APPLICATIONS OF DATA-DRIVEN CLASSIFICATION AND CONNECTIVITY

QUANTIFICATION METHODS IN HIGH-RESOLUTION IMAGE ANALYSIS

A Thesis

by

ELIZA GANGULY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Siddharth Misra
Committee Members,	Alan Yu
	Yuefeng Sun
	I. Yucel Akkutlu
Head of Department,	Jeff Spath

May 2021

Major Subject: Petroleum Engineering

ABSTRACT

Microstructure of a material determines the transport, chemical and mechanical properties. Geological materials and geomaterials are imaged using microscopy tools. The microscopy images are analyzed to better understand the microstructural topology and morphology. Image segmentation is an essential step prior to the microstructural analysis. In this study, we trained a Random Forest classifier to relate certain features corresponding to each pixel and its neighboring pixels in a scanning electron microscopy (SEM) image of shale to a specific component type; thereby developing a methodology to segment SEM images of shale samples into 4 component types, namely, pore/crack, organic/kerogen, matrix and pyrite. We evaluate the generalization capability of the Machine Learning-assisted image-segmentation (MLIS) method by using SEM maps from Wolfcamp and Barnett shale formations. The two formations differ in topology, morphology and distribution of the four components.

The MLIS method is also implemented to classify rock and different fluid phases in micro-CT scans of carbonate core sample undergoing water alternating gas injection with a goal to quantify the three-dimensional fluid connectivity. The three-dimensional connectivity of the fluid phases in porous media plays a crucial role in governing the fluid transport, displacement, and recovery. Accurate three-dimensional quantification of the fluid phase connectivity following each fluid injection stage will lead to better understanding of the efficacy and efficiency of the fluid injection strategies. Two metrics for measuring the connectivity in 3D show robust performance; one uses fast marching

method to quantify average time required for a monotonically advancing wave to travel between any two pixels and the other uses two-point probability function to approximate the average distance between any two connected pixels belonging to the same fluid phase. The two connectivity metrics are applied on the three-dimensional (3D) CT scans of one water-wet Ketton whole-core sample subjected to WAG injection to quantify the evolution of the three-dimensional connectivity of the three fluid phases (oil, water, and gas).

DEDICATION

Dedicated to my grandmother, in loving memory.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Misra for his continued guidance and support throughout the course of this research. I would like to acknowledge the help and consideration of my committee members, Dr. Akkutlu, Dr. Yu, and Dr. Sun, at every step of the thesis defense.

Also, thanks to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my father, mother and brother for their constant encouragement and to my boyfriend for his patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Dr Siddharth Misra and Dr I. Yucel Akkutlu of the Department of Petroleum Engineering and Dr Alan Yu and Dr Yuefeng Sun of the Department of Geology and Geophysics.

The data used in Chapter 2 was provided by the Integrated Core Characterization Center (IC3) at the University of Oklahoma. The data used in Chapter 3 were from the project "Water-wet three-phase flow micro-CT tomograms" uploaded on the Digital Rocks portal by Alessio Scanziani, Kamaljit Singh and Martin Blunt (Imperial College London). The analyses depicted in Chapter 2 and 3 were conducted in part by Yaokun Wu of the Department of Petroleum Engineering and were published in 2019 and 2020.

All other work conducted for the thesis was completed by the student independently.

NOMENCLATURE

SEM	Scanning Electron Microscopy
CT	Computed Tomograghy
MLIS	Machine Learning-assisted Image Segmentation
MDA	Mean decrease Accuarcy
TOC	Total Organic Carbon
WAG	Water Alternating Gas
EOR	Enhanced Oil Recovery
CF	Connectivity Function
FMTT	Fast Marching-based Travel Time
PI	Pre-injection
WF1	Water-flooding #1
GI	Gas Injection
WF2	Water-flooding #2

TABLE OF CONTENTS

LIST OF FIGURES

xi

LIST OF TABLES

CHAPTER I

INTRODUCTION

Data-driven methods can help conduct time-consuming and complex tasks faster, while delivering more accurate results compared to manual efforts. This thesis demonstrates the application of such data-driven methods in two different problems pertinent to the oil and gas industry.

In chapter two, we discuss a machine learning approach to accurately segment the boundary pixels belonging to pore/crack and organic/kerogen components in SEM image of organic-rich shale. Machine learning refers to the technology that uses statistics to find trends in massive amount of data in an autonomous way without being explicitly programmed. This data can be in the form of text, images, numbers, videos, or anything that can be digitally stored. Machine learning is focused on providing systems the ability to learn from experience and improve the decision-making or predictive accuracy over time. Examples of machine learning implemented in everyday activities include internet search engines, spam detector for emails, personalized recommendations on websites and streaming platforms, virtual assistants with voice recognition ability and so on.

On a broader scale, machine learning can be divided into 3 categories: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning encompasses algorithms that require dataset labelled with the information that the model is built to predict. The model learns the labels from the data (called training) and can further classify or predict new, unlabeled data. Unsupervised learning requires large

amount of unlabeled data and uses algorithms to extract meaningful features needed to group or sort the data in real-time without human intervention. Semi-supervised learning provides a middle ground between these two approaches. It can train on a small labeled dataset to help extract features from a larger unlabeled dataset.

In chapter three, different statistical methods are presented that are capable of capturing the evolution of fluid phase connectivity in a core sample during water alternating gas (WAG) injection. The parameter 'connectivity' is utilized in differently in various disciplines. It is a critical estimator in characterizing other physical attributes of a system. Several physics-based and statistical methods have been adopted to quantify connectivity at different levels. This study focuses on measuring the pixel connectivity of a fluid of interest from a binarized core image during multistage injection.

In chapter four, we discuss the limitations of the current methods presented in chapters two and three and recommend ways in which the workflows can be used further for other scenarios.

CHAPTER II

PORE AND KEROGEN DETECTION FROM MICROSCOPIC IMAGES OF SHALES

USING MACHINE LEARNING

Introduction

Literature review

Machine learning have been implemented in different computer vision applications, such as, image segmentation, image classification and object detection. U-net, a deep learning architecture for image segmentation, is used for lung CT image analysis that aids in the detection of malignant lung tumor causing the lethal lung cancer [1]. Algorithms like Random Forest, Support Vector Machines, Neural Networks have been widely used for image classification applications in remote sensing, medical imaging, facial recognition and so on [2,3,4]. Machine learning techniques have been widely implemented in the recent years for interpretation of images in the domain of petroleum engineering and geosciences. These methods can help improve the characterization of subsurface images by identifying underlying patterns that may be obscured through traditional means of visualizing and image processing [5]. Al-Farisi et al. (2019) performed machine learning assisted image recognition using random forest algorithm on 3D µCT and MRI images to determine the lithologies and porosity of a carbonate rock. The porosity was validated with three sets of experimental results and was found to be very close to the experimental porosity value [6]. The time-consuming task of interpreting rock types from petrographic features can be performed in an efficient and accurate way by leveraging machine learning

methods to identify geologic patterns from thin section images [7]. High resolution microscopy images can effectively capture the complex microlevel variations in the rock structure. Analysis of these microstructural components can aid better understanding of the fluid pathways and rock-fluid interaction. Image segmentation is the preliminary step of image processing. Segmentation refers to the process of assigning a label to each pixel present in the image, thereby creating a representation of the image that is more meaningful and easier to analyze.

In this study, the SEM images of shale are classified into four components, namely, pores/cracks, kerogen/organic, matrix and pyrite using machine learning assisted image segmentation (MLIS). Random Forest, an ensemble classification technique, is implemented to accomplish the segmentation. A crucial expectation from machine learning models is the generalization capability, or in other words, the ability to be applicable on data that has not been used to train the model. This project focuses on building a generalizable, robust classification model on limited amount of training data (images) capable of accurately segmenting new, unseen SEM images of shale formations. The study makes use of two SEM maps, which are continuous sequence of SEM images of organic-rich shale samples from Wolfcamp and Barnett formations.

Scientific questions

This chapter aims at answering the following scientific questions:

- How to improve the method of identification of different important components like pores and kerogen from SEM images of shales?

- How can we obtain the pore and kerogen size distribution in shales without conducting petrophysical experiments?

- What are some ways in which a classification model can be optimized in terms of generalization capability?

- How can the effect of imbalanced dataset on the performance of a classifier be reduced?

- How can we identify the features/variables having the highest contribution on the performance of a model on a test data?

Method

The machine learning assisted image segmentation (MLIS) is developed on two continuous: sequence of SEM images (called maps) of shale from two different formations, namely, the Wolfcamp map and the Barnett map. These two formations highly vary in the distribution of their pore network, organic and inorganic components. The goal of this study is to train a robust Random Forest classifier capable of accurately identifying the different components from SEM images of both the formations. **Fig. 1** shows the MLIS workflow built on the two shale maps.
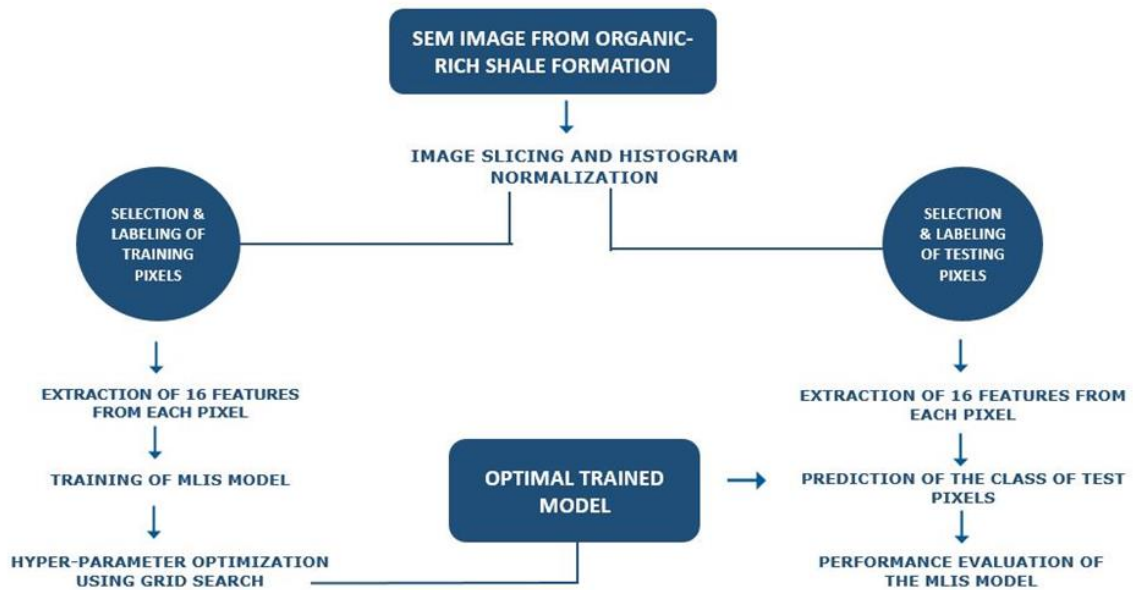
Figure 1. Workflow of the Machine Learning assisted Image Segmentation (MLIS) method

Dataset

The two SEM maps were captured using the FEI Helios Nanolab™ 650 DualBeam™ FIB/SEM machine and FEI SEM MAPS™ software in the Integrated Core Characterization (IC3) lab at the University of Oklahoma. The Wolfcamp map covers an area of 260.6 µm by 2058 µm consisting of 26060×205800 pixels and The Barnett map covers an area of 164.64 µm by 182.42 µm containing 16464×18242 pixels. Wolfcamp and Barnett maps were sliced into 1000 and 56 smaller images respectively with consistent dimensions compatible with the proposed segmentation workflow. All smaller image slices were of dimension 2058×2606 pixels. Image slices are numbered 1 to 1000 in Wolfcamp map and 1 to 56 in The Barnett map. The slices do not contain overlapping pixels.

6

In data-driven methods, training data is chosen such that the distribution of the different classes (in this case, components) in the train data is representative of the distribution of the classes in the entire dataset. It is also important to select adequate number of samples belonging to each class for a robust training of a classification model. In this study, we select 705, 2074, 17373 and 15000 pixels corresponding to the pore/crack, organic/kerogen, matrix and pyrite components, respectively from image slice 90 of The Wolfcamp map, whereas, 912, 8435, 5806 and 5387 pixels of the four components from image slice 35 of The Barnett map for purposes of training (**Fig. 2**).
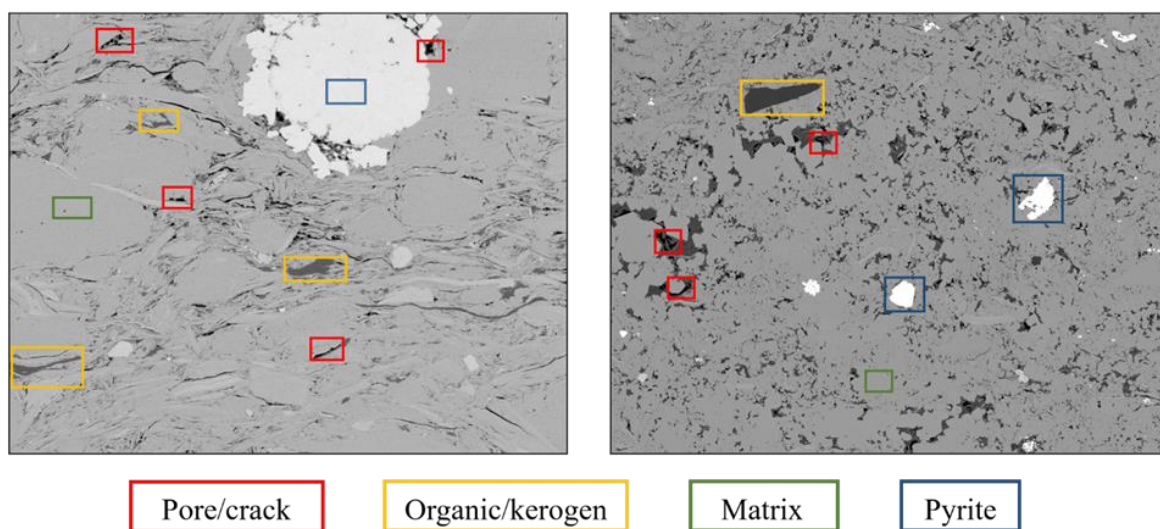
Figure 2. Training image from the Wolfcamp (slice 90: left) and Barnett (slice 35: right) formations (Reprinted from Ganguly et al. 2020)

Feature extraction is then performed on each pixel and the neighboring pixels in the training dataset. Feature extraction is a useful step in case of a large dataset. It helps reduce the amount of redundant data and increases the speed of learning and generalization of the machine learning model. In this case, pixel intensity serves as the primary feature.

7

Fifteen more features have been calculated to ensure a robust learning process for the MLIS model. These features are explained in detail by Misra and Wu (2020) and Wu et al. (2019) and consist of the following [8, 9]:

- Gaussian blur (1 feature): A non-uniform low-pass filter that preserves low spatial frequency and reduces image noise and negligible details in an image that results in a blurred version of the image.

- Difference of Gaussian (1 feature): Acts as a band-pass filter that delineates local structures (e.g. blobs) in image.

- Local statistical information (3 features): Maximum, minimum and mean of intensities of all pixels inside a 3×3 grid centered at the target pixel to be classified.

- Wavelet decomposition (6 features): Wavelet transform enables multiresolution space-scale (time–frequency) analysis of signals. Unlike Fourier transform, wavelet transform preserves local features and are suitable for non-stationary signals.

- Hessian affine region detector (3 features): Hessian affine region detector describes the second-order partial derivatives of the local intensity around a pixel. It is suited for detecting even edges and differentiating between tubular, sheet-like and blob-like structures present in an image.

- Sobel edge detector (1 feature): Sobel operator performs a 2-D spatial-gradient operation on an image to enhance the edges. This helps segmentation of the boundary pixels.

Model development

The sixteen features extracted from each pixel is fed into a Random Forest model for robust classification. Random Forest is an ensemble tree-based supervised learning algorithm. It consists of a set of decision trees and aggregates the votes from the trees to decide the final output. This helps the model outperform the decision from any single tree and makes the algorithm more robust. The Random Forest model used in the MLIS workflow uses the different features as input and classifies each pixel into one of the four classes, namely, pore/crack, organic/kerogen, matrix and pyrite.

Hyper-parameter optimization

Hyper-parameter of a machine learning model is a parameter that is user set prior to the start of the learning process. The set of hyper-parameters define the model architecture and vary from one algorithm to another. Few important hyper-parameters of the Random Forest model are:

n_estimators: The number of individual decision trees that will be constructed in the Random Forest.

max_depth: The maximum depth at which the tree will expand.

min_samples_split: The minimum number of samples required to split an internal leaf node.

min_samples_leaf: The minimum number of samples required to be at a leaf node.

Different combinations of hyper-parameters can be explored to find the optimal model architecture that results in low memorization error and lowest generalization error. This method is called hyper-parameter optimization. In this study, a total of 1512

combinations of hyper-parameters were tried using a grid search method, followed by a 3-fold stratified cross-validation. The optimal set of hyper-parameters were used in the final model.

Performance evaluation

The performance of the classifier was evaluated using precision, recall and weighted F1 score. Precision answers the question, "what proportion of the positive identification was actually correct?" and measures the reliability of the label assigned by a classifier, whereas, recall answers the question, "what proportion of actual positives was identified correctly?" and measures the ability of a classifier to correctly assign a label. F1 score is the harmonic mean of precision and recall. All the three metrics range from 0 to 1, 0 denoting poor classification and 1 denoting robust performance.

$$Precision = \frac{(True\ Positive)}{(True\ Positive + False\ Positive)} \tag{1}$$

$$Recall = \frac{(True\ Positive)}{(True\ Positive + False\ Negative)} \tag{2}$$

The terms True Positive, False Positive and False Negative can be explained with the help of an example where the class of interest in pore/crack. If a pixel belonging to the pore/crack component type is predicted to be pore/crack by the classifier, it is a case of true positive, while, if a pixel belonging to a different component is predicted to be pore/crack by the classifier, it is a case of false positive. Alternatively, true negative is a case where a pixel belonging to a component type other than pore/crack is correctly

10

categorized by the classifier, whereas, false negative is a case where a pixel belonging to the pore/crack component type is incorrectly classified to be a different component. In this study, the dataset is imbalanced, with more pixels belonging to matrix and pyrite components, therefore, F1 score weighted by the support pixel for each component is used to obtain the accuracy.

Result and discussions

In this section, we conduct a detailed comparison of the machine learning assisted image segmentation (MLIS) workflow developed on images from the two shale formations, Wolfcamp and Barnett. We analyze the performances of the two models and present an approach to improve the classification accuracy. The two SEM maps vary in their topology and exhibit slightly different range of pixel intensities as a result of difference in the microscope settings. In order to study the reliability and generalizability of the machine learning models, the model learning is divided into two approaches:

- Approach 1: learning from images of individual formations (includes training on one formation and testing on another)

- Approach 2: learning from images of both the formations combined (includes training on combined dataset from the two formations and testing on each of the formations)

In this document, the model trained on Wolfcamp SEM image is called Model-1, that trained on the Barnett SEM map is called Model-2 and the model trained on the combined dataset is termed Model-3.

11

Approach 1: learning from images of individual formations (includes training on one formation and testing on another)

Model-1 is trained on image slice 90 of SEM map from the Wolfcamp formation. It is deployed on the SEM map from the Barnett formation to identify the inner and boundary region pixels belonging to pore/crack, organic/kerogen, matrix and pyrite components. To evaluate the robustness of the proposed workflow, we train the model on one formation and deploy it on a different formation. The model performance is measured in terms of precision, recall and F1 score. A low precision for a certain component indicates that one or more of the other components present in the image is being misclassified as that particular component.

Is it evident from **Fig. 3** that Wolfcamp and Barnett shale formations significantly vary in the distribution of pore/crack component. While the Wolfcamp formation is characterized by the presence of cracks in the form of thin, elongated strips, the Barnett formation sees a dominance of pores embedded in organic matter. Fig 3. compares the segmentation performance on one image from the Barnett map using Model-1 described in approach 1 versus using a model trained on the Barnett map, which in this case, would generate the ideal performance as the train and test maps are same. The model trained and tested on Barnett map have been elaborated by Misra et al. (2020) [10].
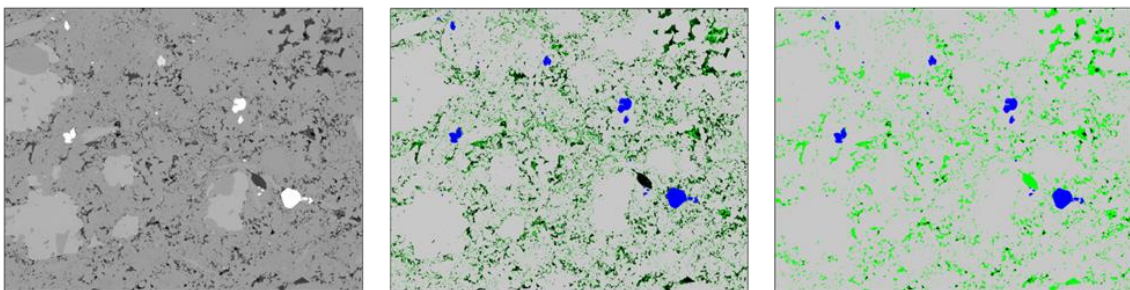
Figure 3**. Comparison of segmentation of slice 15 of the Barnett map (left) using model trained of Wolfcamp map (mid) and model trained on Barnett map (right).**

When MLIS Model-1 is deployed on Barnett map, a low precision of 0.41 for the inner region pixels of pore/crack components is obtained, coupled with a low recall of 0.49 for inner region pixels of the organic/kerogen component. This indicates that a significant number of pixels belonging to organic/kerogen component in the Barnett shale map is being misclassified as pore/crack. The model is trained on strips of black pixels labeled as pore/cracks and fails when organic/kerogen pixels are arranged in the form of thin, elongated structures. The matrix and pyrites show a relatively higher precision and recall, since the variation in pixel intensities between these constituents and pore/cracks and organic/kerogen are more detectable than that between pore/cracks and organic/kerogen. In the Barnett map, the pores also share considerable boundary area with organic/kerogen, thereby resulting in a lower accuracy for the outer region pixels of these two components (**Fig. 4**).
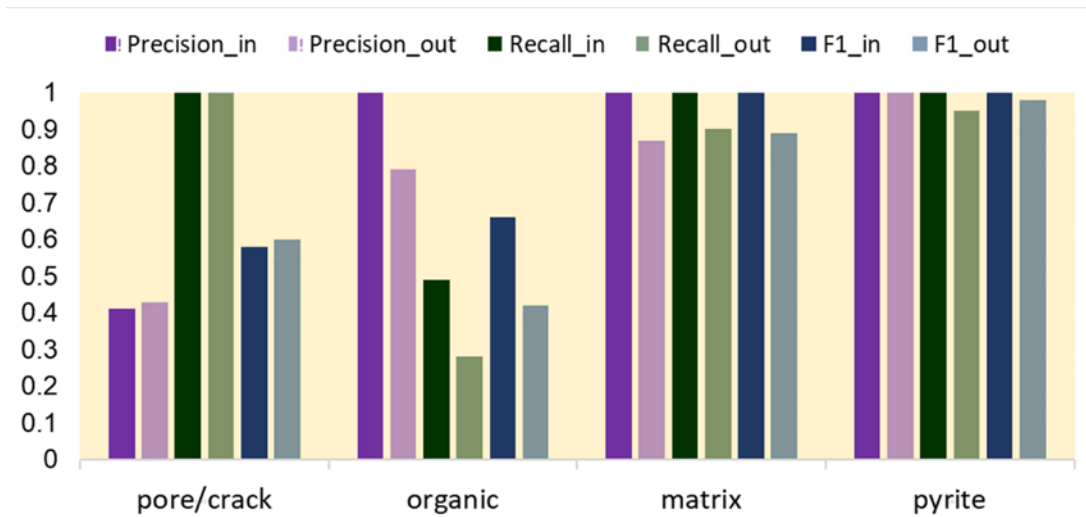
Figure 4. Performance of Model-1 trained on the Wolfcamp map on inner and outer region pixels of the Barnett map.

In the next step, we trained the MLIS model (Model-2) on the images from the Barnett formation and deployed it on the images from the Wolcamp. To determine the robustness of the model, we test it both inner and outer region pixels of the four components in the Wolfcamp SEM map. Model-2 achieves an F1 score of 0.89 when deployed on the inner region pixels and 0.91 when deployed on the outer region pixels of Wolfcamp map. This is a very encouraging performance. These scores are found to be higher than the F1 scores obtained when Model-1 was deployed on Barnett map inner region pixels (F1 score of 0.82) and outer region pixels (F1 score of 0.81). This suggests that The Barnett map has more generalizable statistical features that can serve to better train the segmentation model and The Wolfcamp map has simpler microstructural features that are easier to identify.

Model-2 learns from assorted clusters of pixels labeled as pores and therefore, fails to accurately identify cracks that are present in the Wolfcamp map as elongated strips of pixels. This results in a low recall for pore/crack component along with a low precision for organic/kerogen components (**Fig. 6**). This observation can also be verified from **Fig. 5** which shows the original (far left) and segmented (far right) image slice of Wolfcamp SEM map. We see pixels belonging to cracks being identified as organic/kerogen in the right image. Fig. 5 (mid) shows the ideal segmentation when the model is trained as well as tested on the Wolfacmp formation map (Ganguly et al., 2020) [11].
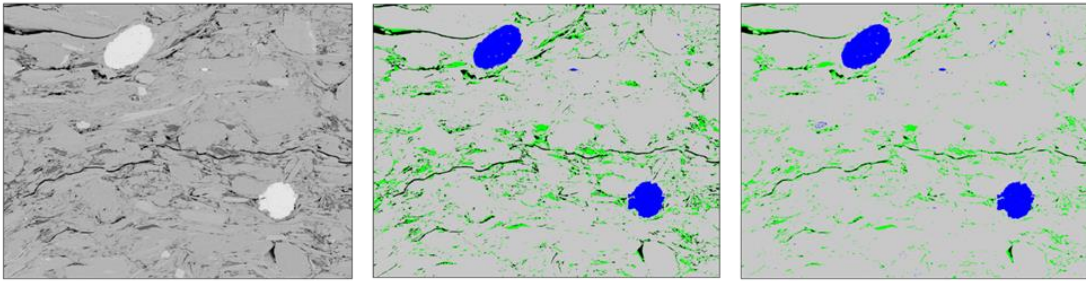
Figure 5. Comparison of segmentation of slice 60 of the Wolfcamp formation (left) using model trained of Wolfcamp map (mid) and model trained on Barnett map (right).
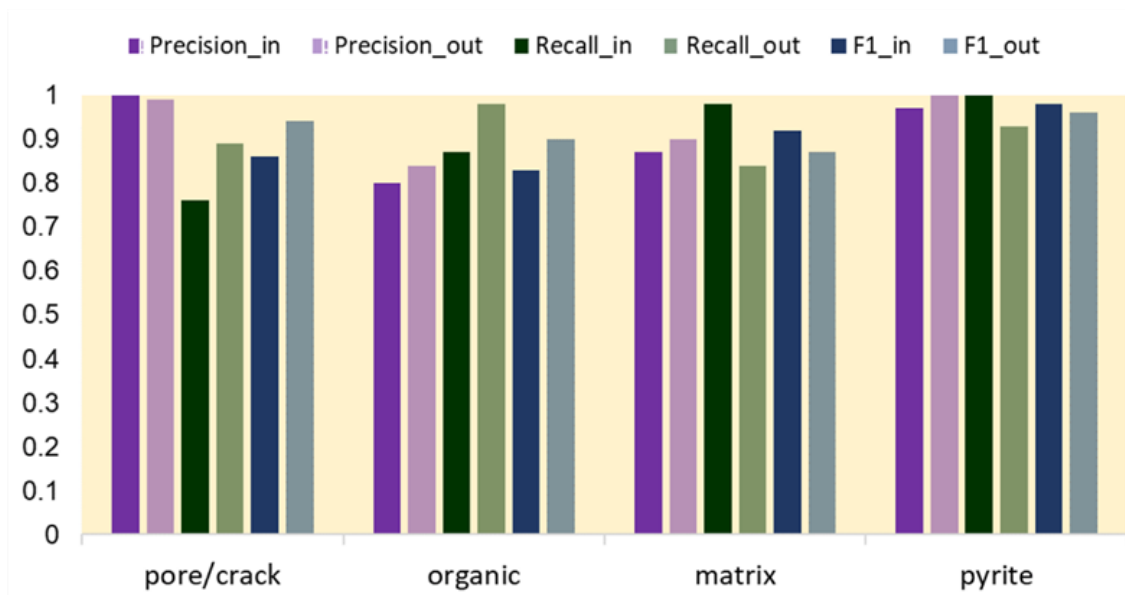
Figure 6**. Performance of Model-2 trained on the Barnett map on inner and outer region pixels of the Wolfcamp map.**

Approach 2: learning from images of both the formations combined (includes training

on combined dataset from the two formations and testing on each of the formations)

In the previous approach, we notice that the models fail to perform well, especially for the

pore/crack and organic/kerogen components, when deployed on a formation other than the

one used for training. In this approach, we train a Random Forest classifier on a combined dataset of pixels selected from SEM images of both the formations. Sixteen features are extracted from training pixel as before. The developed MLIS model is tested on inner and boundary region pixels selected from multiple images from both the formations.

We observe a significant improvement in the performance of the classifier trained on the combined dataset on the inner and boundary region pixels of both pore/crack and
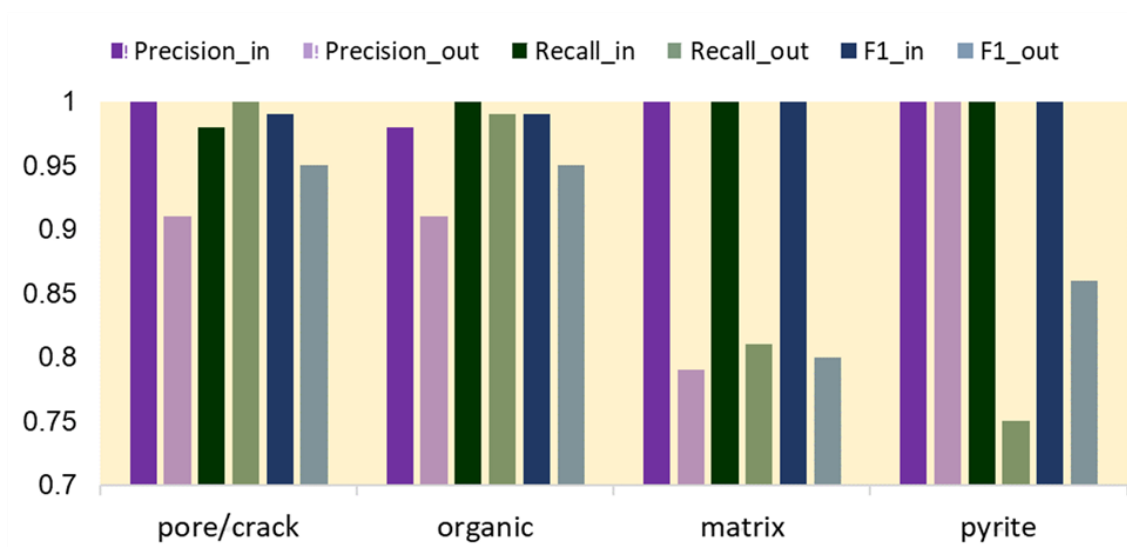
organic/kerogen components of the Wolfcamp map (**Fig. 7**) as well as Barnett map (**Fig. 8**). For inner region pixels, the average F1 score reaches a perfect score of 1.00 for the Wolfcamp formation and of 0.99 for the Barnett formation. For outer/boundary region pixels, the model obtains an average F1 score greater than 0.8 for both the formations, thereby making Model-3 the best performing classifier with respect to generalizability.

Figure 7. Performance of Model-3 trained on the combined dataset on inner and outer region pixels of the Wolfcamp map.
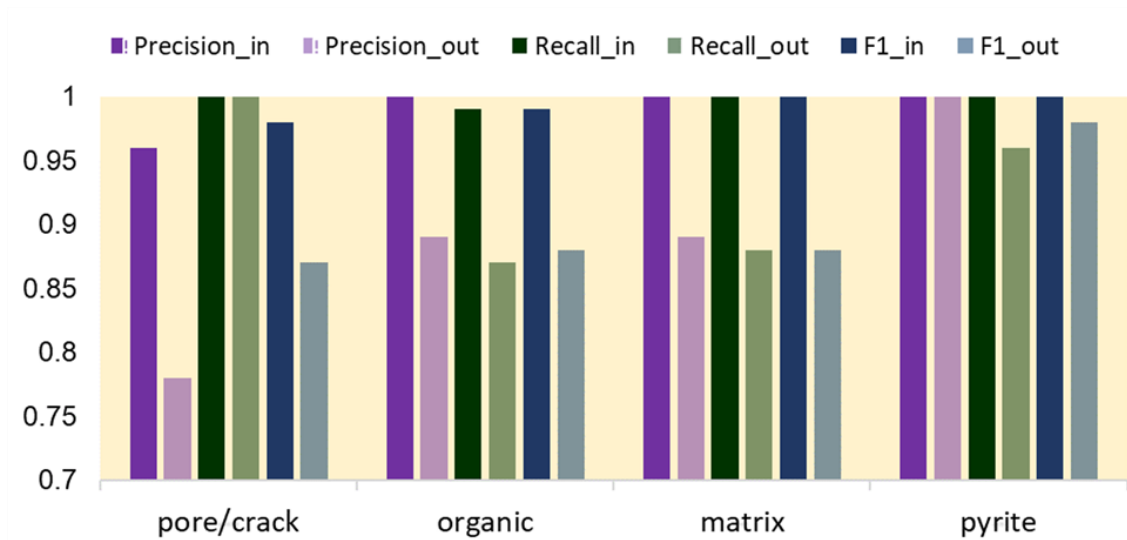
Figure 8. Performance of Model-3 trained on the combined dataset on inner and outer region pixels of the Barnett map.

Fig. 9 compares the generalization capability of the three aforementioned image segmentation models. It is evident that Model-3 trained on the combined dataset exhibits a considerably higher average F1 score for all four scenarios in Fig. 9 and **Fig. 10** compared to the two models trained on individual SEM maps. It can also be noticed that, Model-3 improves the generalization performance on the inner region pixels without compromising on the segmentation of outer region pixels. Therefore, Model-3 can be established as the most robust classifier for the machine learning assisted image segmentation (MLIS) method.
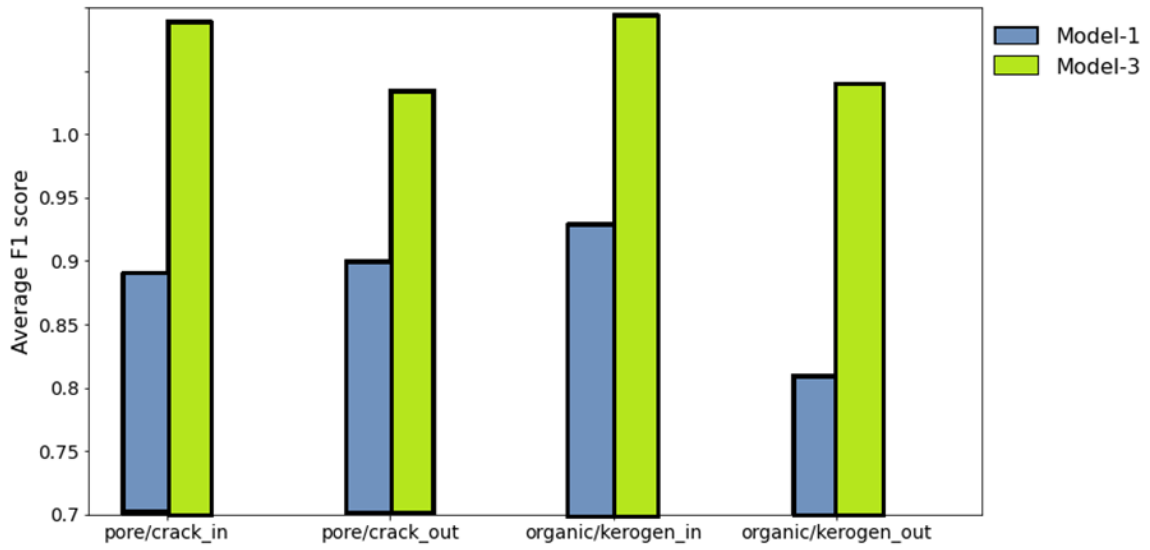
Figure 9. Comparison of the average F1 scores for the pore/crack and organic/kerogen components when Barnett map is segmented using Model-1 (approach 1) and Model-3 (approach 2)
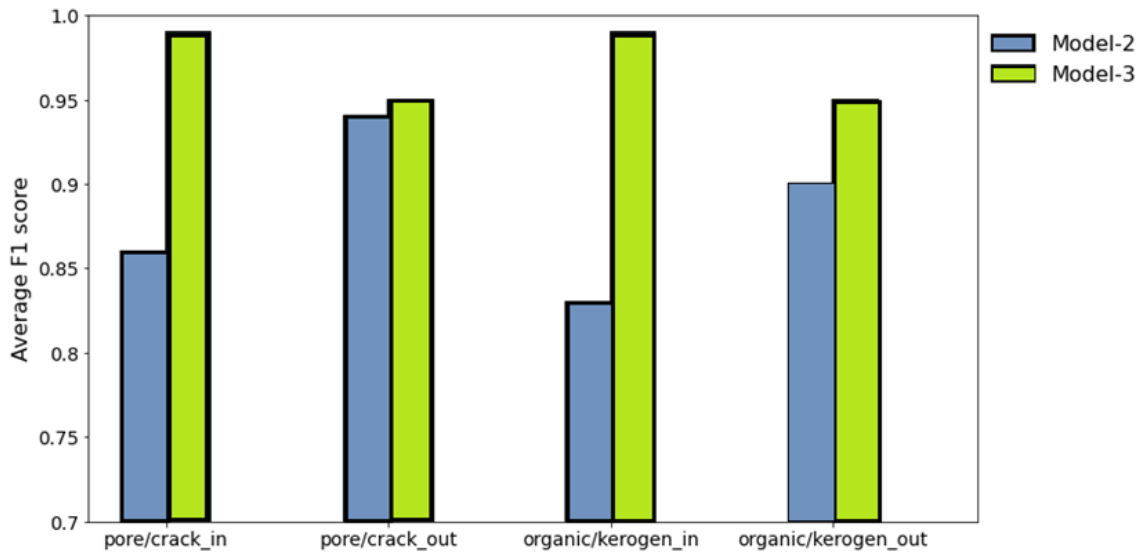
Figure 10. Comparison of the average F1 scores for the pore/crack and organic/kerogen components when Wolfcamp map is segmented using Model-2 (approach 1) and Model-3 (approach 2)

Fig. 11 and **Fig 12** shows a visual comparison of the Barnett map and Wolfcamp map respectively when segmented using approach 1 (left) versus when segmented using approach 2 (right).
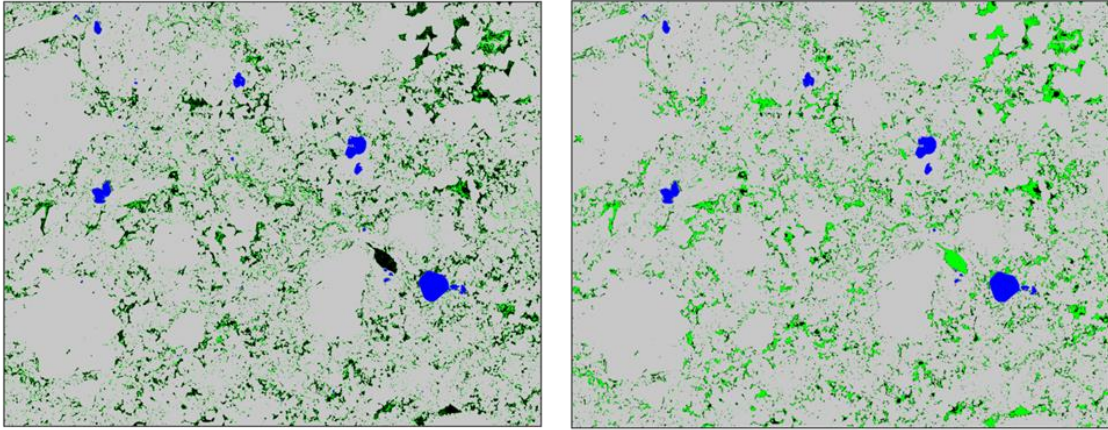
Figure 11. Optimized segmentation performance of the Barnett map from Model-1 (left) to Model-3 (right)
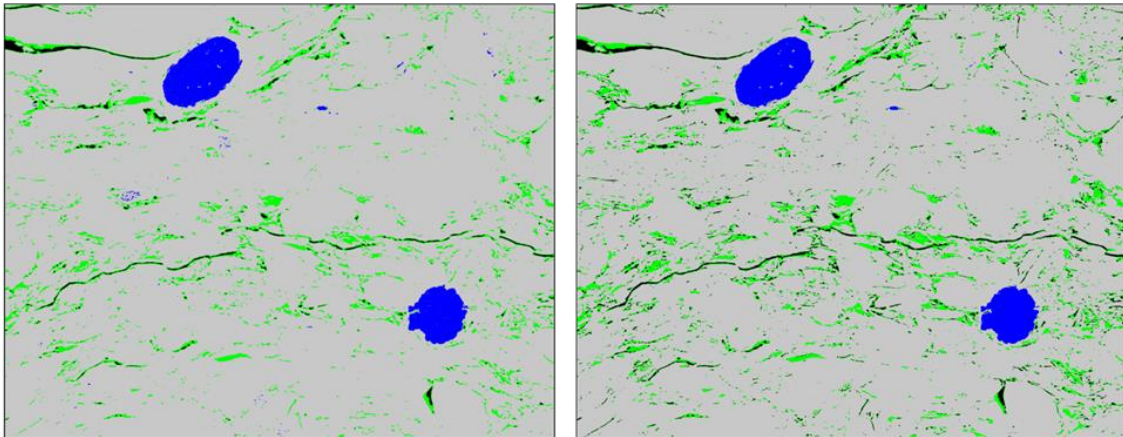
Figure 12**. Optimized segmentation performance of the Wolfcamp map from Model-2 (left) to Model-3 (right)**

Feature selection

In supervised learning, feature selection helps improve the performance of a model and reduces the curse of dimensionality. It involves feature ranking followed by feature

elimination. Feature ranking employs different methods to quantify the contribution of each feature in the overall prediction of the model. Most popular feature ranking methods are mean decrease accuracy (MDA) and mean decrease impurity (MDI). In this study, we use 16 different features to train the each of the three classifier models. We deployed the MDA method to investigate the effect of the features on the prediction of the classifiers. Scikit-Learn compatible black-box estimator called Permutation Importance have been used for the purpose. This estimator ranks a feature based on the decrease in the specified scoring metric when a particular feature is not available (permuted) while testing a trained model. All features are used while training the model and each feature is excluded during deployment of the model on the test set. The more the test accuracy suffers, the higher the importance of that feature in the prediction capability of the model. The estimator outputs a weight associated with each feature to quantify the importance.

In this study, we measure the feature importance for two different cases:

➢ Case I: Testing the performance of Model-3 on pixels from the Wolfcamp shale sample

➢ Case II: Testing the performance of Model-3 on pixels from the Barnett shale sample.

Fig. 13 shows the five most important features for each of the two cases. We observe that in both cases pixel intensity, Gaussian blur and local statistical information like min, max and mean show a high contribution to the prediction performance of the model. Notably, even though Cases I and II are trained on the same training dataset (Model 3), the dominant features change depending on the testing dataset processed by Model 3.
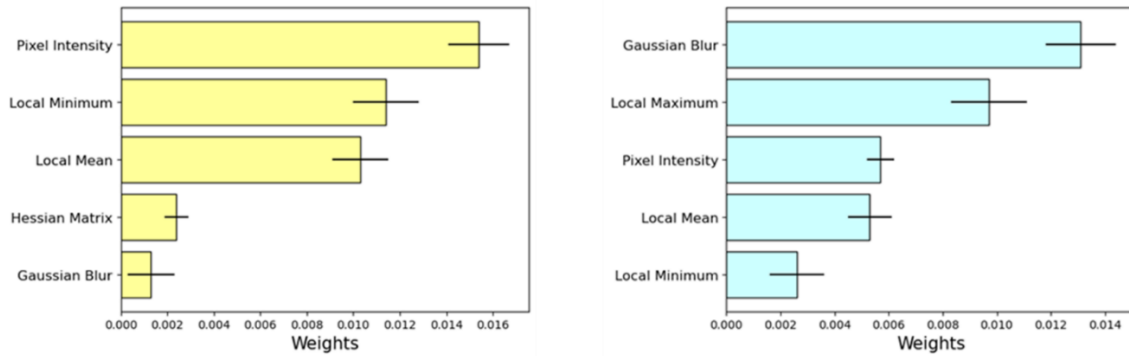
Figure 13: Five most important features ranked by the mean decrease accuracy (MDA) method in case of segmentation of the Wolfcamp map (Case I: left) and Barnett map (Case II: right) using Model-3.

Estimation of Petrophysical Parameters Based on the Identification of Pores, Cracks

and Kerogen

Above sections, describe the development and evaluation of a robust method to identify hydrocarbon storage and transport pathways along with the kerogen distribution in the SEM maps from shale formations. After the identification of pores, kerogen, and cracks, one can quantify various petrophysical and petrological characteristics. On those lines, we quantify porosity distribution, pore size distribution, and kerogen size distribution in the SEM map from the Barnett shale formation. It covers an area of 164.64 µm by 182.42 µm containing 16464×18242 pixels. The map was sliced into 56 image slices, each consisting 2058×2606 pixels and covering an area of 20.58 µm by 26.06 µm. The total porosity from the Barnett map was found to be 0.45%.

Fig. 14 shows a grayscale heatmap with cells representing the 56 (8 rows by 7 columns) slices of the Barnett map. The plot is annotated with the porosity values associated with each image slice. The porosity ranges from 0.1 % to 0.8% in the entire sample representing an area of 164.64 µm by 182.42 µm. The right side of the map is

dominated by low-porosity regions. Both high porosity and low porosity regions are localized and predominantly surrounded by regions having 0.4% porosity. Extremely low and high porosity regions are isolated, whereas 0.5% porosity regions are the most well connected.

0.006	0.006	0.005	0.005	0.005	0.003	0.004
0.005	0.006	0.004	0.004	0.005	0.007	0.003
0.004	0.004	0.006	0.006	0.003	0.003	0.004
0.004	0.001	0.004	0.005	0.006	0.004	0.004
0.006	0.003	0.004	0.005	0.005	0.003	0.003
0.006	0.005	0.005	0.004	0.006	0.004	0.004
0.005	0.005	0.005	0.006	0.008	0.005	0.003
0.004	0.005	0.006	0.004	0.002	0.003	0.005
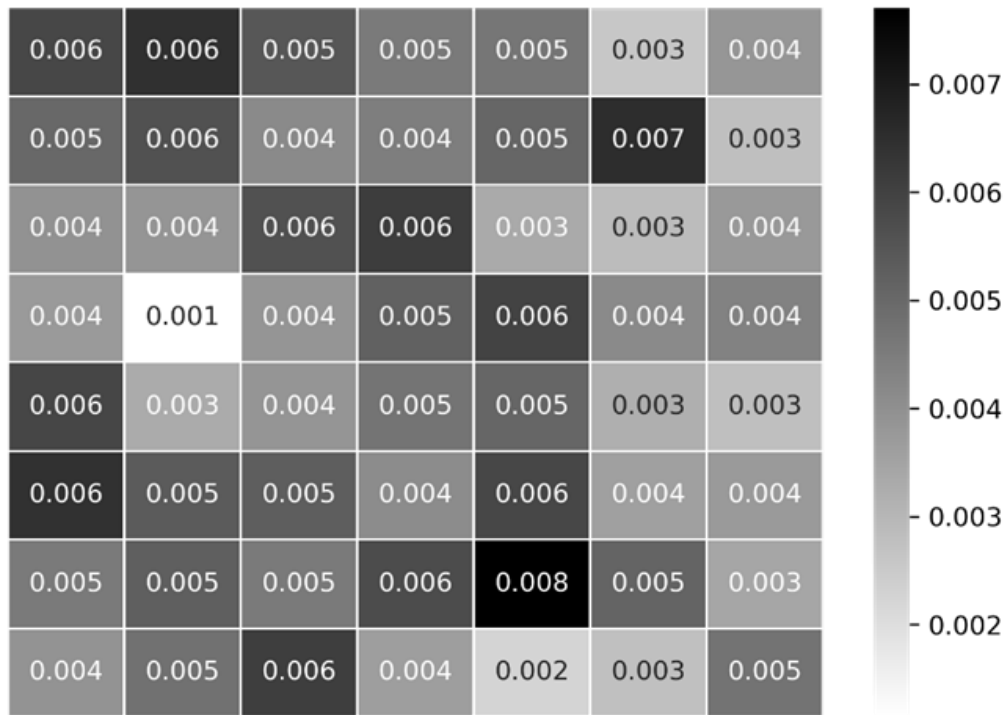
Figure 14. Slice-wise porosity distribution in the Barnett shale map representing variations over an area of 164.64 μm by 182.42 μm.

The pore-size and kerogen-size distributions are quantified with the help of a cluster function that calculates the number of pixels present in each cluster of a specified component. Each pixel covers an area of 100 nm^2. A multiscale pore size distribution was observed with 17% of the pore volume consisting of pore sizes less than 20 pixels (50 nm diameter), 68% of the pore volume made of pixel size between 20 and 500 pixels (50-250 nm diameter) and the remaining 15% pore volume consisting of pore sizes greater than

22

200 pixels. The micro-scale (500 pixels) pore sizes contribute equally to the total pore volume of the Barnett sample. The porosity reported here is primarily embedded in the organic matter.

Fig. 15 shows the Barnett map (a) and the distribution of the three different pore size scales (b: micro-scale pore size < 20 pixels, c: meso-scale pore size between 20 and 500 pixels, d: macro-scale pore size > 500 pixels). Number of macroscale pores is two order of magnitude lower than the number of microscale pores, while the number of mesoscale pores is half of the number of microscale pores.

Fig. 16 plots the distribution of the kerogen cluster sizes present in The Barnett map. Only the large clusters consisting of more than 10,000 pixels are considered for this plot. We observe presence of very large kerogen clusters with sizes more than 100,000 pixels. The map predominantly contains kerogen clusters of size ranging from 10,000 pixels to 20,000 pixels. In the cluster-size range from 100,000 to 200,000 pixels, there are on an average 1 to 2 clusters in the map. These clusters host the organic pores and are important in the calculation of TOC of the shale sample.
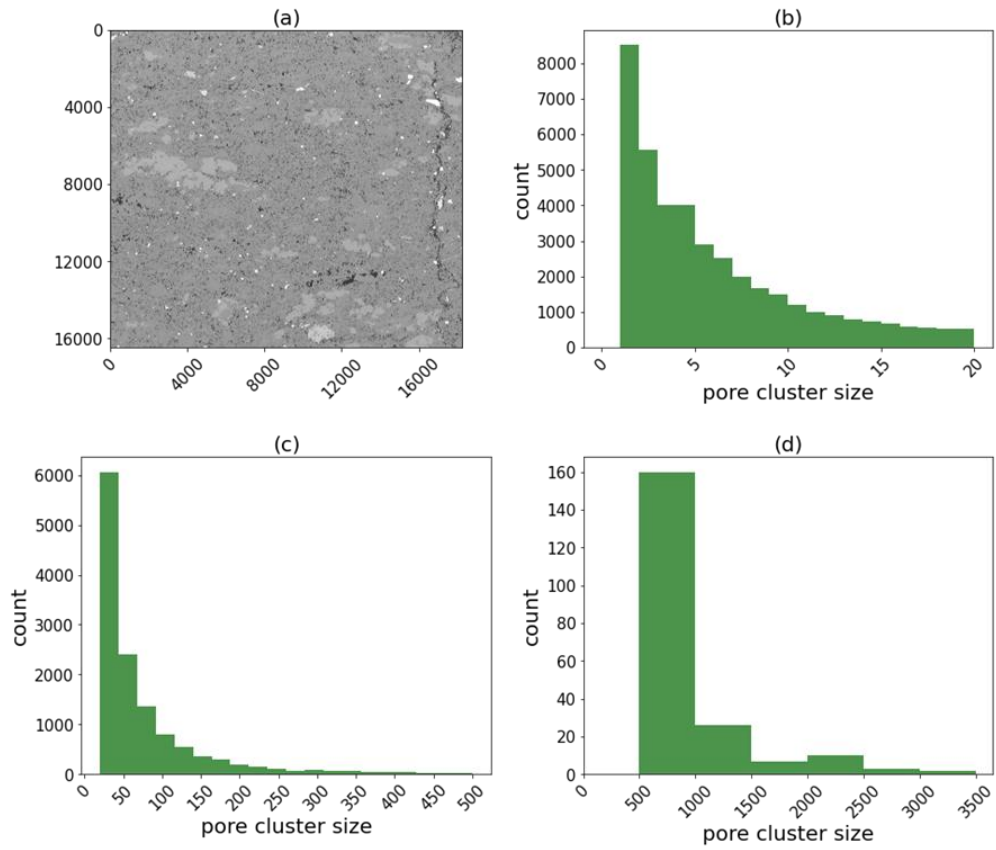
Figure 15. Multiscale pore size distribution of the Barnett shale sample.
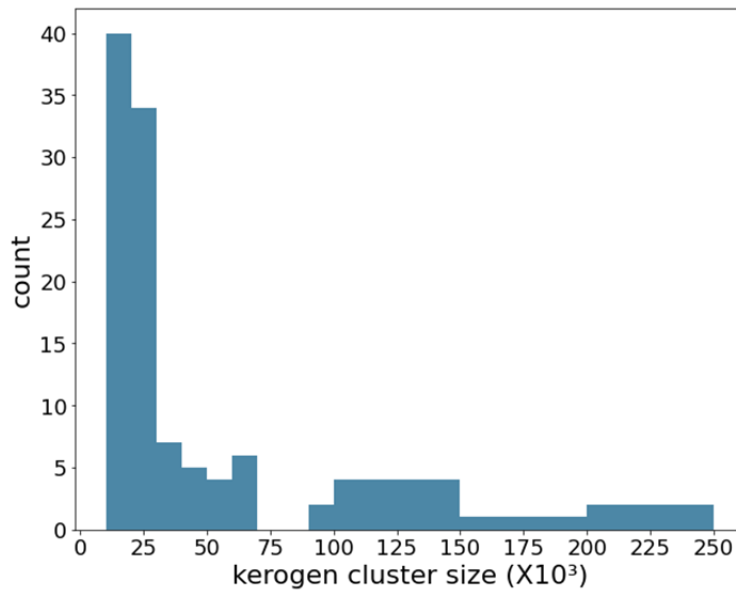
Figure 16. Organic/kerogen size distribution of Barnett shale sample.

Comparison of MLIS method with threshold-based segmentation and Fiji-based

segmentation methods

Here we compare the performance of the proposed machine learning workflow against traditional image segmentation methods, namely threshold-based segmentation and ML-assisted segmentation using the Fiji software. Threshold-based segmentation is a simple but popular method that classifies various components present in an image on the basis of pixel intensity ranging from 0 to 255 (**Fig. 17**). We consider the inner and outer region pixels of Slice 35 of the Barnett map to draw a comparison between the threshold-based method and the proposed MLIS model trained on sixteen different features. We observe a small difference in the F1 scores between the two cases (Table 1 left). However, this difference is much higher in case of outer region pixels (Table 1 right). With exposure to more features in the image, the classification of the interfaces, especially between pore/crack and organic/kerogen constituents, is considerably improved when using the proposed machine learning workflow.
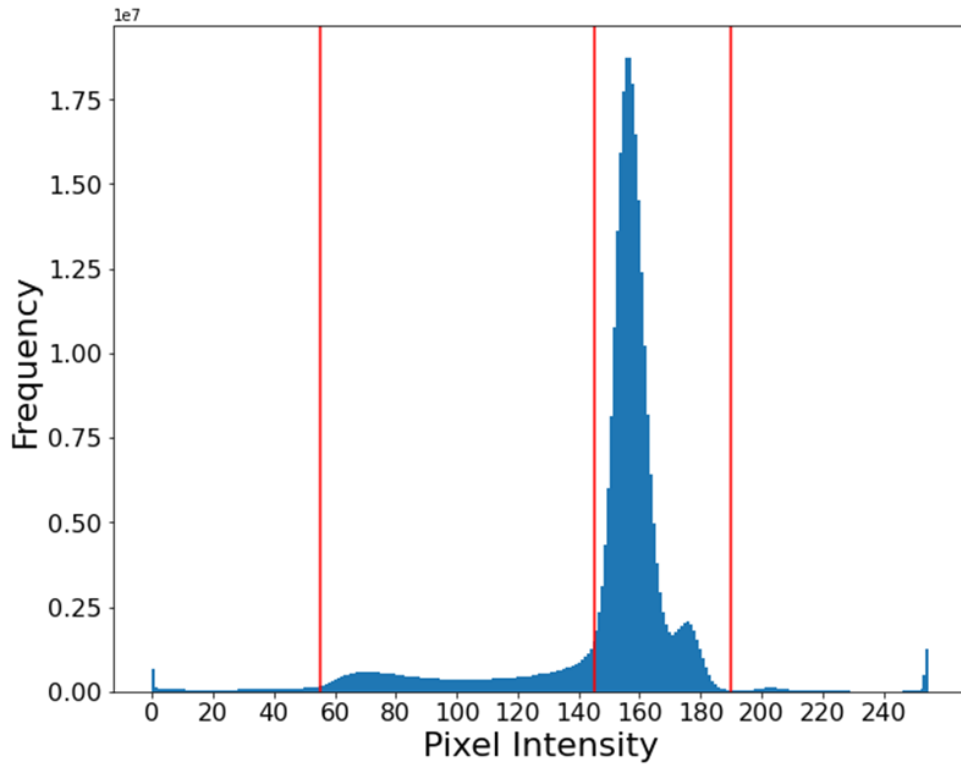
Figure 17. Intensity histogram of Slice 35 of the Barnett map showing the intensity thresholds for each component.

Components	Inner region						Outer region					
	Precision		Recall		F1 score		Precision		Recall		F1 score	
	I	II	I	II	I	II	I	II	I	II	I	II
Pore/ crack	0.96	1.00	1.00	0.97	0.98	0.99	0.29	1.00	0.31	0.88	0.30	0.93
Organic/ kerogen	0.88	0.99	0.99	1.00	0.93	1.00	0.28	0.91	0.29	0.98	0.28	0.95
Matrix	1.00	1.00	0.99	1.00	0.99	1.00	0.24	0.82	0.22	0.97	0.23	0.89
Pyrite	1.00	1.00	1.00	1.00	1.00	1.00	0.83	1.00	0.83	0.93	0.83	0.97

Table 1. Comparison of segmentation performance using threshold-based approach (case I) and machine learning assisted approach (case II) on the inner and outer region pixels of the Barnett map.

Components	Inner region						Outer region					
	Precision		Recall		F1 score		Precision		Recall		F1 score	
	I	II	I	II	I	II	I	II	I	II	I	II
Pore/ crack	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Organic/ kerogen	1.00	1.00	1.00	1.00	1.00	1.00	0.992	1.00	1.00	1.00	0.996	1.00
Matrix	1.00	1.00	1.00	1.00	1.00	1.00	0.994	1.00	0.984	1.00	0.989	1.00
Pyrite	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.993	1.00	0.996	1.00

Table 2. Comparison of performance using Fiji-based segmentation (case I) and proposed MLIS model (case II) on the inner and outer region pixels of the Barnett map.

We also compared the performance of the proposed MLIS workflow with the automated segmentation results from the plugin WEKA (Waikato Environment for Knowledge Analysis) of the open source image processing software Fiji (Table 2). Both the methods have been trained on the same training pixels from the Wolfcamp and Barnett map. Random Forest classifier have been used in both cases keeping the hyper-parameters same. The classifier is trained on the same sixteen set of features. Both models have been tested on the inner and outer region pixels of Slice 10 of The Barnett map. From the inner region, around 1900, 10000, 19000 and 1700 support pixels were chosen for pore/crack, organic/kerogen, matrix and pyrite components, respectively. For those compoenents, support of around 650, 1100, 450 and 400 pixels were selected from the outer region. It

was found that both the methods result in a perfect F1 score of 1.00 with almost zero rate of misclassification for the inner and outer region pixels for the test image. This establishes the robustness of the machine-learning-assisted segmentation methods for accurate classification of the various components in the shale maps.

Effect of test data size on ML-based segmentation

The selection of optimal number of support (or samples) for testing a machine learning model is always associated with a trade-off between selecting large number of support to ensure model robustness and the selecting less support due to imbalance in the dataset. Large number of support from different components for testing enhances the similarity between the probability distribution of features in both training and testing datasets and improves the performance evaluation metrics. However, in imbalanced datasets such as the shale SEM maps, pore/crack and organic/kerogen components that have a lower volume fraction compared to matrix and pyrites, are hard to select resulting in an ununiform distribution of components in the test dataset. Imbalanced testing dataset penalizes the component having an order of magnitude lower number of samples as compared to other components. This is why, it is important to use F1 score weighted by the number of support samples to effectively evaluate the test performance of a classifier. We present a comparison of the performance of Model 2 when tested on small vs. larger number of support pixels (Table 3) of the Barnett map.

In our dataset, we have limited number of pixels for pore/crack. As we increase the support of pore/crack component from 1447 (Table 3: Case I) to 4656 pixels (Table 3: Case II), we observe an decrease in the recall of the pore/crack constituent from 1.00 to

28

0.88. Recall measures the number of times a constituent is being correctly identified by the model. A drop in the recall suggests that Case I support pixels were selected from easily identifiable regions of pore/crack constituent whereas the newly introduced support pixels were selected from regions where the model failed to correctly identify the pore/crack component.

Components	Precision		Recall		F1 score		Support	
Cases	I	II	I	II	I	II	I	II
Pore/crack	0.44	1.00	1.00	0.88	0.61	0.93	1447	4656
Organic/kerogen	1.00	0.97	0.65	1.00	0.79	0.98	5263	23584
Matrix	1.00	1.00	1.00	1.00	1.00	1.00	3701	13835
Pyrite	1.00	1.00	1.00	1.00	1.00	1.00	2396	5154
							Tot:12807	Tot:47229

Table 3. Comparison of segmentation performance of the proposed MLIS method when applied on limited number of support pixels (Case I) against when tested on large number of support pixels (Case II) from slices 15, 26 and 28 of the Barnett map.

In the case of organic/kerogen component, we increase the support from 5263 (Table 3: Case I) to 23584 (Table 3: Case II) pixels. This causes the recall of this component to increase from 0.65 to 1.00 coupled with an increase in the precision of pore/crack component from 0.44 to 1.00. This proves a considerable improvement in the segmentation accuracy of the organic/kerogen component. In Case I, a significant part of the pixels chosen for the test set were being wrongly identified as pore/crack by the model. An increase in the recall of organic/kerogen as well precision of pore/crack proves that the newly introduced support pixels of the organic/kerogen component are being correctly

identified by the classifier. Taking into account the above inferences, it can be said that the Random Forest classifier performs slightly better on the organic/kerogen constituent than on pore/crack. Therefore, increasing the number of support pixels is advantageous to understand the performance of a model on a category (in this case, component), but it is also subjected to the availability and distribution of that category in the data.

Conclusion

Machine learning (ML) can identify pores and cracks representing the hydrocarbon storage and transport pathways as well as kerogen in the scanning electron microscopy (SEM) images of organic-rich shale samples from two shale formations, namely Wolfcamp and Barnett shales. The two shale formations and the corresponding SEM maps differ in topology and distribution of pores, cracks, and kerogen. In our study, pixel intensity, Gaussian Blur and local pixel information (minimum, maximum and mean) are the most important feature out of the 16 features for identifying pores, cracks and kerogen.

Machine learning assisted image segmentation (MLIS) workflow when trained and tested on the same formation exhibits 99% accuracy on inner region pixels and more than 80% accuracy on outer regions. The proposed machine learning workflow do not accurately and reliably identify pores, cracks and kerogen when it is trained on SEM images from one formation and then applied on images from different formation. Model trained on Wolfcamp shale can robustly detect matrix and pyrite in Barnett shale, but performs poorly when identifying the pores, cracks and kerogen. This is primarily due to identification of kerogen in the boundary regions as pores. Moreover, outer region matrix pixels generally get identified as pores, cracks or kerogen. In contrast, the model trained

on Barnett shale can robustly detect pyrite in Wolfcamp shale but exhibits poor performance for the remaining constituents; for example, fails to detect cracks in Barnett shale. This is due to pores and cracks getting wrongly identified as kerogen or matrix; in addition, kerogen is wrongly identified as matrix. Machine learning workflow trained on both the shale formations exhibits the best performance with an average F1 score of 0.99 and 0.91 on the inner-region and outer-region pixels, respectively.

The machine learning assisted identification of pores shows that the porosity ranges from 0.1 % to 0.8% in the 164.64 µm by 182.42 µm region of the Barnett sample. Extremely high porosity and extremely low porosity regions in the Barnett sample are localized. Further, the machine learning assisted identification of kerogen shows that the micro-scale (500 pixels) pore sizes contribute equally to the total pore volume of the Barnett shale sample. The macro-scale porosity of the system is primarily made of the organic pores surrounded by kerogen. Barnett shale sample predominantly contains kerogen clusters of size ranging from 10,000 pixels to 20,000 pixels.

Following scenarios improve the robustness or generalization capability of machine learning assisted identification of pores, cracks and kerogen: (1) large sizes of training and testing datasets sampled from different formations, (2) hyperparameter tuning, (3) use of random forest classifier and unscaled features, (4) simple feature extraction by considering the variations in intensities of neighboring pixels, (5) feature ranking and selection to reduce the dimensionality, and (6) creating a balanced testing dataset and use of weighted F1 score to better evaluate the performance.

References

31

1. Brahim Ait Skourt, Abdelhamid El Hassani, Aicha Majda, Lung CT Image Segmentation Using Deep Neural Networks, Procedia Computer Science, Volume 127, 2018, Pages 109-113, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.01.104.

2. Peña, José M., P. Gutiérrez, C. Hervás-Martínez, J. Six, R. Plant and F. López-Granados. Object-Based Image Classification of Summer Crops with Machine Learning Methods. *Remote. Sens.* 6 (2014): 5019-5041. https://doi.org/10.3390/rs6065019

3. Steven C. H. Hoi, Rong Jin, Jianke Zhu, and Michael R. Lyu. 2006. Batch mode active learning and its application to medical image classification. In Proceedings of the 23rd international conference on Machine learning *(ICML '06). Association for Computing Machinery, New York, NY, USA, 417–424. DOI:* https://doi.org/10.1145/1143844.1143897

4. S. Khan, M. H. Javed, E. Ahmed, S. A. A. Shah and S. U. Ali, Facial Recognition using Convolutional Neural Networks and Implementation on Smart Glasses, *2019 International Conference on Information Science and Communication Technology (ICISCT)*, Karachi, Pakistan, 2019, pp. 1-6, doi: 10.1109/CISCT.2019.8777442.

5. Laura Bandura, Adam D. Halpert, and Zhao Zhang, (2018), "Machine learning in the interpreter's toolbox: Unsupervised, supervised, and deep-learning applications," *SEG Technical Program Expanded Abstracts*: 4633-4637. https://doi.org/10.1190/segam2018-2997015.1

6. Al-Farisi, Omar , Zhang, Hongtao , Raza, Aikifa , Ozzane, Djamel , Sassi, Mohamed , and TieJun Zhang. "Machine Learning for 3D Image Recognition to Determine Porosity and Lithology of Heterogeneous Carbonate Rock." Paper presented at the SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi, UAE, September 2019. doi: https://doi.org/10.2118/196657-MS

7. Jobe, T.D., Vital-Brazil, E., and M. Khait. "Geological Feature Prediction Using Image-Based Machine Learning." *Petrophysics* 59 (2018): 750–760. doi: https://doi.org/10.30632/PJV59N6-2018a1

8. Siddharth Misra, Yaokun Wu, Chapter 10 - Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking, Editor(s): Siddharth Misra, Hao Li, Jiabo He, Machine Learning for Subsurface Characterization, Gulf Professional Publishing, 2020, Pages 289-314, ISBN 9780128177365, https://doi.org/10.1016/B978-0-12-817736-5.00010-7.

9. Yaokun Wu, Siddharth Misra, Carl Sondergeld, Mark Curtis, Jeremy Jernigen, Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales, Fuel, Volume 253, 2019, Pages 662-676, ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2019.05.017.

10. Siddharth Misra, Eliza Ganguly, Yaokun Wu, Chapter 11 - Generalization of machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales, Editor(s): Siddharth Misra, Hao Li, Jiabo He, Machine

Learning for Subsurface Characterization, Gulf Professional Publishing, 2020, Pages 315-338, ISBN 9780128177365, https://doi.org/10.1016/B978-0-12-817736-5.00011-9.

11. Ganguly, Eliza, Misra, Siddharth, and Yaokun Wu. "Generalizable Data-Driven Techniques for Microstructural Analysis of Shales." Paper presented at the SPE Annual Technical Conference and Exhibition, Virtual, October 2020. doi: https://doi.org/10.2118/201554-MS

CHAPTER III

STATISTICAL APPROACH TO QUANTIFY THREE-DIMENSIONAL FLUID

CONNECTIVITY FROM MICRO-CT IMAGES OF CORE UNDERGOING WATER

ALTERNATING GAS (WAG) INJECTION

Introduction

Literature review

The three-dimensional connectivity of the fluid phases in porous media plays a crucial role in governing the fluid transport, displacement, and recovery. Water Alternating Gas (WAG) is a well-established technique that leverages the benefits of Gas Injection and Waterflooding processes to improve the ultimate oil recovery. Accurate three-dimensional quantification of the fluid phase connectivity at each stage of Water Alternating Gas (WAG) injection will lead to better understanding of the efficacy and efficiency of the fluid injection strategies/procedures.

The term 'connectivity' is studied differently in multiple domains of study. In neuroscience, the connection between different regions in the brain is termed as connectivity. The structural connectivity of the brain is measured from diffusion MRI, a non-invasive technique that can characterize the brain tissue [1], On the other hand, in geoscience, connectivity is closely related to permeability of a reservoir [2]. Transport in porous materials in the multiphase regime are governed by the three-dimensional connectivity of the different fluid phases present. Quantification of this parameter and its spatiotemporal variation in the porous material can be used to gain an improved

understanding of the bulk displacement, trapping and recovery mechanisms of the fluid phases.

Efforts have been made to develop metrics for calculating the particle connectivity in multibody systems. General n-point probability functions have been developed to characterize microstructural connectivity in multiphase random media [3]. For a system of suspension of spheres in a uniform matrix, n-point probability function is denoted by Sn. Further study has been conducted to obtain an analytical expression for 2-point probability function to quantify the distribution of equi-sized rods and discs of multiple densities present in a matrix phase [4]. Various morphological descriptors and their limitations have been reviewed, for example, indicator function, cluster function, surface correlation function, nearest neighbor function and pore size distribution function [5]. Conventional soil science typically uses pore size and pore size distribution to characterize soil microstructure. But these metrics can only provide limited information in case of complex structures. Statistical functions like two-point probability function, two-point cluster function and linear function are used as alternate approaches to evaluate connectivity from complex systems [6]. We also come across other applications of local and global static descriptors like investigating the distribution of permeability and connectivity in reservoir models [7]. Flow dynamics are interpreted from several static measures like characteristic path lengths connecting locations in a reservoir to the closest well, distance properties like minimum and differential path lengths and continuity path lines extracted from these distance properties. An alternative to standard geostatistical methods is presented by Western et al. (2001) to evaluate connected features in soil

moisture patterns. The work shows application of probabilistic and computational functions capable of differentiating between connected and disconnected patterns [8].

In this study, we aim at developing higher-order statistical functions to study the spatio-temporal variation in the three-dimensional connectivity of fluid phases during a Water-Alternating-Gas (WAG) process in a porous material. In literature, efforts have been made towards studying the relative permeability for WAG injection in oil reservoirs [9]. Mathematical models have been developed and implemented to examine capillary pressure and hysteresis effects during WAG [10]. However, very limited research have been conducted to quantitatively investigate the connectivity evolution in a three-dimensional system undergoing multi-stage injection. We develop and apply several connectivity metrics on the µCT scans of a water-wet Ketton rock sample. We first present the details of the WAG experiment conducted by scanziani et al. (2018) and describe the different metrics that were tested, their advantages and limitations. We then explain the working of the metrics that proved to the most efficient in capturing the change in three-dimensional fluid phase connectivity after each injection stage. The results obtained from the applied statistical metrics are reported and the observed pore-scale phenomena are explained in terms of the connectivity values and the wettability of the system.

Scientific questions

This study presents a novel method of quantifying three-dimensional connectivity of fluids in a sample undergoing WAG injection. Some of the crucial scientific questions that will be answered in this chapter are as follows:

- How are the distribution and connectivity of the wetting and non-wetting fluid phases changing from one injection stage to another in a water alternating gas operation?

- Which injection sequence is contributing the most to recover the oil phase from the sample?

- In which sections of the sample does maximum recovery and maximum trapping of oil phase occurs?

- How is the connectivity of a certain phase related to the saturation of the phase?

- How can two-point statistics and fast marching algorithm be used to compute pixel connectivity from images?

Method

Description of sample

In this study, the Water-Alternating-Gas experiment is conducted by Scaniziani et al. (2018) [11] on water-wet Ketton carbonate rock sample of diameter 4.9 mm and length 19.5 mm. Porosity of the sample is reported to be 29.8%. The dataset consists of 3050 µCT scans for each of the injection stages, each scan containing 1030×1033 pixels and representing a dimension of 5µm×5µm. the scans are numbered from 0 to 3049. The scans of each stage are divided into 3 parts: bottom (index 0—999), middle (1000—1999) and top (index 2000—3049).

Water-Alternating-Gas (WAG) experiment

Scanziani et al. (2018) performed a 3-stage WAG experiment on the Ketton rock sample. WAG is an enhanced oil recovery (EOR) technique in which water and gas injections are carried out alternately to improve the sweep efficiency. The three stages in this work are referred to as Water Injection #1 (WF1), Gas Injection (GI) and Water Injection #2 (WF2). The pre-injection water and oil saturation in the sample was reported to 37.6% and 62.4% respectively. The injection fluid volume and flow rate are presented in Table 4. Water is injected from the bottom of the sample (index 0) while gas is injected from the top (index 3049).

Injection Stages	Stage Initials	Injected Fluid	Injected Pore Volume	Flow Rate (ml/min)
Water-flooding #1	WF1	H_2O +7%w KI	15	0.015
Gas injection	GI	N_2	5	0.015
Water-flooding #2	WF2	H_2O +7%w KI	5	0.015

Table 4. Pore volume and flow rate of injected fluid at each stage of WAG process.

Image pre-processing

The spatio-temporal evolution of three-dimensional connectivity of oil, water and gas phase is captured with the help of two novel quantification metrics. The image pre-processing can be divided into the following stages:

➢ The dataset of 3050 µCT scans, each with 1030×1033 pixels is converted to 305 image slices, each with 103×103 pixels using the SciPy Zoom functionality to reduce the time complexity associated with estimation of the three-dimensional fluid connectivity.

- Each downscaled scan is segmented into oil, water, rock and gas components using Machine Learning assisted Image Segmentation (MLIS) workflow explained in chapter 2.
- The segmented images are converted to binary images with the fluid phase of interest as white pixels.
- The developed and tested connectivity metrics are applied on the pre-processes images.

In **Fig. 18**, the pre-injection (PI) stage is dominated by presence of non-wetting phase (oil) in the larger pores and wetting phase (water) in the smaller pores. The saturation of oil drops at each stage of the WAG process. Significant displacement of oil from the large pores is observed after Gas Injection (GI) resulting in maximum recovery.
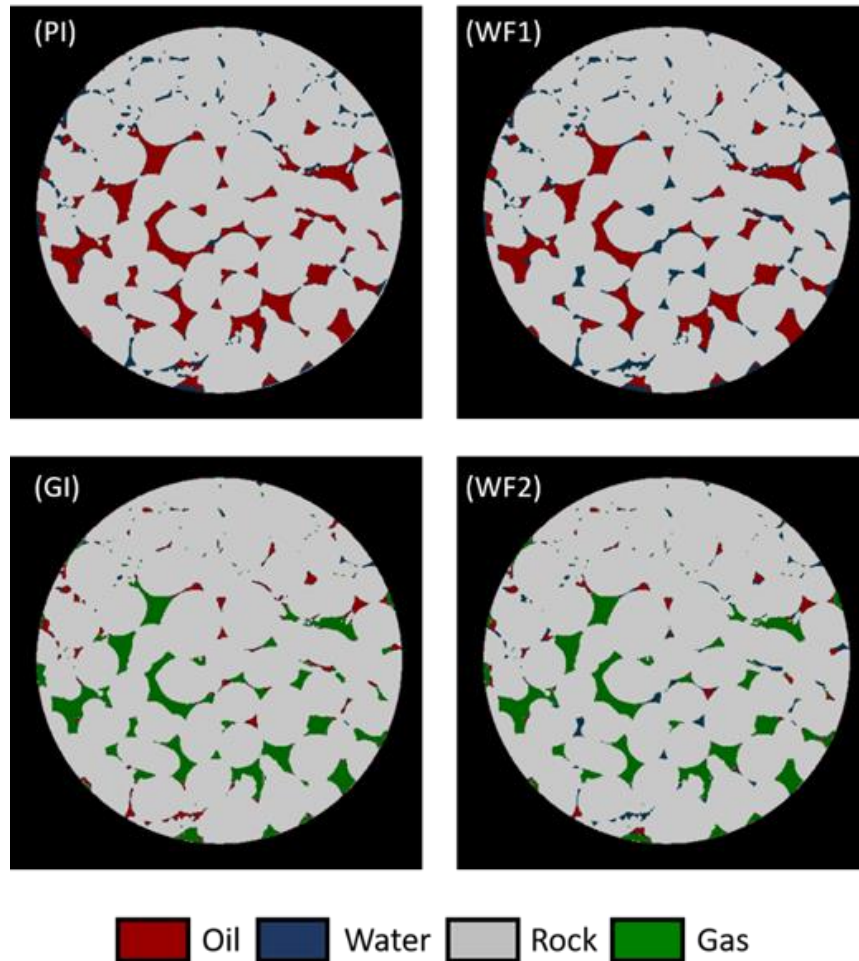
Figure 18. A segmented image slice of the μCT scan dataset showing the distribution of the different fluid phases after each injection stage (Reprinted from Ganguly and Misra 2021)

Connectivity metrics

Three approaches have been established to determine the pixel connectivity from a three-dimensional image, namely, face connectivity, edge connectivity and node connectivity. However, attempts to apply these approaches to quantify the pore-scale connectivity in materials, at a local and global scale, especially during the WAG injections are limited. We develop several mathematical and probabilistic metrics to capture the evolving fluid

41

phase connectivity in the pore space during multi-stage injection. Each metric was scrupulously tested and analyzed and the best performing metrics are applied to accurately capture the change in the three-dimensional connectivity of the fluid phases from one injection stage to another.

The probabilistic functions that are capable to quantifying connectivity, but with limitations, and didn't prove reliable for the current study are as follows:

1. Two-point probability function (S_2): This is defined as the probability that two pixels at a certain distance in an image belongs to the same phase (white or black) in the binary medium. It provides no information regarding the path between the two pixels or the possibility of the two pixels belonging to the same cluster.

2. Two-point cluster function (C_2): This function calculates the probability that two pixels of a particular phase (white or black) at a specific distance belong to the same pixel cluster. The quantity is measured as the ratio of number of pixel pairs belonging to the phase of interest (white) at a certain distance and the total possible pixel pairs (in white and black phase) in the image at the same distance from each other. For each image, this metric gives a distribution of the probability values against the separation distance between pixel pairs.

3. Euler number: Euler number have been implemented in attempt to capture connectivity of a porous medium [12]. It is calculated as the number of discrete objects forming the phase plus the number of enclosed cavities less the number of tunnels through those objects, per unit volume. A higher pixel connectivity of a

certain phase would correspond to higher number of tunnels through the phase, resulting in a lower Euler number. However, this correlation is valid when the volume fraction of the phase of interest does not change. In this study, the volume fraction of each of the oil, water and gas phases are subject to change from one image to another as injection and recovery takes place. Therefore, Euler number does not prove to be a reliable connectivity metric to quantify the changing 3-D connectivity of the fluid phases during the WAG process.

The statistical metrics that could be successfully applied to calculate the fluid phase connectivity from three-dimensional volumes of segmented images are as follows [13]:

➢ *Connectivity Function (CF):* This metric is a slightly modified form of the two-point cluster function (C_2). CF is defined as the probability of a pixel pair of a phase of interest (white) at a separation distance h belonging to the same cluster. Unlike C_2, $CF(h)$ is measured as the ratio of pixel pairs of a certain phase (white) at separation distance h belonging to the same cluster and the total number of pixel pairs of the same phase (white) at the same separation distance. The metric identifies the position of the clusters of pixels of the phase of interest present in an image and determines the Euclidean distance (separation distance) between all possible pixel pairs of the phase of interest. True pairs are the pixel pairs with the same cluster index while false pairs are the ones that have different cluster index. The probability at each separation distance, h, is computed as,

$$P(h) = \frac{true\ pairs}{true\ pairs + false\ pairs}$$

The connectivity function CF(h) is obtained from the area under the probability distribution curve over all possible separation distance (**Fig 19**). This converts the spectral parameter to a single-values scalar parameter denoting the average displacement between two pixels in an image. A higher connectivity corresponds to a higher value of CF.
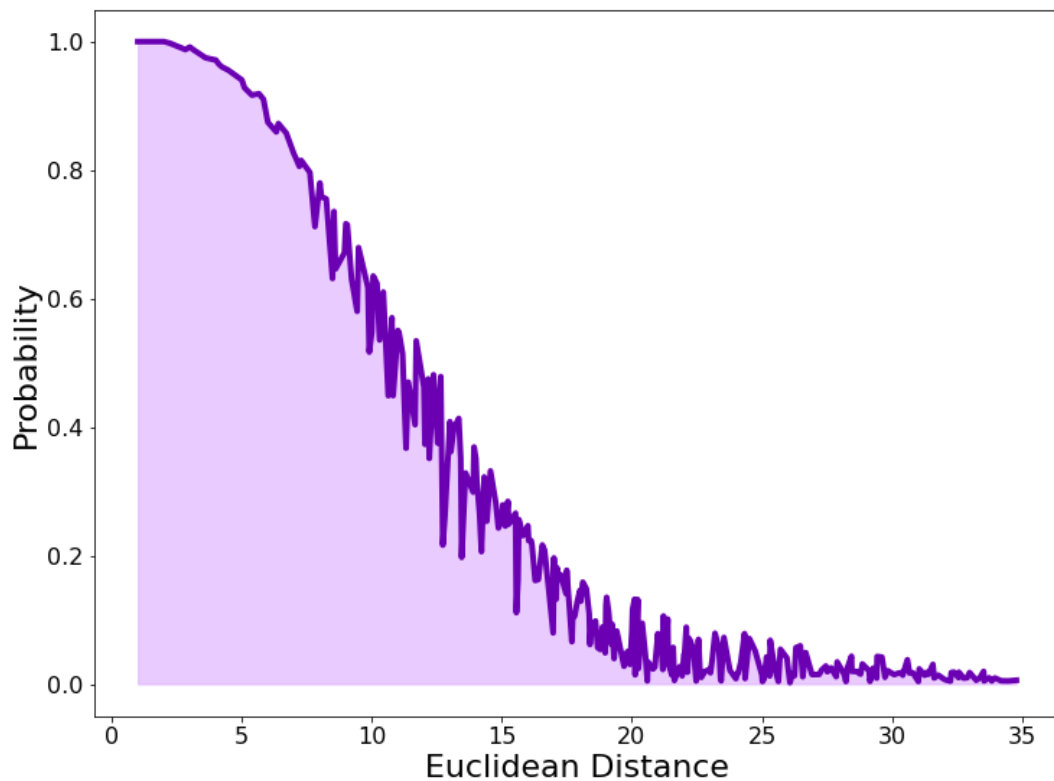

Figure 19. Probability distribution of the pixels belonging to the same cluster at different separation distance (h)

➢ *Fast Marching based Travel Time (FMTT):* The fast marching method models the evolution of a closed surface outward as a function of time at a constant speed. In this study, the algorithm is used to initiate several contours from

random pixels of the phase of interest in the image and the time taken by each

contour to cover all the pixels of that phase belonging to the same cluster is

recorded. The algorithm is designed such that the contour only propagates

through pixels of the phase of interest. The travel speed is pre-set to a high

value for the phase of interest and zero for the background pixels. The travel

time distribution for each contour is compiled into a single travel time

histogram by weighing each histogram using the size of the cluster. The mean

of this travel time distribution is termed as fast marching based travel time

(FMTT) and can be used as a direct indicator of pixel connectivity.

A small value of average connection distance indicates small isotropic, globular

clusters without any tortuosity between connected pixels. A large value of average

connection distance indicates large anisotropic, elongated pathways with high tortuosity

between connected pixels. We select mean instead of median of the travel time histogram

to have higher sensitivity to large clusters because large clusters predominantly contribute

to the overall connectivity. When the mean is very different from the median, it indicates

presence of few large tortuous clusters.

Result and discussions

Two-dimensional connectivity for wetting and non-wetting phases

Studies have been conducted to understand the effect of wettability, relative permeability

and hysteresis during WAG operations [14]. However, limited effort has been given to

describe the wetting and non-wetting phase connectivity in a porous material during such

injection stages. The metrics explained in the previous sections are rigorously tested on synthetic images and applied on each segmented 2-dimensional image slice of the μCt scan dataset. Several pore-scale phenomena such as displacement, trapping and recovery are then studied based on the change in these metric values across the length of the sample.

Multiphase fluid flow in a porous system is affected by different factors like porosity, pore size distribution, connectivity of the pores and fluid phases, capillary pressure and wettability of the system. We deploy the connectivity function and the fast marching based travel time to capture the spatial and temporal variation of fluid phase connectivity from one injection stage to another. The metrics are averaged over every 9 image slices (for a total of 3050 slices) starting from the bottom of the core to reduce the number of samples from 3050 to 339 to aid an easier visualization.

Fig. 20 plots the fast marching based travel time values of every 9 images against a distance marker denoting the distance of each set of image slices from the bottom of the core. It is evident that the oil phase connectivity drops at each injection stage. This is consistent with the saturations reported by Scanziani et al. (2018). Oil saturation drops from 52% to 30.5% after gas injection and to 18% after water-flooding #2. Similar to the saturation, oil phase connectivity also sees a significant drop after gas injection, thereby indicating highest oil recovery at that stage. The two-dimensional connectivity gives us a better resolution for purposes of visualization. Detailed analysis of the two-dimansional connectivity variation in the different stages can be found in Ganguly et al. (2021) [15].
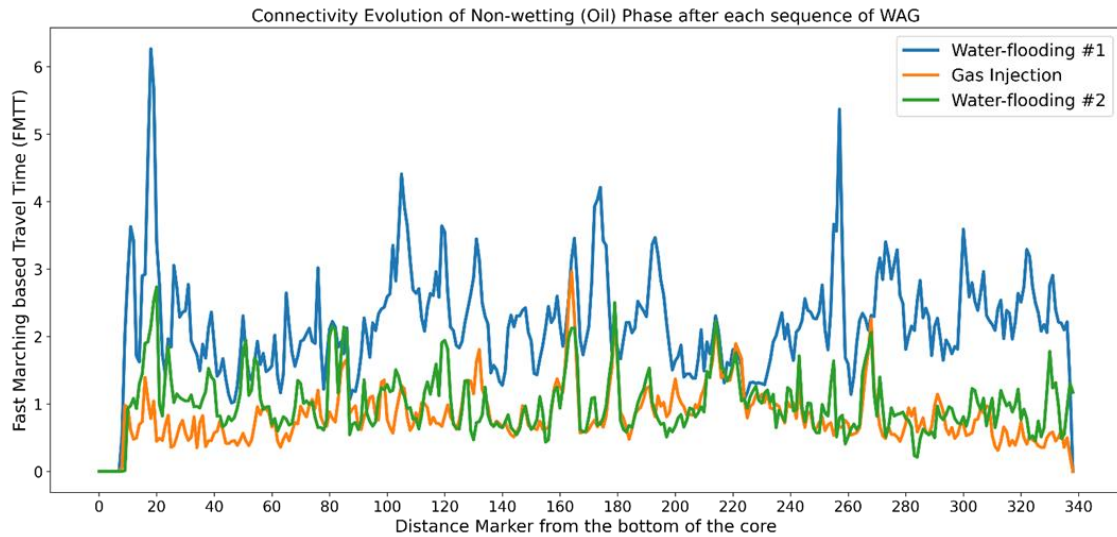
Figure 20. Stage-wise variation in the FMTT values of the oil phase along the length of the core sample.

Fig. 21 compares the fast marching based travel time of the water phase for the different WAG sequences. In a WAG operation, water and gas is injected alternately into the core. As expected, the connectivity of the water phase alternately increases and decreases at each stage. Similar to oil phase, maximum water is also displaced after the gas injection. The pore-scale variation of fluid connectivity along the different zones of the sample are discussed in details later in this document.
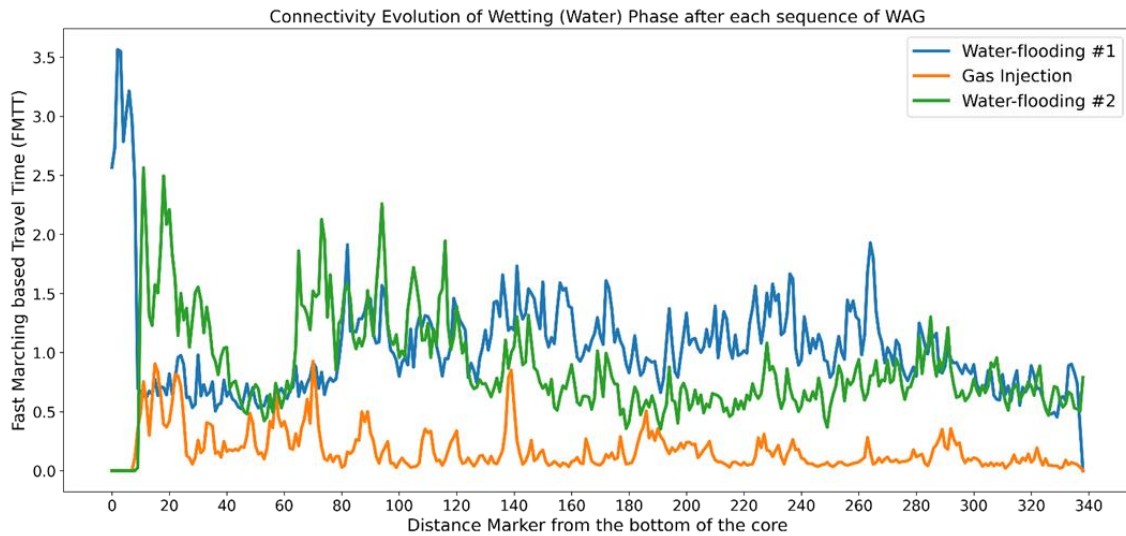
47

Figure 21. Stage-wise variation in the FMTT values of the water phase along the length of the core sample.

Relation between connectivity and saturation of fluid phases

Saturation of as fluid phase indicates the fraction of pore space pixels in an image belonging to that particular phase. These pixels may be present in small, medium or large clusters which may or may or be connected to each other. The saturation and connectivity of fluid pixels are found to be positively correlated only in the case of presence of large connected clusters. Low connectivity coupled with high saturation of a certain fluid phase indicates presence of multiple disconnected clusters. On the other hand, a fluid phase exhibiting high pixel connectivity and low saturation can be characterized by presence of small clusters connected by highly tortuous path.

Fig. 22 shows the oil phase connectivity and saturation for the three WAG sequences. Both the quantities are scaled using min-max scaler for purposes of visualization. As discussed earlier, oil is recovered at each stage of the WAG process,

thereby causing reduction in the saturation. We observe that, as saturation drops from water-flooding #1 to water-flooding#2, the correlation between the two quantities, indicated by the Spearman correlation coefficient, also drops from 0.701 to 0.637. A closer look also reveals that a clear positive correlation exists in sections of the sample with high saturation and connectivity values (large connected clusters) while the correlation is blurred when the value of either of the quantity decreases.
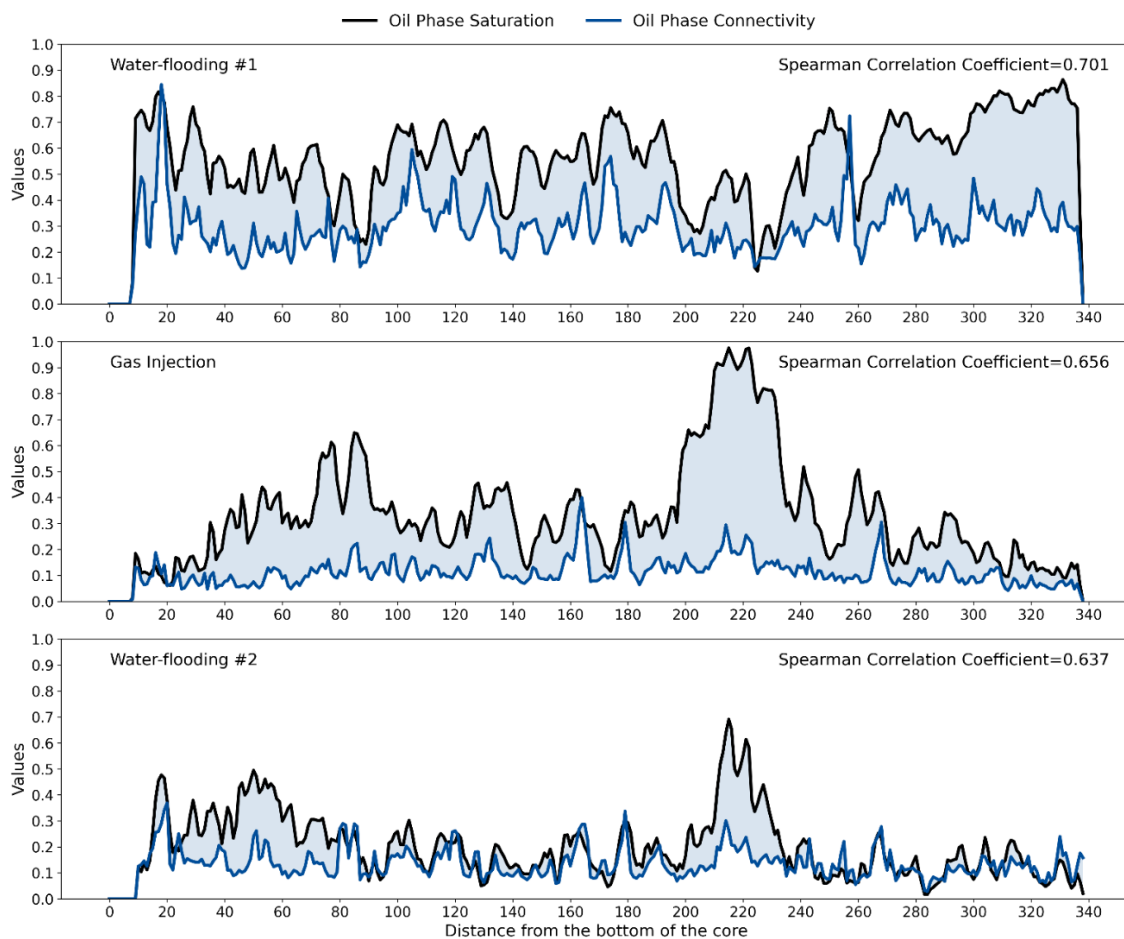

Figure 22. Relation between the oil phase saturation and connectivity for the three stages of the WAG process (Reprinted from Ganguly and Misra 2021)

Correlation between the two metrics: connectivity function (CF) and fast marching

based travel time (FMTT)

In this study, two metrics are developed to accurately quantify the evolution of wetting and non-wetting phase connectivity: Connectivity Function based on the separation distance (CF(h)) and Fast Marching based travel time (FMTT). The two metrics take different approaches to calculate the pixel connectedness of the fluid phase of interest in the segmented image. At a specific separation distance (h), the connectivity function computes the probability that two randomly selected pixels belong to the same cluster of the same phase. On the other hand, Fast Marching method generates the frequency of the travel time taken by a wavefront to travel between two pixels in the same cluster of the same phase. FMTT can be described as the mean of the travel time histogram of all clusters present in an image.

Compared to CF(h), FMTT exhibits higher sensitivity towards tortuosity of the path connecting the two pixels. In **Fig. 23**, the image (a) shows the pixels at position (0,0) and (7,7) connected by a straight path. This image has a CF value of 8.89 and an FMTT value of 1.72. In image (b), the same pixels are connected by a relatively more tortuous path. While the CF value of image (b) remains the same, the FMTT value is found to be 3.34, i.e, an almost two-fold increase from the previous value. For images with highly tortuous connection paths, connectivity function (CF(h)) proves to be a more robust metric. Also, due to random selection of source points, the value of FMTT is prone to change from one iteration to other. Therefore, two images having FMTT values in the range of +/- 0.5 can be approximated to have similar connectivity.
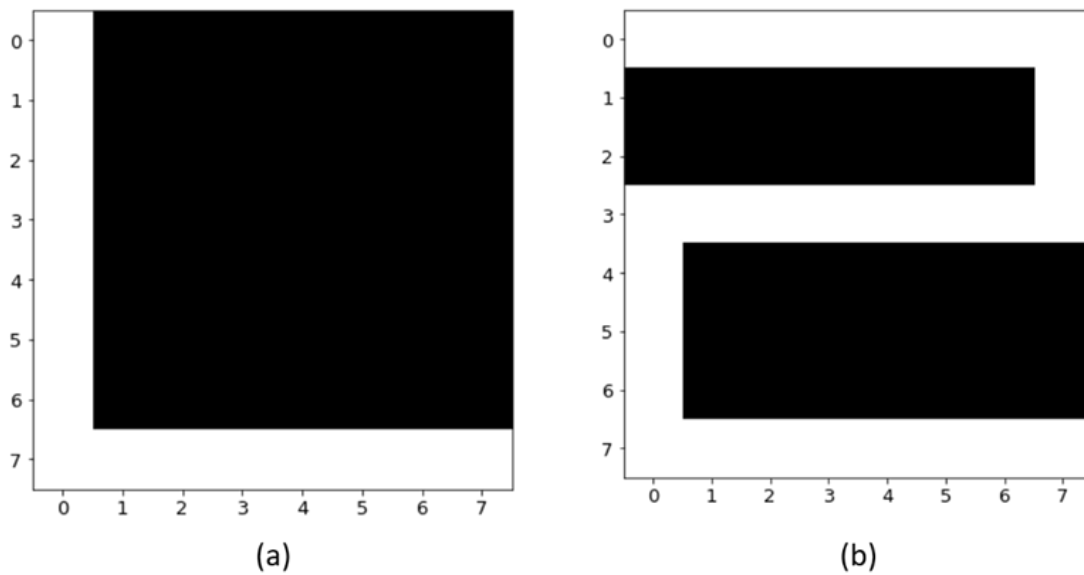
Figure 23. 8×8 binary image showing pixels at positions (0,0) and (7,7) connected by a low tortuosity path (a) and a high tortuosity path (b) (Reprinted from Ganguly and Misra 2021)

Fig. 24 plots the two connectivity metrics for the oil phase in the pre-injection stage. The metric values have been scaled using Min-Max scaling method. It is evident that the two metrics follow the same trend and shows a strong positive correlation.
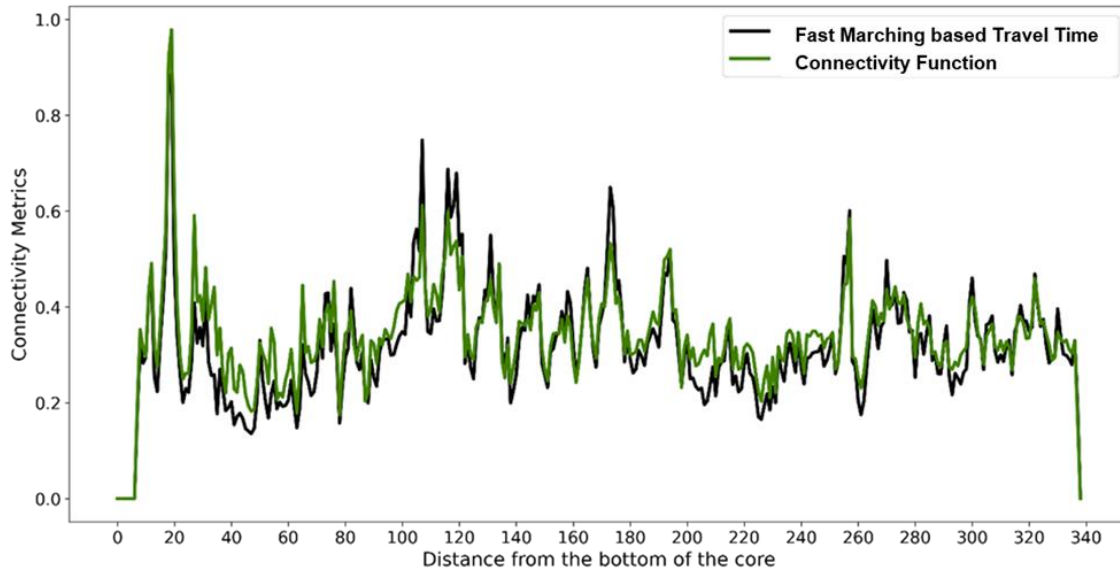
Figure 24. Comparison of the FMTT and Connectivity Function (both scaled using min-max scaling method) for pre-injection oil phase. Correlation between the metrics > 0.8 establishing the reliability of the methods used (Reprinted from Ganguly and Misra 2021)

The correlation between the two proposed connectivity metrics is quantified using the Spearman correlation coefficient for the pre-injection and all other stages of the WAG operation for the oil and water phase data. The Spearman rank-order coefficient is a statistical measure of the monotonic relationship between two datasets. The coefficient values range between -1 and +1 with 0 implying no correlation. For each injection stage, we find a correlation coefficient greater than 0.8 (Table 5) for the two metric values for both phases, indicating that the metrics do not encounter highly tortuous connected pathways in this dataset. This also establishes the reliability of the two proposed metrics in capturing the change in fluid phase connectivity in the various WAG cycles.

Spearman Correlation Coefficients	Oil Phase	Water Phase
Pre-injection (PI)	0.89	0.870
Water-flooding #1 (WF1)	0.876	0.905
Gas Injection (GI)	0.872	0.884
Water-flooding #2 (WF2)	0.904	0.908

Table 5. Spearman Correlation Coefficient of the two metrics for oil and water phases for each injection sequence of the WAG process.

Evolution of three-dimensional connectivity at each injection stage

In this section, we divide the dataset into three-dimensional volumes and study the change in 3-D fluid phase connectivity across the length of the sample. We tested six different resolutions for the z-axis to create the 3-D volume as follows: 1, 5, 15, 25, 35 and 45 image slices. A resolution of 1 image slice at a time corresponds to the 2-D connectivity while the connectivity values converge at a higher resolution (**Fig. 25**). Same can be observed for the volume fraction of the fluid phases (**Fig. 26**). An optimum resolution of 35 slices is used for the three-dimensional connectivity quantification.

We use the SciPy zoom functionality to convert the dataset of 3050 images to 305 images. The connectivity is measured considering 35 slices at a time and moving one slice at a time across the length of the sample (305 slices). Therefore, we obtain (305-35+1=) 271 datapoints for connectivity of the whole-core sample.

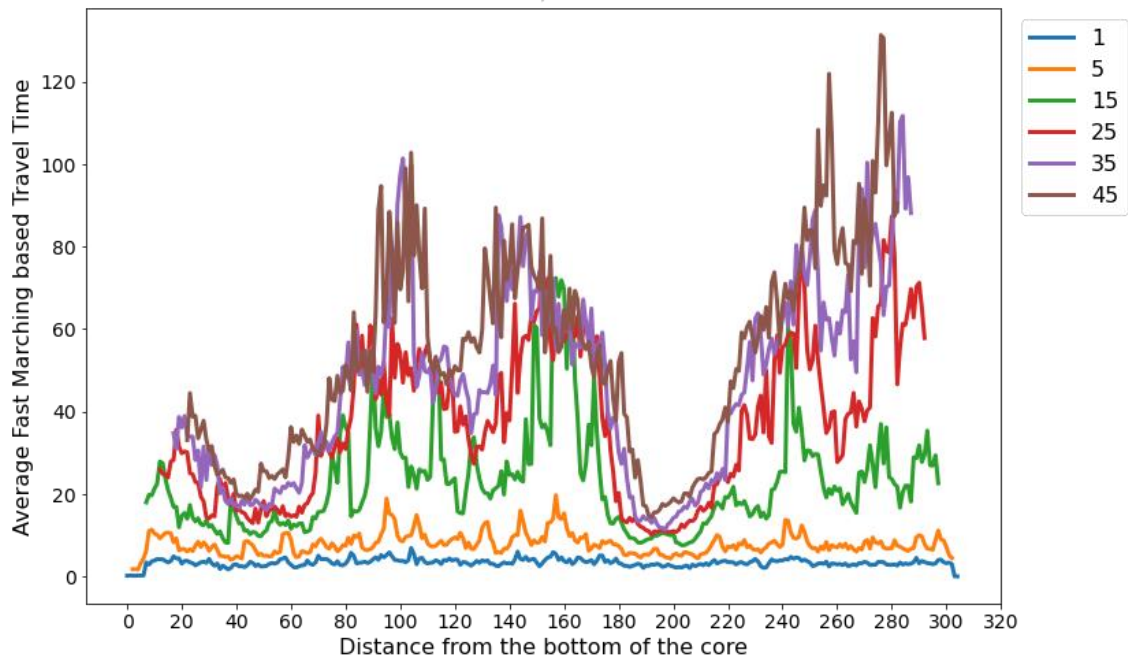**Figure 25. Three-dimensional oil phase connectivity for six different resolutions along the length of the whole-core sample.**
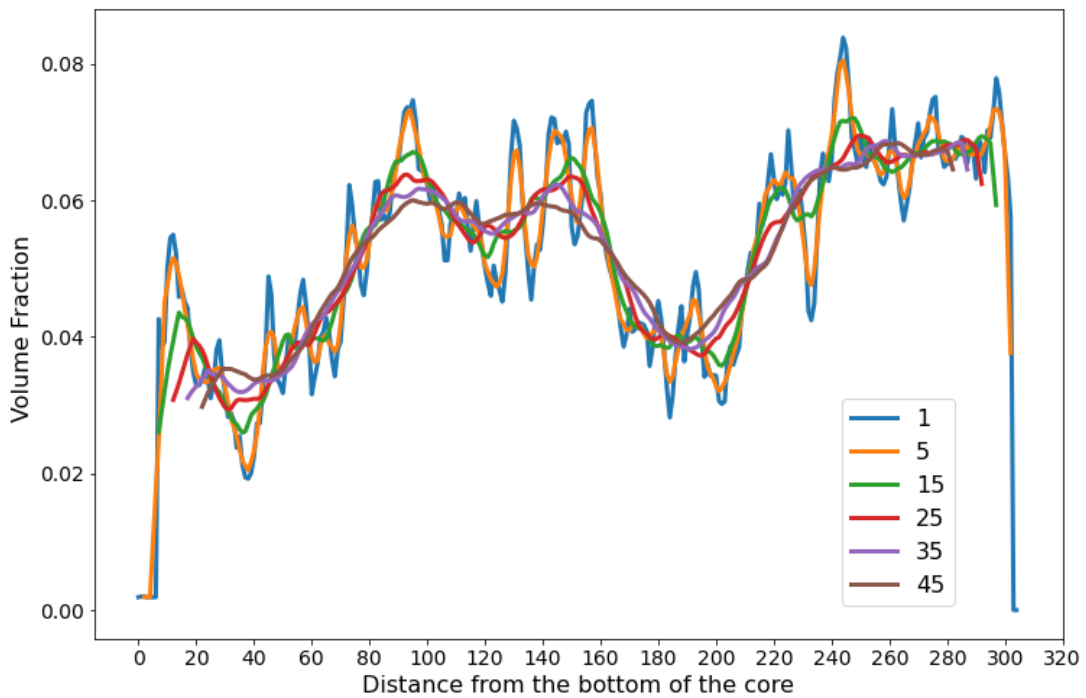

Figure 26. Three-dimensional volume fraction for six different resolutions along the length of the whole-core sample.

Pre-injection stage (PI)

In this study, we aim to characterize the spatial changes in pore-scale connectivity, saturation and displacement of oil, water and gas during the various WAG cycles in terms of pore size and wettability of the rock. We draw a comparison of the fast marching based travel time (FMTT) values of the three different fluid phases at each injection stage with the preceding stage to build a clear picture of how the different fluid clusters evolve at each stage, the sections of the sample that undergo significant displacement and the stages that aid maximum recovery. The sample is this study is a water-wet carbonate core. In the pre-injection stage, the sample consists of wetting (water) and non-wetting (oil) phases with a reported saturation of 37.6% and 62.4% respectively (Scanziani et al., 2018). This is because oil (non-wetting phase) preferably occupies the larger pores while water occupies the medium and smaller sized pores in a water-wet system.

Water-flooding #1

Figs. 27 and 10 shows the evolution of the three-dimensional connectivity (in terms of FMT) of oil and water phase respectively from the pre-injection to the first water-flooding stage. Water is injected from the bottom of the sample in the WF1 stage. The initial distribution of oil and water phases is disrupted as water, being the wetting phase takes up the smaller and medium sized pores. We observe three sections of the sample exhibiting considerable increase in the water phase connectivity due to water-flooding #1: section around distance marker 75, 125 and 200 (**Fig. 28**). An investigation of the oil phase connectivity in the same sections (Fig. 27) reveals that the injected water phase displaces the oil clusters residing in the zone 75 and 125 thereby causing the oil connectivity to

drop. The oil phase connectivity around marker 200 remains low, indicating that the injected water coalesces the previously disconnected water clusters, resulting in an increase in the connectivity.
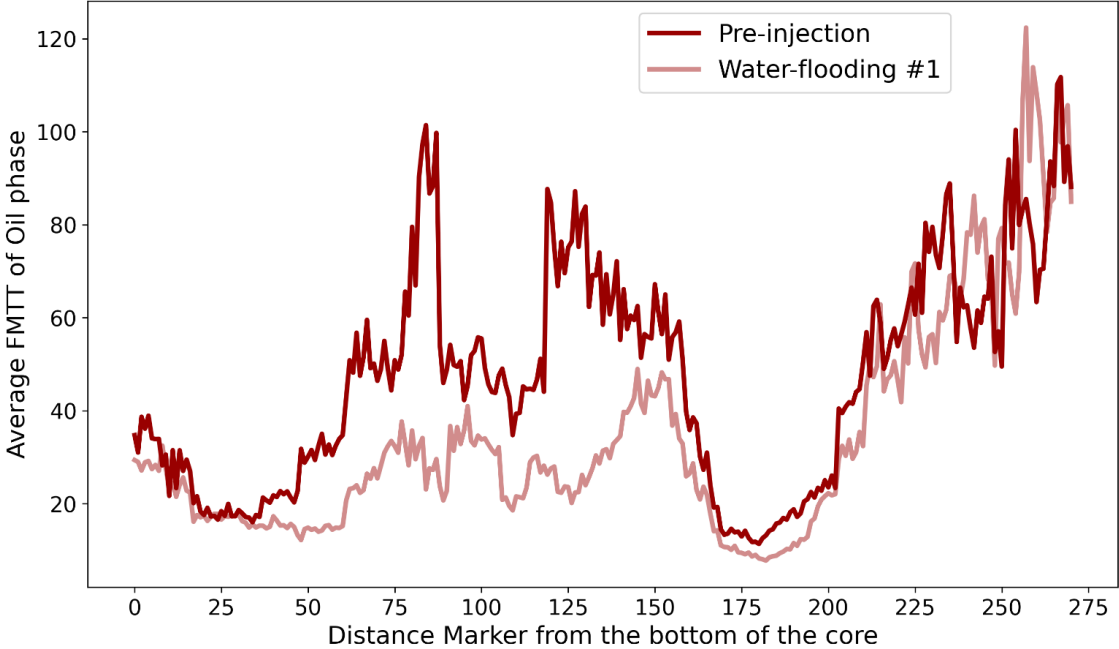

Figure 27. Change in the pre-injection stage 3-D connectivity of oil phase due to the first water-flooding.
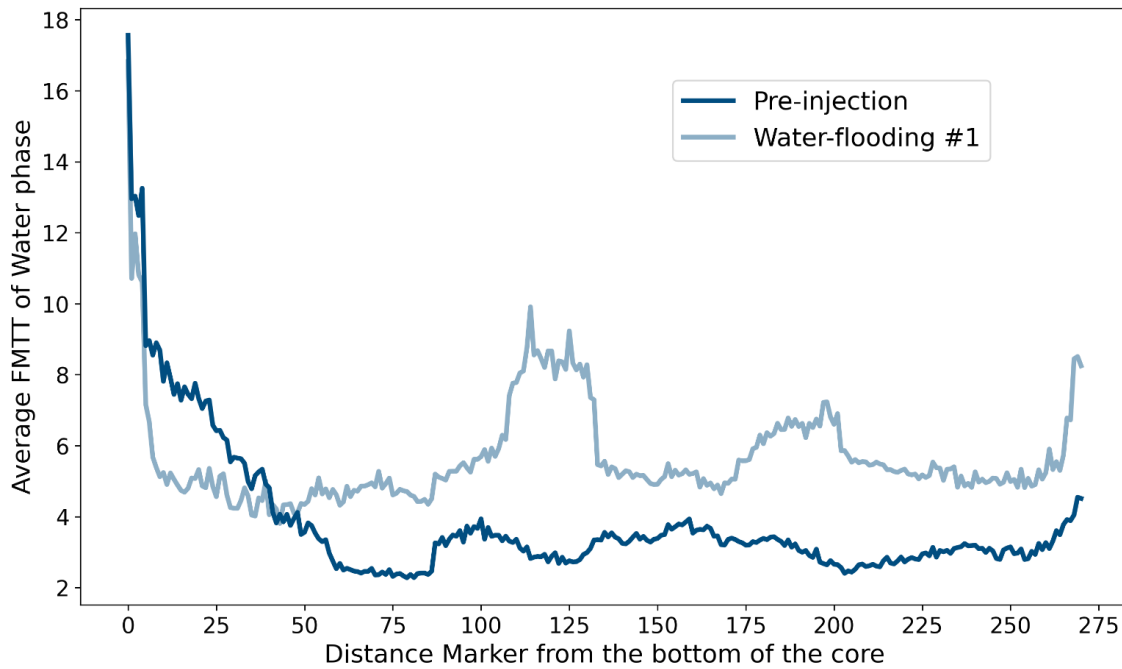
Figure 28. Change in the pre-injection stage 3-D connectivity of water phase due to the first water-flooding.

Another interesting observation is that the oil phase travel time does not significantly change in the section of the sample around distance marker 220—275 (Fig. 27), thereby suggesting trapping of the large oil clusters in the pre space. This can be explained as a result if blockage of the pore throats by the injected wetting phase fluid. Therefore, in this stage the recovery of oil occurs only in parts of the core sample, aligning with a drop in oil saturation from 62.4% to 52% as reported by Scaniziani et al. (2018). This low recovery indicates a need for further fluid injection.

Gas Injection (GI)

Imbibition and drainage mechanisms are observed in the successive cycles of a water alternating gas (WAG) operation. Gas is injected into the porous system from the top of the sample in this stage. **Figs. 29** and **30** shows the change in the oil phase and water phase

connectivity respectively due to gas injection. The significant decrease in the connectivity of both the water and oil phases is consistent with the drop in saturations. Oil phase saturation dropped from 52% to 30.5% while water phase saturation dropped from 48% to 4% (Scanziani et al. 2018) due to gas injection. A drop in both connectivity and saturation suggests maximum recovery of water and oil from the sample after the GI stage.
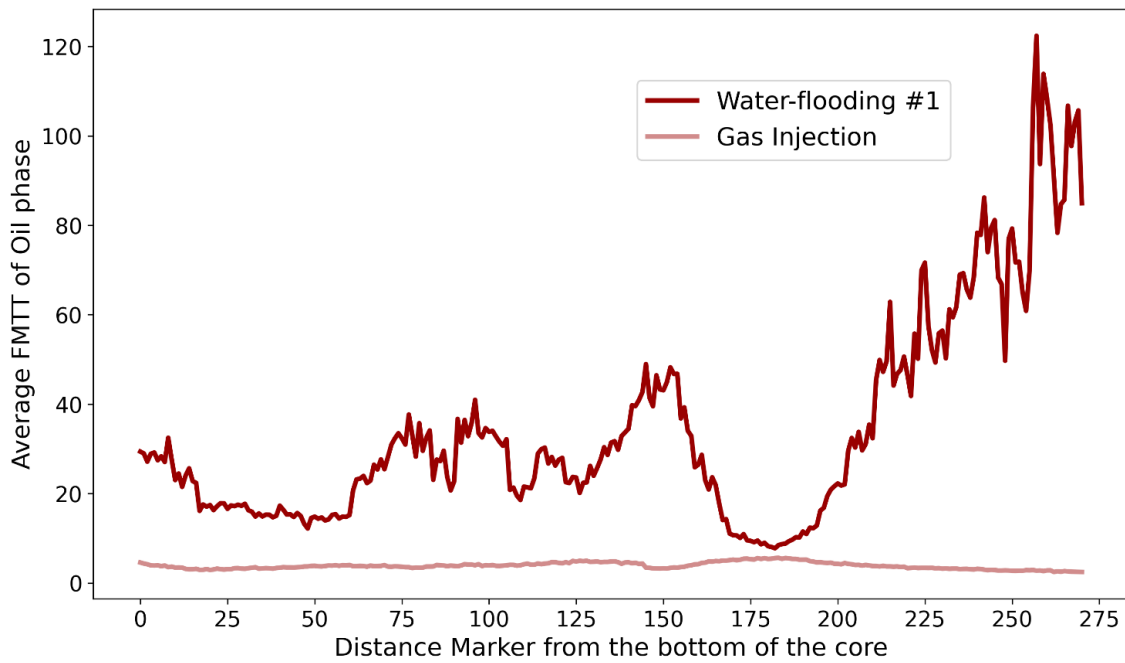

Figure 29. Change in the 3-D connectivity of oil phase from first water-flooding (WF1) to gas injection (GI) stage.
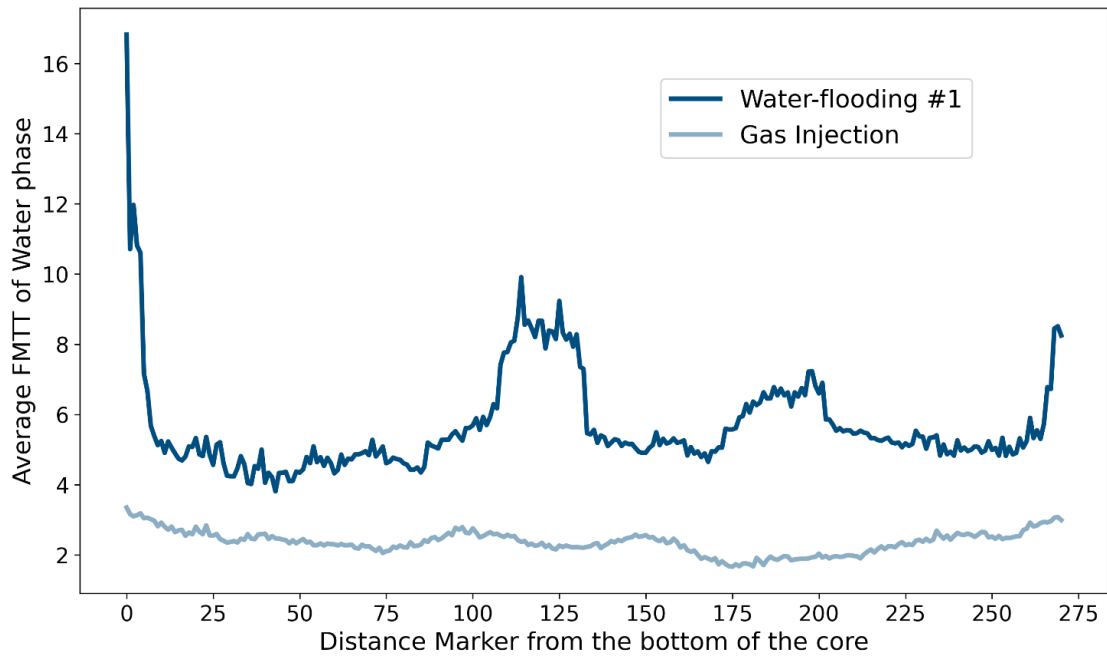
Figure 30. Change in the 3-D connectivity of water phase from first water-flooding (WF1) to gas injection (GI) stage.

Gas is the most non-wetting phase. Therefore, when introduced, gas occupies the center of largest pores and facilitates displacement of the water and oil phases present. Oil phase initially occupying the largest clusters are displaced by the injected gas. This oil, in turn, displaces the water present in the small and medium sized pores, resulting in a negligible residual water saturation. Since both gas and oil are non-wetting phases, the observed behavior can be termed as double drainage (Scanziani, 2018). A high percentage change in the three-dimensional connectivity from WF1 to GI stage proves that the gas injection cycle of the WAG process has the maximum contribution to the ultimate recovery of oil from the carbonate core.

Water-flooding #2

Figs. 31, **32** and **33** shows the change in the oil, water and gas phase respectively from the gas injection stage to the final stage of the WAG sequence. The double drainage

59

phenomena occurring during gas injection is evident in section of the sample around distance marker 150 in the GI plot in Figs. 31 and 33. We notice a low connectivity of the oil phase coupled with a very high connectivity in the gas phase, denoting that injected gas displaced oil clusters from the larger pores, which in turn displaced water from smaller pore spaces.

Figure 31. Change in the 3-D connectivity of oil phase from gas injection (GI) stage to second water-flooding (WF2).

In the final stage, water is injected from the bottom of the sample. The oil connectivity increases in most regions of the sample and decreases in others (Fig. 31), indicating that the injected water mostly contributes to rearrangement of the oil clusters rather than the recovery. Oil phase saturation drops from 30.5% to 18% ensuring partial recovery from small and medium sized pores. Water-flooding #2 failed to recover the few clusters that were trapped after the gas injection.

The re-injected water reconnects and mobilizes the water present in parts of the sample (Fig. 32, distance marker 30—100) and also aids the recovery of the gas phase from the sample. Water being the wetting phase, displaces oil from the smaller pores, which in turn, invades the larger pores and displaces the gas phase present (distance marker 150 and 225—275). Since oil and water phase are more wetting than gas, this mechanism is called double imbibition (Scanziani et al. 2018). In other parts of the sample (Figs. 31 & 33, distance marker 15—30), the displaced oil surrounds the gas phase, thereby trapping the latter and keeping the connectivity intact. The residual gas saturation is 52%.

Figure 32. Change in the 3-D connectivity of water phase from gas injection (GI) stage to second water-flooding (WF2).
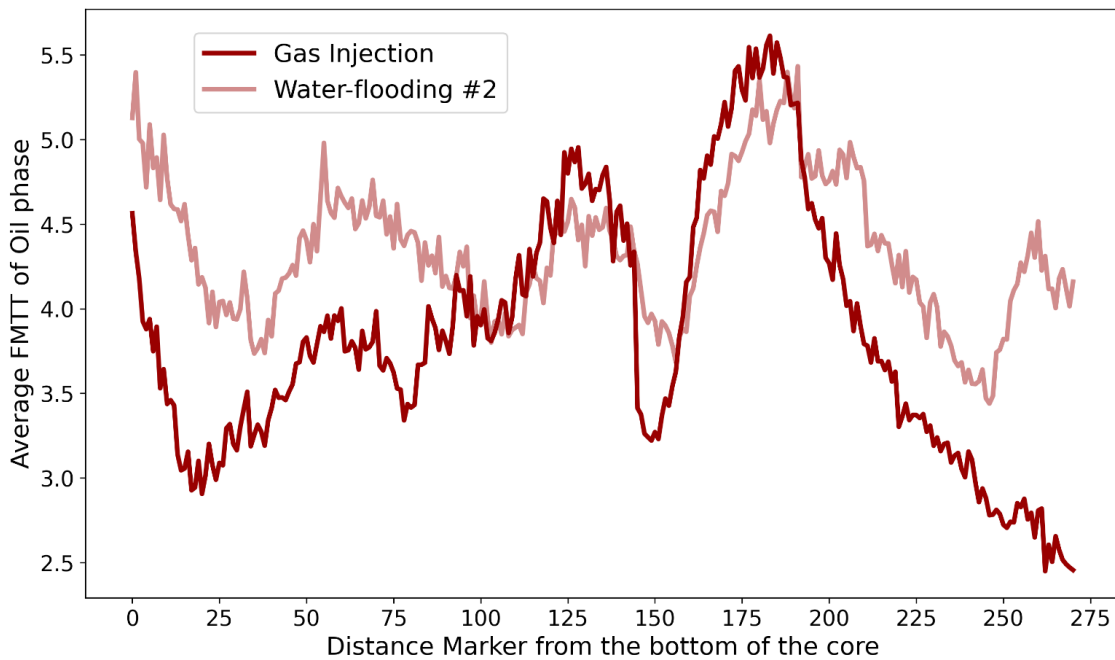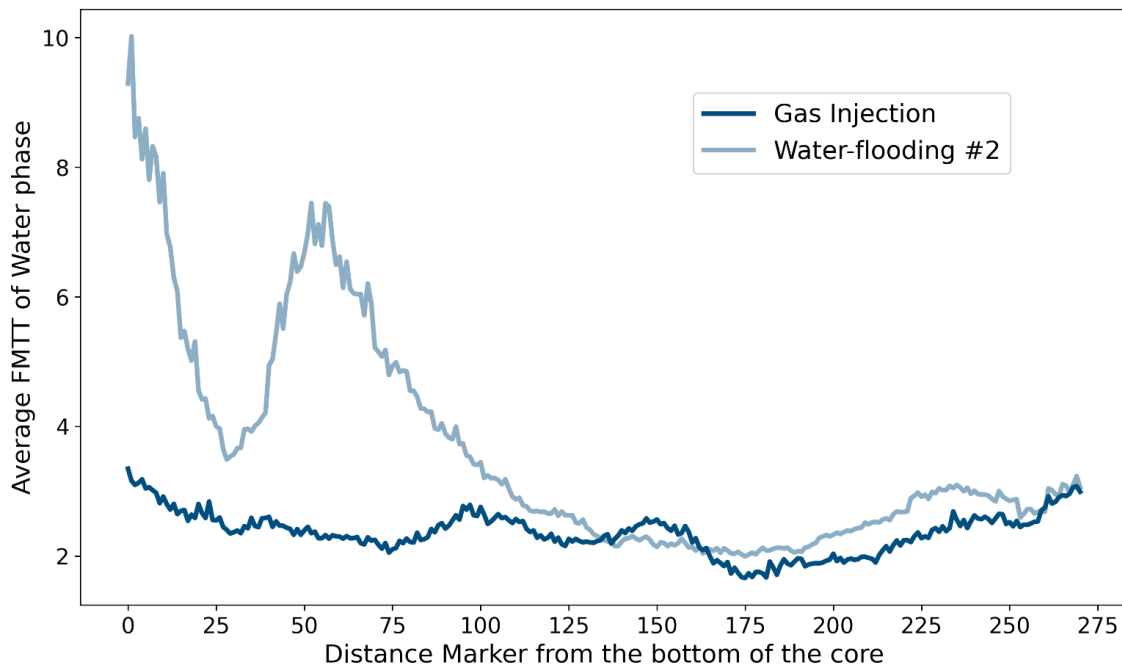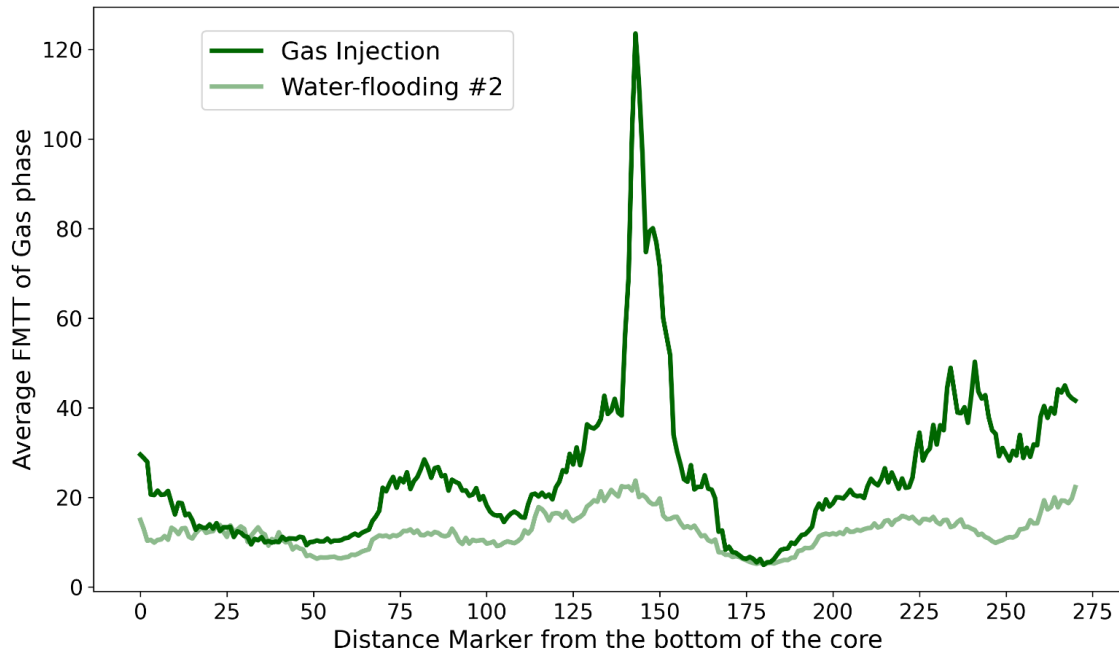
Figure 33. Change in the 3-D connectivity of gas phase from gas injection (GI) stage to second water-flooding (WF2).

At each cycle of the WAG sequence, the connectivity metrics successfully capture the evolution of connectivity of the displaced fluids due to injection of the displacing fluids.

Fig. 34 demonstrates the three-dimensional evolution of an oil phase cluster (red) at each stage of the WAG process [16]. As supported by the two-dimensional (Fig. 20) and three-dimensional (Fig. 27, Fig. 29 and Fig. 31) connectivity evolution plots, the connectedness of the oil cluster decreases at each stage, suggesting oil recovery from each cycle of WAG injection.
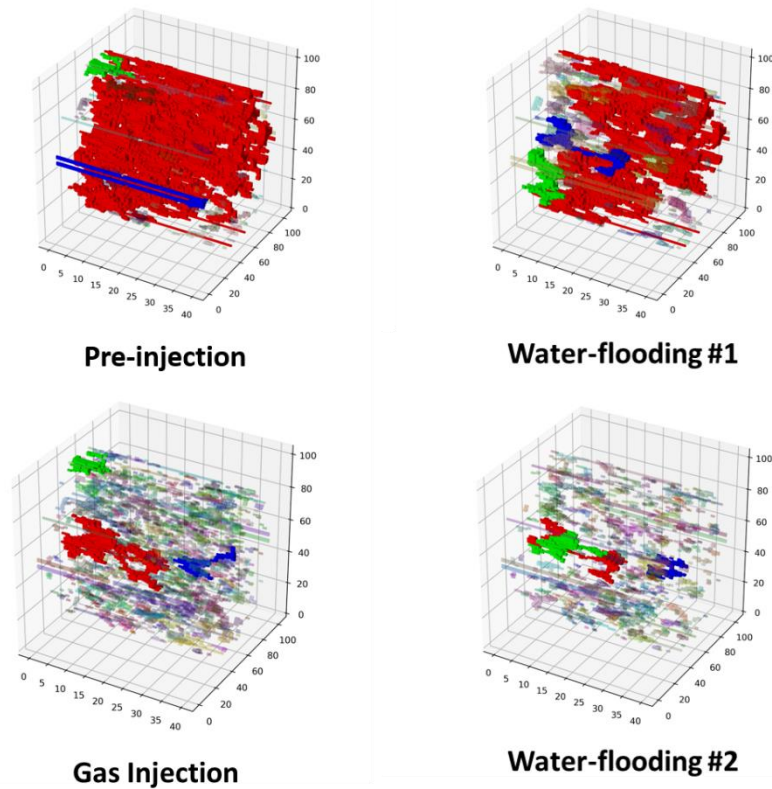
Figure 34. 3D visualization of the displacement and recovery of a large connected oil phase cluster (red) at each injection stage of the water alternating gas (WAG) process (Reprinted from Ganguly and Misra 2021)

Conclusion

In a porous material undergoing water alternating gas (WAG) injection, the change in the two-dimensional and three-dimensional connectivity of the wetting and non-wetting fluid phases can be efficiently captured using the proposed connectivity metrics: connectivity function, CF(h) and fast marching based travel time (FMTT). Each metric takes a different approach to compute the pixel connectivity of a phase of interest, but exhibits a strong positive correlation when the images do not contain highly tortuous connected paths, thereby establishing the reliability of these statistical functions for evaluating the

connectivity of the core sample dataset. In petroleum engineering, high tortuosity contributes to the reservoir heterogeneity. A major difference in the connectivity profile of a phase from the two proposed metrics will shed light on the tortuosity of the connected pathway in the three-dimensional core.

This direct quantification of the connectivity of the fluid phases across the length of the sample help identify the sections of the sample undergoing maximum oil recovery as well as oil trapping. The gas injection stage is found to contribute most to the recovery of oil from the water-wet core. The computed connectivity values at each stage are also found to be consistent with the fluid saturations obtained from the experimental results. Future work would focus on reducing the time complexity of the algorithms to process heavy image datasets.

References

1. Aina Frau-Pascual, Morgan Fogarty, Bruce Fischl, Anastasia Yendiki, Iman Aganj, Quantification of structural brain connectivity via a conductance model, NeuroImage, Volume 189, 2019, Pages 485-496, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2019.01.033.

2. King P.R. (1990) The Connectivity and Conductivity of Overlapping Sand Bodies. In: Buller A.T., Berg E., Hjelmeland O., Kleppe J., Torsæter O., Aasen J.O. (eds) North Sea Oil and Gas Reservoirs—II. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-0791-1_30

3. Torquato, S. & Stell, G. (1982). Microstructure of two-phase random media. I. The n-point probability functions. The Journal of Chemical Physics, Volume 77, Issue 4, 2071-2077. https://doi.org/10.1063/1.444011

4. Torquato, S. & Lado, F. (1985). Characterisation of the microstructure of distributions of rigid rods and discs in a matrix. J. Phys. A: Math. Gen. 18 (1985) 141-148. https://doi.org/10.1088/0305-4470/18/1/025

5. Tashkinov, M. (2017). Statistical methods for mechanical characterization of randomly reinforced media. Mechanics of Advanced Materials and Modern Processes. https://doi.org/10.1186/s40759-017-0032-2

6. Karsanina, M., Gerke, K.M., Skvortsova, E.B. & Mallants, D. (2015). Universal Spatial Correlation Functions for Describing and Reconstructing Soil Microstructure. https://doi.org/10.1371/journal.pone.0126515

7. Hovadik, J & Larue, D. (2007). Static characterizations of reservoirs: Refining the concepts of connectivity and continuity. Petroleum Geoscience - PETROL GEOSCI. 13. 195-211. https://doi.org/10.1144/1354-079305-697

8. Western, A.W., Bloschl, G. & Grayson, R.B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. Water Resources Research, VOL. 37, NO. 1, PAGES 83-97. https://doi.org/10.1029/2000WR900241

9. Shahverdi, H., & Sohrabi, M. (2016, June 1). Relative Permeability Characterization for Water-Alternating-Gas Injection in Oil Reservoirs. Society of Petroleum Engineers. https://doi.org/10.2118/166650-PA

10. Shokufe Afzali, Ali Ghamartale, Nima Rezaei, Sohrab Zendehboudi. Mathematical modeling and simulation of water-alternating-gas (WAG) process by incorporating capillary pressure and hysteresis effects. Fuel, Volume 263, 2020, 116362, ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2019.116362

11. Scanziani, A., Singh, K., Bultreys, T., Bijeljic, B. & Blunt, M.J. (2018). In situ characterization of immiscible three-phase flow at the pore scale for a water-wet carbonate rock. Advances in Water Resources, Volume 121, Pages 446-455. https://doi.org/10.1016/j.advwatres.2018.09.010

12. Renard, P., Allard, D. (2011). Connectivity metrics for subsurface flow and transport. Advances in Water Resources Volume 51, Pages 168-196. https://doi.org/10.1016/j.advwatres.2011.12.001

13. Wu, Y., & Misra, S. (2020, July 20). Quantification of Connectivity in Images. Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2020-2937

14. Shokufe Afzali, Nima Rezaei, Sohrab Zendehboudi. A comprehensive review on Enhanced Oil Recovery by Water Alternating Gas (WAG) injection. Fuel, Volume 227, 2018, Pages 218-246, ISSN 0016-2361. https://doi.org/10.1016/j.fuel.2018.04.015

15. Eliza Ganguly and Siddharth Misra, Spatiotemporal Variations in the Connectivity of Wetting and Nonwetting Phases during Water Alternating Gas Injection, Energy & Fuels 2021 35 (2), 1129-1142, DOI: 10.1021/acs.energyfuels.0c03170

16. Ganguly, Eliza, Misra, Siddharth, and Yaokun Wu. "Evolution of Three Dimensional Fluid Phase Connectivity During Injection." Paper presented at the SPE Annual Technical Conference and Exhibition, Virtual, October 2020. doi: https://doi.org/10.2118/201314-MS

CHAPTER IV

LIMITATIONS AND RECOMMENDATIONS

In this section, the challenges and limitations of the workflows described in chapters two and three are listed. Following that, some recommendations are presented for using this work in the future.

Certain limitations of the Machine Learning-assisted Image Segmentation (MLIS) method introduced in chapter two are as follows:

- In this work, training and test dataset have been created by selecting the pixels from the images with the help of the ImageJ application. It is a manual process and especially challenging for the boundary region pixels.

- The Random Forest model takes approximately 3 seconds to segment one SEM image slice of shale formation. This may be a challenge for batch operations. Time complexity of the method can be reduced by eliminating the low ranked features from training based on the MDA feature ranks.

- The performance of the MLIS model in identifying the boundary region pixels of components with similar pixel intensity has scope for improvement.

- The model does not differentiate organic pores from inorganic pore/crack.

Limitations of the connectivity quantification workflow presented in chapter three can be listed as the following:

- Micro-CT scan images in a dataset may vary in exposure, thereby exhibiting different range of pixel intensities in different image slices. The images need to be

preprocessed to normalize the intensities before performing image segmentation. An approach to preprocess such images is to perform histogram equalization.

- If the connectivity metrics encounter a fluid phase encircling the boundary of the core, as in case of this CT-scan data, that fluid boundary is captured as one large cluster, few orders higher than the connectivity of other clusters present in the image. Therefore, in such case, the image needs to be cropped before applying the metrics.

- The images need to be downscaled to reduce time complexity in case of batch operations.

- The metrics cannot capture connectivity of multiple fluid phases simultaneously and requires the images to be binarized.

- Fast Marching based Travel time (FMTT) depends on the tortuosity of the connected path and overestimates the connectivity of clusters where tortuosity is high.

A few recommendations for using the methods and models presented in this thesis for future works:

- The machine learning assisted image segmentation (MLIS) method for classifying different components in the SEM maps of organic-rich shale, as elaborated in chapter two, can be used to segment images where boundary region pixels play an important role, for example whole-core images. It is important to identify the pixels at the interface of two phases to compute quantities like contact angles and

curvature. The MLIS model outperforms traditional segmentation models in the identification of boundaries or transition zone pixels.

- New features can be extracted to optimize the MLIS model to detect the organic pores embedded in the kerogen components and the inorganic pore/cracks sharing majority of the boundary with matrix in the SEM images of organic-rich shales.

- The connectivity function and fast marching based travel time calculation formulated in chapter three is a generalized approach and can be used to compute the pixel connectivity from binary images. The computational time will vary with the size of the image and the number of clusters of the phase of interest present.