Accepted Manuscript

Characterising particle packings by principal component analysis

Y.T. Feng, Tingting Zhao, Min Wang, D.R.J. Owen

PII: S0045-7825(18)30263-9
DOI: https://doi.org/10.1016/j.cma.2018.05.018
Reference: CMA 11919

To appear in: *Comput. Methods Appl. Mech. Engrg.*

Received date : 1 November 2017
Revised date : 21 May 2018
Accepted date : 22 May 2018

Please cite this article as: Y.T. Feng, T. Zhao, M. Wang, D.R.J. Owen, Characterising particle packings by principal component analysis, *Comput. Methods Appl. Mech. Engrg.* (2018), https://doi.org/10.1016/j.cma.2018.05.018

Characterising Particle Packings by Principal Component Analysis

Y. T. Feng*, Tingting Zhao, Min Wang, D. R. J. Owen
Zienkiewicz Centre for Computational Engineering, Swansea University, UK

Abstract

Particle packings play an important role in the discrete element modelling of particulate systems as different packings can lead to different physical behaviour, and therefore need to be properly characterised and controlled. Apart from a few conventional approaches, there is still a lack of more general, comprehensive and quantitative approaches that can reveal some fundamental features of packings. The current work attempts to develop a novel packing characterising system based on two techniques: digitalised image representation of a packing and subsequent application of Principal Component Analysis to the resulting image. It will prove that the principal components or variances of a packing image can indeed qualify as the signature of the packing, and therefore can be utilised to characterise the packing. Furthermore, a dissimilarity coefficient or a similarity index will be defined which provides a single valued metric to quantitatively compare two packings. Comprehensive investigations for two sets of purposefully generated random packings are conducted to fully understand relationships of their principal variances with packing features. Various issues, including effects of grid resolutions and packing density on principal variances are discussed. Methods of how to apply principal variances to assess spatial homogeneity and isotropy of packings are proposed. The relationship between scaled packings and their principal variances is also considered.

KEYWORDS: particle packing, digitalised image, Principal component analysis, principal variance, spatial homogeneity and isotropy

1 Introduction

Particle packings play an important role in the discrete element method [1] for modelling particulate systems as different packings can lead to different physical behaviour of the system, and therefore need to be properly characterised and controlled before being used for subsequent modelling. This is especially true for packings that are generated in a geometric fashion, see [2, 3, 4, 5, 6, 7, 8], for instance, for packing algorithms belonging to this category. Also in some critical applications, such as pebble bed reactors, detailed packing structures have an important impact on the system performance [9].

Conventional means to characterise a particle packing are limited to: particle size distribution, packing density or porosity. A more computationally expensive method involves the use of a radial correlation function to exploit the spatial distribution of a packing, but it is less effective. It is also possible to check the coordinate number distribution, and spatial isotropy via pairs of particle contacts, or their fabric tensor, when the packing is subject to

*e-mail: y.feng@swansea.ac.uk

(small) external loading. Nevertheless, in our opinion, there is still a lack of more general, comprehensive and quantitative approaches that can reveal some fundamental features of packings. In addition, several Monte Carlo samples of packing in DEM may need to be prepared for a given particle size distribution. However there are no existing approaches available that can compare and quantify the difference or similarity of such packings.

A particle packing can be viewed as a specific spatial variation of solid material (assigned a value of 1) and void (a value of 0). By discretisation of such a packing as a regular grid, and each grid cell is viewed as a pixel with a weighted value, the packing can be represented as a digitalised image with grey-scale ranging from 0 to 1. Then the empirical covariance matrix of the image can be constructed and analysed by some techniques developed in the field of computer graphics and imaging processing [10, 11]. Furthermore, some measures [12, 13] are available that can be applied to compare the (dis)similarity of multiple images in a quantitative manner.

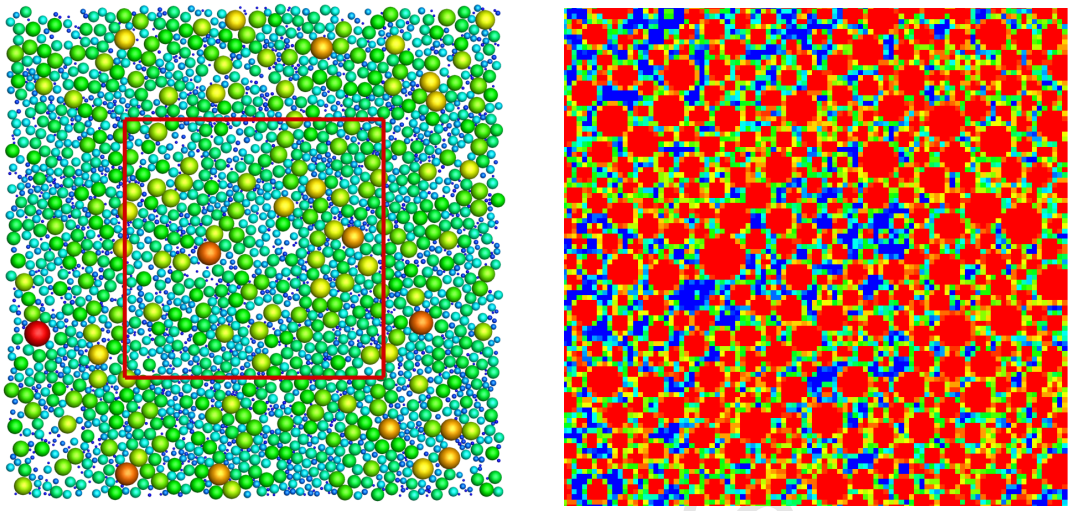
Principal component analysis (PCA) [14, 15], and its many variants under different names such as discrete Karhunen-Loeve or Hotelling transform is one of the most popular linear transform based statistical techniques, and has been extensively used in a wide variety of applications. It has proved to be a powerful tool that is often employed for data analysis in general, and dimension reduction in multi-variance analyses, and pattern recognition in signal and imaging processing in particular.

The current work attempts to develop a novel system that can characterise particle packings by using their principal components or variances obtained from PCA, and particularly it will prove that the principal variances can indeed qualify as the signature of a packing. Furthermore, a dissimilarity coefficient or a similarity index will be defined which provides a single valued metric to quantitatively compare two packings.

The paper is organised as follows. The whole numerical procedure, involving generation of a packing image and subsequently application of PCA, is fully described in Section 2. The principal variances of packings with special configuration features, such as repetition and periodicity, are exploited in Section 3, and the conclusions are numerically validated by some regular packings. In Section 4, a dissimilarity coefficient, or equivalently a similarity index, between two packings is defined based on the introduction of a principal variance function. Then comprehensive investigations for two sets of purposefully generated random packings are conducted to fully understand the relationships of their principal variances with packing features. Various issues, including effects of grid resolutions and packing density on principal variances are discussed. In addition, how to apply principal variances to assess spatial homogeneity and isotropy of packings are proposed. What relationship exists between scaled packings and their principal variances is also considered. Concluding remarks are made in Section 5.

2 Principal Component Analysis

This section is devoted to the full description of the numerical procedure that is involved in principal component analysis of a packing and how such analysis can be applied to characterise particle packings. As only the main principles of PCA are adopted in the current work, some minor modifications are made and different terminologies are used when deemed to be more appropriate. As it is well known that PCA can be derived from Singular Value Decomposition[16], this is also applied for the current development.



(a) a packing with a square analysis window (b) a 100×100 digitalised image

Figure 1: A random particle packing and a digital representation within the analysis window

2.1 Packing digitalisation and formation of packing image

First consider a circular particle assembly $\Omega_p = \bigcup_i \Omega_i$ where Ω_i is the domain of the i-th particle, and arbitrarily choose a rectangular region $\mathcal{A}$ of $L_1 \times L_2$, termed the *analysis window*. The window can be divided into a regular grid of $M \times N$ square cells with spacing $h = L_1/M = L_2/N$. For a grid cell at (i, j) with the area denoted as $\mathcal{A}_{ij}$, compute its average area ratio covered by particles, or grey-scale as

$$a_{ij} = \frac{|\Omega_p \cap \mathcal{A}_{ij}|}{|\mathcal{A}_{ij}|} \tag{1}$$

where $|\Omega|$ denotes the measure or area of a domain Ω; $|\mathcal{A}| = L_1 L_2$; and $|\mathcal{A}_{ij}| = h^2$. An empty cell with no overlapping with any particle has $a_{ij} = 0$; while a cell fully covered by a particle has $a_{ij} = 1$. A cell partially covered by particles has $a_{ij} < 1$. So in general $a_{ij} \in [0, 1]$.

The collection of all the cell average area ratios a_{ij} forms an $M \times N$ matrix $\mathbf{A}_h = \{a_{ij}\}$, which can be viewed as a digitalised grey-scale representation of the original packing Ω_p in the region $\mathcal{A}$, thus is termed as the *packing matrix* or *image*. Figure 1(b) illustrates such a digitalisation of a disc particle packing shown in Figure 1(a). Clearly, $\mathbf{A}_h$ is accurate within particles or void space of the packing, but may introduce approximation around particle boundaries. The accuracy of this representation for the original packing depends on the grid spacing h, and will be accurate in the limit case:

$$\lim_{h \to 0} \mathbf{A}_h = \Omega_p \cap \mathcal{A} \tag{2}$$

2.2 Formulations and Numerical Procedures

The mean value of the packing matrix $\mathbf{A}_h$, i.e. the packing density of the region $\mathcal{A}$, can be computed

$$\rho_{\mathcal{A}} \equiv \frac{|\Omega_p \cap \mathcal{A}|}{|\mathcal{A}|} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij} \tag{3}$$

Let $a(\mathbf{x})$ be the material distribution function with a taking the value of 1 for a point within a particle, and 0 otherwise. It is not difficult to derive that the total variance of a packing in the region $\mathcal{A}$ is related to the packing density by

$$\sigma_{\mathcal{A}} = \frac{1}{|\Omega|}\int_{\Omega}(a-\rho_{\mathcal{A}})^2 d\,\Omega = \rho_{\mathcal{A}}(1-\rho_{\mathcal{A}}) \tag{4}$$

The total variance of the matrix is defined as

$$\sigma_h = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(a_{ij}-\rho_{\mathcal{A}})^2 \le \sigma_{\mathcal{A}} \tag{5}$$

i.e. the total variance of the packing $\sigma_{\mathcal{A}}$ is the upper bound of any packing image.

Let m_j be the mean value of the j-th column of the packing matrix $\mathbf{A}_h$

$$m_j = \frac{1}{M}\sum_{i=1}^{M} a_{ij} \tag{6}$$

which also provides an alternative way to compute the density

$$\rho_{\mathcal{A}} = \frac{1}{N}\sum_{j=1}^{N} m_j \tag{7}$$

By subtraction of its mean from each column vector of $\mathbf{A}_h$, the column centralised matrix $\bar{\mathbf{A}}_N$ of $\mathbf{A}_h$ is obtained as:

$$\bar{\mathbf{A}}_N = \mathbf{A}_h - \mathbf{e}_{_M}\mathbf{m}_N \tag{8}$$

where $\mathbf{e}_{_M}$ is an $M \times 1$ column vector with all its elements being 1's; and $\mathbf{m}_N$ is the $1 \times N$ mean value vector $\mathbf{m}_N = \{m_j\}$.

Define the covariance matrix of $\bar{\mathbf{A}}_N$ as

$$\mathbf{S}_N = \frac{1}{M}\bar{\mathbf{A}}_N^{\,T}\bar{\mathbf{A}}_N \tag{9}$$

where $\mathbf{S}_N$ is a $N \times N$ square matrix. Notice in the above that M instead of $M-1$ is used. Further define the column-wise total variance as

$$\sigma_{_N}^{c} = \frac{1}{N}\mathrm{Tr}(\mathbf{S}_N) = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{S}_N)_{ii} \tag{10}$$

which may be (slightly) different from the total variance σ_h in general.

By solving the eigenvalue problem of $\mathbf{S}_N$, it yields the following matrix decomposition

$$\mathbf{S}_N\mathbf{V}_N = \mathbf{V}_N\mathbf{D}_N \tag{11}$$

with

$$\mathbf{D}_N = \mathbf{V}_N^T\mathbf{S}_N\mathbf{V}_N, \qquad \mathbf{V}_N^T\mathbf{V}_N = \mathbf{I}_N$$

where the diagonal matrix $\mathbf{D}_N = \mathrm{diag}\{d_i\}$ contains all the eigenvalues d_i in descending order, which are termed the *principal variances* (PVs); and $\mathbf{V}_N = \{\mathbf{v}_i\}$ are the orthonormal vectors, termed the *principal modes*. As $\bar{\mathbf{A}}_N$ is column centralised, $\mathbf{S}_N$ is a semi-positive

definite matrix with at least one zero principal variance. It is also well known that the sum of the PVs and the total column-wise variance is related by

$$\frac{1}{N}\sum_{i=1}^{N} d_i = \sigma_N^c \tag{12}$$

$\mathbf{S}_N$ can be recovered from the principal variances and modes as

$$\mathbf{S}_N = \mathbf{V}_N \mathbf{D}_N \mathbf{V}_N^T = \sum_{i=1}^{N-1} d_i \mathbf{v}_i \mathbf{v}_i^T \tag{13}$$

In many applications, only the first few principal variances are needed to approximate $\mathbf{S}_N$ to a reasonable degree, thereby significantly reducing the dimension of the problem concerned. This is often the main objective of PCA, but not an issue for the current problem.

Further define the projection $\mathbf{U}_N$ of $\bar{\mathbf{A}}_N$ onto the space spanned by $\mathbf{V}_N$ as

$$\mathbf{U}_N = \bar{\mathbf{A}}_N \mathbf{V}_N \tag{14}$$

Then it has

$$\mathbf{D}_N = \frac{1}{N}\mathbf{U}_N^T \mathbf{U}_N \tag{15}$$

and $\bar{\mathbf{A}}_N$ and $\mathbf{A}_h$ can be recovered by

$$\bar{\mathbf{A}}_N = \mathbf{U}_N \mathbf{V}_N^T; \quad \mathbf{A}_h = \bar{\mathbf{A}}_N + \mathbf{e}_M \mathbf{m}_N \tag{16}$$

Similarly to $\mathbf{S}_N$, $\bar{\mathbf{A}}_N$ or $\mathbf{A}_h$ can be optimally approximated by the leading principal variances.

The column-wise total variance $\sigma_{_N}^c$, the mean value vector $\mathbf{m}_N$, the principal variance matrix $\mathbf{D}_N$ and the corresponding modes $\mathbf{V}_N$ form a unique set $\mathcal{S}_\mathcal{N}$, termed the column-wise *characteristic set*, that fully determines the packing in the vertical direction

$$\mathcal{C}_\mathcal{N} = \{\sigma_{_N}^c, \mathbf{m}_N, \mathbf{D}_N, \mathbf{V}_N\} \tag{17}$$

As the PVs and the column-wise total variance is related by (12), and the total variance (and also the column-wise variance) is related to the density (see (4)), the PVs play a dominant role to characterise a packing image and therefore can be viewed as the (column-wise) *signature* of the packing.

The above are developed based on the column-wise consideration of the matrix $\mathbf{A}_h$. Equally, another set can be derived following a row-wise consideration. By labelling the mean value (column) vector for the rows of $\mathbf{A}_h$ as $\mathbf{m}_M$, and the row-centralised covariance matrix $\bar{\mathbf{A}}_M$ as

$$\bar{A}_M = \mathbf{A}_h - \mathbf{m}_M \mathbf{e}_N^T \tag{18}$$

where $\mathbf{e}_N$ is a $N \times 1$ vector of all 1's, and the corresponding covariance matrix of $M \times M$ as

$$\mathbf{S}_M = \frac{1}{N}\bar{A}_M \bar{A}_M^T \tag{19}$$

the row-wise total variance can be computed as

$$\sigma_{_M}^r = \frac{1}{M}\,\mathrm{Tr}(\mathbf{S}_M) = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{S}_M)_{ii} \tag{20}$$

The M principal variances and modes can be found as

$$\mathbf{S}_M\mathbf{V}_M = \mathbf{V}_M\mathbf{D}_M, \qquad \mathbf{V}_M^T\mathbf{V}_M = \mathbf{I}_M \tag{21}$$

Consequently, the row-wise characteristic set of $\mathcal{A}$ becomes

$$\mathcal{C}_{\mathcal{M}} = \{\sigma_{_M}^r, \mathbf{m}_M, \mathbf{D}_M, \mathbf{V}_M\} \tag{22}$$

and $\mathbf{D}_M$ are the (row-wise) signature of the packing image. As $\mathbf{m}_N$ and $\mathbf{m}_M$ are generally different even when $N = M$, so $\mathbf{S}_N \neq \mathbf{S}_M$, and thus $D_N \neq D_M$. The difference will depend on N and M, and the packing configuration.

For simplicity, only the column-wise related PCA is discussed below.

2.3 Principal Variance Function

To facilitate the comparison between different sets of principal variances, particularly when they are obtained from different resolutions N, the ordinal number i of a principal variance d_i is mapped from $1 to N$ to a non-dimensional "position" variable $x \in [0, 1]$ by

$$x(i) = \frac{1}{N}\Big(i - \frac{1}{2}\Big) \tag{23}$$

Then a continuous function $\mathsf{d}(x), x \in [0, 1]$, termed the (column-wise) *principal variance function*, can be constructed to interpolate the discrete variances d_i using piecewise linear or higher order interpolation functions such that

$$\mathsf{d}(x_i) = d_i, \quad x_i = x(i), \quad i = 1, \cdots, N \tag{24}$$

The conversion from the ordinal number i to the position x play an essential role to quantitatively compare two images, with different or same resolutions, as demonstrated in Section 4.2.

3 Principal Variances and Modes of Some Special Packings

This section exploits what additional properties of the principal variances and modes that some special packing configurations/images may have. Illustrative examples are provided to validate the theoretical development in this section.

3.1 Permuted, Mirrored Packings

For a given packing matrix $\mathbf{A}_N$, assume that another matrix $\mathbf{A}'_N$ is obtained by permutating columns. Let the permutation be represented by a permutation matrix $\mathbf{P}_N$ of order $N \times N$. As $\mathbf{P}_N^T\mathbf{P}_N = \mathbf{I}_N$ it is not difficult to prove that $\mathbf{A}'_N$ will have the same principal variances but the principal modes are the permutation of the original ones

$$\mathbf{D}'_N = \mathbf{D}_N; \qquad \mathbf{V}'_N = \mathbf{P}_N\mathbf{V}_N \tag{25}$$

A similar conclusion can be drawn if a matrix is obtained from row permutation.

Notice that permutation can mirror or reflect a packing and therefore it can be concluded that the mirror packing has the same principal variances as the original packing.

3.2 Repetitive, Periodic and Symmetric Packings

If the packing has a repetitive or periodic structure where a basic packing unit repeats multiple times along the horizontal direction, PCA can be applied to this basic packing and the principal variances and modes of the whole structure can be readily obtained, as derived below.

Consider the simplest 2-repetition case where the basic structure is repeated twice. Use the same grid spacing for both the basic structure and the whole packing, and let $\mathbf{A}_N$ be the packing matrix of the basic packing. Then the whole packing matrix, $\mathbf{A}_{2N}$, consists of two identical sub-matrices $\mathbf{A}_N$

$$\mathbf{A}_{2N} = [\mathbf{A}_N, \mathbf{A}_N] \tag{26}$$

Let $\mathbf{m}_N$, $\bar{\mathbf{A}}_N$ and $\mathbf{S}_N$ be respectively the column mean vector, the mean centralised matrix and the covariance matrix of $\mathbf{m}_N$ as defined before. Then for $\mathbf{A}_{2N}$, the column mean vector is

$$\mathbf{m}_{_{2N}} = \{\mathbf{m}_N, \mathbf{m}_N\} \tag{27}$$

which leads to the mean centralised matrix

$$\bar{\mathbf{A}}_{2N} = \mathbf{A}_{2N} - \mathbf{e}_M \mathbf{m}_{_{2N}} \tag{28}$$

The covariance matrix takes the form

$$\mathbf{S}_{2N} = \frac{1}{M}\bar{\mathbf{A}}_{2N}^T \bar{\mathbf{A}}_{2N} = \frac{1}{M}\begin{bmatrix} \mathbf{S}_N & \mathbf{S}_N \\ \mathbf{S}_N & \mathbf{S}_N \end{bmatrix} \tag{29}$$

Again let $\mathbf{V}_N$ and $\mathbf{D}_N$ be respectively the principal variances and modes of $\mathbf{S}_N$. Now construct a matrix of order $2N \times N$

$$\mathbf{V}_{2N} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{V}_N \\ \mathbf{V}_N \end{bmatrix} \tag{30}$$

Then

$$\mathbf{S}_{2N}\mathbf{V}_{2N} = \frac{1}{\sqrt{2}}\begin{bmatrix} 2\mathbf{S}_N\mathbf{V}_N \\ 2\mathbf{S}_N\mathbf{V}_N \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 2\mathbf{V}_N\mathbf{D}_N \\ 2\mathbf{V}_N\mathbf{D}_N \end{bmatrix} = \mathbf{V}_{2N}(2\mathbf{D}_N) \tag{31}$$

i.e. $\mathbf{V}_{2N}$ are the principal modes of $\mathbf{S}_{2N}$ and the corresponding principal variances are 2 times $\mathbf{D}_N$. Note, however, that $(\mathbf{V}_{2N}, 2\mathbf{D}_N)$ only account for N or half principal pairs of $\mathbf{S}_{2N}$.

Construct another matrix of order $2N \times N$

$$\mathbf{V}'_{2N} = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{V}_N \\ -\mathbf{V}_N \end{bmatrix} \tag{32}$$

Direct calculation leads to

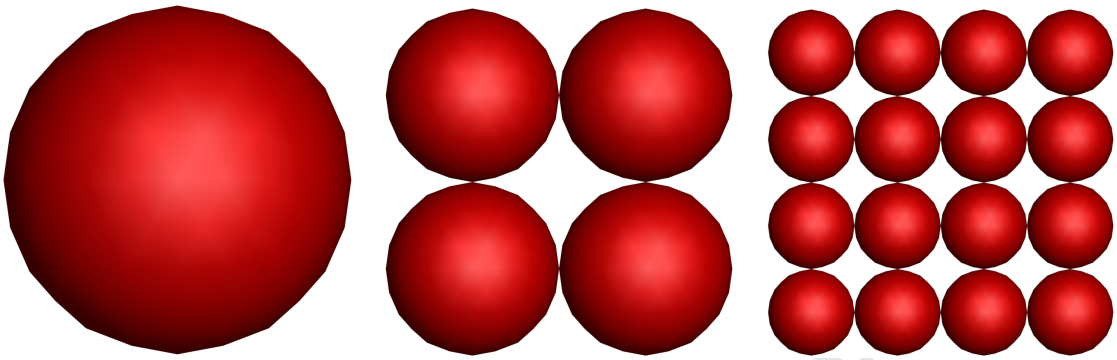
$$\mathbf{S}_{2N}\mathbf{V}'_{2N} = 0 \tag{33}$$

which means that $\mathbf{V}'_{2N}$ are also the principal modes, but the corresponding principal variances are all zero. Combination of (31) and (33) concludes that the half of the principal variances of the whole packing are 2 times those of the basic structure, but the rest are zero.

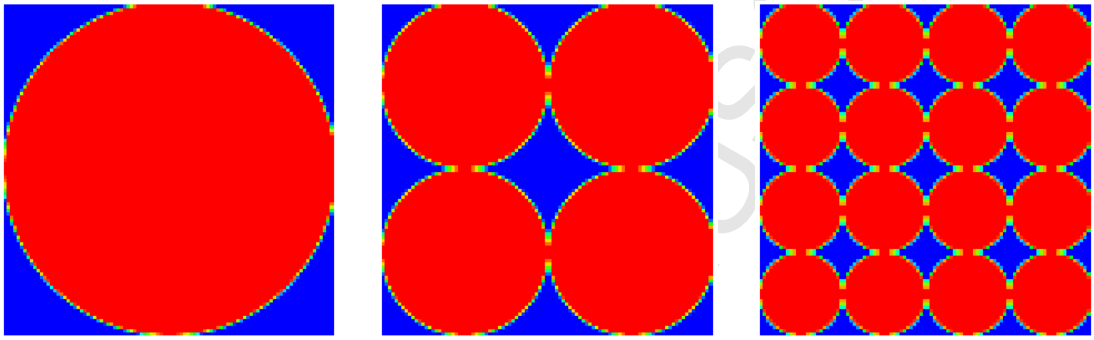
In general, for a m-repetition packing, $1/m$ of the principal variances will be m times those of the basic structure, and the remaining principal variances are zero.

If the repetition occurs in the vertical direction, the whole packing matrix admits the following partition for a 2-repetition packing

$$\mathbf{A}_{2N} = \begin{bmatrix} \mathbf{A}_N \\ \mathbf{A}_N \end{bmatrix} \tag{34}$$



(a) Packing R1 (b) Packing R4 (c) Packing R16



(d) Image R1(N) (e) Image R4(N) (f) Image R16(N)

Figure 2: Three regular packings (a-c) and their digitalised images (d-f) (N=100)

Utilising the fact that $\mathbf{m}_{_{2N}} = \mathbf{m}_N$, one can prove that

$$\mathbf{S}_{2N} = \mathbf{S}_N \tag{35}$$

hence the principal variances and modes remain the same. This repetitive feature will be exploited, nevertheless, when applying the row-wise PCA to the packing matrix.

If a packing is symmetric about a vertical line, one part can be obtained by mirroring the other part against the symmetric line through column permutation, and so both parts will have the same principal variances. Thus for the whole packing, the non-zero principal variances will be twice those from each part.

It is also obvious that an exact scaling of both the packing and the image grid together will lead to the principal variances and modes being unchanged.

3.3 Numerical Validation

Consider three regular packings within a unit square region as show in Figure 2(a-c), where there are respectively 1, 4 and 16 equal-sized particles in the packings that are labelled as R1, R4 and R16 respectively. The three packings have nested or exactly scaled configurations, and have the same packing density $\rho = \pi/4 = 0.7854$. The region is then divided into a $N \times N$ square grid with spacing $h = 1/N$ for each packing, and three resolutions $N = 25, 50$ and 100 are considered. The resulting packing images are labelled as R$i(N)(i = 1, 4, 16; N = 1, 50, 100)$. The images with $N = 100$ are depicted in Figure 2(d-f). The total column-wise variances and first or maximum principal variances of the three packings for different grid resolutions are respectively given in Tables 1 and 2.

Table 1: Total variances of three regular packings with different grid resolutions

Grid		Packing		
N	h	R1	R4	R16
25	0.04	**.1501**	.1362	.1112
50	0.02	.1589	**.1501**	.1362
100	0.01	.1636	.1589	**.1501**

Table 2: The first principal variances of three regular packings with different grid resolutions

Grid		Packing		
N	h	R1	R4	R16
25	0.04	**1.7665**	1.7050	1.5133
50	0.02	3.5778	**3.5329**	3.4099
100	0.01	7.1816	7.1556	**7.0657**

Table 1 shows that the total column-wise variance exhibits a clear pattern: the values along the main diagonal and two off-diagonal lines are the same. This can be easily explained due to the nested relationship between the three packings.

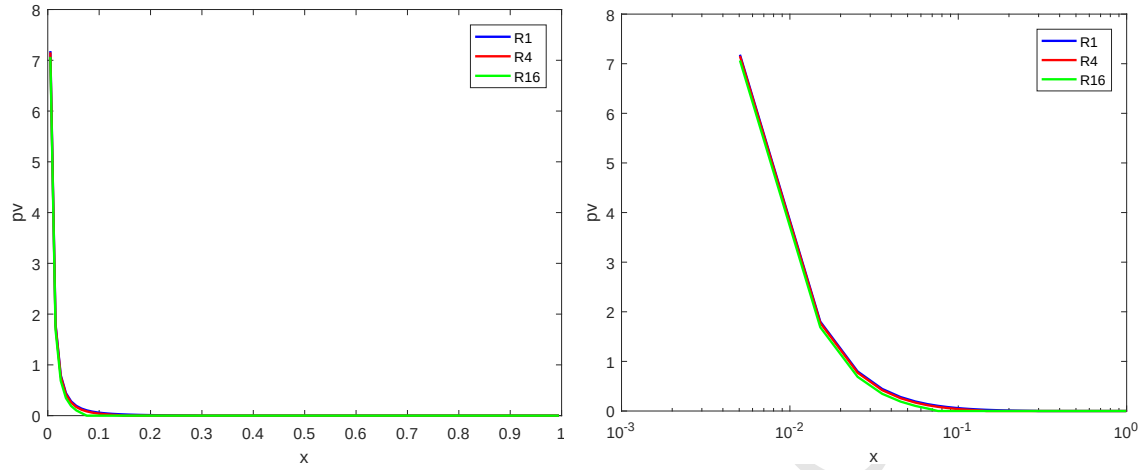
For the first principal variances listed in Table 2, a pattern also emerges along the main diagonal and two off-diagonal lines: the ratio between two consecutive values on a line is exactly 2. This validates the theoretical derivation in this section for repetitive packings. For instance, Image R4(50) is a 2-repetition of Image R1(25) (with an additional 2-repetition in the vertical direction but having no contribution to the PVs), leading to doubled PVs for the former; While Image R16(100) is also a 2-repetition packing of Image R4(50), thus having PVs which are double of those of R4.

It is also observed, but without offering a rigorous proof, that for a quarter of disc which has a diagonal symmetry and is the basic building block of a disc, the number of non-zero principal variances is equal to $\lfloor N(2-\sqrt{2})/2 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function.

The PVs of the three packings with $N = 100$ are plotted in Figure 3(a), and also in Figure 3(b) with a logarithmic scale of x to achieve a zoomed view for the leading PVs. Both Table 2 and Figure 3 show that very similar principal variances are obtained from the three packings with the same grid resolution N. In fact, the dissimilarity coefficients of R4 and R16 against R1 are 0.0081 and 0.0275 respectively. The increased dissimilarity for R16 is due to the reduced relative resolution in terms of ratio r/h.

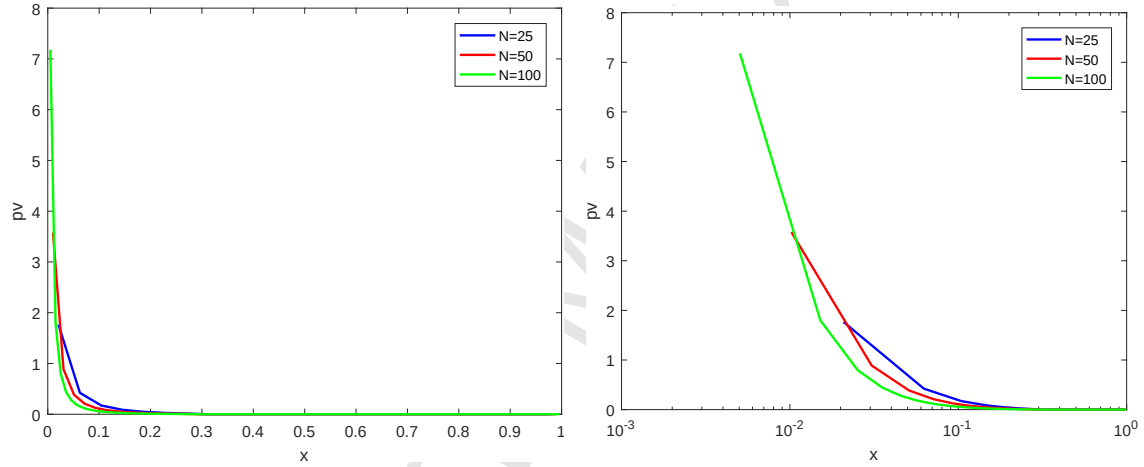
Figure 4(a) displays all the PVs of Packing R1 for the three grid resolutions, with a zoomed view for the main PVs shown in Figure 4(b). Both Table 2 and Figure 4 demonstrate the dependence of the VPs on the grid resolution N. This issue will be further investigated in Section 4.6.

The first three principal modes of R1(100), and the first modes of R1, R4 and R16 with N=100 are plotted in Figure 5(a) and (b) respectively. The symmetric nature of all three modes of R1(100) is evident as expected, while the first modes of R4(100) and R16(100) are compressed and repetitive versions of the base case R1(100).



(a) Normal view (b) Zoomed view

Figure 3: Principal variances of R1, R4 and R16 at N=100



(a) Normal view (b) Zoomed view

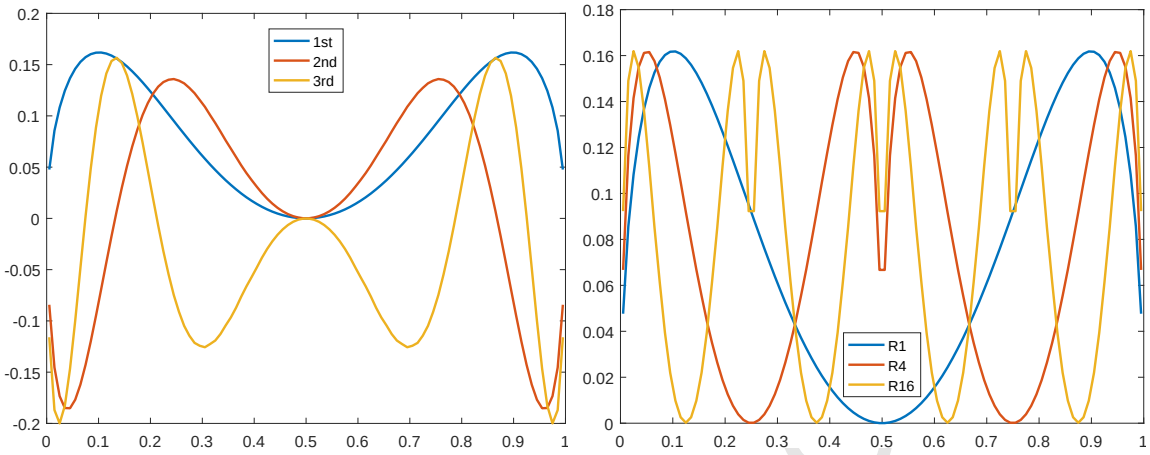
Figure 4: Principal variances of R1 at N=25, 50, 100

4 Packing Characterisation using Principal Variances

As the main signature of a packing, principal variances will be comprehensively exploited in this section to understand how they can be applied to characterise a packing or to quantitatively compare the similarity or difference between different packings. In addition to the three regular packings that have been used to validate some of our theoretical developments for repetitive packings, additional two sets of random packings will be utilised to provide new evidence as the basis for further exploitation. The detail of these packings are given in the next subsection, while a number of characterisation issues will be developed and discussed in other subsections.

4.1 Numerical Examples: Two Sets of Random Packings

Two sets of random but periodic particle packings are generated within the domain $[-0.1, 1.1]\times[-0.1, 1.1]$ with the periodic condition applied to both directions. The first set, or U-set, has



(a) First three modes of R1(100)

(b) First models of R1(100), R4(100) and R16(100)

Figure 5: Some leading principal modes of R1, R4 and R16 at N=100

Table 3: Some properties of two sets of random packings

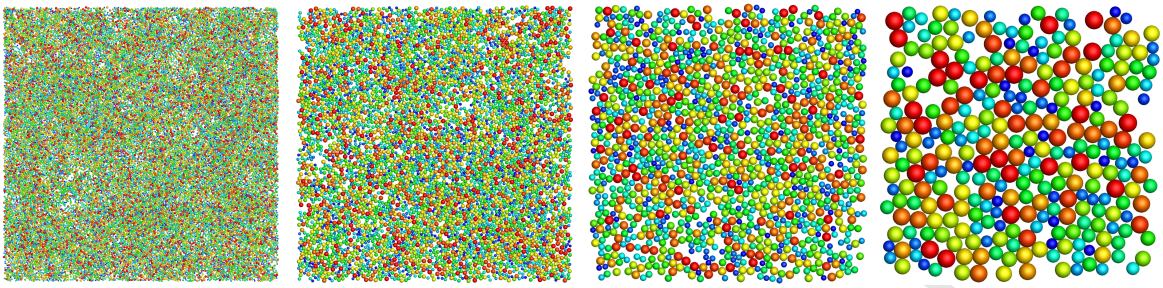
Set	Uniform Distribution (U-set)				Gaussian Distribution (G-set)			
Group Name	U1	U2	U4	U8	G1	G2	G4	G8
Mean Density	.7074	.7132	.7143	.6933	.7141	.7156	.6999	.7067
Particle No.	19617	4894	1216	291	12950	3214	791	210
r_{min}	0.003	0.006	0.012	0.024	0.001	0.002	0.004	0.008
r_{max}	0.005	0.010	0.020	0.040	0.007	0.014	0.028	0.056
r	0.004	0.008	0.016	0.032	0.004	0.008	0.016	0.032

particle sizes uniformly distributed within a range; the second set, or G-set, has particle sizes obeying Gaussian distributions with limited minimum and maximum sizes. Each set has four groups each having the particle size range doubled from the previous group, while within each group 10 random packing samples with the same size distribution are generated. The packings and their images at $N = 100$ of the U set are displayed in Figure 6; while the packings of the G set are displayed in Figure 7.

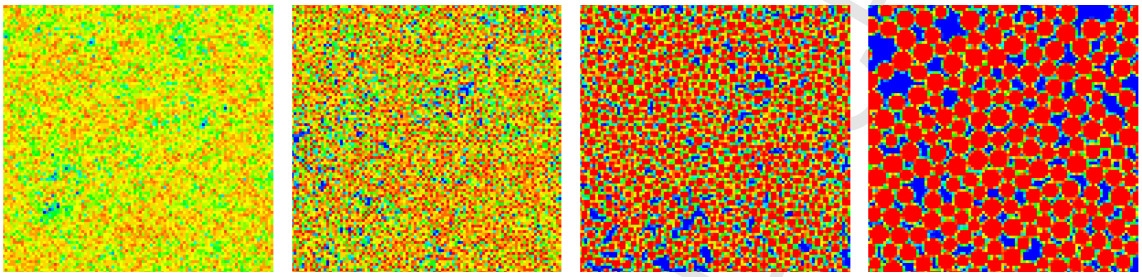
Table 3 lists all the details about the packings up to the group level, including the minimum, maximum and average particle sizes, r_{min}, r_{max}, r, and the average number of particles in each group. In the U-set, the size distributions of groups U2, U4, U8 are respectively 2, 4, and 8 times of the base group U1. This also applies to the G-set. The average particle size r is the same for the corresponding groups between the two sets: Ui and Gi (i=1,2,4,8) have the same r_i. For the G-set, the standard deviation of the Gaussian distribution for a group is taken to be $(r_{max} - r_{min})/2$.

The default analysis window is chosen to be the unit square region $[0, 1] \times [0, 1]$, so $M = N$. Although the packing density is set to be $\rho = 0.7$, the packing density within the analysis window is slightly different from 0.7, as the window is smaller than the packing region. The mean density for each group is listed in Table 3. Because all the mean densities are very close to 0.7, no significant effect on principal variances is expected.

A column-wise PCA is applied to each sample with required grid resolutions N. For each group, the principal variances at a given resolution N are taken to be the average of all the

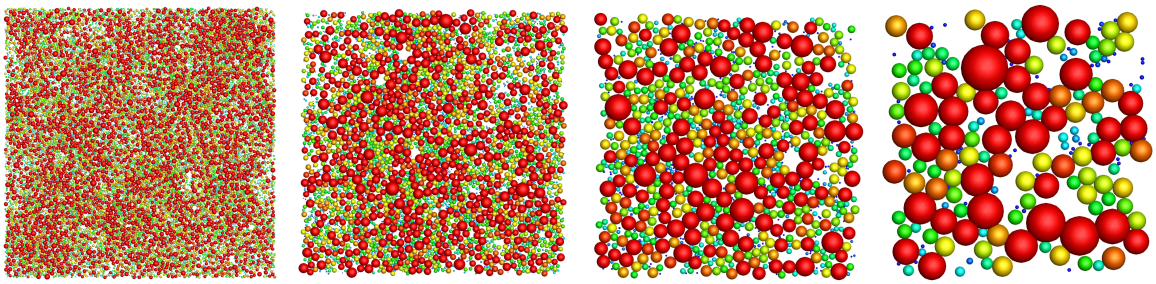


(a) Packing U1 (b) Packing U2 (c) Packing U4 (d) Packing U8

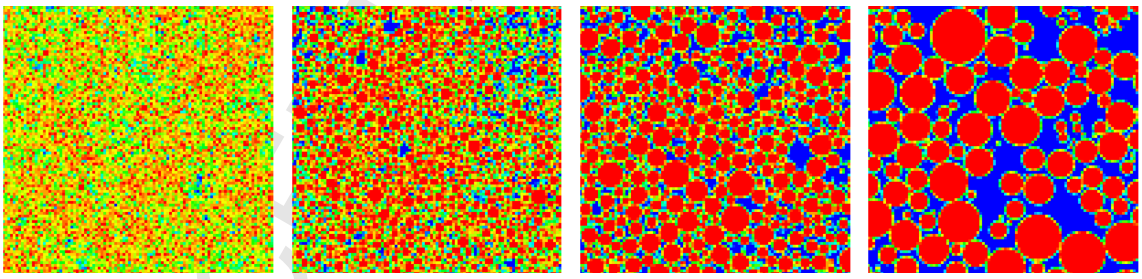


(e) Image U1(N) (f) Image U2(N) (g) Image U4(N) (h) Image U8(N)

Figure 6: U-set: Four uniform packing groups in region [-0.1, 1.1] × [-0.1, 1.1] (a-d); and their digital images (with N=100) within region [0, 1] × [0, 1] (e-h)



(a) Packing G1 (b) Packing G2 (c) Packing G4 (d) Packing G8

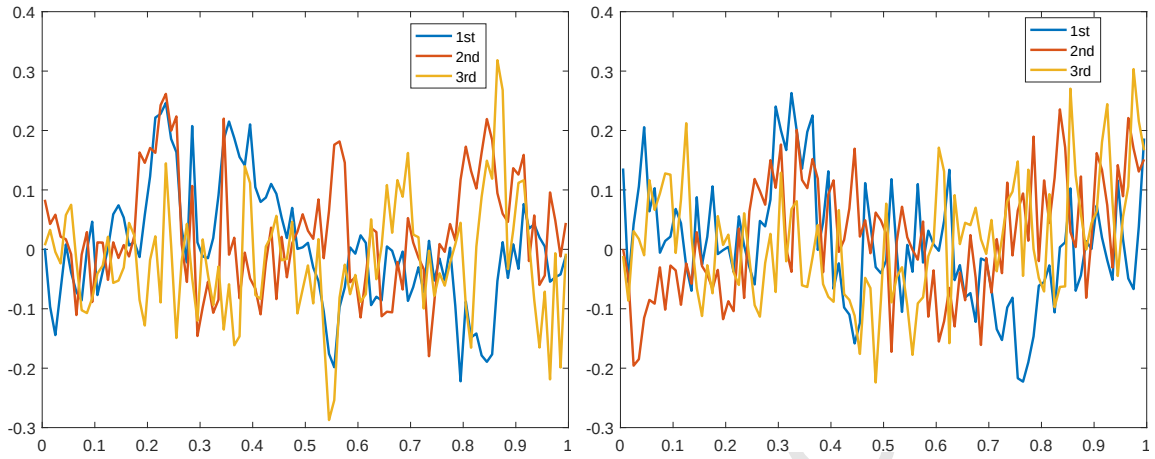


(e) Image G1 (f) Image G2 (g) Image G4 (h) Image G8

Figure 7: G-set: Four Gaussian packing groups in region [-0.1, 1.1] × [-0.1, 1.1] (a-d); and their digital images (with N=100) within region [0, 1] × [0, 1] (e-h)

samples in the group.

The first three principal modes of U1(100) and G1(100) are also plotted in Figure 8(a) and (b) respectively for illustrative purpose. The randomness/irregularity of the modes are the dominant feature compared to the regular packing case (see Figure 5). No further discussion



(a) U1(100) (b) G1(100)

Figure 8: First three principal modes of U1 and G1 at $N = 100$

regarding the principal modes will be conducted.

4.2 Packing Image Similarity

Consider two packing images with their principal variance functions $\mathsf{d}_1(x)$ and $\mathsf{d}_2(x)$ obtained, define a so-called *dissimilarity coefficient* (DC) between these two images as

$$\mathcal{D}_c = \Big[\frac{1}{\Sigma_1 + \Sigma_2}\int_0^1 [\mathsf{d}_1(x) - \mathsf{d}_2(x)]^2 d\,x\Big]^{1/2} \in [0,1] \tag{36}$$

where N_1 and N_2 are respectively the (column) resolutions of the two images; and Σ_1 and Σ_2 are defined as

$$\Sigma_i = \int_0^1 \mathsf{d}_i^2(x)\,dx \quad (i = 1, 2)$$

Consequently, the degree of similarity of these two packing images can be quantified by the *similarity index* $\in [0, 100]$ defined as

$$\mathcal{S}_\mathcal{I} = (1 - \mathcal{D}_c) \times 100 \tag{37}$$

4.3 Dis/Similarity between Packing Samples and Groups

For each packing group, the PVs of all 10 samples are computed and their averages are taken to be the PVs of the group. For illustrative purpose, the principal variance functions of the 10 samples for groups U1 and G1 at three grid resolutions N=100, 400, and 1600 are displayed in Figure 9. Clearly the PV functions of the 10 samples at each set are located within a narrow band around the group mean value where the maximum difference appears at the leading variances but the difference is much reduced for smaller PVs. This indicates that these samples randomly generated from the same distribution indeed have very similar statistical features.

To quantify the difference, the dissimilarity coefficient of 10 samples in each group are calculated based on the formula (36) against their group average for three resolutions: N=100, 400 and 1600. The average dissimilarity coefficients of 10 samples in each group for the

Table 4: Average dissimilarity coefficient of each group in two sets of random packings

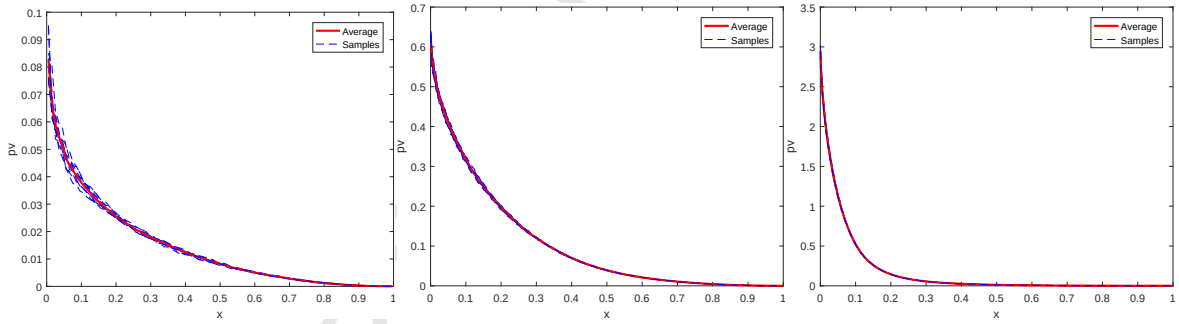
N	U1	U2	U4	U8	G1	G2	G4	G8
1600	0.0094	0.0196	0.0374	0.0621	0.0160	0.0239	0.0401	0.1301
800	0.0101	0.0199	0.0366	0.0599	0.0161	0.0231	0.0371	0.1193
400	0.0126	0.0215	0.0372	0.0602	0.0178	0.0238	0.0371	0.1200
200	0.0201	0.0261	0.0397	0.0622	0.0223	0.0266	0.0387	0.1235
100	0.0378	0.0379	0.0465	0.0662	0.0286	0.0335	0.0435	0.1303

Table 5: Dissimilarity coefficients between groups in two sets of random packings

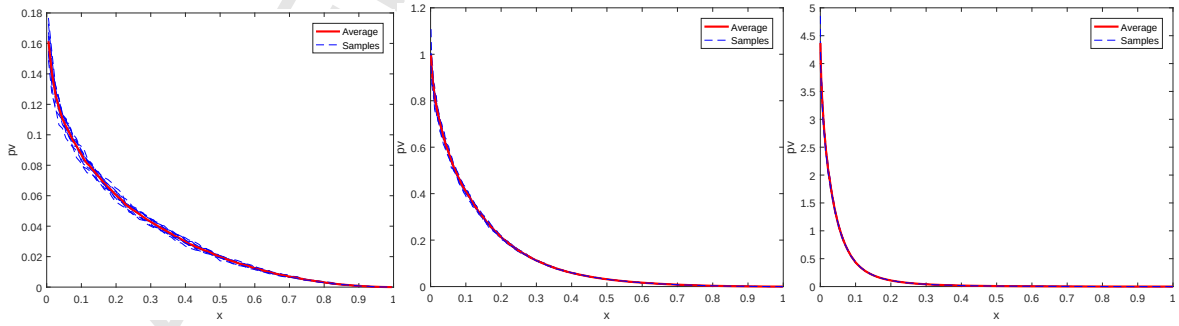
N	U1-G1	U2-G2	U4-G4	U8-G8
1600	0.1610	0.1616	0.1835	0.1471
800	0.1792	0.1672	0.1892	0.1421
400	0.2329	0.1851	0.1970	0.1439
200	0.3459	0.2359	0.2153	0.1482
100	0.5021	0.3427	0.2634	0.1577

three resolutions are provided in Table 4. The DCs between the corresponding groups, Ui-Gi (i=1,2,4,8), of the two distribution sets are also computed and given in Table 5, where their group averaged PVs are used for the calculation.

To visualise the DCs, the sample DCs of the corresponding groups of the two sets for each



(a) U1(100) (b) U1(400) (c) U1(1600)



(a) G1(100) (b) G1(400) (b) G1(1600)

Figure 9: Average principal variance functions of groups U1 and G1 for three resolutions: N=100, 400, 1600

Table 6: Dissimilarity coefficients of Submatrices of 4 packing images at N=1600

Comparison	Within Group			With the whole image		
Submatrix size	800	400	200	800	400	200
U1(1600)	0.012	0.037	0.050	0.010	0.009	0.014
G1(1600)	0.015	0.029	0.052	0.009	0.022	0.022
U4(1600)	0.054	0.095	0.167	0.038	0.069	0.109
G4(1600)	0.140	0.194	0.221	0.139	0.221	0.318

grid resolution, together with the group DC as a reference value, are all plotted in Figure 10. Clearly, dissimilarity reduces when N increases, and in all cases the coefficient seems to converge to a value of around 0.16, or a similarity index = 84, regardless of groups.

It may be concluded: 1) 10 samples within each group have very small differences, and the similarity indices of most groups can reach 99; but the difference increases when the number of particles in the packing decreases. 2) The dissimilarity between the corresponding groups of the two distributions are substantially larger than that of the samples within each group, indicating that the PVs can indeed be utilised to effectively classify packings.

4.4 Packing Uniformity and Isotropy

Principal variances can be employed to exploit the properties of a packing in more detail, such as its spatial uniformity or homogeneity and isotropy.

4.4.1 Uniformity

For one packing, its uniformity in space can be checked by applying a moving analysis window and obtaining the PVs at some selected locations, and then compute the DCs of these PVs by (36). A uniformly small DC indicates that the packing may be statistically homogeneous.

Alternatively, if the image $\mathbf{A}_h$ of a packing within a large analysis window is given, by selecting smaller analysis windows with the same grid spacing h, or by simply selecting sub-matrices/sub-images from $\mathbf{A}_h$, the spatial uniformity of the packing can also be investigated at a smaller scale. To illustrate this approach, U1(1600) and G1(1600) are used as examples. A number of equally-sized sub-matrix blocks are randomly extracted from the two packing images and PCA is applied to every sub-matrix to obtain its PVs. Three different sizes of sub-matrices $800 \times 800, 400 \times 400$, and 200×200, equivalent to analysis windows of $[0.5 \times 0.5], [0.25 \times 0.25]$, and $[0.2 \times 0.2]$, are considered and their numbers are 10, 20, 40 respectively. For the same sized sub-matrices, their dissimilarity coefficients against the average PVs are evaluated, and the averaged dissimilarity coefficients are also obtained.

Figure 11 depicts the average PVs of the three different sized sub-matrices against those of 4 images U1(1600), U4(1600), G1(1600), and G4(1600) respectively. The logarithmic scale for x is used to enlarge the differences around the leading PVs, otherwise almost identical curves are observed at a normal scale. Both dissimilarity coefficients within each sub-matrix group and against the associated whole image are presented in Table 6. Furthermore, the individual dissimilarity coefficients of the sub-matrices against the averaged value of each group are displayed in Figure 12.

Table 6 shows that for U1(1600) and G1(1600), the differences between the submatrices and

their associated whole images are generally very small, and no larger than 5% even with the smallest block size of 200; while for U4(1600) and G4(1600) which have much fewer particles, the differences are higher and reach about $20 \sim 30\%$ for the block size of 200, indicating that it is more difficult to generate statistically equivalent particle packings with a small number of particles.

It can be seen from Figure 12 that the distribution of the dissimilarity coefficient over the entire selection samples for each case is not constant. For U1 and G1 cases, a small level of spatial in-homogeneity exists in the four packings concerned, but for U4 and G4, a large degree of in-homogeneity is observed, as expected.

It is worth in highlighting that the above similarity comparisons between images with different resolutions can not be properly done without using the principal variance function defined earlier.

4.4.2 Isotropy

By comparing a column-wise and a row-wise PCA to a packing image can reveal if the packing within the analysis window is (an)isotropic in these two directions. A more detailed isotropy check of the packing may be conducted by rotating the analysis window from 0° to 180°, and applying PCA to each angle, as illustrated by Figure 13(a). Then the DCs of the PVs of these rotated packings against their averaged PVs will reveal if the given packing is generally isotropic or not in a broader sense. Take G1 as an example, and choose an analysis window of $[0.5 \times 0.5]$, the DCs of the rotated packings with two resolutions N=100 and 400 are displayed in Figure 13(b). It indicates that weak isotropy exists for the packing.

In a similar fashion, strong anisotropy of a packing can also be identified by PVs. Figure 14(a) shows a packing generated from U4 but some particles being removed from a central strip to make it anisotropic. By applying both a column-wise and a row-wise PCA to the packing image at $N = 100$, the two sets of PVs in the two perpendicular directions are attained and displayed in Figure 14(b), demonstrating that the strong anisotropy is indeed captured by significant differences in the leading PVs.

4.5 Packing Density Effects

It is not obvious how packing density affects PVs and more specifically, if two different densities could lead to two very similar PV sets. Such an effect is briefly considered by examples. For each group of U4 and G4, another group of 10 samples is also generated with a smaller packing density of $\rho = 0.65$. The dissimilarity coefficients between the groups with different densities for five resolutions $N = 1600, 800, 400, 200, 100$ are computed and given in Table 7, showing around 10% difference on average. It also appears that reducing the packing density tends to increase the (total and principal) variances. This may be explained by the density-variance relationship (4) from which the maximum total variance is reached when $\rho_{\mathcal{A}} = 0.5$, i.e. achieve a maximum spatial material variance in the packing. As in the current case, reducing the density from 0.7 to 0.6 should increase the variances in general, leading to increased dissimilarity coefficients. This provides further evidence that PVs can indeed be taken as the signature of a packing which can differentiate packings with different densities.

Table 7: Dissimilarity coefficients of two groups with two packing densities

Group	N=1600	N=800	N=400	N=200	N=100
U4	0.0659	0.0670	0.0684	0.0714	0.0783
G4	0.0986	0.0976	0.0983	0.1019	0.1106

4.6 Principal Variances via Grid Resolutions

Many results presented earlier clearly demonstrate a strong dependence of PVs on the grid resolution N or spacing h. Generally speaking, both the total variance and the PVs increase with the increase of N or decrease of h, but their limits exist. Further investigations reveal that the total variance and the leading PVs converge almost linearly towards the limits as $1/N$ or h tends to zero. A similar convergent behaviour has already been observed for DCs within samples of the same group, as shown in Table 4, and between different groups of the two sets as shown in Table 5. Note, however, that these DCs are computed for two packings with the same sized analysis window and grid resolution. While in Section 4.4.1, the comparison is conducted for different sized analysis windows and resolutions but with the same spacing h, showing a convergence of PVs as the size of the selected analysis window increases.

4.7 Principal Variances for Scaled Random Packings

The packing groups within each set considered are deliberately generated to have their size distributions scaled from the base group by a factor of $2^m (m = 1, 2, 3)$ so that the relationship between PVs (and DCs) and the scaling factor can be easily established.

The principal variance functions of U2(800), U4(400) and U8(200) against U1(1600) are plotted in Figure 15(a) and the corresponding DCs against U1(1600) are also shown; while the PV functions of G2(800), G4(400) and G8(200) against G1(1600), together with their DCs, are displayed in Figure 15(b). Notice that the ratio between the average radius and the spacing r/h is kept to be 0.15 for all the cases. The figure clearly indicates that the PVs of U1(1600), U2(800), U4(400) and U8(200) are very similar, with the maximum difference being about 10%, or the minimum similar index = 90. The same applies to G1(1600), G2(800), G4(400) and G8(200). In other words, the PVs of a set of scaled packings may be very close if different resolutions are chosen such that $r * N$ or r/h is a constant.

It is, however, not the case for the regular packings discussed in Section 3.3, where similar PVs are obtained for the same grid resolution N or spacing h. This apparent contradiction can be explained without offering a rigorous proof as follows.

Packings R4 and R16 are scaled and repetitive versions of R1, making them 100% correlated. The relationship in their PVs has been fully established in Section 3.2. While for random packing U1(1600) or G1(1600), when it is split into 4 800×800 blocks, these sub-matrices are (almost) statistically independent but have very similar PVs, making their PV functions very close to that of U1(1600) or G1(1600), as demonstrated earlier in Figure 12. On the other hand, U2(800) or G2(800) is statistically equivalent to a 2-time up-scaled 800×800 block of U1(1600) or G1(1600), thereby having a similar PV function to the block (refer to Section 3.2 for the reason), and therefore also close to U1(1600) or G1(1600).

5 Concluding Remarks

The present work has proposed a Principal Component Analysis based novel methodology to characterise particle packings. It involves first the digitalisation of a packing into a grey-scale image; and then the application of PCA to obtain the principal variances of the image. Another important development is the definition of a dissimilarity coefficient or equivalently a similarity index, by which the degree of (dis)similarity of two packing images can be quantitatively compared and evaluated. This has been made possible by the definition of the principal variance function that maps the ordinal numbers of individual principal variances into a non-dimensional unit interval [0,1]. From comprehensive investigations on the effectiveness of characterising some purposefully generated regular and random packings, it can be concluded that the principal variances are the signature of a packing image.

Note that the packing characterisation through the illustrative examples has been mainly focused on the quantitative comparison between different packings using their principal variances or the dissimilarity coefficient. More work is worth being pursued to gain a full understanding how PVs and other values in the characteristic sets, $\mathcal{C}_{\mathcal{M}}$ and $\mathcal{C}_{\mathcal{M}}$, are directly related to the packing features of a packing, such as packing density or porosity, anisotropy, heterogeneity, and particle size distribution *etc.* Furthermore, the methodology developed can be extended to 3D cases [17] and non-spherical particle packings, and can also be applicable to some other problems in particle systems, which will be reported later.

References

[1] P. A. Cundall and O. D. L. Strack. A discrete numerical model for granular assemblies. *Geotechnique*, 29(1):47-65, 1979.

[2] Y. T. Feng, K. Han, and D. R. J. Owen. Filling domains with disks: An advancing front approach. *International Journal For Numerical Methods in Engineering.* 56: 699-713, 2003.

[3] R. Löhner, and E. Oñate. A general advancing front technique for filling space with arbitrary objects. *International Journal For Numerical Methods in Engineering.* 61(12): 1977-1991, 2004.

[4] K. Han, Y. T. Feng, and D. R. J. Owen. Sphere packing with a geometric based compression algorithm. *Powder Technology*, 155(1):33-41, 2005.

[5] J. P. Latham, A. Munjiza and Y. Lu. On the prediction of void porosity and packing of rock particulates. *Powder Technology.* 125(1): 10-27, 2002.

[6] Y. Sheng, C. J. Lawrence, B. J. Briscoe and C. Thornton. Numerical studies of uniaxial powder compaction process by 3D DEM. *Engineering Computations.* 21(2/3/4): 304-317, 2004.

[7] R. Guises, J. Xiang, J. P. Latham and A. Munjiza. Granular packing: numerical simulation and the characterisation of the effect of particle shape. *Granular Matter.* 11(5): 281-292, 2009.

[8] Jean-Franois Jerier, Didier Imbault, Frederic-Victor Donze and Pierre Doremus. A geometric algorithm based on tetrahedral meshes to generate a dense polydisperse sphere packing *Granular Matter*, 11(1) 4352, 2009.

[9] X. Sun, Y. Dong, P. Hao, L. Shi, F. Li, and Y. T. Feng. Three-dimensional numerical simulation of quasi-static pebble flow. *Advanced Powder Technology* 28(2): 499-505, 2017

[10] H. H. Barret. Foundations of Image Science. 3rd Edition, John Wiley & Sons, New Jersey, UK, 2004.

[11] R. C. Gonzales and R. E. Woods. Digital Image Processing. 2nd Edition, Prentice Hall, 2002.

[12] S. Haykin. Neural Networks, A Comprehensive Foundation. IEEE Press, USA, 1984.

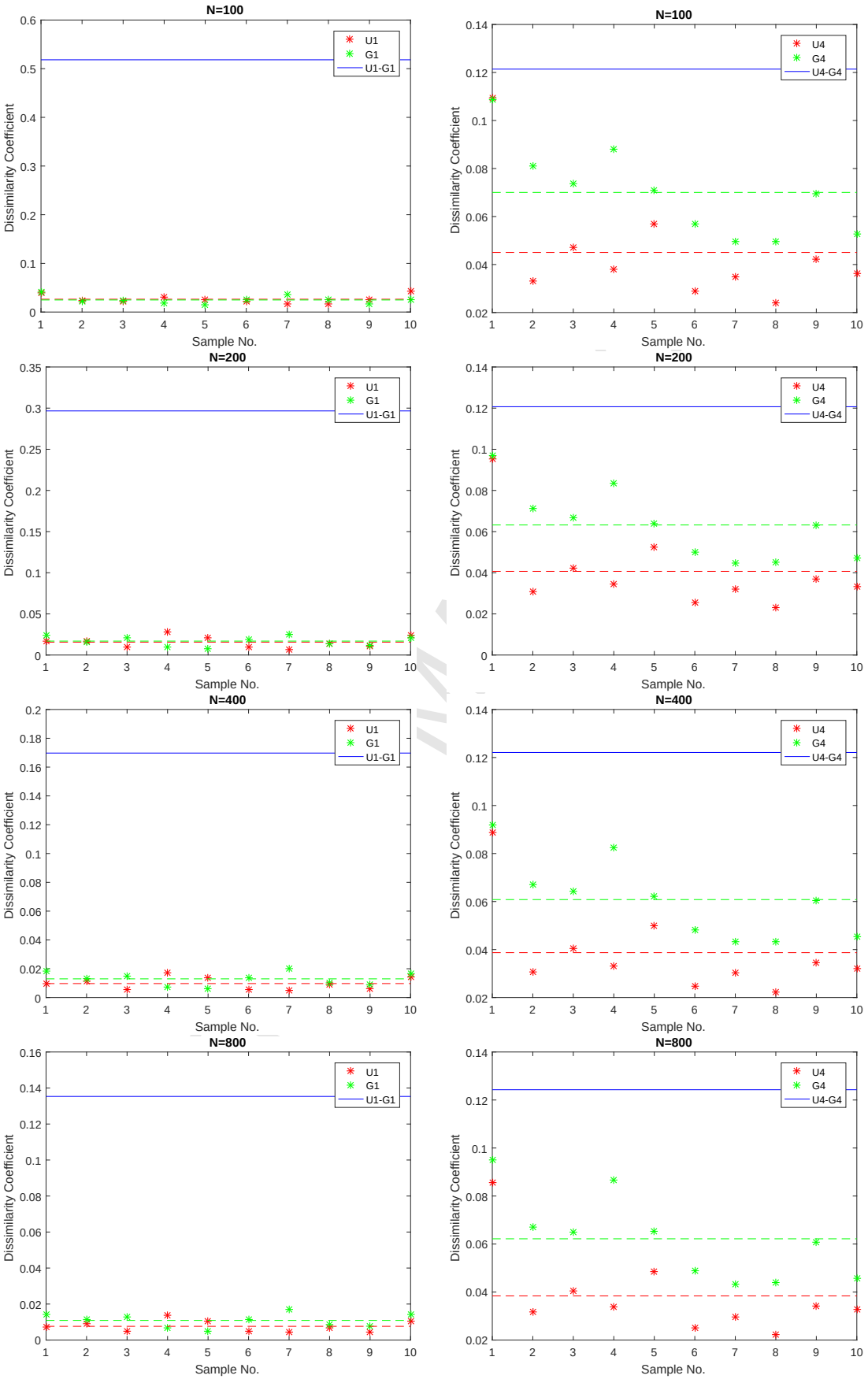
[13] M. Petrou and P. Bosddogianni. Image Processing: The Fundamentals. John Wiley & Sons, UK, 2000.

[14] I. T. Jolliffe. Principal Component Analysis, Springer Series in Statistics, 2nd ed., Springer, NY, 2002

[15] H. Abdi and L. J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics.* 2 (4): 433459, 2010.

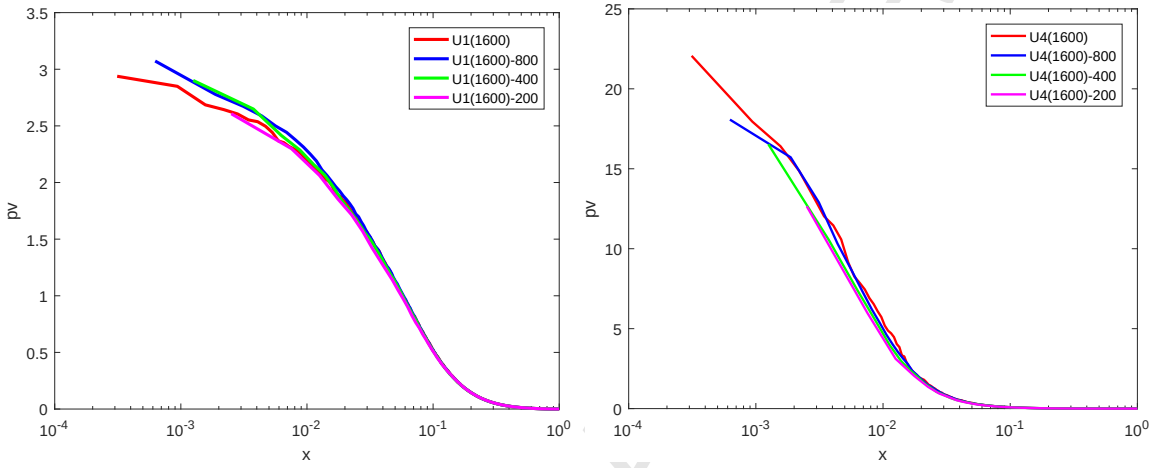
[16] G. H. Golub, C. F. Van Loan. Matrix Computations, 3rd Ed. Johns Hopkins. 1996.

[17] T. Zhao and Y. T. Feng. Characterising 3D Spherical Packings by Principal Component Analysis. *Granular Matter.* 2018 (to appear).



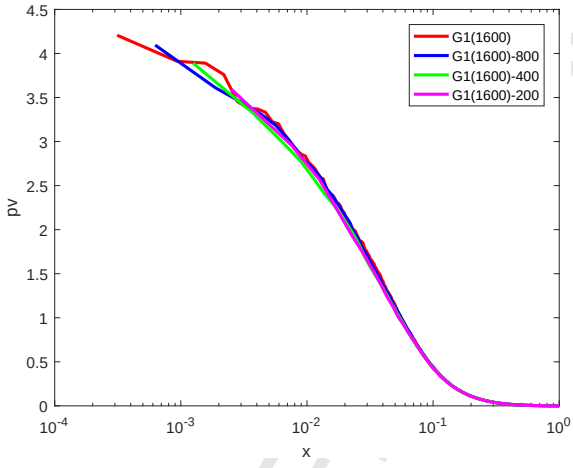
(a) U1, G1 Groups (b) U4, G4 Groups

Figure 10: Dissimilarity coefficients of four packing groups U1, U4, G1 and G4 with four different image resolutions $N = 100, 200, 400$, and 800

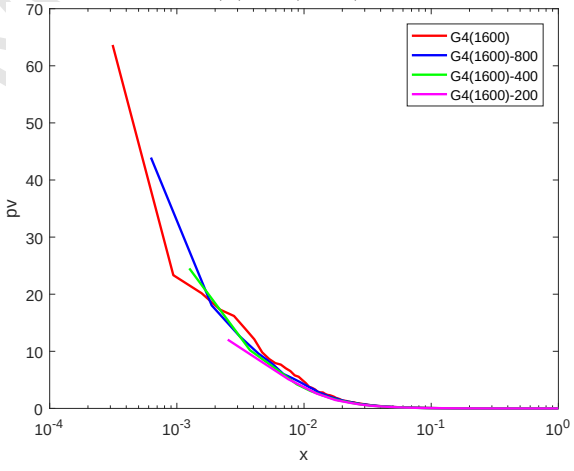


(a) U1(1600)

(b) U4(1600)

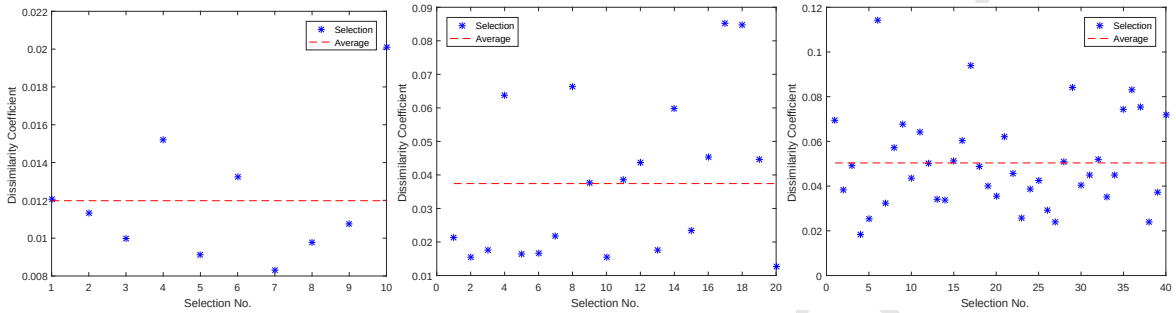


(a) G1(1600)

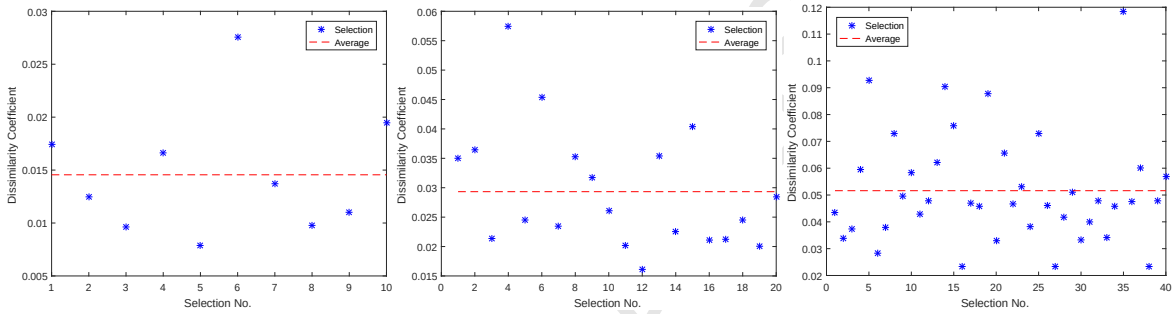


(b) G4(1600)

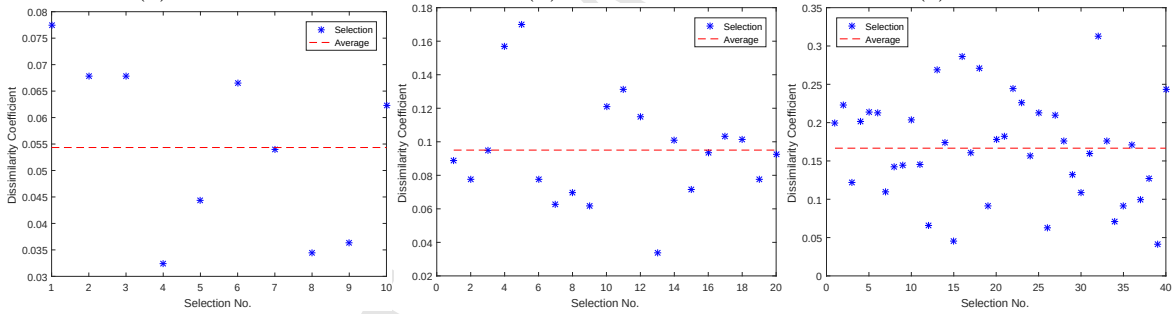
Figure 11: Principal coefficients of four images U1, U4, G1, G4 at N=1600 with three different sub-matrix blocks: $800 \times 800, 400 \times 400$, and 200×200



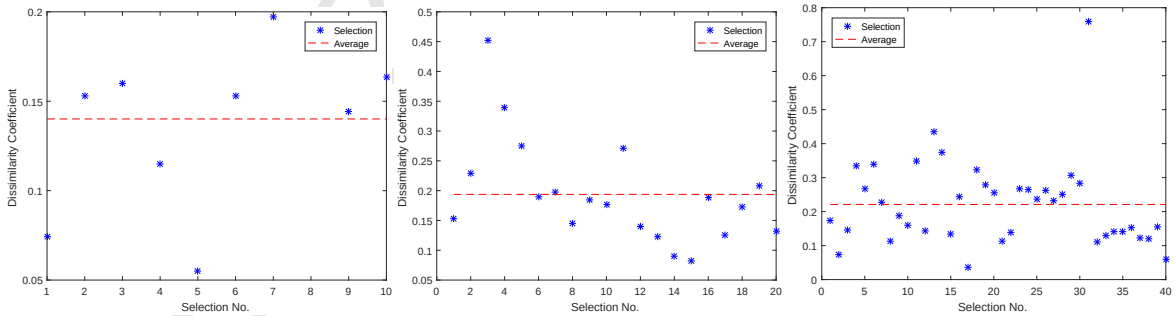
(a) U1-800 (b) U1-400 (c) U1-200



(a) G1-800 (b) G1-400 (c) G1-200

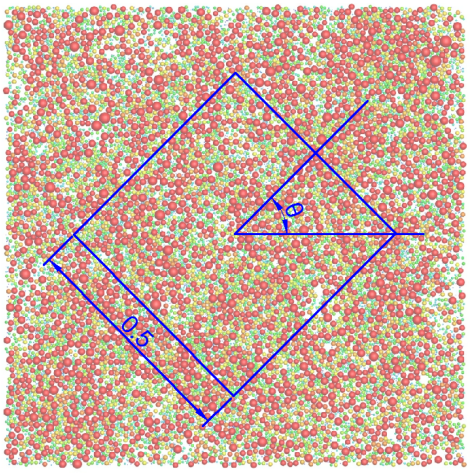


(a) U4-800 (b) U4-400 (c) U4-200

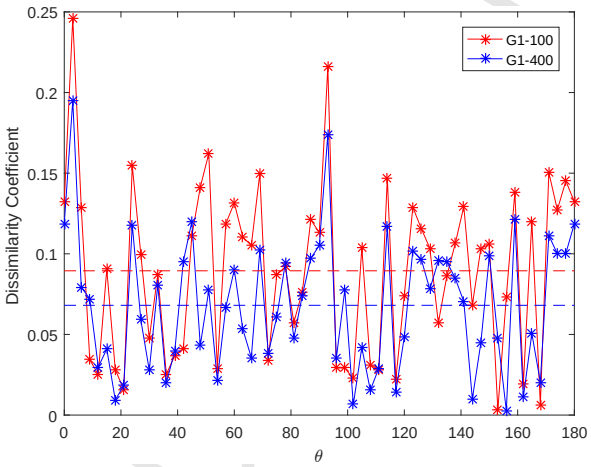


(a) G4-800 (b) G4-400 (c) G4-200

Figure 12: Dissimilarity coefficient distributions at randomly selected positions of groups U1, G1, U4 and G4 for 3 sizes of sub-matrices: N=800, 400, 200

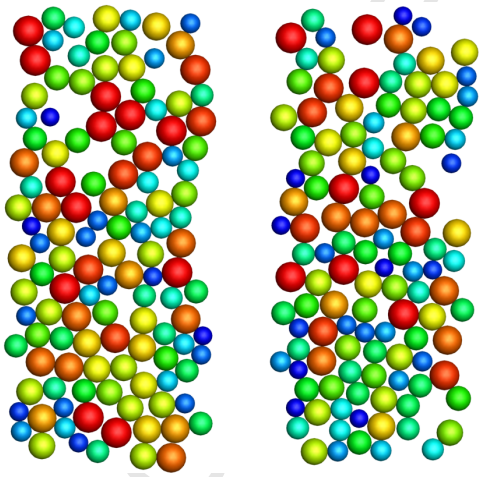


(a) A rotating 0.5×0.5 analysis window

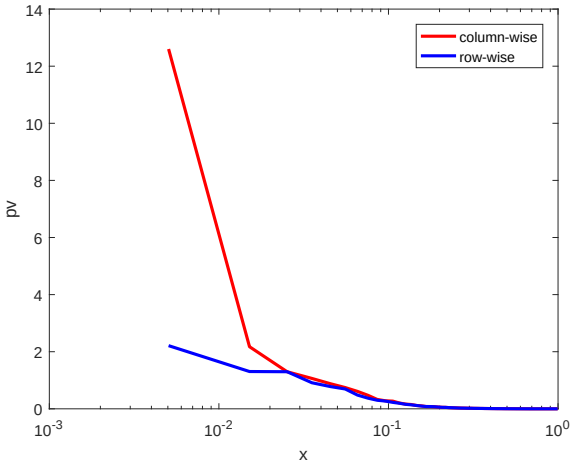


(b) Dissimilarity coefficients for different angles

Figure 13: Isotropic check for G1 using a rotating analysis window

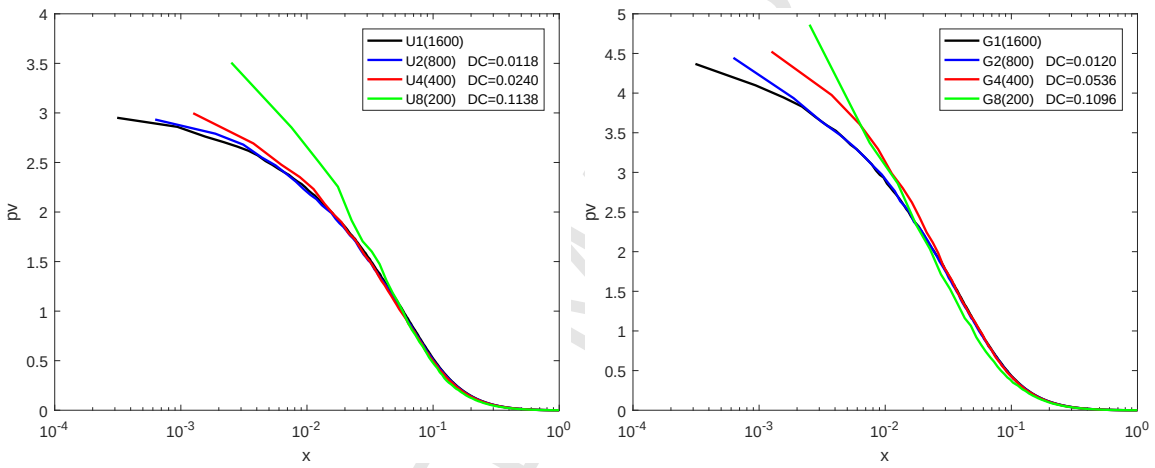


(a) An anisotropic packing



(b) Dissimilarity coefficients for two directions

Figure 14: Packing anisotropic checking in two directions



(a) U-set (b) G-set

Figure 15: Comparison of the principal variance functions and dissimilarity coefficients of scaled packing groups with different resolutions but with the ratio r/h kept the same