ROBUST METRICS OF CONNECTIVITY

Yaokun Wu[1], Siddharth Misra[1], and Rui Liu[1]

[1]Texas A&M University

November 30, 2022

Abstract

Connectivity of material constituents govern the transport, mechanical, chemical, thermal, and electromagnetic properties. Energy storage, recovery and conversion depends on connectivity of material constituents. High-resolution microscopy image of a material captures the microstructural aspects describing the distribution, topology and morphology of various material constituents. In this study, six metrics are developed and tested for quantifying the connectivity of material constituents as captured in the high-resolution microscopy images. The six metrics are as follows: geobody connectivity metric based on percolation theory, Euler number based on integral geometry, indicator variogram based on geostatistics, two-point cluster function, connectivity function, and travel-time histogram based on fast marching method. The performances of these metrics are tested on 3000 images representing six levels of connectivity. The metrics are also evaluated on the organic constituent captured in the scanning electron microscopy (SEM) images of organic-rich shale samples. The connectivity function and travel-time histogram based on fast-marching method are the most robust and reliable metrics. Material constituents exhibiting high connectivity result in large values of average travel time computed using fast-marching method and average connected distance computed using connectivity function. The proposed metrics will standardize and speed-up the analysis of connectivity to facilitate the characterization of properties and processes of energy-relevant materials.

ROBUST METRICS OF CONNECTIVITY

Yaokun Wu, Dr. Siddharth Misra, Rui Liu

Texas A&M University

For any queries contact Dr. Siddharth Misra, Texas A&M University

Abstract

Connectivity of material constituents govern the transport, mechanical, chemical, thermal, and electromagnetic properties. Energy storage, recovery and conversion depends on connectivity of material constituents. High-resolution microscopy image of a material captures the microstructural aspects describing the distribution, topology and morphology of various material constituents. In this study, six metrics are developed and tested for quantifying the connectivity of material constituents as captured in the high-resolution microscopy images. The six metrics are as follows: geobody connectivity metric based on percolation theory, Euler number based on integral geometry, indicator variogram based on geostatistics, two-point cluster function, connectivity function, and travel-time histogram based on fast marching method. The performances of these metrics are tested on 3000 images representing six levels of connectivity. The metrics are also evaluated on the organic constituent captured in the scanning electron microscopy (SEM) images of organic-rich shale samples. The connectivity function and travel-time histogram based on fast-marching method are the most robust and reliable metrics. Material constituents exhibiting high connectivity result in large values of average travel time computed using fast-marching method and average connected distance computed using connectivity function. The proposed metrics will standardize and speed-up the analysis of connectivity to facilitate the characterization of properties and processes of energy-relevant materials.

1 Introduction

Connectivity of a material constituent represents the extent of connectedness depending on the scale of observation ranging from local to global connectedness. In this paper, connectivity describes the topology of connected spaces in material. Connectivity of constituents influence various physical properties, such as electromagnetic, chemical, transport, thermal, and mechanical properties. Connectivity is an important parameter for engineering/scientific studies involving piezoelectric, cement, porous material, brain neuronal networks, ceramics, hydrology, geophysics, and thermoelastic, to name a few. Definition and use of connectivity vary across disciplines. For example, the transfer of sediment from one zone or location to another is defined as the kilometer-scale connectivity in geomorphology, where the connectivity is important for understanding the linkages between river reaches, the influence of sediment sources on channel morphology and the mechanisms of morphological change [1]. In hydrological literature, the hydrological connectivity is related to the physical connection between waterbodies that contributes to the spatial heterogeneity of riverine flood plains across hundreds of meters [2]. In geoscience, the kilometer-scale connectivity is defined as the degree to which subsurface geobodies are connected that governs the transport characteristics [3,4]. Connectivity of an aquifer in the subsurface at kilometer scale influences groundwater flow [5]. The connectivity of catchment and landscape is considered in geomorphology for environmental management [6]. The connectivity of habitats is utilized in spatial ecology for colonization event prediction [7]. At the small scale, the connectivity of pores based on multiple-point statistics facilitates the reconstruction of porous media. [8]. The

connectivity of soil structure helps in soil surface topography detection, as well as morphology and percolation analyses [9].

Connectivity as a physical parameter is relevant and important to various disciplines because of its influence on several physical and chemical properties. Our focus is on the micro-scale/pore-scale connectivity of constituents. Various metrics for connectivity quantification have been proposed in the last 50 years. Few metrics are developed based on percolation theory, which denotes the transition from disconnected clusters to a large spanning cluster [10]. The proportion of percolated cluster is found to be one way of measuring connectivity based on the percolation theory. Percolation-based metrics have been implemented in quantifying the connectivity of pore network [11]. In kilometer-scale subsurface characterization, percolation-based metrics have been used to quantify the connectivity of geobodies [12]. Another concept from integral geometry, such as Euler characteristic, have been used to quantify the shape and structure of topological components [13]. The Euler characteristic has been used as the metric to quantify connectivity of the trabecular network as a measure of bone quality [14]. The Euler characteristic is programmed as a tool to access connectivity of an image for trabeculae in BoneJ, a plugin-in for bone image analysis in ImageJ [15]. One component from Euler characteristic, Betti numbers, has been used for tumor area detection based on the connectivity of cells [16]. As a spatial patterns measurement in geostatistics, the indicator variogram is suggested as an promising tool for connectivity quantification [17]. The indicator variograms are a measure of spatial continuity at a specific threshold, where multiple thresholds can be selected for non-binarized images to calculate indicator variograms and the response serves as the measure of connectivity [18]. However, it was later shown that the indicator variograms are inappropriate and not able to distinguish between lens structure and connected channel structure [19]. Later, connectivity function was proposed by Allard serving as an alternative approach [20]. The connectivity function was used to distinguish the connected and disconnected patterns in soil moisture [21]. The results show the connectivity function can distinguish patterns that share the same response with the indicator variogram. Thus, it serves as a more promising tool that exist in observed spatial fields. Fast marching algorithm introduced by Sethian is a numerical method for tracking the evolution of monotonically advancing fronts in simulated grids [22]. The response from the evolving of wave front is related to the topology and spatial properties. Fast marching method was used to determine paths of anatomical connection between regions of the brain from the map of travel time [23] The application of connectivity quantification can be found across various disciplines. For example, quantification of the pore connectivity captured in the 2D digital images of porous materials [24]. Using such metrics, the segmented X-ray micro-tomography images were analyzed for assessing the connectivity of oil, water, and gas phases during different periods of injection [25]. Currently, there does not exist a standard measure of connectivity of material constituents. Moreover, there has not been a rigorous study on the robustness and reliability of connectivity metrics for composite materials.

In this study, six metrics are developed and tested for quantifying the connectivity of material constituents captured in high-resolution microcopy images. The robustness of the metrics is evaluated by applying the metrics to 3000 synthetic images representing six levels of connectivity and SEM images of organic-rich shale sample. We also study the sensitivity of these metrics to areal fraction and random distribution of a constituent. In this paper, we first introduce the synthetic images and the connectivity metrics used in this study. Following that, we analyze the results of the connectivity metrics when applied on synthetic images and real SEM images. The

effect of areal fraction and random distribution of constituent on the performance of metrics is discussed next.

2 Methodology

2.1 Material

Six types of synthetic binary images, comprising white or black pixels, with specific levels of connectivity are created for evaluating the performances of the six connectivity-quantification metrics. The six types of binary images will be referred as Type 1 to Type 6. These six types of images are created such that the connectivity of the white constituent/component decreases from Type 1 to Type 6 due to the reduction in the connectedness of the white constituent, while ensuring similar areal fraction of the white constituent irrespective of the level of connectivity. The typical binary images representing Type 1 to Type 6 are shown in Figure 1. Each image has a dimension of 200 pixels by 200 pixels. The synthetic dataset contains 500 realizations of each level of connectivity with random location and distribution of the constituent of interest (i.e. the white component/constituent). As a result, there are in total 3000 binary images in the synthetic dataset. The image shown in Figure 1(a) represents Type-1 connectivity, where the image contains ten horizontal and ten vertical bars representing the white constituent randomly distributed over the black constituent in the background. All the bars have the same dimension, i.e. hundred pixels in length and two pixels in width. These white pixels represent the material constituent of interest for which the connectivity is to be quantified, whereas the black pixels represent the background. Due to some overlapping pixels between the white horizontal and vertical bars, all the binary images have slightly less than 4000 pixels representing the white constituent. Consequently, the white constituent in each binary image is approximately 10% fraction of the entire image (~4000/(200×200)). Figure 1(b) represents the Type-2 binary image, which has slightly lower level of connectivity as compared to the Type-1 binary image. All images irrespective of the level of connectivity have the same dimension and fraction of the white constituent. All images belonging to Type 1 contain 10 vertical and 10 horizontal bars representing the white constituent. However, the bars in Type-2 images are half the length of those in Type-1 images; therefore, the numbers of bars in horizontal and vertical directions in Type 2 image are twice those in Type 1 image. The bars in Type-2 images are fifty pixels in length and two pixels in width. Type-2 binary image has 20 vertical and 20 horizontal bars randomly distributed over the black background. The synthetic binary images belonging to Types 3, 4, 5, and 6 are created in the same manner comprising 10% areal fraction of the white constituent, such that each subsequent connectivity type is created by reducing the length of randomly distributed bars and increasing the number of bars to keep the fraction of the white constituent the same while reducing connectivity of material constituent of interest. Each type of binary image representing specific level of connectivity consists of 500 different realizations of randomly distributed bars. As a result, there are in total 3000 binary images in the synthetic dataset.

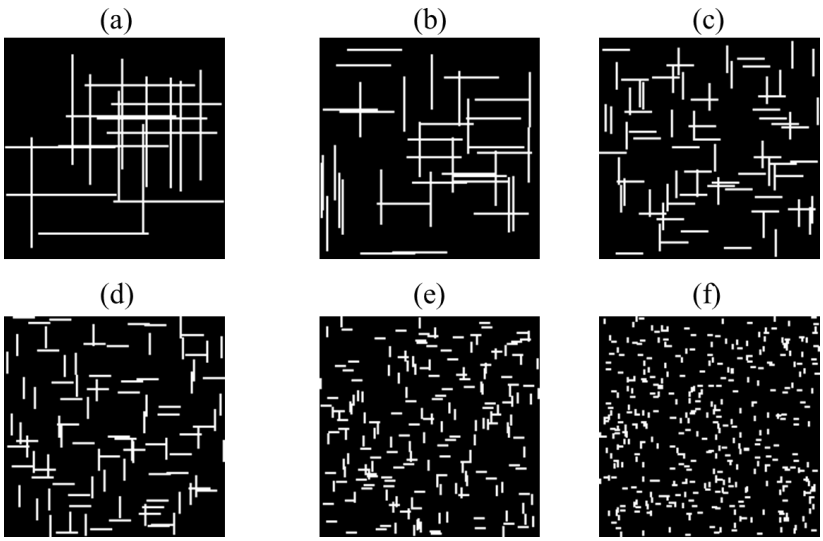

Figure 1. Six synthetic binary images representing Type 1 to Type 6 indexed as (a) to (f), respectively, such that each type represents a specific level of connectivity at approximately 10% areal fraction of the white constituents. Connectivity of the white constituent decreases from Type 1 to Type 6 due to the reduction in the connectedness with reduction in the length of the white bars randomly distributed over the black background. Figure 1(a) represents Type 1 with highest connectivity, whereas Figure 1(f) represents Type 6 with lowest connectivity.

After evaluating the six metrics on binary images corresponding to the six levels on connectivity (shown in Figure 1), the six metrics are applied on portions/slices of a segmented scanning electron microscopy (SEM) image. The segmentation was accomplished using a machine-learning assisted segmentation workflow [30,31]. Thus, the 200-pixel by 200-pixel binary image in this section serves as the real-world data and is taken from a 2000-pixel by 2000-pixel segmented image derived from an SEM image of shale rock sample from Wolfcamp formation. A robust image segmentation is required to convert the SEM images into segmented images that delineates the material constituents [26]. We first convert the segmented images into binary images such that the material constituent that represent organic matter is masked as white pixels and the remaining black pixels represent the background comprising matrix, pores, clays, and other solid minerals. In this study, we quantify the connectivity of the organic constituent shown in white. The two binary images shown in **Figure 2** have the image size of 200 pixels by 200 pixels. The fraction of the organic matter (i.e. the white constituent) in each of the two images is 15%. A visual examination of the images in Figure 2 indicates that the organic matter in the first image has higher connectivity than the second one. Our hypothesis is that the responses of the metrics when applied to two images will reveal the difference in the connectivity of organic matter in the two segmented SEM images. As a result, connectivity of the white constituent can be easily inferred from the response of the metrics instead of human-led visual inspection of each 200-pixel by 200-pixel image slice of the 2000-pixel by 2000-pixel SEM image. The proposed metrics will standardize and speed-up the analysis of connectivity.

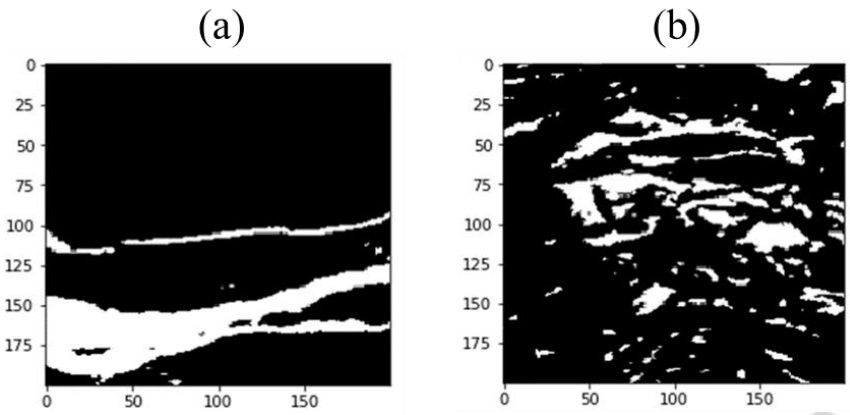

Figure 2. 200-pixel by 200-pixel binary image sliced from the 2000-pixel by 2000-pixel scanning electron microscopy (SEM) image of a shale sample having (a) high connectivity and (b) low connectivity of the organic constituent represented in white.

2.2 Clusters in an image

A cluster in a 2D image is a group of connected pixels. In 2D, two adjacent pixels are connected when they share the same face, referred as the 4-connectivity type, or when they share the same vertex, referred as 8-connectivity type. As shown in the **Figure 3(a),** only four pixels noted by 1 are connected to red pixel when 4-connectivity type is considered. In **Figure 3(b),** all the white pixels are connected to the red one when 8-connectivity type is considered. The group of pixels connected together form a single cluster. Different connectivity type will result in different configurations of clusters. The 8-connectiivty type is used in our study. A sample of clusters identified using the 8-connectivity type are shown in **Figure 4**, where white-constituent pixels in **Figure 4(a)** are grouped into 5 clusters, shown in **Figure 4(b)**. In our study, a 3×3 operating kernel shown in the **Figure 3(b)** when applied on binary image shown in **Figure 4(a)** identifies the clusters and assigns unique indices as shown in **Figure 4(b)**. It is to be noted, image segmentation is essential prior to cluster determination [27].

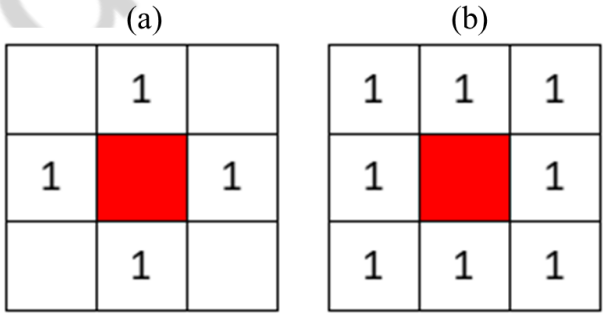

Figure 3. Operating kernels corresponding to (a) 4-connectivity type and (b) 8-connectivity type. These kernels are used to identify all the clusters in a 2D image. Kernel in Figure 3(b) is applied to Figure 4(a) to obtain the Figure 4(b)

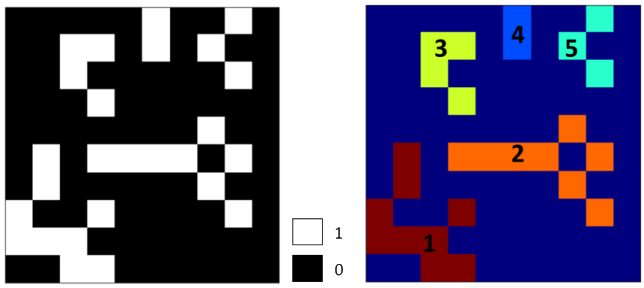

Figure 4. (a) Original synthetic image and (b) all the clusters in the synthetic image identified using the 8-connectivity operating kernel shown in Figure 3(b)

2.3 Metrics based on percolation theory

Connectivity in the percolation theory is defined as the probability of any two cells/points/pixels to belong to the largest percolating cluster [12] specific to the constituent of interest. On those lines, connectivity of a specific constituent is expressed as:

$$Connectivity\ index = \frac{N_L}{\sum_{i=0}^{n} N_i} \tag{1}$$

where the N_L is the number of pixels in the largest cluster, n is the total number of clusters, and N_i is the number of pixels in each of the cluster. Similar to this definition based on percolation theory, geobody connectivity developed for subsurface reservoir models is defined as the ratio of the largest geobody volume to the total grid volume. Following steps are required to calculate connectivity index specific to a constituent:

1. Identify all the clusters and assign a label to each cluster.
2. Determine the number of pixels in each cluster.
3. Use the values determined in the Steps 1 and 2 in Equation 1.

However, the metric in Equation 1 is suitable when the fraction of material constituent of interest is large and there is a single dominant cluster. However, when the fraction of constituent is low and there is no single dominant cluster, a more applicable metric is presented in Equation 2. The metric calculates the connectivity based on all the clusters and probability of pixels being in any one of the clusters, which is expressed as

$$Connectivity\ index = \frac{\sum_{i=0}^{n} N_i^2}{\left(\sum_{i=0}^{n} N_i\right)^2} \tag{2}$$

The numerator is formulated to emphasize the contribution of larger clusters.

2.4 Euler characteristic/number

Euler characteristic/number describes the shape and structure of topological spaces that don't vary with certain types of deformation/distortion, such as stretching, compression, inflation, twisting, and bending without gluing or tearing. Two topologically similar objects will have the same Euler

number. Geometry of an object is not equivalent the topology. In 3D, the Euler number can be calculated as

$$Euler\ Number = \beta_0 - \beta_1 + \beta_2 = number\ of\ vertices - number\ of\ edges + number\ of\ faces \quad (3)$$

β is referred as Betti number. The subscript 0, 1, or 2 of β are all based on the shape of objects represented using grid points or pixels. In 2D image, β_2 does not exist; so, the Euler number corresponds to the difference between the total number of clusters and total number of holes in the grid-based pixel representation, expressed as

$$Euler\ Number = \beta_0 - \beta_1 = number\ of\ clusters\ of\ constituent\ of\ interest - number\ of\ clusters\ of\ the\ background\ material\ completely\ surrounded\ by\ the\ constituent\ of\ interest \quad (4)$$

A hole is a cluster of the background material completely surrounded by the constituent of interest. The calculation of the Euler's number in a binary image is as follows:

1. Identify all the clusters of the constituent of interest.
2. Identify all the clusters of the background material.
3. Calculate the numbers of clusters of the constituent of interest and those of the background material.
4. Solve Equation 4 using values calculated in Step 3.

Negative values or values close to zero for the Euler number indicates high connectivity because of few large-sized dominant clusters of the constituent of interest and limited number of holes formed by the background material. Euler number does not have sufficient resolution for high connectivity. An increase in the areal fraction with a corresponding increase in the connectedness of constituent generally leads to number of holes to be larger than the number of constituent clusters with limited scattered pixels of constituents and background material (hole). This situation will decrease the Euler number to negative values. However, caution is required when using Euler number to quantify connectivity. For example, as the connectedness of clusters increases further the holes can get filled up, resulting in an increase in Euler number to positive values, which is contrary to the general trend of decrease in Euler number with increase in connectivity. Further, Euler number is not suitable for 2-dimensional images when the constituent of interest is distributed as large number of small clusters, especially for lower areal fractions. Issues with Euler number is evident for low areal fraction of the constituent of interest. Large number of scattered pixels, either the constituent of interest or the background material, will result in erroneously large positive or negative values of Euler number with large variance.

2.5 Indicator variogram

Variogram describes the degree of spatial dependence and spatial extension of randomly varying processes or properties. Variogram is the variance of the difference between the values of a property/process measured at two locations. Indicator variogram is calculated based on indicator values at various spatial locations on an indicator map, which is a binary transform of the original map of the continuous stochastic process or random variable based on the selected threshold. Indicator is a binary transform of the spatial distribution of a random variable (process/property) to either 1 or 0 for each spatial location, depending on whether the variable is above or below a threshold. The spatial distribution of a continuous variable can be transformed to an indicator map

based on a selected threshold. The binary images used in this study can be considered as indicator maps where pixels of value 1 (i.e. white pixels) represent constituent of interest and pixels of value 0 represent the background material. We intend to use the indictor variogram to quantify the connectivity of the pixels with the value of 1. The indicator variogram is expressed as

$$\gamma(h) = \frac{1}{2}E\{[I(u+h) - I(u)]^2\} = \frac{1}{2N(h)}\sum_{(u,u+h)}(I(u+h) - I(u))^2 \qquad (5)$$

where *I* is the indicator value (either 1 or 0) based on a specific threshold, *u* is the coordinate vector of a pixel, *h* is the separation distance between two pixels, and *N* is the number of paired pixels at a given distance *h*. Indicator value of 1 represents the constituent of interest and the indicator value of 0 represents the background material. The indicator variogram is related to the two-point spatial-correlation function $S(h)$ in the following manner:

$$\gamma(h) = p - S(h) \qquad (6)$$

where $S(h)$ is the probability of having two pixels located in the constituent of interest at a distance h and p is the areal fraction of the constituent of interest. For estimating reliable indicator variogram, a substantial amount of data is needed. The indicator variogram captures the distribution of pixels in the image but misses the information about clusters in the image. One limitation of indicator variogram is its inability to capture the curvilinearly connected features. Moreover, different patterns of pixel distributions may lead to similar indicator variogram response.

The following steps are used to calculate the two-point spatial-correlation function $S(h)$ and the indicator variogram $\gamma(h)$:

1. Specify the direction for computing the indicator variogram, either along X-axis, Y-axis, X-diagonal or Y-diagonal, as shown in **Figure 5.**
2. Specify the range of separation distance *h* for which indicator variogram $\gamma(h)$ needs to be computed.
3. Randomly select pixel pairs at the specified separation distance *h* in the specified direction.
4. Record total number of pixel pairs selected in the previous step and the number of those pixel pairs that lie in the constituent of interest. Hence, calculate $S(h)$ along the specified direction.
5. Subtract $S(h)$ from the areal fraction of the constituent of interest in the binary indicator map to obtain $\gamma(h)$.
6. Loop through all the separation distance *h* within the specified range to obtain the complete indicator variogram $\gamma(h)$ for the entire range of *h* in one specific direction.
7. Compute $\gamma(h)$ for the four directions, namely X-axis, Y-axis, X-diagonal and Y-diagonal.

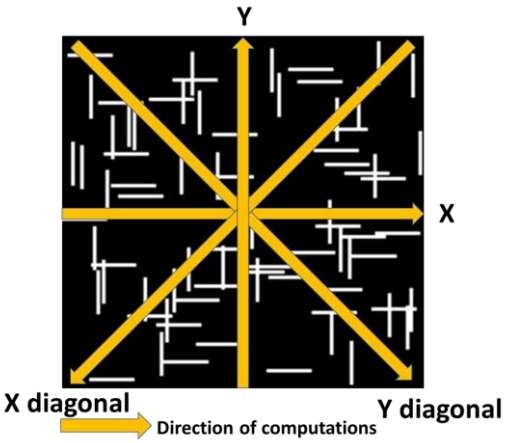

Figure 5. Four directions (X, Y, X-diagonal and Y-diagonal) used to calculate the indicator variogram $\gamma(h)$, two-point correlation function $S(h)$, and two-point cluster function *C2*(*h*)

2.6 Two-point cluster function

The two-point correlation function $S(h)$ can be written as a sum of two contributions [28]:

$$S(h) = C2(h) + D(h) \tag{7}$$

where the $C2(h)$ is the probability of finding two pixels separated by a separation distance *h* in the same cluster of the constituent of interest and $D(h)$ is the probability of finding two pixels at separation distance *h* but in different clusters of the constituent of interest. When the probability of finding two pixels in the same cluster of a constituent is high, there is a dominant cluster and the connectivity of the constituent is high. Calculation of $C2(h)$ requires the same steps as in calculating $S(h)$ except that in the step 4, it needs to be determined whether the pixel pairs lie in the same cluster of the constituent. $C2(h)$ is an indicator of local connectivity within a cluster, whereas $D(h)$ is an indicator of isolated clusters. A high global connectivity requires high $C2(h)$ and low $D(h)$. Consequently, *C2*(*h*) is a better indicator of connectivity as compared to *S*(*h*) and indicator variogram.

2.7 Connectivity function

The connectivity function or connectivity statistics $\tau(h)$ measures the probability of two pixels in the constituent of interest being connected. $\tau(h)$ is a function of separation distance *h* expressed as

$$\tau(h) = Prob(I(u) = I(u+h) = 1 | C(u), C(u+h) = A) \tag{8}$$

where A is a cluster index. Equation 8 mentions that two pixels u and $u+h$ are connected when both belong to constituent of interest and lie in the same cluster. The probability $\tau(h)$ of such connectedness of two pixels can be calculated for all the possible separation distances. Following that, the average connected distance I can be computed as [21]:

$$I = \int_0^\infty \tau(h) dh \tag{9}$$

where I represent the average distance over which pixels are connected. Long average distance represents high connectivity and vice versa. Average connected distance I can be converted to physical distance when the resolution/pixel-dimension of the image is known. Unlike the indicator variogram $\gamma(h)$, two-point correlation function $S(h)$, and two-point cluster function $C2(h)$ that have statistical formulation, $\tau(h)$ has a deterministic formulation. The following steps are required to calculate the connectivity function and the average connected distance I:

1. Locate clusters and assign cluster indices for the constituent of interest.
2. Determine separation distances between all the pixels pairs located in the constituent of interest.
3. For each separation distance calculated above, determine the number of pixel pairs located in the same cluster and the number of pixel pairs whose pixels are located in different clusters.
4. For each separation distance calculated above, calculate $\tau(h)$ as the ratio of number of pixel pairs sharing the same cluster index to the total number of pixel pairs in the constituent irrespective of the cluster index.
5. Compute average connected distance I by calculating the Integral in Equation 9.

2.8 Travel-Time Histogram computed using the Fast Marching Method

The fast-marching method is used to model the evolution of a boundary and interface, also referred as a front. The fast-marching method is a numerical technique to approximate the travel time T of a front moving through a region of varying travel speed $F(x)$. Travel time of a front is computed by solving the Eikonal equation [22] expressed as:

$$F(x)|\nabla T(x)| = 1 \tag{10}$$

where F is the travel speed at pixel location x and T is the travel time of the front arriving at a certain pixel location x. By setting a high travel speed for pixels belonging to the constituent of interest and null travel speed for pixels belonging to the background material, the travel time required for a front to travel between two pixels (i.e. pixel pair) in the constituent of interest at various separation distances can be calculated by solving Equation 10 using fast marching method. Such travel times between several pixel-pairs is a statistical indicator of connectivity, especially the tortuosity of the path connecting the pixels. Large travel times indicate pixels have high global connectivity. Following steps are implemented to compute the travel-time histogram to quantify connectivity of a constituent:

1. Determine pixels belonging to the constituent of interest.
2. Assign a large value of travel speed to the constituent pixels and a zero travel speed to the background pixels. Low values of travel speed will hinder the propagation of front.
3. Randomly select one of the constituent pixels to be source point.
4. Solve Eikonal equation using fast marching to obtain the travel times of a front starting from the source point and traveling to all the constituent pixels.
5. Identify the locations of pixels being reached from a source point. Do not use these identified pixels as source points in the next iterations.
6. Loop from Step 3 to 5 until all the pixels in constituent of interest has a travel time.
7. Loop from Step 3 to 6 for 30 times to obtain statistically significant distribution of travel time responses.
8. Compute the travel-time histogram and the average travel time to quantify the connectivity.

Step 6 prevents bias towards a specific/dominant cluster. Step 6 facilitates comparison of the average travel times for different images. Step 7 is indispensable because it minimizes the effect of random source point selection. However, Step 7 is computationally expensive especially when image size is large. The implementation of fast-marching method is similar to the one used by Ojha et al. [29] to estimate diffusion of pressure propagation in the subsurface.

3 Results and Discussion

The result section demonstrates the connectivity quantification achieved by the six metrics on the six types of synthetic binary images (Figure 1) and the two real SEM images (Figure 2). Connectivity quantified using scalar metrics (single-valued metrics) are presented as the mean and standard deviation computed for the 500 images (realizations) belonging to each connectivity type/level. On the other hand, connectivity quantified using spectral (non-scalar) metrics, such as indicator variogram, are presented as average probability at each separation distance. Connectivity function, travel-time histogram, cluster function, correlation function, and indicator variogram are spectral metrics, whereas Euler number, connectivity index, average connected distance and mean travel time are scalar metrics. Connectivity index, Euler number, and other scalar metrics cannot characterize the anisotropy and scale dependence of connectivity. Indicator variogram, two-point cluster function, connectivity function, and travel time histogram, have scale dependence and can differentiate between local and global connectivity. Indicator variogram and cluster function can quantify the directional nature of connectivity. Connectivity function, connectivity index and Euler number have deterministic formulation. Cluster function and connectivity function share certain similarities in terms of connectedness of pixels when they lie in the same cluster; however, connectivity function has a deterministic formulation. Travel-time histogram represents the tortuosity of connected paths unlike other metrics.

3.1 Quantification of Connectivity of Synthetic Binary Images

3.1.1. Connectivity Index based on percolation theory (Metric 1)

This is a scalar metric. The connectivity results using Equation 2, referred as the connectivity index, are listed in **Table 1**. This metric is computed based on percolation theory. The mean of connectivity index decreases from Type 1 to Type 6, consistent with the reduction in connectivity from Type 1 and 6. Moreover, there is close to 2 orders of magnitude variation in the mean of connectivity index, which indicates a high sensitivity of the metric to connectivity. The standard deviation indicates the variation in the metric over the 500 images per connectivity type. Coefficient of variation (CV) indicates the relative magnitude of standard deviation or the variability of standard deviation to its mean. There is a reduction in CV with reduction in connectivity indicating the variability of the metric for higher connectivity. A good metric should have low CV and low standard deviation especially for higher connectivity (e.g. Type 1). Also, the peak of CV irrespective of the decrease in connectivity from Type 1 to Type 2 and from Type 2 to Type 3, indicates the increase in variability of the metrics for connectivity level represented by the Type 2 images. The large standard deviation of Type 1 images is evident in **Figure 6**. Both the images are from Type 1 but the connectivity in **Figure 6(a)** is larger than in **Figure 6(b)** because of the randomly distributed bars in **Figure 6(a)** intersect with each other forming a single cluster, whereas the bars exhibit fewer intersection in **Figure 6(b)** resulting in more isolated clusters and lower connectivity as compared to **Figure 6(a)**.

Table 1. Connectivity of synthetic binary images shown in Figure 1 quantified using mean, standard deviation (std), and coefficient of variation (CV) of connectivity index based on percolation theory (computed using Equation 2).

	Connectivity Index		
	Mean	Std	CV
Type 1	0.705	0.200	0.28
Type 2	0.211	0.105	0.49
Type 3	0.038	0.009	0.23
Type 4	0.025	0.005	0.2
Type 5	0.009	0.001	0.11
Type 6	0.004	0.000	0

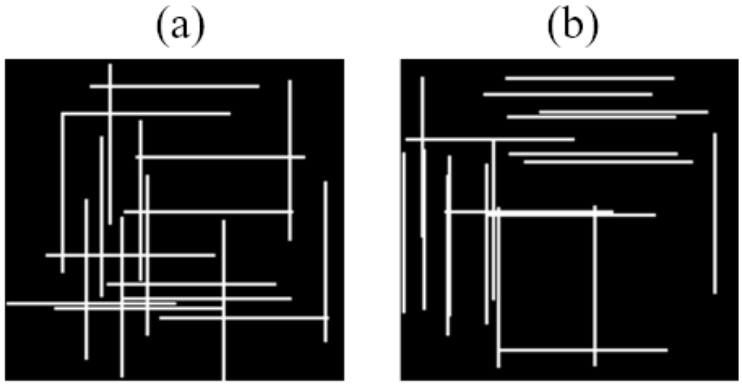

Figure 6. Two Type 1 images with different connectivity due to different random distribution of the white constituent.

3.1.2. Euler Characteristic/Number (Metric 2)

This is a scalar metric. The results from Euler number are shown in the **Table 2.** The average Betti numbers are also included in the table. The average β_0 and β_1 represent the number of clusters and holes, respectively, detected for each conductivity type. As the length of the distributed bars representing the constituent of interest decreases, there is an increase in number of clusters corresponding to the decrease in connectivity. Table 2 shows an increase in β_0 and decrease in β_1 with the decrease in connectivity. The resolution of β_0 improves whereas as that of β_1 decreases with the decrease in connectivity. Overall, there is an increase in Euler number with decrease in connectivity. Both β_0 (i.e. clusters) and β_1 (i.e. holes) influence the variation in Euler number for higher connectivity, whereas β_0 dominates the variation in Euler number for lower connectivity. Compared to Metric 1, this metric exhibits higher CV for higher connectivity (e.g. Type 1) and lower CV for lower connectivity (e.g. Type 6). Consequently, Metric 2 is better than Metric 1 for lower connectivity but Metric 2 exhibits large variability for higher connectivity. Metric 2 also has a high sensitivity to connectivity. Connectivity index and Euler number have higher variability at higher connectivity. Euler number is less reliable than connectivity index at high connectivity, whereas the connectivity index is less reliable at low connectivity.

Table 2. Connectivity of synthetic binary images shown in Figure 1 quantified using Betti numbers and Euler number based on integral topology.

Connectivity type	β_0	β_1	Euler number		
	Mean	Mean	Mean	Std	$\|CV\|$
Type 1	4.914	10.594	-5.68	6.71	1.18
Type 2	15.086	4.626	10.46	5.03	0.48
Type 3	46.116	1.95	44.17	4.91	0.111
Type 4	61.66	1.67	59.99	4.99	0.083
Type 5	142.75	1.232	141.52	5.9	0.041
Type 6	293.912	1.142	292.77	8.14	0.028

3.1.3. Indicator Variogram (Metric 3)

This is a spectral metric generated as a function of separation distance between two pixels. When presenting the indicator variogram, the y-axis is $\gamma(h)$ calculated using Equation 6 and x-axis is the separation distance *h* ranging from 0 to 200, which is the length of the synthetic image. Indicator variogram is calculated as the average over the 500 images for each type of connectivity. At a given separation distance, $\gamma(h)$ can be as high as the fraction of the constituent in the image indicating zero connectivity. When the constituent is fully connected, $\gamma(h)$ is zero. Because of the symmetry in the configuration of the white bars in the synthetic images, indicator variograms are generated in horizontal (X) direction and X-diagonal direction, as shown in **Figure 7**. The six plots correspond to Type 1 to 6, respectively, where red and blue curves represent the response in the horizontal (X) direction and X-diagonal direction, respectively.

Indicator variogram in horizontal direction (red) is highly sensitive to the level of connectivity, whereas that in the diagonal direction (blue) is relatively insensitive. $\gamma(h)$ increases with the increase in the separation distance indicating decrease in connectedness as the separation distance becomes comparable to the length-scale of cluster elements (i.e. the length of bar), beyond which the indicator variogram is insensitive to connectivity and separation distance. Each red curve increases with increase in separation distance *h* until the maximum length of white bars (100, 50, 25, 20, 10, 5 pixels for Type 1 to 6, respectively) are reached for specific image type. At any separation distance, the variogram response increases from Type 1 to 6, which is consistent with the reduction in the connectivity from Type 1 to 6. The directional sensitivity of the indicator variogram (i.e. separation between the horizontal and diagonal responses) is higher for higher connectivity (e.g. Type 1). Indicator variogram is better suited for representing local connectivity as compared to global connectivity because this metric does not differentiate whether pixel pairs belong to same cluster or not. For example, based on the indicator variogram response, Type 1 exhibits extensive local connectivity in horizontal direction that is better than other image types. Indicator variogram has good sensitivity to the connectivity level. The resolution of this metric

decreases with decrease in connectivity and the metric is suited for quantifying higher connectivity.

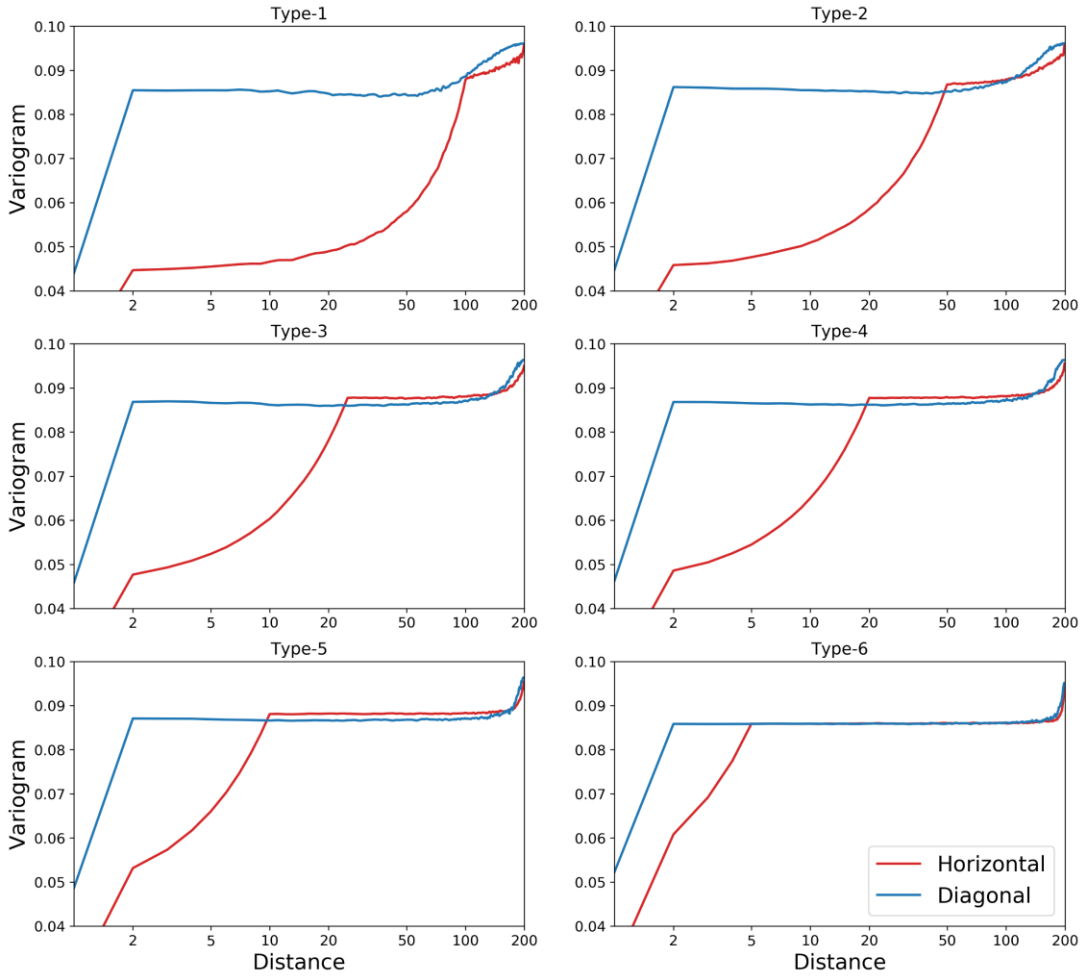

Figure 7. Connectivity of synthetic binary images shown in Figure 1 quantified using indicator variogram versus separation distance. The distance in x-axis is plotted in log scale.

3.1.4. Two-Point Cluster Function (Metric 4)

This is a spectral metric generated as a function of separation distance between two pixels. Due to symmetry in the synthetic images, the two-point cluster function $C2(h)$ is generated in horizontal (X) direction and X-diagonal direction, as shown in **Figure 8**. This metric is generated for separation distance ranging from 0 to 200 pixels and the probability $C2(h)$ ranges from 0 to 1. For visualization purposes, y-axis ranges from 0 to 0.1. Two-point cluster function is calculated as the average over the 500 images for each type of connectivity. The six plots correspond to the six image types with distinct level of connectivity. $C2(h)$ decreases from Type 1 to 6 in both horizontal and diagonal direction indicating the decrease in connectivity. $C2(h)$ in diagonal direction is much lower than that in horizontal direction due to the configuration of the white bars forming the constituent of interest. $C2(h)$ response becomes zero beyond a certain separation distance because pixels in constituent of interest are not connected, which is a major difference

from the $S(h)$. Two-point cluster function has good sensitivity to the connectivity level. The resolution of this metric decreases with decrease in connectivity and the metric is suited for quantifying higher connectivity.

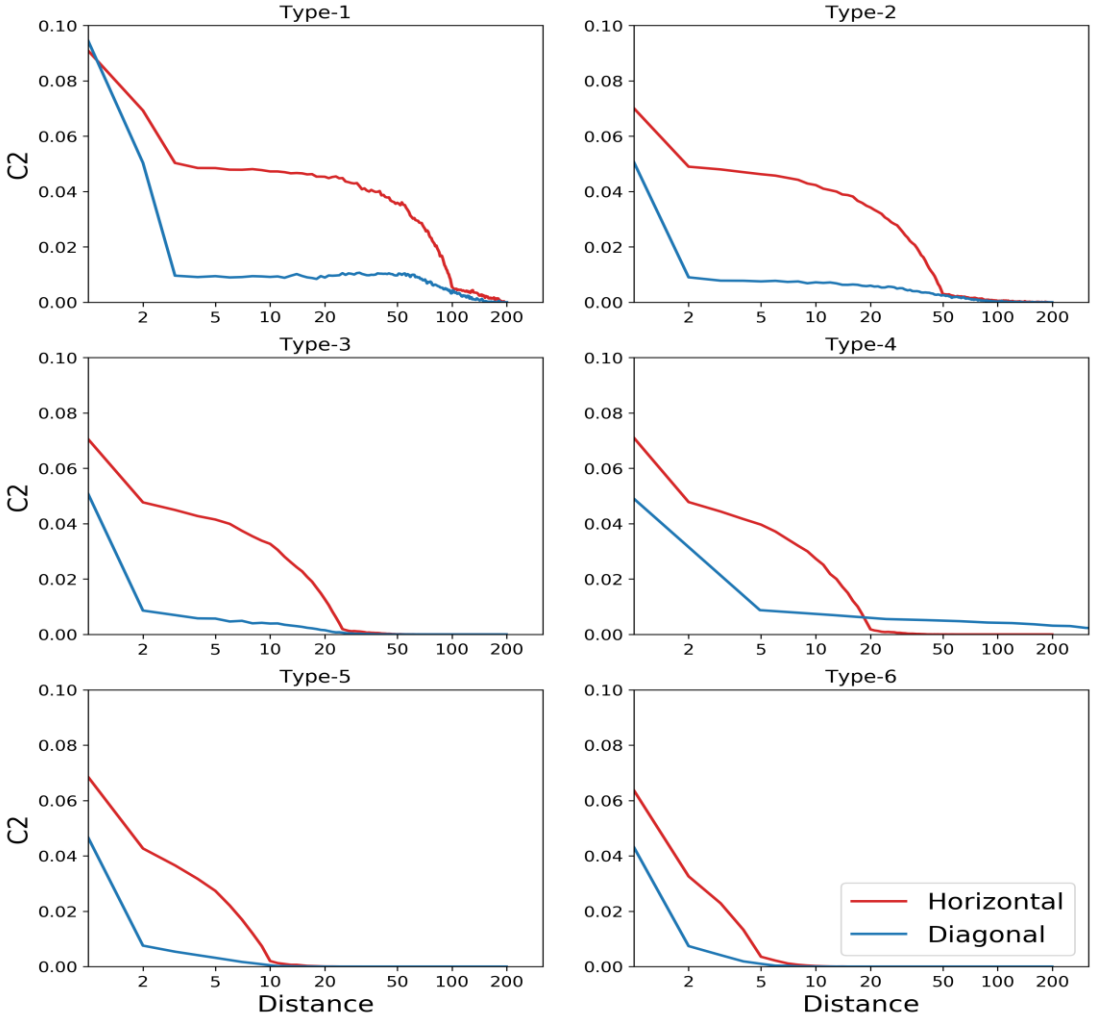

Figure 8. Connectivity of synthetic binary images shown in Figure 1 quantified using two-point cluster function versus separation distance

3.1.5. Connectivity Function (Metric 5)

Connectivity function $\tau(h)$is a spectral metric dependent on separation distance. Unlike cluster function, connectivity function cannot detect anisotropy. Average connected distance I can be derived from connectivity function by computing the area under the curve of connectivity function, as described in Equation 9. Connectivity function is presented in **Figure 9** where the probability $\tau(h)$ is in y-axis and the separation distance is in x-axis. The responses from images representing Type 1 to 6 connectivity are shown in six different colors. Similar to indicator variogram, $\tau(h)$ is calculated as the average over the 500 images for each type of connectivity. $\tau(h)$ decreases with the increase of separation distance. $\tau(h)$ decreases with the increase in connectivity. The Type 1 curve is the most gradually decreasing, whereas the Type 6 shows a drastic decrease within a

separation distance of 10. The local variation in each curve is due to the distinct number of pairs available at specific distances. The area under the curves are also calculated and plotted as bar plot for each connectivity type, which is shown in **Figure 10**. The average connected distance *I* for Type 1 is 141.7 and decreases all the way down to 3.4 for Type 6, which indicates the decrease in connectivity from Type 1 to 6. The resolution of this metric decreases with decrease in connectivity and the metric is suited for quantifying higher connectivity.

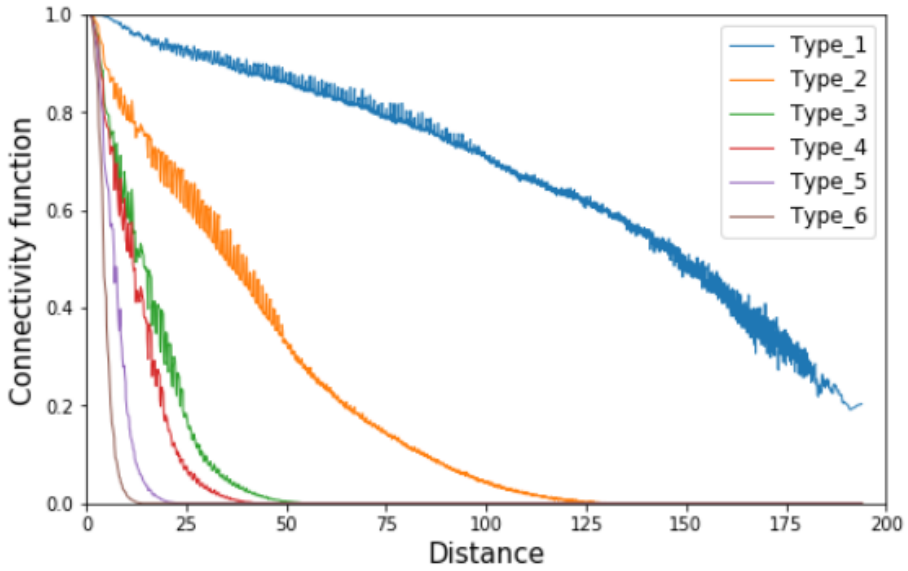

Figure 9. Connectivity of synthetic binary images shown in Figure 1 quantified using connectivity function versus separation distance

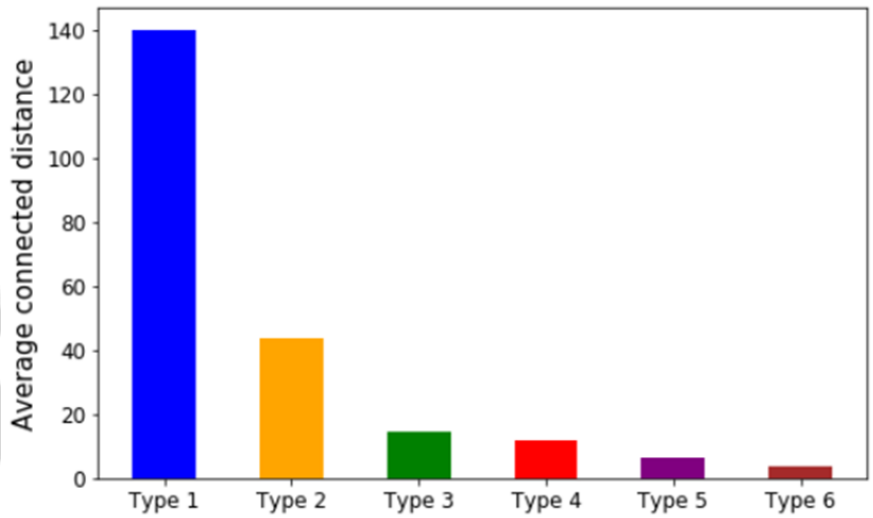

Figure 10. Average connected distance *I* calculated based on the area under the curve of connectivity function response shown in Figure 9.

3.1.6. Travel-Time Histogram (Metric 6)

Travel-time histogram is a spectral metric dependent on the travel time of a propagation front traveling though the constituent of interest. Like connectivity function, travel-time histogram

cannot capture the directional aspects of connectivity. 500 sets of travel-time responses are combined for each connectivity type. The overall travel-time responses are displayed as normalized histogram in **Figure 11.** The x-axis in the plot is the travel time and the y-axis is the frequency for the corresponding travel time bin. The histograms shown in six different colors represent the six connectivity types. The histogram for Type 1 is almost flat, where travel times from low to high are observed with equal frequency (number of occurrence), which indicates that a front can travel quite far from the source point before being stopped by the zero travel-speed edges at the interface of constituent of interest and the background material. The frequency of having short travel-time response gradually increases from Type 1 to 6 with a corresponding decrease in the variance and mean of the travel-time distribution, which indicates the front can only travel shorter distances from the source with the decrease in connectivity. The average travel time is obtained from the histogram for each connectivity type as a direct indicator of connectivity. The average travel time is shown in **Figure 12** as bar plot. The average travel time decreases from Type 1 to 6 suggesting the decrease in connectivity and such a drop is similar to the average connected distance from connectivity function.

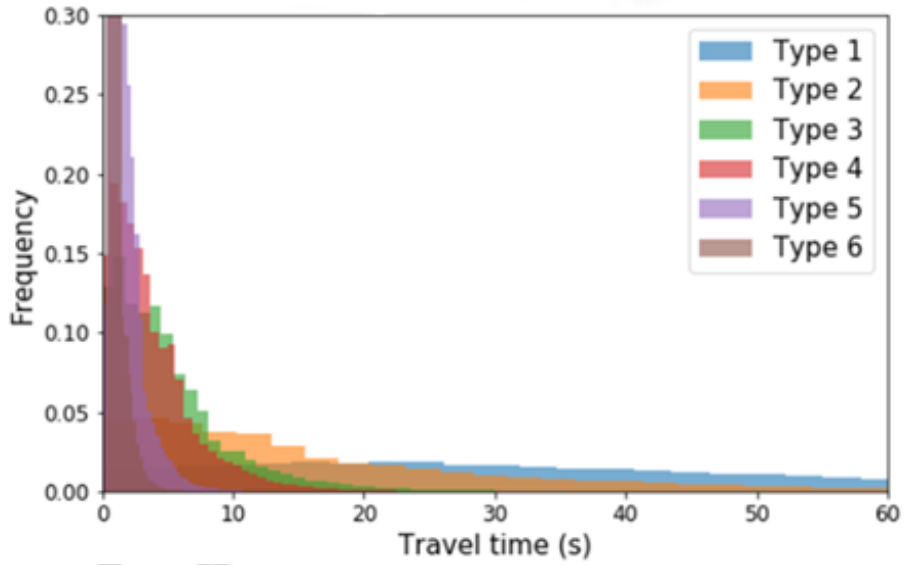

Figure 11. Connectivity of synthetic binary images shown in Figure 1 quantified using travel-time histogram computed using fast marching method

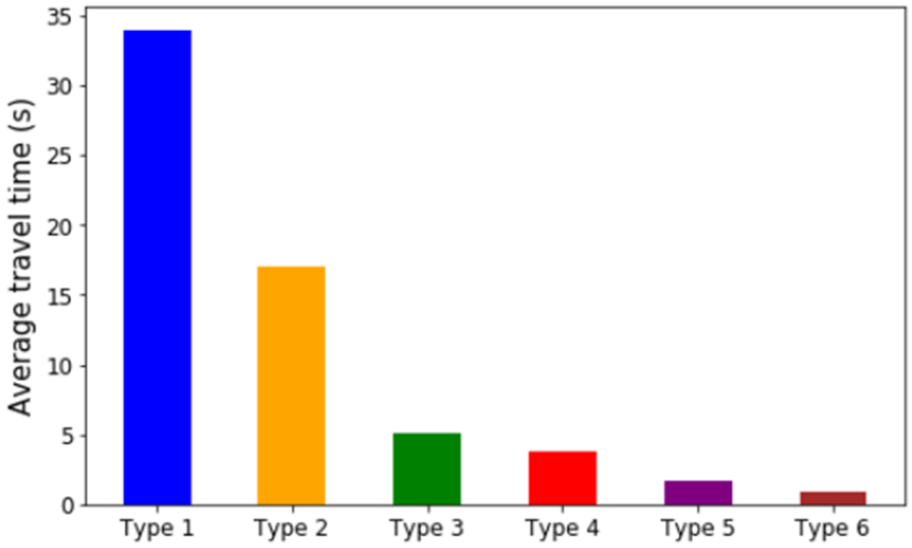

Figure 12. Average travel time for each connectivity type obtained from the travel-time histogram computed using fast marching method

3.2. Quantification of Connectivity of Real Scanning Electron Microscopy (SEM) Images

After testing the metrics on the synthetic images, the performance of the metrics is evaluated on the SEM images of shale samples, as shown in **Figure 2**. **Table 3** lists the connectivity obtained using connectivity index (Equation 2) and Euler number. Both the metrics indicate that **Figure 2(a)** has higher connectivity as compared to **Figure 2(b)**, which is consistent with the visual inspection. Both the metrics exhibit significant resolution in differentiating the connectivity for the two images. These metrics don't have directional and scale dependence.

Table 3. Connectivity of the real SEM images of shales in Figure 2 quantified using connectivity index (Metric 1) and Euler number (Metric 2)

Sample Index	Connectivity Index	Euler number
High, **Figure 2(a)**	0.714	4
Low, **Figure 2(b)**	0.105	76

The indicator variograms are generated in all the four directions, as shown in **Figure 13**. For **Figure 2(a)** representing higher connectivity in horizontal directional, the variograms in **Figure 13(a)** indicate that the local and global connectivity are similar in vertical and the two diagonal directions and are negligible beyond the separation distance of 30 pixels. However, the local and global connectivity in horizontal direction is much higher and the connectivity persists till separation distance of 175 pixels. For **Figure 2(a)**, connectivity in horizontal direction has longer range and higher magnitude as compared to the vertical and the two diagonal directions. In contrast, for **Figure 2(b)**, the connectivity is negligible beyond a separation distance of 15 pixels in all directions, as shown in **Figure 13(b)**. There may be few 100-pixel long clusters in horizontal direction. Based on Figure 13, we can conclude that **Figure 2(a)** has better connectivity than **Figure 2(b)** in all the four directions. Moreover, **Figure 2(a)** is predominantly connected in horizontal direction, while **Figure 2(b)** has isotropic connectivity. We do not present the responses

of two-point cluster function because it will provide similar information as indicator variogram. One disadvantage of two-point cluster function is that is only sensitive to the dominant cluster.

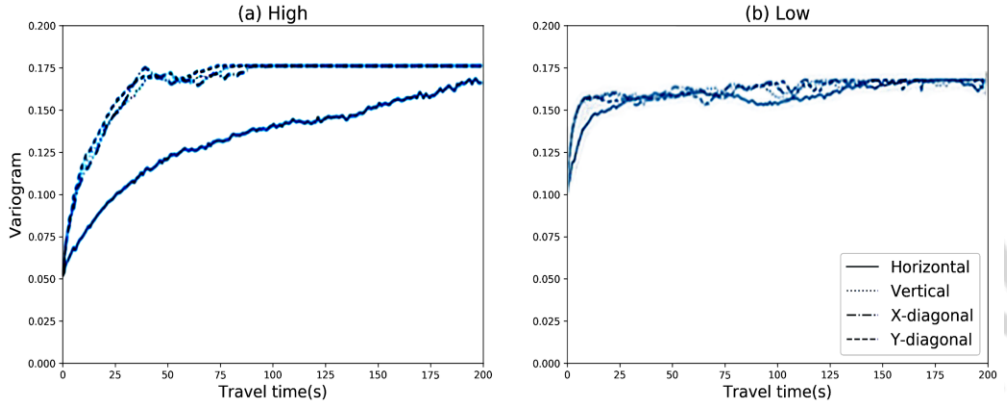

Figure 13. Connectivity of the real SEM images of shales in Figure 2 quantified using indicator variogram

The connectivity function for the two images in **Figure 2** are shown in the **Figure 14**. As separation distance increases, there are fewer pixel pairs in the same cluster beyond a certain separation distance resulting in a sharp drop in connectivity function. The sharp drop in connectivity function is evident in the low-connectivity sample at smaller separation distance. The high-connectivity sample exhibits a flat connectivity response from separation distance of 60 to 200 pixels indicating a dominant cluster spanning the image. Connectivity function has a good resolution for differentiating between high and low connectivity at separation distance greater than 25 pixels. Connectivity function is also a good indicator of local and global connectivity, as shown for the high-connectivity sample, where there are two flat responses, one at small separation distances smaller than 20 pixels and the other for separation distance between 60 to 200 pixels. The area under the curve is determined to be 125.97 and 25.48 for the high-connectivity and low-connectivity images, respectively, indicating the first image to be more connected than the second image. The travel-time histogram for the two samples in Figure 2 are presented in **Figure 15.** For the low connectivity image, the frequency of occurrence goes to zero for travel times larger than 34. The histogram of the low-connectivity sample has narrow spread and low mean travel time. The average travel time for **Figure 2(a)** and **Figure 2(b)** are 24.61 and 6.20, respectively. To conclude, the connectivity of the first sample is significantly larger than the second sample, consistent with visual inspection.

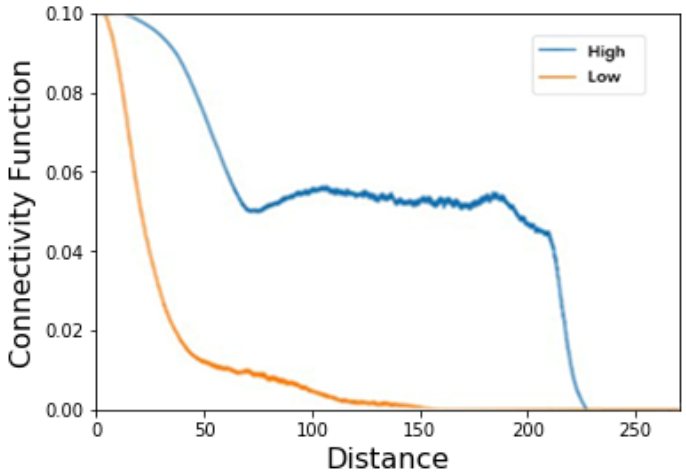

Figure 14. Connectivity of the real SEM images of shales in Figure 2 quantified using connectivity function

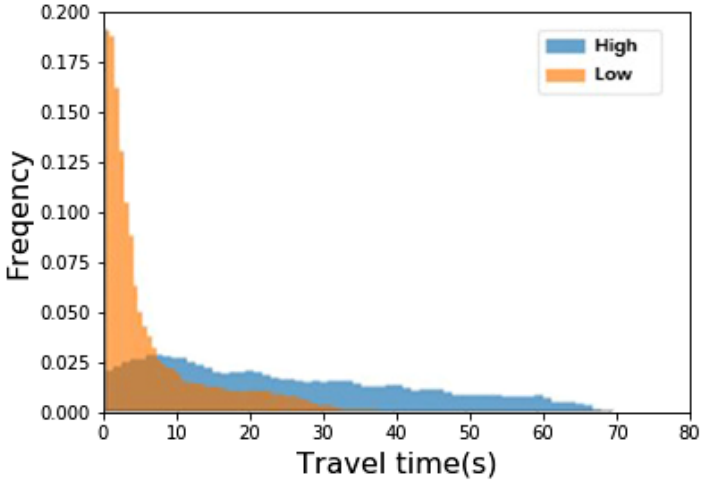

Figure 15. Connectivity of the real SEM images of shales in Figure 2 quantified using travel-time histogram computed using fast marching method

3.3. Effect of Areal Fraction on the Quantification of Connectivity

The fraction of constituent of interest in the synthetic images used so far is 10%. A robust connectivity metric should have higher sensitivity to connectivity and low sensitivity to the areal fraction. To evaluate the effect of areal fraction on connectivity metrics, a new set of synthetic images are created, where randomly distributed yet nonoverlapping white squares with edge length of 12, 9 and 6 are added to each image belonging to Type 1, 3, and 6. This results in an increase of areal fraction from 10% to 20%, 16% and 12%, respectively, without significant change in the global connectivity. For each image belonging to a specific level of connectivity, the areal fraction is altered by keeping the number and the location of the white squares and bars the same and changing the dimensions of the white square. The location of squares varies among the 500 realizations belonging to a specific connectivity type to add statistical randomness to the analysis.

Overall global connectivity decreases from Type 1 to Type 6 because of the reduction in the length of the white bars. Typical images used for this analysis are shown in **Figure 16.**

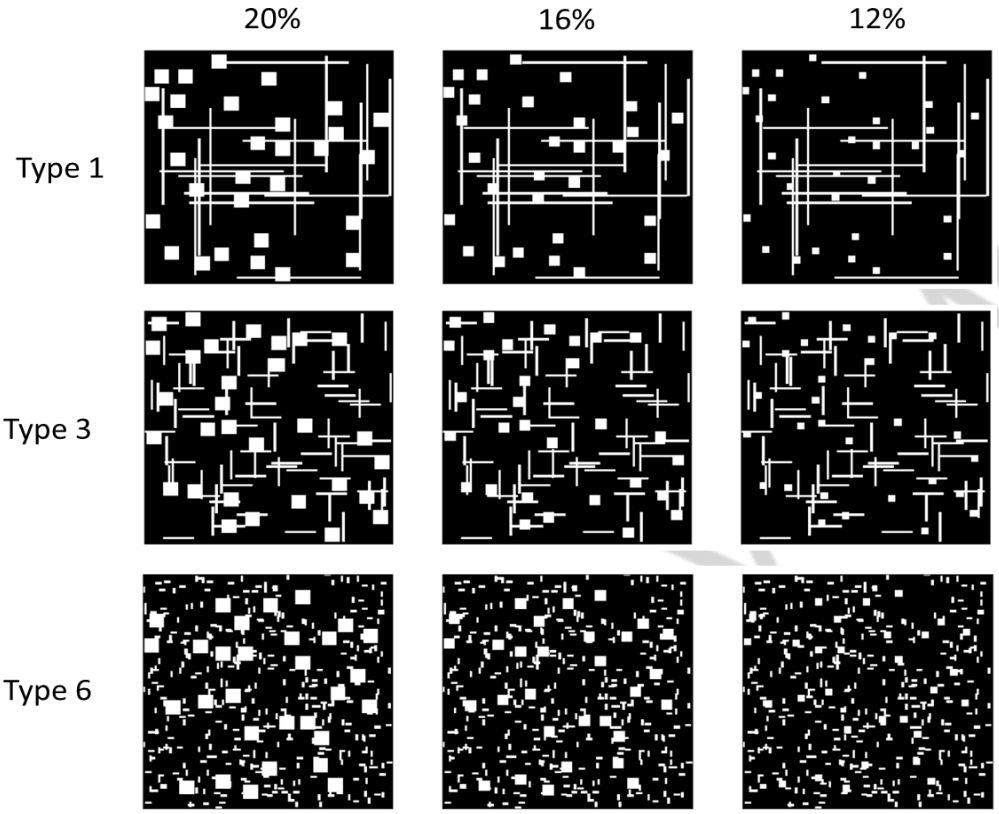

Figure 16. Typical synthetic images with different areal fractions and similar global connectivity. Areal fraction is altered by varying the dimensions of white squares and connectivity is altered by varying the length of the white bars. Leftmost top figure represents highest connectivity and highest areal fraction. Rightmost bottom figure represents lowest connectivity and lowest areal fraction.

Figure 17, 18, and **Table 4** present the performances of connectivity metrics in response to changes in areal fraction and connectivity. In the two-point cluster function $C(h)$ plot (Figure 17), the $C(h)$ is plotted against the distance h, where the solid line is the average $C(h)$ for 500 images per connectivity type and the shaded region represents the variations across the 500 images within the range of ± 2 standard deviations. The connectivity function plot in **Figure 18** is similar to the layout of **Figure 17**, where the average $\tau(h)$ calculated for 500 images per connectivity type is presented in solid line and the variations across the 500 images per connectivity type is presented as the shaded region around the solid line. The area under the curve (AUC) is calculated for the connectivity function and 2-point cluster function and is averaged for each image in the set of 500 images per connectivity type. The average travel time is derived from the travel-time histogram. The mean and coefficient of variation of the scalar metrics are presented in **Table 4.** As compared to connectivity function, cluster function (C2) has higher variability at lower separation distance, and vice versa. Variability of connectivity function and cluster function increases with the increase in connectivity. For Type 1, connectivity function exhibits large variations especially for large

separation distances. Mean connectivity function is relatively insensitive to variations in the areal fraction as compared to the mean cluster function.

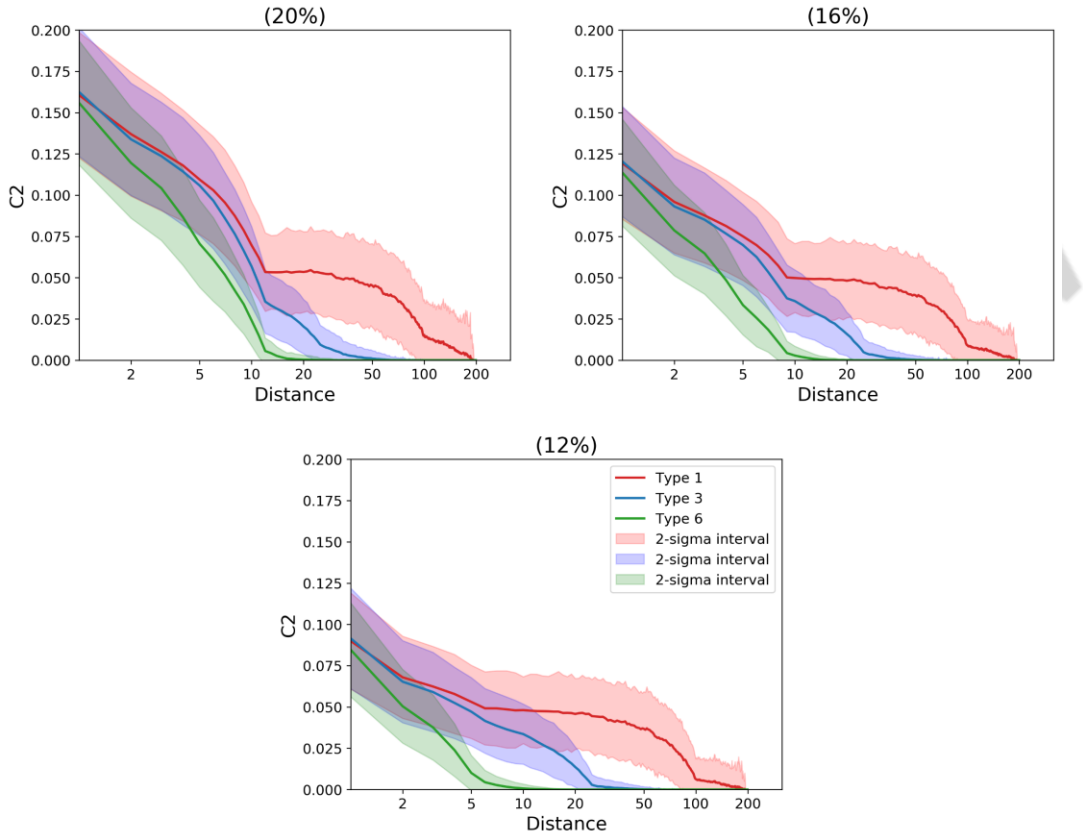

Figure 17. Connectivity of the synthetic images in Figure 16 quantified using the two-point cluster function (C2) to study the effect of areal fraction on the two-point cluster function

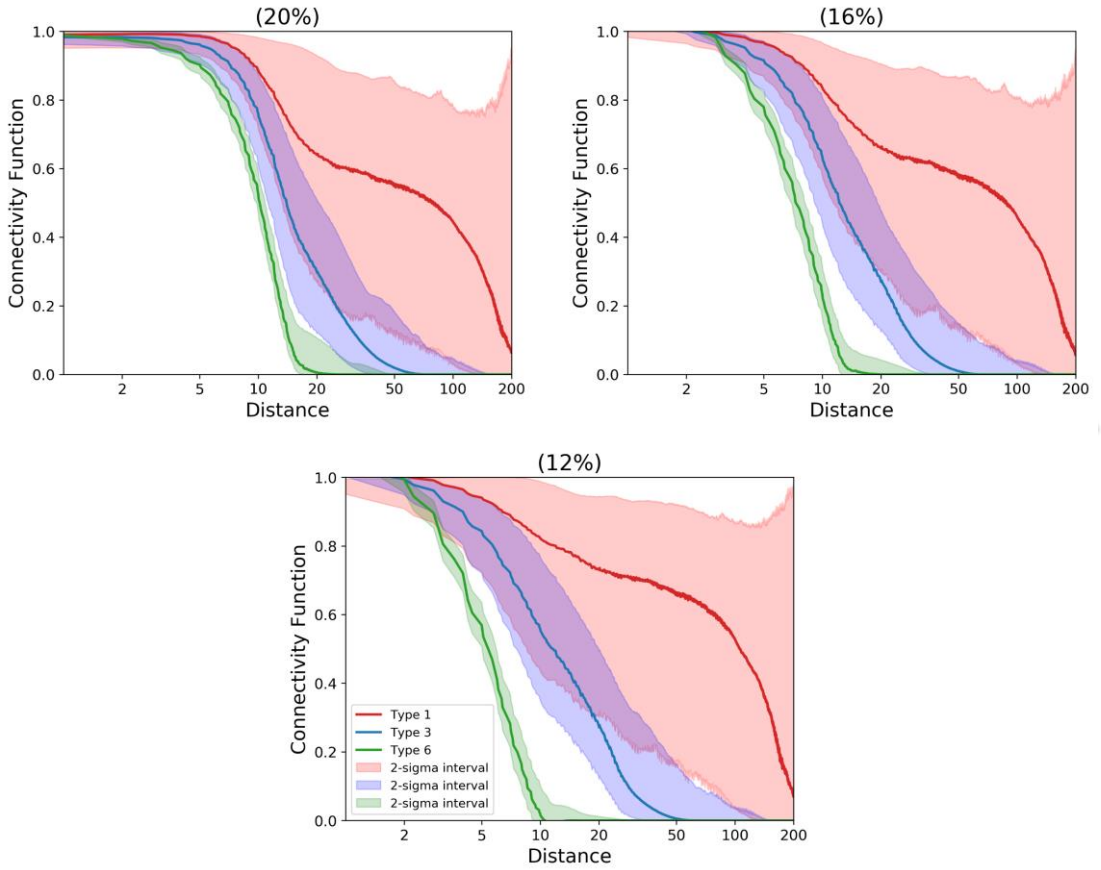

Figure 18. Connectivity of the synthetic images in Figure 16 quantified using the connectivity function to study the effect of areal fraction on the connectivity function

Table 4. Connectivity of the synthetic images in Figure 16 quantified using the five of the six metrics, namely connectivity index, Euler number, area under the cluster function, average connected distance from connectivity function, and mean travel time from the travel-time histogram, to study the sensitivity of the metrics to variations of areal fraction and connectivity type

Connec-tivity	Frac-tion	Connectivity Index		Euler Number		Area Under Cluster Function		Avg. Connected Distance		Mean Travel Time	
		Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Type 1	0.200	0.443	0.271	3.954	1.699	23.150	0.159	90.000	0.293	28.611	0.263
	0.160	0.464	0.274	7.706	0.873	22.020	0.147	90.350	0.294	28.532	0.225
	0.120	0.534	0.277	11.686	0.597	22.990	0.139	101.760	0.311	29.928	0.256
Type 3	0.200	0.040	0.250	46.300	0.123	6.300	0.088	16.770	0.144	5.675	0.171
	0.160	0.033	0.242	52.878	0.111	5.180	0.086	14.650	0.132	5.050	0.152
	0.120	0.032	0.250	59.068	0.098	4.700	0.091	13.630	0.137	4.834	0.145
Type 6	0.200	0.014	0.071	257.960	0.036	3.890	0.055	9.000	0.302	1.855	0.041
	0.160	0.009	0.011	278.792	0.032	2.690	0.173	6.660	0.033	1.363	0.038
	0.120	0.005	0.010	295.902	0.030	1.890	0.067	4.630	0.042	1.026	0.034

A robust connectivity metric should be more sensitive to connectivity type and less sensitive to areal fraction. All metrics are sensitive to connectivity type. Euler number and connectivity index are more sensitive to connectivity variations at lower connectivity, while the area under cluster function and connectivity function are more sensitive to connectivity variations at higher connectivity. The average travel-time is most sensitive to connectivity variations for both low and high connectivity. The average travel time and average connected distance are the least sensitive to the variation in areal fraction at a constant connectivity. Euler number and connectivity index are most sensitive to areal fraction at a constant connectivity. At high connectivity, Euler number exhibits large variations due to the randomness in the distribution of the constituent. Uncertainty in Euler number reduces with reduction in connectivity. Connectivity function and connectivity index also exhibit large uncertainty due to the randomness in the distribution of constituent at high connectivity, which reduces with the decrease in connectivity. Area under the cluster function followed by the average travel time is the most robust to randomness in the distribution of the constituent at high connectivity.

4 Conclusions

Six metrics are developed and tested for quantifying the connectivity of material constituents. The robustness of the metrics is evaluated by applying the metrics to 3000 synthetic images representing six levels of connectivity and SEM images of organic-rich shale sample. Connectivity function, travel-time histogram, cluster function, correlation function, and indicator variogram are spectral metrics, whereas Euler number, connectivity index, average connected distance and mean travel time are scalar metrics. Indicator variogram, two-point cluster function, connectivity function, and travel time histogram, have scale dependence and can differentiate between local and global connectivity. Indicator variogram and cluster function can quantify the directional nature of connectivity. Travel-time histogram represents the tortuosity of connected paths unlike other metrics. Connectivity index and Euler number have good sensitivity to connectivity but higher variability at high connectivity. Euler number is less reliable than connectivity index at high connectivity, whereas the connectivity index is less reliable at low connectivity. Unlike the two-point cluster function, the indicator variogram is better suited for representing local connectivity as compared to global connectivity. The directional sensitivities and resolutions of the cluster function and indicator variogram are higher for higher connectivity. Both average connected distance and mean travel time have good resolution at higher connectivity that decreases with decrease in connectivity. Being scalar metrics, average connected distance and mean travel time can be easily used to compare connectivity of different images. A robust connectivity metric should be more sensitive to connectivity type and less sensitive to areal fraction. The average travel-time is most sensitive to connectivity variations for both low and high connectivity. The average travel time and average connected distance are the least sensitive to the variation in areal fraction at a constant connectivity. Euler number and connectivity index are most sensitive to areal fraction at a constant connectivity. Area under the cluster function followed by the average travel time is the most robust to randomness in the distribution of the constituent at high connectivity.

Acknowledgements

We thank the Financial support from the American Chemical Society's Petroleum Research Foundation (ACS-PRF #59363-DNI Grant). We also thank the University of Oklahoma Research Council's Faculty Investment Program that supported few aspects of this project.

References

[1] J. Hooke, "Coarse sediment connectivity in river channel systems: a conceptual framework and methodology," *Geomorphology*, vol. 56, no. 1-2, pp. 79-94, 2003.

[2] C. Amoros and G. Bornette, "Connectivity and biocomplexity in waterbodies of riverine floodplains," *Freshwater biology*, vol. 47, no. 4, pp. 761-776, 2002.

[3] P. King, "The connectivity and conductivity of overlapping sand bodies," in *North Sea Oil and Gas Reservoirs—II*: Springer, 1990, pp. 353-362.

[4] P. Renard and D. Allard, "Connectivity metrics for subsurface flow and transport," *Advances in Water Resources*, vol. 51, pp. 168-196, 2013.

[5] G. Dagan, *Flow and transport in porous formations*. Springer Science & Business Media, 2012.

[6] G. Brierley, K. Fryirs, and V. Jain, "Landscape connectivity: the geographic basis of geomorphic applications," *Area*, vol. 38, no. 2, pp. 165-174, 2006.

[7] A. Moilanen and M. Nieminen, "Simple connectivity measures in spatial ecology," *Ecology*, vol. 83, no. 4, pp. 1131-1145, 2002.

[8] H. Okabe and M. J. Blunt, "Pore space reconstruction of vuggy carbonates using microtomography and multiple-point statistics," *Water Resources Research*, vol. 43, no. 12, 2007.

[9] J. Koestel, "SoilJ: an ImageJ plugin for the semiautomatic processing of three-dimensional X-ray images of soils," *Vadose Zone Journal*, vol. 17, no. 1, 2018.

[10] A. Hunt, R. Ewing, and B. Ghanbarian, *Percolation theory for flow in porous media*. Springer, 2014.

[11] N. Jarvis, M. Larsbo, and J. Koestel, "Connectivity and percolation of structural pore networks in a cultivated silt loam soil quantified by X-ray tomography," *Geoderma*, vol. 287, pp. 71-79, 2017.

[12] J. M. Hovadik and D. K. Larue, "Static characterizations of reservoirs: refining the concepts of connectivity and continuity," *Petroleum Geoscience*, vol. 13, no. 3, pp. 195-211, 2007.

[13] A. L. Herring, E. J. Harper, L. Andersson, A. Sheppard, B. K. Bay, and D. Wildenschild, "Effect of fluid topology on residual nonwetting phase trapping: Implications for geologic CO2 sequestration," *Advances in Water Resources*, vol. 62, pp. 47-58, 2013.

[14] B. Stampa, B. Kühn, C. Liess, M. Heller, and C.-C. Glüer, "Characterization of the integrity of three-dimensional trabecular bone microstructure by connectivity and shape analysis using high-resolution magnetic resonance imaging in vivo," *Topics in Magnetic Resonance Imaging*, vol. 13, no. 5, pp. 357-363, 2002.

[15] M. Doube *et al.*, "BoneJ: free and extensible bone image analysis in ImageJ," *Bone*, vol. 47, no. 6, pp. 1076-1079, 2010.

[16] K. Nakane, Y. Tsuchihashi, and N. Matsuura, "A simple mathematical model utilizing topological invariants for automatic detection of tumor areas in digital tissue images," in *Diagnostic pathology*, 2013, vol. 8, no. 1, p. S27: BioMed Central.

[17] A. Journel and F. Alabert, "Focusing on spatial connectivity of extreme-valued attributes: Stochastic indicator models of reservoir heterogeneities," *AAPG Bull.;(United States)*, vol. 73, no. CONF-890404-, 1989.

[18] A. W. Western, G. Blöschl, and R. B. Grayson, "How well do indicator variograms capture the spatial connectivity of soil moisture?," *Hydrological processes,* vol. 12, no. 12, pp. 1851-1868, 1998.
[19] S. Krishnan and A. Journel, "Spatial connectivity: from variograms to multiple-point measures," *Mathematical Geology,* vol. 35, no. 8, pp. 915-925, 2003.
[20] D. Allard, "Simulating a geological lithofacies with respect to connectivity information using the truncated Gaussian model," in *Geostatistical simulations*: Springer, 1994, pp. 197-211.
[21] A. W. Western, G. Blöschl, and R. B. Grayson, "Toward capturing hydrologically significant connectivity in spatial patterns," *Water Resources Research,* vol. 37, no. 1, pp. 83-97, 2001.
[22] J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proceedings of the National Academy of Sciences,* vol. 93, no. 4, pp. 1591-1595, 1996.
[23] G. J. Parker, C. A. Wheeler-Kingshott, and G. J. Barker, "Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging," *IEEE transactions on medical imaging,* vol. 21, no. 5, pp. 505-512, 2002.
[24] R. Scharfenberg, K. Meyerhoff, and D. Hesse, "Problems in the determination of pore connectivity by digital image processing," *Chemical engineering science,* vol. 51, no. 10, pp. 1889-1896, 1996.
[25] A. Scanziani, K. Singh, T. Bultreys, B. Bijeljic, and M. J. Blunt, "In situ characterization of immiscible three-phase flow at the pore scale for a water-wet carbonate rock," *Advances in water resources,* vol. 121, pp. 446-455, 2018.
[26] S. Misra, E. Ganguly, and Y. Wu, "Generalization of machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales," *Machine Learning for Subsurface Characterization,* p. 315, 2019.
[27] Y. Wu, S. Misra, C. Sondergeld, M. Curtis, and J. Jernigen, "Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales," *Fuel,* vol. 253, pp. 662-676, 2019.
[28] S. Torquato, J. Beasley, and Y. Chiew, "Two-point cluster function for continuum percolation," *The Journal of chemical physics,* vol. 88, no. 10, pp. 6540-6547, 1988.
[29] S. P. Ojha, S. Yoon, and S. Misra, "Pressure-Transient Responses of Naturally Fractured Reservoirs Modeled Using the Multistencils Fast-Marching Method," *SPE Reservoir Evaluation & Engineering,* 2019.
[30] Wu, Y., & Misra, S. (2019). Intelligent Image Segmentation for Organic-Rich Shales Using Random Forest, Wavelet Transform, and Hessian Matrix. IEEE Geoscience and Remote Sensing Letters.
[31] Misra, S., & Wu, Y. (2019). Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking. Machine Learning for Subsurface Characterization, 289.