Yuanyuan Gao
Department of Industrial Engineering and Management,
College of Engineering,
Peking University,
Beijing 100871, China
e-mail: yuanyuan.gao@pku.edu.cn

Xiaohu Huang
Department of Quantitative Investment,
DaCheng Fund,
Shenzhen 518000, China
e-mail: xiaohuhuang6@163.com

Chao Wang
Department of Industrial and Systems Engineering,
College of Engineering,
The University of Iowa,
Iowa City, IA 52242
e-mail: chao-wang-2@uiowa.edu

Jianguo Wu[1]
Department of Industrial Engineering and Management,
College of Engineering,
Peking University,
Beijing 100871, China
e-mail: j.wu@pku.edu.cn

Estimating Size and Number Density of Three-Dimensional Particles Using Truncated Cross-Sectional Data

The need for estimating three-dimensional (3D) information based on two-dimensional (2D) images has been increasing in numerous fields. It is essential in quality assessment, quality control, and process optimization. However, all the existing methods have not considered the data truncation issue, which is commonly faced in metrology. This paper proposes a new statistical approach to infer size distribution and volume number density (VND) of 3D particles based on 2D cross-sectional images with data truncation considered. In order to estimate the size distribution, a linkage is established between 3D particles and 2D observations with the existence of data truncation. Subsequently, this paper derives the likelihood function of 2D observations and an efficient Monte Carlo expectation-maximization algorithm is developed to estimate the parameters of size distribution. In addition, an explicit relationship between the 3D and 2D particle number densities is established and leveraged to estimate the VND and volume fraction. The effectiveness of the proposed method is demonstrated through both simulation study and real case studies in metal additive manufacturing and metal-matrix nanocomposites manufacturing.
[DOI: 10.1115/1.4051625]

Keywords: size inference, truncated data, Monte Carlo EM, metal additive manufacturing, metal-matrix nanocomposites

1 Introduction

Recent years have witnessed the tremendous increase in the need for measuring three-dimensional (3D) particles in the field of manufacturing, material science, biology, etc. [1–4]. One representative example is the measurement of pores in the metal additive manufacturing (AM). As shown in Figs. 1(*a*) and 1(*b*), the porous defect is in the form of void or simply a lack of solid material, which is one of the most critical quality issues in metal AM [7,8]. It is known that the pores can lead to crack nucleation and propagation, resulting in reduced material stiffness, bending strength, hardness, and fatigue life [9,10]. Consequently, inspecting the porosity is of remarkable significance for quality control and process optimization.

Another example in manufacturing is the measurement of the dispersion of reinforcing nanoparticles in nanocomposites manufacturing [5,11–15]. Working as heterogeneous nucleation agents, the well-dispersed nanoparticles could significantly reduce the grain size and enhance the mechanical properties. However, nanoparticles often agglomerate and form *clusters* in the fabrication process due to high surface energy, large surface-to-volume ratio, and poor wettability in liquid [1,13,16]. For example, Fig. 1(*d*) shows a microstructural image of Al_2O_3 reinforced metal-matrix nanocomposites (MMNCs) with two big clusters, which have undesirable impact on mechanical properties. Thus, it is necessary to measure the cluster size distribution and volume percentage accurately to assess the material quality and further to improve the fabrication process.

Although recent advancement of measurement technology, such as X-ray computed tomography (CT), makes it possible to directly measure the 3D size and volume percentage of voids in 3D printing, it is prohibitive for its high cost and low resolution [17]. Meanwhile, indirect measurements from two-dimensional (2D) microscopic images are still widely used in practical applications, such as scanning electron microscope (SEM) images [18–20]. These methods provide 2D cross sections that can be used to extract certain particle/void distribution information, e.g., whether there are any clustering effects [21] or boundary effects [22] in MMNCs. However, these 2D morphological information directly drawn from microscopic images are not always sufficient to evaluate the product quality. The 3D particle (cluster in MMNCs or void in 3D printing) size distribution and volume number density (VND) (number per unit volume), on the other hand, directly capture the particle information and thus are two critical quantities that need to be estimated in quality assessment. Though 2D particle size and number density are related to the 3D counterpart, the former cannot completely replace the latter. For example, the 2D number density (number per unit area) of clusters shown on SEM images depends on both 3D number density (number per unit volume) and cluster sizes. Larger VND or 3D cluster size tends to result in more clusters shown on SEM images. Therefore, 3D information is more informative, and inferring 3D size and number density based on the observed 2D information can greatly facilitate the quality assessment.

Inferring 3D information from 2D cross-sectional images has been intensively studied in material science, bioscience, and petrography [23]. To estimate the size distribution of spherical 3D particles, Liu et al. [24] first derived the probability density function (PDF) of 2D particle sizes with the assumption of spherical shape and complete spatial randomness (CSR), and then developed a maximum likelihood estimation (MLE) approach to estimate the distribution parameters. However, their approach ignored the sampling bias, i.e., larger particles are more likely to be intersected by SEM images. Consequently, the estimated pdf is for the intersected particles, rather than all the particles in the specimen. Without considering such sampling bias, the size would be overestimated. Later, Wu et al. [1] took into consideration the sampling bias and

[1]Corresponding author.
Manuscript received March 23, 2021; final manuscript received June 25, 2021; published online July 20, 2021. Assoc. Editor: Y. Lawrence Yao.

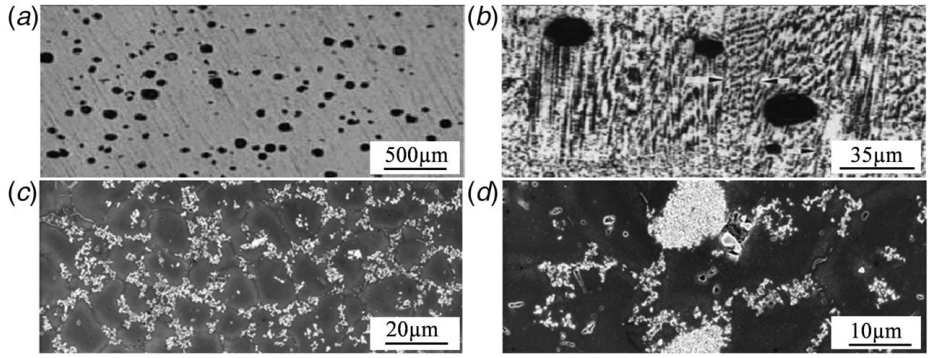

Fig. 1 ((*a*) and (*b*)) SEM images of metal-based AM products showing pores or voids [5], and A206–Al_2O_3 MMNCs (*c*) without and (*d*) with clusters [6]

developed two statistical approaches to estimate the size distributions, namely, the MLE and the method of moment. Their method is a correction of Liu's approach and thus is more accurate. To make the 3D geometric shape more general, Wu et al. [2] assumed that the 3D objects are ellipsoidal and then developed a series of statistical approaches to estimate the distribution parameters and VND of porous defects in metal AM based on cross-sectional images.

However, in almost all practical applications, only sufficiently large 2D particles are selected to estimate the 3D size distributions and number density, which inevitably leads to a data truncation issue. As illustrated in Fig. 2, suppose the particles are spherical and the 2D size (radius) threshold is l_0. Only those 2D particles (circular profiles) with radius larger than l_0 are selected (particle II) in 3D size distribution estimation. There are two causes that lead to data truncation issue. The first one is the resolution of SEM images, which is inherent in all applications. All 2D particles with radius r less than the microscope resolution l_0 will not be observed. The second cause is artificial, in that a threshold is intentionally predetermined in the selection of 2D circular particle profiles. For example, only the 3D nanoparticle clusters with radius larger than a certain threshold, say r_l, are believed to have a detrimental effect on mechanical properties of MMNCs [24]. Since the observed circular profiles with 2D size less than r_l may be from small 3D clusters of no interest ($< r_l$), or even from non-clustered nanoparticles, a predefined threshold l_0 satisfying $l_0 \geq r_l$ has to be set in the selection of 2D samples. Therefore, the data truncation issue has to be considered in order to get an accurate estimation of both 3D size distribution and number density.

To achieve a more accurate estimation, this paper develops a new statistical method to infer both the size distribution and number density of 3D particles by taking into account the data truncation issue, which is the main contribution of this paper. Specifically, the relationship between the pdf of the size of 3D particles and that of intersected ones is derived first conditioning on the existence of data truncation, and then the pdf of observed 3D sizes is derived. Subsequently, an MLE method based on the Monte Carlo expectation-maximization (MCEM) [25] algorithm is developed to estimate the distribution parameters. To infer the VND using the truncated data, it is proved that the observed 2D particles follow a non-homogeneous Poisson process, where the Poisson rate is a function of the 3D VND, 3D particle size, and the data truncation threshold.

The rest of this paper is organized as follows. Section 2 presents the detailed assumption and problem formulation. The linkage of size between 3D and 2D particles considering the data truncation, and the parameter estimation by Monte Carlo expectation-maximization (EM) algorithm are provided in Sec. 3. In Sec. 4, a statistical approach is developed to estimate the VND and volume fraction with truncated data. In Secs. 5 and 6, simulation studies and real case studies on metal AM and MMNCs are presented to evaluate the effectiveness of the proposed method. Section 7 is the conclusion and discussion.

2 Problem Formulation

In this section and Secs. 3 and 4, we formulate the problem and introduce the technical details in the context of MMNCs quality inspection. Before the SEM imaging and analysis, specimens are first cut and then the cutting surfaces are polished. Therefore, the SEM images can be modeled as cross sections of cutting planes with random locations and orientations intersecting the specimen. Figure 3 illustrates the SEM cutting process with/without severe

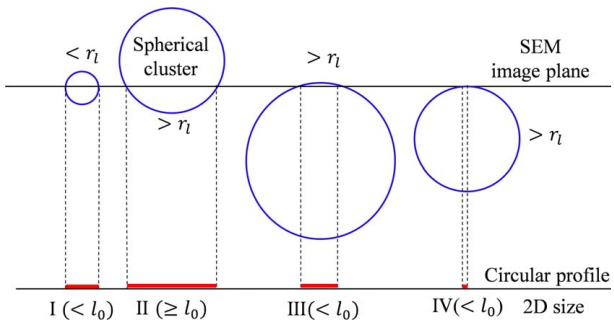

Fig. 2 Illustration of the data truncation issue in the selection of 2D particle samples. r_l is the lower bound of 3D particle radius; and l_0 is the threshold of 2D particle radius.

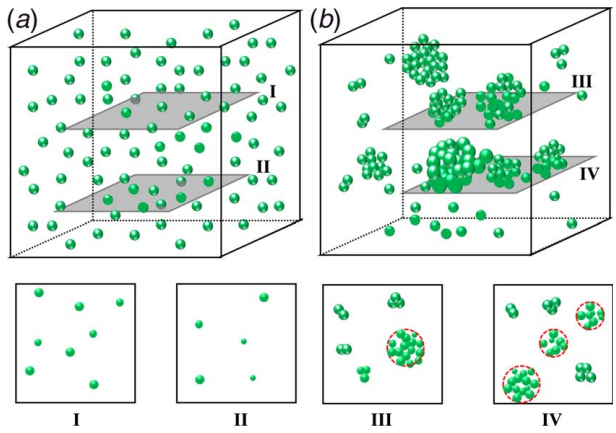

Fig. 3 Illustration of SEM images cutting specimens with and without severe clustering problem: (*a*) no clustering problem and (*b*) clustered. The 2D clusters of size larger than l_0 are marked by circular profiles.

clustering problem. Figure 3(*a*) shows an ideal case where all particles are uniformly distributed and thus no large 2D clusters are observed on the SEM images. Figure 3(*b*) illustrates a specimen with severe clustering issue. If a 3D cluster of size larger than r_l (size threshold to define a severe cluster) is intersected by the image plane and the size of the corresponding 2D profile is larger than l_0, then the 2D cluster will be selected as an effective observation for 3D inference. It is worth noting that in practical applications, the small particles shown on the SEM image may come either from particle clusters or from non-clustered particles. Therefore, in this application, we have to set a threshold l_0 for the 2D observations and it should satisfy $l_0 \geq r_l$; otherwise, the observed 2D clusters may come from 3D ones of size less than r_l. In 3D printing, the data truncation is due to image resolution. Suppose the image resolution is r_0, then $l_0 \geq r_0$. The lower bound of the void size is $r_l = 0$, so $l_0 \geq r_l$ still holds. To maximize the number of effective samples, we can set $l_0 = \max(r_l, r_0)$ in real applications.

To facilitate the inference of 3D quantitative information based on 2D cross-sectional images, we also adopt the two widely used assumptions. The first assumption is that all the formed clusters are spherical with radius following a certain distribution. The spherical shape is commonly used to characterize nanoparticles and clusters in material science [2,24,26,27]. The reason is that in the nano or microscale, the spherical particle/clusters have the smallest surface-to-volume ratio and are thermodynamically more stable [28]. The second assumption is that all the nanoparticle clusters are uniformly distributed in the metal matrix with CSR [16,21,27]. CSR has been widely used in the field of point pattern analysis [2,16]. It can be achieved in a homogeneous specimen. This assumption is the foundation for all 3D information inference problems based on 2D cross sections. It indicates that once a cluster is intersected by an image plane, the distance between the cutting plane and the center of that cluster follows a uniform distribution.

Let R, Z_c, and R_c be the radius of the 3D cluster, the distance between the SEM image plane and the center of the cluster, and the radius of the 2D circular profile, respectively, as illustrated in Fig. 4. It is straightforward that $R^2 = R_c^2 + Z_c^2$. Suppose there are n truncated observations $\boldsymbol{r_c} = \{r_{c1}, \ldots, r_{cn}\}$ with $r_{ci} \geq l_0$ for $i = 1, \ldots, n$. Then the problem is to estimate the pdf $f_R(r)$ and VND λ_0 based on the truncated observations. The technical details of estimating the size distribution and VND will be provided in Secs. 3 and 4, respectively.

3 Size Distribution Inference and Parameter Estimation

In this section, several commonly used parametric distributions are assumed for the size distribution, and an MCEM algorithm is developed to find the MLE of the distributional parameters. Without loss of generality, three distributions are considered, uniform distribution, truncated normal distribution with a lower bound, and truncated log-normal distribution with a lower bound. The lower bounds r_l of the distributions are assumed known or are predefined, and $l_0 \geq r_l$. Note that if the lower bound is unknown, it is easy to show that when $l_0 \geq r_l$, the lower bound is not estimable, due to the fact that truncated data with $r_c \geq l_0$ only provides information for 3D clusters with size $r \geq l_0$. In the following subsections, the pdf and likelihood functions are derived, and then an MCEM is developed for intractable likelihood functions. To make the method more general, we release the constraint of $l_0 \geq r_l$ in Secs. 3 and 4. Although in the case of MMNCs we have to keep the constraint, in other application contexts, the constraint may not be necessary when the 2D samples are all from the target distribution.

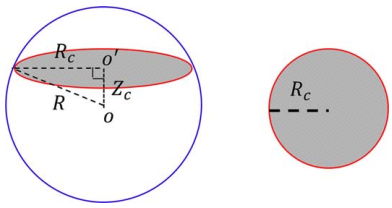

Fig. 4 The relationship among R, Z_c, and R_c

3.1 Derivation of the Probability Density Function and Likelihood Function. Suppose $\boldsymbol{\theta}$ is the vector of distribution parameters of 3D clusters and I is a binary variable describing the SEM cutting status with $I_i = 1$ representing that the cluster i is intersected by the image plane, and $I_i = 0$ otherwise. Based on Ref. [1], the SEM cutting process is a biased sampling process with size distribution of intersected 3D clusters, given by

$$f(r|\boldsymbol{\theta}, I=1) = \frac{f(r|\boldsymbol{\theta})P(I=1|r)}{P(I=1)} = \frac{f(r|\boldsymbol{\theta})(r/r_u)}{\int_{r_l}^{r_u}(r/r_u)f(r|\boldsymbol{\theta})dr} = \frac{rf(r|\boldsymbol{\theta})}{E(r|\boldsymbol{\theta})} \quad (1)$$

where r is the radius of 3D clusters, $f(r|\boldsymbol{\theta})$ is the assumed parametric distribution, $E(r|\theta)$ is the mean or expectation of the 3D cluster size, and r_u and r_l are the nonnegative upper and lower bounds of the cluster size, respectively. Since all clusters are uniformly distributed, the distribution of z_c can be intuitively obtained as

$$f(z_c|r, I=1) = \frac{1}{r}1_{[0,r]}(z_c) \quad (2)$$

where $1_{[0,r]}(z_c)$ is an indicator function, i.e., $1_{[0,r]}(z_c) = 1$ if $z_c \in [0, r]$ and 0 otherwise. Since $r_c = \sqrt{r^2 - z_c^2}$, the conditional pdf $f(r_c|r, I = 1)$ can be derived through variable transformation:

$$f(r_c|r, I=1) = \begin{cases} \dfrac{r_c}{r\sqrt{r^2 - r_c^2}}, & 0 \leq r_c \leq r \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Therefore, conditioning on the truncated data $r_c \geq l_0$, the distribution can be derived as

$$f(r_c|r, I=1, r_c \geq l_0) = \begin{cases} \dfrac{r_c}{r\sqrt{r^2 - r_c^2}}\dfrac{1}{P(r_c \geq l_0|r, I=1)}, & l_0 \leq r_c \leq r \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Based on Eq. (3), the probability $P(r_c \geq l_0|r, I = 1)$ can be derived as

$$P(r_c \geq l_0|r, I=1) = 1 - \int_0^{l_0} \frac{r_c}{r\sqrt{r^2 - r_c^2}}dr_c = \frac{\sqrt{r^2 - l_0^2}}{r} \quad (5)$$

Therefore,

$$f(r_c|r, I=1, r_c \geq l_0) = \begin{cases} \dfrac{r_c}{\sqrt{r^2 - r_c^2}}\dfrac{1}{\sqrt{r^2 - l_0^2}}, & l_0 \leq r_c \leq r \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The conditional complete-data likelihood can be derived as

$$f(r_c, r|I=1, r_c \geq l_0, \boldsymbol{\theta}) = \frac{f(r|I=1, \boldsymbol{\theta})P(r_c \geq l_0|I=1, r)}{P(r_c \geq l_0|I=1, \boldsymbol{\theta})}$$

$$f(r_c|r, I=1, r_c \geq l_0) = \frac{r_c f(r|\boldsymbol{\theta})}{g(l_0, \boldsymbol{\theta})\sqrt{r^2 - r_c^2}}1_{[l_0,r]}(r_c) \quad (7)$$

where

$$g(l_0, \boldsymbol{\theta}) = \int_{\max(l_0, r_l)}^{r_u} \sqrt{r^2 - l_0^2}f(r|\boldsymbol{\theta})dr \quad (8)$$

where r_u is the upper bound of the 3D size distribution. Note that if $l_0 = 0$, i.e., there is no data truncation, $g(l_0, \boldsymbol{\theta}) = E(r|\boldsymbol{\theta})$, Eq. (7) is exactly the one obtained in Ref. [1]. The marginal likelihood can

be calculated by integrating out r from Eq. (7):

$$f(r_c|I=1,\, r_c \geq l_0,\, \boldsymbol{\theta}) = \int_{\max(r_c, r_l)}^{r_u} \frac{r_c f(r|\boldsymbol{\theta})}{\sqrt{r^2 - r_c^2}\, g(l_0, \boldsymbol{\theta})} 1_{[l_0, r]}(r_c) dr \quad (9)$$

Given the observations $\boldsymbol{r_c} = \{r_{c1},\, r_{c2},\, \ldots,\, r_{cn}\}$, the likelihood function can be expressed by

$$L(\boldsymbol{\theta}|\boldsymbol{r_c}) = \prod_{i=1}^{n} f(r_{ci}|I_i = 1,\, r_{ci} \geq l_0,\, \boldsymbol{\theta}) \quad (10)$$

The MLE estimation of distribution parameters can thus be obtained through

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \{L(\boldsymbol{\theta}|\boldsymbol{r_c})\} \text{ or } \boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \{\log [L(\boldsymbol{\theta}|\boldsymbol{r_c})]\} \quad (11)$$

For the simple case where R follows uniform distribution $U[r_l, r_u]$, Eq. (9) can be analytically obtained as

$$f(r_c|I=1,\, r_c \geq l_0,\, \boldsymbol{\theta}) = \frac{r_c 1_{[l_0, r_u]}(r_c)}{(r_u - r_l) g_u(l_0, \boldsymbol{\theta})} \log \frac{\sqrt{r_u^2 - r_c^2} + r_u}{\sqrt{\max(r_c, r_l)^2 - r_c^2} + \max(r_c, r_l)} \quad (12)$$

where

$$g_u(l_0, \boldsymbol{\theta}) = \frac{1}{2(r_u - r_l)} \left[l_0^2 \log \frac{\sqrt{\max(l_0, r_l)^2 - l_0^2} + \max(l_0, r_l)}{\sqrt{r_u^2 - l_0^2} + r_u} + \left(r_u \sqrt{r_u^2 - l_0^2} - \max(l_0, r_l) \sqrt{\max(l_0, r_l)^2 - l_0^2} \right) \right] \quad (13)$$

From Eqs. (12) and (13), we can see that when $l_0 \geq r_l$, the likelihood function does not involve r_l. In other words, the likelihood function will not change when r_l varies in the interval $[0,\, l_0]$. It is also true for other parametric distributions, which can be seen from Eqs. (7) and (8). This is what we expected and mentioned earlier, since the truncated data with $r_c \geq l_0$ only provide information for 3D clusters of size $r \geq l_0$.

Since the likelihood function is tractable for uniform distribution, the distribution parameter, i.e., the upper bound, can be easily obtained through Eq. (10) using various optimization techniques. However, for other distributions, the likelihood function is not tractable analytically, due to the intractable integrals in Eqs. (8) and (9). In the following subsection, a Monte Carlo-based EM algorithm (MCEM) is developed for 3D size inference with truncated data.

3.2 Monte Carlo Expectation-Maximization Algorithm for Size Distribution Estimation. The EM algorithm is commonly used to find the maximum likelihood parameters of a statistical model when the marginal likelihood function is intractable [25]. Typically, there are incomplete data or latent variables, e.g., the 3D cluster size r in this paper, and the conditional expectation of the log-likelihood with both observed data and missing data is optimized iteratively. Specifically, there are two steps in the EM algorithm, the expectation step (E-step) and the maximization step (M-step). The E-step is to calculate the expectation of the complete log-likelihood with respect to the predictive density of the incomplete data given the observed data and the estimated distribution parameters at the current step. At the M-step, the distribution parameters are updated by maximizing the expectation of the log-likelihood obtained at the E-step. Suppose the incomplete data are $\boldsymbol{r} = \{r_1, \ldots, r_n\}$, which are the corresponding 3D cluster sizes of the observed 2D clusters. The EM algorithm at the kth iteration can be expressed as

$$\text{E-step: } Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) = E_{\boldsymbol{R}|\boldsymbol{R}_c, \hat{\boldsymbol{\theta}}^{(k)}} (\log L(\boldsymbol{\theta}|\boldsymbol{r},\, \boldsymbol{r_c},\, r_{ci} \geq l_0,\ \ i = 1, \ldots, n)) \quad (14)$$

$$\text{M-step: } \hat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \{Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)})\} \quad (15)$$

where $Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)})$ is the conditional expectation of the log-likelihood and is often called Q-function and $\hat{\boldsymbol{\theta}}^{(k)}$ is the vector of the estimated distribution parameters at iteration k. Based on Eq. (7), the complete log-likelihood can be expressed by

$$\log\ L(\boldsymbol{\theta}|\boldsymbol{r},\, \boldsymbol{r_c},\, r_{ci} \geq l_0,\ \ i = 1, \ldots, n) = \sum_{i=1}^{n} \log \left(\frac{r_{ci} f(r_i|\boldsymbol{\theta})}{g(l_0, \boldsymbol{\theta}) \sqrt{r_i^2 - r_{ci}^2}} \right) \quad (16)$$

The conditional predictive density function is simply

$$f(r_i|r_{ci}, \hat{\boldsymbol{\theta}}^{(k)}) \propto f(r_i,\, r_{ci}|\hat{\boldsymbol{\theta}}^{(k)},\, I_i = 1) = \frac{r_{ci} f(r_i|\hat{\boldsymbol{\theta}}^{(k)})}{E(r_i|\hat{\boldsymbol{\theta}}^{(k)}) \sqrt{r_i^2 - r_{ci}^2}} 1_{[r_{ci},\, r_u]}(r_i) \quad (17)$$

Therefore, the Q-function can be written as

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} \int_{\max(r_{ci}, r_l)}^{r_u^{(k)}} f(r_i|r_{ci}, \hat{\boldsymbol{\theta}}^{(k)}) [\log f(r_i|\boldsymbol{\theta}) - \log g(l_0, \boldsymbol{\theta})] dr_i + C \quad (18)$$

where C is a constant that is not related to $\boldsymbol{\theta}$.

Clearly, the Q-function is not tractable for most of parametric distributions due to intractable integrals in Eq. (18). A natural way to overcome this challenge is the Monte Carlo sampling approach [19], where the Q-function can be approximated through random samples drawn from the conditional predictive density function $f(r_i|r_{ci}, \hat{\boldsymbol{\theta}}^{(k)})$, $i = 1, \ldots, n$. However, the direct sampling from the conditional predictive density function is very difficult, if not impossible. Instead, the importance sampling can be used to draw random samples. A simple importance function can be selected as $f(r_i|\hat{\boldsymbol{\theta}}^{(k)})$, and the weight function is simply $1_{[r_{ci},\, r_u^{(k)}]}(r_i)/\sqrt{r_i^2 - r_{ci}^2}$. However, using this importance may result in many samples with zero weight, e.g., $r_i < r_{ci}$, which will significantly reduce the sampling efficiency. To overcome this problem, a truncated normal density function $f(x;\, \mu_i,\, \sigma_i^2,\, [r_{ci},\, \infty))$ can be used instead as the importance function. Suppose N samples are drawn for each observed particle, denoted by r_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, N$. Then the unnormalized weight w_{ij} and normalized weight W_{ij} of each sample can be calculated by

$$w_{ij} = \frac{1_{[r_{ci},\, r_u^{(k)}]}(r_{ij})}{f(r_{ij};\, \mu,\, \sigma^2,\, [r_{ci},\, \infty)) \sqrt{r_{ij}^2 - r_{ci}^2}}$$
$$W_{ij} = \frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}} \quad (19)$$

With these weighted samples, the Q-function can be approximated by

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) \approx \left[\sum_{i=1}^{n} \sum_{j=1}^{N} W_{ij} \log\ f(r_{ij}|\boldsymbol{\theta}) \right] - n \log\ g(l_0, \boldsymbol{\theta}) + C \quad (20)$$

If the 3D cluster size follows a truncated normal distribution with mean μ, variance σ^2, and a known lower bound r_l, then the

approximated Q-function can be computed as

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) \approx -\frac{\sum_{i=1}^{n}\sum_{j=1}^{N} W_{ij} r_{ij}^2 - 2\mu \sum_{i=1}^{n}\sum_{j=1}^{N} W_{ij} r_{ij}}{2\sigma^2} - n\left[\log\ \Delta\sigma + \frac{\mu^2}{2\sigma^2} + \log\ g(l_0, \boldsymbol{\theta})\right] + C \tag{21}$$

where

$$\Delta = 1 - \Phi\left(\frac{r_l - \mu}{\sigma}\right) \tag{22}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. If r follows a truncated log-normal distribution, i.e., $r \sim \ln N(\mu, \sigma^2 | r \geq r_l)$, then the approximated Q-function can be computed by

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)}) \approx -\frac{\sum_{i=1}^{n}\sum_{j=1}^{N} W_{ij}(\log\ r_{i,j})^2 - 2\mu \sum_{i=1}^{n}\sum_{j=1}^{N} W_{ij} \log\ r_{i,j}}{2\sigma^2} - n\left(\log\ P + \frac{1}{2}\log\ \sigma^2 + \frac{\mu^2}{2\sigma^2} + \log\ g(l_0, \boldsymbol{\theta})\right) + C \tag{23}$$

where

$$P = 1 - \Phi\left(\frac{\log\ r_l - \mu}{\sigma}\right) \tag{24}$$

At the M-step, the Q-function needs to be maximized to find the optimal value $\hat{\boldsymbol{\theta}}^{(k+1)}$. However, for both Eqs. (21) and (23), the analytical form of the partial derivatives $\partial Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)})/\partial \boldsymbol{\theta}$ does not exist, due to the intractable integration term $\log\ g(l_0, \boldsymbol{\theta})$. Therefore, the Q-functions cannot be optimized by simply solving the nonlinear equations $\partial Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(k)})/\partial \boldsymbol{\theta} = 0$. To address this issue, we discretize the feasible region with equal space and then evaluate the Q-function at each grid point to find the optimal value. This approach is very effective as the dimension of $\boldsymbol{\theta}$ is very low. In this method, the fixed parts of the Q-functions, i.e., the second term of Eqs. (21) and (23), do not change along iterations. Therefore, they only need to be calculated once, which can be done before the EM algorithm starts. Since $\log\ g(l_0, \boldsymbol{\theta})$ cannot be computed analytically, the Monte Carlo simulation is used instead to get approximated values.

The MCEM algorithm will be terminated once the estimated distribution parameters converge. The following commonly used stopping criterion is chosen:

$$\max_i \left|\frac{\hat{\theta}_i^{(k)} - \hat{\theta}_i^{(k-1)}}{\hat{\theta}_i^{(k-1)}}\right| < \epsilon \tag{25}$$

where ϵ is a small positive constant. Since the MCEM algorithm involves randomness caused by Monte Carlo simulation, we use a more conservative stopping criterion, i.e., the inequality (25) is satisfied for three consecutive iterations.

4 Volume Number Density and Volume Fraction Estimation

VND is an important quality parameter, which can be further utilized to estimate the volume fraction. The problem of VND estimation based on 2D observations without any data truncation issue has been well studied [2,16]. However, how to estimate the VND based on truncated data has not been studied. When there is no data truncation, it has been shown that the number of 2D observations on a cutting image can be modeled as a non-homogeneous Poisson process. When there exists data truncation, similar result can be derived, which is given in Proposition 1.

Proposition 1. *Under the spherical and CSR assumptions stated in Sec. 2, the number of clusters n shown on a cutting image of area A and satisfying $r_c \geq l_0$ can be modeled as a non-homogeneous Poisson process with density $\lambda(t)$*

$$\lambda(t) = 2A\lambda_{3D}\int_{\max\left\{\sqrt{t^2+l_0^2}, r_l\right\}}^{r_u} f(r|\boldsymbol{\theta})dr \tag{26}$$

where $t \leq \sqrt{r_u^2 - l_0^2}$ is the distance from the center of the spherical particle to the image plane, and λ_{3D} is the VND parameter. The total number of 2D observations follows a Poisson distribution with Poisson rate:

$$\lambda_{2D} = 2A\lambda_{3D}\int_{\max\{r_l, l_0\}}^{r_u} \sqrt{r^2 - l_0^2} f(r|\boldsymbol{\theta})dr \tag{27}$$

The proof of Proposition 1 is given in Appendix A. We can see that if $l_0 = 0$, or equivalently if there is no data truncation, Eq. (27) is simply $\lambda_{2D} = 2A\lambda_{3D}E(r|\boldsymbol{\theta})$, which is exactly the same result as in Ref. [16]. Note that Eq. (27) can actually be expressed as

$$\lambda_{2D} = 2A\lambda_{3D}E\left(\sqrt{(r^2 - l_0^2)^+}|\boldsymbol{\theta}\right) \tag{28}$$

where

$$(r^2 - l_0^2)^+ = \begin{cases} r^2 - l_0^2, & r \geq l_0 \\ 0, & r < l_0 \end{cases}$$

Since n is an unbiased estimator of λ_{2D}, i.e., $E(n|\lambda_{2D}) = \lambda_{2D}$, the VND λ_{3D} can be unbiasedly estimated given the distribution parameter $\boldsymbol{\theta}$:

$$\hat{\lambda}_{3D} = \frac{n}{2AE\left(\sqrt{(r^2 - l_0^2)^+}|\boldsymbol{\theta}\right)} \tag{29}$$

In practice, the true distribution parameters are unknown, therefore the estimated value $\hat{\boldsymbol{\theta}}$ using the proposed approach in Sec. 3 can be used. The estimation variance of the normalized parameter $\hat{\lambda}_{3D}/\lambda_{3D}$ equals to its mean squared error here, and thus can be used to characterize the estimation accuracy. It can be obtained as

$$\mathrm{Var}\left(\frac{\hat{\lambda}_{3D}}{\lambda_{3D}}\right) = \frac{\lambda_{2D}}{\left[2AE\left(\sqrt{(r^2 - l_0^2)^+}|\boldsymbol{\theta}\right)\right]^2}\frac{1}{\lambda_{3D}^2} = \frac{1}{\lambda_{2D}} \tag{30}$$

As we can see, λ_{2D} plays an important role on the estimation accuracy. The larger the λ_{2D} is, the higher the estimation accuracy will be. Increasing the image size A or reducing the truncation threshold l_0 can improve the estimation.

Another important quality parameter is the volume fraction, which captures the percentage of nanoparticles clustered or porosity in 3D printing. Let ρ denote the volume fraction. Without data truncation, it has been shown that the mean ρ can be unbiasedly estimated by the area fraction of circular intersections, i.e., $E(\rho) = E(A_c/A)$, where A_c is the total area of intersections [2]. However, when the data truncation issue exists, the area fraction is not measurable. Instead, we can use an alternative approach, which is described in Proposition 2.

Proposition 2. *Under the spherical and CSR assumptions stated in Sec. 2, the mean and variance of the volume fraction can be calculated as*

$$\begin{aligned} E(\rho) &= \frac{4\pi\lambda_{3D}}{3}E(r^3|\boldsymbol{\theta}) \approx \frac{4\pi\hat{\lambda}_{3D}}{3}E(r^3|\hat{\boldsymbol{\theta}}) \\ \mathrm{Var}(\rho) &= \frac{16\pi^2\lambda_{3D}}{9V}E(r^6|\boldsymbol{\theta}) \approx \frac{16\pi^2\hat{\lambda}_{3D}}{9V}E(r^6|\hat{\boldsymbol{\theta}}) \end{aligned} \tag{31}$$

where V is the volume of the specimen.

The proof of Proposition 2 is provided in Appendix B. As we can see from Eq. (31), when $V \to \infty$, $\mathrm{Var}(\rho) \to 0$. In practical applications, $V \gg \lambda_{3D} E(r^6|\boldsymbol{\theta})$, therefore, the volume fraction can be treated as a deterministic parameter.

5 Numerical Studies

In this section, Monte Carlo-based simulation is conducted to verify the effectiveness of the proposed approach. We investigate the size parameter estimation of the uniform, truncated normal, and log-normal distribution in Sec. 5.1 and the joint estimation of the size distribution, particle number density, and volume fraction in Sec. 5.2.

5.1 Estimation of Size Distribution Parameters. The truncated 2D observations are simulated using the following procedure: (1) generate 1×10^5 3D particles from the parametric distribution, which is sufficient enough for this simulation; (2) the weight of each 3D sample is assigned with its size; (3) the samples being cut are resampled according to their weights to obtain n intersected 3D samples (importance sampling for Eq. (1)); (4) the sizes of the 2D cross sections of the corresponding 3D samples are randomly generated based on the uniform distribution given in Eq. (2) and the geometric relationship $R^2 = R_c^2 + Z_c^2$; and (5) the intersected particles with 2D size smaller than the threshold l_0 are ignored and the remaining 2D samples are used for 3D inference.

Three types of size distributions are considered here, including uniform, truncated normal, and log-normal distribution (a special truncated log-normal with $r_l = 0$). The parameters for each distribution are given in Table 1. For the MCEM algorithm, the Monte Carlo sample size in the E-step is set as $N = 1000$ and the convergence stopping threshold is set as $\delta = 0.005$. To speed up the convergence of iteration, the initial values are obtained by directly fitting the distribution using the 2D samples. The truncation threshold is set to $l_0 = 3$.

To illustrate the spatial distribution, in Fig. 5 we provide three representative images with 200 circles simulated from the three size distributions. Note that in the simulation of cross-sectional images, we may get some overlapped circles. As both the number density and size are small, overlapping is not common. We just ignore this phenomenon and directly use the sizes and total number of circles obtained from simulation instead of extraction from images for 3D inference.

Table 1 Model parameters for three types of distributions

Type of distribution	r_l	r_u	μ	σ
Uniform	2	10	–	–
Truncated normal	0	∞	6	2
Log-normal	0	∞	2	0.5

We compare the proposed method with Wu's method [1], which ignored the data truncation issue. Figure 6 shows the estimated pdfs, histograms of the generated 3D samples, the intersected 3D samples, and the truncated 2D samples on the images. It can be obviously seen that the proposed method outperforms Wu's method remarkably on the estimation accuracy for truncated normal distribution and log-normal distribution. As expected, Wu's method overestimates the size distributions. The reason is that the truncated data tend to be large than the un-truncated one and Wu's method fails to consider such truncation. While for the uniform distribution, the upper bound estimated by both methods has little difference. In fact, the estimated upper bound for both methods is very close to $\max\{r_{c1}, \ldots, r_{cn}\}$. Wu et al. [1] provided a thorough discussion about using $\max\{r_{c1}, \ldots, r_{cn}\}$ to estimate r_u. Note that if r_l is unknown and $l_0 \le r_l$ (to make sure the lower bound is estimable), the proposed method will produce a much better estimation for the lower bound.

To evaluate the proposed MLE approach under different truncation thresholds and sample sizes, we conduct numerical experiments with $l_0 = 3$ and 5, the number of 2D observations before truncation $n_b = 100, 200, \ldots, 1000$ for the three distributions. Each simulation is repeated $S = 50$ times to reduce the randomness. Figure 7 shows the estimation accuracy for three distributions, which is measured by the relative standard error (RSE) of a parameter c given by $\mathrm{RSE} = \sqrt{1/S \sum_{i=1}^{S} ((\hat{c}_i - c)/c)^2}$.

As expected, the estimation accuracy for $l_0 = 3$ is higher than that of $l_0 = 5$, especially for the mean parameters. The reason is that larger truncation threshold leads to more information loss, e.g., less observations. This phenomenon is more obvious for the log-normal distribution. The reason is that a much larger proportion of particles are neglected as the threshold increases from 3 to 5 for the log-normal distribution. In general, the estimation accuracy decreases as the threshold l_0 increases.

5.2 Joint Estimation of Size Distribution, Volume Number Density, and Volume Fraction. In this section, the joint estimation of the size distribution, VND, and volume fraction of particles are presented, where the log-normal distribution is chosen to model the size distribution. We assume that the number of particles follow a Poisson distribution with density parameter $\lambda_0 = 14.99$, and the Poisson parameter of the intersected particles is $\lambda = 0.4A$, where A is the image size. In order to evaluate how image size A and truncated threshold l_0 affect the accuracy of estimation, we considered $A = 600{,}900\ \mathrm{mm}^2$ and $l_0 = 4.2 \times 10^{-3}$, 7.0×10^{-3} mm, respectively, in our numerical study and compare the proposed method with Wu's method without considering the data truncation [1]. For each group of simulation, we run the simulation for 20 times to reduce the randomness. The Monte Carlo EM algorithm parameters are the same as in Sec. 5.1.

Table 2 shows the mean values of the estimated parameters, the relative errors (in parentheses), and the true values (in the bottom

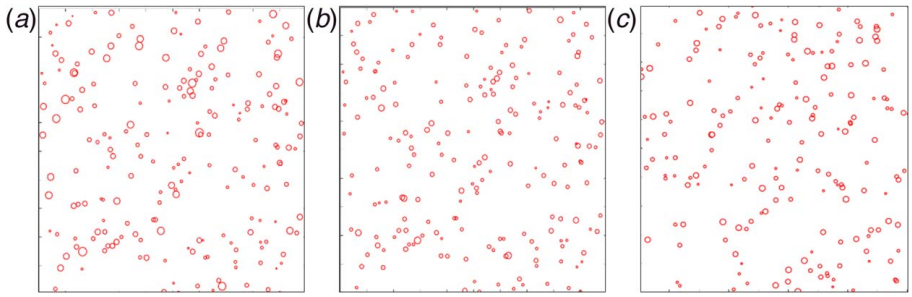

Fig. 5 The simulated 2D observations for three types of size distributions after truncation: (*a*) uniform, (*b*) truncated normal, and (*c*) log-normal distribution. The image size is 1000 × 1000.

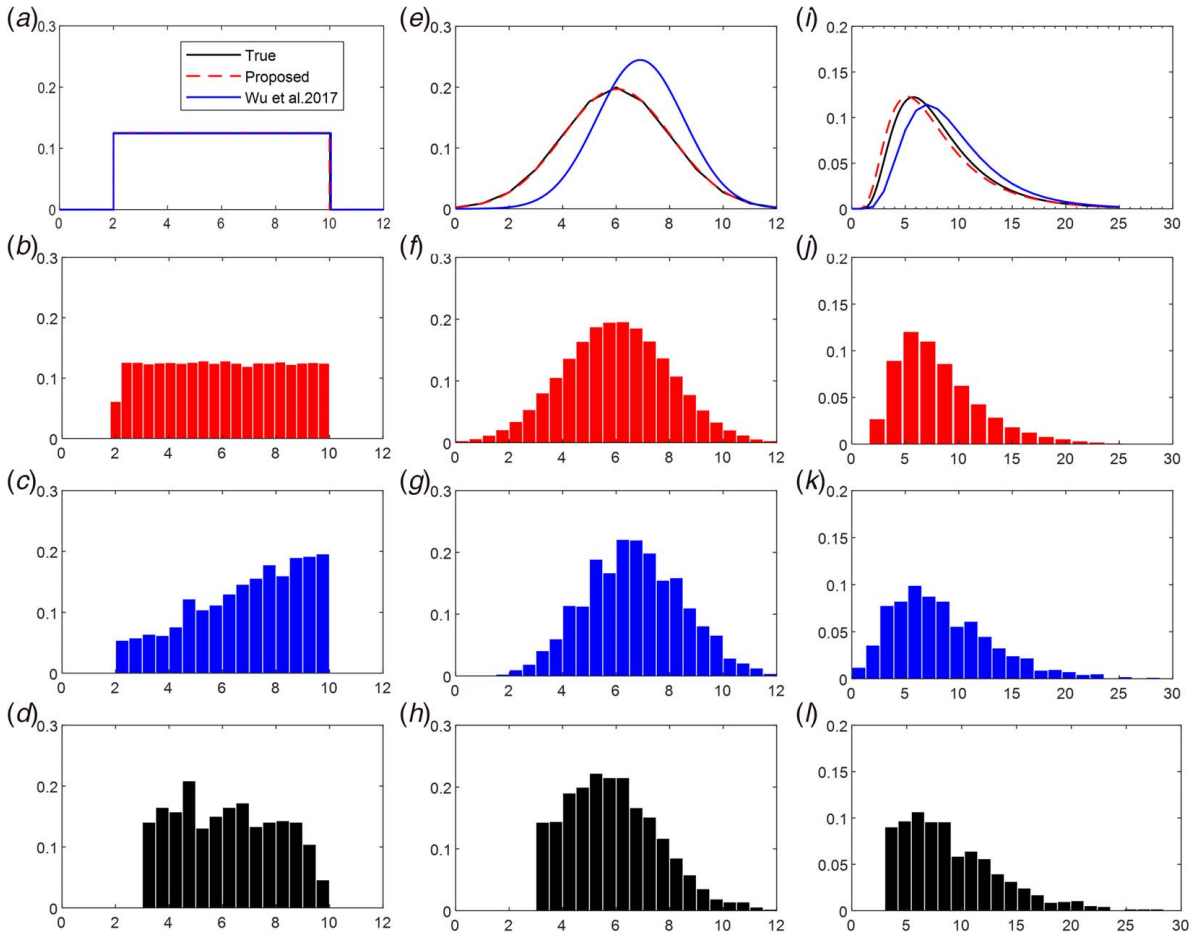

Fig. 6 Comparison of the proposed method with Wu et al. [1] for (left column: (*a*)–(*d*)) uniform distribution, (middle column: (*e*)–(*h*)) truncated normal distribution, and (right column: (*i*)–(*l*)) log-normal distribution. The first row ((*a*), (e), and (*i*)) shows the true and estimated pdfs; the second row ((*b*), (*f*), and (*j*)) shows the normalized histograms of the generated 3D samples from the true distribution; the third row ((*c*), (*g*), and (*k*)) shows the normalized histograms of the 3D samples cut by the microscopic images; the bottom row shows the normalized histograms of the 2D cross sections with threshold $l_0 = 3$. The sample size for the bottom two rows is 200.

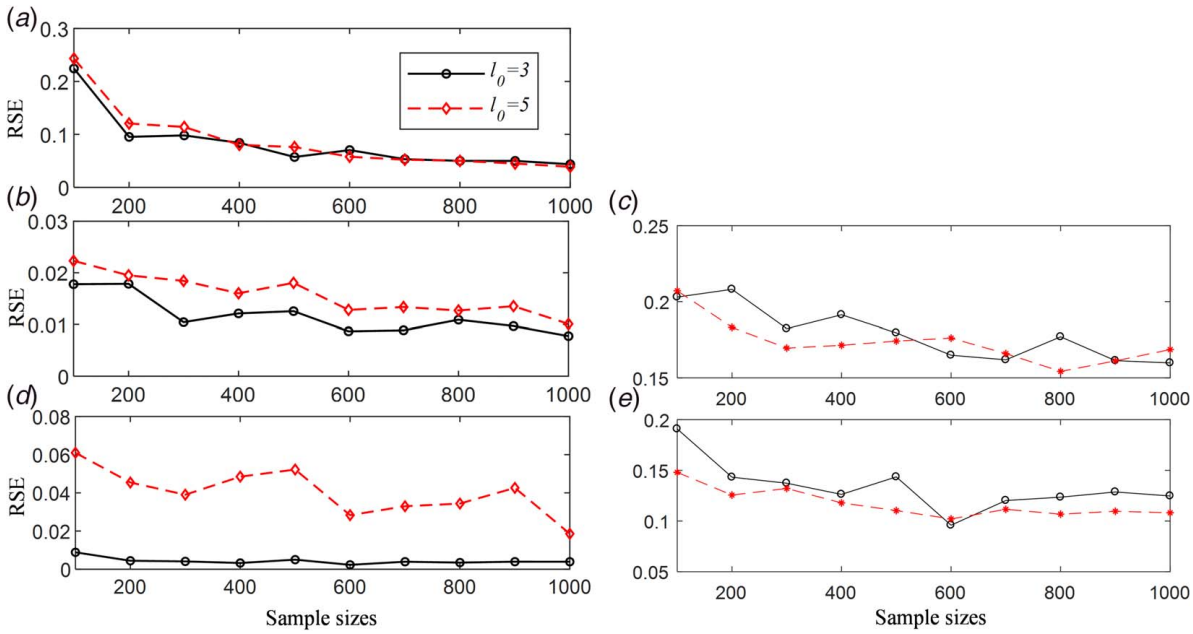

Fig. 7 The RSE of parameters estimation for three distributions: (*a*) RSE of the upper bound r_u estimation error; (*b*) and (*c*) RSE of the mean and variance parameter σ for the truncated normal; and (*d*) and (*e*) RSE of the mean and variance parameter σ for the log-normal distribution

Table 2 Comparison of the proposed method with Wu's method in the size distribution, VND, and volume fraction estimation

		The proposed method				Wu's method [1]			
A	l_0	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\rho}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\rho}$
600	4.2×10^{-3}	2.039 (1.95%)	0.4920 (1.6%)	12.28 (18.1%)	0.0235 (7.80%)	2.175 (8.75%)	0.420 (16%)	11.63 (22.5%)	0.0258 (18.3%)
	7.0×10^{-3}	1.900 (5%)	0.5320 (6.4%)	11.67 (22.2%)	0.0185 (15.1%)	2.357 (17.9%)	0.364 (27.2%)	9.784 (34.8%)	0.0376 (72.5%)
900	4.2×10^{-3}	2.028 (1.4%)	0.4750 (5%)	12.84 (14.4%)	0.0218 (0%)	2.168 (8.4%)	0.431 (13.8%)	12.11 (19.3%)	0.0282 (29.4%)
	7.0×10^{-3}	1.9610 (1.95%)	0.5230 (4.6%)	11.17 (25.5%)	0.0211 (3.2%)	2.346 (17.3%)	0.361 (27.8%)	9.784 (34.8%)	0.0352 (61.5%)
True value		2	0.5	15	0.0218	2	0.5	15	0.0218

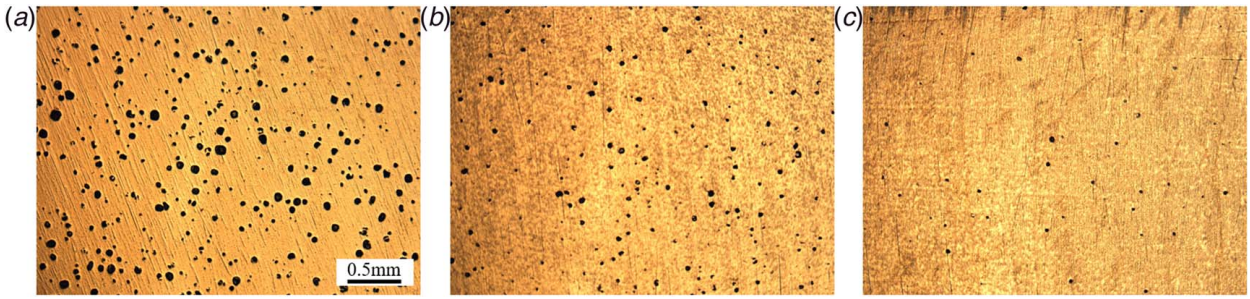

Fig. 8 The cross-sectional images of SLM-produced Ti–6Al–4V specimens with laser power 120 W and scan speed: (*a*) 360 mm/s, (*b*) 480 mm/s, and (c) 600 mm/s

row). Obviously, the estimation accuracy of all the parameters increases with the image size rising for both the proposed method and Wu's model which does not consider the data truncation. It is expected since larger images contain more particles. Second, increasing the threshold l_0 will cause more particles truncated, which leads to lower estimation accuracy. Comparing these two methods, we can see that in Wu's method, the mean parameter μ is overestimated and the VND λ is slightly underestimated, which leads to overestimation of the volume fraction ρ. In contrast, by considering the data truncation issue, the proposed approach has effectively mitigated the data truncation effects and achieved a much higher estimation accuracy.

6 Real Case Studies

6.1 Porosity Assessment in Additive Manufacturing. In this section, we apply the proposed method to estimate the size distribution, VND, and porosity in the metal AM. Metal AM, such as selective laser melting (SLM) and electron beam melting, utilizes high energy laser beam or electron beam as heating sources in the fabrication process. Porosity is one of the most severe quality issues, and it is essential to evaluate the severity of the porosity for quality inspection and process optimization. The pore formation mechanisms have been intensively studied, and all these studies showed that the global energy density is the most important factor affecting the porosity [29–31]. Excessive energy density, i.e., $E > E_{opt}$ where E_{opt} is the optimal energy density, causes vaporization of the material, which leads to spherical or near-spherical pores due to gas bubble formation. At the region $E < E_{opt}$, the porosity is incurred by incomplete melting of the raw powders due to the lack of fusion of powder particles, resulting in complex and elongated pores. In this case, we only consider pores with spherical shapes under the condition of $E > E_{opt}$.

6.1.1 Experiments Setup. The SLM samples are fabricated using Ti–6Al–4V powder with laser power 120 W and scan speed of 360, 480, and 600 mm/s, respectively [7]. For each specimen, cross-sectional images are obtained at different locations with the size 3.6 mm × 2.7 mm (1920 pixel × 2560 pixel, the area $A = 9.72\ \text{mm}^2$). Figure 8 shows one representative microscopic image for each specimen. It is clear that the porosity decreases when the scan speed increases, or consequently the global energy density E

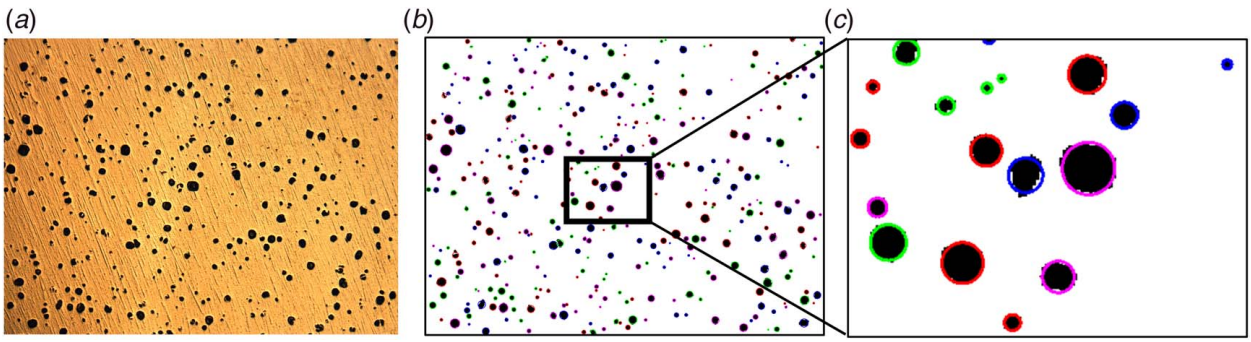

Fig. 9 (*a*) Raw image for 120 W and 360 mm/s, (*b*) extract boundaries and fitted circles, and (c) zoomed voids

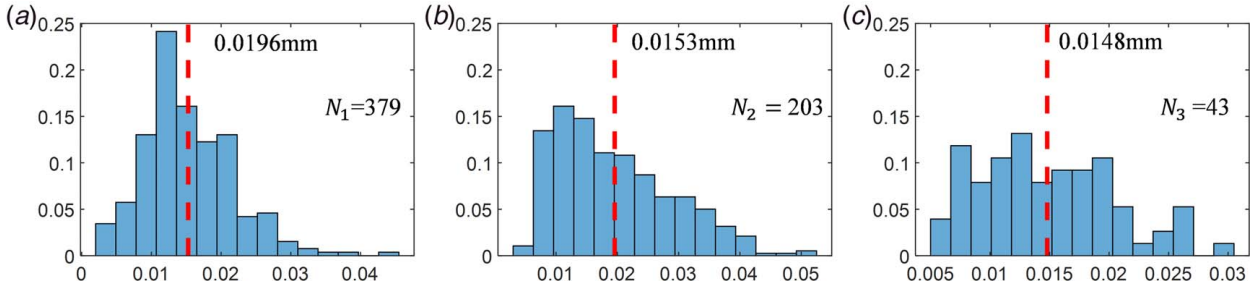

Fig. 10 The histograms of radius for three specimens: (*a*) 360 mm/s, (*b*) 480 mm/s, and (*c*) 600 mm/s

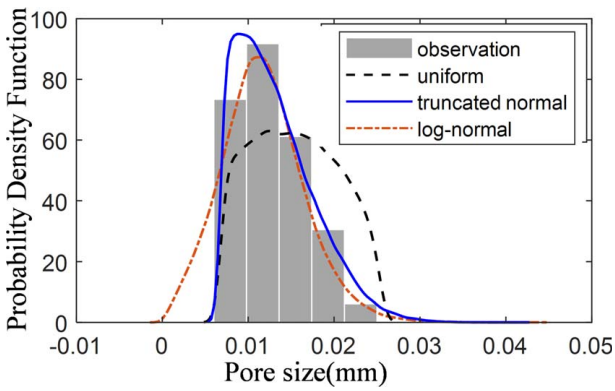

Fig. 11 Pdfs of the simulated 2D sizes from three MLE-estimated models in comparison with the normalized histogram of real observations

decreases, which is consistent with the porosity formation mechanism.

6.1.2 Image Preprocessing and Two-Dimensional Void Extraction. In order to obtain the 2D sizes, we first convert raw images into binary images, then the boundaries or contours of defects are extracted, after which the pores are fitted to circles using a nonlinear least-square method by minimizing the sum of squared distances from the extracted data points to the circle boundary. The whole process is shown in Figs. 9(*a*)–9(*c*). At last, the numbers and radii of the voids on cross-sectional images are obtained to inference the size distribution, VND, and porosity in 3D space. The number of fitted circles is 379, 203, 43, respectively, for the three specimens. Figure 10 illustrates the radius distributions of fitted circles and the vertical dashed lines denote the mean values, respectively, for the three specimens.

6.1.3 Size Distribution Selection. In order to select an appropriate model for size distribution of pores, we propose to use MLE to fit the three candidate models, i.e., uniform, truncated normal, and log-normal distribution, and then simulate 2D observations using the fitted models. It is expected that if a candidate model is the true model, the simulated observations would follow the same distribution as the actual observations. Therefore, we use Kolmogorov–Smirnov (KS) test to evaluate the difference of simulated distribution and actual distribution and then perform a hypothesis testing on the proximity of these two distributions. The KS statistic is

$$D_{n_1,n_2} = \sup_x |F_{1,n_1}(x) - F_{2,n_2}(x)|$$

where $F_{1,n_1}(x)$ and $F_{2,n_2}(x)$ are two distribution functions. Clearly, D_{n_1,n_2} is the maximum of the vertical difference between these two distribution functions.

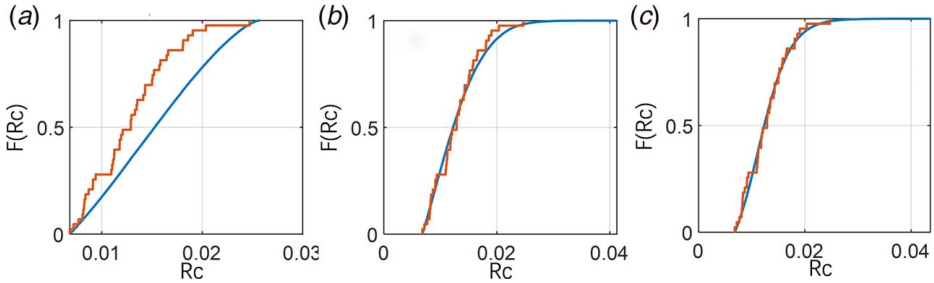

Fig. 12 The empirical distribution function of the observations in comparison with these of the simulated ones from (*a*) the uniform, (*b*) truncated normal, and (*c*) log-normal distributions

Table 3 Comparison of the proposed method with Wu's method in the size distribution, VND, and volume fraction estimation

	The proposed method				Wu's method [1]			
No.	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\rho}$ (%)	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$	$\hat{\rho}$ (%)
1	2.56	0.46	722.5	5.87	2.62	0.42	720.3	6.08
2	2.34	0.40	662.5	2.11	2.47	0.33	656.5	2.16
3	2.32	0.39	201.7	0.44	2.45	0.31	201.8	0.47

Table 4 Comparison of the porosity estimation using different methods

Specimen	The proposed approach (%)	Wu's method (%)	Area fraction (%)	X-ray CT (%)	Archimedes (%)
1	5.87	6.08	6.01	6.63	5.85
2	2.11	2.16	1.34	1.03	1.99
3	0.44	0.47	0.47	0.22	0.46

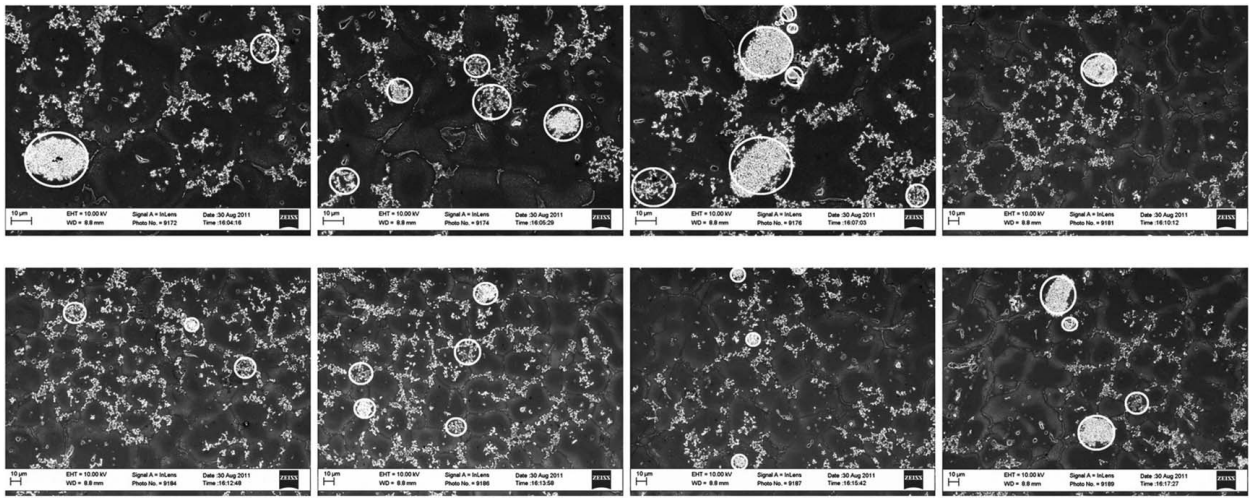

Fig. 13 Nanoparticle clusters shown in SEM images of MMNCs fabricated using the ultrasonic cavitation-assisted casting process

Based on the results of KS test, log-normal distribution is finally selected to model the size distribution of pores for all three specimens. To illustrate the model selection process, we take specimen 3 as an example. The truncation threshold is selected to be the minimum radius of all the observed samples, which is $l_0 = 2.8 \times 10^{-3}$. Figure 11 presents the normalized histogram of the 2D observations shown in Fig. 8(*c*) and simulated pdfs based on three size distributions. Obviously, the log-normal and truncated normal distributions are much closer to the true observations than the uniform distribution. The KS test results are shown in Fig. 12. The test statistics are 0.2758, 0.1094, and 0.0798, respectively, and the corresponding *p*-values are 0.0022, 0.6575, and 0.9381, respectively, which quantitatively demonstrate that the log-normal distribution is the best one. As a result, we utilize log-normal distribution to model the size distribution of pores.

6.1.4 Results. Table 3 shows the estimated parameters utilizing the proposed approach and Wu's method [1], which fails to consider the data truncation issue. We can easily see that with the scanning speed increasing, the size, VND, and porosity are all reduced, which is consistent with the pore formation mechanism. On the other hand, compared with the proposed approach, Wu's method overestimated the size distribution while underestimated the volume number densities slightly due to not considering the data truncation, which consequently leads to larger porosity estimation.

To evaluate the accuracy of the porosity estimation, we compare the estimation results with several existing porosity inspection method [7,32], including (1) area fraction, (2) X-ray CT, and (3) Archimedes.

Table 4 shows the porosity estimation results. Among them, the Archimedes method is to directly determine the density of the specimens by measuring the mass in air and water, which can be considered as a benchmark to evaluate the accuracy of other methods. Therefore, the true porosity of each specimen can be regarded as 5.85%, 1.99%, and 0.46%, respectively. Clearly, the proposed method has the highest accuracy among all the other inspection methods in most cases (specimen 1 and specimen 2), which indicates the necessity of considering the truncation effects. The error is due to insufficient pores for implementing the proposed method. The X-ray CT method estimates the porosity by calculating the area fraction of voids on X-ray slices. The X-ray CT method estimates the porosity by calculating the area fraction of voids on X-ray slices. It significantly underestimates the porosity for the second and third specimens and shows the lowest accuracy. The reason may be that due to the resolution issue, this method is incapable of detecting small voids. The area fraction method also has a good performance, especially on the first and third specimens.

6.2 Cluster Estimation in Metal-Matrix Nanocomposites

6.2.1 Experiments and Observations. In this section, we apply the proposed model to estimate the distribution of nanoparticle clusters of A206–Al_2O_3 MMNCs, which was fabricated using

Table 5 Observed radii of cross sections of nanoparticle clusters from SEM images

Obs. no.	*Rc* (nm)	Obs. no.	*Rc* (nm)	Obs. no.	*Rc* (nm)
1	11,667	11	7222	21	4444
2	5556	12	5556	22	16,667
3	16,667	13	5556	23	10,556
4	10,556	14	5556	24	6111
5	15,833	15	11,667	25	9444
6	6944	16	2778	26	4167
7	9444	17	1667	27	4722
8	10,000	18	2778	28	7778
9	11,111	19	11,944	29	8333
10	7778	20	8333	30	5278

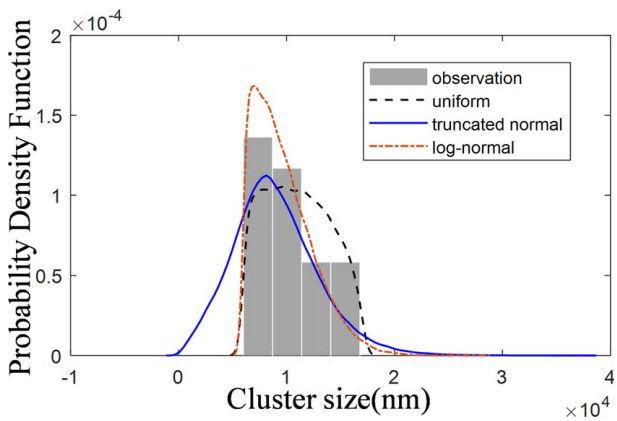

Fig. 14 Pdfs of the simulated 2D sizes from three MLE-estimated models in comparison with the normalized histogram of real observations

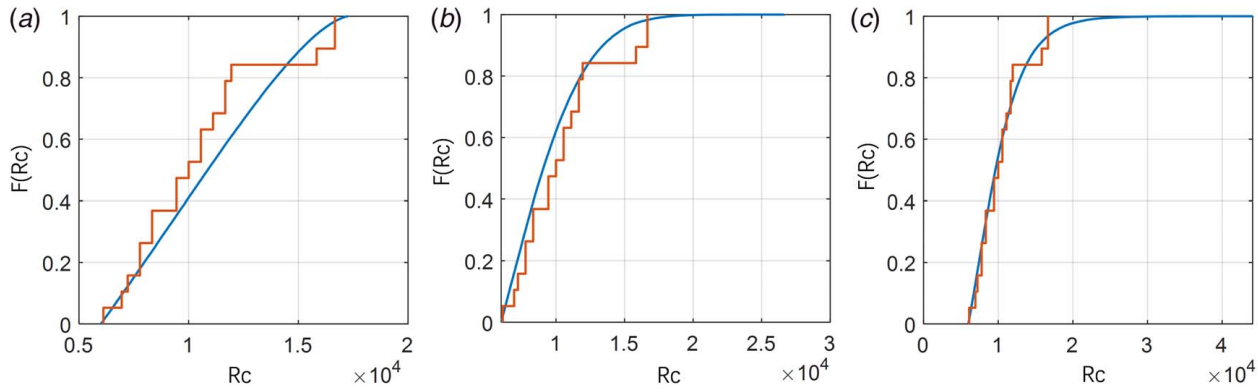

Fig. 15 The empirical distribution function of the 30 observations in comparison with these of simulated ones from the (*a*) uniform, (*b*) truncated normal, and (c) log-normal distributions

ultrasonic cavitation-based casting technology [33]. A206 is a type of aluminum–copper alloy with a chemical composition of Al (93.5–95.3%), Cu (4.2–5.0%), Mn (0.2–0.5%), Mg (0.15–0.35%), and Fe ($\leq 0.1\%$). The nanoparticle clusters Al_2O_3 incorporated in this alloy help to improve the hot tearing resistance of A206 [34]. Figure 13 shows eight SEM images, and 30 nanoparticle clusters are observed in total. The area for each image is $1.617 \times 10^4\ \mu m^2$. The cluster sizes are listed in Table 5. Similar to Ref. [24], we set the lower bound $r_l = 6000$ nm, above which the cluster is considered to be detrimental to the material properties. As stated in Sec. 2, the threshold l_0 is also set to be 6000 nm to maximize the sample size. Clusters smaller than 6000 nm will not be considered, and thus, 19 clusters are left to estimate the parameters of size and VND.

6.2.2 Size Distribution Selection and Estimation Results. Similar to the case of metal addictive manufacturing, we first select the appropriate model through KS test for size distribution. Ultimately, log-normal distribution has been chosen according to the comparison results. Figure 14 presents the normalized histogram of the 30 clusters and simulated pdfs of the three candidate models. Figure 15 illustrates the empirical accumulative distribution function of the 30 observations in comparison with that of the candidate distributions. The KS test statistics are 0.2329, 0.1809, and 0.1066 for the uniform, truncated normal, and log-normal distributions, and the corresponding p-values are 0.2212, 0.5206, and 0.9753, respectively. Therefore, we estimate the parameters of log-normal distribution based on the observations on the SEM images. The estimated log-normal distributions are log-normal ($\mu = 9.13$, $\sigma = 0.36$) by the proposed method and log-normal ($\mu = 9.2$, $\sigma = 0.31$) by Wu's method [1]. The KS test statistic of Wu's model is 0.1130, slightly larger than that of the proposed method, implying that the proposed model outperforms Wu's method. The number density is estimated as $\hat{\lambda} = 8.5 \times 10^{-6}/\mu m^3$ and the volume fraction is estimated as $\hat{\rho} = 0.050$.

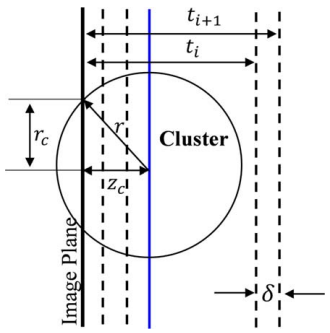

Fig. 16 Modeling the intersected particles with $r_c \geq l_0$ as a non-homogeneous Poisson process

7 Conclusion and Discussion

In this paper, we developed new statistical approaches to inferring particles in 3D space based on 2D cross-sectional images with data truncation problem. In the proposed model, we first establish the relationship between the pdf of 3D clusters and those of observed circular cross sections on size with existence of data truncation. Then, the likelihood function of 2D observations is derived and an efficient Monte Carlo EM algorithm is developed to estimate the parameters of size distribution. In addition, the linkage between 3D and 2D particle number density is also established with data truncation considered for density estimation. For the VND estimation, we prove that the number of particles shown on cross-sectional images can be modeled by a non-homogeneous Poisson process. Ultimately, the volume fraction of 3D particles can be estimated based on the size and density estimations. Both the simulation and real case studies in metal addictive manufacturing and MMNCs show that the proposed method is capable of effectively inferring 3D particles based on 2D cross-sectional images when there exists data truncation.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtained from the corresponding author upon reasonable request. The authors attest that all data for this study are included in the paper.

Funding Data

- This work was supported by the Natural Science Foundation of China under Grant Nos. 51875003 and 71932006.

Appendix A: Proof of Proposition 1

The particles are uniformly distributed on both sides of the image plane. Now we only consider the right side of the image plane, as shown in Fig. 16. Divide the specimen into equally sized cells with a sufficiently small thickness δ. Suppose the upper bound of the particle radius is r_u, then we only need to consider the cells with distance $t \leq \sqrt{r_u^2 - l_0^2}$; otherwise, the particle will not intersect

the image plane or will not result in a cross section with size $r_c \geq l_0$ if intersected.

Let $N(t_i)$ denote the number of particles that are within distance $t \leq t_i$ to the image plane and are intersected by the image plane with $r_c \geq l_0$. Then the number of particles with $r_c \geq l_0$ and coming from the ith cell can be calculated as $N_i = N(t_i) - N(t_{i-1})$, $i \geq 1$. Define $N_0 = 0$ and $t_0 = 0$, then we have

(1) $N(0) = 0$ by definition;
(2) N_1, N_2, … are independent (independent increments property);
(3) Based on the CSR assumption, N_i, $i \geq 1$, follow Poisson distributions with Poisson rate $\lambda_i = A\delta\lambda_{3D} \Pr\left(\sqrt{r^2 - t_i^2} \geq l_0\right)$. Therefore,

$$\lim_{\delta \to 0} \frac{\Pr(N_i = 1)}{\delta} = A\lambda_{3D} \Pr\left(\sqrt{r^2 - t_i^2} \geq l_0\right), \quad \lim_{\delta \to 0} \frac{\Pr(N_i > 1)}{\delta} = 0$$

Based on the above properties, the particles intersecting the image plane from the *right side* and satisfying $r_c \geq l_0$ can be modeled as a non-homogeneous Poisson process with density function

$$\lambda(t) = A\lambda_{3D} \Pr\left(\sqrt{r^2 - t^2} \geq l_0\right) = A\lambda_{3D} \int_{\max\left\{\sqrt{t^2 + l_0^2}, r_l\right\}}^{r_u} f(r|\boldsymbol{\theta}) dr, \quad t \leq \sqrt{r_u^2 - l_0^2}$$

Therefore, the Poisson rate of 2D observations can be calculated by

$$\lambda_{2D} = 2\int_0^{\sqrt{r_u^2 - l_0^2}} \lambda(t) dt = 2A\lambda_{3D} \int_0^{\sqrt{r_u^2 - l_0^2}} \int_{\max\left\{\sqrt{t^2 + l_0^2}, r_l\right\}}^{r_u} f(r|\boldsymbol{\theta}) dr dt$$

$$= 2A\lambda_{3D} \int_{\max\{r_l, l_0\}}^{r_u} f(r|\boldsymbol{\theta}) \sqrt{r^2 - l_0^2} dr = 2A\lambda_{3D} E\left(\sqrt{(r^2 - l_0^2)^+} | \boldsymbol{\theta}\right)$$

Appendix B: Proof of Proposition 2

Suppose there are N particles with volume V_i, $i = 1, \ldots, N$, in the specimen of volume V. Then $N \sim \text{Poisson}(\lambda_{3D} V)$ and $\rho = \sum_{i=1}^{N} V_i / V$. Clearly, $\sum_{i=1}^{N} V_i$ follows a compound Poisson distribution. The mean and variance of ρ can be calculated as

$$E(\rho) = \frac{E(N)E(V_i)}{V} = \lambda_{3D} E(V_i) = \frac{4\pi\lambda_{3D}}{3} E(r^3|\boldsymbol{\theta})$$

$$\text{Var}(\rho) = \frac{E(N)E(V_i^2)}{V^2} = \frac{16\pi^2\lambda_{3D}}{9V} E(r^6|\boldsymbol{\theta})$$

References

[1] Wu, J., Yuan, Y., and Li, X., 2017, "Size Distribution Estimation of Three-Dimensional Particle Clusters in Metal-Matrix Nanocomposites Considering Sampling Bias," ASME J. Manuf. Sci. Eng., **139**(8), p. 081017.
[2] Wu, J., Yuan, Y., Gong, H., and Tseng, T.-L., 2017, "Inferring 3D Ellipsoids Based on Cross-Sectional Images With Applications to Porosity Control of Additive Manufacturing," IISE Trans., **50**(7), pp. 570–583.
[3] West, M. J., Slomianka, L., and Gundersen, H. J., 1991, "Unbiased Stereological Estimation of the Total Number of Neurons in the Subdivisions of the Rat Hippocampus Using the Optical Fractionator," Anat. Rec., **231**(4), pp. 482–497.
[4] West, M. J., 2002, "Design-Based Stereological Methods for Counting Neurons," Prog. Brain Res., **135**, pp. 43–51.
[5] Wu, J., Zhou, S., and Li, X., 2015, "Ultrasonic Attenuation Based Inspection Method for Scale-Up Production of A206–Al_2O_3 Metal Matrix Nanocomposites," ASME J. Manuf. Sci. Eng., **137**(1), p. 011013.
[6] Buchbinder, D., Schleifenbaum, H., Heidrich, S., Meiners, W., and Bültmann, J., 2011, "High Power Selective Laser Melting (HP SLM) of Aluminum Parts," Phys. Procedia, **12**(1), pp. 271–278.
[7] Gong, H., 2013, "Generation and Detection of Defects in Metallic Parts Fabricated by Selective Laser Melting and Electron Beam Melting and Their Effects on Mechanical Properties," University of Louisville, Louisville, KY.
[8] Aboulkhair, N. T., Everitt, N. M., Ashcroft, I., and Tuck, C., 2014, "Reducing Porosity in AlSi10Mg Parts Processed by Selective Laser Melting," Addit. Manuf., **1**, pp. 77–86.
[9] Li, S. J., Murr, L. E., Cheng, X. Y., Zhang, Z. B., Hao, Y. L., Yang, R., Medina, F., and Wicker, R. B., 2012, "Compression Fatigue Behavior of Ti–6Al–4V Mesh Arrays Fabricated by Electron Beam Melting," Acta Mater., **60**(3), pp. 793–802.
[10] Kruth, J. P., Froyen, L., Van Vaerenbergh, J., Mercelis, P., Rombouts, M., and Lauwers, B., 2004, "Selective Laser Melting of Iron-Based Powder," J. Mater. Process. Technol., **149**(1), pp. 616–622.
[11] Wu, J., Liu, Y., and Zhou, S., 2016, "Bayesian Hierarchical Linear Modeling of Profile Data With Applications to Quality Control of Nanomanufacturing," IEEE Trans. Autom. Sci. Eng., **13**(3), pp. 1355–1366.
[12] Wu, J., Zhou, S., and Li, X., 2013, "Acoustic Emission Monitoring for Ultrasonic Cavitation Based Dispersion Process," ASME J. Manuf. Sci. Eng., **135**(3), p. 031015.
[13] Liu, Y., Wu, J., Zhou, S., and Li, X., 2016, "Microstructure Modeling and Ultrasonic Wave Propagation Simulation of A206–Al_2O_3 Metal Matrix Nanocomposites for Quality Inspection," ASME J. Manuf. Sci. Eng., **138**(3), p. 031008.
[14] Cao, G., Konishi, H., and Li, X., 2008, "Mechanical Properties and Microstructure of Mg/SiC Nanocomposites Fabricated by Ultrasonic Cavitation Based Nanomanufacturing," ASME J. Manuf. Sci. Eng., **130**(3), p. 031105.
[15] Le, B., Khaliq, J., Huo, D., Teng, X., and Shyha, I., 2020, "A Review on Nanocomposites. Part 1: Mechanical Properties," ASME J. Manuf. Sci. Eng., **142**(10), p. 100801.
[16] Zhou, Q., Zhou, J., De Cicco, M., Zhou, S., and Li, X., 2014, "Detecting 3D Spatial Clustering of Particles in Nanocomposites Based on Cross-Sectional Images," Technometrics, **56**(2), pp. 212–224.
[17] Park, H., and Miwa, K., 2003, "X-Ray Computed Tomography for Micro Porosity in AZ91D Alloy," Mater. Trans., **44**(11), pp. 2326–2333.
[18] Kiss, L., Söderlund, J., Niklasson, G., and Granqvist, C., 1999, "New Approach to the Origin of Lognormal Size Distributions of Nanoparticles," Nanotechnology, **10**(1), p. 25.
[19] Wei, G. C., and Tanner, M. A., 1990, "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms," J. Am. Stat. Assoc., **85**(411), pp. 699–704.
[20] Babinsky, E., and Sojka, P. E., 2002, "Modeling Drop Size Distributions," Prog. Energy Combust. Sci., **28**(4), pp. 303–329.
[21] Diggle, P. J., 2013, *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, CRC Press, New York.
[22] Zeng, L., Zhou, Q., De Cicco, M. P., Li, X., and Zhou, S., 2012, "Quantifying Boundary Effect of Nanoparticles in Metal Matrix Nanocomposite Fabrication Processes," IIE Trans., **44**(7), pp. 551–567.
[23] Baddeley, A., and Jensen, E. B. V., 2004, *Stereology for Statisticians*, CRC Press, Boca Raton, FL.
[24] Liu, H., Zhou, S., and Li, X., 2013, "Inferring the Size Distribution of 3D Particle Clusters in Metal Matrix Nanocomposites," ASME J. Manuf. Sci. Eng., **135**(1), p. 011013.
[25] Dempster, A. P., Laird, N. M., and Rubin, D. B., 1977, "Maximum Likelihood From Incomplete Data Via the EM Algorithm," J. R. Stat. Soc. Ser. B Methodol., **39**(1), pp. 1–38.
[26] Seo, Y.-H., and Kang, C.-G., 1995, "The Effect of Applied Pressure on Particle-Dispersion Characteristics and Mechanical Properties in Melt-Stirring Squeeze-Cast SiCp/Al Composites," J. Mater. Process. Technol., **55**(3–4), pp. 370–379.
[27] Russ, J. C., and Neal, F. B., 2015, *The Image Processing Handbook*, CRC Press, Inc., Boca Raton, FL.
[28] Khodashenas, B., and Ghorbani, H. R., 2015, "Synthesis of Silver Nanoparticles With Different Shapes," Arabian J. Chem., **12**(8), pp. 1823–1838.
[29] Weingarten, C., Buchbinder, D., Pirch, N., Meiners, W., Wissenbach, K., and Poprawe, R., 2015, "Formation and Reduction of Hydrogen Porosity During Selective Laser Melting of AlSi10Mg," J. Mater. Process. Technol., **221**, pp. 112–120.
[30] Kabir, M. R., and Richter, H., 2017, "Modeling of Processing-Induced Pore Morphology in an Additively-Manufactured Ti–6Al–4V Alloy," Materials, **10**(2), p. 145.
[31] Vilaro, T., Colin, C., and Bartout, J. D., 2011, "As-Fabricated and Heat-Treated Microstructures of the Ti–6Al–4V Alloy Processed by Selective Laser Melting," Metall. Mater. Trans. A, **42**(10), pp. 3190–3199.
[32] Slotwinski, J. A., Garboczi, E. J., and Hebenstreit, K. M., 2014, "Porosity Measurements and Analysis for Metal Additive Manufacturing Process Control," J. Res. Natl. Inst. Stand. Technol., **119**, pp. 494–528.
[33] Yang, Y., and Li, X., 2007, "Ultrasonic Cavitation Based Nanomanufacturing of Bulk Aluminum Matrix Nanocomposites," ASME J. Manuf. Sci. Eng., **129**(3), pp. 497–501.
[34] Choi, H., Cho, W. H., and Konish, H., 2013, "Nanoparticle-Induced Superior Hot Tearing Resistance of A206 Alloy," Metall. Mater. Trans. A, **44**(4), pp. 1897–1907.