

# The Correlation between Statistical Descriptors of Heterogeneous Materials

Shaoqing Cui <sup>a</sup>, Jinlong Fu <sup>a</sup>, Song Cen <sup>b</sup>, Hywel R Thomas <sup>a</sup>, and Chenfeng Li <sup>a,c,\*</sup>

<sup>a</sup>Zienkiewicz Centre for Computational Engineering, College of Engineering, Swansea University Bay Campus, Swansea SA1 8EN, United Kingdom

<sup>b</sup>School of Aerospace Engineering, Tsinghua University, Beijing 100084, China

<sup>c</sup>Energy Safety Research Institute, College of Engineering, Swansea University Bay Campus, Swansea SA1 8EN, United Kingdom

## Abstract

Heterogeneous materials such as rocks and composites are comprised of multiple material phases of different sizes and shapes that are randomly distributed through the medium. The random microstructure is typically described by using various statistical descriptors, which include volume fraction, two-point correlation function, and tortuosity, to name a few. Capturing different morphological features, a large number of statistical descriptors are proposed in different research fields, such as material science, geoscience and computational engineering. It is well known that these statistical descriptors are not independent from each other, but until recently it remains unclear what descriptors are more similar or more different. In particular, it is extremely difficult to look for quantified relations between various descriptors, since they are often defined in very different formats. The lack of quantified understanding of descriptors' relations can cause uncertainties or even systematic errors in heterogeneous materials studies. To address this issue, we propose a novel and generic correlation analysis strategy and establish, for the first time, the quantified relations between various statistical descriptors for heterogeneous materials. Based on data science techniques, our approach consists of three operational steps: data regularization, dimension reduction and correlation analysis. A total of 41 statistical descriptors are collected and analyzed in this study, which is readily extensible to include other new descriptors. The generic and quantified correlation results are compared with other established descriptor relations that are obtained from analytical analysis or physical intuition, and good agreements are observed in all cases. The quantified relations between various descriptors are summarized in a single correlation graph, which provides useful guiding information for the characterization, reconstruction and property prediction of heterogeneous materials.

**Keywords:** heterogeneous material, random media, morphology, statistical descriptor, characterization, reconstruction, property prediction

---

\* Corresponding Author, E-mail: [c.f.li@swansea.ac.uk](mailto:c.f.li@swansea.ac.uk)

# 1. Introduction

Multiphase heterogeneous materials (or random media) such as concrete, rocks, porous media, biological/artificial membranes, and composite materials are ubiquitous in both natural environment and engineering applications. A heterogeneous material is composed of two or more material phases (or voids) that are randomly distributed in the material medium. As a result, the microstructure of heterogeneous material is random in nature, and heterogeneous material samples are all different, even if they are of the same material type and sampled from the same source. The morphological features of heterogeneous materials are typically characterized by various statistical descriptors, including the volume fraction, the pore-size distribution function and the two-point correlation function, to name a few. These statistical descriptors capture the random morphological features from different aspects, e.g. connectedness, clustering, periodicity, and fractal etc. In recent years, the rapid development of various microscopy imaging techniques, such as X-ray micro-computed tomography (Micro-CT), magnetic resonance imaging (MRI) and scanning electron microscopy (SEM), focused ion beam and scanning electron microscopy (FIB-SEM), transmission electron microscopy (TEM), atomic force microscopy (AFM) and small-angle X-ray scattering (SAXC), has made the acquisition of high quality digital microstructures of heterogeneous materials increasingly more affordable and accessible, while the demand for digitalized heterogeneous material models is continuously growing at an even faster pace. As a result, quantitative morphological characterization is becoming more and more important for the research and application of heterogeneous materials in broad engineering fields, especially for energy, environmental, material, and chemical sectors.

The morphological descriptors not only help the characterization of heterogeneous materials, but they also support high-fidelity digital reconstruction (Zhang, et al., 2019) (Li, et al., 2019) (Feng, et al., 2014 & 2016) (Rahman, 2008) (Jiao, et al., 2008) (Fullwood, et al., 2008). Digital microstructures obtained through microscopy imaging are typically of small size, due to the hardware limitation from imaging equipment, and sometimes are only available in 2D form. In practice, the costs associated with sample collection, preparation and imaging also add a constraint on the quantity of digital microstructures that can be directly obtained from scanning of real-world samples. On the other hand, using scanned 2D/3D microstructures as the reference and guided by appropriate morphological descriptors, statistical reconstruction techniques can mass produce digital microstructures that share the same morphological features as the original real-world samples, without the limitation in the size, dimension and quantity of microstructures. This is particularly important for heterogeneous materials whose material properties can only be rigorously evaluated as an ensemble in the statistical context.

Another important use of morphological descriptors is for property prediction and material design. It has long been recognized that the complex microstructural morphology of heterogeneous materials has a profound impact on their macroscopic properties (Brown, 1955) (Rubinstein & Torquato, 1988) (Rubinstein & Torquato, 1989) (Torquato, 2000). With the increasingly more accessible microscopy imaging techniques, it is desirable to predict the physical properties of heterogeneous materials from their microstructural information. Such prediction is often much cheaper and quicker than performing complicated physical tests in

labs or fields. In addition, for manmade heterogeneous materials such as composites and foams, once the linkage between the microstructural morphology and the specific physical property is established, it becomes much easier to design and optimize the material structure for targeted properties. The relations between micromorphological characterization and macroscopic material properties have been extensively researched for various heterogeneous materials. For example, it is found that the particle nearest-neighbor density function is strongly correlated with the fluid permeability of porous media (Keller, et al., 1967) (Rubinstein & Torquato, 1988) (Rubinstein & Torquato, 1989) and some other examples are listed in Table 1.

Table 1. The relations between morphological descriptors and material properties

<b>Morphological Descriptor</b>	<b>The related material property</b>	<b>Reference</b>
N-point probability functions	Conductivity	(Brown, 1955)
	Elastic moduli	(Beran, 1968) (Torquato & Lado, 1985)
	Trapping constant	(Torquato & Rubinstein, 1989)
	Fluid permeability	(Prager, 1961) (Berryman & Milton, 1985) (Rubinstein & Torquato, 1989)
Surface correlation functions	Trapping constant	(Doi, 1976) (Rubinstein & Torquato, 1988)
	Fluid permeability	(Doi, 1976) (Rubinstein & Torquato, 1989)
Lineal path function & Chord-length density function	Knudsen diffusion and Radiative transport	(Ho & Strieder, 1979) (Tokunaga, 1985) (Tassopoulos & Rosner, 1992)
	Fluid motions in sedimentary rock	(Underwood, 1970)
Surface-particle correlation function	Fluid permeability	(Torquato & Beasley, 1987)
	Pyroconductivity	(Torquato & Rintoul, 1995)

Over the past several decades and in different research fields, many statistical descriptors have been proposed to characterize the random morphology of various heterogeneous materials. Some descriptors focus more on the lower-order statistical features, e.g. the mean and the grain-size distribution, while others emphasize more on the higher-order characteristics, e.g. the connectivity and the cluster correlation. Based on intuition, it has been long recognized that various statistical descriptors are not independent from each other, and some are more akin (or different) than others. However, it remains unclear to academic and industrial communities which morphological descriptors are more (or less) correlated to each other and to what extent they are correlated. These descriptors are expressed in different formats, i.e. numbers, vectors, matrices or functionals. Furthermore, some of them have a clear mathematical definition, while others do not. These facts make it very difficult to study the correlations of various morphological descriptors in a mathematically rigorous manner. On the other hand, without accurate knowledge on the relations between various descriptors, it is hard, if not impossible, to reliably quantify morphological features of heterogeneous materials or to interpret the physical properties in relation to the underpinning random microstructure.

By conducting a mathematically rigorous correlation analysis for various morphological descriptors collected from different research fields, this study aims to improve the understanding of the relations between different descriptors. A major challenge for the correlation analysis of morphological descriptors is caused by the intrinsic data heterogeneity: various descriptors are defined in different formats, making it very difficult to conduct quantified cross comparisons. To overcome this problem, we propose a novel correlation analysis methodology based on image processing and machine learning to study morphological descriptors that are heterogeneous in their own representations. The results of correlation analysis provide quantified evidence to guide the use of different descriptors in the characterization, reconstruction, and property prediction of heterogeneous materials. The rest of the paper is organized as follows. Section 2 provides a brief overview of various morphological descriptors, and the proposed new methodology for correlation analysis is explained in Section 3. The results and discussions are presented in Section 4, followed by the concluding remarks in Section 5.

## 2. Overview of Morphological Descriptors

Owing to the diversity and the random nature of heterogeneous materials, a wide range of statistical descriptors have been proposed to characterize their morphological features in different research fields, such as petroleum, material, and environmental engineering. Different statistical descriptors capture different morphological characteristics. For example, the volume fraction merely represents the volume percentage of individual material phases present in the random medium, while the grain-size distribution function captures the percentage of each material phase in different sizes. It is true that no single descriptor can capture fully the morphological features for all arbitrary heterogeneous materials. It is also true that different descriptors are not independent from each other. Instead, they are correlated, and some are more similar between each other than with others. Unfortunately, due to the data heterogeneity involved, quantified relations between various morphological descriptors have not been established, which hampers the research progress in characterization, reconstruction, property prediction and optimization of heterogeneous materials.

To draw unbiased and generic conclusions, a target of this study is to include in the analysis as many descriptors as possible from diverse literatures. Table 2 summarizes the morphological descriptors collected and analyzed in this study. The list is not exhaustive, but it has included all morphological descriptors we were able to identify and implement during this study. Due to the space limit, we cannot present detailed definitions for all these descriptors, and readers are referred to the corresponding references listed in Table 2.

Table 2. Morphological descriptors analyzed in this study

ID	Descriptor	Reference	ID	Descriptor	Reference
1	Volume Fraction	(Yeong & Torquato, 1998a,b) (Wang, et al., 1999)	2	Specific Internal Surface Density	(Yeong & Torquato, 1998a) (Kikkinides & Politis, 2014)

3	Grain-size Distribution Function	(Fullwood, 2010)	4	Pore-size Distribution Function	(Torquato & Avellaneda, 1991) (Torquato & Lu, 1993)
5	Core Distribution Probability	(Wang & Chen, 2007)	6	Directional Growth Probability	(Wang & Chen, 2007)
7	Orientation Distribution Function	(Bunge, 1982)	8	Two-point Correlation Function	(Debye, 1949) (Torquato, 2002)
9	Two-point Cluster Correlation Function	(Cule & Torquato, 1999) (Torquato, 2002)	10	Lineal-path Function	(Lu & Torquato, 1992a,b) (Yeong & Torquato, 1998a) (Strebelle, 2002)
11	Chord-length Density Function	(Torquato & Lu, 1993) (Roberts & Torquato, 1999)	12	Aspect Ratio Distribution Function	(Jennings & Parslow, 1988)
13	Radial Distribution Function	(Schiemann, 1997) (Jiao, et al., 2010)	14	Integral Correlation Function	(Pyrz & Bochenek, 2003)
15	Topological Entropic Descriptor	(Bochenek, et al., 2004)	16	Microcanonical Entropic Descriptor	(Piasecki, 2000) (Piasecki, 2009)
17	Local Percolation Probability	(Hilfer, 1991) (Manwart, et al., 2000)	18	Local Porosity Distribution Function	(Hilfer, 1991) (Hilfer, et al., 1997)
19	Tortuosity	(Carman, 1937) (Watanabe & Nakashima, 2001)	20	Voronoi Polygon Area	(Melro, et al., 2008)
21	Nearest-neighbor Distribution Function (1)_ep	(Hertz, 1909) (Reiss, et al., 1959)	22	Nearest-neighbor Distribution Function (2)_ev	(Hertz, 1909) (Reiss, et al., 1959)
23	Nearest-neighbor Distribution Function (3)_hp	(Hertz, 1909) (Reiss, et al., 1959)	24	Nearest-neighbor Distribution Function (4)_hv	(Hertz, 1909) (Reiss, et al., 1959)
25	Nearest-neighbour Orientation Distribution Function	(Sundararaghavan, et al., 2015)	26	Ripley's K Function	(Ripley, 1976&1977)
27	Coarseness	(Lu & Torquato, 1990)	28	Contiguity	(Gurland, 1958) (Han & Dawson, 2005)
29	Point/q-particle Correlation Function	(Torquato, 1986) (Torquato, 2000)	30	Surface-particle Correlation Function	(Torquato & Beasley, 1987) (Torquato & Rintoul, 1995)
31	Surface Correlation Function (1)_SSC	(Torquato, 2002)	32	Surface Correlation Function (2)_SVC	(Torquato, 2002)
33	Maximum Grain Radius	(Wang, et al., 1999) (Rollett, et al., 2007)	34	Minimum Grain Radius	(Wang, et al., 1999) (Rollett, et al., 2007)

35	Mean		36	Variance	(Cowan, 2002)
37	Standard Deviation	(Bland & Altman, 1996)	38	Autocovariance	(Hoel, 1984)
39	Autocorrelation Function	(Priestley, 1982)	40	Power Spectral Density Functions	(Wiener, 1964)
41	Margin Distribution Function				

Among these statistical descriptors, some carry the lower-order morphological information, while others carry the higher-order morphological information. Some descriptors are sensitive to topological connectivity, while others are strongly related to the spatial correlation of particles, et al. Based on the intuition of morphological features captured by individual statistical descriptors, several feature groups can be approximately formed and are summarized in Table 3.

Table 3. Intuitive grouping of morphological descriptors

Descriptors capturing the interrelation between clusters/pores	Nearest-neighbor Distribution Function	Radial Distribution Function
	Ripley's K Function	Integral Correlation Function
	Point/q-particle Correlation Function	Voronoi Polygon Area
	Surface-particle Correlation Function	
Descriptors capturing the correlation between points	Two-point Correlation Function	Power Spectral Density Function
	Two-point Cluster Correlation Function	Pore-size Distribution Function
	Lineal-path Function	Nearest-neighbour Distribution Functions
	Chord-length Density Function	Point/q-particle Correlation Function
	Autocovariance	Autocorrelation Function
	Specific Internal Surface Density	Chord-length Density Function
Descriptors capturing the geometries of the clusters/pores	Grain-size Distribution Function	Contiguity
	Pore-size Distribution Function	Surface-particle Correlation Function
	Directional Growth Probability	Surface Correlation Function
	Two-point Cluster Correlation Function	Maximum Grain Radius
	Lineal-path function	Minimum Grain Radius
	Topological Entropic Descriptor	Microcanonical Entropic Descriptor
Descriptors capturing the disorder degree of clusters/pores	Coarseness	
	Directional Growth Probability	Orientation Distribution Function

Descriptors capturing the orientation of inclusions	Nearest-neighbor Distribution Function	Lineal-path Function
	Two-point Cluster Correlation Function	
Descriptors capturing the connectivity of the clusters	Chord-length Density Function	Tortuosity
	Local Percolation Probability	

### 3. Correlation Analysis

#### 3.1. Analysis strategy for heterogeneous descriptors

Among the various morphological descriptors listed in Table 2, some have clear mathematical expressions, while others are only defined as computer algorithms that operate on 2D/3D microstructural images. These descriptors are represented in different data formats, as scalars, vectors, matrices and functionals. Various statistical descriptors are different, but they are not independent from each other (Torquato, 2000). Hence, a mathematically rigorous understanding for the relations between different descriptors is desirable, but till now this is only possible in very limited cases where explicit mathematical formulations exist and are defined in compatible forms. A quantified understanding for the relations between various morphological descriptors remains missing. The lack of compatible mathematical formulations and data formats present a major challenge to studying the relations of morphological descriptors. To overcome the problem posed by intrinsic data heterogeneity, we propose a novel correlation analysis strategy based on image processing and machine learning. As shown in Figure 1, the new analysis strategy comprises three operational steps:

- **Step 1:** Data regularization to transform the raw data of various morphological descriptors into a unified format
- **Step 2:** Dimension reduction to extract characteristic features and reduce the dimension of regularized data
- **Step 3:** Correlation analysis via classification and ranking

The details of these operational steps are explained in the following subsections.



Figure 1. Analysis strategy for heterogeneous descriptors

#### 3.2. Data regularization through heterogeneous material modelling

To regularize the heterogeneous morphological descriptors, we transform the morphological information captured by individual statistical descriptors that are of different expressions and formats into heterogeneous material images of the same size and resolution. Thus, instead of

directly dealing with various statistical descriptors that are heterogeneous in nature, this transformation allows the correlation analysis to be performed on unified data sets, i.e. the microstructural images of identical size and resolution. A schematic illustration of the data regularization workflow is given in Figure 2. First, a reference sample from measurement (e.g. Micro-CT scan) is selected as the representative microstructure, which should contain as many as possible morphological features captured by the involved descriptors. For instance, to study the grain-size distribution function or the orientation distribution function, the representative microstructure must contain enough clearly separated particles. By doing so, the morphological information carried by each involved descriptor can be potentially measured through the representative microstructure. Next, all involved descriptors  $D_i$  are individually extracted from the representative microstructure, where the subscript  $i \in \{1, 2, \dots, 41\}$  denotes the  $i$ -th descriptor as defined in Table 2. It should be noted that descriptors  $D_i$  are of different data formats depending on their definitions. Finally, using the representative microstructure as reference and guided separately by each descriptor  $D_i$ ,  $i \in \{1, 2, \dots, 41\}$ , we reconstruct a series of heterogeneous material sample sets  $S_i = \{R_i^j | j = 1, 2, \dots, N_S\}$ , each comprising  $N_S$  reconstructed samples. The size  $N_S$  of reconstructed sample sets should be chosen sufficiently large to avoid bias, and we found  $N_S = 20$  is sufficient to ensure statistical convergence in our analysis. These reconstructed sample sets  $\{S_i | i = 1, 2, 3, \dots, 41\}$  form a database that transforms the statistical information of individual descriptors with different formats into a unified format, i.e. the microstructural images of identical size and resolution.

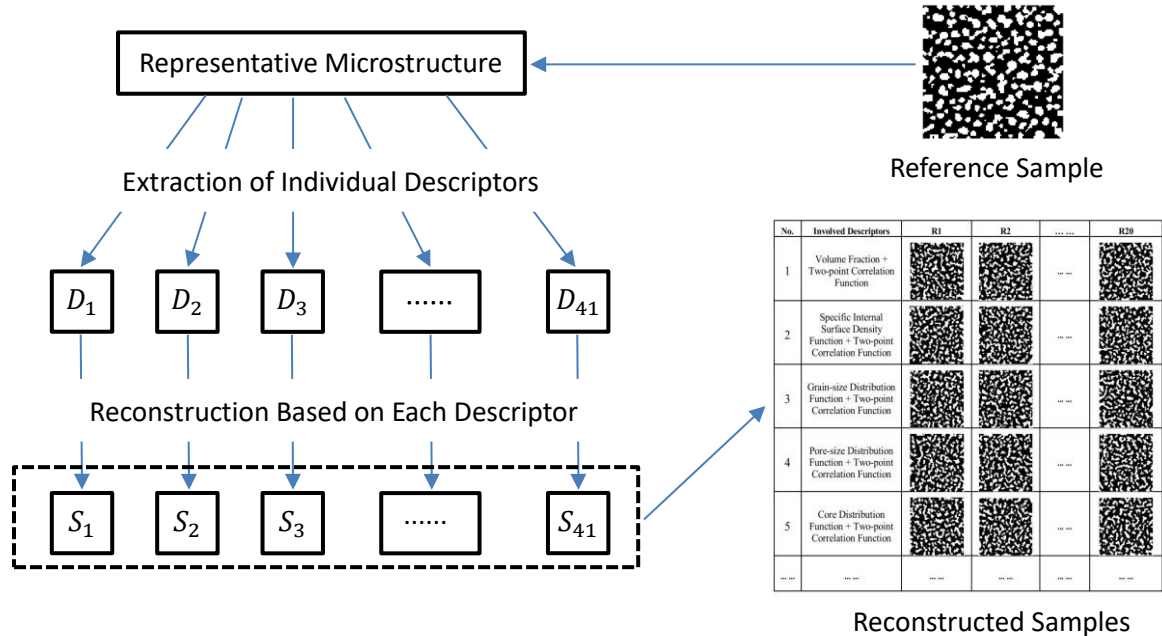


Figure 2. Data regularization through heterogeneous material modelling

Heterogeneous material modelling has been extensively researched in the past several decades (Zhang, et al., 2019) (Li, et al., 2019) (Yang, et al., 2018) (Quey & Renversade, 2018) (Cui, et al., 2016 & 2017) (Li, et al., 2015) (Kikkinides & Politis, 2014) (Chen, Jiao & Torquato, 2014) (Guo, et al., 2014) (Tahmasebi & Sahimi, 2013), and a range of modelling techniques have been developed to reconstruct digital samples that share the same statistical features as



the given reference (Cui, et al., 2015). Among the various heterogeneous material modelling techniques, the stochastic optimization method (Yeong & Torquato, 1998) (Manwart, Torquato, & Hilfer, 2000) (Ryu & Li, 2012) (Li, 2014) is arguably the most flexible and robust approach. This method is not particularly fast in terms of computational efficiency but with abundant computing power, it can achieve very high reconstruction accuracy for almost any type of heterogeneous materials. For these reasons, the stochastic optimization method is adopted in this study for the sample reconstruction task. The basic idea is simple, as shown in Figure 3. Given a reference sample and an initial guess, a second-order objective function is defined to measure the difference/error between guess and reference samples in terms of specific descriptors. Then, the current guess is randomly mutated to create a new guess, and the objective function is recomputed to accept (if it reduces the objective function) or reject (if it increases the objective function) the new guess. This process of guess, judge and re-guess continues until the error measured by the objective function drops below a user-specified tolerance, and the last guess is taken as the reconstruction result.

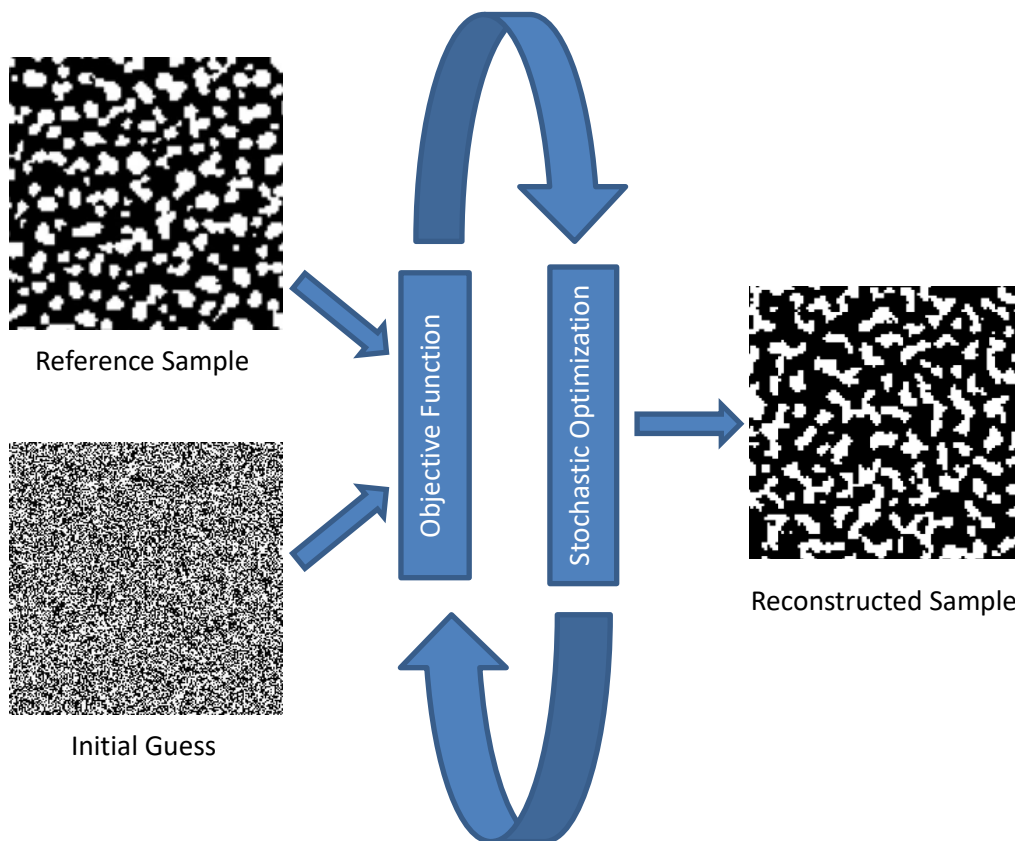


Figure 3. Stochastic optimization method

The objective function  $E$  of the stochastic optimization method is typically defined as the sum of squared errors between reference and reconstructed samples measured by a set of user-selected descriptors:

$$E = \sum_i [D_i(\Omega) - D_i(\hat{\Omega})]^2 \quad (1)$$

where  $D_i(\cdot)$  denotes the  $i$ -th descriptor incorporated in the objective function,  $\Omega$  denotes the reference sample, and  $\widehat{\Omega}$  denotes the trial reconstruction. Thus, the reconstruction process shown in Figure 3 practically transforms the statistical information captured by the descriptors  $D_i(\cdot)$  into reconstructed digital images of the same size and resolution. However, it is worth to note that some basic descriptors (e.g. the mean  $D_{35}(\cdot)$  and the variance  $D_{36}(\cdot)$ ) may contain too little morphological information and are not sufficient to support the reconstruction alone. On the other hand, the two-point statistics (i.e. two-point correlation function and autocorrelation functions) carries the most fundamental and essential morphological information for the statistical reconstruction (Torquato, 2002) and in almost all literatures of heterogeneous material modelling, it is always adopted either to reconstruct samples or to verify the quality of reconstruction.

Indeed, for heterogeneous material reconstruction using the stochastic optimization method, the two-point correlation function  $D_8(\cdot)$  is typically included as part of the objective function (Eschricht, et al., 2005) (Li, et al., 2012) (Alexander, et al., 2009) (Capek, et al., 2009) (Capek, et al., 2011). Therefore, in this study, a combination of the two-point correlation function and each involved descriptor is employed to reconstruct the samples, such that the objective function  $E_i$  corresponding to the  $i$ -th descriptor  $D_i(\cdot)$  is defined as follows:


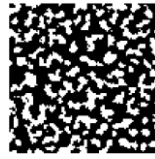
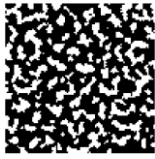






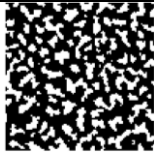





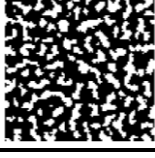

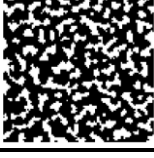
$$E_i = \omega_1 [D_i(\Omega) - D_i(\widehat{\Omega})]^2 + \omega_2 [D_8(\Omega) - D_8(\widehat{\Omega})]^2 \quad (2)$$

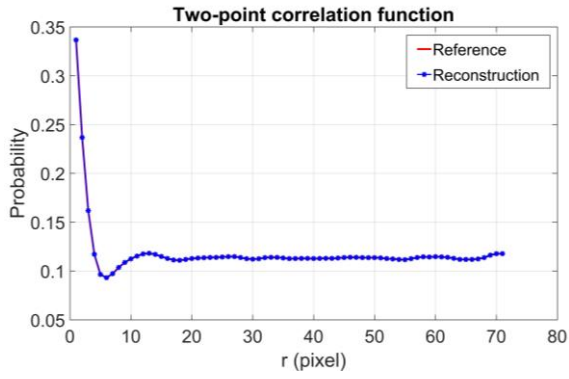
where  $D_8(\cdot)$  denotes the two-point correlation function,  $\Omega$  denotes the reference sample, and  $\widehat{\Omega}$  denotes the trial reconstruction. The first and second terms in Eq. (2) represent the squared errors of the involved descriptor and the two-point correlation function, respectively. All descriptors  $D_i$ ,  $i = 1, 2, \dots, 41$ , are regularized such that their values are at the same order of magnitude. The weight parameters are set as  $\omega_1 = 0.5$  and  $\omega_2 = 0.5$  in our study, which practically results in both terms getting minimized (or nearly minimized) during the reconstruction iterations. Note that the two-point correlation function is a special case of a more general descriptor series including the  $N$ -point correlation function (Torquato & Stell, 1982) (Adams et al., 1989) (Quintanilla, 2006) (Baniassadi, et al., 2012), the multiple-point statistics function (Caers, 2001) (Strebelle, et al., 2003), the pair correlation function (Schiemann, 1997) (Jiao, et al., 2010), the cross-correlation function (Tahmasebi & Sahimi, 2013) and the field-field correlation function (Teubner & Strey, 1987) (Roberts, 1997a, b). Therefore, we do not explicitly consider these descriptors in the correlation analysis, as their relations with other descriptors are represented by the two-point correlation function that is already included as part of the objective function for all descriptors.

Corresponding to each selected reference sample and each morphological descriptor in Table 2, a set of 20 2D digital microstructures at the resolution 100-by-100 are constructed to avoid errors caused by singular realization. Four reference samples of different morphological features are used in this study for the correlation analysis of descriptors, which provide the same correlation analysis results. As an illustration of the reconstruction database, Table 4 shows the reconstruction results from one reference sample. Visually, all realizations look similar. This is because the essential information captured by the two-point correlation function

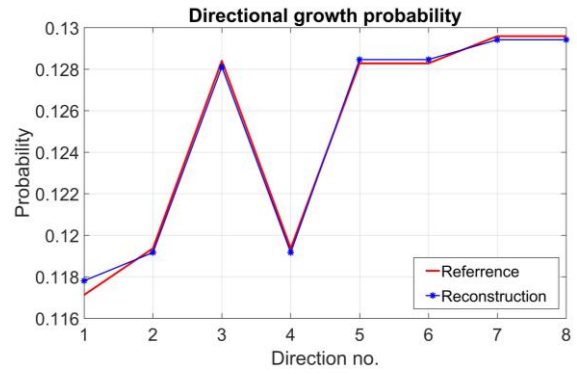
is fully preserved in all realizations. The other descriptors only provide additional information of microstructure. The differences between the reconstruction and the reference can be examined by using various descriptors, as shown in Figure 4. It can be observed that not only the two-point correlation but also the other descriptors are accurately preserved by the reconstructed samples. It is noted that all involved descriptors have been tested for all corresponding samples but due to the limitation in space, only a small portion of the validations are shown in Figure 4. Hence, the information carried by each descriptor is successfully transformed into the reconstructed microstructures with uniform size and resolution.

Table 4. Reconstructed samples

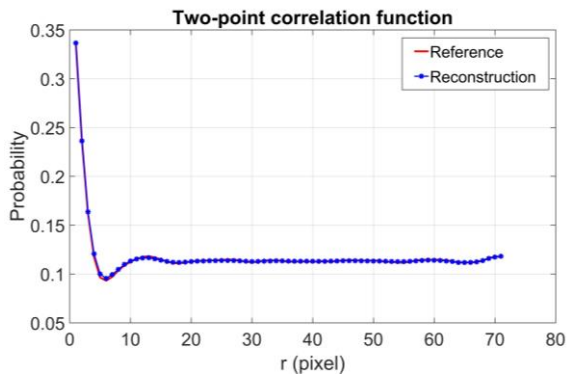
No.	Involved Descriptors	R1	R2	...	R20
1	Volume Fraction + Two-point Correlation Function			...	
2	Specific Internal Surface Density Function + Two-point Correlation Function			...	
3	Grain-size Distribution Function + Two-point Correlation Function			...	
...	...	...	...	...	...
39	Autocorrelation function + Two-point Correlation Function			...	
40	Power spectral density functions (SDF) + Two-point Correlation Function			...	
41	Margin Distribution Function + Two-point Correlation Function			...	



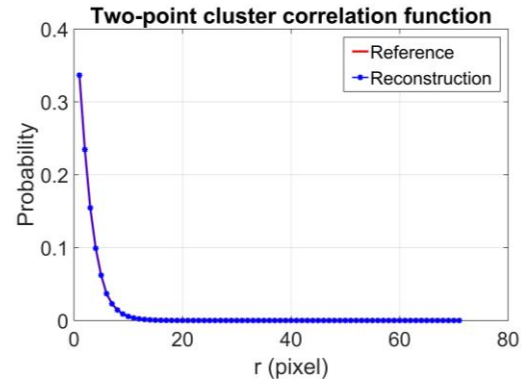
(a1)



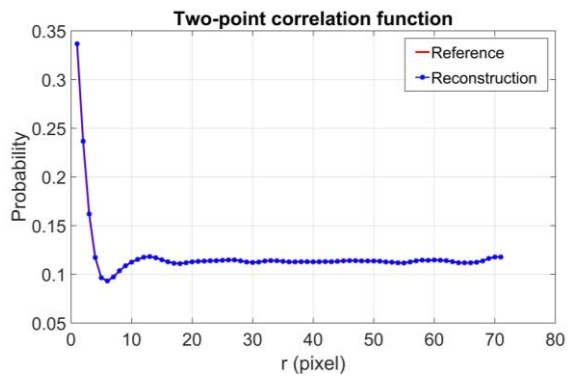
(b1)



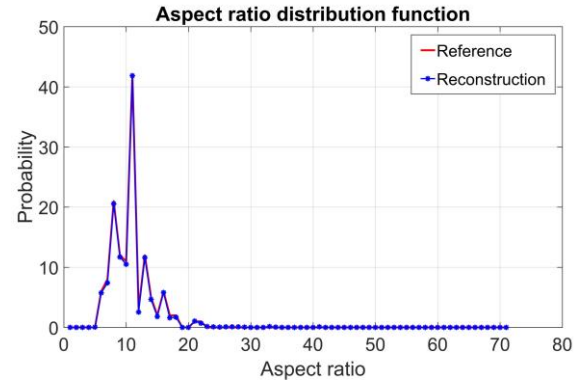
(a2)



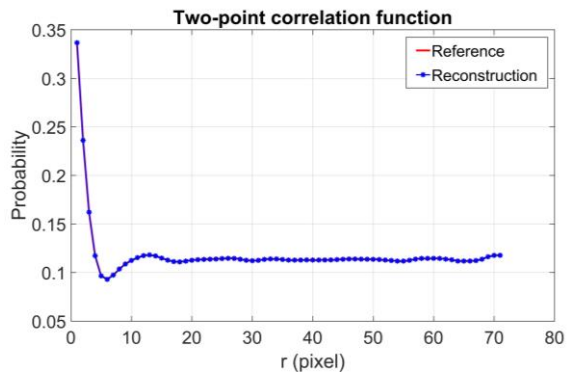
(b2)



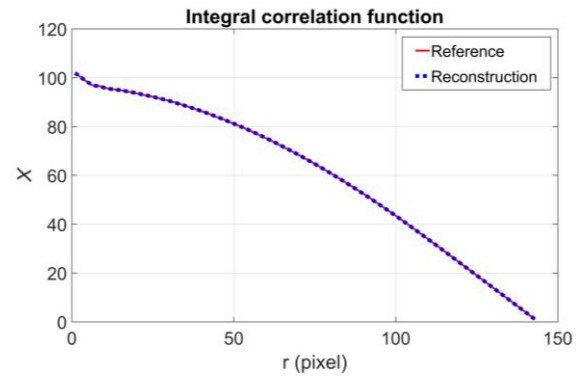
(a3)



(b3)



(a4)



(b4)

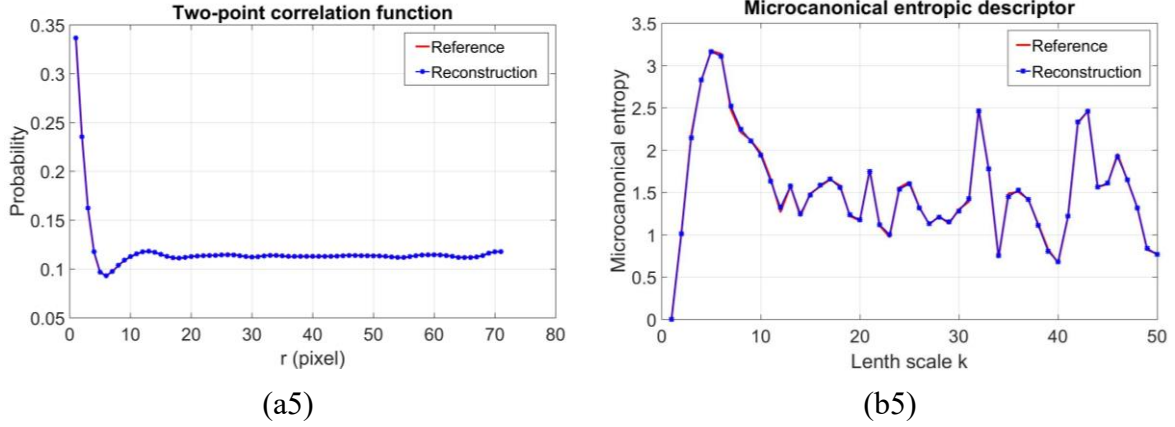


Figure 4. Accuracy validation for the reconstruction database in Table 4. (a1) The effective two-point correlation function of dataset 6. (b1) The effective directional growth probability of dataset 6. (a2) The effective two-point correlation function of dataset 9. (b2) The effective two-point cluster correlation function of dataset 9. (a3) The effective two-point correlation function of dataset 12. (b3) The effective aspect ratio distribution function of dataset 12. (a4) The effective two-point correlation function of dataset 14. (b4) The effective integral correlation function of dataset 14. (a5) The effective two-point correlation function of dataset 16. (b5) The effective microcanonical entropic descriptor of dataset 16.

### 3.3. Dimension reduction using texture spectrum

To ensure the analysis is generic and unbiased, different types of heterogeneous materials are chosen as reference samples for the descriptor transformation described in Section 3.2 and the digital microstructures are all reconstructed with sufficient size and resolution to fully cover all morphological features that can potentially be captured by a descriptor. As a result, the data sets obtained from descriptor transformation are uniform and representative, but they are too large to be analyzed directly. Therefore, we introduce the texture spectrum approach to reduce the dimensionality of image data in a neutral manner, without boosting or suppressing the morphological information embedded in the microstructural images. Texture is an important spatial feature that is frequently used for image recognition and classification. The texture spectrum approach is a statistical method for texture analysis, which was first proposed by (He & Wang, 1990, &1991) (Wang & He, 1990). This approach is built on two concepts: texture unit and texture spectrum. The local texture information of an image can be captured by the texture unit, while the texture spectrum reveals the global feature of the image.

$E_1$	$E_2$	$E_3$
$E_8$		$E_4$
$E_7$	$E_6$	$E_5$

Figure 5. Texture unit

As shown in Figure 5, each pixel in a digital image has eight nearest neighbouring pixels surrounding it, and these pixels together compose the smallest complete unit. A set containing nine elements  $V = \{V_0 V_1 V_2 V_3 V_4 V_5 V_6 V_7 V_8\}$  is used to denote the unit, where  $V_0$  denotes the value of central pixel and  $V_i$  ( $i = 1, 2, \dots, 8$ ) denotes the values of the neighbouring pixels. The set contains eight elements corresponding to the eight neighbouring pixels is defined as the texture unit  $TU = \{E_1 E_2 E_3 E_4 E_5 E_6 E_7 E_8\}$ , where the value of  $E_i$  is determined by the following rules:

$$E_i = \begin{cases} 0, & \text{if } V_i < V_0 \\ 1, & \text{if } V_i = V_0 \\ 2, & \text{if } V_i > V_0 \end{cases} \quad (3)$$

As the digital reference samples in this study are binary images, the possible intensity values of the pixels  $V_i$  are 0 and 1. In this case, the value of  $E_i$  is determined by simpler rules as follows:

$$\begin{aligned} \text{For } V_0 = 1, \quad E_i &= \begin{cases} 0 & \text{if } V_i < V_0 \\ 1 & \text{if } V_i = V_0 \end{cases}, \quad \text{equivalent to } E_i = V_i \\ \text{For } V_0 = 0, \quad E_i &= \begin{cases} 1 & \text{if } V_i = V_0 \\ 2 & \text{if } V_i > V_0 \end{cases}, \quad \text{equivalent to } E_i = V_i + 1 \end{aligned} \quad (4)$$

Therefore, each element  $E_i$  of  $TU$  has one of two possible values, which means the combination of all eight elements generates  $2^8 = 256$  possible texture units.

The texture spectrum is defined as the occurrence frequency distribution of all texture units over the whole image. For binary images, the 256 texture units can be ordered and labelled as follows:

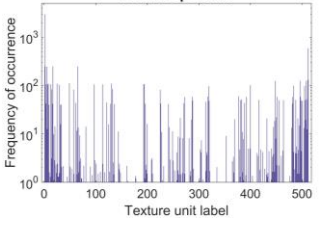
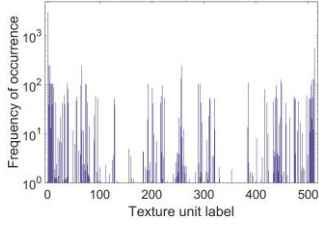
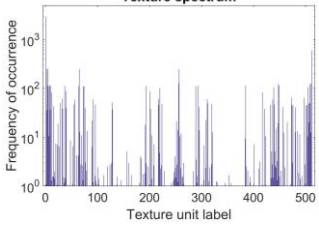
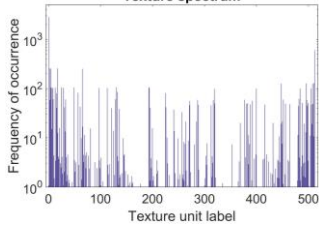
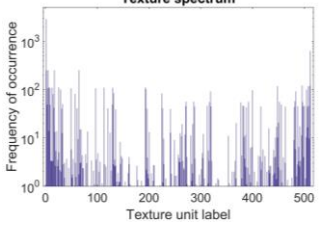
$$N_{TU} = \sum_{i=1}^8 E_i \cdot 2^{i-1} \quad (5)$$

The eight elements are ordered clockwise as shown in Figure 5. In this way, all 256 texture units can be labelled by consecutive integers in the interval of  $[0, 255]$ , and it is a one-to-one correspondence between the patterns of  $TU$  and the corresponding  $N_{TU}$  values. The 256 texture units represent local features of texture and capture the intensity-value difference between central pixels and their neighbouring pixels. The occurrence frequency of texture units over the whole image can be calculated to reveal the global information of the reference microstructure. For images with different textures, the corresponding texture spectrums distinguish them in a statistical sense and therefore can be used for image classification. As there are two possible values for each pixel (i.e. 0 and 1) of a binary image, the texture spectrum used in this study is the combination of two texture spectrums, where the texture spectrum for the central pixels of '0' is followed by the texture spectrum for the central pixels of '1'. Therefore, the final texture spectrum is defined as the interval  $[0, 511]$ .

The texture spectrum  $T_i$  is extracted from each set  $S_i$  (or row) of the reconstructed samples in Table 4, and for illustration purpose some examples are shown in Table 5. Specifically, the texture spectrum is first calculated for each sample, and then the overall texture spectrum  $T_i$  for the whole sample set  $S_i$  (containing 20 samples) is obtained by taking the

average. The texture spectrum is essentially a histogram of the texture unit label, of which each label index corresponds to a unique texture pattern. Therefore, the overall texture spectrum for the whole sample set can also be obtained by counting the occurrences of texture patterns directly on all 20 reconstructed samples. These two ways of calculating  $T_i$  are equivalent and give the same result. Therefore, instead of dealing with a large number of binary images (Table 4) that capture the statistical information of heterogeneous morphological descriptors (Table 2), the correlation analysis can now be carried out on the texture spectrums (Table 5) whose data sets are of much smaller dimensionality.

Table 5. Texture spectrums of the reconstructed samples in Table 4

No.	Involved Descriptors	Effective texture spectrum	No.	Involved Descriptors	Effective texture spectrum
1	Volume Fraction + Two-point Correlation Function	 <p style="text-align: center;"><math>T_1</math></p>	4	Pore-size Distribution Function + Two-point Correlation Function	 <p style="text-align: center;"><math>T_4</math></p>
2	Specific Internal Surface Density Function + Two-point Correlation Function	 <p style="text-align: center;"><math>T_2</math></p>	5	Core Distribution Function + Two-point Correlation Function	 <p style="text-align: center;"><math>T_5</math></p>
3	Grain-size Distribution Function + Two-point Correlation Function	 <p style="text-align: center;"><math>T_3</math></p>	6	... ..	... ..

### 3.4. Correlation analysis with K-means clustering

After the data regularization step described in Section 3.2 and the dimension reduction step described in Section 3.3, the 41 morphological descriptors  $D_i$  listed in Table 2 are now converted into 41 texture spectrums  $T_i$ , as illustrated in Table 5. That is, the task of correlation analysis of morphological descriptors can now be performed on a set of 41 texture spectrums

$T_i$ , which corresponds to the morphological descriptor  $D_i$  and is obtained from the corresponding sample set  $S_i$  with 20 reconstructed samples to avoid bias.

Noting that each texture spectrum is expressed as a vector of 512 entries or a point in a 512-dimensional space, we use  $K$ -means clustering in this study to perform the correlation analysis for the texture spectrums. Widely used in signal processing and data mining,  $K$ -means clustering (Lloyd, 1982) is a standard method of vector quantization, and in the wider context of machine learning it can be viewed as an unsupervised method. Given a set of observations  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ , where each observation is a  $d$ -dimensional real vector, i.e. a point in the  $d$ -dimensional space, the goal of the algorithm is to partition these  $n$  observations into  $k$  ( $\leq n$ ) clusters in which each observation belongs to the cluster with the nearest mean. In other words, the data space formed by  $n$  observations is portioned into  $k$  Voronoi cells such that an observation is always closer to the centre of the cell it falls in than the centres of other cells.

Specifically, starting with an initial estimation for  $k$  means  $\mathbf{m}_1^{(1)}, \mathbf{m}_2^{(1)}, \dots, \mathbf{m}_k^{(1)}$  with a partition  $\{P_1^{(1)}, P_2^{(1)}, \dots, P_k^{(1)}\}$ , these  $n$  observations are partitioned into  $k$  ( $\leq n$ ) sets  $\{P_1, P_2, \dots, P_k\}$  through an iterative procedure:

- Assignment step: Assign each observation to the cluster whose mean has the least squared Euclidean distance, i.e.

$$P_i^{(t)} = \left\{ \mathbf{x}_p \mid \left\| \mathbf{x}_p - \mathbf{m}_i^{(t)} \right\|^2 \leq \left\| \mathbf{x}_p - \mathbf{m}_j^{(t)} \right\|^2 \forall j, 1 \leq j \leq k \right\} \quad (6)$$

where the superscript  $t$  denotes the current iteration step. The assignment step allocates each observation  $\mathbf{x}_p$  to the cluster with the closest centroid (i.e. mean). If there are more than one closest centroid (mean) for some observations, these observations can be randomly assigned to one of the corresponding clusters. An example of a 2D cluster assignment is shown in Figure 6.

- Update step. Update the  $k$  means sets  $P_i$  by calculating the new means to be the centroids of the observations in the new clusters, i.e.

$$\mathbf{m}_i^{(t+1)} = \sum_{\mathbf{x}_j \in P_i^{(t)}} \mathbf{x}_j / \left| P_i^{(t)} \right| \quad (7)$$

where  $\left| P_i^{(t)} \right|$  denotes the number of points in the set  $P_i^{(t)}$ .

- Repeat the above two steps until the stopping criterion is met (i.e. the cluster assignments do not change, or the maximum number of iterations is reached).

It should be noted that the clustering result from the above algorithm depends on the initial choice of  $k$  means  $\mathbf{m}_1^{(1)}, \mathbf{m}_2^{(1)}, \dots, \mathbf{m}_k^{(1)}$ , and it may not reach the global optimum. However, by having enough runs of the algorithm with randomized initial sets of  $k$  means, the optimum clustering can be effectively achieved.



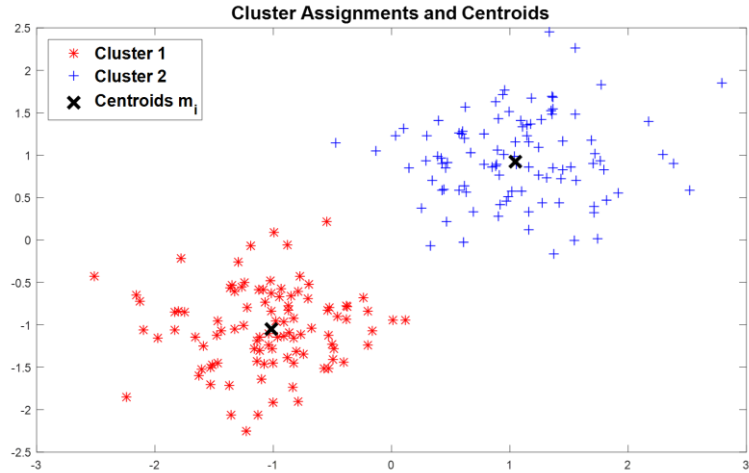


Figure 6. Example of a 2D cluster assignment

By utilizing the  $K$ -means clustering technique, the ‘descriptor points’ can be categorized into groups of any fixed number. The descriptors categorized in the same group are more similar (or more correlated) with each other, while descriptors categorized in different groups are more different (or less correlated) with each other. In this study, the number of groups is set from 2 to 15 such that the categorization is performed at different levels from coarse to fine. The grouping results from 2 to 15 are summarized in Tables 6, where the letters A, B, ..., O denote the cluster, the ID numbers 1, 2, ..., 41 represent the corresponding descriptors as defined in Table 2, and the descriptors that change categories as the grouping refines are highlighted. To save space, Table 6 only lists the results of 2-group, 3-group, 4-group, and 15-group clustering.

Table 6  $K$ -means clustering of morphological descriptors

(6.1) Two-groups categorization

A	1	3	5	7	8	9	11
	12	13	14	15	16	18	19
	20	21	22	23	24	25	26
	27	28	29	30	31	32	33
	34	35	36	37	38	39	40
	41						
B	2	4	6	10	17		

(6.2) Three-groups categorization

A	1	8	9	13	14	16	18
	19	20	21	27	28	29	31
	33	34	35	36	37	38	39
	40	41					
B	2	4	6	10	17		
C	3	5	7	11	12	15	22
	23	24	25	26	30	32	

(6.3) Four-groups categorization

A	1	8	9	13	14	16	18
	19	20	21	27	28	29	31
	33	34	35	36	37	38	39
	40	41					
B	2	4	6				
C	3	5	7	11	12	15	22
	23	24	25	26	30	32	
D	10	17					

.....

(6.14) Fifteen-groups categorization

A	8	18	27	31	33	34	35
	36	37	38	39	40	41	
B	4	6					
C	5	25					
D	10						
E	12	22	24	26	30		
F	9						
G	11						
H	13	15	20	21	23		
I	32						
J	28						
K	7						
L	3						
M	2						
N	1	14	16	19	29		
O	17						

The full clustering results are summarized in Table 7, where the individual descriptors are represented by the columns and their classification into different numbers of categories are represented the rows. Closely correlated descriptors are placed in nearby columns and represented by a color code. From the top row to the bottom row, the descriptor classification changes from 2 groups to 15 groups, and the color code clearly demonstrates the clustering evolution, which indicates the relationship between descriptors.

Table 7. Correlation graph of all 41 descriptors in Table 2

	8	18	27	31	33	34	35	36	37	38	39	40	41	1	16	28	14	19	29	13	20	21	15	23	12	22	24	26	30	32	9	11	7	3	5	25	4	6	2	10	17								
2-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	
3-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	C	C	C	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	C	B	B	B	B	B
4-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	C	C	C	C	C	C	A	C	C	C	C	C	C	C	C	C	C	C	C	B	B	B	D	D	
5-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	C	C	C	C	C	C	C	C	C	B	B	B	D	D
6-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	E	E	E	E	E	E	E	E	E	E	F	F	C	C	C	C	C	C	C	C	C	C	C	C	B	B	B	D	D	
7-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	E	E	E	E	E	E	E	E	E	E	F	F	G	C	C	C	C	C	C	C	C	C	C	B	B	B	D	D		
8-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H	H	H	H	H	H	E	E	E	E	E	E	F	G	C	C	C	C	C	C	C	C	B	B	B	D	D			
9-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	H	H	H	H	H	H	H	H	E	E	E	E	E	I	F	G	C	C	C	C	C	C	C	C	C	B	B	B	D	D			
10-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	J	H	H	H	H	H	H	H	H	E	E	E	E	E	E	I	F	G	C	C	C	C	C	C	C	C	B	B	B	D	D			
11-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	J	H	H	H	H	H	H	H	H	E	E	E	E	E	I	F	G	K	C	C	C	C	C	C	C	C	B	B	B	D	D			
12-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	J	H	H	H	H	H	H	H	H	E	E	E	E	E	I	F	G	K	L	C	C	C	C	C	C	C	B	B	B	D	D			
13-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	J	H	H	H	H	H	H	H	H	E	E	E	E	E	I	F	G	K	L	C	C	C	C	C	C	B	B	M	D	D				
14-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	N	N	J	N	N	N	H	H	H	H	H	E	E	E	E	E	I	F	G	K	L	C	C	C	C	C	C	B	B	M	D	D				
15-Groups	A	A	A	A	A	A	A	A	A	A	A	A	A	N	N	J	N	N	N	H	H	H	H	H	E	E	E	E	E	I	F	G	K	L	C	C	C	C	C	C	B	B	M	D	O				

Legend

1	Volume Fraction	2	Specific Internal Surface Density	3	Grain-size Distribution Function	4	Pore-size Distribution Function	5	Core Distribution Probability
6	Directional Growth Probability	7	Orientation Distribution Function	8	Two-point Correlation Function	9	Two-point Cluster Correlation Function	10	Lineal-path Function
11	Chord-length Density Function	12	Aspect Ratio Distribution Function	13	Radial Distribution Function	14	Integral Correlation Function	15	Topological Entropic Descriptor
16	Microcanonical Entropic Descriptor	17	Local Percolation Probability	18	Local Porosity Distribution Function	19	Tortuosity	20	Voronoi Polygon Area
21	Nearest-neighbor Distribution Function (1)_ep	22	Nearest-neighbor Distribution Function (2)_ev	23	Nearest-neighbor Distribution Function (3)_hp	24	Nearest-neighbor Distribution Function (4)_hv	25	Nearest-neighbour Orientation Distribution Function
26	Ripley's K Function	27	Coarseness	28	Contiguity	29	Point/q-particle Correlation Function	30	Surface-particle Correlation Function
31	Surface Correlation Function (1)_SSC	32	Surface Correlation Function (2)_SVC	33	Maximum Grain Radius	34	Minimum Grain Radius	35	Mean
36	Variance	37	Standard Deviation	38	Autocovariance	39	Autocorrelation Function	40	Power Spectral Density Functions
41	Margin Distribution Function								

## 4. Result Discussions

### 4.1. Accuracy validation

Some descriptors have explicit mathematical expressions, while others do not. This inconsistency makes it impossible to conduct analytical correlation analysis uniformly for all descriptors. However, for some specific descriptors, existing analytical studies have revealed strong correlations between them. For example, the information of volume fraction is fully contained in the two-point correlation function, and as a result the combination of the volume fraction and the two-point correlation function should represent identical statistical information as the sole two-point correlation function. Table 8 provides a summary of various analytical relations between specific descriptors and their associated references. Moreover, based on physical insights, some descriptors are also known to have strong links with each other, and these physics-based correlation results are summarized in Table 9. To validate the accuracy of the proposed correlation analysis, we compare our generic results with the special cases shown in Table 8 and Table 9. The comparison results are shown in Table 10, from which it can be observed all known correlations are accurately identified by our machine-learning based correlation analysis. This comparison is not a comprehensive validation, but to an extent it does confirm the effectiveness of the proposed generic correlation analysis strategy and the accuracy of its results.

Table 8. The analytically derived relations

The descriptors with validated strong links	References
Volume Fraction $D_1$ – Two-point Correlation Function $D_8$	(Torquato, 1999)
Coarseness $D_{27}$ – Local Porosity Distribution Function $D_{18}$	(Lu & Torquato, 1990)
Autocorrelation Function $D_{39}$ – Two-point Correlation Function $D_8$	(Jiao, et al.,2008)
Autocorrelation Function $D_{39}$ – Power Spectrum Density Function $D_{40}$	(Wiener, 1964)
Autocorrelation Function $D_{39}$ – Autocovariance $D_{38}$	(Hoel, 1984)
Integral Correlation Function $D_{14}$ – Radial Distribution Function $D_{13}$	(Bochenek & Pyrz, 2004)

Table 9. The relations between morphological descriptors based on their physical insights

The descriptors with strong links	The related physical insight
Nearest-neighbour Orientation Distribution Function $D_{25}$ - Orientation Distribution Function $D_7$	The orientations of the clusters
Volume Fraction $D_1$ - Local Porosity Distribution Function $D_{18}$	The component ratio
Nearest-Neighbour Distribution Function(1)_ep $D_{21}$ - Nearest- Neighbour Distribution Function(3)_hp $D_{23}$	Opposite to each other

Nearest-Neighbour Distribution Function(2)_ev $D_{22}$ - Nearest-Neighbour Distribution Function(4)_hv $D_{24}$	Opposite to each other
---	------------------------

Table 10. Comparison between the proposed generic correlation analysis and the special case correlation results in Table 8 and Table 9

<b>Correlated descriptors</b>	<b>Our correlation analysis grouping</b>
Volume Fraction $D_1$ – Two-point Correlation Function $D_8$	From 2-groups categorization to 8-groups categorization
Integral Correlation Function $D_{14}$ – Radial Distribution Function $D_{13}$	From 2-groups categorization to 13-groups categorization
Coarseness $D_{27}$ – Local Porosity Distribution Function $D_{18}$	From 2-groups categorization to 15-groups categorization
Autocorrelation Function $D_{39}$ – Two-point Correlation Function $D_8$	From 2-groups categorization to 15-groups categorization
Autocorrelation Function $D_{39}$ – Power Spectrum Density Function $D_{40}$	From 2-groups categorization to 15-groups categorization
Autocorrelation Function $D_{39}$ – Autocovariance $D_{38}$	From 2-groups categorization to 15-groups categorization
Nearest-neighbour Orientation Distribution Function $D_{25}$ - Orientation Distribution Function $D_7$	From 2-groups categorization to 10-groups categorization
Volume Fraction $D_1$ - Local Porosity Distribution Function $D_{18}$	From 2-groups categorization to 13-groups categorization
Nearest-Neighbour Distribution Function(1)_ep $D_{21}$ - Nearest-Neighbour Distribution Function(3)_hp $D_{23}$	From 2-groups categorization to 15-groups categorization
Nearest-Neighbour Distribution Function(2)_ev $D_{22}$ - Nearest-Neighbour Distribution Function(4)_hv $D_{24}$	From 2-groups categorization to 15-groups categorization

## 4.2. Correlation between morphological descriptors

Table 6 and Tables 7 summarize the correlation grouping for all descriptors listed in Table 2. Some descriptors are more correlated with each other and others are less correlated, and all these correlations are relative. At different precision levels, these descriptors are categorized into different numbers of groups, ranging from 2 to 15. The correlation between descriptors can be examined by making lateral and longitudinal comparisons in Tables 6.1-6.14. The lateral comparison focuses on the classification at a fixed precision level. The descriptors categorized into the same group are more correlated than the descriptors in different groups, and such correlation can be examined at different precision levels. The longitudinal comparison focuses on the changing trend when the precision of classification increases or decreases. As the precision increases, the descriptors that are removed from the current groups have less correlation than the descriptors that remain in the current groups. For instance, the surface-

particle correlation function  $D_{30}$ , the surface-surface correlation function  $D_{31}$ , and the surface-void correlation function  $D_{32}$  are all classified into Group A in Table 6.1. However, at the next level of precision, the surface-surface correlation function  $D_{31}$  is moved to another group. This implies that the surface-particle correlation function  $D_{30}$  is more correlated to the surface-void correlation function  $D_{32}$  than to the surface-surface correlation function  $D_{31}$ .

A graphical illustration for the correlation of all 41 descriptors is given in Table 7, where the columns represent individual descriptors  $D_i$ ,  $i = 1, 2, \dots, 41$ , the rows represent their classifications, and the letters A, B, C, ..., O represent the same clustering as summarized in Table 6. Closely correlated descriptors are placed in nearby columns and represented by a color code. From the top row to the bottom row, the descriptor classification changes from 2 groups to 15 groups, and the color code clearly demonstrates the clustering evolution, which indicates the relationship between descriptors.

## 5. Conclusion

This study proposes a flexible and robust method for uniform correlation analysis of various statistical descriptors that characterize the morphology of heterogeneous materials. Based on image processing and machine learning, the new correlation analysis strategy comprises three operational steps: data regularization, dimension reduction and correlation analysis. Specifically, the data regularization is achieved by heterogeneous material modelling that transforms the statistical information represented by individual morphological descriptors of different formats into reconstructed microstructure images of uniform size and resolution. The dimension reduction is achieved by using the texture spectrum approach that converts the large data set of microstructural images into a texture spectrum vector of moderate dimension. The correlation analysis is achieved via  $K$ -means clustering that categorizes the descriptor points into different numbers of groups. The new method overcomes the fundamental challenge caused by the intrinsic data heterogeneity associated with various morphological descriptors. For the first time, the quantitative relations between various morphological descriptors are uniformly established. The improved understanding of interrelation of morphological descriptors can benefit widely the characterization, reconstruction, and property prediction/optimization of heterogeneous materials. A total of 41 morphological descriptors are analysed in this study, but the method is generic and can be readily extended to include other descriptors. In addition, the idea of the proposed unified correlation analysis strategy could be beneficial to other engineering problems with the presence of data heterogeneity.

This study is not without limitation. First, this study uses texture spectrum for dimension reduction and  $K$ -means clustering for correlation analysis, but there are other ways to perform dimension reduction on image data sets and to carry out correlation analysis on feature vectors. Different methods for dimension reduction and correlation analysis may result in slightly different results, which is worth for future studies. Secondly, this study relies on 2D reference samples and 2D microstructure reconstruction to transform heterogeneous statistical descriptors into unified data sets (i.e. binary images of the same size and resolution) but at a

much higher computational cost, using 3D reference samples and 3D microstructure reconstruction is arguably more representative and more accurate for the correlation analysis. This will also be pursued in our future work.

## 6. Acknowledgement

The authors would like to thank the support from China Scholarship Council (CSC Number: 201608440279), Swansea University (Zienkiewicz Scholarship), and the Royal Society (Ref.: IECnNSFCn191628).

## Reference

- Adams, B.L., Canova, G.R., Molinari, A., (1989). A statistical formulation of viscoplastic behavior in heterogeneous polycrystals. *Textures and Microstructures*, 11: p. 57–71.
- Alexander, S. K., Fieguth, P., Ioannidis, M. A., & Vrscay, E. R. (2009). Hierarchical Annealing for Synthesis of Binary Images. *Mathematical Geosciences*, 41(4): p. 357-378.
- Brown, J.W.F. (1955). Solid mixture permittivities. *J. Chem. Phys.* 23, 4.
- Beran, M.J. (1968). *Statistical Continuum Theories*. Switzerland: Interscience Publishers.
- Berryman, J.G. & Milton, G.W. (1985). Normalization constraint for variational bounds on fluid permeability, *J.Chem. Phys*, 83: p. 754-760.
- Bunge, H. J. (1982). *Texture Analysis in Materials Science: Mathematical Methods*, Butterworth& Co, London, UK.
- Bland, J.M., Altman, D.G. (1996). Statistics notes: measurement error. *BMJ*, 312 (7047): p. 1654.
- Bochenek, B., & Pyrz, R. (2004). Reconstruction of random microstructures - a stochastic optimization problem. *Computational Materials Science*, 31(1-2): p. 93-112.
- Baniassadi, M., Mortazavi, B., Hamedani, H. A., Garmestani, H., Ahzi, S., Fathi-Torbaghan, M., . . . Khaleel, M. (2012). Three-dimensional reconstruction and homogenization of heterogeneous materials using statistical correlation functions and FEM. *Computational Materials Science*, 51(1): p. 372-379.
- Carman, P. C. (1937). Fluid Flow Through Granular Beds. *Trans. Inst. Chem. Eng.*, 15, p. 150–166.
- Cule, D., & Torquato, S. (1999). Generating random media from limited microstructural information via stochastic optimization. *Journal of Applied Physics*, 86(6): p. 3428-3437.
- Caers, J., (2001). Geostatistical reservoir modelling using statistical pattern recognition. *Journal of Petroleum Science and Engineering*, 29 (3–4): p.177–188.
- Cowan, G. (2002). *Statistical data analysis*. Oxford: Clarendon Press.
- Capek P., Hejtmanek V., Brabec L., Zikanova A., and Kocirik M. (2009), Stochastic Reconstruction of Particulate Media Using Simulated Annealing: Improving Pore Connectivity. *Transport in Porous Media*. 76, 179.
- Capek, P., Hejtmanek, V., Kolafa, J., & Brabec, L. (2011). Transport Properties of Stochastically Reconstructed Porous Media with Improved Pore Connectivity. *Transport in Porous Media*, 88(1), p. 87-106.
- Chen, D., Jiao, Y., Torquato, S., Equilibrium Phase Behavior and Maximally Random Jammed State of Truncated Tetrahedra. *Journal of Physical Chemistry B*. 2014, 118 (28), 7981-7992.
- Cui, S., Li, C., Owen, D., (2015). A comparison study of statistical reconstruction of heterogeneous materials, *Proceedings of the 23rd UK conference of the Association for Computational Mechanics in Engineering*.
- Cui, S., Li, C., Owen, D., (2016). Stochastic reconstruction of heterogeneous media, *Proceedings of Euromech Colloquium 584 Multi-uncertainty and multi-scale methods and Related Applications*.
- Cui, S., Li, C., Owen, D., (2016). Multi-scale weave algorithm for statistical reconstruction of heterogeneous materials, *Proceedings of European Congress on Computational Methods in Applied Sciences and Engineering*.
- Cui, S., Li, C., Owen, D., (2017). Statistical reconstruction of heterogeneous materials, *Proceedings of Computational Modelling of Multi-uncertainty and Multi-scale Problems*.
- Doi, M. (1976). A new variational approach to the diffusion and the flow problem in porous media, *J. Phys. Soc. Japan*, 40: p. 567-572.
- Debye, P., and Bueche, A. M., (1949), Scattering by an Inhomogeneous Solid. *J. Appl. Phys.*, 20: p. 518–525.
- Eschricht, N., Hoinkis, E., Madler, F., Schubert-Bischoff, P., & Rohl-Kuhn, B. (2005). Knowledge-based reconstruction of random porous media. *Journal of Colloid and Interface Science*, 291(1): p. 201-213.
- Fullwood, D.T., Niezgodna, S.R. & Kalidindi, S.R. (2008). Microstructure reconstructions from 2-point statistics

- using phase-recovery algorithms, *Acta Materialia*, 56, p. 942-948.
- Fullwood, D. T., Niezgoda, S. R., Adams, B. L., & Kalidindi, S. R. (2010). Microstructure sensitive design for performance optimization. *Progress in Materials Science*, 55(6): p. 477-562.
- Feng, J.W., Li, C.F., & Owen, D.R.J. (2014). Statistical reconstruction of two-phase random media, *Computers and Structures*, 137: p. 78-92.
- Feng, J.W., Li, C.F. & Owen, D.R.J. (2016). Statistical reconstruction and Karhunen-Loeve expansion for multiphase random media, *International Journal for Numerical Methods in Engineering*. 105: p. 3-32.
- Gurland, J. (1958). The measurement of grain contiguity in two-phase alloys. *Trans. Metall. Soc. AIME*, 212 452.
- Guo, E.-Y.; Chawla, N.; Jing, T.; Torquato, S.; Jiao, Y., Accurate modeling and reconstruction of three-dimensional percolating filamentary microstructures from two-dimensional micrographs via dilation-erosion method. *Materials Characterization* 2014, 89, 33-42.
- Hertz, P. (1909). Über den gegenseitigen durchschnittlichen Abstand von Punkten die mit bekannter mittlerer Dichte im Raume angeordnet sind. *Math. Ann.* 67: p. 387-98.
- Ho, F.G. & Strieder, W. (1979). Asymptotic expansion of the porous medium, effective diffusion coefficient in the Knudsen number, *J. Chem. Phys.*, 70: p. 5635-5639.
- Hoel, P. G. (1984). *Mathematical Statistics* (Fifth ed.). New York: Wiley. ISBN 0-471-89045-6.
- He, D. C. & Wang, L. (1990). Texture Unit, Texture Spectrum, and Texture Analysis. *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, 28(4).
- He, D. C. & Wang, L. (1991). Texture feature based on Texture Spectrum. *Pattern Recognition*, 24(5): p.391-399.
- Hilfer, R. (1991). Geometric and dielectric characterization of porous media. *Physical Review B*, 44(1): p. 60-75.
- Hilfer, R., Rage, T., & Virgin, B. (1997). Local percolation probabilities for a natural sandstone. *Physica A*, 241: p. 105-110
- Han, T. S., & Dawson, P. R. (2005). Representation of anisotropic phase morphology. *Modelling and Simulation in Materials Science and Engineering*, 13(2): p. 203-223.
- Jennings, B.R., Parslow, K., (1988). Particle size measurements: the equivalent spherical diameter. *P. Roy. Soc. A Math. Phys.*, 419: p.139-149.
- Jiao, Y., Stillinger, F. H., & Torquato, S. (2008). Modeling heterogeneous materials via two-point correlation functions. II. Algorithmic details and applications. *Physical Review E*, 77(3).
- Jiao, Y., Stillinger, F. H., & Torquato, S. (2010). Geometrical ambiguity of pair statistics: Point configurations. *Physical Review E*, 81: p. 11105.
- Keller, J.B., Rubinfeld, L. & Molyneux, J. (1967). Extremum principles for slow viscous flows with applications to suspensions. *J. Fluid Mech.*, 30(1): p. 97-125.
- Kikkinides, E. S., & Politis, M. G. (2014). Linking pore diffusivity with macropore structure of zeolite adsorbents. Part I: three dimensional structural representation combining scanning electron microscopy with stochastic reconstruction methods. *Adsorption-Journal of the International Adsorption Society*, 20(1): p. 5-20.
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*. 28 (2): 129-137.
- Lu, B., & Torquato, S. (1990). Local volume fraction fluctuations in heterogeneous media. *J. Chem. Phys.*, 93, 3452.
- Lu, B. & Torquato, S. (1992a). Lineal-path function for random heterogeneous materials. *Physical Review A*, 45(2): p. 922-929.
- Lu, B. & Torquato, S. (1992b). Lineal-path function for random heterogeneous materials. *Physical Review A*, 45(10): p. 7292-7301.
- Li, D. S., Tschopp, M. A., Khaleel, M., & Sun, X. (2012). Comparison of reconstructed spatial microstructure images using different statistical descriptors. *Computational Materials Science*, 51: p. 437-444.
- Li, D. (2014). Review of Structure Representation and Reconstruction on Mesoscale and Microscale. *Journal of Metals*, 66(3): p. 444-454.
- Li, L., Sheng, L.M. & Proust, G. (2015). Generalised Voronoi tessellation for generating microstructural finite element models with controllable grain-size distributions and grain aspect ratios, *Internal Journal For Numerical Methods in Engineering*, 103: p. 144-156
- Li, X., Liu, Z.L., Cui, S.Q., Luo, C.C., Li, C.F., & Zhuang, Z. (2019). Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning, *Computer Methods in Applied Mechanics and Engineering*, 347: p. 735-753.
- Manwart, C., Torquato, S., & Hilfer, R. (2000). Stochastic reconstruction of sandstones. *Physical Review E*, 62(1): p. 893-899.
- Prager, S. (1961). Viscous flow through porous media, *Phys. Fluids*, 4: p. 1477-1482.
- Priestley, M. B. (1982). *Spectral analysis and time series*. London, New York: Academic Press. ISBN 0125649010.



- Piasecki, R. (2000) Entropic measure of spatial disorder for systems of finite-sized objects. *Physica A*, 277: p. 157–173.
- Pyrz, R. & Bochenek, B. (2003). On application of stochastic optimization to reconstruction of spatial random microstructures. Proceedings of the 5th. world congress of structural and multidisciplinary optimization, 2003
- Piasecki, R. (2009). A versatile entropic measure of grey level inhomogeneity. *Physica A: Statistical Mechanics and its Applications*, 388(12): p. 2403-2409.
- Quey, R. & Renversade, L. (2018). Optimal polyhedral description of 3D polycrystals: Method and application to statistical and synchrotron X-ray diffraction data. *Computer Methods in Applied Mechanics and Engineering*, 330: p. 308-333.
- Quintanilla, J. (2006). Measures of clustering in systems of overlapping particles. *Mechanics of Materials*, 38(8-10): p. 849-858.
- Reiss, H., Frisch, H.L., Lebowitz, J.L. (1959). Statistical Mechanics of Rigid Spheres. *J. Chem. Phys.* 31: p. 369–380.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, 13: p. 255-266.
- Ripley, B. D. (1977). Modelling spatial patterns, *Journal of the Royal Statistical Society, Series B*, 39: p. 172-192.
- Rubinstein, J., & Torquato, S. (1988). Diffusion-controlled reactions: Mathematical formulation, variational principles, and rigorous bounds. *J. Chem. Phys.* 88 (10).
- Rubinstein, J., & Torquato, S. (1989). Flow in random porous media: mathematical formulation, variational principles, and rigorous bounds. *J. Fluid Mech.*, 206: p. 25-46.
- Roberts, A. P. (1997a). Statistical reconstruction of three-dimensional porous media from two-dimensional images. *Physical Review E*, 56(3): p. 3203-3212.
- Roberts, A. P. (1997b). Morphology and thermal conductivity of model organic aerogels. *Phys. Rev. E*, 55: p. 1282.
- Roberts, A. P. and Torquato, S. (1999). Chord-distribution functions of three-dimensional random media: Approximate first-passage times of Gaussian processes, *Phys. Rev. E*, 59: p.4953-4963.
- Rollett, A. D., Lee, S. B., Campman, R., & Rohrer, G. S. (2007). Three-dimensional characterization of microstructure by electron back-scatter diffraction. *Annual Review of Materials Research*, 37: p. 627-658.
- Rahman, S. (2008). A random field model for generating synthetic microstructures of functionally graded materials, *Internal Journal For Numerical Methods in Engineering*, 76(7): p. 972-993.
- Ryu, S., & Li, D. (2012). OPTIMIZING STOCHASTIC PROCESS FOR EFFICIENT MICROSTRUCTURE RECONSTRUCTION. *Proceedings of the 1st International Conference on 3d Materials Science*, p. 171-176.
- Schiemann, A. (1997). Ternary positive definite quadratic forms are determined by their theta series, *Math. Ann.* 308: p.507-517.
- Strebelle S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34: p. 1-21.
- Strebelle, S., Payrazyan, K., Caers, J., (2003). Modeling of a deepwater turbidite reservoir conditional to seismic data using principal component analysis and multiple-point geostatistics. *SPE Journal*, 8 (3): p. 227–235.
- Sundararaghavan, V., Kumar, A., & Sun, S. (2015). Crystal plasticity simulations using nearest neighbor orientation correlation function. *Acta Materialia*, 93: p. 12-23.
- Tassopoulos, M. & Rosner, D. E. (1992). Simulation of vapor diffusion in anisotropic particulate bubbles, *Proc. R. Soc. Lond. A*, 226: p. 34-37.
- Teubner, M. & Strey, R. (1987). Origin of the scattering peak in microemulsions. *J. Chem. Phys.* 87,3195.
- Torquato, S., & Stell, G. (1982). Microstructure of two-phase random media. I. The n-point probability functions. *J. Chem. Phys.* 77(4).
- Tokunaga, T.K. (1985). Porous media gas diffusivity from a free path distribution model, *J. Chem. Phys.* 82, p. 5298-5299.
- Torquato, S. & Lado, F. (1985). Characterisation of the microstructure of distributions of rigid rods and discs in a matrix. *J. Phys. A: Math. Gen.*, 18: p. 141-148.
- Torquato, S. (1986). Bulk properties of two-phase disordered media. III.. New bounds on the effective conductivity of dispersions of penetrable spheres, *J. Chem. Phys.* 84: p. 6345-6359.
- Torquato, S. & Beasley, J. D. (1987). Bounds on the permeability of a random array of partially penetrable spheres, *Phys. Fluids*, 30: p. 633-641.
- Torquato, S. & Avellaneda, M. (1991). Diffusion and reaction in heterogeneous media: Pore size distribution, relaxation times, and mean survival time. *The Journal of Chemical Physics*, 95(9).
- Torquato, S., & Lu, B. (1993). Chord-length distribution function for two-phase random media. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics*, 47(4): p. 2950-2953.
- Torquato, S. & Rintoul, M. D. (1995). Effect of the interface on the properties of composite media, *Phys. Rev. E*,

- 47: p. 2950-2953.
- Torquato, S. (1999). Exact conditions on physically realizable correlation functions of random media, *The Journal of Chemical Physics*, 111: p. 8832.
- Torquato, S. (2000). *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer, New York, USA, 2000.
- Torquato, S. (2002). Statistical description of microstructures. *Annual Review of Materials Research*, 32: p. 77-111.
- Tahmasebi, P., & Sahimi, M. (2013). Cross-Correlation Function for Accurate Reconstruction of Heterogeneous Media. *Physical Review Letters*, 110(7).
- Underwood, E. E. (1970). *Quantitative Stereology*, Addison – Wesley, Reading, Massachusetts.
- Wiener, N. (1964). *Time Series*. M.I.T. Press, Cambridge, Massachusetts. p. 42.
- Wang, L. & He, D. C. (1990). Texture Classification Using Texture Spectrum. *Pattern Recognition*, 23(8): p.905-910.
- Wang, Z. M., Kwan, A. K. H., & Chan, H. C. (1999). Mesoscopic study of concrete I: generation of random aggregate structure and finite element mesh. *Computers & Structures*, 70(5): p. 533-544.
- Watanabe, Y & Nakashima, Y (2001). Two-Dimensional Random Walk Program for the Calculation of the Tortuosity of Porous Media. *J-Stage*, 43(1): p. 13-22. Wang, M., & Chen, S. (2007). Electroosmosis in homogeneously charged micro- and nanoscale random porous media. *Journal of Colloid and Interface Science*, 314(1): p. 264-273.
- Yeong, C. L. Y., & Torquato, S. (1998a). Reconstructing random media. *Physical Review E*, 57(1): p. 495-506.
- Yeong, C. L. Y., & Torquato, S. (1998b). Reconstructing random media. II. Three-dimensional media from two-dimensional cuts. *Physical Review E*, 58(1): p. 224-233.
- Yang, M., Nagarajan, A., Liang, B., & Soghrati, S. (2018). New algorithms for virtual reconstruction of heterogeneous microstructures. *Computer Methods in Applied Mechanics and Engineering*, 338: p. 275-298.
- Zhang, W.L., Song, L. & Li, J.J. (2019). Efficient 3D reconstruction of random heterogeneous media via random process theory and stochastic reconstruction procedure. *Computer Methods in Applied Mechanics and Engineering*, 354: p. 1-5.