} \tag{10}$$

where p is the applied pressure (in MPa).

Micromeritics AutoPore 9520 (Fig. 7) was used to analyze the pore size distribution of cubic samples at 1 cm in length. For an MIP analysis, samples were oven-dried at 60°C oven for 48-hour drying, the same way as for other tests. When the samples cooled down to room temperature, the samples weight was recorded as the oven-dry weights. In the next step, samples

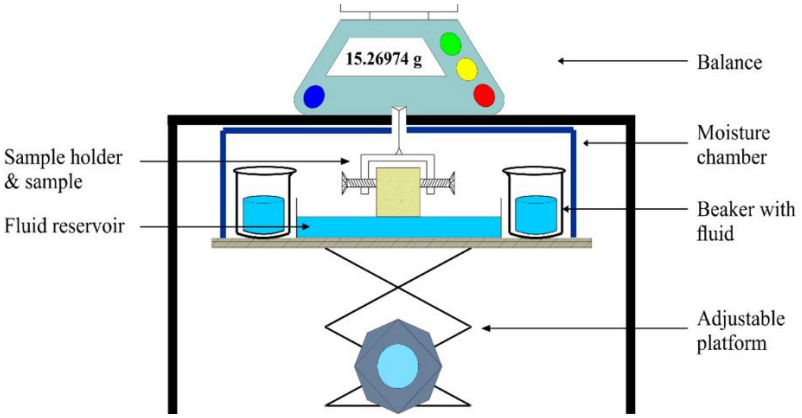
were assembled with a penetrometer and weighed as the assembly mass. The penetrometer is an analyzer tube connecting the sample with the MIP station. The penetrometer needs to be covered with vacuum grease for a good vacuum. The penetrometer was inserted into the low-pressure port to be evacuated to 0.05 mm Hg (0.05 torr, 6.666 pascals). After evacuation, the mercury invaded the penetrometer under 5 psi pressure. The pressure usually ranged from 5 psi to 30 psi for low-pressure analysis while 30 psi to 60,000 psi for high-pressure analysis. During the analyses, the mercury intrusion system will stay at a specific pressure for a while, which is the equilibrium time, to stabilize the volume reading and then move to the next pressure point, either increasing or decreasing the pressure. Each equilibrium time is set as 10 seconds for low-pressure and 60 seconds for high-pressure analyses. When completing the low-pressure analysis, the whole penetrometer assembly was weighted to obtain the bulk volume of the sample. Then the penetrometer was placed in the high-pressure port for subsequent analyses.

Figure 7: Micromeritics AutoPore 9520.

4.5 Liquid imbibition tests

The spontaneous imbibition (SI) experiments were executed in a controlled laboratory environment utilizing either cubic or cylindrical specimens with sides sealed using epoxy, maintained at a consistent temperature of 23°C. The procedural nuances and subsequent data analysis adhered to methodologies delineated by Hu et al. (2001) and Wang et al. (2021). The spontaneous imbibition setups and the sketch illustration of spontaneous imbibition are shown in Fig. 8 (A) and (B). Alterations in sample mass relative to time were meticulously captured by a high-precision balance, boasting a resolution of 0.01 mg (Rad Wag AS 60/220. R2). Each specimen was treated with a quick-drying epoxy on all facets save for two antipodal surfaces. Subsequently, the sample was suspended beneath a balance using a designated holder. To augment the ambient relative humidity and curtail sample evaporation, dual water beakers were strategically positioned within the imbibition chamber. Initiation of an imbibition test involved leveraging an adjustable jack to elevate the reservoir of deionized water, ensuring the sample's lowermost 1mm was immersed.

(A)


15.26974 g
Balance
Moisture chamber
Sample holder & sample
Fluid reservoir
Beaker with fluid
Adjustable platform

(B)

Figure 8: Spontaneous imbibition tests (A) Setups in the lab (B) Illustration of spontaneous imbibition setup (Wang et al., 2021).

5 Experimental results

5.1 Mineral compositions

Based on XRD analysis, the mineral compositions of 15 natural rocks have been identified and are outlined in Table 2. The studied rocks are arranged in order of type, and their permeability, ranging from low to high. In this study, the sandstone lithologies, including Bandera Gray, Scioto, Crab Orchard, Kentucky, Nugget, Parker, Berea, Boise, and Idaho Gray, exhibit a substantial modal abundance of quartz and feldspar, spanning from 34% to 90%. Conversely, the investigated carbonate rocks encompassing Indiana limestone, Silurian dolomite, Alabama marble, and Carthage marble, display an exceedingly dominant presence of calcite or dolomite, constituting nearly 100% of their mineralogical composition.

Of particular note, the Bandera Gray sandstone is distinguished by the relatively elevated occurrence of a distinctive mineral, Ankerite. This mineralogical trait is atypical within the context of sandstone formations and is seldom encountered in other sandstone varieties. Within the realm of geological relationships, an intriguing observation pertains to the interplay between porosity and permeability. In the case of sandstone formations, a distinct pattern emerges wherein porosity does not exhibit a strong linear correlation with permeability. This phenomenon is evident in instances such as Crab Orchard sandstone and Kentucky sandstone, where despite their similar porosities, their permeabilities differ significantly, registering values of 0.08 mD and 4 mD, respectively.

Another noteworthy illustration of this pattern can be observed in the Berea sandstone and Boise sandstone counterparts. The Berea sandstone manifests higher porosity, accommodating around 18% void space within the rock matrix, whereas the Boise sandstone holds a porosity of 8%. Surprisingly, the Berea sandstone showcases relatively low permeability, measuring 220 mD, whereas the Boise sandstone exhibits a substantially higher permeability of 500 mD. This intriguing phenomenon underscores the intricate and multifaceted nature of the porosity-permeability relationship within these sandstone formations.

Table 2 Measured XRD composition of 15 natural rocks

		Composition (%)											Geophysics		
	Name	Quartz	K-feldspar	Feldspar	Calcite	Dolomite	Pyrite	Gypsum	Anhydrite	Ankerite	Siderite	Clay	Poro(%)	Perm(mD)	Homo/Heter
Sandstone	Bandera Gray sandstone	52.1	0.5	17.3	0.3	0	0.3	0	1.7	14.5	0	13.2	17.69	0.002	Heter
	Scioto sandstone	76.2	5.8	11	0.2	0	0	0	0.2	0.1	0	6.5	0.1433	0.01	Homo
	Crab Orchard Sandstone	90	0	5	0	0	0	0	0	0	0	5	6.76	0.08	Homo
	Kentucky sandstone	71.6	1.4	20.4	0.2	0	0	0	0.5	0	0.3	5.8	5.07	4	Homo
	Nugget sandstone	93.4	3.4	0	1.2	0	0	0	0	0.3	0	1.7	12.54	10	Heter
	Parker sandstone	85.2	0.4	5.5	0.1	0.2	0	0	0	0	0	8.6	7.32	11	Heter
	Berea sandstone	82.2	4.1	3.7	0.4	0	0	0	1.4	0.1	0.6	7.4	18.24	220	Homo
	Boise sandstone	34.9	17.6	35.1	6.7	0	0	0	0.2	3.8	0.4	1.2	8.38	500	Heter
	Idaho Gray Sandstone	49.2	18.4	28.8	0	0	0	0.4	0.3	1.5	0	1.4	28.23	6000	Homo
Limestone	Indiana Limestone (Perm 2 mD)	0	0	0	100	0	0	0	0	0	0	0	13.85	3	Homo
	Indiana Limestone (Perm 200 mD)	0	0	0	100	0	0	0	0	0	0	0	8.99	200	Homo
	Silurian dolomite	1.9	0	0.5	2.3	95.2	0	0	0	0	0	0	0.0895	70	Heter
Marble	Alabama Marble	0	0	0	99	0	0	0	0	0	0	0	0.84	0.002	Homo
	Carthage Marble	0.3	0	0	99.7	0	0	0	0	0	0	0	0.91	0.003	Heter
Carbonate	Edwards Brown Carbonates	2.3	0	0	23.6	70.5	0	0	0	0	0	3.6	21.25	100	Heter

5.2 Contact angle

Contact angles of deionized water (DIW) were determined utilizing the sessile drop contact angle methodology. Table 3 provides an overview of the data encompassing contact angle measurements at the water-air interface with respect to the surface of the rock specimens. The measurements reveal notable trends, particularly in the context of Alabama and Carthage marbles, where substantial contact angles are observed.

Within the realm of wettability or hydrophilicity, the assessment pivots upon the magnitude of the contact angle. A contact angle measuring below 90 degrees denotes an escalated level of wettability or hydrophilicity, indicating a pronounced affinity for water. Conversely, when

contact angles surpass 90 degrees, they point towards diminished wettability, implying a proclivity for hydrophobic behavior in response to the water employed in the tests. Evidently, the measurements unveil a consistent hydrophilic characteristic across all rock samples that underwent testing.

Table 3 Acquisition of the contact angle measurements between the water-air interface and the surface of the rock sample.

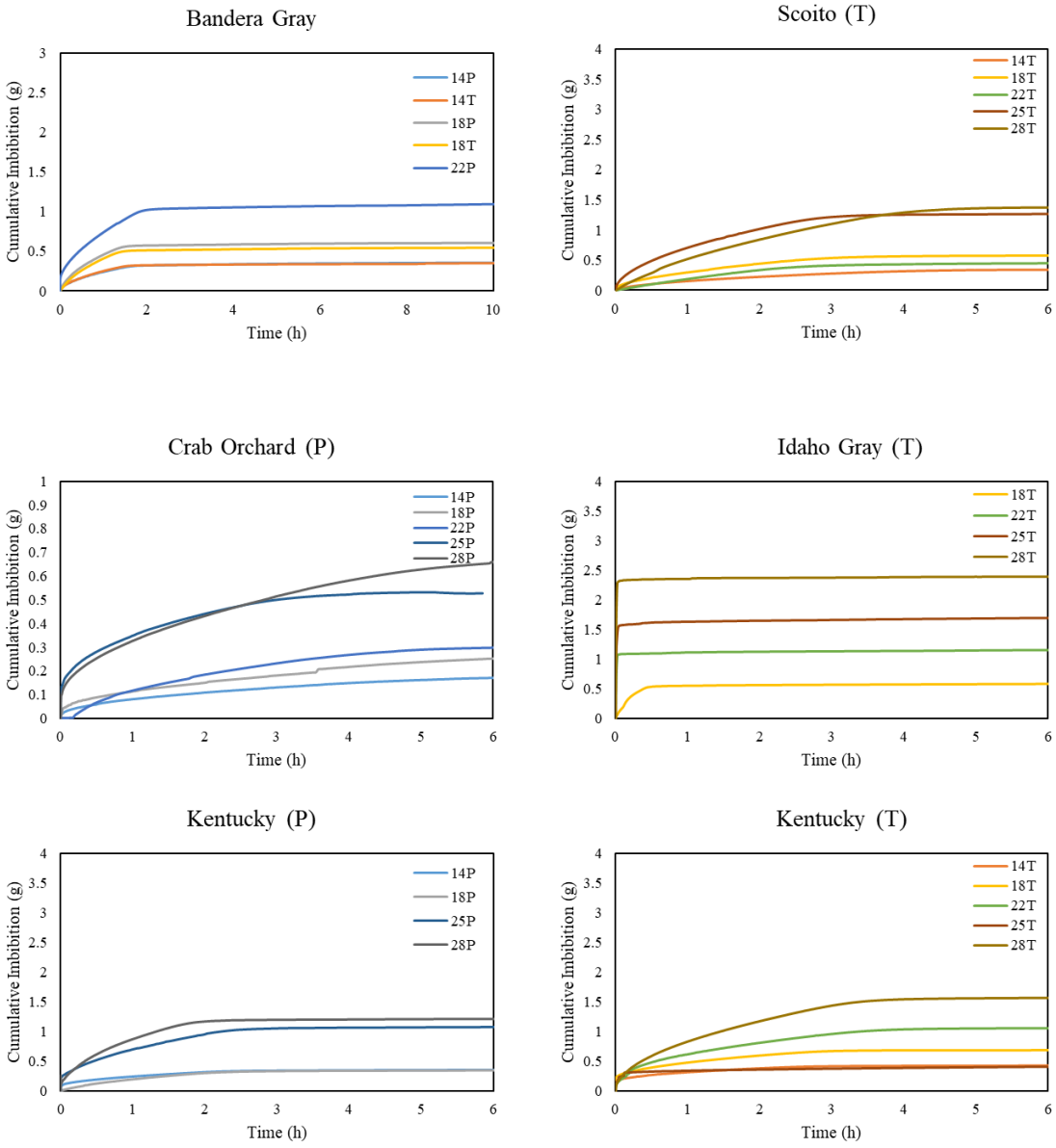
Sample	Contact angle (°)	Sample	Contact angle (°)
Alabama	79.695	Indiana Perm 2	38.956
Bandera Gray	25.560	Indiana Perm 200	27.407
Berea	32.421	Kentucky	47.215
Boise	49.015	Nugget	33.803
Carthage	75.072	Parker	27.350
Crab Orchard	25.573	Scioto Sandstone	24.534
Edwards Brown	57.529	Silurian	47.429
Idaho Gray	5.000		

5.3 Imbibition tests of natural rock samples

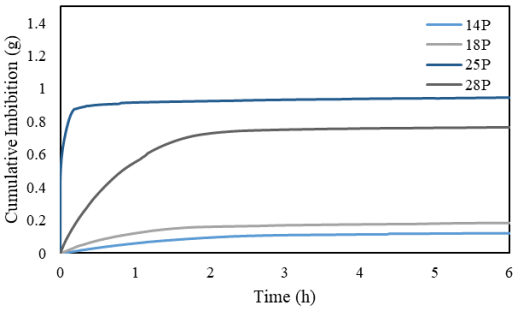
The imbibition process of 15 kinds of rock samples is tested and the corresponding cumulative imbibed water weight is recorded. The collected data are shown in Fig. 8.

Fig. 8 (C) delineates the spontaneous imbibition of water by Alabama Marble. This particular type of marble is characterized by its relatively low porosity and large contact angle, compared to other rocks used in this study. Upon contact with water, the capillary force entraps the water within the rock's surface. However, due to the tight matrix, the movement of imbibed water throughout the entire process is significantly impeded. In a similar vein, Fig. 8 (C) illustrates Carthage Marble, which exhibits lower porosity and permeability compared to Alabama Marble. Consequently, the addition of imbibed water occurs slowly, even during the initial stages,

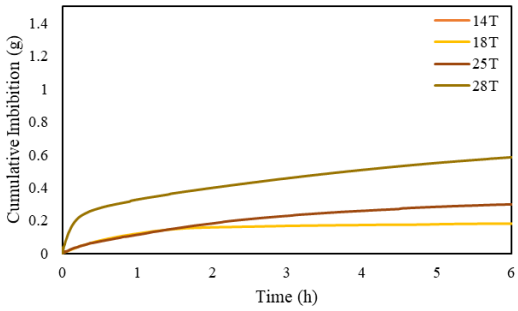
resulting in only minor changes in the imbibition slope. The slope of imbibed water in the marble matrix is not as pronounced as that seen in sandstone matrices.



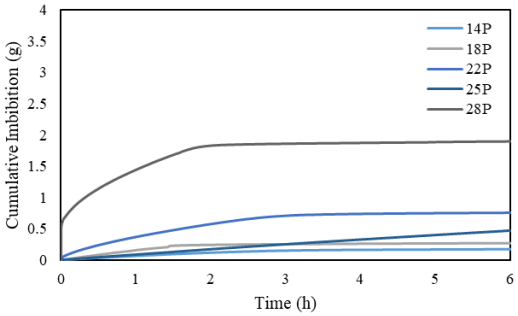
Nugget (P)



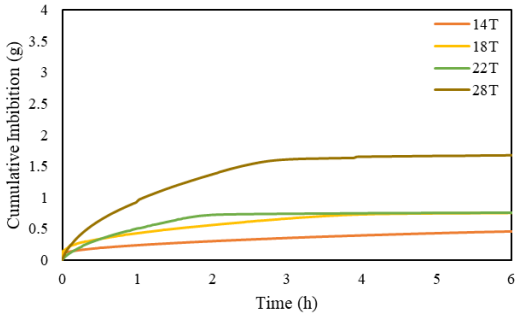
Nugget (T)



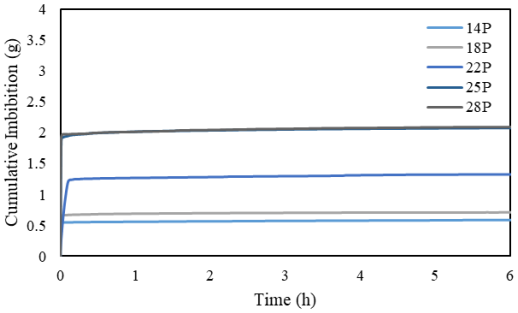
Parker (P)



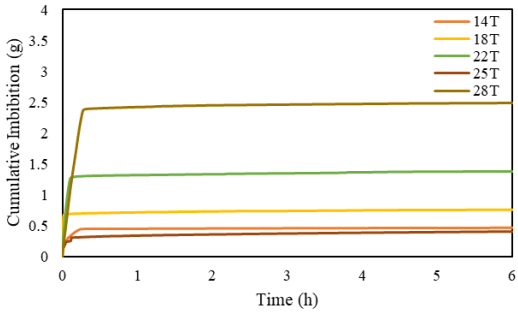
Parker (T)



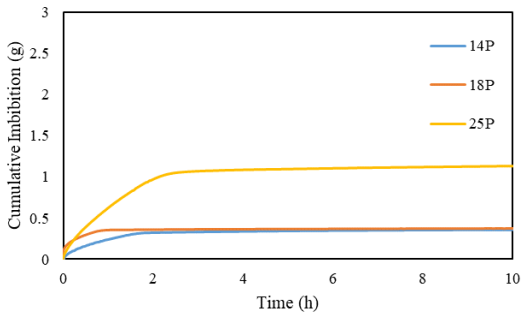
Boise (P)



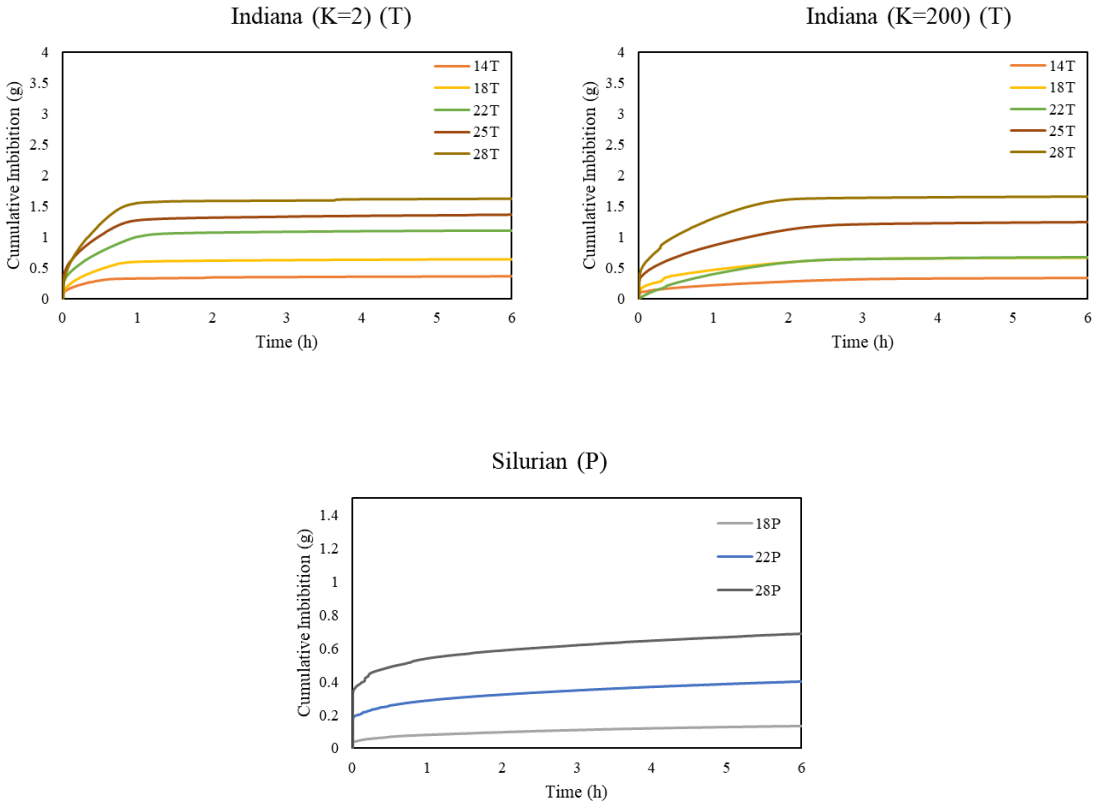
Boise (T)



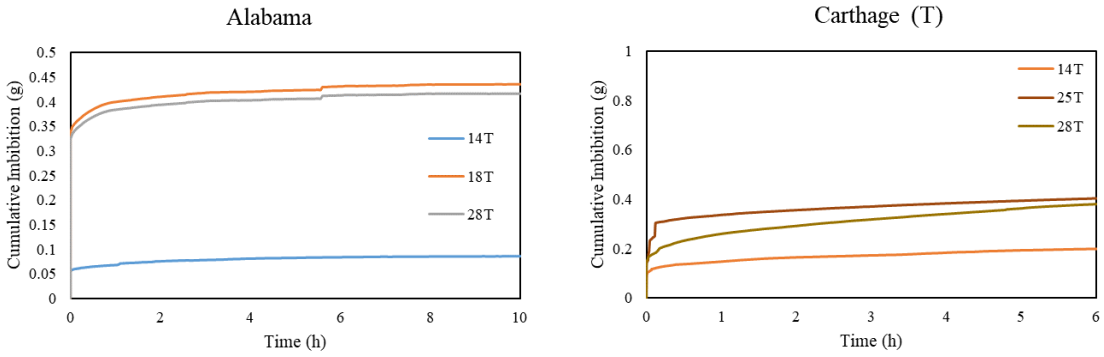
Berea



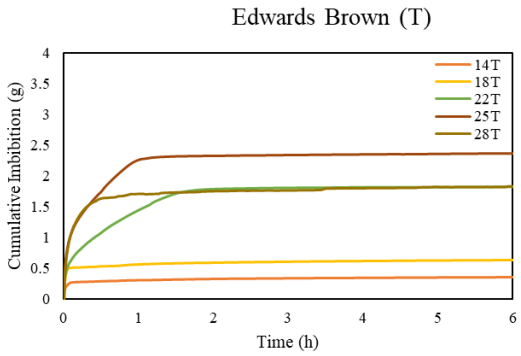
(A) Sandstone



(B) Limestone



(C) Marble



(D) Carbonate

Figure 9 Cumulative imbibition weight gains over experimental times; sample ID is presented as the diameter (e.g., The rock sample is cut into a 14 mm diameter size, with the layer oriented along the P-direction, referred to as size 14T) for a cylindrical core.

Fig. 9 represents a compilation of the imbibed water weights measured across 15 types of naturally occurring rocks, demonstrating the unique imbibition characteristics of different rock types. The spontaneous imbibition test was divided into two stages, noise phase (seconds to a few minutes) and matrix phase (several minutes to hours). During the initial phase, the interaction between the sample surface and water can potentially induce vibrations, resulting in an alteration in the recorded weight. Following this vibrational phase, the balance reading solely reflects the weight change from the imbibition of the matrix and any potential fractures (Hu et al., 2019; Ning et al., 2022).

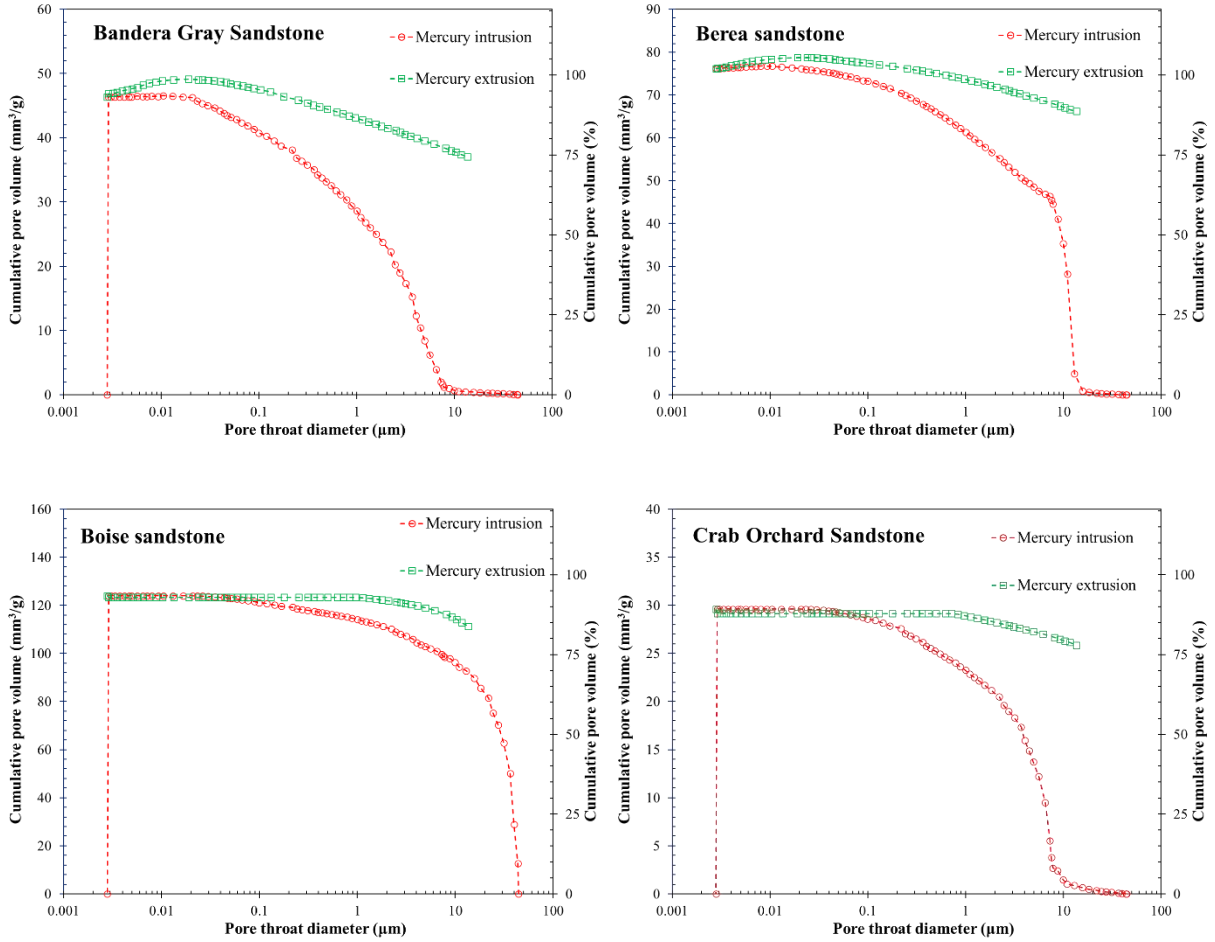
Fig. 9 (A) delineates the spontaneous imbibition of water by Alabama Marble. This particular type of marble is characterized by its relatively low porosity and large contact angle, compared to other rocks used in this study. Upon contact with water, the capillary force entraps the water within the rock's surface. However, due to the tight matrix, the movement of imbibed water throughout the entire process is significantly impeded.

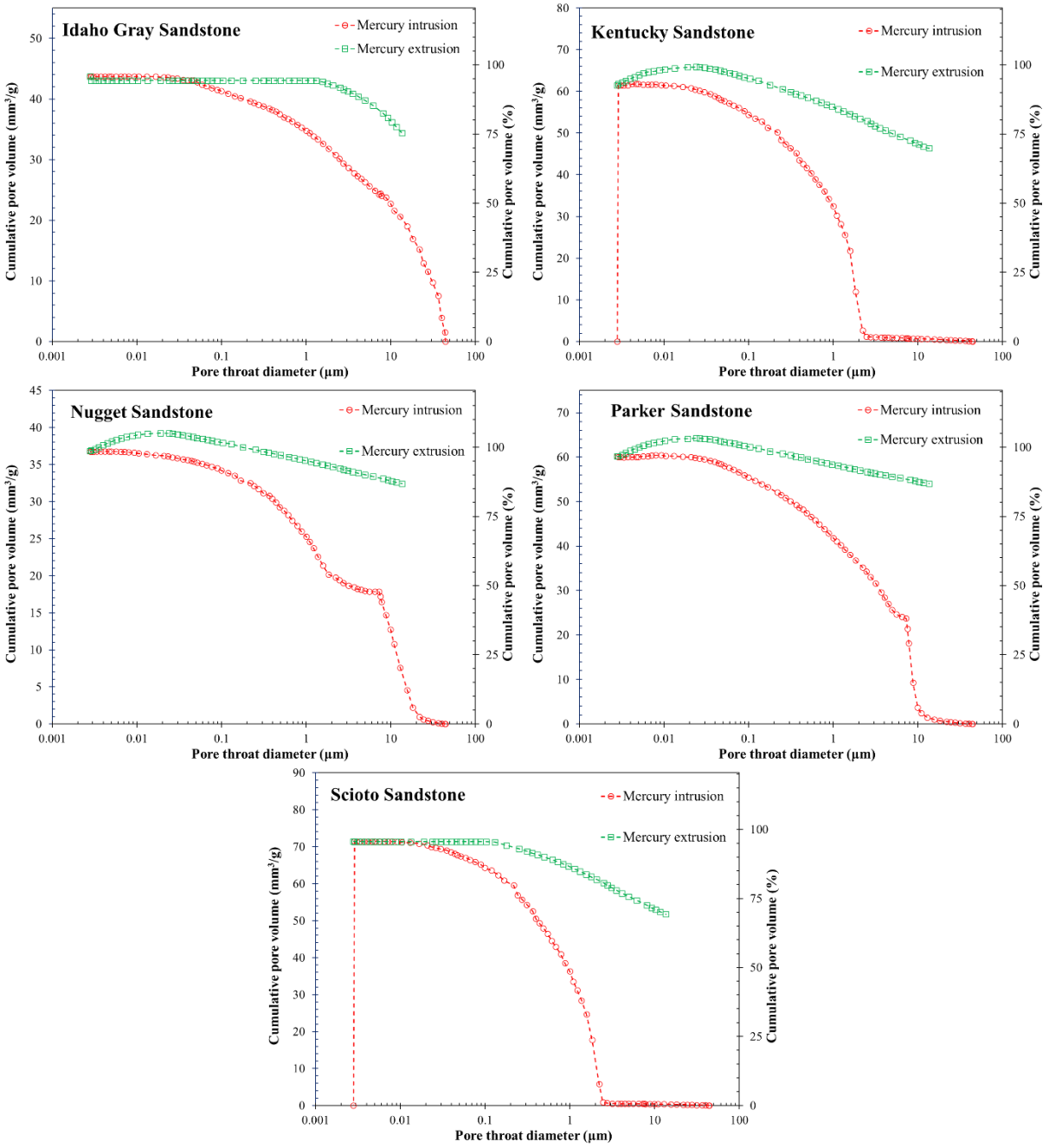
In a similar vein, Fig. 9 (A) illustrates Carthage Marble, which exhibits lower porosity and permeability compared to Alabama Marble. Consequently, the addition of imbibed water occurs slowly, even during the initial stages, resulting in only minor changes in the imbibition slope. The slope of imbibed water in the marble matrix is not as pronounced as that seen in sandstone matrices.

Other figures of Fig. 9 (B) demonstrate the water imbibition process in a variety of sandstones, namely, Bandera Gray, Berea, Boise, Crab Orchard, Idaho Gray, Kentucky, Nugget, Parker, Scioto, and Silurian. These rock types share a commonality in their cumulative imbibed water records. They all exhibit two distinguishable stages of water imbibition, both of which are prolonged compared to the same stages in marble. Notably, the rate of imbibed water weight gain remains relatively stable, and there is a sudden slow down through the process. Therefore, the imbibition process within the matrix phase can be demarcated into two stages. The first stage is characterized by the filling of main pores which pore diameter is over 1 μm. The collection of pore diameter is shown in Fig. 9. The subsequent stage involves water vapor or imbibed water filling the minor or exceedingly small pores in sandstones, which can be concluded in Fig. 10. The measurements of the penetrated mercury volume for 15 types rocks are documented in Fig. 10. The pressure is incrementally augmented until even the finer pores become impregnated with mercury. The volumes of mercury intrusion at distinct pressure levels are then employed to construct a capillary pressure versus saturation curve. Despite the minor pores' less significant role in the initial stage of pore filling, their impact remains relevant throughout both stages.

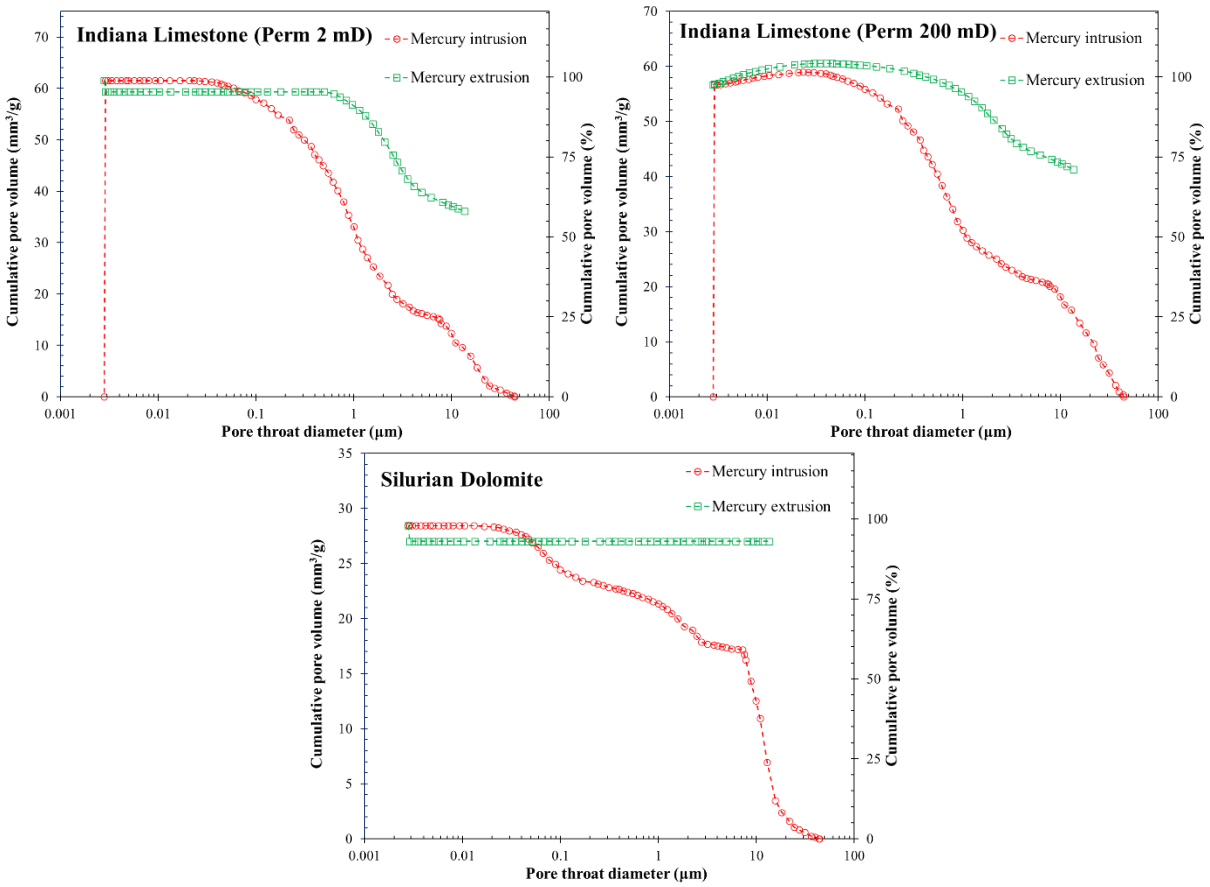
Fig. 9 (C) and (D) exhibit the imbibition process in Edwards Brown carbonate and Indiana limestones. Notably, these limestones display identical imbibition characteristics. The primary

pore filling phase in these limestones is relatively brief, which is congruent with their lower porosity when compared to typical sandstones.

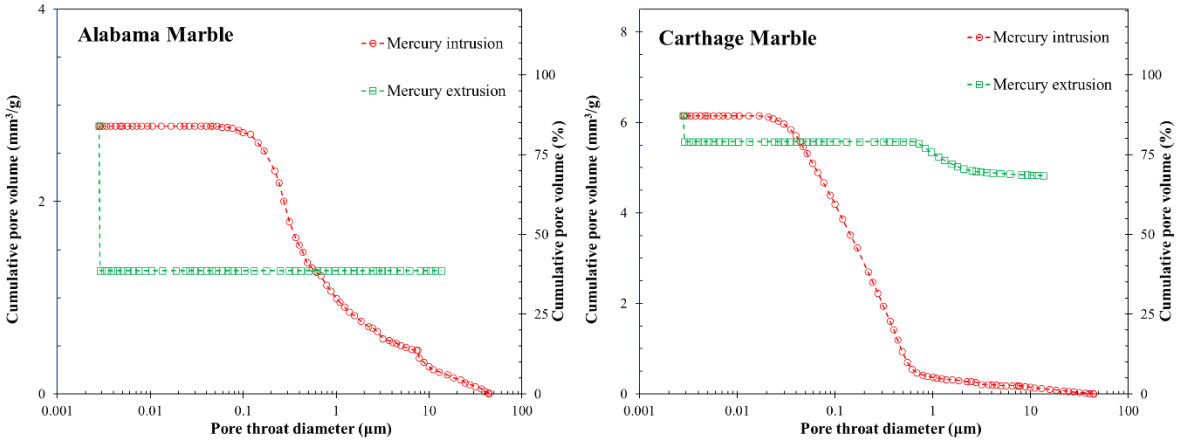



Idaho Gray Sandstone
Mercury intrusion
Mercury extrusion
Cumulative pore volume (mm³/g)
Cumulative pore volume (%)
Pore throat diameter (μm)
Kentucky Sandstone
Mercury intrusion
Mercury extrusion
Cumulative pore volume (mm³/g)
Cumulative pore volume (%)
Pore throat diameter (μm)
Nugget Sandstone
Mercury intrusion
Mercury extrusion
Cumulative pore volume (mm³/g)
Cumulative pore volume (%)
Pore throat diameter (μm)
Parker Sandstone
Mercury intrusion
Mercury extrusion
Cumulative pore volume (mm³/g)
Cumulative pore volume (%)
Pore throat diameter (μm)
Scioto Sandstone
Mercury intrusion
Mercury extrusion
Cumulative pore volume (mm³/g)
Cumulative pore volume (%)
Pore throat diameter (μm)

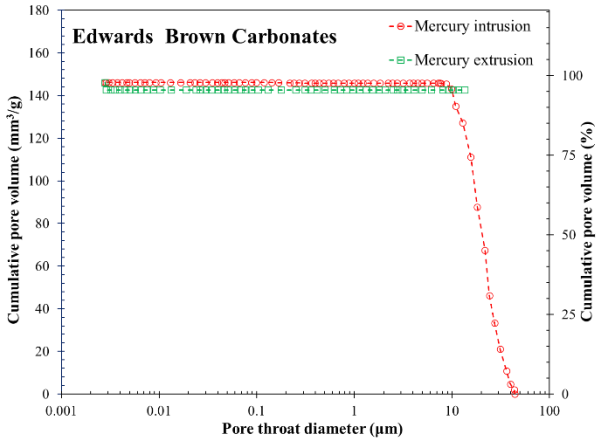
(A) Sandstone



(B) Limestone



(C) Marble



(D) Carbonate

Figure 10 Mercury intrusion porosimetry in 15 kinds of natural rocks.

6. Simulation and prediction models

In the realm of geology, the utilization of machine learning models to simulate imbibition tests brings forth a host of advantages that enhance our understanding of porous materials' behavior. Imbibition, a process involving the infiltration of liquids into porous substances, is a complex phenomenon influenced by intricate interactions at the pore scale. Machine learning models offer the capability to capture and analyze intricate relationships within vast datasets, enabling the prediction and interpretation of imbibition behaviors that might otherwise be challenging to deduce through traditional analytical methods. By leveraging these models, researchers and geologists can simulate imbibition processes with a higher level of accuracy, uncover hidden patterns, and identify critical factors affecting fluid penetration, such as pore structure, wettability, and rock composition. These simulations not only provide insights into fluid dynamics within porous media but also facilitate the optimization of resource extraction, reservoir management, and the design of enhanced oil recovery strategies, ultimately contributing to more informed decision-making and advancements in geologic sciences.

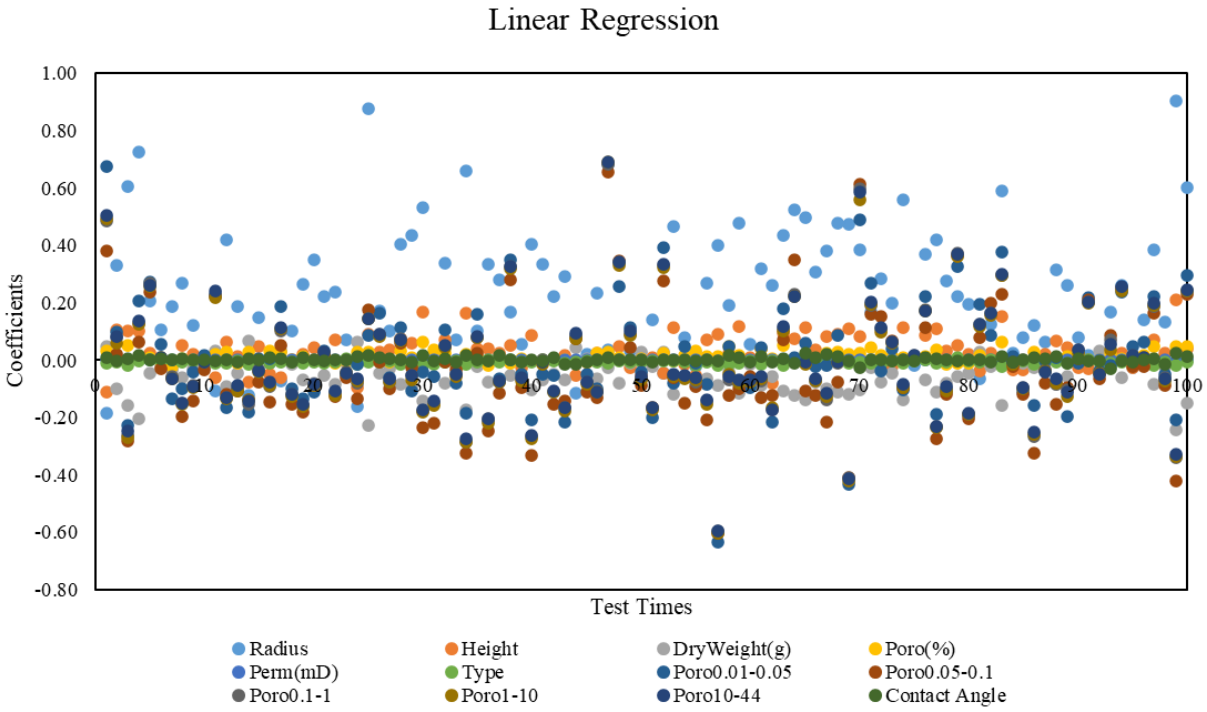
Table 4 enumerates the mean squared error (MSE) of each natural rock type tested within various machine learning models, including linear regression, Decision Tree, Random Forest, and Gradient Boosting regressions. The MSE quantifies the average squared discrepancies between the actual and predicted values within the dataset, thereby offering an estimation of the dispersion of residuals. A lower MSE and a higher coefficient of determination (R2) signify a more accurate and explanatory model, indicating the model's capacity to predict the simulation outcomes effectively.

Table 4 The evaluation of imbibition model of natural rocks in machine learning models

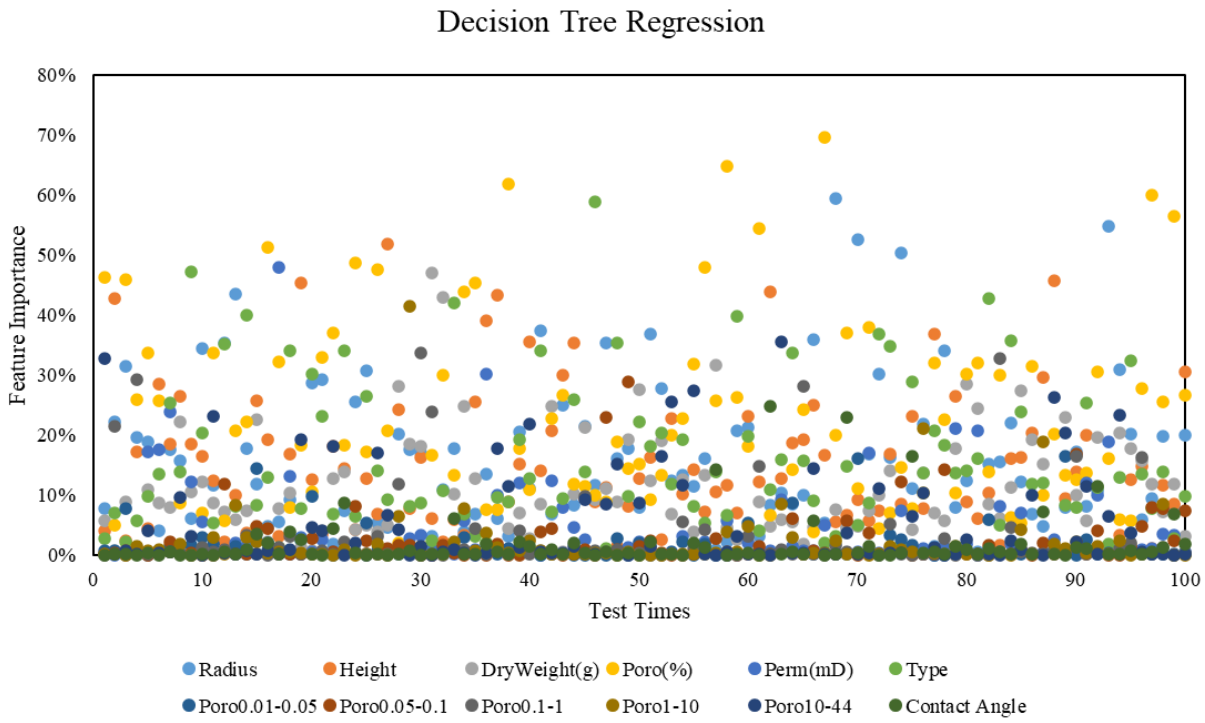
	Linear		Decision Tree		Random Forest		Gradient Boosting	
Name	MSE	R^2	mse	r2	mse	r2	mse	r2
Alabama	0.01319	-0.9854	0.00198	0.70147	0.00112	0.83175	0.00194	0.70854
Bandera Gray	0.0014	0.65123	0.00463	-0.1553	0.00319	0.20515	0.0036	0.10159
Berea	0.12725	0.75383	0.04838	0.9064	0.03267	0.93679	0.03824	0.92602
Boise	9.96221	0.12795	0.5469	0.95213	0.48451	0.95759	0.5464	0.95217
Carthage	0.00642	0.23761	0.00079	0.90596	0.00056	0.93298	0.00067	0.92092
Crab Orchard	0.01063	0.58832	0.00557	0.78427	0.00324	0.87454	0.00487	0.81141
Edwards Brown	0.39483	0.77972	0.06689	0.96268	0.05018	0.972	0.04925	0.97252
Idaho Gray	0.01051	0.54718	0.00241	0.89621	0.00289	0.87552	0.00239	0.897
Indiana200	0.37272	0.28407	0.02049	0.96065	0.00598	0.98851	0.01777	0.96587
Indiana2	0.01878	0.90799	0.00927	0.95456	0.00611	0.97004	0.00627	0.96926
Kentucky	0.16925	0.4668	0.03765	0.88138	0.02297	0.92764	0.02863	0.90981
Nugget	3.89E-05	0.55437	8.70E-06	0.90032	8.10E-06	0.9071	8.65E-06	0.90088
Parker	2.15E-05	0.78655	2.32E-06	0.97702	1.81E-06	0.98206	1.89E-06	0.98123
Scioto	0.01529	0.62668	0.01652	0.59654	0.01087	0.73458	0.01067	0.73942
Silurian	0.02074	0.19785	0.00406	0.84294	0.00201	0.92219	0.00382	0.85215

The focal column of interest pertains to the rate at which water weight gain occurs in the models through imbibition. The diverse rock types serve as training features. Within each data row, information includes the rock's designation, layer direction, radius, height, pore throat distribution, contact angle, and the current rate of imbibed water weight gain. As the models

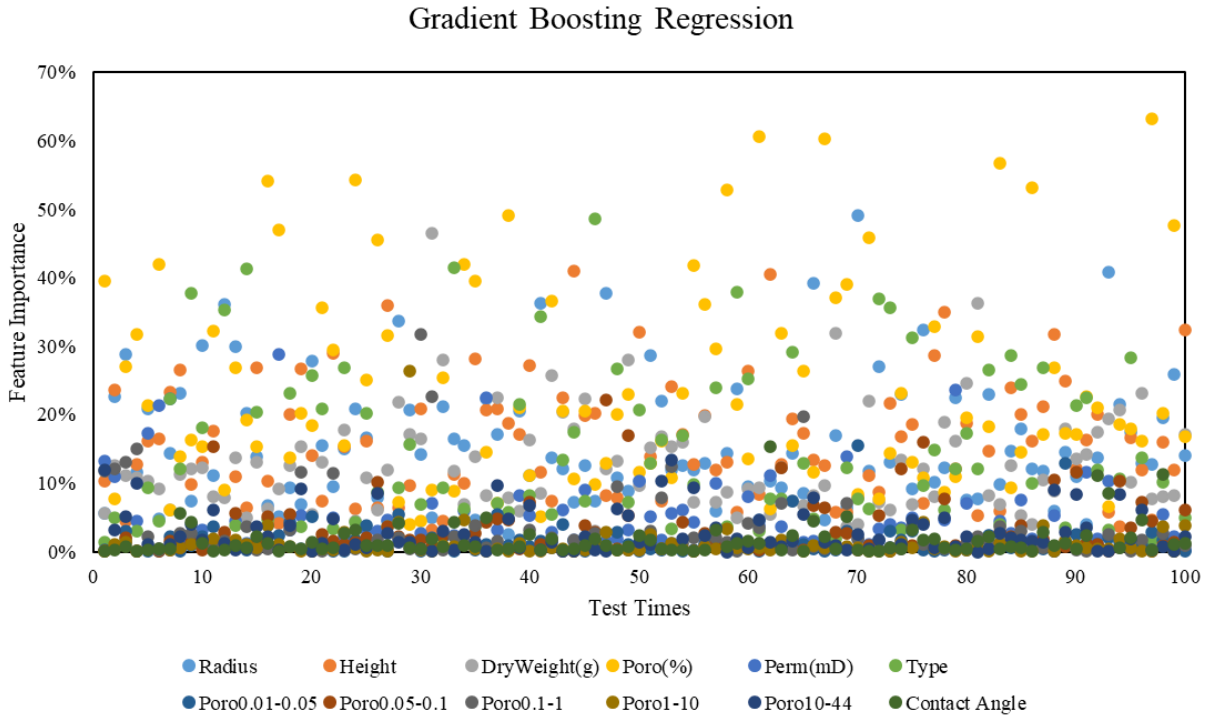
accumulate a sufficient historical record, they are poised to forecast the forthcoming state within a dynamic process. As evidenced in Table 4, the decision tree, gradient boosting, and random forest models exhibit high R2 values. This observation implies that for each rock type, a suitable machine learning model can be discerned, capable of accurately depicting the water imbibition process.



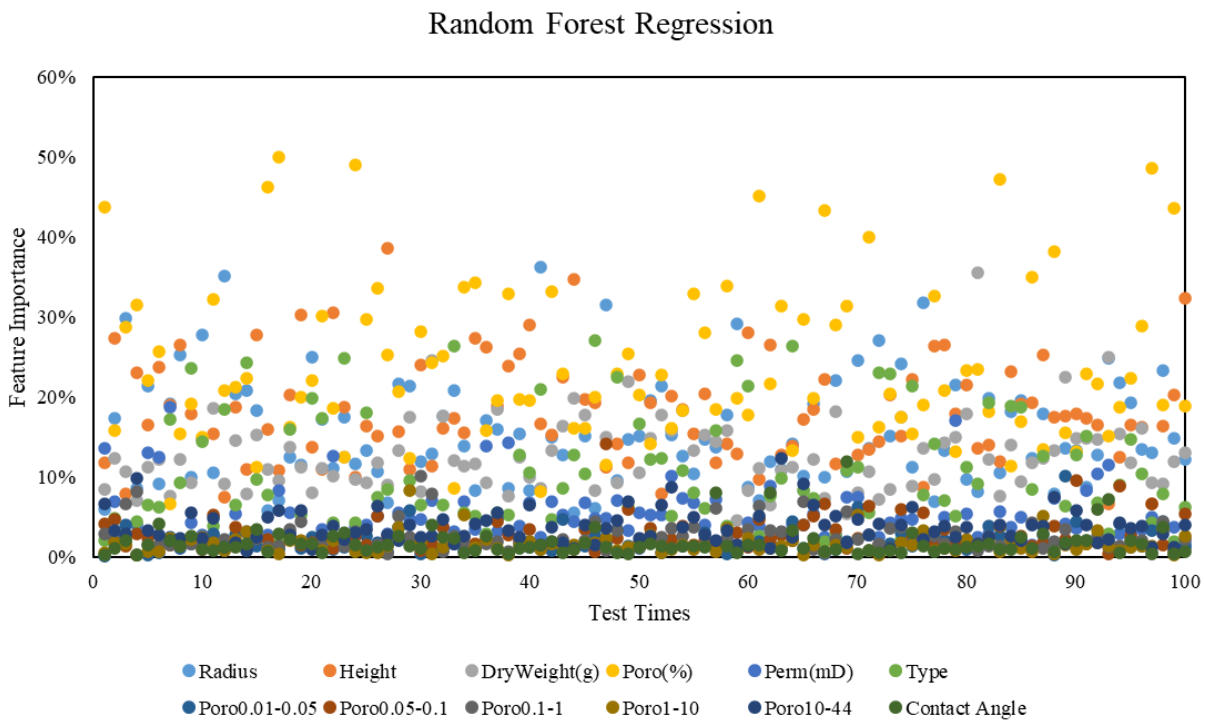
(A)


Decision Tree Regression
Feature Importance
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
10
20
30
40
50
60
70
80
90
100
Test Times
Radius
Height
DryWeight(g)
Poro(%)
Perm(mD)
Type
Poro0.01-0.05
Poro0.05-0.1
Poro0.1-1
Poro1-10
Poro10-44
Contact Angle

(B)


Gradient Boosting Regression
Feature Importance
70%
60%
50%
40%
30%
20%
10%
0%
0
10
20
30
40
50
60
70
80
90
100
Test Times
Radius
Height
DryWeight(g)
Poro(%)
Perm(mD)
Type
Poro0.01-0.05
Poro0.05-0.1
Poro0.1-1
Poro1-10
Poro10-44
Contact Angle

(C)



(D)

Figure 11: The 100 test times collections of coefficients and feature importance in regressor models. (A) coefficients in linear regressor model. (B) Feature importance in Decision Tree Regression model. (C) Feature importance in Gradient Boosting Regression model. (D) Feature importance in Random Forest Regression model.

Fig. 11 displays the aggregation of coefficients and feature importance from 100 test runs in various regression models. If the model risks being influenced by an excessive amount of data from a single rock type, a precautionary measure is taken. All data sourced from the monitoring recorder will be redistributed across 100 distinct files, each carefully curated to contain a balanced representation of every rock variety. This approach ensures that the model's predictive capabilities remain uninfluenced by the dominance of specific rock characteristics, allowing it to

forecast the subsequent steps based on the current status impartially. Fig. 11 (A) Represents the coefficients determined in the linear regression model. Fig. (B) Shows the feature importance as identified in the Decision Tree Regression model. Fig. (C) Illustrates the feature importance as determined in the Gradient Boosting Regression model. Fig. (D) Demonstrates the feature importance as identified in the Random Forest Regression model. These four figures delineate the coefficients or feature importance for four types of regression models, encompassing data related to imbibed water, porosity, permeability, and pore-throat distribution. Despite the significant physical disparity among the various rock types, the composite results derived from the simulations do not fare as well as those generated from separate simulations conducted via machine learning, as indicated in Table 3. As inferred from these figures, the size of the rock samples considerably influences the velocity of water imbibition. Comparatively, the pore-throat distribution and contact angle exert less pronounced effects.

7. Conclusions

Chapter IV meticulously explores the imbibition process across diverse rock types, including a variety of sandstones, limestones and marbles, revealing distinct behaviors in marbles, sandstones, and limestones, with specifics such as Alabama marble's restrained water uptake due to its low porosity and high contact angle. Through rigorous analysis, machine learning models have been authenticated as potent tools for predicting these processes, where lower mean squared error (MSE) and higher R^2 values signify enhanced model accuracy. Crucially, while aggregate simulation results did not supersede the efficacy of individual machine learning models, this investigation notably augments our comprehension of water imbibition mechanics in various rock matrices and underscores the predictive proficiency of computational models in this domain.

Alabama marble displayed spontaneous water imbibition with its low porosity and high contact angle impeding the process. Carthage marble showed a lower porosity and permeability with minor slope alterations. Sandstones shared a two-stage imbibition process with a stable rate of water weight addition and a sudden shift in the middle. The first stage involved filling the main pores, whereas the second filled the minor ones. Indiana and Carthage limestones presented a brief primary pore-filling phase due to lower porosity but exhibited a similar second phase to sandstones.

The use of machine learning models to predict these processes effectively was validated. The Mean Squared Error (MSE) of each rock type in various models provided a dispersion of residuals. Lower MSE and higher R^2 were indicative of a more accurate model, capable of predicting simulation outcomes effectively. This chapter discusses the coefficients and feature importance of various regression models. Despite significant physical differences among rock types, combined simulation results were less effective than individual machine learning models. Rock sample size considerably influenced the velocity of water imbibition, with pore-throat distribution and contact angle having lesser effects. The figures consistently reveal a recurring outcome. Among the factors examined, including sample size, porosity, and permeability, these primarily govern the rate of water weight gain through imbibition. Meanwhile, the impact of rock type and contact angle remains relatively marginal across these three machine learning models. This study enhances our comprehension of water imbibition within a diverse range of 15 natural rock types, as well as the effectiveness of machine learning models in forecasting such intricate processes.

References

Alhammadi, A. M., AlRatrout, A., Singh, K., Bijeljic, B., Blunt, M. J., 2017. In situ characterization of mixed-wettability in a reservoir rock at subsurface conditions. Scientific Reports, 7(1), 10753.

Al-Ghamdi, A. H., 2019. X-ray diffraction and gamma-ray analysis of rock samples from Haradh Region in Saudi Arabia. Journal of Radiation Research and Applied Sciences, 12(1), 87–92.

Austad, T., Shariatpanahi, S.F., Strand, S., Aksulu, H., Puntervold, T., 2015. Low salinity EOR effects in limestone reservoir cores containing anhydrite: a discussion of the chemical mechanism. Energy & Fuels 29 (11), 6903–6911.

Anovitz, L. M., Cole, D. R., 2015. Characterization and analysis of porosity and pore structures. Reviews in Mineralogy and Geochemistry, 80(1), 61–164.

Bernstein, J., 2007. Analytical techniques for studying and characterizing polymorphs. Polymorphism in Molecular Crystals, 94–150.

Cai, J., Li, C., Song, K., Zou, S., Yang, Z., Shen, Y., Liu, Q., Liu, Y., 2020. The influence of salinity and mineral components on spontaneous imbibition in tight sandstone. Fuel, 117087.

Lohr, C. D., Hackley, P. C., 2018. Using mercury injection pressure analyses to estimate sealing capacity of the Tuscaloosa marine shale in Mississippi, USA: Implications for carbon dioxide sequestration. International Journal of Greenhouse Gas Control, 78, 375–387.

Liu, D., Gu, Z., Liang, R., Su, J., Ren, D., Chen, B., Yang, C., 2020. Impacts of pore-throat system on fractal characterization of tight sandstones. Geofluids, 2020, 1–17.

Liu, D., Ren, D., Du, K., Qi, Y., Ye, F., 2021. Impacts of mineral composition and pore structure on spontaneous imbibition in tight sandstone. Journal of Petroleum Science and Engineering, 201, 108397.

Georgiadis, A., Berg, S., Maitland, G., Ott, H., 2012. Pore-scale micro-CT imaging: Cluster size distribution during drainage and imbibition. Energy Procedia, 23, 521–526.

Granados, A., Sordo-Ward, A., Paredes-Beltrán, B., Garrote, L., 2021. Exploring the role of reservoir storage in enhancing resilience to climate change in Southern Europe. Water, 13(1), 85.

Guo, J., Li, M., Chen, C., Tao, L., Liu, Z., Zhou, D., 2020. Experimental investigation of spontaneous imbibition in tight sandstone reservoirs. Journal of Petroleum Science and Engineering, 193, 107395.

Rezaei G., S., Amrouche, F., Santos, R. G., Greenwell, H. C., Cubillas, P., 2020. A new framework to quantify the wetting behavior of carbonate rock surfaces based on the relationship between zeta potential and contact angle. Energies, 13(4), 993.

Mandile, A. J., Hutton, A. C., 1995. Quantitative X-ray diffraction analysis of mineral and organic phases in organic-rich rocks. International Journal of Coal Geology, 28(1), 51–69.

Mehana, M., Al Salman, M., Fahes, M., 2018. Impact of salinity and mineralogy on slick water spontaneous imbibition and formation strength in shale. Energy & Fuels 32 (5), 5725–5735.

Mitchell, P., Al Hosani, I., Al Mehairi, Y., Kalam, M. Z., 2008. Importance Of Mercury Injection Capillary Pressure (MICP) Measurements at Pseudo Reservoir Conditions. Abu Dhabi International Petroleum Exhibition and Conference. ISBN: 978-1-55563-204-5

Peng, S., Zhang, T., Loucks, R. G., Shultz, J., 2017. Application of mercury injection capillary pressure to mudrocks: Conformance and compression corrections. Marine and Petroleum Geology, 88, 30–40.

Sheng, G., Zhao, H., Su, Y., Javadpour, F., Wang, C., Zhou, Y., Wang, H., 2020. An analytical model to couple gas storage and transport capacity in organic matter with noncircular pores. Fuel, 268, 117288.

Wang, D., Wang, X., Ge, H., Sun, D., Yu, B., 2020. Insights into the Effect of Spontaneous Fluid Imbibition on the Formation Mechanism of Fracture Networks in Brittle Shale: An Experimental Investigation. ACS Omega.

Wang, Q., Hu, Q., Ning, X., Ilavsky, J., Kuzmenko, I., Tom, T., 2021. Spatial heterogeneity analyses of pore structure and mineral composition of Barnett Shale using X-ray scattering techniques. Marine and Petroleum Geology, 134, 105354.

Yin, X., Jiang, S., Chen, S., Wu, P., Gao, W., Gao, J., Shi, X., 2020. Impact of rock type on the pore structures and physical properties within a tight sandstone reservoir in the Ordos Basin, NW China. Petroleum Science, 17(4), 896–911.

Zhu, P., Dong, Y., Chen, M., Li, Z., Han, B., Wang, J., Cui, Y., 2020. Quantitative evaluation of pore structure from mineralogical and diagenetic information extracted from well logs in tight sandstone. Journal of Natural Gas Science and Engineering. 80.103376.

Chapter IV

Gas Diffusion Experiments and Predictive Models with Machine Learning Methods

Xuewei Ning[a], Khawaja Hasnain Iltaf[a], Xiao Shi[b], Qinhong Hu[a], Yunxiang Bai[c]

a Department of Earth and Environmental Sciences, the University of Texas at Arlington, TX 76019, USA

b Department of Computer Science and Engineering, the University of Texas at Arlington, TX 76019, USA

c Department of Mathematics, 104 E. University Circle

Lafayette, LA 70503, USA

Key words: Gas diffusion, porous media, machine learning, Random Forest regression, Gradient Boosting regression, Linear regression

Abstract

Gas diffusion is a crucial process in the exploration and production of petroleum. In this work, we investigate the use of machine learning methods, specifically linear regression, Gradient Boosting regression, Decision Tree Regression and Random Forest regression, to simulate diffusion velocity and predict gas concentration changes during diffusion process in 15 different types of rocks, including heterogenous and homogeneous structures. Regression models are widely used in scientific analysis and prediction due to their simplicity and interpretability. I applied four regression models to predict gas diffusion velocity in totally different types of rocks and compare their performance to experimental results. Four regression models are used to rank the importance of attributes in analyzing the results of gas diffusion rate. The attributes considered include rock type, sample radius and height, porosity, permeability, gas-tracer concentration difference. The results show that Random Forest model is an effective tool in ranking the most important attributes for gas diffusion. This indicates that the tested gas concentration difference is the main factor that affects gas diffusion in rocks. The most important attributes identified by regression models are the tested gas concentration difference between the air and inner monitored chamber space. The second feature importance is the size (radius and height) of rocks used in the gas diffusion tests. In this research, machine learning methods are used to predict the gas diffusion in different rocks of various sizes. From this database, the model aggregates all pertinent geophysical properties, providing researchers with a tool that obviates the need for redundant gas diffusion tests, thereby conserving valuable research hours. Within the disciplines of engineering and industrial sectors, the study of gas diffusion presents intricate ramifications. The implications extend beyond geological research and prominently manifest in areas such as Environmental Monitoring and Regulation, and the Oil and Gas Industry. Delving

into gas diffusion research, there's an emerging trend towards the development of advanced sensors and monitoring systems adept at detecting the migration of gases in varied environments. These technological breakthroughs are quintessential for industries exposed to hazardous gases, including chemical processing plants and oil refineries. Such sophisticated sensor systems provide real-time insights into gas concentration levels, ensuring a safeguarded operational milieu and reducing potential hazards. In the context of the oil and natural gas sectors, a profound understanding of gas diffusion mechanisms is vital for forecasting the behavioral patterns of gases within geological reservoirs, pipelines, and containment facilities. This in-depth comprehension is central to refining approaches concerning gas extraction, transportation, and storage. In conclusion, our study demonstrates the potential of machine learning methods to effectively predict gas diffusion in different types of rocks, providing valuable insights into the gas diffusion process.

1. Introduction

Gas diffusion plays a crucial role in many areas of science and technologies, including enhanced-oil-recovery process in petroleum production (Mukherjee et al., 2020; Cui et al., 2009; Fu et al., 2020), in polymer electrolyte membrane fuel cells (Gigos et al., 2015), gas purification and chemical processing in environmental science (Shen et al., 2005; Neil et al., 2022), and CO_2 geological sequestration in geosciences (Hatiboglu and Babadagli, 2010). It is important to explore gas diffusion in various types of rocks for a better understanding of the gas migration process (Gillham et al., 1984; Peng et al. 2012). The migration of gas through rocks is a complex process that depends on the physical properties of the rock, including porosity, permeability, and composition (Choquette et al., 1970; Alfredsson et al., 2008; Tiab et al., 2016). Therefore, to understand this process, it is necessary to examine gas diffusion in various types of rocks with

both homogeneous and heterogeneous nature. Studies have shown that rock heterogeneity can have a significant impact on gas diffusion, and collecting data from multiple types of rocks can provide a more comprehensive understanding of the process (Zhao et al., 2009; Liu et al., 2012; Boon et al., 2017; Faramarzi et al., 2020; Ning et al., 2021; Yang et al., 2022).

In recent decades, many analytical, numerical, probability, and mathematical models have been applied in investigating and solving geological, optimization production, and engineering problems (Zhao et al., 2019; Zhao et al., 2020; Ning et al., 2020). In recent years, machine learning models, including supervised and unsupervised models, have been successfully applied to solve various problems, such as reservoir characterization, oil and gas production optimization, environmental pollution modeling, manufacturing technologies in conventional industries and even in exploring disease-altered cardiac electrical activity (Dramsch et al., 2020; Ye et al., 2020; Sircar et al., 2021; Tariq et al., 2021; Kan et al., 2023). Machine learning methods offer distinct advantages over traditional numerical or probability-based approaches (Yuan et al., 2021; Qian et al., 2019). Unlike numerical methods that often rely on intricate mathematical equations and explicit assumptions, machine learning methods harness the power of data-driven insights (Dane et al., 2002; Sarker, 2021.). These methods can automatically detect intricate patterns, relationships, and trends within large and complex datasets, enabling them to uncover hidden insights that might be difficult to capture using conventional approaches. Furthermore, machine learning techniques are adept at handling non-linear and high-dimensional data, which is often encountered in real-world scenarios. They have the ability to adapt and generalize from the data, making them suitable for various applications even when the underlying relationships are complex and not well-defined. Moreover, machine learning models can continuously learn and improve as they're exposed to more data, whereas traditional numerical methods may require

manual adjustments or recalibrations. This adaptability allows machine learning models to evolve alongside changing data patterns.

In this study, the applications of machine learning methods, specifically linear regression, Gradient Boosting regression, Decision Tree Regression and Random Forest regression, were investigated to predict gas diffusion in 15 different types of rocks. By utilizing Lasso regression to rank the significance of rock properties, the most important factors influencing gas diffusion were identified. The proposed machine learning models have the potential for further improvement by expanding the dataset to encompass more rock types and by adding other physical properties, such as porosity, permeability, and composition, which can be measured via techniques such as Mercury injection capillary pressure (MICP) and X-ray computed tomography (CT). The dataset is enriched with empirical observations obtained from the aforementioned methodologies, thereby enhancing the prognostic precision of specific machine learning methodologies encompassing parameters such as the disparity in gas concentration between the external and internal compartments of the monitored enclosure, contact angle measurements, distribution of pore-throat dimensions, porosity levels, permeability characteristics, as well as dimensional attributes of the specimen (height and radius). The model further addresses the issue of concentration fluctuations arising from turbulent gas motion during testing, in tandem with mitigating monitored errors. The incorporation of the z-score technique is introduced to ameliorate the recorded data series, thereby diminishing adverse influences on the fidelity of the model's accuracy.

The integration of machine learning techniques holds immense promise in diverse industrial sectors, including geological processes, fuel cells, and gas sequestration within petroleum production. In geological contexts, machine learning enables the accurate prediction and

modeling of gas diffusion phenomena, aiding in the understanding of subsurface gas migration, reservoir behavior, and geological storage. In the realm of fuel cells, machine learning can optimize gas diffusion pathways, enhancing the efficiency of energy conversion processes. Moreover, in gas sequestration for petroleum production, machine learning offers the potential to predict gas diffusion patterns within reservoirs, facilitating efficient gas injection and extraction strategies. By harnessing machine learning's capacity to analyze complex data sets and uncover intricate patterns, these industries can significantly improve their operations, optimize resource utilization, and contribute to more sustainable and effective practices.

2. Rock samples

In this research, 15 kinds of rock samples were collected from 14 different places in United States, which includes Alabama Marble, Bandera Gray sandstone, Berea sandstone, Boise sandstone, Carthage Limestone, Crab Orchard Sandstone, Edwards Brown Carbonates, Idaho Gray, Indiana Limestone (Perm 2 mD), Indiana Limestone (Perm 200 mD), Kentucky sandstone, Nugget sandstone, Parker sandstone, Scioto sandstone, Silurian dolomite (shown in the Figure 1).

Figure 1: 15 various kinds of rock sample's location map

Cylinder cores were cut out from larger rocks for fitting into the experimental setups of diffusion tests. Various physical characteristics satisfy the investigation in gas diffusion with porosity, permeability, sample size, pore throat distribution, composition, heterogeneous and homogeneous structure. The data frame incorporates distinct attributes such as gas concentration difference, porosity, permeability, and sample size. Pore throat diameter data will be divided into five columns representing the percentage of pores within diameter ranges of 0.01-0.05, 0.05-0.1, 0.1-1, 1-10, and 10 to 44 μm. The structural nature, whether heterogeneous or homogeneous, will be denoted using textual features within the data sets.

3 Hypothesis

1) The gas diffusion coefficient relates to the size of the rock, mineral composition, and pore-throat frequency. A larger cross-section, shorter path length helps gas go through the porous media faster than common sense. The different mineral compositions may cause various levels of gas adsorption and desorption, not studies in our work.
2) The setup of gas diffusion is sealed during monitored time and the results are not affected by the temperature fluctuation. The present rate of gas diffusion velocity is ascertained by evaluating the variance between the specific gas concentration within the identified chamber and the corresponding gas concentration within the source. Importantly, the velocity of gas diffusion remains unaffected by the initial surroundings or initial gas concentrations, indicating that the prediction process remains independent of time.

4. Gas Diffusion Experiment Sets

The foundational configuration for rock sample measurements closely resembled the methodology for soil samples, albeit with the substitution of the stainless-steel sample holder for

a PVC core holder. Prior to experimentation, the diffusion chamber underwent an initial purging with nitrogen, followed by the measurement of oxygen concentration within the chamber to ascertain the rate of atmospheric oxygen intake through the sample. Subsequently, the analysis of outcomes was conducted using the equation outlined by the Rolston and Moldrup's method (Dane et al., 2002; Peng et al., 2012).

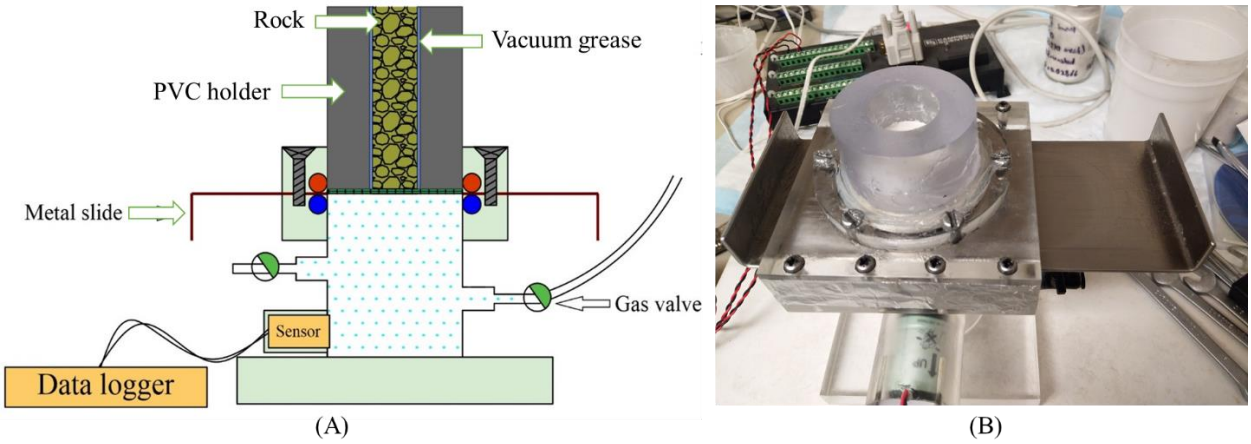


Figure 2: Experiments set up. (a) the Sketch of and gas diffusion apparatus. (b) Gas diffusion set up

Prior to conducting gas diffusion tests, it is imperative to conduct comprehensive assessments of the physical attributes of each rock sample. The geophysical details pertinent to the target samples have been expounded upon in Chapter III. The porosity of the rocks was determined through liquid immersion porosimetry, while rock composition underwent X-ray diffraction (XRD) analysis. Permeability evaluations were carried out using KCl or nitrogen, alongside measurements of porosity and pore-size distribution through mercury intrusion porosimetry (MIP). Subsequently, the gas diffusion experiments were tailored to dimensions compatible with the designated setup, as illustrated in Figure 2. The setup employs a sensor to record the current

concentration of the intended gas molecules within the sealed chamber. Additionally, the gas valve facilitates the injection of nitrogen or oxygen during the testing phase.

Beyond the confines of geology, a range of applications and inquiries span diverse fields. Notably, the migration of gaseous elements within materials like concrete, characterized as porous media, can be impelled by gradients in pressure or concentration, as documented by Sercombe et al. (2007). The intricate landscape of gas diffusion through porous media is commonly categorized into three distinct modes or mechanisms, each associated with its unique diffusion coefficient: free-molecule or Knudsen diffusion, molecular or ordinary diffusion, and surface diffusion, as expounded by Mason et al. (1983). Nonetheless, surface diffusion is frequently omitted from general models due to its relatively minor influence on outcomes.

The experimental arrangement comprises two principal components. The first component encompasses a Plexiglas chamber situated at the lower section, an oxygen sensor strategically positioned to the chamber's left, and a metal slide located at the front. This configuration serves the purpose of facilitating gas diffusion experiments and is depicted here alongside a sand sample encased within a stainless-steel holder. The metal slide is adjustable, allowing for the manipulation of chamber openness or closure to govern experiment conditions. The second component encompasses a PVC core holder assembly designed for use with consolidated samples. This assembly functions independently from the gas diffusion setup, serving to securely hold the sample in place during testing. During the process of gas diffusion through the porous media, a monitoring system integrated within the chamber is interconnected with a computer, capturing oxygen concentration through electronic signals. Upon collecting comprehensive data for each distinct rock type with varying sizes, the machine learning models initiate an analysis of

the salient importance attributed to every influencing factor, encompassing gas concentration, sample size, and permeability.

5. Basic Theories of Gas Diffusion and Assumptions

Fick's law describes the process of gas diffusion, stating that the rate of diffusion is directly proportional to the concentration gradient and the surface area, while inversely proportional to the thickness of the barrier. Equations (1) and (2) show the two forms of Fick's law.

$$\overrightarrow{J_D} = -D\ gradC \tag{1}$$

$$\frac{\partial C(x,t)}{\partial t}\mathrm{dx} = j_D(x,t) - j_D(x+dt,t) = D\frac{\partial C(x,t)}{\partial x} + D\frac{\partial C(x+dx,t)}{\partial x} \tag{2}$$

where $\overrightarrow{J_D}$ is the $diffusion\ current$; D is the proportionality factor between the current and grad C, named the coefficient of diffusion. Expanding $C(x+dx,t)$ into a power series of dx, with dx tending toward zero, which shown in the Equation (3)

$$\frac{\partial C(x,t)}{\partial t} = \mathrm{D}\frac{\partial^2 C(x,t)}{\partial x^2} \tag{3}$$

Thus, in anisotropic medium, the three-dimensional diffusion of Fick's Law can be written as Equation (4):

$$\frac{\partial C}{\partial t} = D_x\frac{\partial^2 C}{\partial x^2} + D_y\frac{\partial^2 C}{\partial y^2} + D_z\frac{\partial^2 C}{\partial z^2} \tag{4}$$

Eq. (4) describes the migration of substances under conditions when filtration and interaction are absent between rocks or existing substances.

Previous literature investigates the theoretical and empirical gas velocity. The velocity distribution of Poiseuille flow is shown as Eq. (5), which is under a no-slip boundary condition

(Mortensen et al., 2005), where v represents the free gas velocity, m/s; R_f represents pore radius, m; μ_g represents gas viscosity, Pa·s; ∇p is the pressure gradient in flow direction, Pa/m.

$$v(r) = \frac{R_f^2}{4\mu_g}(1 - r^2)\nabla p \tag{5}$$

The previous study investigated the relationship between diffusivity and porosity with Archie's law (Peng et al., 2012). Notably, their findings indicated a significant correlation between the cementation factor m value in Archie's law and pore size, with finer pore sizes exhibiting larger m values. Though, for most rocks characterized by a volumetric mean pore diameter (ln d50) below 1.3 μm, a linear regression equation effectively captures the variation of m with ln d50. The results also suggested a possible association with narrower pore size distributions. Eq. (6) shows the diffusivity in Archie's Law, defined as the ratio of $\frac{D_p}{\mathrm{D}_a}$, correlate with porosity. D_p and D_a are the gas diffusion coefficients in porous media and in air, which is shown in Eq. (7). τ , σ, ϕ_a, m are tortuosity, constrictivity factor, air-occupied porosity and cementation factor, respectively.

$$D' = \frac{D_p}{\mathrm{D}_a} = \phi_a^m \tag{6}$$

$$D_p = \tau\sigma\phi\mathrm{D}_a \tag{7}$$

However, the previous numerical methods focus on the explanation of some specific aspects of porous media characteristics and these models are unable to analyze the complexity of features. The importance of impact factors varies according to their noticeable features, and they are hard to quantify in only a few equations (Bagalkot et al., 2017; Liu et al., 2020; Wang et al., 2021). Diffusion, within the realm of physics, refers to the phenomenon where particles within a

substance exhibit a tendency to disperse from regions of high concentration to regions of low concentration. This process is observed in both solids and liquids and is predominantly influenced by the inherent random thermal motion of particles. However, many numerical methods have inevitable drawbacks, such as they only can analyze single or simple influencers. The solution only gives one result instead of the most possible result. The geological problems are intricate because of the distinct variations in their physical characteristics. Thus, in this research, the gas diffusion model adopts and compares four machine learning models to simulate and predict how the concentration changes when multiple conditions change.

Machine learning surpasses numerical or probability models in several ways. Firstly, machine learning can effectively handle complex and non-linear relationships within data, making it more suitable for capturing intricate patterns and interactions. In contrast, numerical or probability models often rely on simplified assumptions and linear relationships, limiting their ability to accurately model real-world complexities. Secondly, machine learning models can adapt and improve over time through continuous learning. They can update their knowledge and adjust their behavior based on new data, feedback, and experiences. This adaptability allows machine learning models to evolve and deliver more accurate predictions or decisions compared to static numerical or probability models. Furthermore, machine learning models excel at handling large-scale and diverse datasets. They can process and analyze vast amounts of data quickly, uncovering valuable insights and making predictions at a scale that would be challenging for traditional numerical or probability models. This scalability makes machine learning well-suited for big data applications. Lastly, machine learning offers the advantage of automation, enabling the automatic extraction of knowledge and insights from data without the need for manual intervention. This automation saves time, reduces human error, and allows for the efficient

analysis of large datasets, providing a significant advantage over numerical or probability models that often require manual calculations and assumptions.

6 Case studies in 15 kinds of rock samples

The gas diffusion test setup for rock samples and measurement procedures was first described in Rolston and Moldrup (2002)'s research work. The experiment setup was introduced in Figure 1. Their research results were concluded in the following equation:

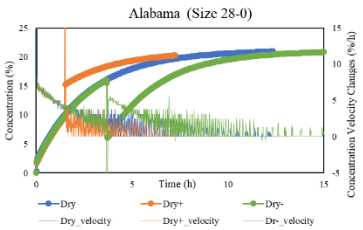
$$C_r = \frac{C_t - C_s}{C_0 - C_s} = \sum_{n=1}^{\infty} \frac{2hexp\left(-\frac{D_p \alpha_n^2 t}{\phi_a}\right)}{L(\alpha_n^2 + h^2) + h} \tag{9}$$

Eq. (9) shows the relationship between the oxygen concentration in diffusion chamber along with time. C_t and C_s represent the oxygen concentration at any time of test (t=time) and the atmospheric oxygen concentration at a normal condition, respectively. α_n is the positive root of $(\alpha_n)\tan(\alpha_n L) = h$, $h = \phi_a/\alpha$, the Eq. (9) will be reduced as Eq. (10) when $n \geq 2$:

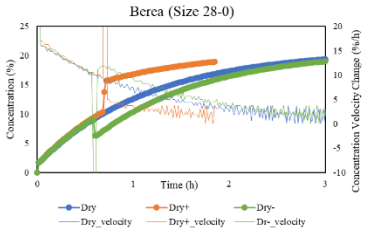
$$lnC_r = -\frac{D_p \alpha_1^2}{\phi} t + ln\frac{2h}{L(\alpha_1^2 + h^2) + h} \tag{10}$$

Though lnC_r is a linear function with time t, this equation only considered the oxygen concentration changes based on the initial concentration in the diffusion chamber. The existing function employed fails to adequately capture the gas velocity when sudden changes in concentration occur. It is widely recognized that different types of rocks possess unique physical properties, highlighting the significance of the gas diffusion coefficient as a fundamental characteristic for characterizing gas transport capabilities within rocks. In numerical methods, the coefficient heavily relies on the selected test duration, which introduces potential errors stemming from variations among individuals conducting the tests. Moreover, the initial and final

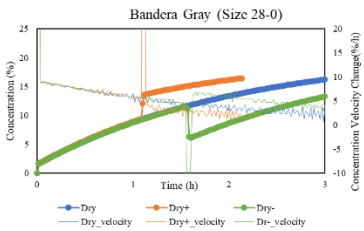
stages of the test tend to be overlooked due to the log plot utilized. Several test records indicate deviations from the expected linear function (how can machine learning address this?), with the monitored gas concentration exhibiting bending patterns instead. Consequently, the exploration of new methodologies becomes imperative to effectively address this issue.



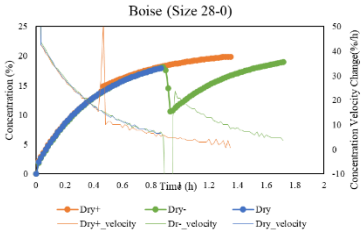
(A)



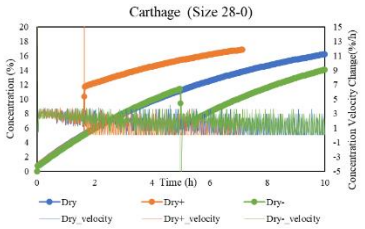
(B)



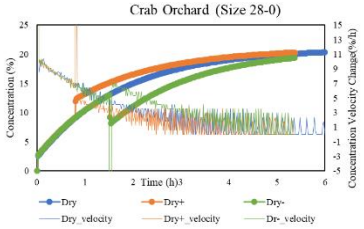
(C)



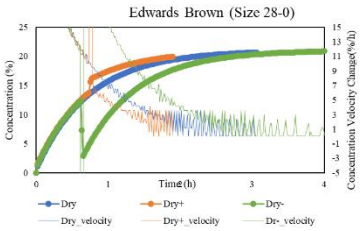
(D)



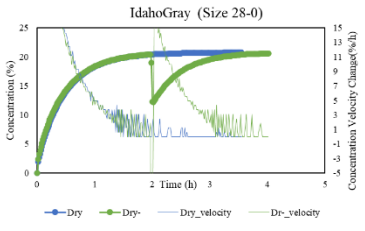
(E)



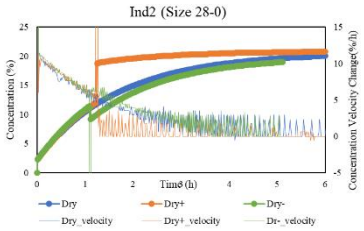
(F)



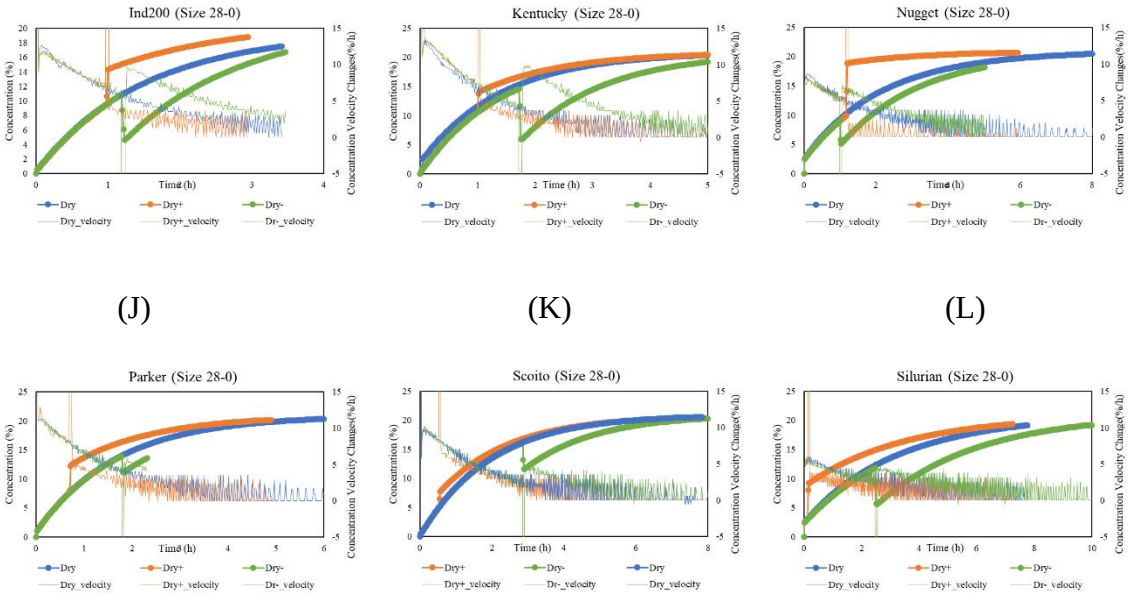
(G)



(H)



(I)



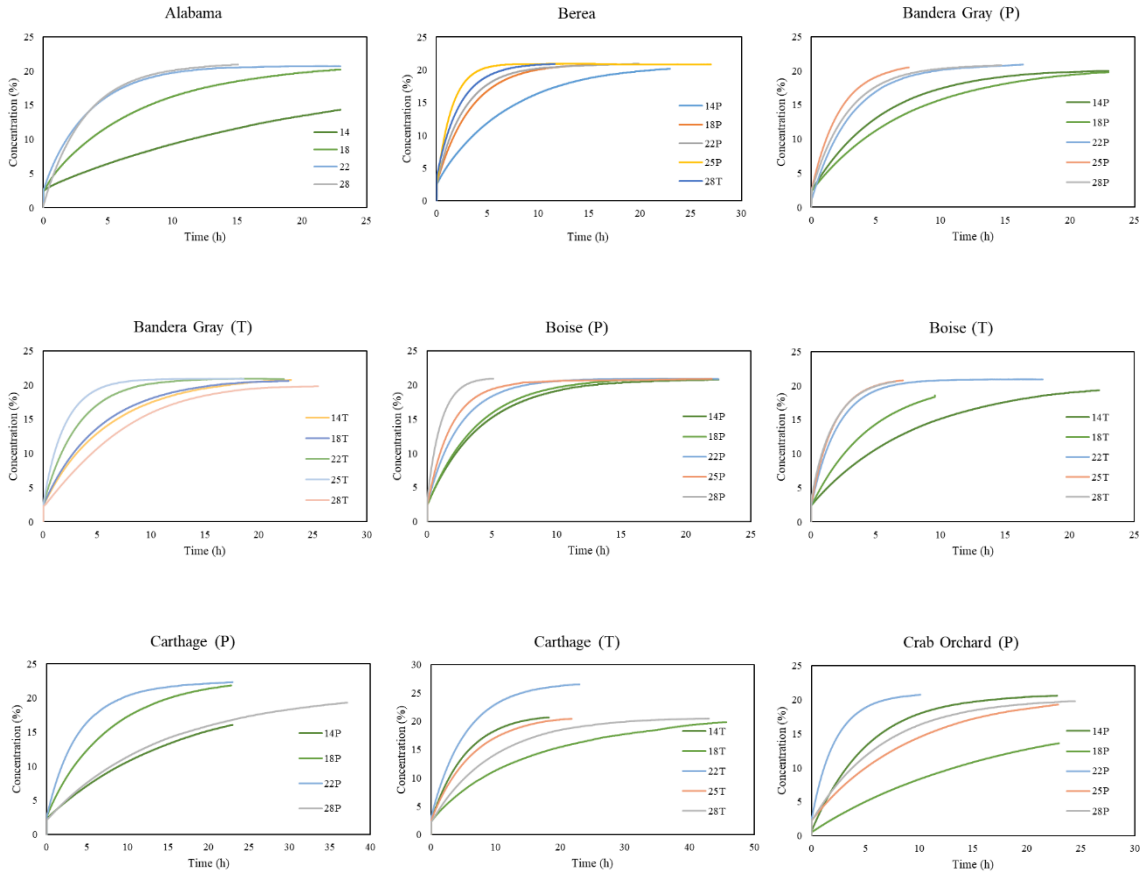
(J) (K) (L)

(M) (N) (o)

Figure 3: Oxygen concentration changes (cut by size-28 mold) and rate changes vs. Time for (A) Alabama Marble; (B) Bandera Gray sandstone; (C) Berea sandstone; (D) Boise sandstone; (E) Carthage Limestone; (F) Crab Orchard Sandstone; (G) Edwards Brown Carbonates; (H) Idaho Gray Sandstone; (I) Indiana Limestone (Perm 2 mD); (J) Indiana Limestone (Perm 200 mD); (K) Kentucky sandstone; (L) Nugget sandstone; (M) Parker sandstone; (N) Scioto sandstone; (O) Silurian dolomite. In sequences stated above (sandstones, carbonates, marbles)

Figure 3 shows the gas diffusion and velocity changes along with time in 15 different kinds of rock. The gas coefficients are different from each other. Figure 3 contains oxygen diffusion rate changes data. Each figure includes three tests (rock sample No. Name-28), the rock sample in consistent dry condition without any interruption, sudden increase and decrease of oxygen concentration in the diffusion chamber with oxygen sensor. Though the oxygen concentration continues to increase in every stage as long as there are no human-made changes, the rate of diffusion becomes slower along with time. And from the figure, we can see that the velocity of

diffusion decreases more drastically when the concentration difference is larger. It seems that the oxygen concentration is changing smoothly based on those figures, which is expected as according to the theory that the gas diffusion is driven by the concentration gradient. But the jagged line of gas diffusion rate shows that the gas molecules pass through the porous media randomly (random is the way for diffusion).



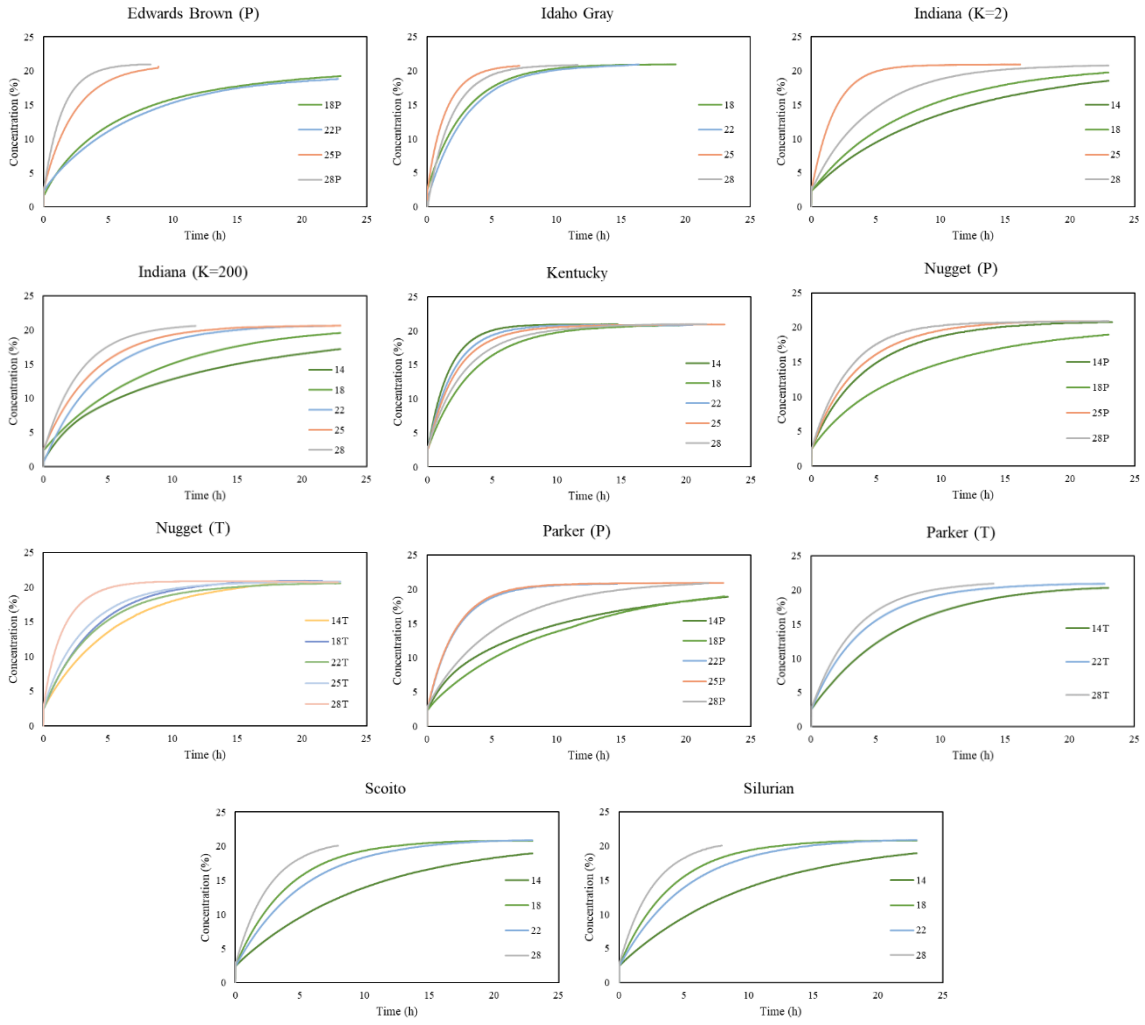
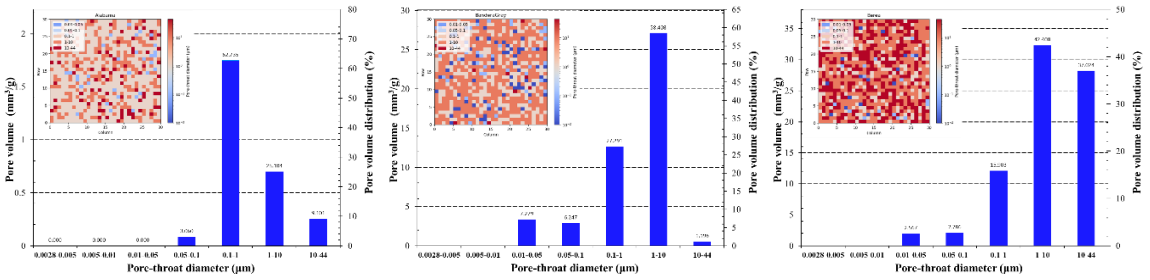


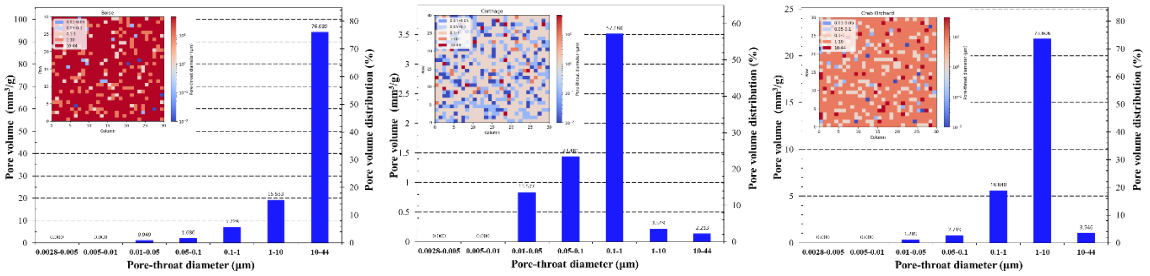
Figure 4: Collection of oxygen concentration data during gas diffusion test in 15 kinds of rock samples

Figure 4 illustrates the alterations in oxygen concentration within the hermetically sealed chamber throughout the course of the gas diffusion examination. Each illustration within this figure represents a unique classification of rock. Subsequent to the collection of this data, it will be converted into numerical form and incorporated into various machine learning models. Although the illustrations may initially appear to depict continuous lines, it must be emphasized that these lines are, in fact, comprised of thousands of individual data points. Each of these points signifies a recorded value of the oxygen concentration at a specific moment in time.

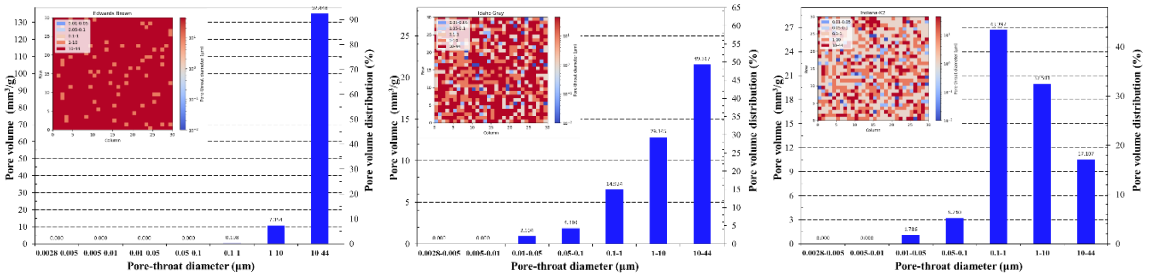
Within the machine learning models, the data undergoes a shuffling process, retaining only those properties that adequately encapsulate the prevailing conditions at the time of measurement. Theoretical considerations suggest that identical states should predict analogous tendencies in the following stages of the process.



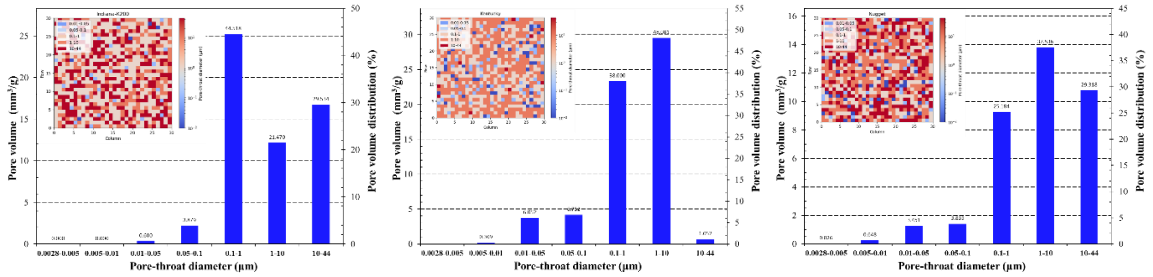
(A) (B) (C)



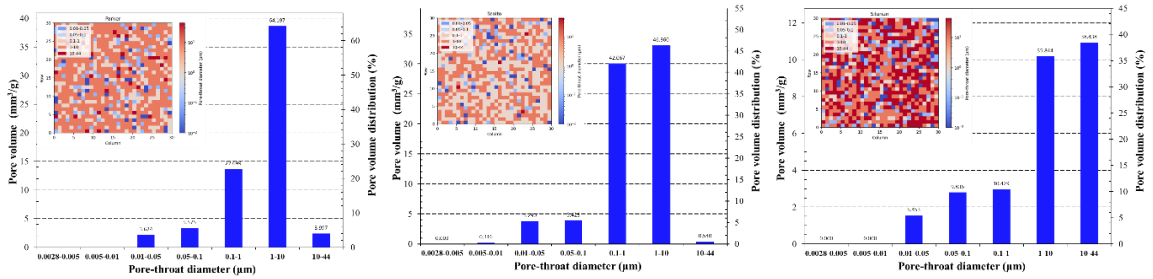
(D) (E) (F)



(G) (H) (I)



(J) (K) (L)



(M) (N) (O)

Figure 5: Collection of pore-throat size distribution in rock samples. (A) Alabama Marble (B) Bandera Gray sandstone (C) Berea sandstone (D) Boise sandstone (E) Carthage Limestone (F) Crab Orchard Sandstone (G) Edwards Brown Carbonates (H) Idaho Gray Sandstone (I) Indiana Limestone (Perm 2 mD) (J) Indiana Limestone (Perm 200 mD) (K) Kentucky sandstone (L) Nugget sandstone (M) Parker sandstone (N) Scioto sandstone (O) Silurian dolomite

Figure 5 displays the pore-throat size distribution in a set of 15 rock samples. The histogram figures, concluded by the mercury intrusion porosimetry, provide a clear representation of the distribution of pore sizes. The majority of pore diameters are concentrated within the range of 0.1-44 micrometers. In each histogram, the left-top colorful picture represents a visualized depiction of the pore diameter distribution. The colors within the figure indicate the prevalence of specific pore diameters within each type of rock. Higher color intensity, particularly in red, suggests the dominance of larger pores in the rock structure. Conversely, a higher proportion of

blue in the figure (E), Carthage Marble, indicates a lower porosity compared to the other rock samples. These histogram figures provide valuable insights into the pore-size distribution patterns across different rock types, aiding in the characterization and understanding of the porous nature of these rocks.

To explore the correlation between pore-size distribution and gas diffusion rate, additional data on pore sizes obtained from the mercury injection capillary pressure (MIP) test are incorporated into the existing datasets. These new variables related to pore throat distribution are included as impacted factors in the regression models. By incorporating the pore-size distribution information, we aim to analyze its influence on gas diffusion and further understand the relationship between these variables within the regression framework.

7. Simulation and prediction models

7.1 Compare rationale models for gas diffusion experiments and model procedures.

Gas diffusion is a crucial process in the exploration and production of petroleum. Understanding gas diffusion in different types of rocks is essential for better comprehension of the gas migration process. In this study, the use of supervised machine learning techniques was explored, specifically linear regression Gradient Boosting regression model, Decision Tree Regression model, and Random Forest regression model, for predicting gas diffusion in 15 types of rocks. Supervised machine learning models are utilized to explore the relationship between gas diffusion and impacted features due to their capability to capture complex patterns and make predictions based on labeled training data, enabling a comprehensive analysis of the influencing factors. By leveraging historical data with known outcomes, these models offer valuable insights into the intricate dynamics between gas diffusion and the relevant features.

Table 1 Pros and Cons for tested regression models

Model Name	Pros	Cons
Linear Regression Model	1. Simplicity 2. Efficiency 3. Well-defined assumptions 4. Stability	1. May not capture complex nonlinear relationships. 2. Underfitting 3. Limited to Continuous Outputs 4. Sensitive to outliers 5. Limited Expressiveness 6. Overfitting
Gradient Boosting Regression	1. High Predictive Power 2. Identify the most influential variables. 3. Combines multiple models, reducing bias and variance and improving generalization. 4. Regularization 5. Handles Outliers 6. Handling missing data	1. Computationally Intensive 2. Hyperparameter Tuning 3. Overfitting Risk 4. Complexity 5. Prone to Bias 6. Limited Parallelization 7. Black Box Nature
Decision Tree Regression	1. Interpretability 2. Non-linearity 3. Identify the most influential variables. 4. Handle Missing Values 5. Robust to Outliers	1. Overfitting 2. Instability 3. Limited Expressiveness 4. High Variance 5. Prone to Biased Sampling
Random Forest regression	1. Not influenced by outliers 2. Handle both linear and non-linear relationships 3. Provide high accuracy and balance the bias-variance trade-off well	1. Not easily interpretable 2. Computationally intensive for large datasets

Table 1 presents an overview of the pros and cons associated with four regression models, highlighting the trade-offs inherent in each approach. Linear Regression is valued for its simplicity and interpretability, but it struggles to handle complex relationships. Gradient Boosting Regression addresses concerns related to overfitting but may encounter challenges in

terms of scalability. Decision Tree Regression excels in capturing non-linear relationships but may be subject to instability and computational expenses. Random Forest Regression achieves high accuracy yet sacrifices interpretability and computational efficiency. When selecting a regression model, researchers and practitioners should carefully consider the characteristics of their datasets and specific requirements. Future research could explore hybrid approaches or model ensembles to leverage the strengths of multiple models while mitigating their respective limitations (Zemel et al., 2000). Overall, the choice of regression model depends on factors such as the complexity of relationships, interpretability needs, computational constraints, and desired predictive performance.

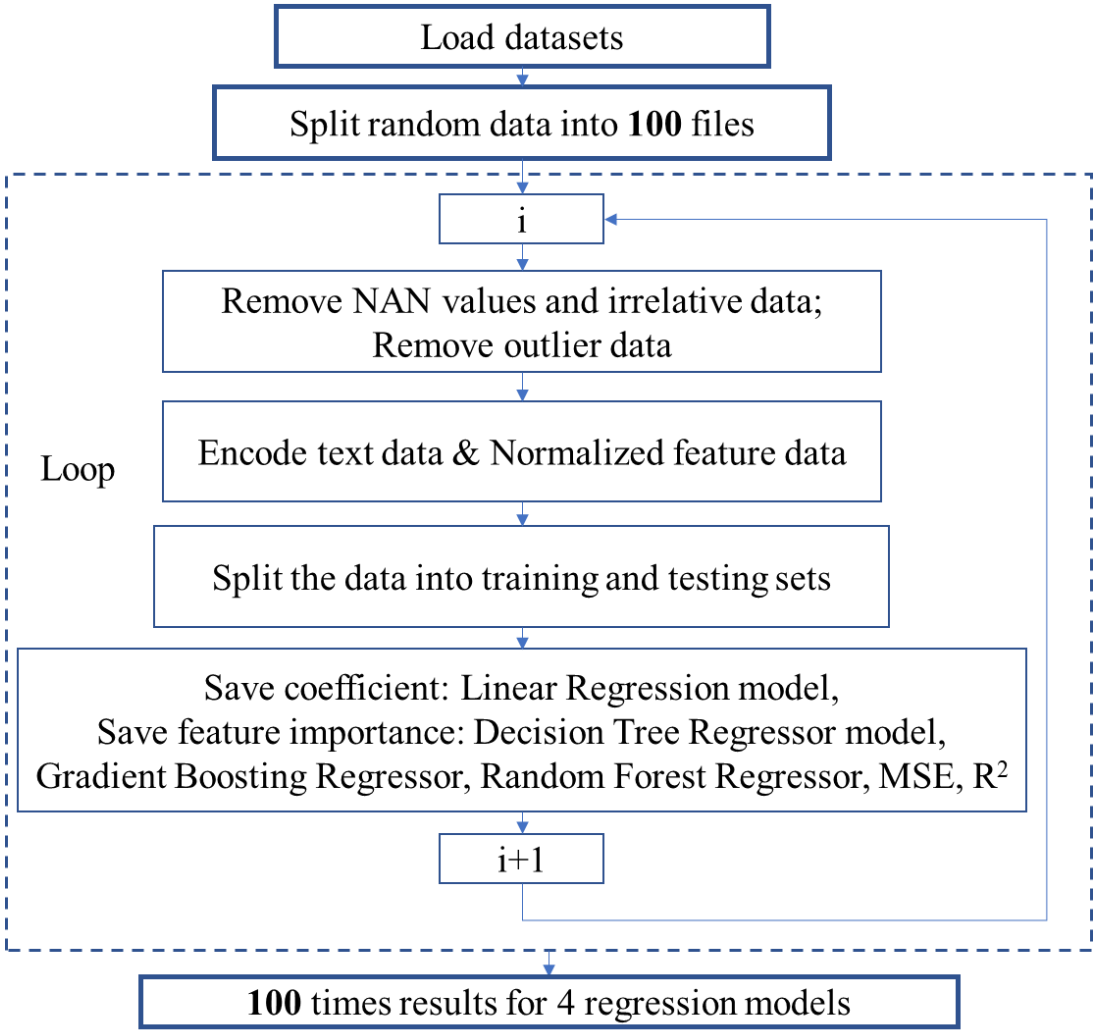


Figure 6: Sequence diagram of calculating feature importance and evaluating the models.

In this chapter, a linear regression model is applied to obtain the coefficients of features and used three regression models to rank the importance of each feature in predicting gas diffusion in rocks. Fig. 6 illustrates the sequence of the machine learning model. The proposed model follows a systematic sequence of steps for data processing and regression modeling. In the initial stage, the datasets, including the data presented in Figure 3, are imported into the machine learning system. To further explore the relationship between gas diffusion and physical properties, additional data on various sizes of rock samples is required to train machine learning models. Supplementary sample data is the gas diffusion concentration record for all types of rocks mentioned in Fig. 3. Subsequently, the complete dataset is randomly partitioned into 100 separate files for further analyses. Considering the unequal quantitative representation of the dataset across different rock types, it becomes necessary to remove the influence of rock type on the analysis. By doing so, we can focus solely on the impact factors that contribute to gas diffusion, such as gas concentration difference, sample size, and other physical properties of tested rocks, without the confounding effect of varying rock types. This approach allows for a more accurate and unbiased assessment of the attributes that significantly influence gas diffusion in rocks. Then, a loop is initiated, wherein each iteration involves the preprocessing of the data. Within each iteration, two crucial data cleaning procedures take place. Firstly, missing values (NAN) and irrelevant data, for example, the weight of the sample, are eliminated from the dataset. Secondly, outlier data are detected and removed using the z-score method. Once the cleaning is completed, the text data is encoded, and the feature data is normalized to ensure consistent scaling. The dataset is subsequently divided into training and testing sets. Following this, three regression models are sequentially applied. Firstly, a Linear Regression model is

employed, and the coefficients are saved for further analyses. Next, a Decision Tree Regressor model is utilized, followed by a Gradient Boosting Regressor model and a Random Forest Regressor model. For each model, the feature importance, mean squared error (MSE), and R-squared (R^2) values are saved to evaluate their performance. The loop continues until all 100 files have been processed, and the entire sequence is concluded.

MSE (mean squared error) and R^2 (R-squared) are commonly used metrics to assess the performance of regression models. They provide different insights into the quality and effectiveness of the models and are often used together to gain a comprehensive understanding of the model's performance. Smaller MSE values indicate better model performance, as they represent a smaller average discrepancy between the predicted and actual values. R^2 ranges from 0 to 1, with a higher value indicating a better fit. R^2 provides an assessment of how well the model captures the variability of the data. Using both MSE and R^2 together can provide a more comprehensive evaluation of the model's performance. While MSE focuses on the magnitude of the errors, R^2 provides insight into the overall goodness of fit. A model with a low MSE but a low R^2 may have low prediction errors but may not explain much of the variance in the data. Conversely, a model with a high R^2 but a high MSE may explain a lot of the variance but still have relatively large prediction errors.

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST}$$

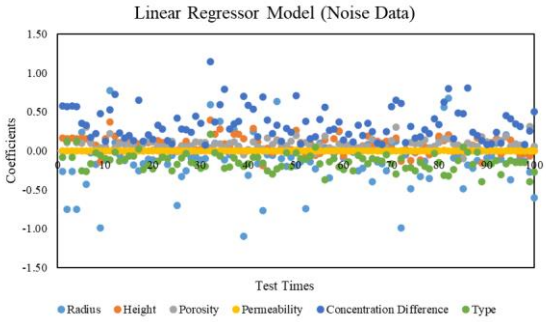
where y_i represents the actual value of the target variable for the i-th observation in the dataset. $\hat{y}_i$ represents the predicted value of the target variable for the i-th observation. $\bar{y}$ represents the

mean value of the target variable across all observations. Parameter n represents the total number of observations in the dataset. SSR is the sum of the squared differences between the predicted values and the mean value of the target variable. SST is the sum of the squared differences between the actual values and the mean value of the target variable.

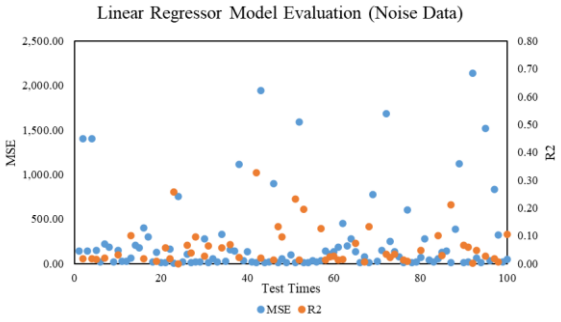
In this study, we have collected three datasets and applied four different regression models to analyze and compare their performance. The datasets were derived from the data presented in Fig. 3 and Fig. 4, which illustrates the variations in oxygen concentration and velocity over time. The first datasets are the unsorted datasets, and the pore-size distribution is not considered for simplicity. The second datasets are the data trim the invalid or bias data. For example, in the gas diffusion experiment procedure, the plate between the chamber and the rock sample needs to be closed immediately once the sensor starts recording the oxygen concentration in the diffusion chamber. The oxygen concentration in the chamber will increase drastically, the rate of the gas diffusion will increase too. However, the extremely high rate cannot represent the normal gas diffusion coefficient and is unable to reflect the rock properties. Thus, trimming extremely bias data is necessary. The third datasets are escalated from the second data set, which not only removes the noise data, but also adds the pore-size distribution obtained from the MIP tests.

By using Equation (10) and referring to the methods used in previous papers (Hu et al., 2003; Peng et al., 2012), previous results heavily relied on manual analysis and suffered from significant biases when determining the appropriate coefficients. Furthermore, the coefficient may not fit the model well at the beginning and the last end of the test. To address this limitation, we aim to enhance the reliability of the predictions by leveraging machine learning regression models.

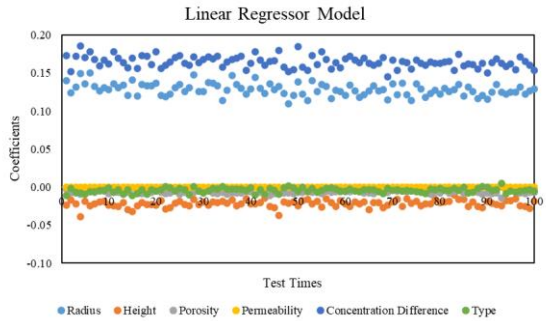
7.2 Linear Regressor Model Test Results



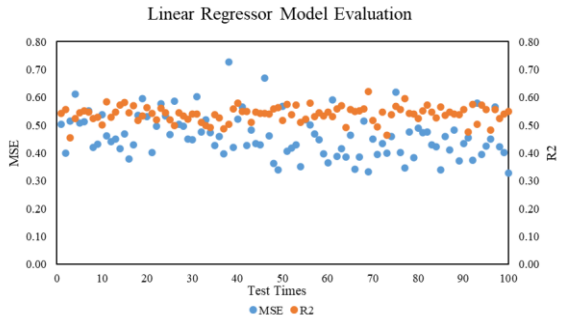
(A)



(B)

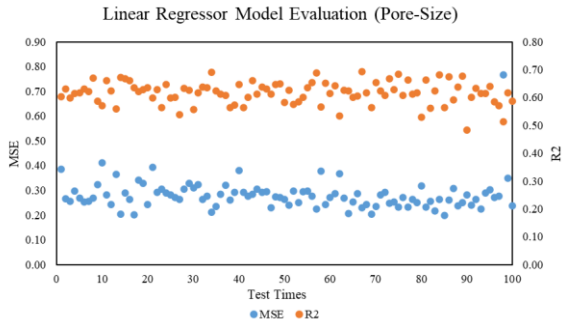


(C)



(D)

(E)



(F)

Figure 7: The 100-file collections of linear regressor model coefficient and evaluation with noise data and without noise data. (A) collection of coefficients in linear regressor model with noise data. (B) model evaluation based on noise data. (C) collection of coefficients in linear regressor model without noise data (D) model evaluation based on no-noise data. (E) Evaluation collection of coefficients in linear regressor model with pore-size distribution impact factors. (F) model evaluation based on pore-size distribution impact factors.

Fig.7 shows the results of the linear regression model. This figure contains the comparisons of the three datasets. (A) (C) and (E) are the coefficients collection of the impacted factors in linear regression model. (B) (D) and (F) are the evaluations of linear regression based on the MSE and R^2. By comparing figures (A), (C), and (E), it is evident that the coefficients exhibit less variance as the data becomes more specific. Notably, there are significant differences in the values of radius and height among these figures. Radius appears to have a positive effect on gas velocity, indicating that as the radius increases, the gas velocity tends to increase. Conversely, sample height has a negative impact on gas diffusion rate, suggesting that as the height increases, the gas diffusion rate tends to decrease. Among these three models, the concentration difference between the rock samples' two sides emerges as the most influential factor, despite the variations in its value range. This implies that the concentration difference has a substantial impact on diffusion rate. The second most influential factor is sample radius, as gas molecules pass through the sample's cross-sectional area, which is proportional to the square of the radius. Consequently, the impact of the radius on diffusion rate is equivalent to the impact of the square of the radius itself.

Fig. 7 (B) displays an irrational value for the MSE, rendering it unnecessary to analyze further due to its anomalous nature. However, a comparison between Fig. 7 (D) and (E) reveals clear patterns. It is evident that the MSE decreases from the range of 0.4-0.6 in Fig. 7 (D) to 0.2-0.3 in

Fig. 7 (E). Simultaneously, the R^2 value increases from the range of 0.4-0.5 in Fig. 7 (D) to 0.5-0.6 in Fig. 7 (E). These trends indicate improved model performance in terms of lower prediction errors and increased explained variance when transitioning from Fig. 7 (D) to Fig. 7 (E).

When the datasets are untrimmed, the regression model exhibits coefficients that span a wider range of values, leading to more random and unpredictable results. This phenomenon can be attributed to the presence of outliers or influential observations within the dataset, which introduce substantial variability and disrupt the model's ability to capture consistent patterns. Consequently, the model's predictions become less reliable and less interpretable. The inclusion of these untrimmed datasets may lead to biased parameter estimates and reduced overall model performance.

7.3 Decision Tree Regression Model

Decision Tree Regression Model. Decision Tree regression is a powerful tool in geological analyses that helps us predict the velocity of rocks based on key factors. In our dataset, these factors include the type of rock, sample radius and height, as well as properties like porosity, permeability, and concentration difference. By understanding how these factors influence velocity flux, we can make accurate predictions. Imagine the decision tree as a flowchart that guides us through a series of questions. At each step, I chose the most informative factor to split the data. For example, I started by asking if the rock type is sedimentary or igneous. Depending on the answer, I followed the corresponding branch and continued with other questions, such as whether the radius is above or below a certain threshold. The goal is to create groups of rocks that have similar velocities within each subset.

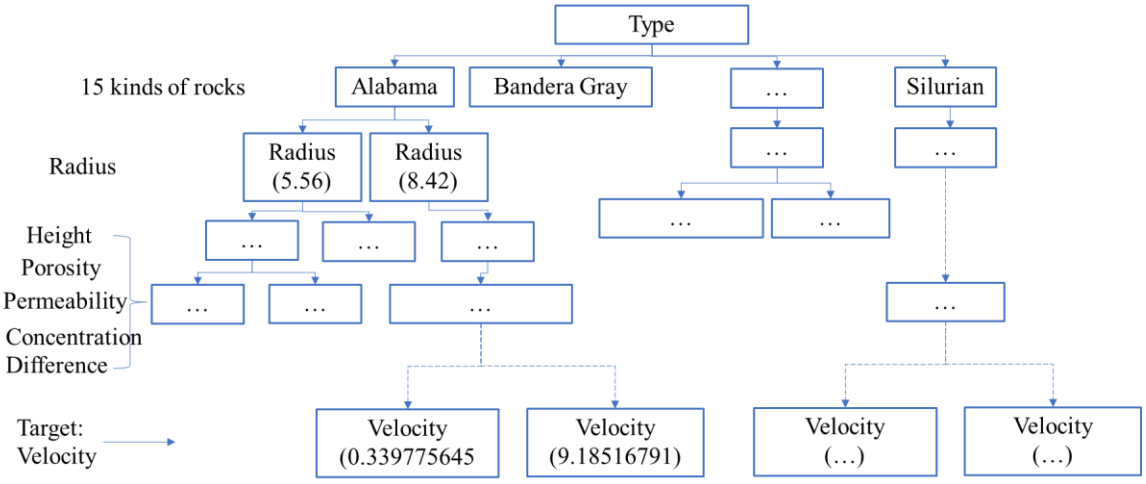
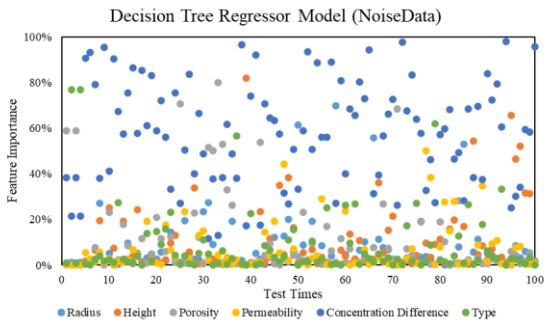
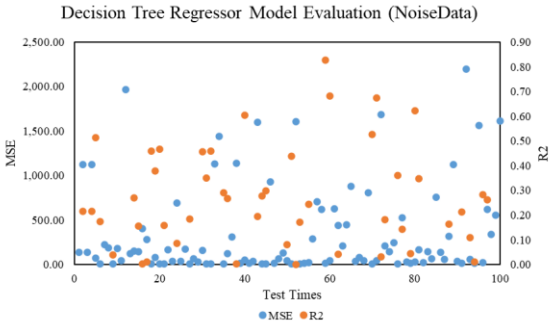


Figure 8: The flowchart of Decision Tree Regression Model

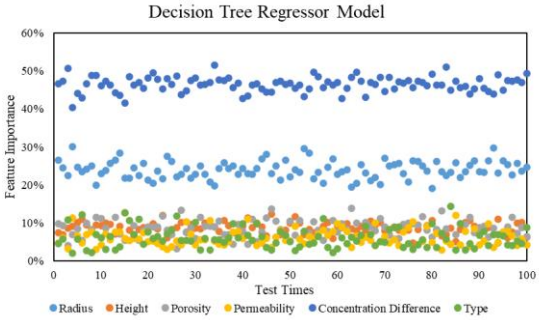
Figure 8: illustrates the flowchart depicting the sequential steps involved in implementing the I look for patterns in the dataset that allow us to make reliable predictions. By recursively splitting the data based on different factors of rock geophysical characterizations, we can create a tree-like structure where each branch represents a different combination of factors. To predict the velocity of a new rock, we simply follow the decision path down the tree until we reach a leaf node. At that point, we assign the mean or median gas diffusion rate of the rocks within that group as our prediction. This way, we can estimate the velocity based on the specific characteristics of the rock. Decision tree regression has advantages in geology because it provides interpretable results. We can easily visualize and understand the decision-making process, helping us gain insights into the relationships between the factors and gas diffusion rate. Moreover, decision trees can handle different types of geological data, including numerical measurements and categorical rock types. However, it's important to be cautious of overfitting, where the decision tree becomes too complex and fits the training data too closely. This can lead to poor predictions on new data. To address this, we can use techniques like pruning to simplify the tree or ensemble methods like Random Forests, which combine multiple decision trees to improve accuracy. Decision Tree

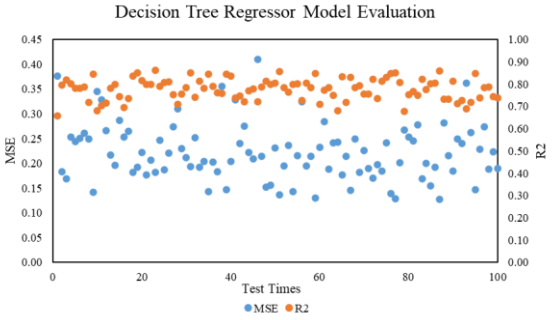
regression is a valuable tool in geology that allows us to predict rock gas diffusion rate based on a set of influential factors. By constructing a hierarchical decision tree, we can uncover the relationships between rock properties (sample radius is not a rock property, rather an experimental one) and gas diffusion rate, providing valuable insights for geological analysis and exploration.



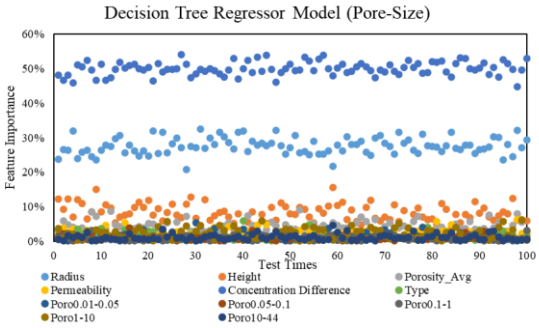


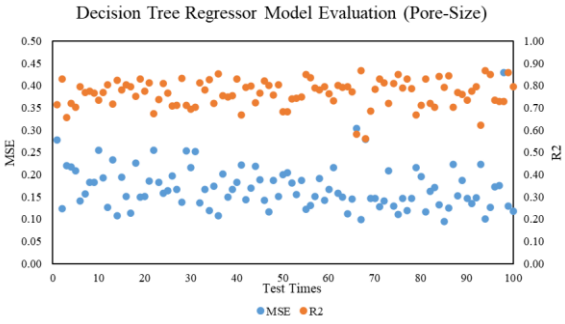
(A) (B)





(C) (D)





(E) (F)

Figure 9: The 100-file collections of decision tree regressor model (A) collection of coefficients of decision tree regression with noise data (B) collection of coefficients of decision tree regression without noise data (C) Evaluation collection of coefficients of decision tree regression with noise data (D) Evaluation collection of coefficients of decision tree regression without noise data

Fig. 9 showcases the outcomes of the Decision Tree regressor model, offering a comparative analysis across three distinct datasets. The feature importance values for the impacted factors within the decision tree regressor model are presented in Fig. 9 (A), (C), and (E). Complementarily, Fig. 9 (B), (D), and (F) provide a comprehensive evaluation of the model's performance based on MSE and R^2. In line with the outcomes obtained from the linear regression model, Fig.9 (A) and (B) in the present analysis exhibit irrational MSE values that hinder further analysis due to their anomalous nature. Nonetheless, a comparative assessment of Fig. 9(D) and (E) reveals notable patterns in the evaluation results. Specifically, the MSE values demonstrate a decrease from the range of 0.15-0.35 in Fig.9 (D) to 0.1-0.25 in Fig. 9 (F). Simultaneously, the R^2 values display an increase from the range of 0.65-0.85 in Fig. 9 (D) to 0.7-0.9 in Fig. 9 (F). It is worth noting that throughout the evaluation, the concentration difference and the size of the rock sample consistently exhibit substantial importance, while the significance of the remaining impact factors appears relatively low in Fig. 9 (E). These findings emphasize the effectiveness of the decision tree regression method in accurately predicting gas diffusion concentration. However, even after removing all noise data, the evaluation results still exhibited unreliable outcomes in the form of inconsistent predictions. Specifically, in Figure 9d, the unreliable results were observed in several instances, which were subsequently reduced to two instances in Fig. 9

(F) during the 100-times tests. Despite efforts to eliminate noise and enhance the robustness of the analysis, the presence of unreliable outcomes suggests underlying complexities or unaccounted factors that influence the predictive performance of the model. These factors could include outliers, rare events, or data instances with unique characteristics that the model struggles to generalize from. Therefore, the presence of unreliable outcomes underscores the importance of thorough feature selection, data preprocessing, and continuous model refinement to account for these latent complexities and unaccounted factors.

7.4 Gradient boosting regression model

The Gradient Boosting regression model stands as a potent and adaptable machine learning approach, demonstrating remarkable proficiency in predictive endeavors, particularly when confronted with intricate relationships and data characterized by noise. The model functions through iterative refinement of predictions, achieved by amalgamating numerous weak learners, often in the form of decision trees. Subsequent learners within this progression are dedicated to rectifying errors committed by their predecessors. This iterative mechanism facilitates a gradual

enhancement of predictive accuracy while accommodating the detection of intricate data patterns.

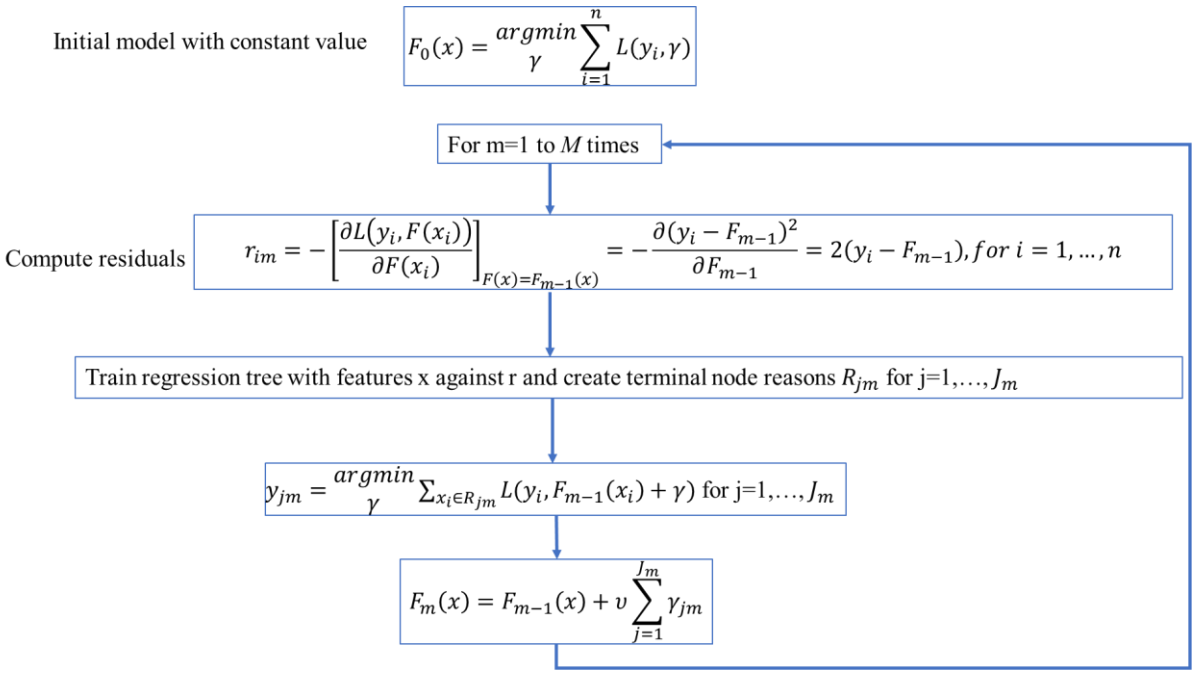


Figure 10: The flowchart of gradient boosting regression model

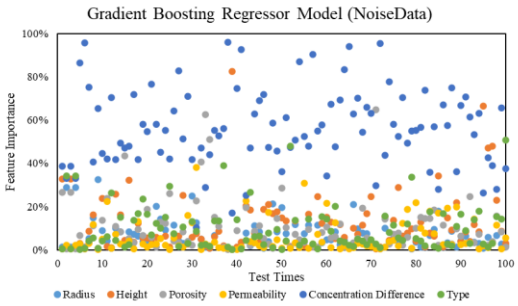
Fig. 10 provides a detailed flowchart outlining the implementation of gradient boosting regression. This variant of the gradient boosting algorithm is specifically tailored for regression tasks and employs a squared loss function. The algorithm begins by initializing the prediction, F_0, with a constant value, and the loss function, denoted as L, is defined as the squared loss. The goal is to find the optimal value of γ (gamma) that minimizes the sum of the loss function, denoted as ΣL. This is achieved by calculating the derivative of ΣL with respect to γ and setting it equal to zero, resulting in the equation $\partial\Sigma L/\partial\gamma = 0$. An intriguing finding is that the value of γ that minimizes ΣL corresponds to the mean value of the target variable, y. Consequently, F_0 is set as the mean of y in order to leverage this insight. The subsequent steps are iterated M times, where M represents the number of trees to be constructed and m denotes the index of each tree. Residuals, represented as $r_i m$, are computed by calculating the derivative of the loss function with respect to the previous prediction, Fm_{-1}, and multiplying it by -1. These residuals serve as

negative gradients, providing guidance on both the direction and magnitude required to minimize the loss function. The term "residuals" is used due to their resemblance to residuals encountered in traditional regression analysis. Equations are solved for each individual sample, denoted by i, to compute the residuals. These equations incorporate the previous prediction, F_{m-1}, and the objective is to determine the values of $r_i m$. The algorithm then searches for the optimal value of $\gamma_j m$ that minimizes the loss function for each terminal node (leaf) within the tree. This involves aggregating the loss over all samples, x_i, that belong to the corresponding terminal node, $R_j m$. The optimal value of $\gamma_j m$ is determined as the average of the residuals, $r_i m$, within the terminal node, $R_j m$. In essence, $\gamma_j m$ represents the prediction value for each terminal node, which corresponds to the average of the target values (in this case, residuals) within that node. The prediction of the combined model, Fm, is then updated. When a given sample, x, falls into a specific terminal node, $R_j m$, the corresponding value of $\gamma_j m$ is selected and added to the previous prediction, F_{m-1}, resulting in the updated prediction, Fm. The learning rate, represented by ν, controls the contribution of the additional tree prediction, γ, to the combined prediction, Fm. It ranges between 0 and 1, with a smaller learning rate reducing the impact of the additional tree prediction and potentially mitigating overfitting to the training data.

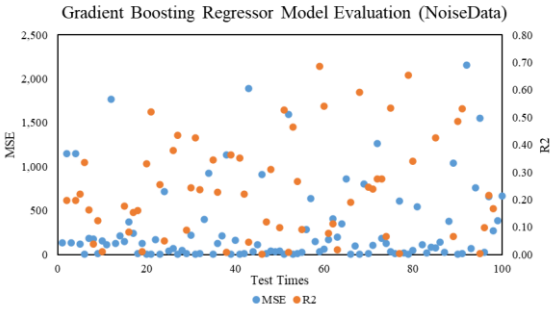
$$F_{i+1} = \begin{cases} F_i + \upsilon \cdot \gamma_2 & if\ x < given\ number \\ F_i - \upsilon \cdot \gamma_2 & otherwise \end{cases}$$

$$r_{i+1} = y - F_i$$

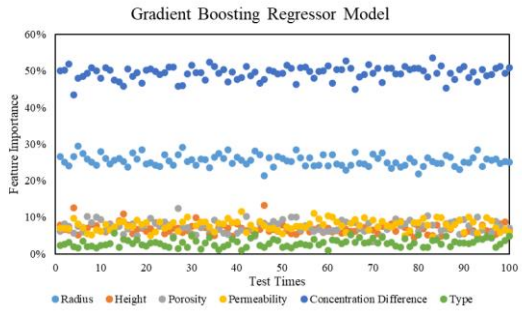
$$F_0(x) = \underset{\gamma}{argmin} \sum_{i=1}^{n} L(y_i, \gamma)$$


Gradient Boosting Regressor Model (NoiseData)
Feature Importance
Test Times
Radius
Height
Porosity
Permeability
Concentration Difference
Type

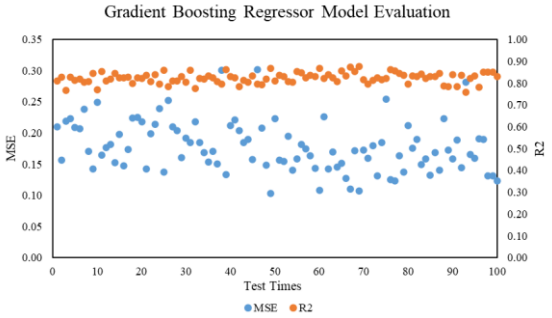
(A)


Gradient Boosting Regressor Model Evaluation (NoiseData)
MSE
R2
Test Times
MSE
R2

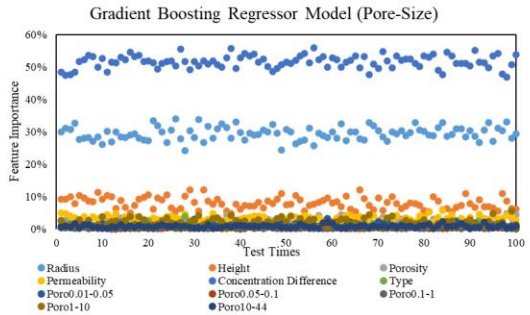
(B)


Gradient Boosting Regressor Model
Feature Importance
Test Times
Radius
Height
Porosity
Permeability
Concentration Difference
Type

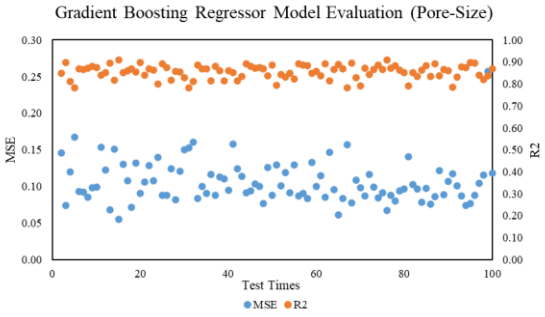
(C)


Gradient Boosting Regressor Model Evaluation
MSE
R2
Test Times
MSE
R2

(D)


Gradient Boosting Regressor Model (Pore-Size)
Feature Importance
Test Times
Radius
Height
Porosity
Permeability
Concentration Difference
Type
Poro0.01-0.05
Poro0.05-0.1
Poro0.1-1
Poro1-10
Poro10-44

(E)


Gradient Boosting Regressor Model Evaluation (Pore-Size)
MSE
R2
Test Times
MSE
R2

(F)

Figure 11: The 100-file collections of gradient boosting regressor model coefficient and evaluation with noise data and without noise data. (A) collection of coefficients in linear regressor model with noise data. (B) model evaluation based on noise data. (C) collection of coefficients in linear regressor model without noise data (D) model evaluation based on no-noise data. (E) Evaluation collection of coefficients in linear regressor model with pore-size distribution impact factors. (F) model evaluation based on pore-size distribution impact factors.

Fig. 11 presents a comprehensive analysis of the decision tree regressor model, providing insights into its performance across three distinct datasets. Fig. 10 (A), (C), and (E) depict the feature importance values for the impacted factors within the model, offering a quantification of their relative significance. Concurrently, Fig. 10(B), (D), and (F) offer an in-depth evaluation of the model's performance through the examination of key metrics such as MSE and R^2. Notably, when comparing the decision tree regressor model with the linear regressor model, a distinctive feature is observed in the more concentrated range of impact factor values. For instance, the impact factor of "concentration difference between inside and outside of the chamber" exhibits values primarily within the range of 45% to 55% in the decision tree regressor model (Fig. 8 (E)), whereas the same datasets manifest a highly concentrated range of 50% to 55% in the decision tree regressor model (Fig. 10 (E)). Another better improvement is the MSE values in both datasets, no-noise dataset (Figure 10 d, and pore-size distribution datasets (Fig. 10 (F)). The MSE values in these 100-time tests, fewer results are obviously out of the lines compared to linear and decision tree regressor models. Consequently, the interpretability of the gradient boosting regressor model is slightly higher than that of the previous two models. Furthermore, the R^2 values demonstrate an improvement from the range of [0.6,0.9] to [0.8,0.9], indicating

enhanced stability and higher reliability of the results across all datasets within the gradient boosting regressor model.

7.5 Random Forest model

A random forest model is a machine learning algorithm that is commonly used for both classification and regression tasks. It belongs to the family of ensemble learning methods, which combine multiple individual models to make predictions. The sketch theory about random forest is shown in Fig. 12.

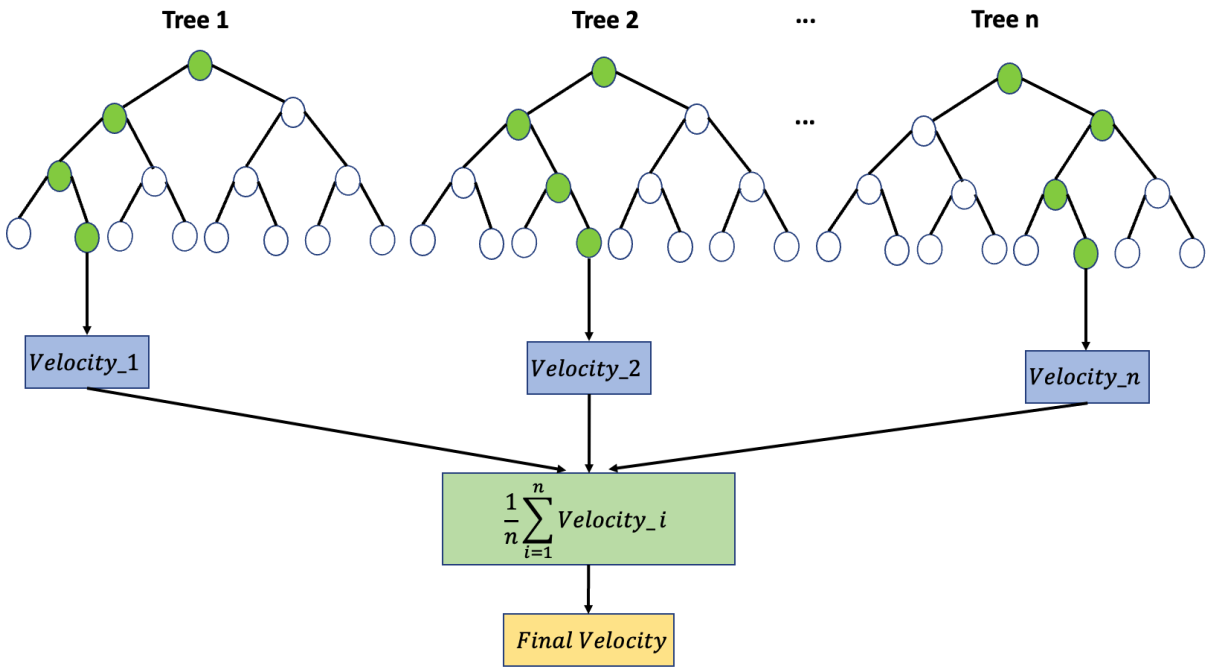
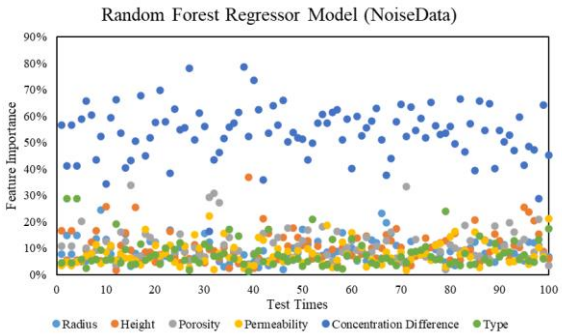


Figure 12: The sketch of random forest regression model

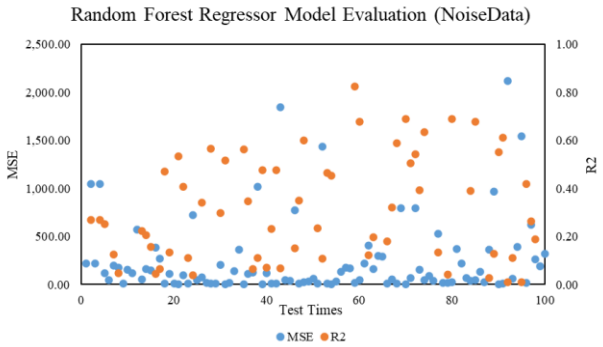
The random forest algorithm constructs a collection of decision trees during the training phase. Each decision tree is built on a randomly selected subset of the training data and a subset of features. This randomness in selecting subsets helps to reduce overfitting and improves the generalization ability of the model.

When making predictions with a Random Forest regression, each individual decision tree in the forest independently predicts the target variable, and the final prediction is determined by aggregating the predictions of all the trees. For regression tasks, the predictions of all the trees are averaged to obtain the final prediction.

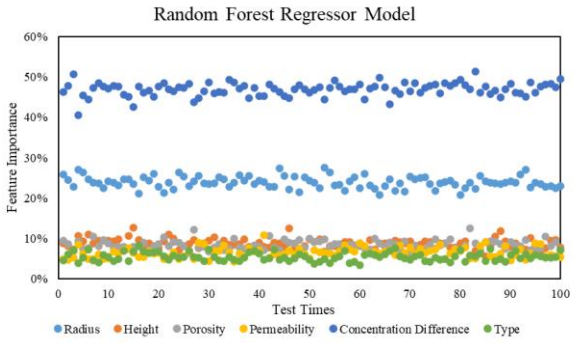
Random forests have several advantages. They are robust against overfitting, handle high-dimensional data well, and can capture complex relationships between features and the target variable. They are also capable of handling missing values and outliers without preprocessing the data. Additionally, random forests can provide measures of feature importance, which can be helpful for feature selection or understanding the underlying data. Overall, random forests are popular and widely used in various domains due to their versatility, accuracy, and interpretability.

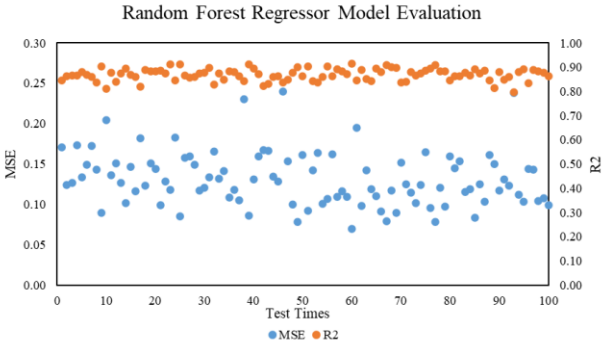


(A)

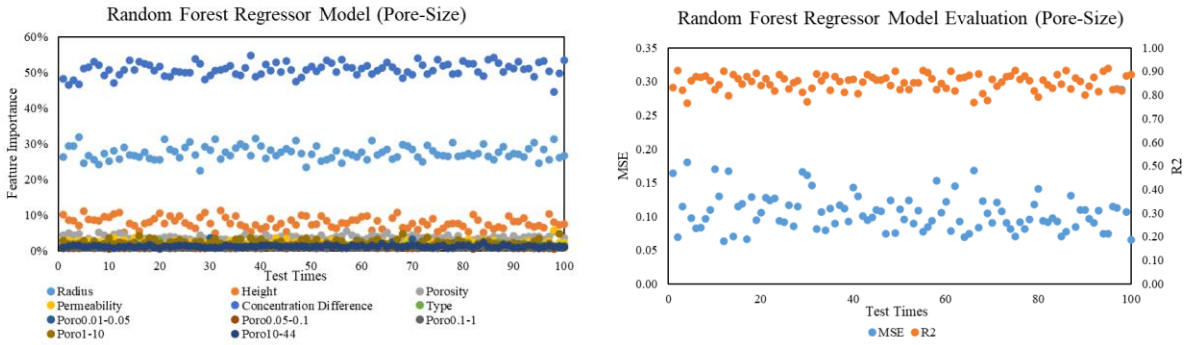


(B)





(C) (D)



(E) (F)

Figure 13: The 100-file collections of random forest regressor model coefficient and evaluation with noise data and without noise data. (A) collection of coefficients in linear regressor model with noise data. (B) model evaluation based on noise data. (C) collection of coefficients in linear regressor model without noise data (D) model evaluation based on no-noise data. (E) Evaluation collection of coefficients in linear regressor model with pore-size distribution impact factors. (F) model evaluation based on pore-size distribution impact factors.

Fig. 13 presents a comparative analysis of the random forest regressor model in relation to other regression models, namely linear, decision tree, and gradient boosting regressor models. The random forest model demonstrates narrower ranges of feature importance values for most impact factors, indicating improved predictability and precision in the simulation results. Notably, attributes such as radius and height consistently exhibit similar levels of importance across all machine learning models. Additionally, the concentration difference between sample two sides emerges as a highly influential factor, with a consistent importance attribution of 50% across the models. Although some variability is observed in the importance values of other impact factors, such as porosity, permeability, and type, their values remain within reasonable ranges.

7.6 Comparison of Four models

Fig. 14 provides a comprehensive evaluation of three machine learning regression models, offering an overview of their performance in the regression task. Figure (A) and (D) are the evaluation results of initial raw data in three models. Both (A) and (D) show bias and unreliable results. MSE is too high and R^2 is too low for a good model. The box in the plot represents the interquartile range (IQR). The box top line and the bottom line mean the first quartile (Q1) and the third quartile (Q3). The length of the box indicates the spread of the central 50% of the data. The median is typically represented by a line or dot inside the box. Though the medians in three models are almost the same, there are too many significantly outlier data. Figure (B) and (E) are the evaluation results of the trimmed data in three models. Compared to (A) and (D), we can observe that outlier data becomes less. In random forest regressor model, MSE is the least and R2 is the highest value which means that this model is the most suitable one for gas diffusion velocity prediction. Figure (C) and (F) are the evaluation results of the data with pore-size distribution in three models. After adding the pore-size distribution, the models are more predictable and more accurate. Compared to the second dataset, the MSE median values of decision tree, gradient boosting, random forest models dropped 25.32%, 42.27%, 18.39%, respectively. In the same method, compared the evaluation results with two datasets, the R^2 performance better in gradient boosting model with an increase of 4.58% and a bit worse in decision tree and random forest, which decrease 1.40% and 1.47%, respectively.

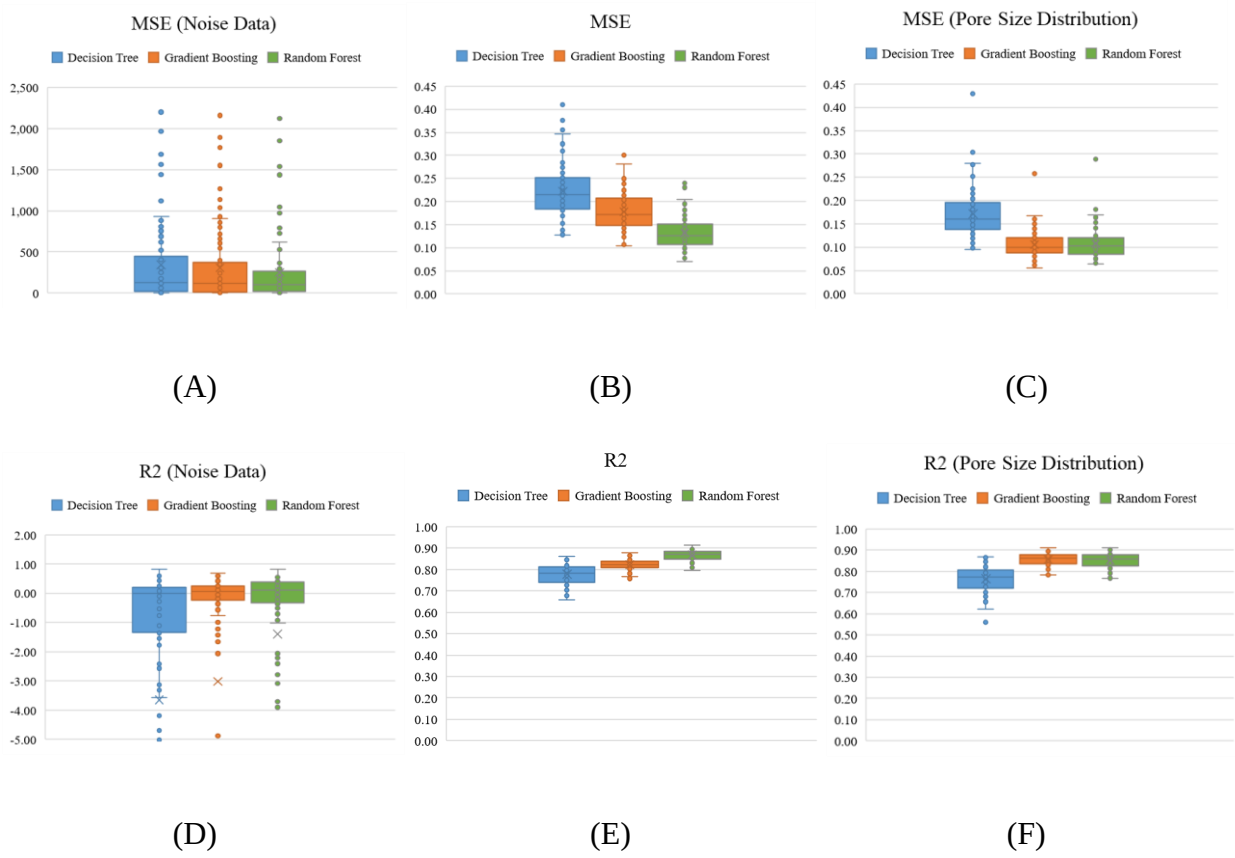
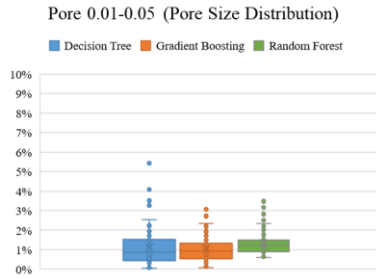


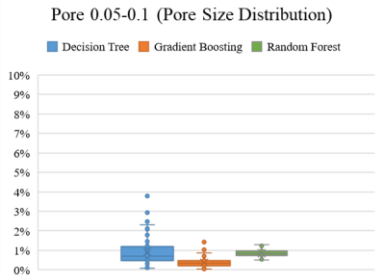
Figure 14: Collection of three machine learning regression models evaluation (A) MSE of machine learning regressor models with noise data. (B) MSE of machine learning regressor models without noise data. (C) MSE of machine learning regressor models with pore-size distribution and without noise data. (D) R2 of machine learning regressor models with noise data. (E) R2 of machine learning regressor models without noise data. (F) R2 of machine learning regressor models with pore-size distribution and without noise data.

Fig. 15 illustrates the feature importance of pore size within different ranges. When considering Fig. 4, it is observed that the majority of rock samples have pore diameters concentrated between 0.1 and 44 micrometers. However, certain rock samples exhibit low porosity and permeability, indicating that the pore diameters within the range of 10 to 44 micrometers may constitute a smaller proportion of the total pores. Conversely, rock samples with higher porosity may lack

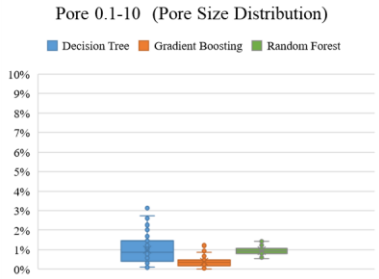
pores in the 0.1-micrometer range. Consequently, the pore diameter range of 1 to 10 micrometers exerts the most significant influence on the models. Although the median pore diameter values within this range are similar across the three regressor models, random forest regression displays the narrowest range. The analysis of over 50,000 datasets, divided into 100 files, reveals that pore diameter plays an important role, although its influence is relatively lower compared to factors such as concentration difference or rock sample size. The pore diameter's importance percentage within these five ranges is estimated to be approximately 1%, 0.8%, 0.9%, 2%, and 1%, respectively. Overall, random forest regression exhibits the narrowest prediction range, suggesting that this model offers more precise and reliable interpretations.



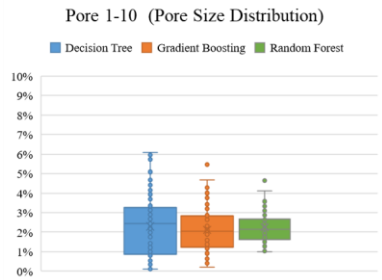
(A)



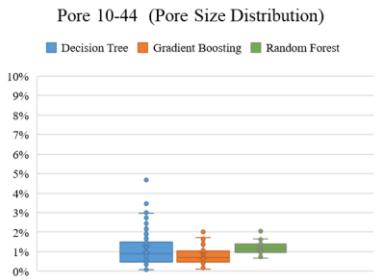
(B)



(C)



(D)



(E)

Figure 15: Collection of impact factors in three machine learning regression models evaluation (A) Pore diameter in range 0.01-0.5. (B) Pore diameter in range 0.05-0.1. (C) Pore diameter in range 0.1-10. (D) Pore diameter in range 10-44.

8. Conclusions

This study emphasizes the proficiency of the Random Forest regression technique in predicting gas diffusion patterns and selecting salient attributes. The research delineates that the disparity in concentration between gas samples, the dimensions of the rock, permeabilityof the rock type are substantial factors governing the diffusion of gas through rock structures. A notable contribution of this research lies in its demonstration that complex geological phenomena can be both quantitatively analyzed and empirically validated through the application of machine learning models. Moreover, the research observes that at the initiation of the gas diffusion process, the presence of noise data may exert a deleterious effect on the performance of the regression model. As the concentration difference between the two facets of the sample converges towards a minimal disparity, the gas diffusion mechanism may descend into a state of stochastic turbulence, which could subsequently induce unpredictable fluctuations in the directional movement of the diffusing gas. Consequently, the excision of noise data originating from the test set itself emerges as an essential preprocessing measure that must be undertaken prior to the deployment of the machine learning model.

Machine learning techniques, with their ability to handle complex relationships and process large-scale data, offer advantages over traditional numerical or probability models (our diffusion work/results are based on analytical derivation). These techniques provide accurate predictions, efficient analysis, and valuable insights across various domains. Furthermore, our study showcases the applicability of our model in predicting gas diffusion in large-scale reservoirs. The

accurate predictions generated by our model have the potential to greatly assist petroleum engineers in reservoir management and production optimization activities.

In conclusion, this study substantiates the efficacy of machine learning methodologies in forecasting gas diffusion across varied rock types. Employing the Random Forest Model, the research identifies key factors that enhance our understanding of gas diffusion behavior within rock formations. These insightful discoveries offer valuable insights applicable to reservoir management, where they can be utilized to optimize processes related to oil and gas extraction. By doing so, efficiency and economic viability in this sector can be significantly improved.

References

Alfredsson, H. A., Hardarson, B. S., Franzson, H., Gislason, S. R., 2008. CO2 sequestration in basaltic rock at the Hellisheidi site in SW Iceland: stratigraphy and chemical composition of the rocks at the injection site. Mineralogical Magazine, 72(01), 1–5.Bagalkot, N., Hamouda, A. A., 2017. Experimental and Numerical Method for Estimating Diffusion Coefficient of the Carbon Dioxide into Light Components. Industrial & Engineering Chemistry Research, 56(9), 2359–2374.

Boon, M., Bijeljic, B., Krevor, S., 2017. Observations of the impact of rock heterogeneity on solute spreading and mixing. Water Resources Research, 53(6), 4624–4642.

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001).

Kan, C., Ye, Z., Zhou, H., Cheruku, S., 2023. DG-ECG: Multi-stream deep graph learning for the recognition of disease-altered patterns in electrocardiogram. Biomedical Signal Processing and Control, 80(2), 104388, ISSN 1746-8094

Philip W. Choquette, Lloyd C. P., 1970. Geologic Nomenclature and Classification of Porosity in Sedimentary Carbonates. AAPG Bulletin, 54.Cui, X., Bustin, A. M. M., Bustin, R. M., 2009. Measurements of gas permeability and diffusivity of tight reservoir rocks: different approaches and their applications. Geofluids, 9(3), 208–223.

Dramsch, J. S. (2020). 70 years of machine learning in geoscience in review. Advances in Geophysics.

Dramsch, J. S., 2020. 70 years of machine learning in geoscience in review. Advances in Geophysics. 61, 1-55.

Faramarzi, N., Sadeghnejad, S.,2020. Fluid and rock heterogeneity assessment of gas condensate reservoirs by wavelet transform of pressure-transient responses. Journal of Natural Gas Science and Engineering, 103469.

Fu, Q., Cudjoe, S., Barati, R., Tsau, J.-S., Li, X., Peltier, K., Bradford, K., 2020. Investigating the Role of Diffusion in Hydrocarbon Gas Huff-n-Puff Injection- An Eagle Ford Study. Journal of Petroleum Science and Engineering, 108146.

Gigos, P. A., Faydi, Y., Meyer, Y., 2015. Mechanical characterization and analytical modeling of gas diffusion layers under cyclic compression. International Journal of Hydrogen Energy, 40(17), 5958–5965.

Gillham, R. W., Robin, M. J. L., Dytynyshyn, D. J., Johnston, H. M., 1984. Diffusion of nonreactive and reactive solutes through fine-grained barrier materials. Canadian Geotechnical Journal, 21(3), 541–550.

Hatiboglu, C. U., Babadagli, T., 2010. Experimental and visual analysis of diffusive mass transfer between matrix and fracture under static conditions. Journal of Petroleum Science and Engineering, 74(1-2), 31–40.

Hu, Q., Wang, J. S. Y.,2003, Aqueous-phase diffusion in unsaturated geologic media: A review, Crit. Rev. Environ. Sci. Technol., 33(3), 275–297.

Liu, A., Liu, P., Liu, S., 2020. Gas diffusion coefficient estimation of coal: A dimensionless numerical method and its experimental validation. International Journal of Heat and Mass Transfer, 162, 120336.

Liu, G., Zhao, Z., Sun, M., Li, J., Hu, G., Wang, X., 2012. New insights into natural gas diffusion coefficient in rocks. Petroleum Exploration and Development, 39(5), 597–604.

Mason, E. A., Malinauskas, A. P., 1983. Gas transport in porous media: the dusty-gas model / E.A. Mason and A.P. Malinauskas. Elsevier.

Mortensen, N. A., Okkels, F., Bruus, H., 2005. Reexamination of Hagen-Poiseuille flow: Shape dependence of the hydraulic resistance in microchannels. Physical Review E, 71(5).

Mukherjee, S., Dang, S. T., Rai, C. S., Sondergeld, C. H., 2020. Measurement of Oil-Gas Diffusivity at Reservoir Conditions for Huff-n-Puff EOR in Shales.

Neil, C.W., Boukhalfa, H, Xu, H., Ware, S.D., Ortiz, J., Avendaño, S., Harp, D., Broome, S., Hjelm, R.P., Mao, Y., Roback, R., Brug, W.P., Stauffer, P. H., 2022. Gas diffusion through variably-water-saturated zeolitic tuff: Implications for transport following a subsurface nuclear event. Journal of Environmental Radioactivity, V250, 106905. ISSN 0265-931X

Ning, X., Feng, Y., Wang, B., 2020. Numerical simulation of Channel Fracturing Technology in Developing Shale Gas Reservoirs. Journal of Natural Gas Science and Engineering, 103515.

Ning, X., Ewing, R., Hu, Q., Wang, Q., Zhang, X., 2021. A new Model for Simulating the Imbibition of a Wetting-Phase Fluid in a Matrix-Fracture Dual Connectivity System. Geofluids, Volume 2022, Article ID 7408123

Peng, S., Hu, Q., Hamamoto, S., 2012. Diffusivity of rocks: Gas diffusion measurements and correlation to porosity and pore size distribution. Water Resources Research, 48.

Qian, F., Chen, L., Li, J., Ding, C., Chen, X., Wang, J., 2019. Direct prediction of the Toxic Gas Diffusion Rule in a Real Environment Based on LSTM. International Journal of Environmental Research and Public Health, 16(12), 2133.

Wang, Q., Hu, Q., Ning, X., Ilavlky, J., Kuzmenko, I., Tom, T., 2021. Spatial heterogeneity analyses of pore structure and mineral composition of Barnett Shale using X-ray scattering techniques. Marine and Petroleum Geology, Volume 134, 105354.

Dane, J. H., Topp, C. G., Rolston, D. E., Moldrup, P., 2002. 4.3 Gas Diffusivity. SSSA Book Series.

Sarker, I. H., 2021. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Computer Science, 2(6).

Sercombe, J., Vidal, R., Gallé, C., Adenot, F.,2007. Experimental study of gas diffusion in cement paste. Cement and Concrete Research, 37(4), 579–588.

Shen, Z., Yang, J., Hu, X., Lei, Y., Ji, X., Jia, J., Wang, W., 2005. Dual Electrodes Oxidation of Dye Wastewater with Gas Diffusion Cathode. Environmental Science & Technology, 39(6), 1819–1826.

Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., Oza, H.,2021. Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research, 6(10).

Tariq, Z., Aljawad, M.S., Hasan, A. et al.,2021. A systematic review of data science and machine learning applications to the oil and gas industry. J Petrol Explor Prod Technol 11, 4339–4374

Tiab, D., Donaldson, E. C., 2016. Porosity and Permeability. Petrophysics, 67–186.

Verbeek, M., 2017. Using linear regression to establish empirical relationships. IZA World of Labor 2017: 336

Wang, H., Yang, X., Du, F., Wang, G., Wang, Y., Zhao, W., Wang, H., 2021. Calculation of the diffusion coefficient of gas diffusion in coal: The comparison of numerical model and traditional analytical model. Journal of Petroleum Science and Engineering, 205, 108931.

Wang, Q., Hu, Q., Ning, X., Ilavsky, J., Kuzmenko, I., Tom, T., 2021. Spatial heterogeneity analyses of pore structure and mineral composition of Barnett Shale using X-ray scattering techniques. Marine and Petroleum Geology, Volume 134, 105354, ISSN 0264-8172.

Yang, Y.; Stenby, E.H.; Shapiro, A.A.; Yan, W. Diffusion Coefficients in Systems Related to Reservoir Fluids: Available Data and Evaluation of Correlations. Processes 2022, 10, 1554.

Ye, Z., Liu, C., Tian, W., Kan, C., 2020. A Deep Learning Approach for the Identification of Small Process Shifts in Additive Manufacturing using 3D Point Clouds. Procedia Manufacturing, 48, 770–775.

Yuan, X., Suvarna, M., Low, S., Dissanayake, P. D., Lee, K. B., Li, J., Ok, Y. S., 2021. Applied machine Learning for Prediction of CO_2 Adsorption on Biomass Waste-Derived Porous Carbons. Environmental Science & Technology, 55(17), 11925–11936.

Zamel, N., Li, X., Shen, J., 2009. Correlation for the effective Gas Diffusion Coefficient in Carbon Paper Diffusion Media. Energy & Fuels, 23(12), 6070–6078.

Zemel, R., Pitassi, T., 2000. A gradient-based boosting algorithm for regression problems. Advances in neural information processing systems.

Zhao, W., Cheng, Y., Pan, Z., Wang, K., Liu, S., 2019. Gas diffusion in coal particles: A review of mathematical models and their applications. Fuel, 252, 77–100.

Zhao, H., Lin, C., Xingkai, Z., Ning, X., 2020. Uncertainty analysis and Optimization in Cyber-Physical Systems of Reservoir Production. Big Data Analytics for Cyber-Physical Systems, 215-229

Zhao, H., Xu, L., Guo, Z., Liu, W., Zhang, Q., Ning, X., Shi, L., 2019. A new and fast waterflooding optimization workflow based on INSIM-derived injection efficiency with a field application. Journal of Petroleum Science and Engineering, 179, 1186–1200.

Zhang, M., Chakraborty, N., Karpyn, Z. T., Emami-Meybodi, H., Ayala, L. F., 2021. Experimental and numerical study of gas diffusion and sorption kinetics in ultra-tight rocks. Fuel, 286, 119300.

Chapter V

Conclusions

This thesis undertakes a methodical exploration encompassing multi-scale and diverse approaches to the petrophysical characterization and fluid flow analysis of various natural rock formations. Drawing insights from Chapters II to IV, the ensuing key observations can be distilled as follows:(1) In Chapter II, the distribution of fractures significantly influences the imbibition process. In scenarios where multiple crossed fractures exist, they create efficient conduits that amplify conductivity within the matrix. The preferential movement of water along fractures results in a larger mass of imbibed water over time in rocks with a P-direction compared to those with a T-direction. In cases where horizontal fractures are present, water movement is impeded, even in a homogeneous matrix. The upper grid cells are isolated from the water percolated from the lower grid cells due to the heterogeneous fractures, underscoring their inhibitory effect.

(2) Sensitivity analysis underscores the percolation probability's role in dictating the imbibition rate. Vapor saturation and matrix grid cell probability wield substantial influence over the imbibed water mass, particularly during the early stages of imbibition. These effects are more pronounced in rock samples featuring T-direction fractures.

(3) The dissimilar rates of water imbibition can be attributed to fracture distribution. P-fractures, predominantly situated in contact with the bottom layer, exhibit higher water saturation in grid cells and subsequently lead to a higher probability of matrix cell saturation. Conversely, T-fractures, while saturating the bottom layer, impede the upward movement of water, resulting in

a more restrained distribution. The steady imbibition rate may be attributed to the absence of fractures, mitigating the acceleration of water imbibition.

(4) In Chapter III, this study delves into the distinctive behaviors of various rock types during water imbibition, revealing valuable insights into their unique characteristics. Alabama marble, marked by its low porosity and high contact angle, exhibited hindered water imbibition due to its intrinsic properties. Carthage marble displayed lower porosity and permeability, resulting in minor slope changes during the imbibition process. Sandstones showcased a dual-stage imbibition process, characterized by a steady rate of water weight addition in the initial phase and an abrupt shift in the subsequent phase, indicating a transition from main pore filling to minor pore saturation. Indiana and Carthage limestones experienced a brief primary pore-filling stage due to their lower porosity but shared a similar second phase akin to sandstones. Rock sample size emerged as a significant influencer of water imbibition velocity, with the distribution of pore throats and contact angle exhibiting comparatively lesser effects. The consistent outcomes across figures underscored the prominence of sample size, porosity, and permeability in governing the rate of water weight gain through imbibition. In contrast, the influence of rock type and contact angle remained relatively marginal across the three machine-learning models.

(5) Chapter IV underscores the efficacy of the Random Forest regression technique in predicting gas diffusion behaviors and extracting influential attributes. It highlights that the discrepancy in gas sample concentrations, rock dimensions, permeability, and intrinsic rock properties plays a pivotal role in governing gas diffusion through rock structures. A significant contribution of this research lies in its successful demonstration that intricate geological phenomena can be quantitatively analyzed and empirically validated through the implementation of machine learning models. This study identifies that noise data in the early stages of the gas diffusion

process could potentially hamper the regression model's performance. As the concentration difference between the two aspects of the sample decreases, the gas diffusion mechanism may enter a state of stochastic turbulence, inducing unpredictable fluctuations in gas movement. Thus, the preprocessing elimination of noise data from the test set emerges as a critical step before employing the machine learning model.

In summary, machine learning techniques, with their capacity to handle intricate relationships and process extensive datasets, offer distinct advantages over traditional numerical or probability models. These techniques deliver accurate predictions, efficient analyses, and valuable insights across diverse domains. Additionally, our study demonstrates the model's applicability in predicting gas diffusion within large-scale reservoirs. The precision of our model's predictions can significantly aid petroleum engineers in optimizing reservoir management and production processes.

Biographical Information

Originally from Dongying, Shandong, China, Xuewei Ning began her academic journey at Yangtze University, where she completed her Bachelor of Science degree in Petroleum Engineering in 2017. She then pursued a Master's program at the same institution, subsequently expanding her studies to the United States. In May 2020, she successfully achieved dual Master's degrees from both Yangtze University and the University of Louisiana at Lafayette, where she studied Petroleum Engineering.

Later in August 2020, Xuewei joined the University of Texas at Arlington, where she embarked on her doctoral journey in the Earth and Environmental Science Department, with a focus on geosciences. Her research explores fluid flow and gas diffusion in natural rocks, employing both experimental methods and simulation studies. This specialized area of study allows her to contribute meaningfully to our understanding of geophysical processes within natural rock formations.