Korean J. Chem. Eng., 40(12), 3079-3086 (2023)
DOI: 10.1007/s11814-023-1521-0

pISSN: 0256-1115
eISSN: 1975-7220

Reconstruction of particle size distribution from cross-sections

Jihoon Oh[*,‡], Dongjae Kim[,***,‡], Seunggeon Lee[*], and Jaewook Nam[*,****,†]**

*School of Chemical and Biological Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
**Department of Chemical Engineering, Soonchunhyang University, 22 Soonchunhyang-ro, Shinchang-myeon, Asan-si, Chungcheongnam-do 31538, Korea
***Department of Electronic Materials, Devices, and Equipment Engineering, Soonchunhyang University, 22 Soonchunhyang-ro, Shinchang-myeon, Asan-si, Chungcheongnam-do 31538, Korea
****Institute of Chemical Process, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
(*Received 10 March 2023 • Revised 23 June 2023 • Accepted 29 June 2023*)

Abstract−Controlling the microstructure enables higher energy density and lower energy consumption of a battery. Although particle size distribution is an important property of microstructures, its study is hindered by limited analytical tools. In this study, we precisely estimate the 3-dimensional (3D) spherical size distribution from a 2-dimensional circular size distribution. Here, we introduce the least absolute shrinkage and selection operator (LASSO) regularization method to handle the existing issues in 3D reconstruction efficiently. Using a virtual structure from various predefined distributions, we demonstrate that the LASSO regression outperforms other regularization methods in predicting the original distribution. Finally, we suggest an effective number of cross sections, that is, the minimum required number of cross sections, for 3D reconstruction consisting of spherical particles.

Keywords: LASSO Regression, Stereology, 3D Reconstruction, Wicksell's Corpuscle Problem, Regularization, The Number of Cross-sections

INTRODUCTION

The battery industry has undergone dramatic development owing to the increased demand for diverse applications such as electronic vehicles [1,2], electronics [2,3], and energy storage devices [4,5]. Despite this rapid development, current industrial demands still require batteries with higher energy density and lower energy consumption [6]. Microstructures are considered promising for enhancing battery performance [7-9]. Various studies have shown that controlling microstructural properties, such as particle size distribution, porosity, and tortuosity is correlated with battery performance [10-13]. However, limited methods for analyzing microstructures have hindered progress in related research.

Several cross-sectional SEM images are commonly used to investigate microstructural properties [14]. Especially in investigating particle size distribution, Wicksell's corpuscle problem [15] estimates the 3-dimensional (3D) particle size distribution from the available 2-dimensional (2D) particle size distribution data obtained from cross-sectional images. It is also a known example of stereology [14,16-18]: For example, Saltikov's transformation [16,17,19], Cruz-Olive's transformation [20], and Underwood's transformation [21,22]. These transformations can be applied differently depending on the uniform and geometric bin-size systems. The former implies that all bins in the histogram are uniform, and the latter implies that the bin sizes vary as a geometric sequence. In particular, analyzing the size distribution using a geometric bin-size system is notably used in several fields, such as the size distribution of active materials in electrodes and sieve analyses [23,24].

Based on the above transformation matrices, the 3D particle size distribution can be estimated using the product of the inverse of the transformation matrix and the 2D available particle size distribution. Nevertheless, this direct method has unphysical issues such as negative frequencies or small oscillations in some bins [25]. In particular, we discovered that these problems are more severe in geometric bin-size systems. Several regularization methods have been introduced to manage these issues, including expectation maximization (EM), smoothed expectation maximization (EMS), and wavelet regularization [26-28]. Although these regularization methods can provide better results, they still offer inaccurate solutions.

Therefore, in this study, we developed a new regularization method that is applied to Wicksell's corpuscle problem with a geometric bin-size system: LASSO regression [29]. First, the LASSO equation was developed for application to Wicksell's problem. We then constructed a virtual structure to obtain the 2D diameter distribution and 3D ground-truth information. From the 2D diameter distribution of the virtual structure, the 3D diameter distribution was estimated using various regularization methods, and the accuracy of LASSO regression was compared with that of other methods. Finally, we investigated the required number of cross-sections to reasonably reconstruct the entire distribution based on the LASSO regression.

METHODS

1. Theoretical Background: Uniform Interval Width of Bins

We briefly introduce here the background method for estimat-

†To whom correspondence should be addressed.
E-mail: jaewooknam@snu.ac.kr
‡These authors contributed equally to this work.

ing the number density distribution of spheres ($\mathbf{N}'$) from that of circles in cross-sectional images ($\mathbf{n}'$). Note that the description in this section is for a linear-scale bin with a uniform interval width. In the following section, we extend the theory to a geometric scale bin, which is our primary target for the remainder of the paper.

We assume that $\mathbf{n}'$ is the distribution of circles obtained by random sectioning of 3D structures composed of spheres with distribution $\mathbf{N}'$. Each element of the vectors represents the absolute number density in predefined bins, where the bins are divided by the diameters of the sphere and circle. N_i' and n_i' are the number densities of the i^{th} bins of the spheres and circles, respectively. According to Pabst et al. [25], the number density distribution of circles can be expressed as the product of the transformation matrix $\mathbf{B}^{lin}$ and the number density distribution of spheres.

$$\mathbf{n}' = \mathbf{B}^{lin}\mathbf{N}'. \tag{1}$$

Let us denote B_{ij}^{lin} as the i^{th} row and j^{th} column element of $\mathbf{B}^{lin}$. Based on Pabst et al. [25], the transformation matrix can be modeled as the product of the statistical weight (w_j) and conditional probability (P_{ij}) as follows:

$$B_{ij}^{lin} = w_j P_{ij}, \tag{2}$$

w_j is a weight related to the probability of choosing the j^{th} bin of the sphere, simply considering the representative diameter of each bin, D_j. Meanwhile, P_{ij} is the probability that the j^{th} bin of $\mathbf{N}'$ converted into the i^{th} bin of $\mathbf{n}'$. Owing to its geometry, P_{ij} can be expressed as [30]

$$P_{ij} = \frac{\sqrt{D_j^2 - d_{i-1}^2} - \sqrt{D_j^2 - d_i^2}}{D_j}, \tag{3}$$

where d_i denotes the representative diameter of the i^{th} bin of the circle. Substituting Eq. (3) into Eq. (2) results in

$$B_{ij}^{lin} = \sqrt{D_j^2 - d_{i-1}^2} - \sqrt{D_j^2 - d_i^2}. \tag{4}$$

Here we are looking at the evenly spaced bins; therefore, the interval width of the bins (Δ) is constant. The first bin starts at zero, and the representative diameters (D_j and d_i) are assumed to be the upper-limit values for each bin. This means $D_j = j\Delta$ and $d_i = i\Delta$. Furthermore, different representative values for D_j can be used as $\Delta(j-\alpha)$, where α is a shifting parameter indicating the distance of the representative diameter from the upper limit of the bin [25].

Substituting these relationships into Eq. (4), and B_{ij}^{lin} is expressed as follows:

$$B_{ij}^{lin} = \Delta\cdot\left[\sqrt{(j-\alpha)^2 - (i-1)^2} - \sqrt{(j-\alpha)^2 - i^2}\right] \text{ for } (i<j), \tag{5}$$

$$B_{ij}^{lin} = \Delta\cdot\left[\sqrt{(2-2\alpha)i - 1 + \alpha^2}\right] \text{ for } (i=j), \tag{6}$$

$$B_{ij}^{lin} = 0 \text{ for } (i>j). \tag{7}$$

As mentioned, circles are obtained from the cross-sectional images by random sectioning of the 3D sphere structure. Therefore, the diameter of the circle never exceeds the diameter of the sphere (Eq. (7)), resulting in the upper triangular matrix $\mathbf{B}^{lin}$.

2. Geometric Scale Interval Width of Bin (Geometric Bin-size System)

It should be emphasized that Eqs. (5)-(7) are only valid for uniform bin intervals. However, in some cases, the bin size cannot be designed equally, as in the sieve analysis of powder [24]. Assuming that the bin interval increases geometrically, the basic governing equation and derivation procedure are identical to those in Eqs. (1)-(4); however, we used $\mathbf{B}^{geo}$ and B_{ij}^{geo} instead of $\mathbf{B}^{lin}$ and B_{ij}^{lin}, respectively.

$$\mathbf{n}' = \mathbf{B}^{geo}\mathbf{N}'. \tag{8}$$

$$B_{ij}^{geo} = \sqrt{D_j^2 - d_{i-1}^2} - \sqrt{D_j^2 - d_i^2}. \tag{9}$$

However, in the geometric bin-size system, D_j and d_i are not simple functions of Δ because Δ is no longer constant. The lower limits of the first bin for both $\mathbf{N}'$ and $\mathbf{n}'$ are denoted as d_{min}. It should not be zero; otherwise, other bin values cannot be defined. Then, D_j and d_i are expressed as

$$D_j = d_{min} + \sum_{k=1}^{j}\Delta k - \alpha\Delta_j, \tag{10}$$

$$d_i = d_{min} + \sum_{k=1}^{i}\Delta k. \tag{11}$$

Δ_k is the interval width of the k^{th} bin. The corresponding transformation matrix is obtained by substituting Eqs. (10) and (11) into Eq. (9):

$$B_{ij}^{geo} = \sqrt{\left(d_{min} + \sum_{k=1}^{j}\Delta_k - \alpha\Delta_j\right)^2 - \left(d_{min} + \sum_{k=1}^{i-1}\Delta_k\right)^2} - \sqrt{\left(d_{min} + \sum_{k=1}^{j}\Delta_k - \alpha\Delta_j\right)^2 - \left(d_{min} + \sum_{k=1}^{i}\Delta_k\right)^2} \text{ for } (i<j), \tag{12}$$

$$B_{ii}^{geo} = \sqrt{\left(d_{min} + \sum_{k=1}^{j}\Delta_k - \alpha\Delta_j\right)^2 - \left(d_{min} + \sum_{k=1}^{i-1}\Delta_k\right)^2} \text{ for } (i=j), \tag{13}$$

$$B_{ij}^{geo} = 0 \text{ for } (i>j). \tag{14}$$

3. Applying LASSO Regression to Wicksell's Corpuscle Problem

The inverse matrices of $\mathbf{B}^{lin}$ and $\mathbf{B}^{geo}$ exist because they are upper triangular matrices. However, it turns out that the $\mathbf{N}'$ estimated by the inverse matrix of Eqs. (1) and (8) shows undesirable results, such as negative frequencies or oscillations. These results are attributed to the large condition number of the transformation matrix, leading to oscillations that cause negative frequency values. Thus, various regularization methods have been introduced to remove oscillations. In this study, we used LASSO regression [29] to address these unphysical issues. LASSO regression is a well-known effective regularization method for eliminating noise by shrinking small values and is therefore widely used for various applications such as denoising [31], image reconstruction [32], and machine learning [33,34].

Before applying LASSO regression to Eq. (8), we slightly modified the equation because the values of $\mathbf{n}'$ were too small, that is, vulnerable to round-off errors. Herein, we changed the "number density distribution" to "number distribution." Based on this definition, the number *density* distributions $\mathbf{N}'$ and $\mathbf{n}'$ can be expressed as

$$\mathbf{N}' = \frac{\mathbf{N}}{L^3}, \tag{15}$$

$$\mathbf{n}' = \frac{\mathbf{n}}{L^2}. \tag{16}$$

N and n are vectors representing the number distributions whose elements are the actual numbers of spheres and circles in each bin, respectively. We assume that the spheres are distributed in cubic space, whose side length is L. Recall that spheres are distributed in 3D space, whereas circles are in the 2D plane obtained by cross-sectioning. Eqs. (1) and (8) are converted into

$$\mathbf{n} = \frac{1}{L}\mathbf{B}^{lin}\mathbf{N}, \tag{17}$$

$$\mathbf{n} = \frac{1}{L}\mathbf{B}^{geo}\mathbf{N}. \tag{18}$$

One can modify Eqs. (15)-(18) for a noncubic system. In this study, we focused on geometric-scale bin cases, that is, Eq. (18). See Fig. S3 in the Supplementary Material for the uniform bin-size cases.

In this study, **N** was estimated from **n** using LASSO regression based on Eq. (18). The objective function is expressed as

$$\underset{\mathbf{N}}{argmin} \left\| \frac{1}{L}\mathbf{B}^{geo}\mathbf{N} - \mathbf{n} \right\|_2^2 + \lambda \|\mathbf{N}\|_1 \text{ subject to } {}^{\forall}N_i \geq 0, \tag{19}$$

where $\|\cdot\|_2$ and $\|\cdot\|_1$ are the notations for 2-norm and 1-norm, respectively, λ is a regularization parameter, and N_i denotes each component in vector **N**. To address the previously mentioned unphysical issues, we add the regularization term ($\lambda\|\mathbf{N}\|_1$) to reduce the noisy oscillation in the solution **N**. Also, the constraint (${}^{\forall}N_i \geq 0$) in the optimization statement prevents having negative values. To solve the optimization problem, we adopted an interior-point algorithm implemented in the MATLAB®-built function "fmincon." The numerical experiments were performed using MATLAB® 2020b on a computer with an Intel Core i7-10750H CPU at 2.60 GHz and 16 GB RAM.

4. Virtual Structure Generation

In this study, 3D virtual structures were used to obtain the 2D circle distribution and the 3D sphere ground-truth. Based on these structures, we demonstrate the proposed LASSO regression-based stereology. In summary, we mimicked a Tetris-like strategy, similar to Shi and Zhang's work [35], to generate high-packing-density structures without overlapping particles. First, we generated an empty cube with L=500 (arbitrary unit) aligned in Cartesian coordinates: x and y axes in the horizontal plane and the z axis in the vertical direction. Then, the diameter distribution of the spheres was determined. We considered the normal, bimodal, lognormal, and Rayleigh distributions. The radius of a sphere was randomly selected from the considered distribution. The initial position of the sphere was arbitrarily located at the top of the system (z=L). The sphere then dropped vertically until it hit other existing spheres or the bottom plane (z=0), and the corresponding z-height was collected. Each sphere was dropped ten times at different initial positions, and then the sphere was pinned at the minimum z-height location among the ten trials. This entire procedure was repeated until the target packing density was attained. After generating the structure, the circle diameter distribution was obtained by the xz plane or yz plane sectioning of the structure. The circular diameter was calculated using the Pythagorean theorem with coordinate information of the spherical center.

RESULTS AND DISCUSSION

Fig. 1 illustrates the entire procedure of LASSO-based stereology. First, a 3D structure consisting of randomly distributed spheres was obtained using microscopy or numerically generated virtual structures. We considered only virtual structures to demonstrate the effectiveness of the proposed method. Circular diameter distributions were obtained by collecting several cross-sections. Using Eq. (19), the 3D sphere diameter distribution was reconstructed from the circle diameter distribution.

1. Demonstration

The estimated distribution of spheres was calculated using three major procedures, as shown in Fig. 2. The first step was to obtain the number distribution of circles **n**. The range of the bins was set to the minimum diameter of the circle and maximum diameter of the sphere. Subsequently, the number of bins m could be determined arbitrarily. We set m to 25 considering the resolution in sieve analysis. It is defined as the ratio of the range of each bin to the median of each bin. For grain size distribution analysis, the empirically recommended resolution is close to or less than 0.25 [36]. Interestingly, 25 bins are typically used in Saltikov's transformation [16]. The number distribution of the circles was obtained by sectioning the virtual structure system, as shown in Fig. 2(a), and redistributing it according to a previously chosen bin size.

The number of cross-sections, which is a user-defined parameter, determines the accuracy of the solution because a larger sample ensures an analogy with the population. In Section 3.2 we performed numerical experiments to determine the desirable number of

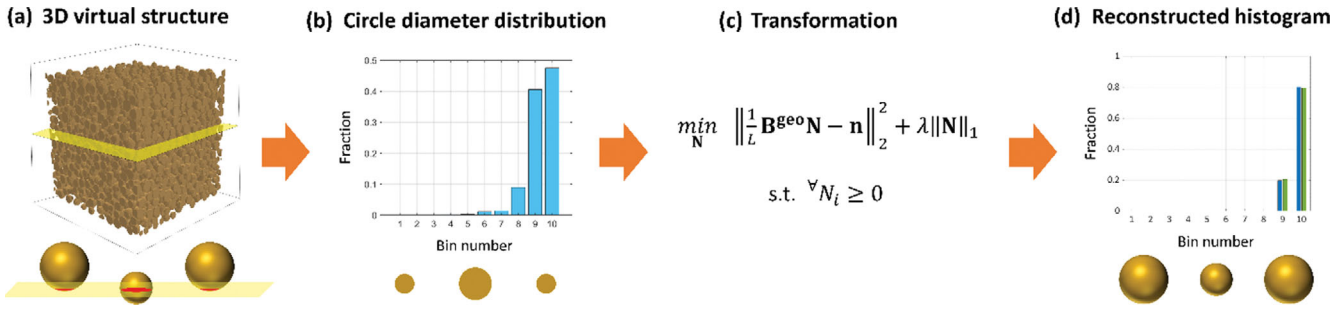

Fig. 1. Overall process of the numerical experiment.

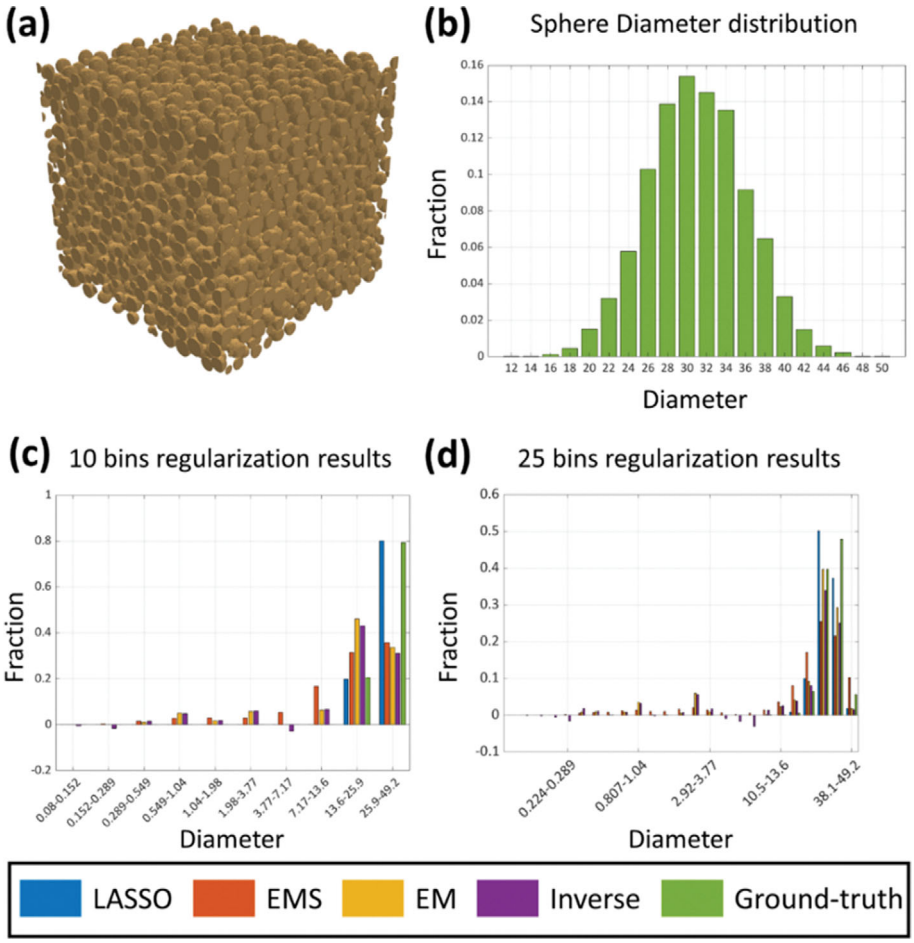

Fig. 2. Virtual structure and reconstructed histograms. (a) an example of virtual structure, (b) the spherical diameter distribution of the virtual structure in (a) (linear bin-size), (c) a reconstructed histogram when the number of bins is ten (geometric bin-size), and (d) a reconstructed histogram when the number of bins is 25 (geometric bin-size).

cross-sections. Therefore, we selected 15 cross-sections to obtain **n**. During sectioning, simple random sampling was used such that all positions could be equivalently selected. During our study, we found that iso-distance sampling, which means that the distance between cross-sections is uniform, misrepresents the internal configuration of the system, such as the shape of a virtual structure.

The second step was to solve the equations based on **n** and $\mathbf{B}^{geo}$. In this study, we consider solutions using four different methods. The inverse method is based on the product of the inverse matrix of the transformation matrix $(\mathbf{B}^{geo})^{-1}$ and circular distribution **n**. The LASSO method is based on solving LASSO Eq. (19). Here, the hyperparameter λ is chosen to minimize the sum of the squared errors (SSE), which is used as a standard for representing the accuracy. More information regarding the SSE is provided in Section 3.2. To test the validity of the LASSO method, we considered the solutions from the EMS and EM methods as well. The EMS solution, which is based on Wilson's work [27], was solved by applying a smoothed EM algorithm to Eq. (18). The EM solution was derived by adopting the EM algorithm in Eq. (18) [27]. In all methods, α can be arbitrarily chosen; however, in Fig. 2(c) and 2(d), we chose $\alpha=0$, which corresponds to Saltikov's transformation.

The last step involves normalizing the above four solutions and displaying them with the ground-truth as a histogram. Fig. 2(b) shows an example of the ground-truth, normal distribution (μ=15, σ=2.5). Each of the four solutions and the ground-truth were normalized by dividing them by the sum of their vector elements. These normalization results are referred to as fractions in this study. We then plotted a histogram with these fraction results, as shown in Fig. 2(c) and 2(d), where the number of bins m is 10 and 25, respectively. Herein, note that the geometric bin-size was used. From these figures, we can qualitatively confirm that the LASSO solution is the closest to the ground-truth regardless of the number of bins.

2. Methods Comparison

In this study, SSE is defined as the square of the 2-norm value for the difference between the solution and the ground-truth, as shown in Eq. (20).

$$\mathrm{SSE}=\sum_i(\mathrm{N}_i^{cal}-\mathrm{N}_i^{gt})^2. \tag{20}$$

$\mathbf{N}^{cal}$ is a vector estimated by each method and $\mathbf{N}^{gt}$ is the ground-truth. Many other studies have used goodness-of-fit (GoF) tests, such as the chi-square or Kolmogorov-Smirnov tests, as the quantitative standard for accuracy [37,38]. These GoF tests are based on com-

Table 1. Information about distributions

Distributions	PDF	Parameters
Normal	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ=15, σ=2.5
Bimodal	Summation of the two normal distributions	μ_1=10, σ_1=2.5 μ_2=20, σ_2=2.5
Lognormal	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$	μ=2.69, σ=0.17
Rayleigh	$f(x)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right)$	σ=11.97

paring the estimated curve with a probability density function (PDF); however, recall that our primary purpose is to analyze geometric bin-size systems. Since converting a discrete histogram to a continuous PDF is not feasible (as far as the authors' knowledge), especially for geometric bin-size systems, we directly compared the histogram by SSE instead of GoF tests.

We considered four distributions (normal, bimodal, lognormal, and Rayleigh distributions) for the quantitative accuracy analyses, and Table 1 summarizes detailed information on these distributions. A bimodal distribution was generated by combining two normal distributions with the same variance but different mean values. We set the parameters of each distribution so that the mean value of each distribution became 15 for consistent analyses. Note that the mean values are μ for the Normal distribution, $\frac{\mu_1+\mu_2}{2}$ for the bimodal distribution, $\exp\left(\mu+\frac{\sigma^2}{2}\right)$ for the lognormal distribution, and $\sigma\sqrt{\frac{\pi}{2}}$ for the Rayleigh distribution.

We generated 30 virtual structures for each distribution to evaluate the performances of the four regularization methods with α=0. For each virtual structure, 10,000 SSE values were estimated. Notably, each SSE result was obtained from different random cross-sections of the virtual structure.

Fig. 3 presents the mean SSE of each method applied to the four distributions. The standard deviations are represented by the black bars. From Fig. 3, we confirmed that the LASSO solution has the smallest mean SSE values and standard deviations for all ground-truth distributions. We believe that these are derived from the original feature of LASSO, which effectively reduces the variance of data [33].

We also investigated the influence of the number of bins, the results of which are shown in Fig. 4. Except for the number of bins, all parameters and experimental procedures are the same as those in Fig. 3. In Fig. 4, although the tendencies of the graphs differ for each distribution, the mean SSE value monotonically decreases as the number of bins increases for all distributions. We suspect that the numerical error caused by LASSO regularization decreases as the number of bins increases because the histogram becomes a more continuous curve. However, using too many bins is not feasible for practical applications, e.g., 10-25 bins were used for typical sieve analysis. Therefore, 25 bins were chosen for the remainder of the analyses.

In general, the regularization parameter determines the regularization performance. Because the proposed method (Eq. (19)) is based on LASSO regularization, the influence of λ on the LASSO method should be investigated. Experiments were conducted on 30 samples of each distribution, and the optimized λ values in each sample were determined when the result showed the smallest SSE. The findings are summarized in Tables S1 and S2 of the Supplementary Material. Although the optimal λ value varied depend-

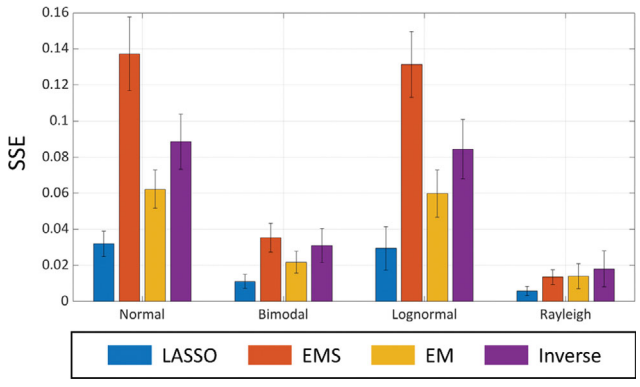

Fig. 3. SSE analysis results in geometric bin-size system.

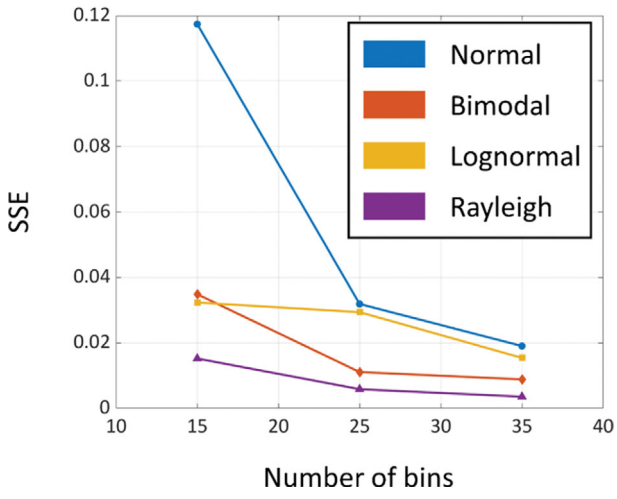

Fig. 4. Mean SSE according to the number of bins.

ing on sample type and distribution, we found that the optimal λ ranged from 10^{-2} to 1. Furthermore, we empirically discovered that the proposed method with an optimal λ provided the smallest SSE, regardless of the α value (Supplementary Fig. S2).

In addition to the number of bins and the regularization parameter, the number of cross-sections can affect LASSO performance. We investigated the correlation between the SSE of the proposed methods and the number of cross-sections. Fig. S4 in the Supplementary Material presents the results. In all types of distributions, the proposed methods provided the smallest mean SSE regardless of the number of cross-sections. The 15 cross-sections in Fig. S4 indicate that the SSE converges toward a specific value. Therefore, we conclude that the proposed method outperforms all other methods considered in this study.

3. Effective Number of Cross-sections

In this section, we investigate the effective number of cross-sections, which is the minimum required number of cross-sections that suitably reconstructs the original 3D distribution. The proportion of non-outliers in the entire SSE data of the LASSO method was used to determine the effective number of cross-sections. Here, a data point is considered an outlier if it does not exist between a value of 1.5 the interquartile range (IQR) away from the first quartile (Q1) and the third quartile (Q3): Q1-1.5IQR~Q3+1.5IQR. The proportion of non-outliers was evaluated by implementing 3,000 samples: a combination of 30 virtual structures and 100 different sections per number of cross-sections. Fig. S5 in the Supplementary Material shows each result overlaid on a one coordinate, and Fig. 5 presents the average value of Fig. S5.

The proportion of non-outliers increases with the number of cross-sections because reconstruction is performed by utilizing the given structure as much as possible if many cross-sections are used. When the number of cross-sections is sufficiently large in all types of distributions, the graph approximately converges toward unity. Here, we would like to emphasize the steep slope of the graph for a small number of cross-sections. When the proposed method is applied to a geometric bin system, this suggests that 3D distribution reconstruction is possible with a small number of sections. Furthermore, the overall shapes of the graphs are almost identical.

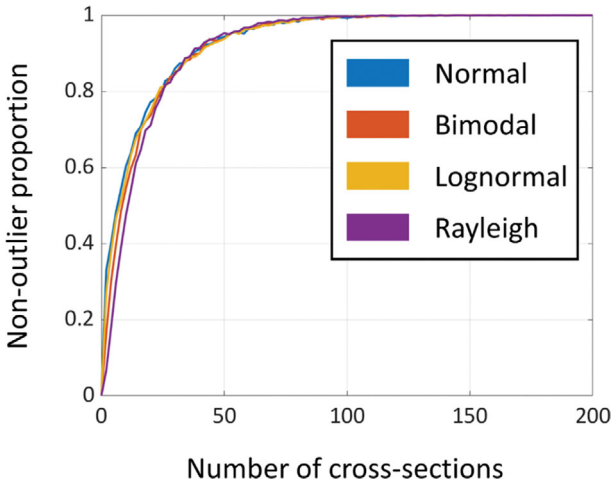

Fig. 5. Average non-outlier proportion curves according to the number of cross-sections.

Here, the effective number of cross-sections is defined as the point at which the proportion of non-outliers is 0.95. The effective number of cross-sections for each distribution was calculated as follows: 54 for a normal distribution, 54 for a bimodal distribution, 56 for a lognormal distribution, and 52 for a Rayleigh distribution. The 52-56 cross-sections would be fascinating options for reconstructing the 3D distribution. However, in practice, the number of cross-sectional images could not be obtained owing to the cost or time limit of microscope use. In this case, we suggest 15 cross-sections because they correspond to a proportion of non-outlier values around 0.7 (see Fig. 5). Empirically, we figured out that 15 cross-sections lead to suitable results. Indeed, in Section 3.1, we used this number to compare the qualitative performances of various regularization methods.

4. In-depth Analysis for Lognormal Distribution

We investigated the impact of the standard deviation of the lognormal distribution on the effective number of cross-sections. Based on the literature, Depriester and Kubler [38] discovered that as a variance of the sphere distribution increases, a larger number of cross-sections are required. Although their experiments used a uniform bin-size system, we performed similar experiments to observe the difference between the uniform and geometric bin-size system. Here, we set the lognormal distributions having the parameters summarized in Table 2. The mean value of the lognormal distribution ($\hat{E}$) was set to 15, and the standard deviations of the lognormal distribution (s) were changed to s=2.5, 5, 7.5, and 10. We used the proposed method to reconstruct 3D distribution for 300,000 random cases per standard deviation s, and the SSE was recorded as a quantitative index.

Fig. 6 shows the mean and standard deviation of the SSE distributions according to the number of cross-sections. The line represents the mean SSE, and the shaded area indicates one standard deviation. As shown in Fig. 6, we confirmed that the SSE steeply decreased in the first stage and approached a constant value after approximately 50 cross-sections. Similarly, the standard deviation of the SSE exhibited the same pattern as the mean of the SSE. Note that this coincides with Fig. 5. These results support that the effective number of cross-sections can be approximately 50, regardless of the standard deviation of the lognormal distribution.

As shown in Fig. 6, one can observe that the means and standard deviations of the SSE decreased as the standard deviation of the lognormal distribution increased. Note that the standard deviation of the SSE (red shaded area) indicates the uncertainty of the methods. Thus, the reconstructed 3D distribution would be different by which set of cross-sections was considered even if the same sample is provided. Interestingly, this is the opposite of the work of

Table 2. Detailed information about lognormal distributions

PDF	Parameters
$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$	$\hat{E}$=15, s=2.5 (μ=2.69, σ=0.17)
	$\hat{E}$=15, s=5 (μ=2.66, σ=0.32)
	$\hat{E}$=15, s=7.5 (μ=2.60, σ=0.47)
	$\hat{E}$=15, s=10 (μ=2.52, σ=0.61)

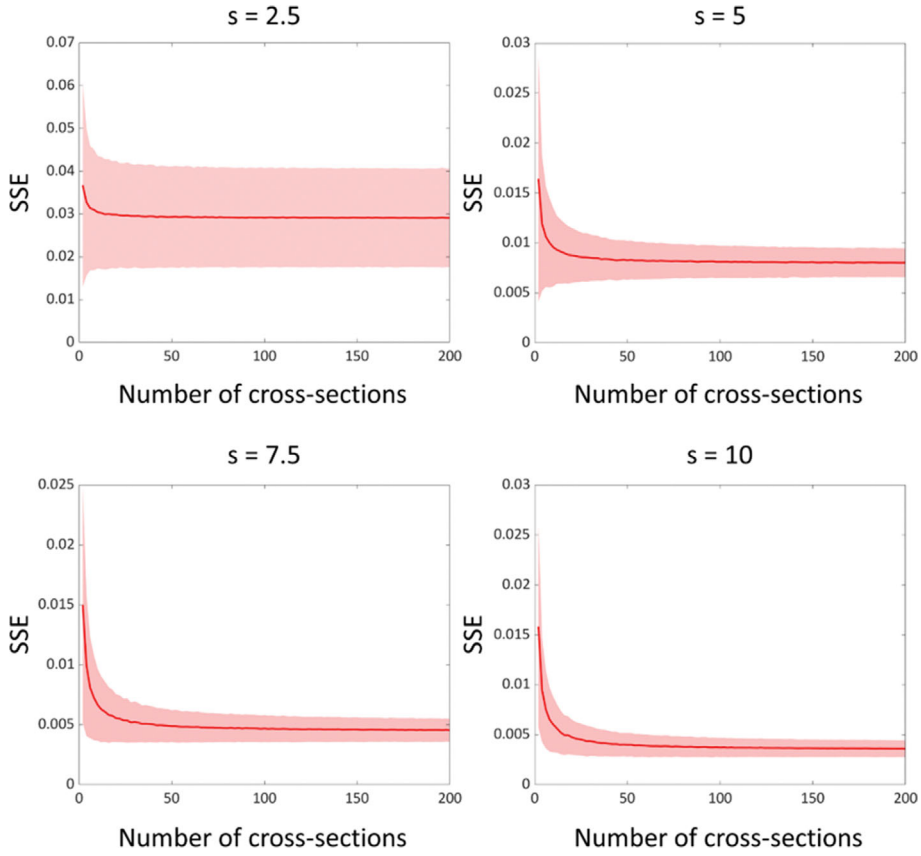

Fig. 6. Mean and standard deviation of SSE concerning the number of cross-sections in lognormal distribution.

Depriester and Kubler. We believe that this difference is due to the difference in the bin scale. Depriester and Kubler used a uniform bin-size system, whereas we focused on a geometric bin-size system. To clarify this difference, we are currently conducting a theoretical study to elucidate the fundamentals of these deviations.

FINAL REMARKS

In this study, we first applied LASSO regression to Wicksell's corpuscle problem in a geometric bin-size system. The accuracy of the LASSO method was tested by comparing its SSE values with those of other regularization methods for various ground-truth distributions. These comparison tests showed that the proposed method provided the smallest mean SSE values and SSE variances for all distributions. In addition, we propose an effective number of cross-sections from the average proportion of the non-outlier curve. Beyond the effective number, all the proportion curves became significantly insensitive, regardless of the distribution and virtual structure. Therefore, the number can be considered as a part of important guidelines to reconstruct the original 3D particle size distribution from the available 2D cross-sectional images.

We considered Saltikov's and Cruz-Olive's transformations; however, there are other transformation methods for various conditions, such as Underwood's transformation [21,22]. Currently, we are evaluating the performance of our reconstruction method when using the transformation matrix from Underwood's transformation. In addition, we are working on developing methods for estimation optimal λ without ground-truth information. Finally, we will expand the proposed method to anisotropic and irregularly shaped particles found in various industrial applications including battery electrodes.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. NRF-2018R1A5A1024127, NRF-2023R1A2C2004002, and NRF-2021M3H4A6A01041234), and Soonchunhyang University Research Fund (No. 20221186).

NOMENCLATURE

N′ : the number density distribution of spheres
n′ : the number density distribution of circles
N : the number distribution of spheres
n : the number distribution of circles
$\mathbf{B}^{lin}$: transformation matrix in uniform bin-size system
$\mathbf{B}^{geo}$: transformation matrix in geometric bin-size system
w : statistical weight of each bin
P : conditional probability matrix
D : representative spherical diameter of each bin
d : representative circular diameter of each bin

$\mathbf{U}$: the upper limit value of each bin
Δ : the size of bin
α : shifting parameter
d_{min} : the lower limits of the first bin
L : the length of virtual structure
λ : LASSO parameter
m : the number of bins
μ : the mean of certain distribution
σ : the standard deviation of certain distribution
$\mathbf{N}^{cal}$: the calculated number distribution of spheres
$\mathbf{N}^{gt}$: the number distribution of spheres for ground-truth
$\hat{E}$: the mean of lognormal distribution
s : the standard deviation of lognormal distribution

Abbreviations

3D : 3-dimensional
2D : 2-dimensional
LASSO : least absolute shrinkage and selection operator
EM : expectation maximization
EMS : smoothed expectation maximization
SSE : sum of squared errors
IQR : interquartile range
Q1 : first quartile
Q3 : third quartile.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Jihoon Oh: Conceptualization, Methodology, Software, Formal analysis, Writing- original draft. **Dongjae Kim:** Conceptualization, Methodology, Investigation, Writing- reviewing and editing. **Seunggeon Lee:** Software, Investigation. **Jaewook Nam:** Methodology, Supervision, Writing- reviewing and editing.

SUPPORTING INFORMATION

Additional information as noted in the text. This information is available via the Internet at http://www.springer.com/chemistry/journal/11814.

REFERENCES

1. A. M. Andwari, A. Pesiridis, S. Rajoo, R. Martinez-Botas and V. Esfahanian, *Renew. Sust. Energ. Rev.*, **78**, 414 (2017).
2. B. Dunn, H. Kamath and J.-M. Tarascon, *Science*, **334**, 928 (2011).
3. G. Zubi, R. Dufo-López, M. Carvalho and G. Pasaoglu, *Renew. Sust. Energ. Rev.*, **89**, 292 (2018).
4. M. Bragard, N. Soltau, S. Thomas and R. W. De Doncker, *IEEE Trans. Power Electron.*, **25**, 3049 (2010).
5. B. Kang and G. Ceder, *Nature*, **458**, 190 (2009).
6. J. Janek and W. G. Zeier, *Nat. Energy*, **1**, 1 (2016).
7. D. R. Nevers, S. W. Peterson, L. Robertson, C. Chubbuck, J. Flygare, K. Cole and D. R. Wheeler, *J. Electrochem. Soc.*, **161**, A1691 (2014).
8. P. Shearing, R. Bradley, J. Gelb, N. Brandon and P. Withers, *Microsc. Microanal.*, **17**, 1672 (2011).
9. D. Kim, S. Lee, W. Hong, H. Lee, S. Jeon, S. Han and J. Nam, *Microsc. Microanal.*, **25**, 1139 (2019).
10. Z. Chen, W. Zhang and Z. Yang, *Nanotechnology*, **31**, 012001 (2019).
11. S. T. Taleghani, B. Marcos, K. Zaghib and G. Lantagne, *J. Electrochem. Soc.*, **164**, E3179 (2017).
12. M. Kishimoto, H. Iwai, M. Saito and H. Yoshida, *ECS Trans.*, **25**, 1887 (2009).
13. M. M. Majdabadi, S. Farhad, M. Farkhondeh, R. A. Fraser and M. Fowler, *J. Power Sources*, **275**, 633 (2015).
14. K. Wiencek, T. Skowronek and B. Khatemi, *Metall. Foundry Eng.*, **31**, 167 (2005).
15. V. Wernert, B. Coasne, P. Levitz, K. L. Nguyen, E. J. Garcia and R. Denoyel, *Chem. Eng. Sci.*, **264**, 118136 (2022).
16. M. A. Lopez-Sanchez and S. Llana-Fúnez, *J. Struct. Geol.*, **93**, 149 (2016).
17. D. Depriester and R. Kubler, *Image Anal. Stereol.*, **38**, 213 (2019).
18. N. Keiding and S. T. Jensen, *Biometrics*, **28**, 813 (1972).
19. S. A. Saltikov, in *Stereology*, Springer (1967).
20. L. M. C. Orive, *J. Microsc.*, **112**, 153 (1978).
21. E. E. Underwood, *Quantitative stereology*, Addison-Wesley Publ. Co., Reading (1970).
22. E. Underwood, in *Stereology and quantitative metallography*, ASTM International (1972).
23. E. F. Maher and N. M. Laird, *J. Aerosol Sci.*, **16**, 557 (1985).
24. M. Konert and J. Vandenberghe, *Sedimentology*, **44**, 523 (1997).
25. W. Pabst and T. Uhlířová, *Ceram. Silik*, **61**, 147 (2017).
26. K. C. G. Chan and J. Qin, *Biometrika*, **103**, 273 (2016).
27. J. Wilson, *J. Stat. Comput. Simul.*, **31**, 195 (1989).
28. S. Champier and L. Grammont, *Inverse Probl.*, **18**, 79 (2002).
29. R. Tibshirani, *J. R. Stat. Soc., B: Stat. Methodol.*, **58**, 267 (1996).
30. D. L. Sahagian and A. A. Proussevitch, *J. Volcanol. Geotherm. Res.*, **84**, 173 (1998).
31. M. J. Wainwright, *IEEE Trans. Inf. Theory*, **55**, 2183 (2009).
32. S. Jiang, J. Liu, G. Zhang, Y. An, H. Meng, Y. Gao, K. Wang and J. Tian, *IEEE Trans. Biomed. Eng.*, **66**, 1361 (2018).
33. R. Muthukrishnan and R. Rohini, *2016 IEEE international conference on advances in computer applications (ICACA)*, (2016).
34. D. Wang, Y. Che, C. Li, Y. Chen, H. Yin and C. Zhang, *2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (2021).
35. Y. Shi and Y. Zhang, *Appl. Phys. A*, **92**, 621 (2008).
36. P. Bowen, *J. Dispersion Sci. Technol.*, **23**, 631 (2002).
37. R. Blödner, P. Mühlig and W. Nagel, *J. Microsc.*, **135**, 61 (1984).
38. D. Depriester and R. Kubler, *J. Struct. Geol.*, **151**, 104418 (2021).

Supporting Information

Reconstruction of particle size distribution from cross-sections

Jihoon Oh[*,‡], Dongjae Kim[,***,‡], Seunggeon Lee[*], and Jaewook Nam[*,****,†]**

*School of Chemical and Biological Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
**Department of Chemical Engineering, Soonchunhyang University, 22 Soonchunhyang-ro, Shinchang-myeon, Asan-si, Chungcheongnam-do 31538, Korea
***Department of Electronic Materials and Devices Engineering, Soonchunhyang University, 22 Soonchunhyang-ro, Shinchang-myeon, Asan-si, Chungcheongnam-do 31538, Korea
****Institute of Chemical Process, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
(*Received 10 March 2023 • Revised 23 June 2023 • Accepted 29 June 2023*)

S1. Additional Derivation of the Representative Diameter in Geometric Bin-size System

As the width of the k^{th} bin is Δ_k and the minimum value of the first bin is d_{min}, the upper limit of k^{th} bin, U_k, is expressed as

$$U_k = d_{min} + \sum_{k=1}^{j} \Delta_k. \tag{S1}$$

The representative diameter is $\alpha\Delta_j$ away from the upper limit. Therefore, D_j and d_i can be expressed as

$$D_j = d_{min} + \sum_{k=1}^{j} \Delta k - \alpha\Delta_j, \tag{10}$$

$$d_i = d_{min} + \sum_{k=1}^{i} \Delta k. \tag{11}$$

S2. Supplementary Figures & Tables

Table S1. The mean and standard deviation of optimized λ values for obtaining the smallest SSE in different numbers of bins and distributions. The experiments were conducted over 30 samples of each distribution. The mean and standard deviation of all distributions were 15 and 2.5, respectively. Here, m is the number of bins

	Normal	Bimodal	Lognormal	Rayleigh
m=15	0.228 (±0.256)	0.454 (±0.303)	1.225 (±0.613)	0.041 (±0.031)
m=25	0.995 (±0.389)	0.104 (±0.031)	1.126 (±0.351)	0.009 (±0.005)
m=35	0.589 (±0.249)	0.077 (±0.023)	0.804 (±0.369)	0.011 (±0.005)

Table S2. The mean and standard deviation of optimized λ values for obtaining the smallest SSE in different standard deviations of the lognormal distribution. Here, m is the number of bins, and s is the standard deviation of the lognormal distribution

	s=2.5	s=5	s=7.5	s=10
Lognormal (m=25)	1.126 (±0.351)	0.532 (±0.233)	0.135 (±0.054)	0.024 (±0.009)

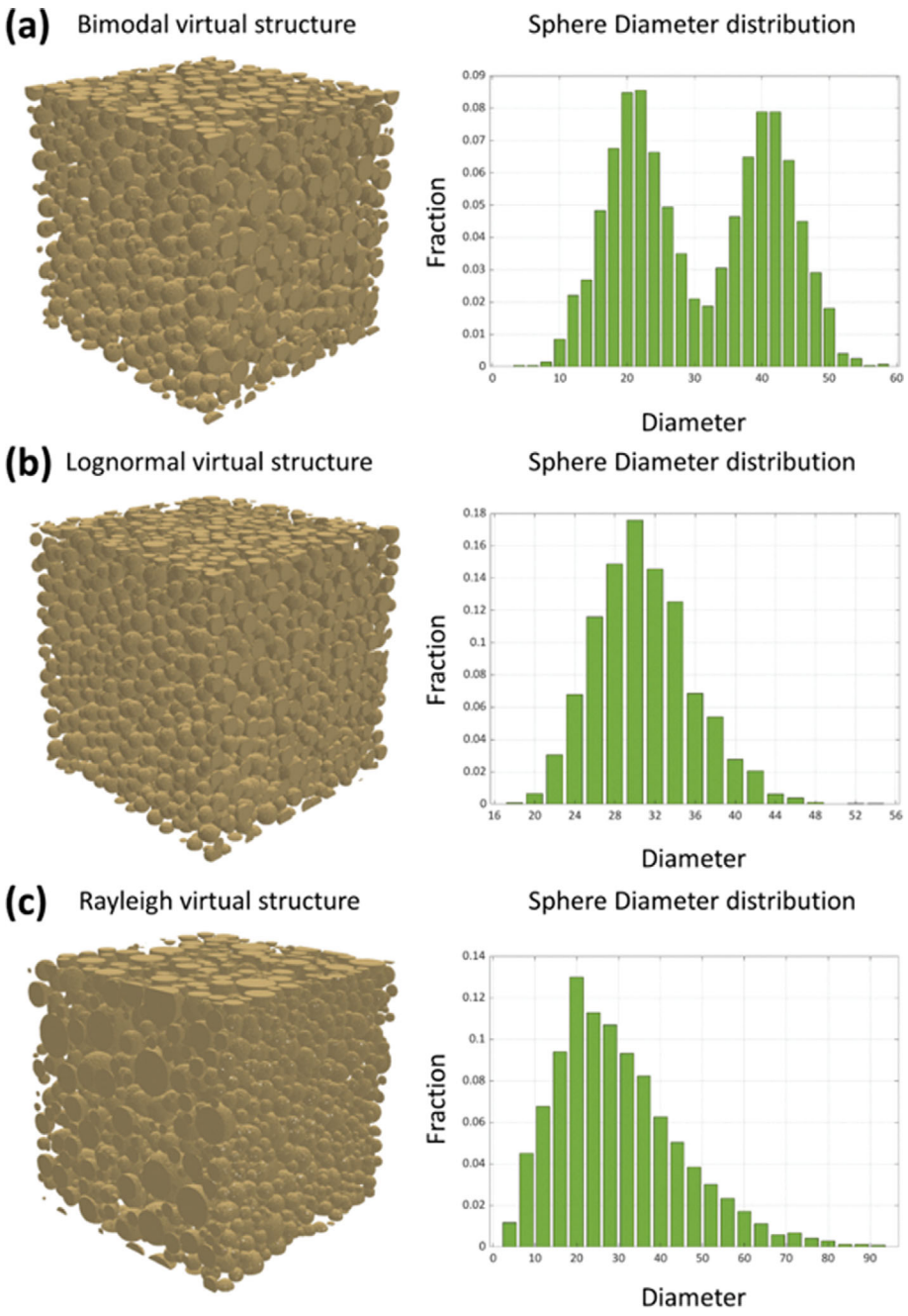

Fig. S1. Virtual structures with various ground-truth distribution. (a) bimodal distribution, (b) lognormal distribution, and (c) Rayleigh distribution.

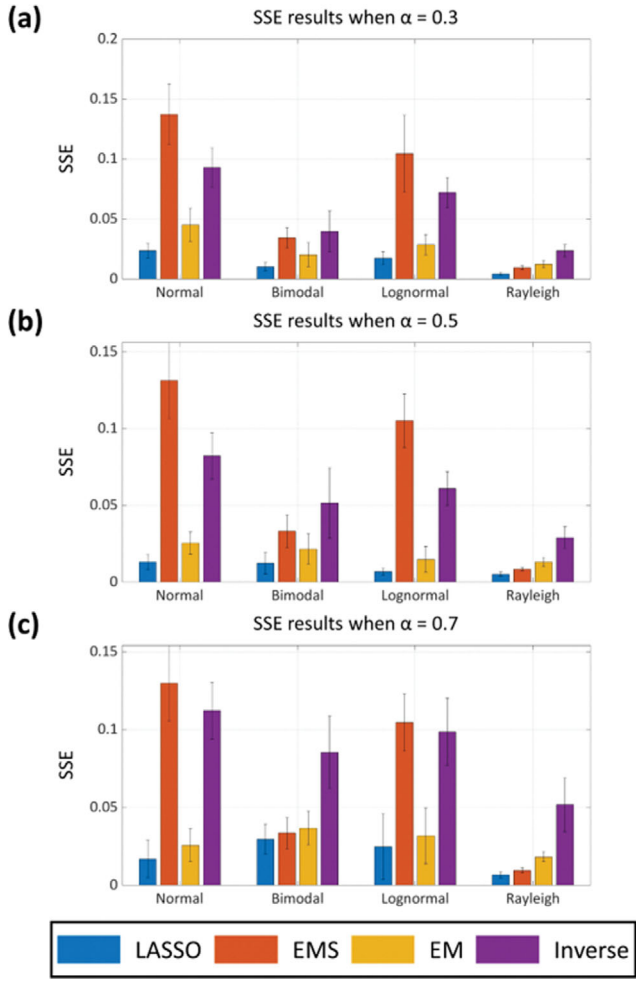

Fig. S2. SSE values according to α. (a) α=0.3, (b) α=0.5, and (c) α= 0.7.

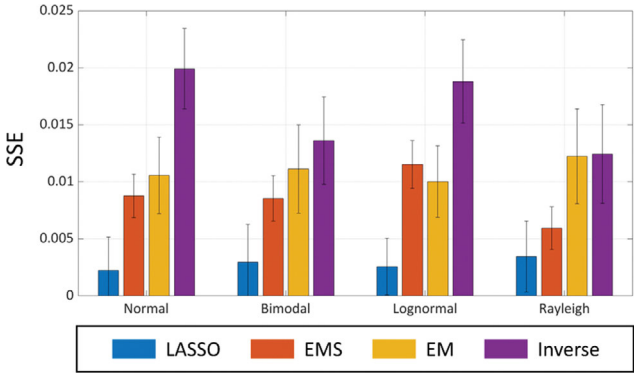

Fig. S3. SSE analysis results in uniform bin-size system.

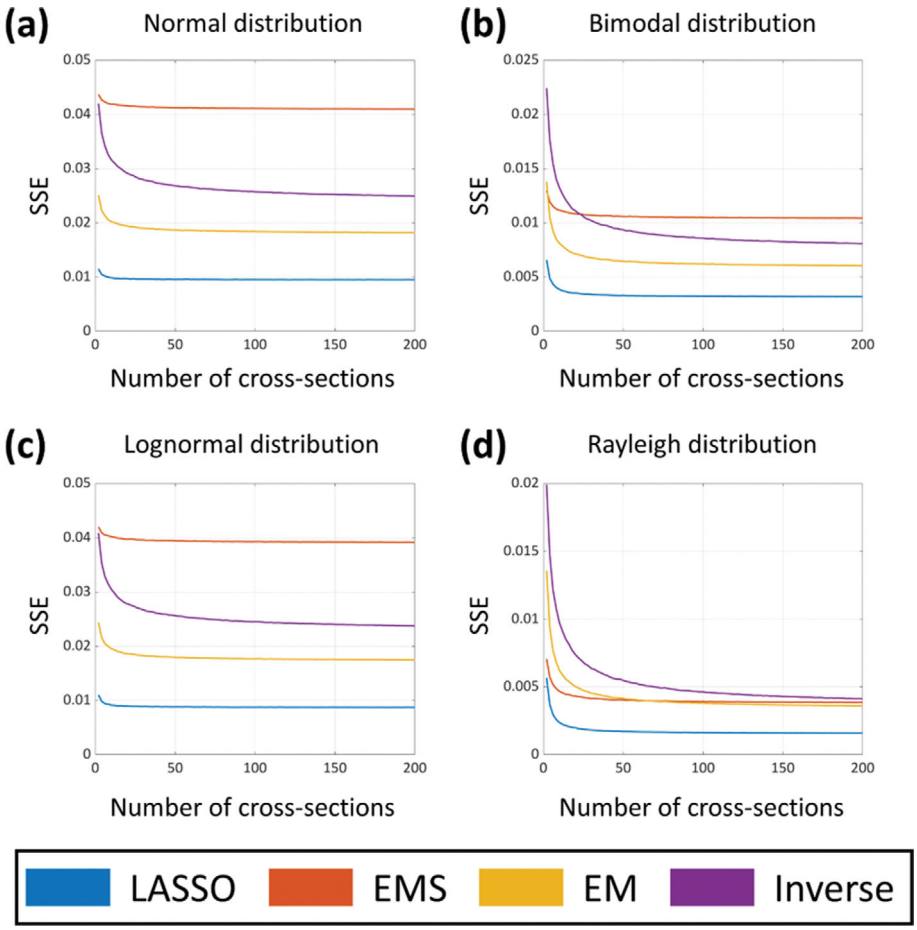

Fig. S4. SSE analysis results according to the number of cross-sections. (a) normal distribution, (b) bimodal distribution, (c) lognormal distribution, and (d) Rayleigh distribution.

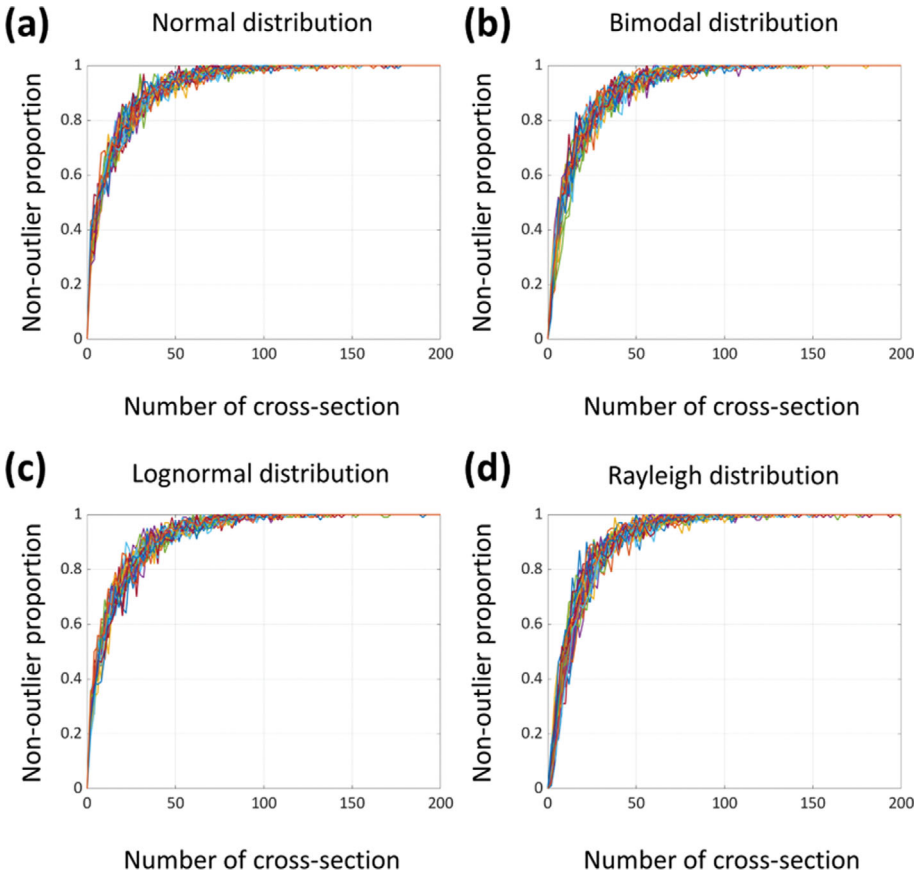

Fig. S5. Non-outlier proportion curves of each virtual structure. (a) normal distribution, (b) bimodal distribution, (c) lognormal distribution, and (d) Rayleigh distribution.

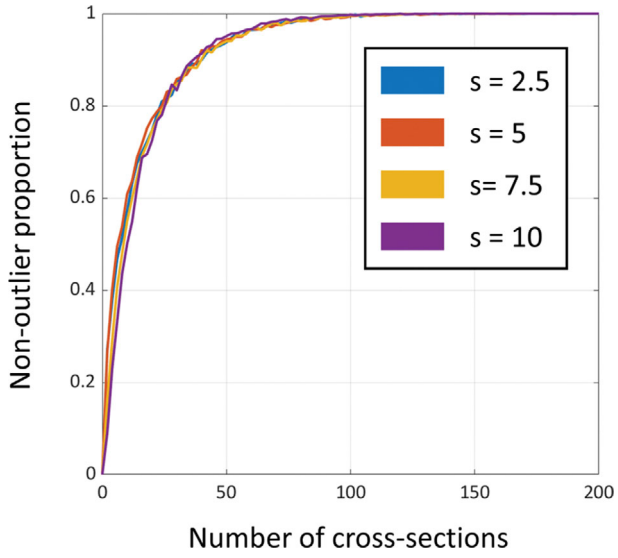

Fig. S6. Average non-outlier proportion curves according to the number of cross-sections in lognormal distribution.