AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2020WR027473

Special Section:
Big Data & Machine Learning in Water Sciences: Recent Progress and Their Use in Advancing Science

Using Unsupervised Machine Learning to Characterize Capillary Flow and Residual Trapping

Hailun Ni[1] and Sally M. Benson[1]

[1]Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

Abstract Recent research has highlighted the impact mesoscale heterogeneity can have on larger-scale multiphase fluid flow properties. However, currently, there is no consensus on how to quickly and reliably analyze coreflooding experimental data to gain insights into mesoscale capillary-dominated flow behaviors and how rock petrophysical properties affect such behaviors. In this study, we combine a machine learning-based clustering method with physics-based hypothesis testing to analyze multistage steady-state coreflooding experimental data. By examining five different sets of coreflooding data, we have found that under capillary-dominated flow regimes, voxel-level CO_2 saturation fields are much more heterogeneous as capillary equilibrium is approached. CO_2 saturation time series behaviors are distinctively different between high and low permeability voxels. However, the effect of mesoscale permeability greatly diminishes under viscous-dominated flow regimes. By computing the maximum and optimal number of clusters on the CO_2 saturation time series data, we are able to differentiate between these two flow regimes. Through this analysis method, it is now also possible to identify regions within a sandstone core with capillary heterogeneity trapping behaviors. It is confirmed that capillary heterogeneity trapping indeed occurs upstream of capillary barriers. Furthermore, we have discovered a new type of capillary heterogeneity trapping—one that occurs within the capillary barrier itself, which can cause parts of the core to achieve residual CO_2 saturations much higher than expected based on existing models of residual gas trapping.

Key Points:
- An automated unsupervised machine learning method is developed to differentiate flow regimes and identify capillary heterogeneity trapping
- Capillary- and viscous-dominated flow regimes can be differentiated by combining time series clustering and hypothesis testing
- This work shows the promise of machine learning methods for analyzing large data sets from coreflooding experiments

Supporting Information:
- Supporting Information S1
- Table S1

Correspondence to:
H. Ni,
hni@stanford.edu

Citation:
Ni, H., & Benson, S. M. (2020). Using unsupervised machine learning to characterize capillary flow and residual trapping. *Water Resources Research*, *56*, e2020WR027473. https://doi.org/10.1029/2020WR027473

Received 8 MAR 2020
Accepted 24 JUN 2020
Accepted article online 22 JUL 2020

1. Introduction

CO_2 geologic storage is an important approach for reducing greenhouse gas emissions (IPCC, 2005, 2014; Rogelj et al., 2018). Recent research has shown that mesoscale (millimeter-scale) heterogeneity can have a substantial influence on larger-scale multiphase flow and trapping behaviors in porous media. Mesoscale heterogeneity can significantly affect saturation distribution, relative permeability, and capillary pressure measurements at the larger scale during both drainage and imbibition (Chang & Yortsos, 1992; Corey & Rathjens, 1956; Graue, 1994; Honarpour & Saad, 1994; Huppler, 1970; Saad et al., 1995; Shi et al., 2011a). Mesoscale heterogeneity can also slow CO_2 plume spreading by a factor of 2 (Li & Benson, 2015). Furthermore, mesoscale capillary barriers can residually trap a significant fraction of the nonwetting phase fluid (CO_2 or oil) (Huang et al., 1995; Kortekaas, 1985; Krevor et al., 2011; Meckel et al., 2015; Saadatpoor et al., 2010; Trevisan et al., 2017), thereby reducing CO_2 geologic storage leakage risks (Zahasky & Benson, 2016). It is therefore essential for us to better understand how mesoscale heterogeneity affects CO_2/water flow and trapping in real reservoir rocks in order to achieve safe long-term CO_2 geologic storage. In particular, this study uses a data-driven approach to answer the following two questions: (1) How to accurately differentiate between capillary- and viscous-dominated flow regimes in the presence of mesoscale heterogeneity without relying on computing capillary numbers? (2) How to identify, quantify, and visualize capillary heterogeneity trapping, a type of residual trapping induced by mesoscale heterogeneity, inside real sandstone cores during coreflooding experiments?

Nowadays, advanced noninvasive imaging techniques such as medical X-ray computed tomography (CT) scanning allow us to calculate porosity, permeability, and saturation at the millimeter scale (Akin & Kovscek, 2003; Krause & Benson, 2015). Therefore, a large amount of voxel-level saturation data can be generated when multiphase flow experiments are conducted at various fractional flows in order to calculate steady-state drainage and imbibition relative permeability curves. Conventional methods of analyzing voxel-level CO_2/water coreflooding data include calculating the core-averaged relative permeability, capillary pressure, and residual trapping relationships (Akbarabadi & Piri, 2015; Hingerl et al., 2016;

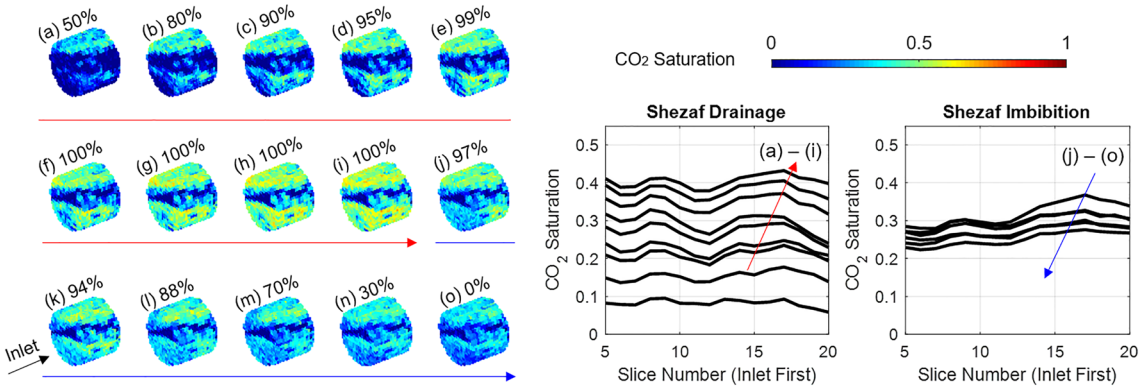

Figure 1. 3-D CO_2 saturation maps and 1-D slice-average CO_2 saturation profiles along the length of the central portion of the Shezaf core sample for 15 different CO_2 fractional flow values scanned at steady state. Stages (a) to (i) are drainage, and stages (j) to (o) are imbibition. The percentage numbers represent the CO_2 fractional flow values used for the stages. For stages (f) to (i), the total flow rate of the 100% CO_2 injection is increased at each stage. Inlet is to the left. The red arrows indicate the direction of scan sequence during drainage, and the blue arrows indicate the direction of scan sequence during imbibition.

Krause, 2012a, 2012b; Krevor et al., 2015, 2012; Ni et al., 2018, 2019; Niu et al., 2015; Pentland et al., 2011; Pini & Benson, 2013; Ruprecht et al., 2014). In addition, one can also visualize the three-dimensional (3-D) or two-dimensional CO_2 saturation fields as well as visualizing the one-dimensional slice-average CO_2 saturation profiles along the length of the core to get a sense of the rock core's lamination direction and other internal structures (Akbarabadi & Piri, 2013; Krevor et al., 2012; Shi et al., 2011b; Zuo & Benson, 2014). Furthermore, one can calculate a series of statistics from the voxel-level property and saturation fields to gauge the rock core's directional correlation length and degree of heterogeneity (Ni et al., 2019). However, currently, there is no consensus on how we can quickly and more systematically gain greater insights into the physical flow processes happening at the mesoscale throughout the entire experiment.

Taking the coreflooding data obtained from a Shezaf sandstone as an example, we can visualize the data as shown in Figure 1. The internal structures of this core sample are clearly visible from the 3-D CO_2 saturation maps. From the 1-D CO_2 saturation profiles, we can get a sense of the core's degree of heterogeneity as well as its residual trapping capacity. While conventional analysis methods yield useful information by analyzing the spatial aspect of the voxel-level data, they fail to take advantage of the temporal aspect of the data and cannot provide further insights into how the underlying petrophysical fields affect the resulting CO_2 saturation field at each stage of the experiment. Currently, no existing coreflooding data analysis methods can identify how different regions of the core sample behave similarly or differently and what factors affect such differences throughout the entire multistage experimental procedure. Furthermore, one capability of particular interest is to be able to identify regions in a sandstone core with different types of residual trapping occurring, such as capillary heterogeneity trapping. At the pore scale, CO_2 is residually trapped through snap off and bypass; while at the mesoscale, above pore-scale residual level CO_2 can be trapped behind mesoscale capillary barriers (Krevor et al., 2015). Ni et al. (2019) have calculated the relative contribution to CO_2 residual trapping from the two different residual trapping mechanisms, but no previous studies have been done to identify actual regions with capillary heterogeneity trapping and visualize such regions in a rock core or on an initial-residual saturation (IR) plot.

In this study, we employ an unsupervised learning-based method to quickly and methodically analyze steady-state coreflooding data. This method enables one to visualize how different parts of the core display different CO_2/water flow and trapping behaviors. Together with voxel-level porosity and permeability data, this information can be used to explain how rock properties influence the multiphase flow data. This unsupervised learning-based method involves clustering CO_2 saturation time series data at the voxel level. Note that in this study, we use "time series" to refer to the CO_2 saturation data series obtained by varying the CO_2 fractional flow values sequentially in time as illustrated in Figure 1. Therefore, all CO_2 saturation data are acquired at steady state. Unlike traditional clustering methods where the maximum number of clusters is determined arbitrarily, in this work, the novel analysis method combines the time series clustering

Table 1
Dimensions, Petrophysical Property Values, Lamination Directions, and the Experimental Conditions of the Sandstone Cores Analyzed in the Current Study (Reynolds et al., 2018)

Core name/sandstone type	Berea	Shezaf	Bunter	Ormskirk
Core diameter (cm)	5.08	4.9	3.81	3.81
Core length (cm)	12	6.6	15.1	12.7
Core average porosity	0.19	0.14	0.26	0.27
Core absolute permeability (md)	214	79	2,200	12,100
Capillary entry pressure (Pa)	4,922	4,120	1,964	1,097
Experiment temperature (°C)	50	50	53	33
Experiment pressure (MPa)	9	9	13.1	12.7
Experiment salinity (mol/kg)	0	0	1	4.32
Voxel size for analysis (mm^3)	$2 \times 2 \times 2$	$2 \times 2 \times 2.5$	$2 \times 2 \times 3$	$2 \times 2 \times 3$
Lamination direction	Parallel to the flow	No obvious lamination	Perpendicular to the flow	Perpendicular to the flow

technique with hypothesis testing to find the maximum number of clusters at which to automatically terminate the clustering algorithm. This approach ensures that all clusters have statistically significant differences in porosity or permeability mean/median values and no redundant clusters are produced (Aghabozorgi et al., 2015; Albalate & Minker, 2013; James et al., 2013; Kaufman & Rousseuw, 1990; Kyan et al., 2014; Liao, 2005). We are not aware of any other similar data analysis methods present in the coreflooding experiment field.

2. Data and Methods

2.1. Experimental Data Acquisition

Experiments have been conducted on four sandstone core samples with different types and degrees of heterogeneity. The core samples come from four different sandstone formations: Berea, Shezaf, Bunter, and the Ormskirk formations. Table 1 shows the information on all core samples used in the current study. All sandstone cores analyzed in this study are quartz dominated and water wet (Ni et al., 2019; Reynolds et al., 2018).

Experimental data from two separate sources are used here. The first set of sandstone core sample data, Berea and Shezaf, is obtained from coreflooding experiments conducted in our laboratory. For this set of core samples, low flow rate steady-state CO_2/water coreflooding data are obtained from 15 stages of CT scans, including both drainage and imbibition. As illustrated in Figure 1, there are five stages of CO_2/water coinjection drainage scans with increasing CO_2 fractional flow. Then the coinjection drainage scans are followed by four stages of 100% CO_2 injection scans with increasing total flow rates. After all the drainage stages are concluded, there are six coinjection imbibition stages which then follow with decreasing CO_2 fractional flow. For detailed CO_2 fractional flow values and flow rates used for all experiments covered in this study, see supporting information. Only data in the central portion of these core samples are used in the analysis to avoid experimental artifacts caused by the inlet and outlet boundary conditions. For more details of the core samples, the experimental procedure, and data uncertainty, see Ni et al. (2019) and the associated supplementary material.

The second set of sandstone core sample data, Bunter and Ormskirk, is coreflooding data published online (Reynolds et al., 2018). Both low and high flow rate multistage drainage and imbibition coreflooding experiments have been conducted on this set of core samples. More specifically, for the current study, the following experimental data are used: Bunter high flow rate experiment (B3B4), Bunter low flow rate experiment (B5B6), and Ormskirk high flow rate experiment (O1O2). Experiments in this second set only have coinjection drainage and imbibition scans without the multiple 100% CO_2 injection stages that the first set has. For this set of core samples, data from the entire domain are used in the analysis. For more information on this set of coreflooding data, see Reynolds et al. (2018).

One of the goals of this work is to differentiate flow regimes, and capillary number is one commonly used means. However, there exists a wide range of capillary number definitions that emphasize different

Table 2
Maximum Values of the Four Different Capillary Numbers for Both Drainage and Imbibition for All Experiments

	$N_{c,1}$		$N_{c,2}$		$N_{c,3}$		$N_{c,4}$	
Rock type	Drainage	Imbibition	Drainage	Imbibition	Drainage	Imbibition	Drainage	Imbibition
Berea	1.3×10^{-7}	7.7×10^{-7}	7.6×10^{-7}	3.6×10^{-6}	0.32	1.51	0.09	0.56
Shezaf	7.0×10^{-8}	2.1×10^{-7}	4.4×10^{-7}	1.8×10^{-6}	1.48	5.94	0.27	0.82
Bunter low rate	4.0×10^{-9}	4.6×10^{-8}	1.5×10^{-6}	1.6×10^{-6}	0.56	0.59	3.0×10^{-4}	0.003
Bunter high rate	4.0×10^{-7}	4.6×10^{-6}	5.0×10^{-5}	4.3×10^{-5}	19.2	16.2	0.03	0.34
Ormskirk high rate	6.6×10^{-7}	7.1×10^{-6}	1.5×10^{-4}	1.4×10^{-4}	141	135	0.02	0.19
Capillary-dominated flow regime criteria	$N_{c,1} \leq 10^{-6}$ (Berea: $N_{c,1} \leq 10^{-5}$)		$N_{c,2} < 2 \times 10^{-5}$		$N_{c,3} < 1$~100 (Bunter: $N_{c,3} < 5$; Ormskirk: $N_{c,3} < 30$)		$N_{c,4} \ll 1$	

Note. The criteria for judging capillary-dominated flow regimes are shown in the last row for all capillary numbers (Chatzis & Morrow, 1984; Morrow et al., 1988; Reynolds et al., 2018; Reynolds & Krevor, 2015; Suekane et al., 2010; Virnovsky et al., 2004; Vizika & Payatakes, 1989; Yokoyama & Lake, 1981; Zhou et al., 1994, 2010).

perspectives. Here, we compare four such definitions. The first one is the most commonly used conventional capillary number,

$$N_{c,1} = v\mu/\sigma, \tag{1}$$

where v is the total Darcy velocity, μ is the viscosity of the majority fluid flowing (for simplicity, the non-wetting phase viscosity is used for drainage and the wetting phase viscosity for imbibition), and σ is the CO_2/water interfacial tension at experimental temperature, pressure, and salinity (Lake et al., 2014; Morrow et al., 1988; Suekane et al., 2010; Vizika & Payatakes, 1989; Zhou et al., 2010). The second one is an alternative form of the conventional capillary number derived with Darcy's law:

$$N_{c,2} = k_w \Delta P / L\sigma, \tag{2}$$

where k_w is the permeability measured with water or brine, ΔP is the pressure drop across the core, and L is the core length (Chatzis & Morrow, 1984). Both $N_{c,1}$ and $N_{c,2}$ measure the relative strength of viscous forces and local capillary forces at the pore scale (Hilfer et al., 2015; Lake et al., 2014). However, $N_{c,2}$ includes information about the permeability of the rock that is not considered in $N_{c,1}$. The third capillary number is as follows:

$$N_{c,3} = \Delta P h / L \Delta P_c, \tag{3}$$

where h is lamination thickness and ΔP_c is the standard deviation in the voxel-level capillary entry pressures (Reynolds et al., 2018; Reynolds & Krevor, 2015; Virnovsky et al., 2004). $N_{c,3}$ allows direct comparison between the viscous pressure drop and the degree of capillary heterogeneity (Reynolds & Krevor, 2015). Finally, the fourth one is the transverse capillary number:

$$N_{c,4} = H^2 v\mu / L p_c^* k_w, \tag{4}$$

where H is the core diameter and p_c^* the capillary entry pressure (Hilfer et al., 2015; Kuo & Benson, 2015; Reynolds & Krevor, 2015; Yokoyama & Lake, 1981; Zhou et al., 1994). $N_{c,4}$ measures the relative importance of viscous-driven longitudinal flow versus capillary-driven transverse flow caused by mesoscale heterogeneity (Reynolds & Krevor, 2015). The main difference between $N_{c,1}$, $N_{c,2}$ and $N_{c,3}$, $N_{c,4}$ is that the former two are more related to fluid properties and fluid-fluid interaction while the latter two are more related to the degree of heterogeneity of the domain. One advantage of $N_{c,2}$, $N_{c,3}$ over $N_{c,1}$ and $N_{c,4}$ is that both $N_{c,2}$ and $N_{c,3}$ take different CO_2 fractional flow values into consideration through the ΔP term.

The maximum capillary number values for both drainage and imbibition for all five experiments in this study are tallied in Table 2. For details on calculation, see the supporting information. The criteria for judging capillary-dominated flow regimes for the four different capillary numbers are also indicated in Table 2.

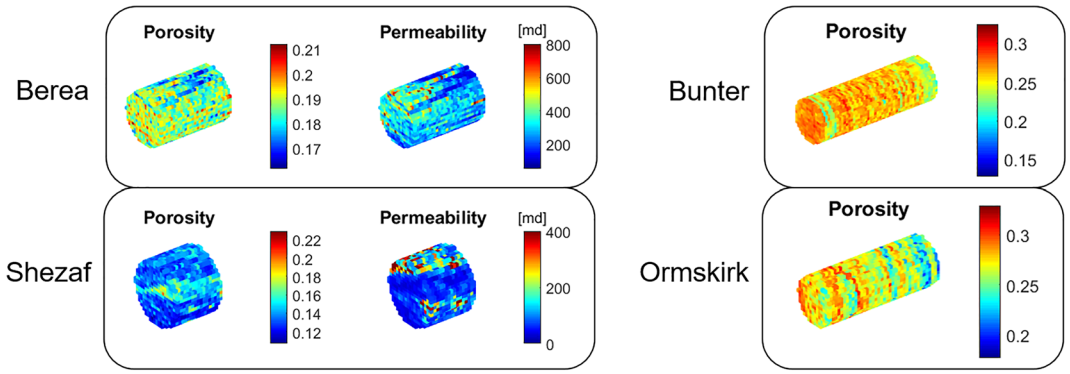

Figure 2. 3-D voxel-level porosity and permeability maps for the core samples covered in this study.

From Table 2, it is apparent that the different capillary numbers show conflicting results in most of the experiments. Therefore, it is unclear which capillary number can best differentiate between the capillary- and viscous-dominated flow regimes in the presence of mesoscale heterogeneity. While Reynolds et al. (2018) employed a manual approach by visually examining the CO_2 saturation cross-section images, in this study, we propose an automatic statistical approach to differentiate flow regimes as will be discussed later.

From the CT scans, we can calculate the CO_2 saturation maps, as shown in Figure 1. We can also obtain both the voxel-level porosity and the permeability maps, as shown in Figure 2. These three maps are the necessary inputs into the clustering method, although it is also possible for the clustering procedure to work with only the CO_2 saturation and the porosity data. While both the CO_2 saturation and the porosity maps can be easily obtained through CT image manipulation, the voxel-level permeability map is obtained using the extended Krause's method (Krause et al., 2013; Krause, 2012a; Krause, 2012b; Krause & Benson, 2015). All voxel-level permeability maps used in the current study are obtained by history matching to the single 95% CO_2 fraction flow scan. For more details on the calculation of the three maps, refer to Ni et al. (2019).

2.2. Statistical Analysis Method

This section describes the statistical analysis method applied to the coreflooding data. When we have multistage coreflooding scan data, each voxel in the core sample has a multistage time series that represents the CO_2 saturation evolution trend throughout the coreflooding experiment. The multistage time series data of each voxel can be thought of as a multidimensional data point, with the voxel's CO_2 saturation value at each stage as the data point's individual dimension or attribute. To efficiently analyze the thousands of voxel-level time series data points, we use unsupervised machine learning. It is a statistical method that works directly on the data without doing regression or classification. More specifically, we use clustering algorithms on the CO_2 saturation time series data. Clustering methods identify patterns in data and group voxels with similar patterns together. For detailed descriptions of the statistical approach used here, see the supporting information. Table 3 provides a summary of the factors considered in designing the unsupervised learning approach described here, along with decisions made regarding how to address them.

As shown in Table 3, we perform the clustering algorithm directly on the raw voxel-level CO_2 saturation time series data without any preprocessing steps such as detrending or scaling. Two different clustering methods are used in this analysis method: K-means and hierarchical clustering. These clustering methods are selected because they are easy to implement and, at the same time, yield the most useful results for our purposes. In K-means clustering, after the number of clusters, K, has been assigned, the algorithm first randomly initializes K centroids. Then at each iteration, it calculates the distance between each data point and the K centroids and assigns the data point to the closest centroid. After all data points are assigned to a cluster, the algorithm then averages all data points within the cluster and assigns that average as the new centroid for the next iteration. Eventually when the centroid locations no longer change, the algorithm will have converged, and the clustering is complete (James et al., 2013). In hierarchical clustering, pairwise distance between all data points is first computed. Initially, all data points are treated as separate clusters.

Table 3
Available Options Considered for All the Main Statistical Analysis Decisions

Main decisions		Available options considered
1	What features to select?	✓ Raw time series data
		• Add first derivatives
		• Add second derivatives
2	Whether or not to preprocess the data?	✓No preprocessing
		• Preprocessing
		• Detrend
		• Scaling
		• Standardization
		• Normalization
3	Which clustering algorithm to select?	✓K-means
		✓Hierarchical
4	What metric to use for pairwise distance?	✓Euclidean distance (for K-means clustering)
		✓Spearman distance (for hierarchical clustering)
		• Other types of distance metrics
5	How to average time series in a cluster for K-means clustering?	✓Arithmetic averaging
		• K-medoids
		• Other types of averaging methods
6	How to measure dissimilarity in between clusters for hierarchical clustering?	✓Ward linkage
		• Other types of linkages
7	How to determine the maximum number of clusters?	• Arbitrary determination
		✓Hypothesis testing on cluster properties
		• Parametric
		✓Two-sample unpaired t-test
		✓One-way analysis of variance
		• Nonparametric
		✓Wilcoxon rank-sum test (Mann-Whitney U test)
		✓Kruskal-Wallis test
8	How to determine the optimal number of clusters?	• Use an evaluation criterion
		✓Gap
		• Silhouette
		• Calinski-Harabasz
		• Davies-Bouldin

Note. The chosen options are marked with tick marks. For more details and discussion, see supporting information (Aghabozorgi et al., 2015; Albalate & Minker, 2013; Anderson & Darling, 2010; Batagelj, 1988; Caliński & Harabasz, 1974; Davies & Bouldin, 1979; Geher & Hall, 2014; Gibbons & Chakraborti, 2011; James et al., 2013; Kaufman & Rousseuw, 1990; Kyan et al., 2014; Liao, 2005; McClave & Sincich, 2017; NIST/SEMATECH, 2013; Rani, 2012; Roelofsen, 2018; Rousseeuw, 1987; Sheskin, 2011; Tibshirani et al., 2001; Ward, 1963).

Then the algorithm works by continually grouping the next two closest clusters until all data points have been grouped into a single cluster. The algorithm then stops and the clustering results can be visualized with a dendrogram. This is known as agglomerative hierarchical clustering. The resulting dendrogram can be cut at any level to produce different clustering results (James et al., 2013). We have conducted both clustering methods on the coreflooding data and found that K-means clustering consistently produces more evenly sized clusters, which is desirable for characterizing flow behaviors, whereas hierarchical clustering tends to produce clusters with highly uneven sizes, which is desirable for identifying any capillary heterogeneity trapping behaviors.

The distance metric used to group the time series data into clusters measures how far apart the data points are from each other. Many different metrics for measuring pairwise distance exist (Aghabozorgi et al., 2015; Liao, 2005; Roelofsen, 2018). The most commonly used distance metric is the Euclidean distance. For this problem, it is defined below for two n-dimensional vectors s_i and s_k

$$d_E = \sqrt{\sum_{j=1}^{n} \left(s_{i,j} - s_{k,j}\right)^2}, \tag{5}$$

where s_i and s_k correspond to voxel-level saturation time series data with n stages each. Other distance metrics include correlation-based metrics such as the Spearman distance (Kaufman & Rousseuw, 1990; McClave & Sincich, 2017). With the Euclidean distance, time series with similar values will be grouped

together. Whereas with the Spearman distance, time series with similar trends will be grouped together (James et al., 2013). For K-means clustering, it is necessary to select a method for averaging the many time series data points in a cluster in order to calculate the centroid for the next iteration step. For time series of equal lengths, arithmetic averaging is often used for each dimension of the data points in the same cluster (Aghabozorgi et al., 2015). Arithmetic averaging is also preferred because it closely approximates the volume-averaged CO_2 saturation value at each stage for a cluster. For hierarchical clustering, because at each iteration step, smaller clusters that are closer together are joined together to form larger clusters, it is necessary to select a linkage method to judge how close or similar the clusters are to each other. In this study, Ward linkage is used because it yields clusters with the most well-defined trapping behaviors. The Ward linkage works by grouping two clusters with the least incremental error sum of squares at each hierarchical clustering step (Kaufman & Rousseuw, 1990; Liao, 2005; Ward, 1963).

In order to determine the maximum number of clusters with which to conduct both the K-means and the hierarchical clustering algorithms, at each iteration step when the number of clusters, K, is increased, we compare the mean/median of the clusters' porosity and permeability properties to ensure that they are significantly different from each other for at least one of the two property fields. When this is no longer true, then further increasing the number of clusters is not necessary. After the maximum number of clusters is reached, one can then use the gap evaluation criterion to find the optimal number of clusters based on the clustering results from the K-means method to differentiate flow regimes (Tibshirani et al., 2001). The optimal number of clusters would be less than or equal to the maximum number of clusters. We use hypothesis testing to compare cluster mean/median properties to ensure that all clusters have significantly different porosity or permeability values at the 5% significance level. For a detailed description of the statistical analysis decisions, see supporting information.

Using K-means clustering with the squared Euclidean distance metric and arithmetic averaging, we are able to characterize the CO_2/water flow regime. More specifically, we investigate how mesoscale permeability affects capillary flow and how to distinguish between capillary- and viscous-dominated flow regimes using the gap value. Using hierarchical clustering with the Spearman distance metric and the Ward linkage, we are able to characterize CO_2 residual trapping behaviors. More specifically, we want to identify regions with different trapping behaviors such as capillary heterogeneity trapping.

A flow chart for the coreflooding data cluster analysis method is shown in Figure 3. The method starts by conducting either K-means or hierarchical clustering depending on whether we want to characterize flow or trapping. For each number of clusters, K, we verify the assumptions for using parametric hypothesis test methods for comparing cluster means. If the assumptions are not met, then we use nonparametric methods. If all the cluster means/medians based on the voxel-level porosity or the permeability fields are significantly different from each other, then we proceed to the next K. Otherwise, we have already reached the maximum number of clusters. Next, there are several postclustering steps that need to be implemented. After K-means clustering, we can evaluate the optimal number of clusters in order to distinguish flow regimes. After hierarchical clustering, we can superimpose the clustering results onto IR plots in order to identify regions with capillary heterogeneity trapping. After either clustering method, cluster-average time series behaviors and cluster index core maps should be plotted and examined to identify any different flow and trapping behaviors.

3. Theory

3.1. Explanation Behind the Capillary-Dominated CO_2 Saturation Time Series Behaviors

This section details the theory that can be used to motivate the clustered CO_2 saturation time series behaviors that will be presented later in this study. In general, the more heterogeneous a core sample is, the more differently we can expect its clusters to behave from the core average behavior and from each other. This is due to mesoscale heterogeneity.

For coreflooding experiments conducted under capillary-dominated conditions, saturation time series behaviors can be explained by capillary heterogeneity under the assumption of capillary equilibrium. For this example, we use the Leverett (1941) J-function scaling relationship to illustrate different CO_2 saturation time

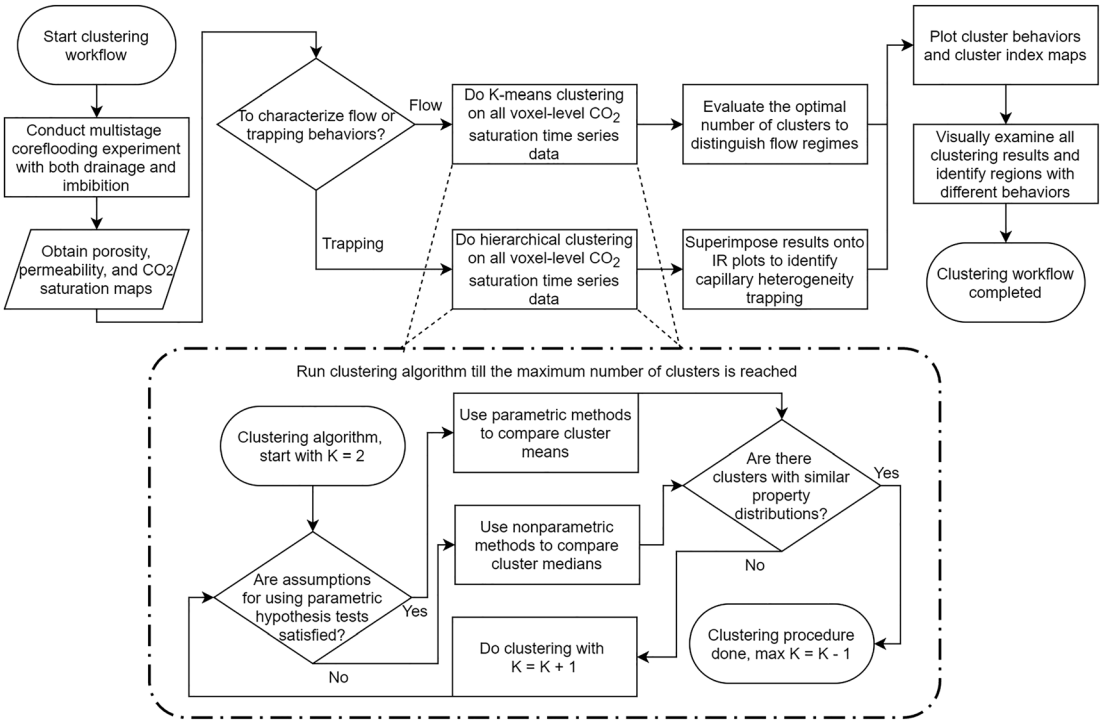

Figure 3. Flow chart of the voxel-level CO_2 saturation time series data cluster analysis method.

series behaviors for both drainage and imbibition. A linear trapping relationship is used in this illustrative example to compute the imbibition scanning curves for the voxels (Pini & Benson, 2017).

Figure 4a illustrates the resulting CO_2 saturation time series behaviors during both drainage and imbibition for three voxels with different permeability values when the domain is at capillary equilibrium. It can be observed that for rocks where the Leverett J-function scaling relationship applies, CO_2 saturation is higher for higher permeability rocks. This is due to the fact that at capillary equilibrium, the voxels' distinct capillary pressure curve controls CO_2 saturation as illustrated in Figure 4a. Furthermore, the illustration also shows that the drainage CO_2 saturation time series curve would be concave upward for low permeability voxels and concave downward for high permeability voxels. When the capillary pressure is high enough, all drainage CO_2 saturation time series always eventually become concave downward, as dictated by the shape of the drainage capillary pressure curves.

As illustrated in Figure 4a, when we reduce the capillary pressure stepwise by imbibition of a wetting phase, the higher the voxel permeability, the steeper the decline in the saturation time series curve. On the other hand, voxels with low permeability do not decrease in saturation very much. From this illustrative example, we can see that the CO_2 saturation time series behaviors of the voxels are all dictated by their individual capillary pressure curves when at or close to capillary equilibrium.

3.2. Explanation Behind the Capillary Heterogeneity Trapping CO_2 Saturation Time Series Behaviors

The previous illustration demonstrates the expected CO_2 saturation time series behaviors when capillary equilibrium is approached. However, during imbibition, when there exists a capillary barrier downstream, the capillary barrier can effectively cause upstream voxels to have higher than expected residual CO_2 saturation at the end of imbibition. These voxels with high initial CO_2 saturation and higher than expected trapping demonstrate typical capillary heterogeneity trapping behaviors (Jackson & Krevor, 2019).

To illustrate how capillary heterogeneity trapping manifests in the CO_2 saturation time series behaviors, we continue with the previous three-voxel example in Figure 4a. The main difference in this case is that now the relative location of these voxels matters. Capillary heterogeneity trapping can only be induced when the

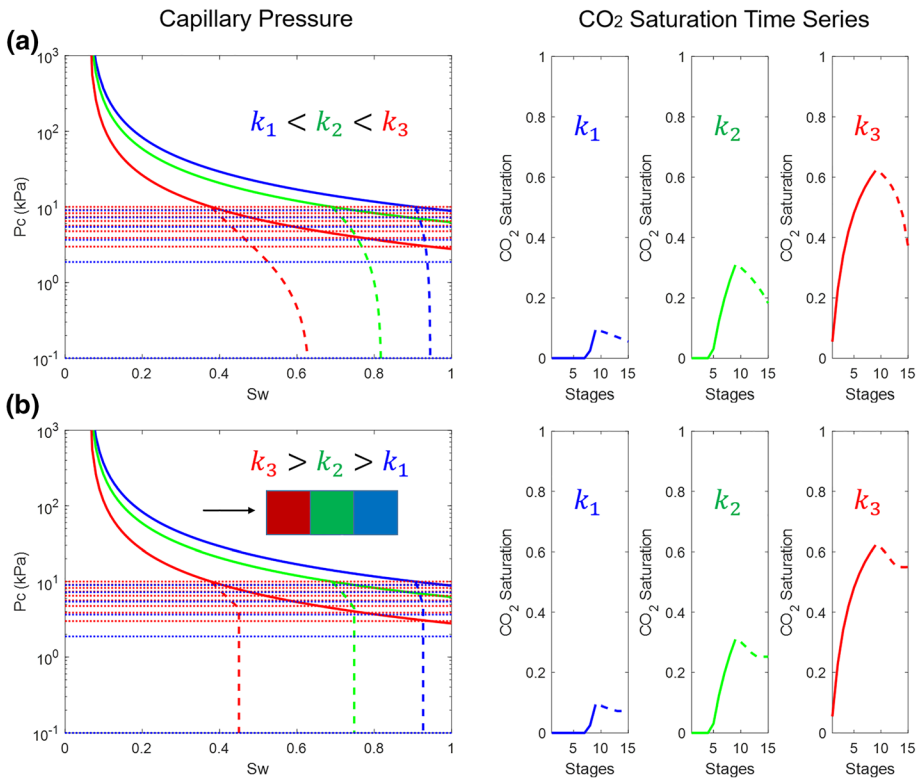

Figure 4. (a) Capillary equilibrium CO_2 saturation time series illustration for both drainage and imbibition. (b) Capillary heterogeneity trapping CO_2 saturation time series illustration for imbibition. (left) Representative capillary pressure curves for the three voxels with different permeability values: blue (100 md), green (200 md), and red (1,000 md); solid lines = drainage; dashed lines = imbibition. The dotted lines show the capillary pressure steps taken: red = drainage; blue = imbibition. Each continuum-scale voxel here has its own well-defined permeability, saturation, and capillary pressure values. (right) Corresponding CO_2 saturation time series plots for the three voxels.

low-permeability voxel, or the capillary barrier, is downstream of the other voxels, as illustrated in the inset schematic in Figure 4b. If a nonzero pore-scale snap-off capillary pressure is imposed onto the voxel-level capillary pressure curves, then as the domain capillary pressure decreases during imbibition, the low-permeability voxel k_1 (blue) will reach its snap-off capillary pressure first. And as voxel k_1's CO_2 saturation is cut off from the continuous CO_2 phase and becomes residually trapped, it effectively blocks CO_2 flow from upstream voxels due to capillary forces (Jackson & Krevor, 2019). Consequently, the resulting CO_2 saturation time series curves of both the medium (k_2) and the high permeability (k_3) voxels are more flattened during imbibition compared to Figure 4a as the presence of a downstream capillary barrier prevents further reduction in their CO_2 saturation values.

4. Results and Discussion

4.1. Use K-Means Clustering to Characterize Capillary Flow

This section presents the results from performing K-means cluster analysis on the five coreflooding data sets in order to characterize capillary flow behaviors and to differentiate between capillary- versus viscous-dominated flow regimes. For each experiment, the maximum number of clusters is reported. As the number of clusters increases, the clusters become more well-defined and distinct from each other. In this study we only present the clustering results of $K \leq 4$ scenarios because they contain all of the essential information and it is easier to see the main features of the clusters. For all experimental results presented, the flow direction is from left to right. The clustering results for the $K = 3$ scenario for all five coreflooding data sets are shown together for easy comparison in Figure 5.

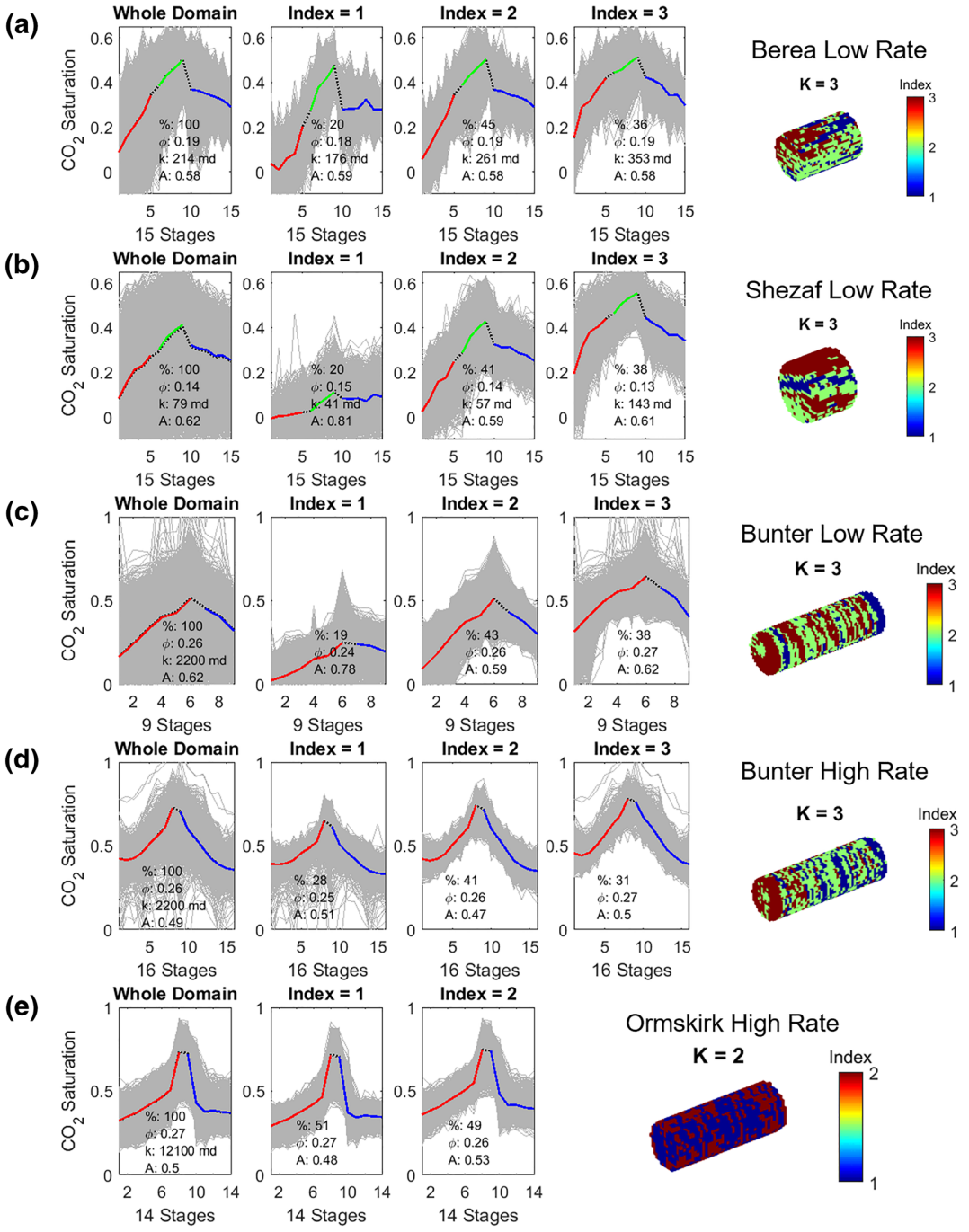

Figure 5. Clustered CO_2 saturation time series data for the K = 3 scenario, with the 3-D cluster index map for all five coreflooding data sets. (a) Berea. (b) Shezaf. (c) Bunter low rate. (d) Bunter high rate. (e) Ormskirk high rate. The first column shows the time series data for the entire study area of the core; the following columns show the clustered time series. In each subplot, the colored line is the cluster-average time series, while the gray lines are all the individual time series in that cluster. The color of the cluster-average time series represents the three experimental stages: red = coinjection drainage with increasing CO_2 fractional flow; green = 100% CO_2 injection drainage with increasing flow rates; and blue = coinjection imbibition with decreasing CO_2 fractional flow. Underneath the colored lines lie black dotted lines representing the volume-averaged CO_2 saturation for all voxels in the clusters. % = percentage of voxels included in the cluster; ϕ = cluster-average porosity; k = cluster-average permeability; A = cluster-average linear trapping coefficient.

In each time series subplot in Figure 5, several metrics have been calculated for the clusters shown. They are the percentage of voxels included in the current cluster (%), the cluster-average porosity (ϕ, arithmetic mean), and permeability (k, geometric mean) values, as well as the linear trapping coefficient, A, computed from the cluster-average initial and residual CO_2 saturation values. The equation for A is as follows:

$$A=\frac{S_{CO_2,r}}{S_{CO_2,i}}, \tag{6}$$

where $S_{CO_2,i}$ is the initial CO_2 saturation reached at the last drainage scan and $S_{CO_2,r}$ is the residual CO_2 saturation reached at the last imbibition scan for the cluster. Having a high linear trapping coefficient is equivalent to having a high trapping ability, which means that a high percentage of the initial CO_2 in place can be trapped as residual CO_2 after imbibition. Both the linear and the Land (1968) trapping models are simple models that are commonly used in the CO_2 storage field (Pini & Benson, 2017). The Land's trapping coefficient is not included in this analysis as it tends to severely underestimate residual trapping at low initial CO_2 saturation. This problem is addressed in detail in Ni et al.'s (2019) supplementary material. Therefore, in this study we use the linear trapping coefficient to more accurately measure the trapping ability of different clusters.

4.1.1. Berea Sandstone

The Berea sandstone core has an average permeability of 200 md and laminations parallel to the flow direction. The maximum number of clusters for Berea is 4. By comparing the 3-D cluster index map in Figure 5a with the 3-D property maps in Figure 2, not surprisingly, the spatial features of the clusters are quite similar to the permeability map (Krause et al., 2013).

From Figure 5a, it can be seen that different clustered time series have distinctively different drainage and imbibition behaviors consistent with the illustrative example in Figure 4a. In Figure 5a, we observe that as the average permeability of the three clusters increases from left to right, the cluster-averaged coinjection drainage time series curves (red) become more concave downward, and the imbibition time series curves (blue) become more sloping downward. We can also notice that the blue imbibition curve for the cluster index = 1 after the initial decrease in saturation then increases slightly before decreasing again. It is unclear what the cause may be. One possible explanation is that there is a dynamic displacement and accumulation of CO_2 from upstream, which later gets displaced during subsequent imbibition stages. This is the only core where this pattern was observed. Finally, it can be seen that cluster-average trapping coefficients are very similar among the different clusters for Berea.

4.1.2. Shezaf Sandstone

Shezaf is an 80 md sandstone core sample with no obvious laminations, although it does have a high porosity, low permeability band that cuts through the middle of the core sample, as can be seen in the 3-D porosity and permeability maps in Figure 2 and the 3-D cluster index map in Figure 5b. The maximum number of clusters for Shezaf is 7. Shezaf is a highly heterogeneous core sample as can be seen from the range of the voxel-level CO_2 saturation time series curves (gray) in the "Whole Domain" subplot in Figure 5b.

From Figure 5b, we can see that the clustered CO_2 saturation time series curves behave very differently from each other and from the core-average CO_2 saturation behavior. Even though there is a greater diversity among the clusters, we can still observe the same time series concaving and sloping behaviors seen in the Berea core samples. Same as before, all CO_2 saturation time series behaviors closely correlate with the cluster's average permeability values. It can be clearly seen that clusters with higher permeability values achieve higher CO_2 saturation during drainage, which is consistent with the illustration in Figure 4a. We can also discern significant differences among the cluster-average trapping coefficients for the Shezaf sample. Generally low permeability clusters tend to have greater trapping abilities. This does not necessarily mean that low permeability clusters contribute to a larger fraction of the trapped gas volume, but rather, it only means that for these clusters, a higher percentage of the initial CO_2 saturation is residually trapped after imbibition. Because the low permeability portions of the rock do not achieve high saturations during drainage, these regions are not expected to contribute a substantial amount to trapping unless they create capillary barriers which increase trapping elsewhere in the core.

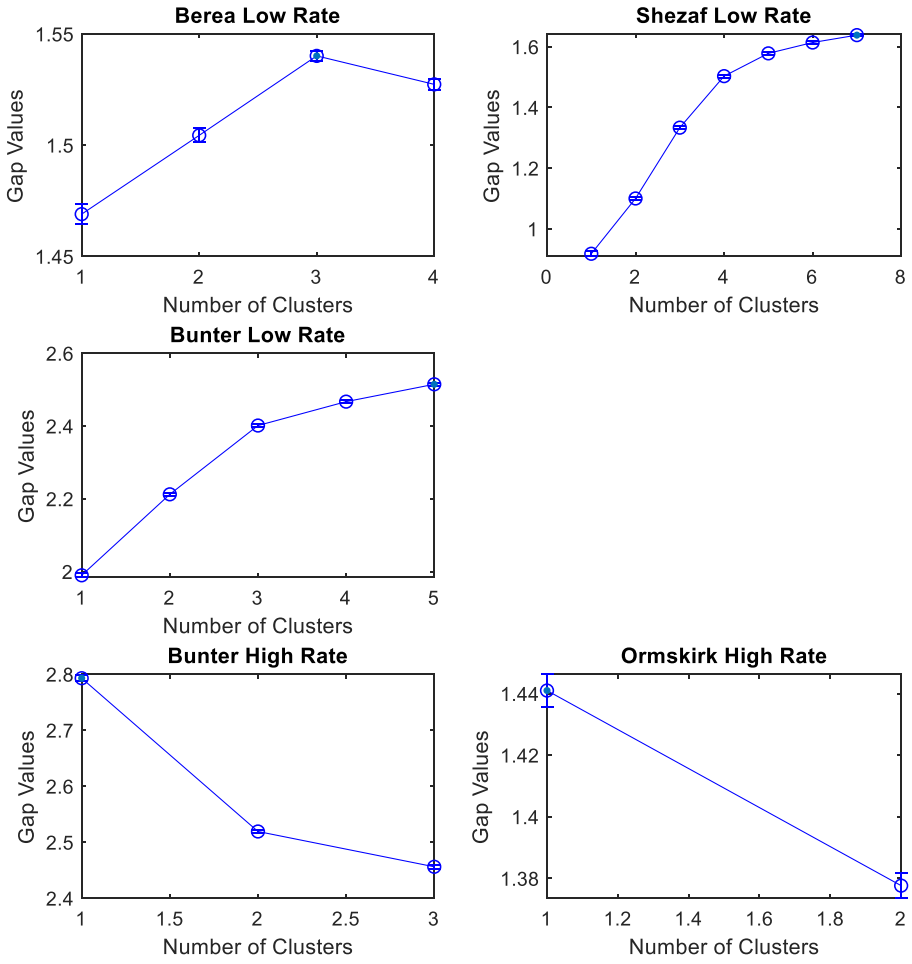

Figure 6. Gap values calculated for all clustering results. The uncertainty bounds represent one standard error. The blue dots inside the circles indicate the optimal numbers of clusters. The optimal numbers of clusters are as follows: Berea = 3; Shezaf = 7; Bunter low rate = 5; Bunter high rate = 1; and Ormskirk = 1.

4.1.3. Bunter Sandstone

The Bunter sandstone core sample has a permeability of 2.2 darcy and laminations perpendicular to the flow direction (Reynolds et al., 2018). The voxel-level permeability map is not computed for either the Bunter or the Ormskirk core sample, so the clustering procedure is done using only the 3-D porosity map to determine the maximum cluster number. For the Bunter sandstone core, two separate experiments have been conducted, a low and a high flow rate experiments. The clustering results for both of the Bunter experiments are shown in Figures 5c and 5d. For all the Bunter and the Ormskirk experiments, the last imbibition stage has nonzero CO_2 fractional flow values that are close to zero, so the calculated trapping coefficient values, though approximate, should still be valid proxies of the clusters' true trapping abilities.

From the low flow rate clustering results in Figure 5c, again, we observe the familiar concaving behaviors in both the coinjection drainage (red) and the imbibition (blue) cluster-average CO_2 saturation time series curves. Voxel-level permeability values are unavailable for Bunter; however, based on the clustering results of the previously analyzed cores, we can infer the values. The cluster with a concave upward red curve, low drainage CO_2 saturation, an almost flat blue curve, and a high trapping ability should have a high capillary entry pressure and therefore a low permeability value. Conversely, clusters with the opposite behaviors then are likely to have low capillary entry pressures and therefore high permeability values. The Bunter low flow rate clustering results are also similar to the Shezaf results in that both core samples are highly heterogeneous and significant differences can be observed for the trapping abilities of the various clusters.

From the high flow rate clustering results in Figure 5d, we can see that unlike the clustering results on the previously analyzed core samples, the cluster-average CO_2 saturation time series curves here are all highly similar among the different clusters. We also no longer observe the same concaving behaviors as before. Additionally, the maximum number of clusters is only 3 for the Bunter high rate data. As a comparison, the Bunter low flow rate clustering results have a higher maximum cluster number of 5. This is because when the flow rate is low, capillary forces dominate and push the domain toward capillary equilibrium. When the flow rate is high, viscous forces dominate and achieve a more homogeneous CO_2 saturation field with a highly heterogeneous capillary pressure field (Shi et al., 2011b; Reynolds & Krevor, 2015). By comparing the two Bunter experiments in Figures 5c and 5d, it is easy to distinguish between the two flow regimes based on the clustering results as the viscous-dominated case demonstrates homogeneous CO_2 saturation time series behaviors and the capillary-dominated case demonstrates heterogeneous CO_2 saturation time series behaviors.

Comparison between the trapping coefficients for the high and low flow rate experiments demonstrates that trapping coefficients are higher for the capillary-dominated flow regimes. This is true for all of the clusters. For example, trapping coefficients for the viscous-dominated regime range from 0.47 to 0.51 (core-averaged value of 0.49), while they range from 0.59 to 0.78 (average value of 0.62) for the capillary-dominated regime. This is because as the ratio between viscous forces and capillary forces increases beyond a certain critical threshold, capillary desaturation happens (Lake et al., 2014).

4.1.4. Ormskirk Sandstone

The Ormskirk sandstone core has a permeability of 12.1 darcy and laminations perpendicular to the flow direction. For this core sample, the clustering procedure is only done on the high flow rate experimental data. Figure 5e shows the clustering results for the case with the maximum cluster number of 2. Similar to the Bunter high flow rate clustering results, Figure 5e demonstrates the typical viscous-dominated case where the CO_2 saturation field is highly homogeneous.

4.1.5. How to Use the Gap Value to Differentiate Flow Regimes

As illustrated by the clustering results on both the Bunter high flow rate and low flow rate experimental data, the flow regime has a significant influence on the CO_2 saturation time series behaviors. We would expect the viscous-dominated case to have a homogeneous CO_2 saturation field and the capillary-dominated case to have a heterogeneous CO_2 saturation field because capillary equilibrium is approached. An easy way to differentiate between the two flow regimes is to calculate the optimal number of clusters with the gap value based on the results from K-means clustering. When plotting an error measurement, such as the total distance to centroids, against an increasing number of clusters, the error always reduces due to more flexibility present. However, often times, the error reduces much more rapidly at some small numbers of clusters, and the error curve therefore forms an "L" shape. The gap statistic can identify the optimal number of clusters that is located at the "elbow" of the error curve. The optimum occurs at the smallest number of clusters that is within one standard error of the largest gap value (James et al., 2013; Tibshirani et al., 2001).

When examining the gap values computed for all the previous clustering results in Figure 6, one can easily distinguish between the capillary-dominated cases (Berea, Shezaf, and Bunter low rate) and the viscous-dominated cases (Bunter high rate and Ormskirk) based on both the maximum number of clusters and the gap value trends. The maximum number of clusters is greater for the capillary-dominated cases because their CO_2 saturation fields are more heterogeneous so that more subgroups within the data can be found. The capillary-dominated cases also all have an increasing gap value trend with the largest gap value occurring at a high number of clusters, whereas the two viscous-dominated cases have a decreasing gap value trend with a single cluster identified as the optimal number of clusters, indicating that no obvious subgroups can be found in the data because the CO_2 saturation field is too homogeneous.

This analysis result indicates that out of the four different capillary numbers we have previously investigated in Table 2, only $N_{c,2}$ correctly identifies the low rate experiments to be capillary dominated and the high rate experiments to be viscous dominated. $N_{c,1}$ correctly identifies the capillary-dominated experiments but not the viscous-dominated ones. $N_{c,3}$ is hard to generalize to different types of sandstone as its threshold value for judging capillary-dominated flow regimes is heterogeneity dependent. Lastly, $N_{c,4}$ is inconsistent for both flow regimes.

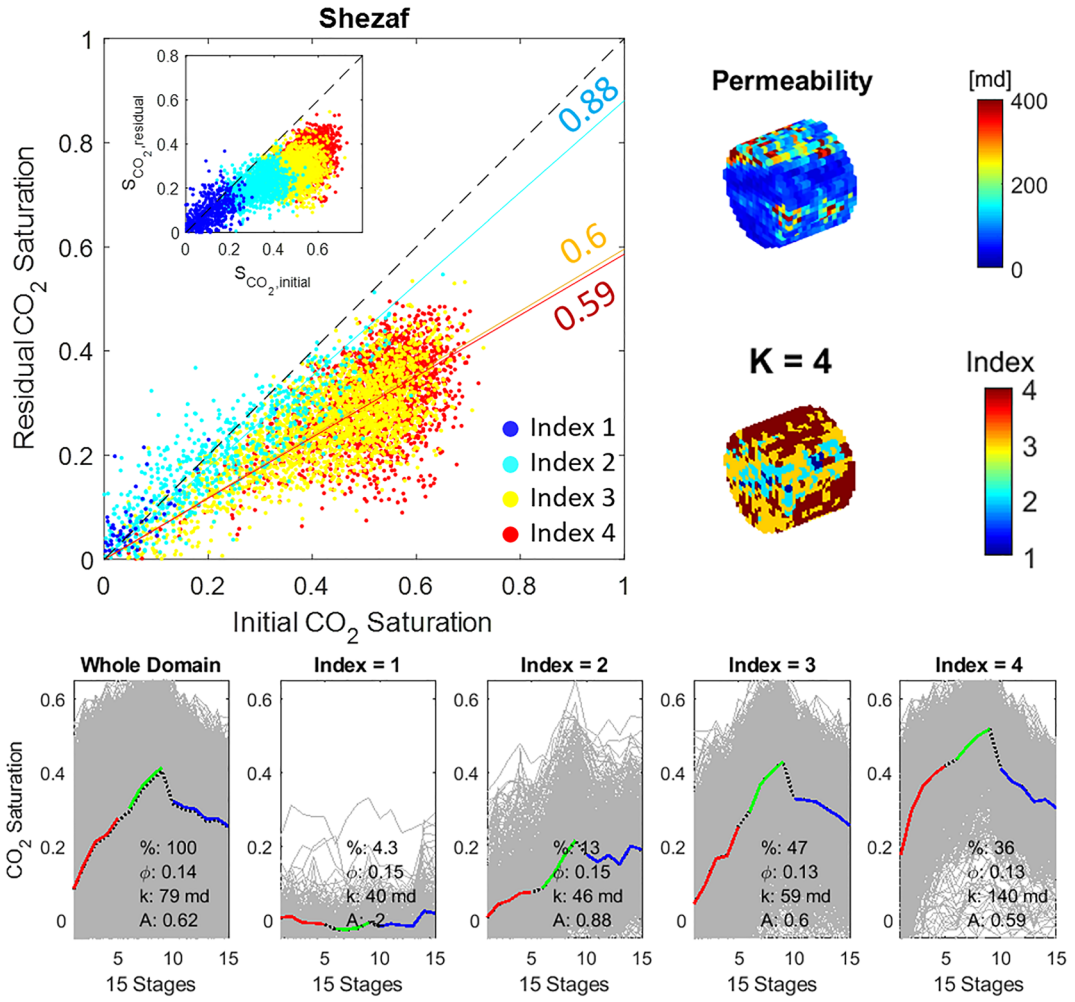

Figure 7. Results from hierarchical clustering performed on the Shezaf experiment to characterize residual CO_2 trapping. (top left) Large IR plot with hierarchical clustering K = 4 scenario superimposed. Black dashed line = 100% trapping line. Colored lines = cluster-average linear trapping relationships. (inset) K-means clustering K = 4 scenario superimposed on the same IR plot. (top right) 3-D permeability map and the cluster index map. (bottom row) K = 4 time series clustering results. For a detailed figure description, see Figure 5.

4.2. Use Hierarchical Clustering to Characterize Residual Trapping

4.2.1. Shezaf Sandstone

For the Shezaf experimental data, the hierarchical clustering procedure has also been applied in order to characterize residual trapping. The results are shown in Figure 7. By comparing the larger IR plot with the inset IR plot in Figure 7, we can see that the hierarchical clustering method used for characterizing trapping produces drastically different clusters as compared to the K-means clustering method used to characterize flow. Furthermore, it can be observed that although cluster index = 2 has many low permeability voxels with low initial CO_2 saturation that are simply following the expected capillary equilibrium behaviors, the cluster also includes many voxels with relatively high initial CO_2 saturation and has a much higher linear trapping coefficient than the core-average value. This cluster exhibits capillary heterogeneity trapping behaviors consistent with the medium-permeability group as shown in the previous illustration in Figure 4b.

4.2.2. Bunter Sandstone

The results of applying the hierarchical clustering method on the Bunter low rate experimental data to characterize residual trapping are shown in Figure 8. Again, we see how different the results are from the different clustering methods by comparing the large and the inset IR plots. In Figure 8, the clustering results for the Bunter data clearly show that the regions of the core with low porosity (cluster index = 1, likely with high

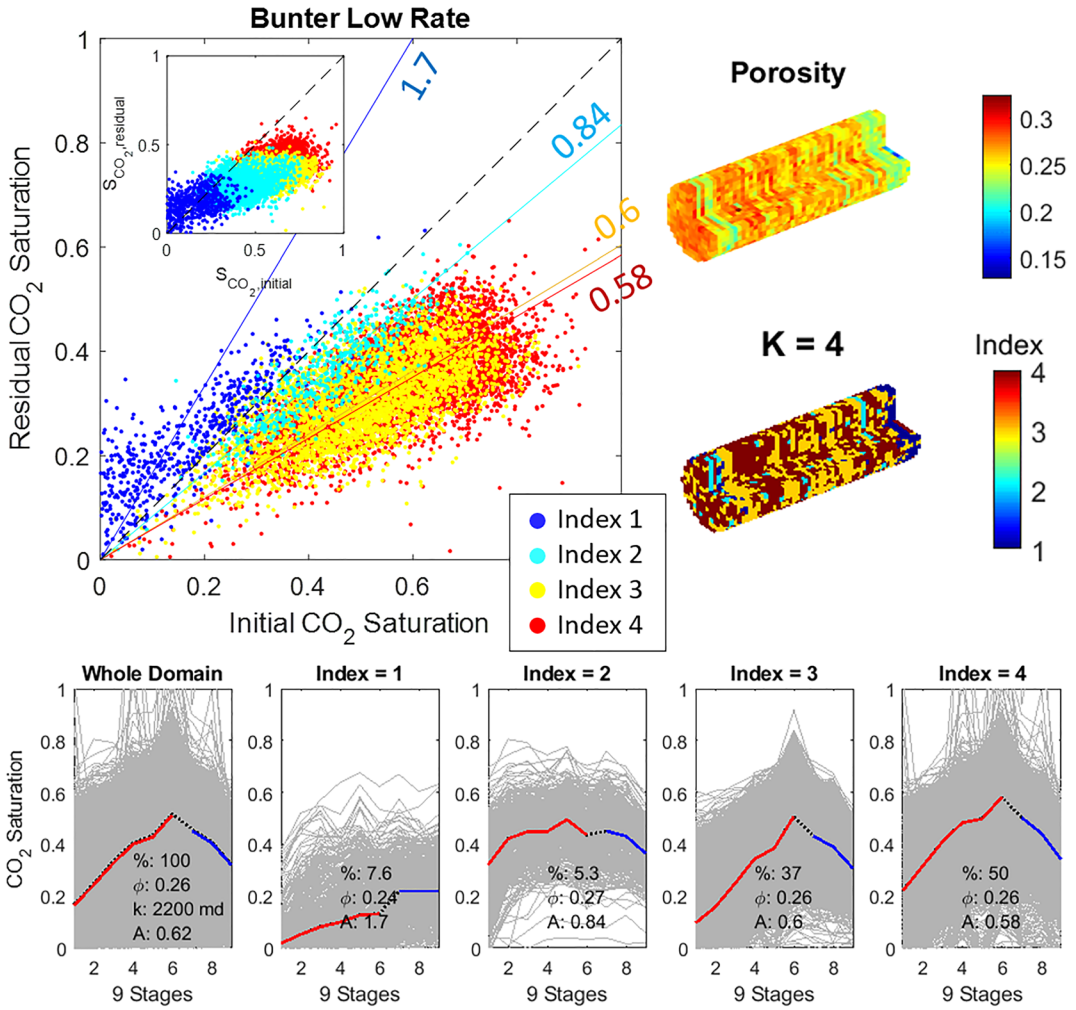

Figure 8. Results from hierarchical clustering performed on the Bunter low rate experiment to characterize residual CO_2 trapping. (top left) Large IR plot with hierarchical clustering K = 4 scenario superimposed. Black dashed line = 100% trapping line. Colored lines = cluster average linear trapping relationships. (inset) K-means clustering K = 4 scenario superimposed on the same IR plot. (top right) 3-D porosity map and the cluster index map. (bottom row) K = 4 time series clustering results. For a detailed figure description, see Figure 5.

capillary entry pressure and low permeability) trap above 100% of the initial CO_2 saturation that is present at the end of global drainage. We can clearly see that almost all voxels within the index = 1 cluster (deep blue) in the IR plot show above 100% trapping. While some may argue that the initial CO_2 saturation values plotted in an IR plot should strictly be the highest CO_2 saturation achieved during the entire experiment so that no voxels would have above 100% trapping, here we choose initial CO_2 saturation to be the CO_2 saturation field achieved after the last global drainage stage to demonstrate the fact that capillary barriers, unlike the rest of core, may reach maximum CO_2 saturation during global imbibition.

From Figure 8, it can be clearly seen that cluster index = 2 has high initial CO_2 saturation (likely also high permeability) and, at the same time, has a much higher linear trapping coefficient than clusters with index = 3 and 4 which follow the core-average linear trapping model. The trapping behavior demonstrated by this cluster is consistent with the capillary heterogeneity trapping behavior of the high-permeability group shown in the previous illustration in Figure 4b. As indicated by the 3-D cluster index map and the 3-D porosity map, two main regions displaying such capillary heterogeneity trapping characteristics are both situated upstream of low porosity layers that are likely capillary barriers.

4.2.3. Capillary-Dominated Imbibition and Capillary Heterogeneity Trapping Behaviors

It has been previously hypothesized that the global imbibition process is a complex phenomenon involving both local drainage and local imbibition at the same time (Ni et al., 2019). Evidence for this comes from the increase in CO_2 saturation in some voxels during water injection. Now this hypothesis has been verified by applying the hierarchical clustering analysis to the Bunter low flow rate experimental data. For cluster index = 1 in Figure 8, it appears that due to the high capillary entry pressure of the low porosity layers, the imbibing water does not have enough forces to push the upstream CO_2 all the way through these capillary barriers. As a result, upstream CO_2 that has been swept toward these layers ends up been trapped within such layers. This is another form of capillary heterogeneity trapping, one that happens when the capillary barriers themselves are thick enough to trap more CO_2 than expected based on capillary equilibrium theories.

Both Figures 7 and 8 show that the cluster with index = 2 has a significantly higher linear trapping coefficient than the core-average coefficients. These clusters are also associated with capillary heterogeneity trapping as previously illustrated. From the cluster index = 2 subplot in the bottom rows of Figures 7 and 8, we can see that although voxels in this cluster reach different initial CO_2 saturation values during drainage, they invariably lose little CO_2 during imbibition due to the presence of capillary barriers. Note that when the same method is applied to the Berea experimental data, no clusters with significantly higher trapping coefficients are observed. This result indicates that the presented hierarchical clustering method works best with highly heterogeneous core samples in identifying capillary heterogeneity trapping regions. This result also implies that for certain sandstone types, the contribution of the capillary heterogeneity trapping mechanism to the total amount of residually trapped CO_2 may be less than we previously thought (Jackson et al., 2019; Ni et al., 2019). While for some sandstone types such as the Mount Simon sandstone, the contribution of capillary heterogeneity trapping can be as high as 44% (Krevor et al., 2011, 2012; Ni et al., 2019), for others such as the Berea sandstone, this analysis suggests that the contribution of capillary heterogeneity trapping may be too small to detect using the clustering methods developed here.

5. Conclusions

In this study, we have developed a novel method based on time series clustering and hypothesis testing to analyze coreflooding CO_2 saturation time series data. The method enables us to visualize and analyze how mesoscale petrophysical properties affect capillary-dominated flow and residual trapping behaviors. By applying the novel method to a variety of sandstone core samples, we have obtained the following conclusions.

1. Under capillary-dominated flow regimes, during CO_2/water coinjection drainage, the CO_2 saturation time series are concave downward for high permeability clusters and are concave upward for low permeability clusters. During imbibition, high permeability clusters have downtrending CO_2 saturation time series curves, whereas low permeability clusters display a flat line. Leverett J-function scaling combined with the assumption of capillary equilibrium can explain these CO_2 saturation time series behaviors during both drainage and imbibition.
2. The CO_2 saturation time series behaviors are only distinctive during capillary-dominated flow regimes. When the flow regime is viscous dominated, the differences in time series behaviors among the different clusters are greatly diminished. We present a criterion for distinguishing between these flow regimes based on how the gap value changes with an increasing number of clusters.
3. Global imbibition involves both local drainage and local imbibition events, as illustrated by the low porosity (likely low permeability) regions in the Bunter core where their CO_2 saturation can increase during global imbibition due to trapping of displaced CO_2 from upstream. This form of capillary heterogeneity trapping traps extra CO_2 within the capillary barriers themselves.
4. We can also identify regions with typical capillary heterogeneity trapping using the clustering results. These are regions with high initial CO_2 saturation that can achieve higher than expected trapping due to capillary heterogeneity. Such regions can be identified for both Shezaf and Bunter sandstone cores.

Conflict of Interest

The authors have no known financial conflicts of interest.

Data Availability Statement

Data sets for this study can be found in the data repository of Ni (2020) and the paper of Reynolds et al. (2018).

Acknowledgments
This work was supported by the Stanford University's Global Climate & Energy Project and the GeoCquest project funded by BHP.

References

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering—A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Akbarabadi, M., & Piri, M. (2013). Relative permeability hysteresis and capillary trapping characteristics of supercritical CO_2/brine systems: An experimental study at reservoir conditions. *Advances in Water Resources*, *52*, 190–206. https://doi.org/10.1016/j.advwatres.2012.06.014

Akbarabadi, M., & Piri, M. (2015). Co-sequestration of SO_2 with supercritical CO_2 in carbonates: An experimental study of capillary trapping, relative permeability, and capillary pressure. *Advances in Water Resources*, *77*, 44–56. https://doi.org/10.1016/j.advwatres.2014.08.011

Akin, S., & Kovscek, A. R. (2003). Computed tomography in petroleum engineering research. *Geological Society of London, Special Publication*, *215*(1), 23–38. https://doi.org/10.1144/GSL.SP.2003.215.01.03

Albalate, A., & Minker, W. (2013). *Semi-supervised and unsupervised machine learning*. London, UK and Hoboken, NJ: ISTE Ltd and John Wiley & Sons, Inc. https://doi.org/10.1002/9781118557693

Anderson, T. W., & Darling, D. A. (2010). A test of foodness of fit. *Journal of the American Statistical Association*, *49*, 765–769.

Batagelj, V. (1988). Generalized Ward and related clustering problems. In H. H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 67–74). North-Holland, Amsterdam: Elsevier Science Publishers. http://www.academia.edu/download/31188171/ward.pdf

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Chang, J., & Yortsos, Y. C. (1992). Effect of capillary heterogeneity on Buckley-Leverett displacement. *SPE Reservoir Engineering*, *7*(02), 285–293. https://doi.org/10.2118/18798-PA

Chatzis, I., & Morrow, N. R. (1984). Correlation of capillary number relationships for sandstone. *Society of Petroleum Engineers Journal*, *24*(05), 555–562. https://doi.org/10.2118/10114-PA

Corey, A. T., & Rathjens, C. H. (1956). Effect of stratification on relative permeability. *Journal of Petroleum Technology*, *8*(12), 69–71. https://doi.org/10.2118/744-G

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Geher, G., & Hall, S. (2014). *Straightforward statistics: Understanding the tools of research*. Oxford: Oxford University Press.

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference*. Boca Raton, FL: Chapman & Hall/CRC Press, Taylor & Francis Group.

Graue, A. (1994). Imaging the effects of capillary heterogeneities on local saturation development in long corefloods. *SPE Drilling and Completion*, *9*(01), 57–64. https://doi.org/10.2118/21101-PA

Hilfer, R., Armstrong, R. T., Berg, S., Georgiadis, A., & Ott, H. (2015). Capillary saturation and desaturation. *Physical Review E*, *92*(6), 063023. https://doi.org/10.1103/PhysRevE.92.063023

Hingerl, F. F., Yang, F., Pini, R., Xiao, X., Toney, M. F., Liu, Y., & Benson, S. M. (2016). Characterization of heterogeneity in the Heletz sandstone from core to pore scale and quantification of its impact on multi-phase flow. *International Journal of Greenhouse Gas Control*, *48*, 69–83. https://doi.org/10.1016/j.ijggc.2015.12.037

Honarpour, M. M., & Saad, N. (1994). Influence of small-scale rock laminations on core plug oil/water relative permeability and capillary pressure, in: University of Tulsa Centennial Petroleum Engineering Symposium. Society of Petroleum Engineers, pp. 117–127. https://doi.org/10.2118/27968-MS

Huang, Y., Ringrose, P. S., & Sorbie, K. S. (1995). Capillary trapping mechanisms in water-wet laminated rocks. *SPE Reservoir Engineering*, *10*(04), 287–292. https://doi.org/10.2118/28942-PA

Huppler, J. D. (1970). Numerical investigation of the effects of core heterogeneities on waterflood relative permeabilities. *Society of Petroleum Engineers Journal*, *10*(04), 381–392. https://doi.org/10.2118/2874-PA

IPCC (2005). *IPCC special report on carbon dioxide capture and storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change. IPCC Special Report on Carbon Dioxide Capture and Storage*. Cambridge, United Kingdom and NewYork, NY, USA: Cambridge University Press.

IPCC (2014). *Climate change 2014: Mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jackson, S. J., & Krevor, S. (2019). Characterization of hysteretic multiphase flow from the mm to m scale in heterogeneous rocks. E3S Web Conf. 89, 02001. https://doi.org/10.1051/e3sconf/20198902001

Jackson, S. J., Lin, Q., Krevor, S. (2019). Representative elementary volumes, hysteresis and heterogeneity in multiphase flow from the pore to continuum scale. Submitt. available EarthArXiv 1–33.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*, *Springer texts in statistics*. New York, NY: Springer. https://doi.org/10.1007/978-1-4614-7138-7

Kaufman, L., & Rousseuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis* (Vol. 47, p. 788). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.2307/2532178

Kortekaas, T. (1985). Water/oil displacement characteristics in crossbedded reservoir zones. *Society of Petroleum Engineers Journal*, *25*(06), 917–926. https://doi.org/10.2118/12112-PA

Krause, M., Krevor, S., & Benson, S. M. (2013). A procedure for the accurate determination of sub-core scale permeability distributions with error quantification. *Transport in Porous Media*, *98*(3), 565–588. https://doi.org/10.1007/s11242-013-0161-y

Krause, M. H. (2012a). Modeling and investigation of the influence of capillary heterogeneity on multiphase flow of CO_2 and brine. Stanford University.

Krause, M. H., & Benson, S. M. (2015). Accurate determination of characteristic relative permeability curves. *Advances in Water Resources*, *83*, 376–388. https://doi.org/10.1016/j.advwatres.2015.07.009

Krause, M. H. H. (2012b). Modeling and investigation of the influence of capillary heterogeneity on relative permeability. SPE 160909-STU 8–10. https://doi.org/10.2118/160909-STU

Krevor, S., Blunt, M. J., Benson, S. M., Pentland, C. H., Reynolds, C., Al-Menhali, A., & Niu, B. (2015). Capillary trapping for geologic carbon dioxide storage—From pore scale physics to field scale implications. *International Journal of Greenhouse Gas Control*, *40*, 221–237. https://doi.org/10.1016/j.ijggc.2015.04.006

Krevor, S. C. M., Pini, R., Li, B., & Benson, S. M. (2011). Capillary heterogeneity trapping of CO_2 in a sandstone rock at reservoir conditions. *Geophysical Research Letters*, *38*, L15401. https://doi.org/10.1029/2011GL048239

Krevor, S. C. M., Pini, R., Zuo, L., & Benson, S. M. (2012). Relative permeability and trapping of CO_2 and water in sandstone rocks at reservoir conditions. *Water Resources Research*, *48*, W02532. https://doi.org/10.1029/2011WR010859

Kuo, C.-W., & Benson, S. M. (2015). Numerical and analytical study of effects of small scale heterogeneity on CO_2/brine multiphase flow system in horizontal corefloods. *Advances in Water Resources*, *79*, 1–17. https://doi.org/10.1016/j.advwatres.2015.01.012

Kyan, M., Muneesawang, P., Jarrah, K., & Guan, L. (2014). *Unsupervised learning: A dynamic approach*. Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118875568

Lake, L. W., John, R. T., Rossen, W. R., & Pope, G. A. (2014). *Fundamentals of enhanced oil recovery* (2nd ed.). Richardson, TX: Society of Petroleum Engineers.

Land, C. (1968). Calculation of imbibition relative permeability for two- and three-phase flow from rock properties. *SPE Journal*, *8*(02), 149–156. https://doi.org/10.2118/1942-PA

Leverett, M. C. (1941). Capillary behavior in porous solids. *Transactions of AIME*, *142*(01), 152–169. https://doi.org/10.2118/941152-G

Li, B., & Benson, S. M. (2015). Influence of small-scale heterogeneity on upward CO_2 plume migration in storage aquifers. *Advances in Water Resources*, *83*, 389–404. https://doi.org/10.1016/j.advwatres.2015.07.010

Liao, T. W. (2005). Clustering of time series data—A survey. *Pattern Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

McClave, J., & Sincich, T. (2017). *Statistics* (13th ed.). New York, NY: Pearson.

Meckel, T. A., Bryant, S. L., & Ganesh, P. R. (2015). Characterization and prediction of CO_2 saturation resulting from modeling buoyant fluid migration in 2D heterogeneous geologic fabrics. *International Journal of Greenhouse Gas Control*, *34*, 85–96. https://doi.org/10.1016/j.ijggc.2014.12.010

Morrow, N. R., Chatzis, I., & Taber, J. J. (1988). Entrapment and mobilization of residual oil in bead packs. *SPE Reservoir Engineering*, *3*(03), 927–934. https://doi.org/10.2118/14423-PA

Ni, H. (2020). Data for: Using unsupervised machine learning to characterize capillary flow and residual trapping. https://doi.org/10.17632/9rtxgrv9db.2

Ni, H., Boon, M., Garing, C., & Benson, S. M. (2018). Effects of correlation length and lamination direction on CO_2 residual trapping ability for different sandstone types, in: 14th International Conference on Greenhouse Gas Control Technologies. pp. 1–9.

Ni, H., Boon, M., Garing, C., & Benson, S. M. (2019). Predicting CO_2 residual trapping ability based on experimental petrophysical properties for different sandstone types. *International Journal of Greenhouse Gas Control*, *86*, 158–176. https://doi.org/10.1016/j.ijggc.2019.04.024

NIST/SEMATECH (2013). Bartlett's Test [WWW Document]. E-handb. Stat. Methods. URL https://www.itl.nist.gov/div898/handbook/eda/section3/eda357.htm (accessed 2.19.19).

Niu, B., Al-Menhali, A., & Krevor, S. C. (2015). The impact of reservoir conditions on the residual trapping of carbon dioxide in Berea sandstone. *Water Resources Research*, *51*, 2009–2029. https://doi.org/10.1002/2014WR016441

Pentland, C. H., El-Maghraby, R., Iglauer, S., & Blunt, M. J. (2011). Measurements of the capillary trapping of super-critical carbon dioxide in Berea sandstone. *Geophysical Research Letters*, *38*, L06401. https://doi.org/10.1029/2011GL046683

Pini, R., & Benson, S. M. (2013). Simultaneous determination of capillary pressure and relative permeability curves from core-flooding experiments with various fluid pairs. *Water Resources Research*, *49*, 3516–3530. https://doi.org/10.1002/wrcr.20274

Pini, R., & Benson, S. M. (2017). Capillary pressure heterogeneity and hysteresis for the supercritical CO_2/water system in a sandstone. *Advances in Water Resources*, *108*, 277–292. https://doi.org/10.1016/j.advwatres.2017.08.011

Rani, S. (2012). *Handbook of basic statistical concepts: For scientists and pharmacists*. Oxford: Alpha Science International.

Reynolds, C. A., Blunt, M. J., & Krevor, S. (2018). Multiphase flow characteristics of heterogeneous rocks from CO_2 storage reservoirs in the United Kingdom. *Water Resources Research*, *54*, 729–745. https://doi.org/10.1002/2017WR021651

Reynolds, C. A., & Krevor, S. (2015). Characterizing flow behavior for gas injection: Relative permeability of CO_2-brine and N_2-water in heterogeneous rocks. *Water Resources Research*, *51*, 9464–9489. https://doi.org/10.1002/2015WR018046

Roelofsen, P. (2018). Master thesis business analytics time series clustering.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., et al. (2018). Mitigation pathways compatible with 1.5°C in the context of sustainable development. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, et al. (Eds.), *Global warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change* (pp. 93–174).

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *29*(4), 97–104. https://doi.org/10.1177/003754977702900403

Ruprecht, C., Pini, R., Falta, R., Benson, S., & Murdoch, L. (2014). Hysteretic trapping and relative permeability of CO_2 in sandstone at reservoir conditions. *International Journal of Greenhouse Gas Control*, *27*, 15–27. https://doi.org/10.1016/j.ijggc.2014.05.003

Saad, N., Cullick, A. S., & Honarpour, M. M. (1995). Immiscible displacement mechanisms and scale-up in the presence of small-scale heterogeneities, in: SPE annual technical conference and exhibition. Society of Petroleum Engineers, pp. 795–804. https://doi.org/10.2523/30779-MS

Saadatpoor, E., Bryant, S. L., & Sepehrnoori, K. (2010). New trapping mechanism in carbon sequestration. *Transport in Porous Media*, *82*(1), 3–17. https://doi.org/10.1007/s11242-009-9446-6

Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). New York: CRC Press.

Shi, J.-Q., Xue, Z., & Durucan, S. (2011a). Supercritical CO_2 core flooding and imbibition in Tako sandstone—Influence of sub-core scale heterogeneity. *International Journal of Greenhouse Gas Control*, *5*(1), 75–87. https://doi.org/10.1016/j.ijggc.2010.07.003

Shi, J.-Q., Xue, Z., & Durucan, S. (2011b). Supercritical CO_2 core flooding and imbibition in Berea sandstone—CT imaging and numerical simulation, in: GHGT-10. *Energy Procedia*, 5001–5008. https://doi.org/10.1016/j.egypro.2011.02.471

Suekane, T., Zhou, N., Hosokawa, T., & Matsumoto, T. (2010). Direct observation of trapped gas bubbles by capillarity in sandy porous media. *Transport in Porous Media*, *82*(1), 111–122. https://doi.org/10.1007/s11242-009-9439-5

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *78*(8), 1073–1098. https://doi.org/10.1016/j.scico.2012.08.004

Trevisan, L., Pini, R., Cihan, A., Birkholzer, J. T., Zhou, Q., González-Nicolás, A., & Illangasekare, T. H. (2017). Imaging and quantification of spreading and trapping of carbon dioxide in saline aquifers using meter-scale laboratory experiments. *Water Resources Research*, *53*, 485–502. https://doi.org/10.1002/2016WR019749

Virnovsky, G. A., Friis, H. A., & Lohne, A. (2004). A steady-state Upscaling approach for immiscible two-phase flow. *Transport in Porous Media*, *54*(2), 167–192. https://doi.org/10.1023/A:1026363132351

Vizika, O., & Payatakes, A. C. (1989). Parametric experimental study of forced imbibition in porous media. *PhysicocChemical Hydrodynamics*, *11*, 187–204.

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Yokoyama, Y., & Lake, L. (1981). The effects of capillary pressure on immiscible displacements in stratified porous media. https://doi.org/10.2118/10109-MS

Zahasky, C., & Benson, S. M. (2016). Evaluation of hydraulic controls for leakage intervention in carbon storage reservoirs. *International Journal of Greenhouse Gas Control*, *47*, 86–100. https://doi.org/10.1016/j.ijggc.2016.01.035

Zhou, D., Fayers, F. J., & Orr, F. M. Jr. (1994). Scaling of multiphase flow in simple heterogeneous porous media. *SPE Reservoir Engineering*, *12*(03), 173–178. https://doi.org/10.2118/27833-PA

Zhou, N., Matsumoto, T., Hosokawa, T., & Suekane, T. (2010). Pore-scale visualization of gas trapping in porous media by X-ray CT scanning. *Flow Measurement and Instrumentation*, *21*(3), 262–267. https://doi.org/10.1016/j.flowmeasinst.2010.05.002

Zuo, L., & Benson, S. M. (2014). Process-dependent residual trapping of CO_2 in sandstone. *Geophysical Research Letters*, *41*(8), 2820–2826. https://doi.org/10.1002/2014GL059653.Received

References From the Supporting Information

Chiquet, P., Daridon, J.-L., Broseta, D., & Thibeau, S. (2007). CO_2/water interfacial tensions under pressure and temperature conditions of CO_2 geological storage. *Energy Conversion and Management*, *48*(3), 736–744. https://doi.org/10.1016/j.enconman.2006.09.011

Georgiadis, A., Maitland, G., Trusler, J. P. M., & Bismarck, A. (2010). Interfacial tension measurements of the (H_2O + CO_2) system at elevated pressures and temperatures. *Journal of Chemical & Engineering Data*, *55*(10), 4168–4175. https://doi.org/10.1021/je100198g

Lemmon, E. W., McLinden, M. O., & Friend, D. G. (2018). Thermophysical properties of fluid systems [WWW Document]. NIST Chem. WebBook, NIST Stand. Ref. Database Number 69. https://doi.org/10.18434/T4D303

Steinbach, M., Ertöz, L., & Kumar, V. (2004). The challenges of clustering high dimensional data. In L. T. Wille (Ed.), *New directions in statistical physics* (pp. 273–309). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-08968-2_16

The MathWorks Inc. (2018). Hierarchical clustering [WWW Document]. MATLAB & Simulink.

The MathWorks Inc. (2006). k-means clustering [WWW Document]. MATLAB Doc. URL https://www.mathworks.com/help/stats/kmeans.html#bues3lh (accessed 2.14.19).