Measurement 238 (2024) 115291

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Numerical investigation on sample population requirement for mean particle diameter analysis

Qiwen Jin [a,b], Zhiming Lin [a], Yingchun Wu [a], Xuecheng Wu [a,b,*]

[a] State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou 310027, China
[b] Ningbo Research Institute, Zhejiang University, Ningbo 315100, China

ARTICLE INFO

Keywords:
Particle characterization
Sample size
Gates–Gaudin–Schuhmann
Rosin–Rammler
Simulation
Uncertainty evaluation

ABSTRACT

The determination of the minimum particle number needed to meet error specifications and confidence levels poses a fundamental challenge in particle size analysis. Conventional models, primarily designed for log-normal distributions, may yield inaccuracies when applied to other distribution functions. This study introduces a numerical approach to explore the necessary sample size for measuring different types of mean diameters. The methodology involves distribution conversion, sample generation, repeated sampling, and error estimation. Specifically, Gates–Gaudin–Schuhmann (GGS) and Rosin–Rammler (RR) distributions serve as representative models in this investigation. The impact of sample size, span ratio, and boundary sizes on relative errors is examined. Additionally, an empirical model is formulated, simplifying the calculation of the requisite sample size when the relative error and mass-based span ratio are provided. The proposed method is also verified using experimental particle size data that follow the RR distribution.

1. Introduction

Particle size distribution (PSD) and mean size are among the most fundamental and critical characteristic parameters of particulate matter [1]. Reliable characterization and measurement of particle size are indispensable processes in many scientific researches and industrial productions. Sampling is the first step in particle characterization, aimed at obtaining small samples that effectively represent the physical and chemical properties of the entire bulk [2]. However, sampling errors are often underestimated or even ignored. These errors include fundamental errors and segregation errors [3]. The fundamental error is related to the discrete nature of the particles and determines the minimum amount of sample required for accurate characterization. Therefore, to ensure measurement reliability, a key issue is determining the appropriate sample size. This is particularly important in image-based particle measurement techniques, as it can guide the balance between accuracy and measurement cost [4,5]. In applications of online particle monitoring, understanding the required minimum sample size also helps improve the time resolution capability of the measurement system [6,7].

Regarding the issue of particle sample size, existing research mainly unfolds through three dimensions including theoretical derivation, simulation study, and experimental data analysis. Masuda and Iinoya [8] have studied the sample size problem theoretically and derived the relationship between particle number required and relative error under a given confidence level, based on utilization of the distribution width. They used standard deviation of distribution to represent the distribution width which is called span ratio in this work. Their theory has been confirmed by computer simulation [9] and adopted by the ISO guidance (ISO 13322-1:2014) for static image analysis [10]. However, Masuda's theory has several limitations. (1) The particle size distribution should follow an idealized number-weighted log-normal distribution (LND). (2) It is applicable only to single-component particle systems. (3) The geometric standard deviation of the distribution should be less than 1.6 [11]. (4) It can only resolve the uncertainty level of mean particle size. There have been a number of reported approaches to address these limitations.

Assumption of LND may lead to a misleading sample size evaluation may be introduced when dealing with skewed or multi-modal distributions [5]. In fact, many materials do not adhere to the idealized log-normal distribution but have upper and lower bounds or follow other distributions such as Gates–Gaudin–Schuhmann (GGS) or Rosin-Rammler (RR) [12,13] distributions. Additionally, in some cases, PSDs may not conform to any distribution function. Yoshida et al. [14] have proposed a new theoretical equation to calculate fundamental uncertainty region based upon a LND truncated by the maximum and minimum size. In their further work [15,16], Tschebyscheff theory was introduced to determine the uncertainty region of poly-disperse

* Corresponding author at: State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou 310027, China.
E-mail address: wuxch@zju.edu.cn (X. Wu).

https://doi.org/10.1016/j.measurement.2024.115291
Received 14 June 2024; Accepted 9 July 2024
Available online 18 July 2024

Nomenclature

δ	Relative error at 95% confidence level
Γ	Upper incomplete gamma function
λ	Scale parameter of distribution function
σ	Standard deviation
σ_m	Mass-based span ratio
σ_n	Number-based span ratio
${}_2F_1$	Hypergeometric function
b	Coefficient of linear regression
D_{32}	Sauter diameter
D_{m50}	Mass median diameter
D_m	Mass mean diameter
D_{n50}	Number median diameter
D_n	Number mean diameter
e	Relative error in a sampling
F_m	Mass-weighted cumulative distribution function
f_m	Mass-weighted probability density function
F_n	Number-weighted cumulative distribution function
f_n	Number-weighted probability density function
k	Coefficient of linear regression
N	Sample size
n	Spread parameter of distribution function
P	Polynomial function
R^2	Correlation coefficient
x	Particle size
x_{max}	Upper bound of particle size
x_{min}	Lower bound of particle size
sim	Simulation results
the	Theoretical results
ARM	Acceptance-rejection method
CDF	Cumulative distribution function
GGS	Gates–Gaudin–Schuhmann distribution
ITM	Inverse transform method
LND	Log normal distribution
PA	Pareto distribution
PDF	Probability density function
PSD	Particle size distribution
RR	Rosin-Rammler distribution

particles and a picket-fence distribution composed of two kinds of nearly mono-disperse particles. Moreover, Yoshida et al. [17] proposed a theoretical equation of uncertainty region over all particle diameter range on the basis of Masuda's theory. Endo [18] provided a theoretical solution for estimating the confidence intervals for geometric mean size and geometric standard deviation. A theoretical evaluation of the uncertainty of percentile values for number weighted distribution data was introduced by Matsuyama [19].

For particle system following GGS, RR, or other distributions in which theoretical derivation of the confidence level could be difficult. Computer simulations can be used to obtain quantitative relationships between the sample size and errors, and to fit empirical formulas ultimately [11,20]. Paine [20] developed empirical formulate for the critical number of particles which must be counted to prevent the error from blowing up through simulations based on LND. Camalan [21] conducted simulations by create particle population from GGS model and used Kolmogorov–Smirnov and Chi-Square Goodness-of-Fit tests to compare the population and its samples. As for experimental data obtained by specific measurement technique, PSD may not conform to any specific distribution function. In such cases, non-parametric statistical analysis is commonly utilized for assessing uncertainty and determining the required sample size [4]. The bootstrap method, featuring with non-parametric, flexibility, and simplicity, is a statistic method to evaluate confidence level of statistical quantities. Applications of the bootstrap method for uncertainty and required sample size estimation have been reported in early studies [5,22–24]. Although many efforts have been taken to solve the particle sample size problem, any theoretical or empirical relationships between the sample size and errors have not yet been established for GGS and RR distributions.

The aim of this work is to establish a relationship between the relative error of various mean diameters and the sample size at a specific confidence level. A simulation approach is developed to generate particle populations following the expected distribution functions, particle sampling, and error estimation. GGS and RR are taken as two examples and empirical equations are proposed.

2. Numerical methods

Fig. 1 illustrates the simulation procedure. In industrial production processes, particles are generally abundant. However, when measuring particle size, a small number of samples are often obtained through sampling to represent the overall particle size. Therefore, it can be approximated as sampling from an infinitely large particle population. To simulate this process, a large number of random numbers following a certain distribution are first generated as the total sample of particles to be measured. Subsequently, a random sampling approach is employed to obtain a small sample from the particle population. In simulations, particles are assumed to be homogeneous, spherical, and solid. Numerical simulations are conducted on Matlab (Version R2021b).

2.1. Conversion of distribution functions

In the study conducted by Masuda and Iinoya [8], the relation between the particle sample size N, the relative error δ, and the standard deviation σ was determined theoretically based upon an assumption of log normal distribution which is number-weighted. However, many kinds of particle, probably produced by breakage or pulverization, are commonly characterized with sieving analysis in which the mass-weighted cumulative distributions are obtained. Camalan [25] has deduced the conversion method to convert mass-weighted distributions to their number-weighted equivalents, including Gates–Gaudin–Schuhmann, Pareto (PA), and Rosin-Rammler distributions. In this work, doubly-truncated GGS and RR distributions with a lower bound x_{min} and an upper bound x_{max} are investigated and the cumulative distribution functions (CDF) and the probability density functions (PDF) are expressed as the following equations.

For GGS distribution,

$$F_m(x) = \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right)^n \tag{1}$$

$$f_m(x) = \frac{n\left(x - x_{min}\right)^{n-1}}{\left(x_{max} - x_{min}\right)^n} \tag{2}$$

$$F_n(x) = \frac{{}_2F_1\left(3, n; 1+n; \left(\frac{x_{min}-x}{x_{min}}\right)\right)\left(x - x_{min}\right)^n}{{}_2F_1\left(3, n; 1+n; \left(\frac{x_{min}-x_{max}}{x_{min}}\right)\right)\left(x_{max} - x_{min}\right)^n} \tag{3}$$

$$f_n(x) = \frac{n x_{min}^3\left(x - x_{min}\right)^{n-1} x^{-3}}{{}_2F_1\left(3, n; 1+n; \left(\frac{x_{min}-x_{max}}{x_{min}}\right)\right)\left(x_{max} - x_{min}\right)^n} \tag{4}$$

For RR distribution,

$$F_m(x) = \frac{\exp\left(-\left(\frac{x_{min}}{\lambda}\right)^n\right) - \exp\left(-\left(\frac{x}{\lambda}\right)^n\right)}{\exp\left(-\left(\frac{x_{min}}{\lambda}\right)^n\right) - \exp\left(-\left(\frac{x_{max}}{\lambda}\right)^n\right)} \tag{5}$$

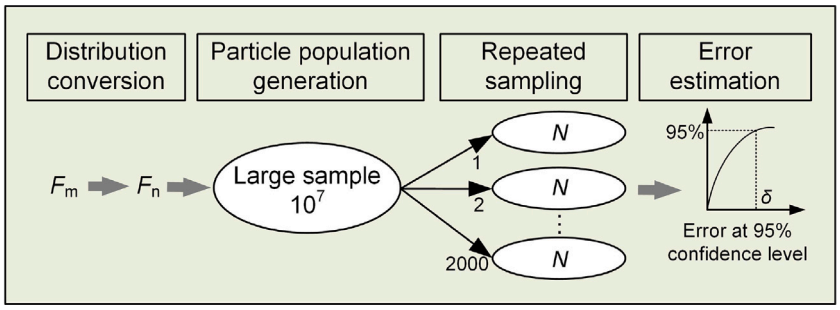

Fig. 1. Schematic illustration of the simulation procedure.

$$f_m(x) = \frac{n x^{n-1} \exp\left(-\left(\frac{x}{\lambda}\right)^n\right)}{\lambda^n \left[\exp\left(-\left(\frac{x_{\min}}{\lambda}\right)^n\right) - \exp\left(-\left(\frac{x_{\max}}{\lambda}\right)^n\right)\right]} \tag{6}$$

$$F_n(x) = \frac{\Gamma\left(\frac{n-3}{n}, \left(\frac{x_{\min}}{\lambda}\right)^n\right) - \Gamma\left(\frac{n-3}{n}, \left(\frac{x}{\lambda}\right)^n\right)}{\Gamma\left(\frac{n-3}{n}, \left(\frac{x_{\min}}{\lambda}\right)^n\right) - \Gamma\left(\frac{n-3}{n}, \left(\frac{x_{\max}}{\lambda}\right)^n\right)} \tag{7}$$

$$f_n(x) = \frac{n x^{n-4} \exp\left(-\left(\frac{x}{\lambda}\right)^n\right)}{\lambda^{n-3} \left[\Gamma\left(\frac{n-3}{n}, \left(\frac{x_{min}}{\lambda}\right)^n\right) - \Gamma\left(\frac{n-3}{n}, \left(\frac{x_{max}}{\lambda}\right)^n\right)\right]} \tag{8}$$

where F_m is the mass-weighted CDF, f_m the mass-weighted PDF, F_n the number-weighted CDF, and f_n the number-weighted PDF. $n > 0$ is the spread parameter indicating the spread of particle sizes. λ is the scale parameter indicating the overall particle size with $x_{\min} < \lambda < x_{\max}$. ${}_2F_1\left(3, n; 1+n; \left(\frac{x_{\min}-x}{x_{\min}}\right)\right)$ is the hypergeometric function. $\Gamma\left(\frac{n-3}{n}, \left(\frac{x}{\lambda}\right)^n\right)$ is the upper incomplete gamma function.

2.2. Generation of particle populations

Acceptance-rejection method (ARM) and inverse transform method (ITM) are two typical methods to generate populations of particles following an expected distribution function. In addition, Camalan [21] has generated a population of particles mass discretization in which the cumulative distribution values confirm the mass-weighted CDF at the edges of each size fractions.

ARM [26] is suitable for generation of particle populations that the number-weighted PDF is known while the number-weighted CDF is difficult to calculate. Assume that the expected particle population **X** conforms number-weighted PDF $f_n(x)$ and there exists another frequency distribution $g(x)$ and a constant c with $c \cdot g(x) \geq f_n(x)$. One can use the following procedures to generate the expected particle sample:

(1) Generate random number pairs $\mathbf{Y} \sim g(\mathbf{Y})$.

(2) Generate uniform random number $u \in [0, 1]$.

(3) If $u \leq f_n(\mathbf{Y})/[cg(\mathbf{Y})]$, $\mathbf{Y} \rightarrow \mathbf{X}$, or reject **Y**.

(4) Repeat above procedures until the valid particle number satisfies the goal.

The selection of c is critical to improve the generation efficiency of particle sample. The efficiency would decrease greatly if c is too large. It should be noted that one can realize the algorithm with a simple and efficient method in which $g(x) = 1$ and $c = \max[f_n(x)]$.

ITM [19,27] is suitable for generating particle samples that the number-weighted CDF is known and the inverse function can be resolved. However, the method has limitations in mathematical expressions of number-based cumulative distribution.

Suppose the expected particle population **X** with a number-weighted PDF $f_n(x)$ has a continuous and strictly increasing CDF $F_n(x)$. $F_n(x)$ and $f_n(x)$ are related by

$$F_n(x) = \int_{x_{\min}}^{x} f_n(x)\mathrm{d}x \tag{9}$$

Denote by $F_n^{-1}(x)$ the inverse of $F_n(x)$. The expected particle population **X** can be generated as follows:

(1) Generate uniformly distributed random number u belonging to [0,1].

(2) Return $\mathbf{X} = F_n^{-1}(y)$. That is, an inverse transformation is used to convert the random number $u \in [0, 1]$ to a random variate of the desired distribution.

In this work, the inverse function of the hypergeometric function is difficult to solve mathematically. The inverse of the upper incomplete gamma function cannot be computed when $n < 3$. Thus, ARM is utilized to generate the particle populations.

2.3. Particle sampling

For each distribution function, a large population containing 10^7 particles is generated. Fig. 2 shows that the number and mass weighted CDFs of the generated populations are consistent with the theoretical CDFs calculated from distribution functions for GGS (Eqs. (1) and (3)) and RR (Eqs. (5) and (7)) distributions. Each simulation is repeated for 2×10^3 times with random sampling from the large population by using the built-in function *datasample* in Matlab. The simulated sample size N ranges from 10^3 to 10^5 and the spread parameter n ranges from 1.2 to 6.0. The lower and upper bound of particle size are $x_{\min} = 1$ and $x_{\max}$ ranging from 20 to 160 (note that the unit is omitted), respectively. The scale parameter λ of the RR distribution is $\left[x_{\min} + \alpha(x_{\max} - x_{\min})\right]$ while α ranges from about 20% to 80%. The particle sample is represented with $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$.

2.4. Error estimation

Relative errors of five commonly used mean diameters, including mass median diameter D_{m50}, number median diameter D_{n50}, mass mean diameter D_m, number mean diameter D_n, and Sauter mean diameter D_{32}, are analyzed in this work. It is worth noting that, strictly speaking, the median diameter should not be confused with the mean particle diameter. For simplicity, we collectively refer to them as mean diameter in this work. The statistical definition and physical interpretation of the mentioned diameters are listed in Table 1. The CDF curve gives easy access to percentile PSD parameters such as D_{n50} or D_{m50}, representing the 50% point in the number or mass weighted cumulative undersize PSD, respectively [1]. D_n, D_m, and D_{32} are defined by the Moment-Ratio method [28]

$$\bar{D}_{p,q} = \left[\frac{\sum_i n_i D_i^p}{\sum_i n_i D_i^q}\right]^{1/(p-q)} \tag{10}$$

where p and q have real values and $p \neq q$. n_i is the number of particles in ith size class. For example, the mass mean diameter D_m is defined by $\bar{D}_{3,0}$ with $p = 3$ and $q = 0$.

The superscripts *the* and *sim* represent the theoretical results and simulated results, respectively. Theoretical results are calculated using the original number weighted PDF. D_{m50}^{the} and D_{n50}^{the} can be determined

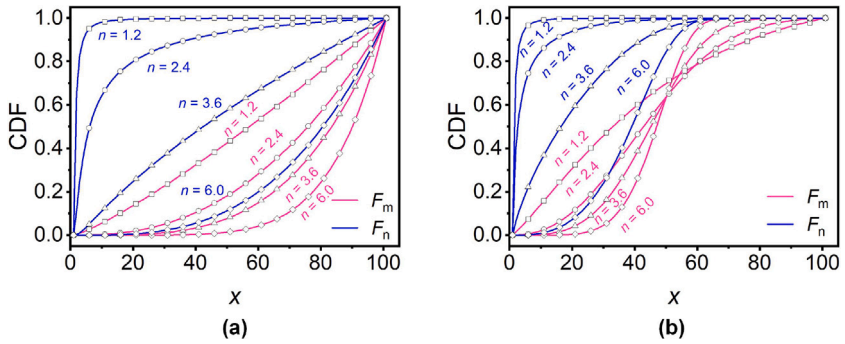

Fig. 2. The number and mass weighted CDFs of the generated populations (10^7) with various spread parameter (symbols) and theoretical CDFs calculated from distribution functions (solid lines). $x_{min} = 1$, $x_{max} = 101$, $\lambda = 50$. (a) GGS, (b) RR.

Table 1
Definition of the median and mean diameters.

Diameter	Symbol	Definition	Interpretation
Number median diameter	D_{n50}	$F_n(D_{n50}) = 50\%$	Fifty percent (by number) of the particles in the sample measured have diameters greater than D_{n50}.
Mass median diameter	D_{m50}	$F_m(D_{m50}) = 50\%$	Fifty percent (by mass) of the particles in the sample measured have diameters greater than D_{m50}.
Number mean diameter	D_n	$\bar{D}_{1,0}$	Arithmetic average diameter of all the particles in the sample.
Mass mean diameter	D_m	$\bar{D}_{3,0}$	Diameter of a spherical particle with a mass equal to the mean mass of all the particles in the sample.
Sauter mean diameter	D_{32}	$\bar{D}_{3,2}$	Diameter of a spherical particle with the same volume/surface area ratio as the entire ensemble.

through solving the equations that $F_m(D^{the}_{m50}) = 50\%$ and $F_n(D^{the}_{n50}) = 50\%$ with the built-in *vpasolve* function in Matlab. Moreover, D^{the}_m, D^{the}_n, and D^{the}_{32} are calculated by the following equations (continuous expressions of Eq. (10))

$$D^{the}_m = \left[\int_{x_{min}}^{x_{max}} x^3 f_n(x)\,dx\right]^{1/3} \tag{11}$$

$$D^{the}_n = \int_{x_{min}}^{x_{max}} x f_n(x)\,dx \tag{12}$$

$$D^{the}_{32} = \frac{\int_{x_{min}}^{x_{max}} x^3 f_n(x)\,dx}{\int_{x_{min}}^{x_{max}} x^2 f_n(x)\,dx} \tag{13}$$

where integrals are calculated with the built-in *integral* function in Matlab.

The spread parameter n fundamentally affects the span of the particle size distribution. This study delineates two types of span ratios based on mass and number distributions. Additionally, to maintain consistency with the definition by Masuda [8], the logarithmic span ratios are defined as follows

$$\sigma_m = \ln\left(D_{m84}/D_{m50}\right) \tag{14}$$

$$\sigma_n = \ln\left(D_{n84}/D_{n50}\right) \tag{15}$$

The logarithmic span ratios σ_m and σ_n of GGS and RR distribution are shown in Fig. 3. It is demonstrated that σ_m decreases as n increases, whereas σ_n initially increases and then decreases. It should be noted that the definition of span ratio is not unique. Other commonly used forms include $(D_{90} - D_{10})/D_{50}$ and D_{90}/D_{10} [1]. The span ratios derived from these two definitions are also plotted in Fig. 3. As can be seen, the three span ratios exhibit consistent trends, but there are differences in their magnitudes. For the span ratios of number weighted distributions, the values initially increase and then decrease, reaching a maximum around $n = 2$ to 3. Details of span ratios are introduced in Appendix A.

The mean diameters in each simulation are calculated with statistic analysis. The particle size range is divided into a series of intervals with edges of $\{x_{min}, D^{the}_{m1}, D^{the}_{m2}, \ldots, D^{the}_{m99}, x_{max}\}$ for D^{sim}_{m50} evaluation and $\{x_{min}, D^{the}_{n1}, D^{the}_{n2}, \ldots, D^{the}_{n99}, x_{max}\}$ for D^{sim}_{n50} evaluation. Note that the interval widths are not uniform since the uniform intervals may introduce greater artificial statistical errors. The particles in the sample are statistically analyzed to obtain mass and number weighted CDF curves. As is shown in Fig. 4, D^{sim}_{m50} and D^{sim}_{n50} can be determined by interpolation method. D^{sim}_m, D^{sim}_n, and D^{sim}_{32} are calculated by the following definitions

$$D^{sim}_m = \left[\frac{\sum_{i=1}^{N} x_i^3}{N}\right]^{1/3} \tag{16}$$

$$D^{sim}_n = \frac{\sum_{i=1}^{N} x_i}{N} \tag{17}$$

$$D^{sim}_{32} = \frac{\sum_{i=1}^{N} x_i^3}{\sum_{i=1}^{N} x_i^2} \tag{18}$$

The relative error in a sampling is defined as

$$e = \left|\frac{D^{sim} - D^{the}}{D^{the}}\right| \tag{19}$$

In the repeated sampling, the distributions of the relative errors are obtained, as is shown in Fig. 5 which is statistical result of 2×10^3 points for each mean diameter. The straight line in Fig. 5 represents the 95% confidence level. The intersection point of the 95% confidence line and distribution curve determines an error value δ which is regarded as the relative error for the specific $F_m(x)$ and mean diameter. Moreover, the error value δ means that 95% data should be in the region of relative error $\pm\delta$ when the particle number N, the function $F_m(x)$, and the mean diameter type are known.

Reliable error evaluation requires a sufficient number of simulations. The impact of the number of samples on the error evaluation results is shown in Fig. 6. The results indicate that when the number of samples is less than 1000, there are significant fluctuations in the results. When the number of samples exceeds 2000, the results become relatively stable, although minor fluctuations still exist. Similar results were observed under different parameters. Therefore, to ensure the reliability of the results while avoiding excessively slow computation speeds, this study selects 2000 sampling times.

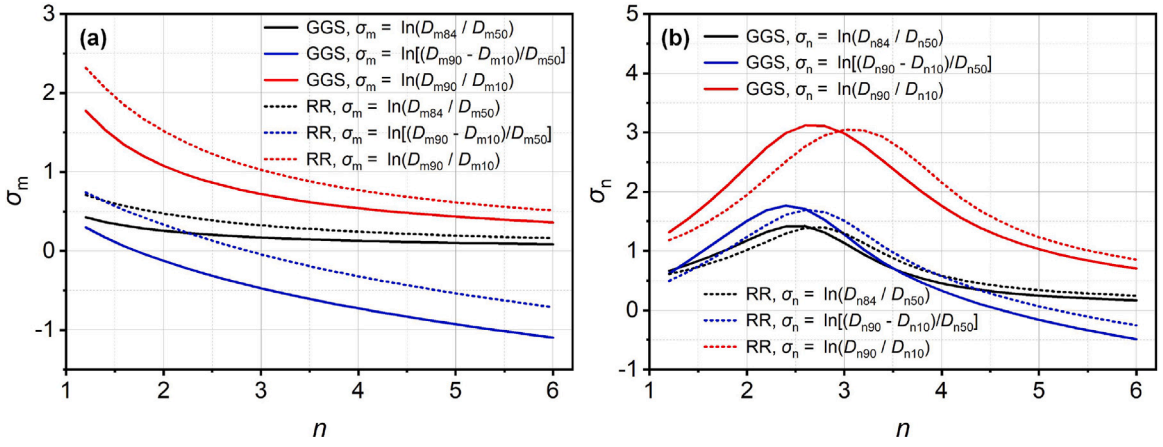

Fig. 3. Mass-based and number-based logarithmic span ratios of GGS and RR by different definitions. (a) σ_m, (b) σ_n.

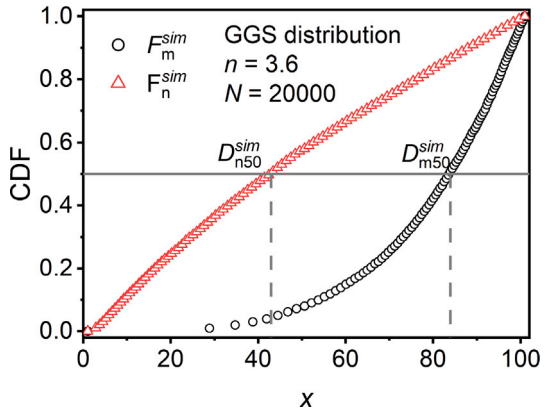

Fig. 4. Mass weighted and number weighted CDF of a simulated sample case.

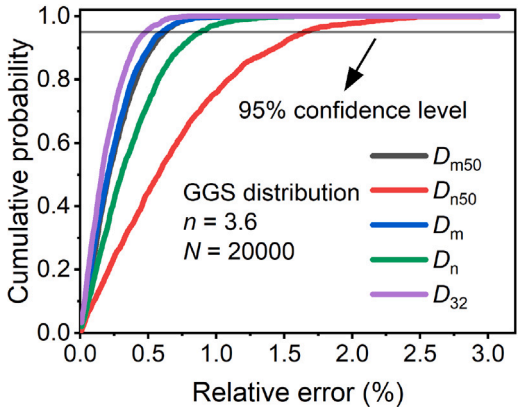

Fig. 5. Cumulative probability distributions of e. The horizontal straight line represents the 95% confidence level.

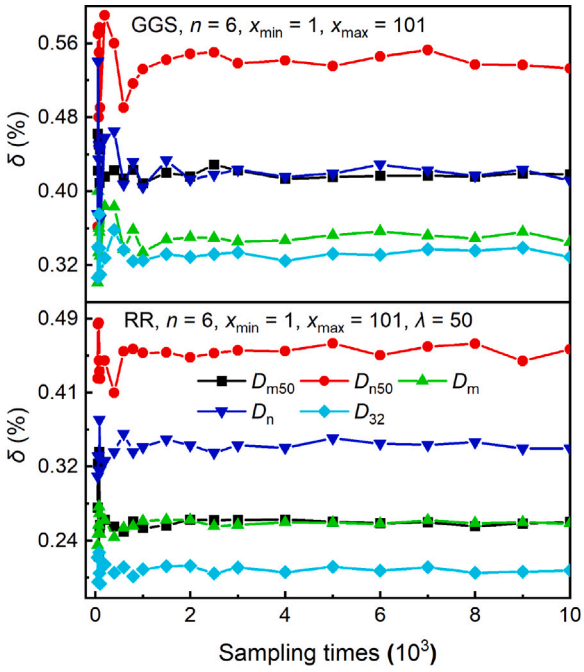

Fig. 6. Effect of sampling times on the estimated errors.

3. Simulation results

3.1. Effect of sample size

Figs. 7 and 8 show the relationship between the sample size N and relative error δ in a logarithmic coordinate for GGS and RR distribution, respectively. The common logarithms of sample size $\log N$ and relative error $\log \delta$ exhibit obvious linear relationships for the five types of mean diameters except for D_{m50}, D_m, and D_{32} when $n = 1.2$, as shown in Figs. 7(a) and 8(a). The fundamental sampling errors calculated by Gy's theory in Appendix B also exhibit similar linear relationship, as shown in Fig. B.1(a).

The linear relationships mean that the relationship between $\log N$ and $\log \delta$ can be represented by a linear function which is confirmed in Section 3.4. Once the expressions for the slope and intercept of the line are obtained, the mathematical relationship between $\log N$ and $\log \delta$ can be determined. Figs. 7 and 8 also imply that the slope of the line may to be a constant, while the intercept is determined by n.

3.2. Effect of span ratios

Figs. 9 and 10 show the relationships between the span ratios σ_m and σ_n and the relative errors δ. The relative errors of D_{m50}, D_m, and D_{32} increase with σ_m, while the relative errors of D_{n50} and D_n increases initially with an increase of σ_m and then decreases, similar to the tendency of σ_n in Fig. 3. The relative errors of the five types of mean diameter are unique for each σ_m. However, for σ_n, a span ratio may result in two error values, as illustrated in Figs. 9(b) and (d) and 10(b) and (d). The results indicate that δ of D_{n50} and D_n is not related to

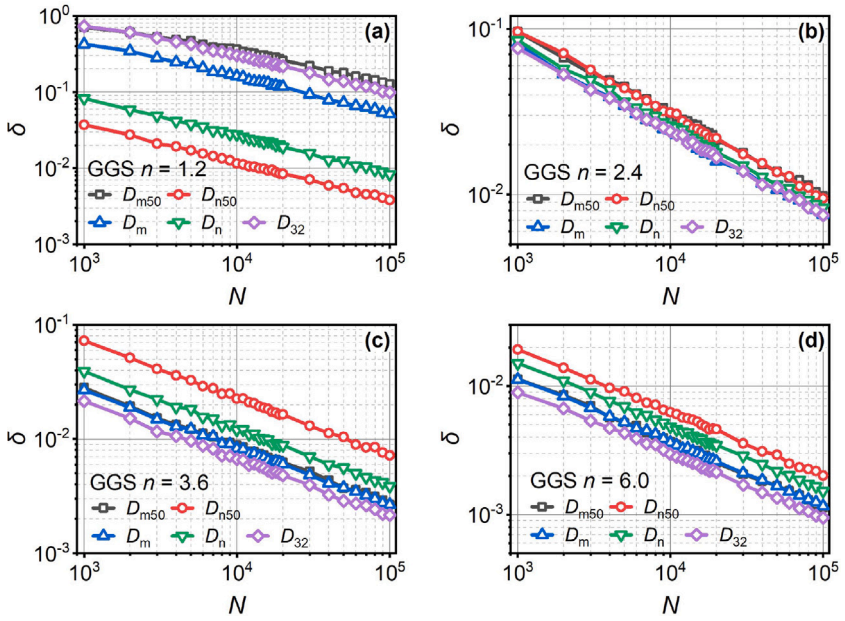

Fig. 7. Particle sample size versus relative errors for GGS distribution. $x_{min} = 1$ and $x_{max} = 101$. (a) $n = 1.2$, (b) $n = 2.4$, (c) $n = 3.6$, and (d) $n = 6.0$.

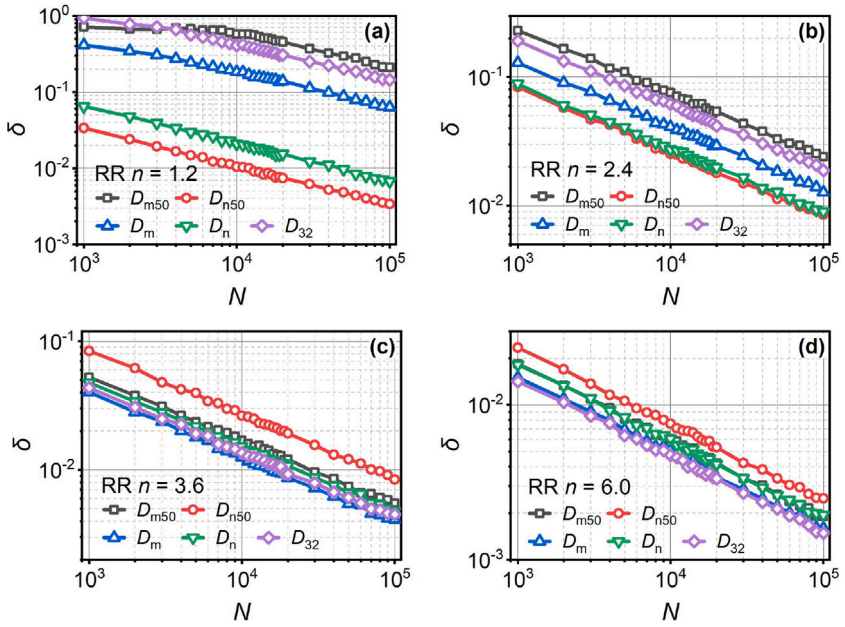

Fig. 8. Particle sample size versus relative errors for RR distribution. $x_{min} = 1$, $x_{max} = 101$, and $\lambda = 50$. (a) $n = 1.2$, (b) $n = 2.4$, (c) $n = 3.6$, and (d) $n = 6.0$.

σ_n strictly. There might be an optimal parameter to characterize the distribution span, which corresponds uniquely to the error value, but this parameter is still unclear.

Since σ_m and δ have a functional relationship, meaning a unique δ corresponds to a given σ_m, σ_m can be used to represent the intercept of the aforementioned linear function.

3.3. Effect of boundary sizes

In Appendix A, analysis indicates that for GGS, in addition to n, the span ratio σ_m is also affected by x_{max}/x_{min}, whereas for RR, σ_m is influenced by both x_{max}/x_{min} and λ/x_{min}. Effects of x_{max}/x_{min} and λ/x_{min} on the relative errors of the five mean diameters were investigated numerically. The sample size N was set to be 10^4 and $x_{min} = 1$.

Fig. 11(a) shows the variation of relative errors for different n as x_{max}/x_{min} increases from 20 to 160 for GGS. The results indicate that for D_{m50}, D_m, and D_{32}, a larger n reduces the impact of x_{max}/x_{min} on the relative errors. When $n = 3.6$ and $n = 6.0$, the errors remain unchanged as x_{max}/x_{min} increases. However, when $n = 1.2$ and $n = 1.6$, this impact is significant. For D_{n50}, the influence of x_{max}/x_{min} is relatively small, particularly when n is either very small or very large. For D_n, the impact of x_{max}/x_{min} can be neglected only when n is relatively large. Similar results are shown in Fig. 11(b) for RR. Fig. 11(c) shows the effect of λ/x_{min} ranging from 20 to 80 on the relative errors for RR. Overall, λ/x_{min} does have a significant impact only when n is very small.

The above results illustrate that errors are influenced by multiple factors, and using span ratio alone may not be sufficient in certain special cases. There might be a more suitable parameter to characterize

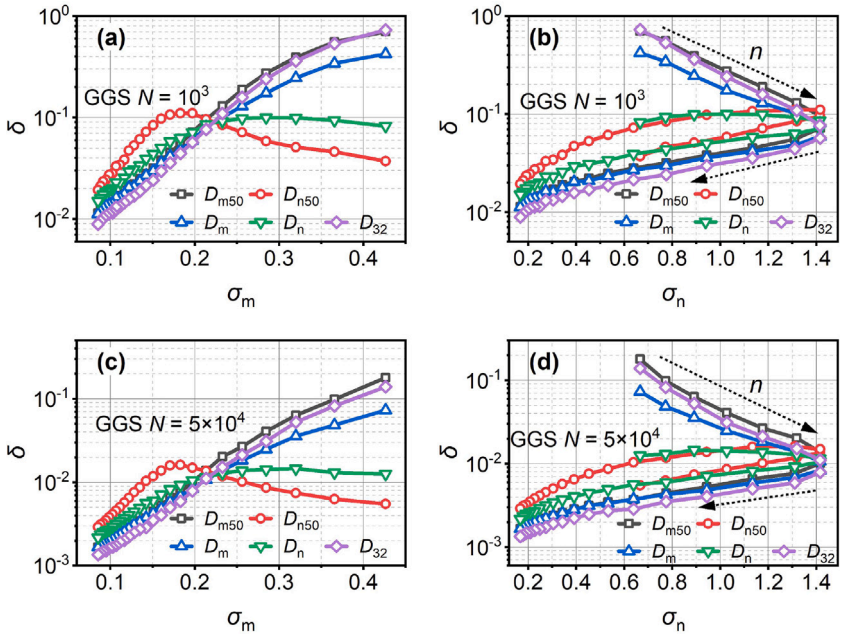

Fig. 9. Span ratio of distribution function versus relative errors for GGS distribution. $x_{min} = 1$ and $x_{max} = 101$. (a) σ_m with $N = 10^3$, (b) σ_n with $N = 10^3$, (c) σ_m with $N = 5\times10^3$, and (d) σ_n with $N = 5\times10^3$.

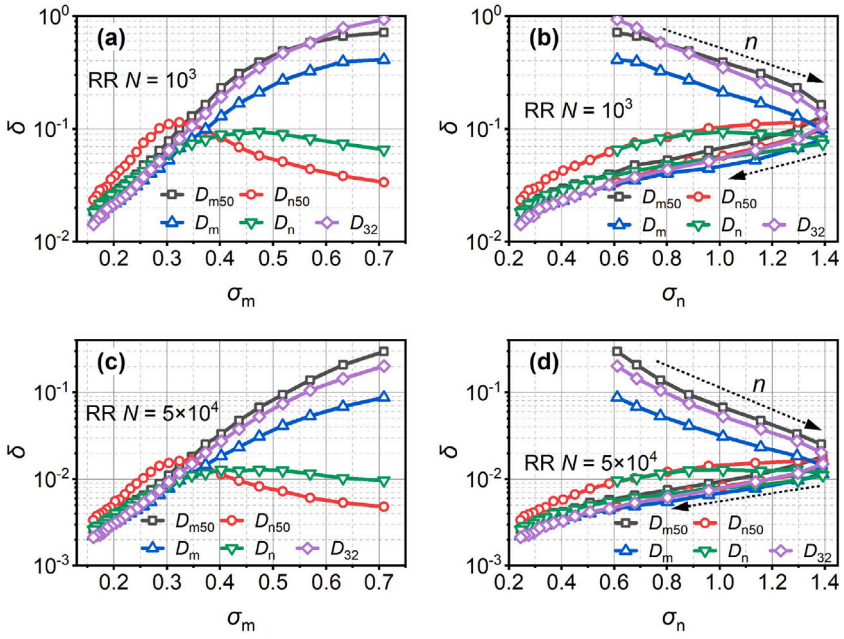

Fig. 10. Span ratio of distribution functions versus relative errors for RR distribution. $x_{min} = 1$, $x_{max} = 101$, and $\lambda = 50$. (a) σ_m and $N = 10^3$, (b) σ_n with $N = 10^3$, (c) σ_m with $N = 5\times10^3$, and (d) σ_n with $N = 5\times10^3$.

the particle size distribution, encompassing both the span ratio and the boundary sizes. However, finding such a parameter could be highly challenging.

3.4. Empirical equations

Based on the influence of sample size and span ratio on the relative error, the error can be empirically represented by a linear function, with the intercept related to the span ratio. Firstly, linear regressions were performed for $\log N$ and $\log\delta$ that $\log N = k\cdot\log\delta + b$. Figs. 12 and 13 show the coefficients k and b and correlation coefficient R^2 of linear regressions. The results are similar for GGS and RR distribution. The coefficient k is observed to be about -2 for the number-based mean diameters D_{n50} and D_n with R^2 larger than 0.998. Actually, $k = -2$ means that δ decreases proportionally to $N^{-0.5}$ [8,11,20,24]. When concerning the mass-based mean diameters, it seems that k is smaller than -2 with R^2 smaller slightly than the others at a small n. Fig. 12(a) shows that k can be regarded as -2 for D_{m50}, D_m, and D_{32} when $n > 1.6$ of GGS distribution. As for RR distribution, $n > 2.0$ facilitates stronger linear relationships, as shown in Fig. 13(a). Then, linear regressions were performed again by using $\log N = -2\cdot\log\delta + b'$. In this way, we obtained new coefficients b' as we fix the value of k to -2. After that, we used Curve Fitting Tool of Matlab to fit the relationship between

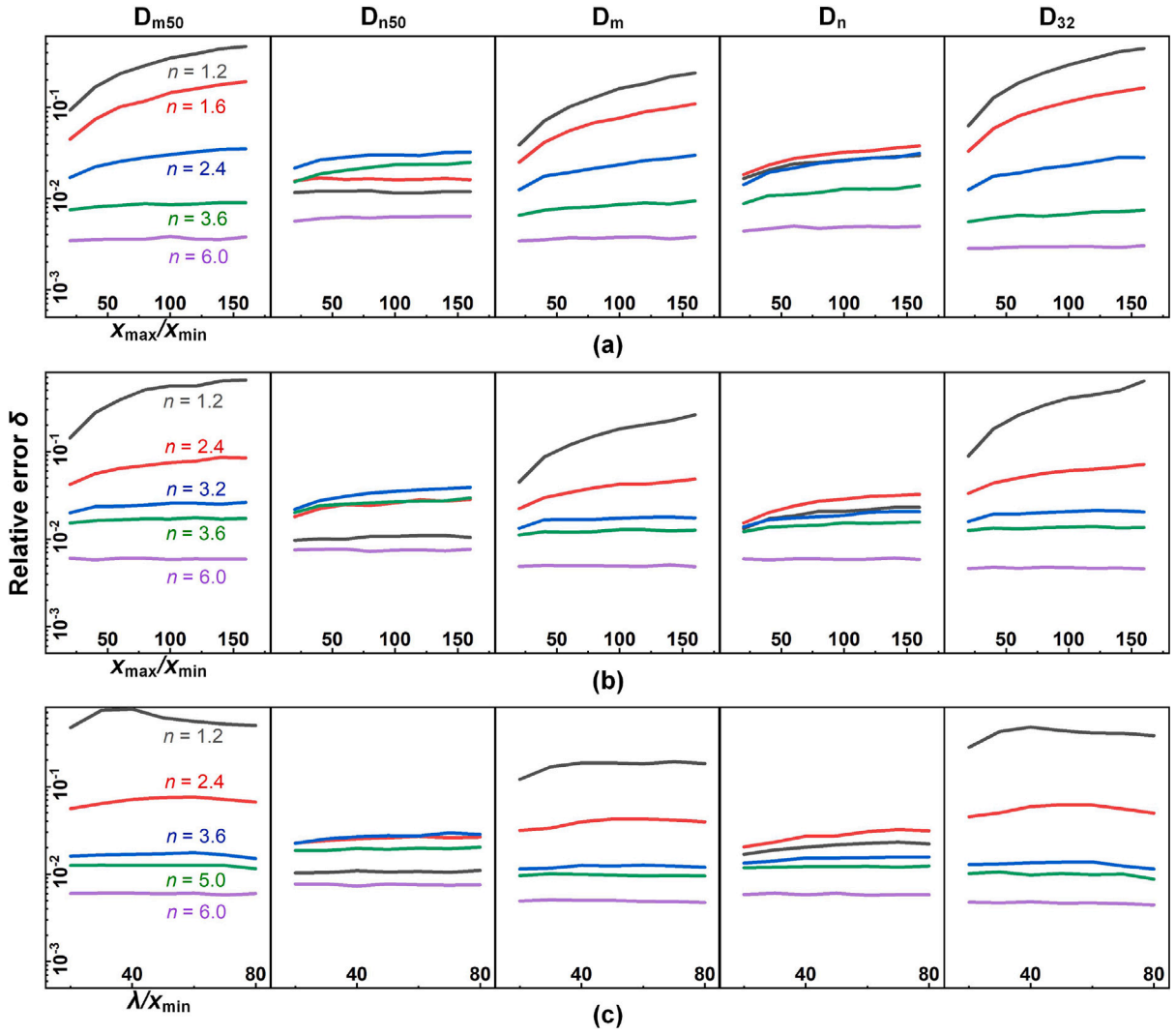

Fig. 11. Effect of x_{max}/x_{min} and λ/x_{min} on relative errors. (a) GGS with $x_{min} = 1$ and $N = 10^4$. (b) RR with $x_{min} = 1$, $\lambda = 50$ and $N = 10^4$. (c) RR with $x_{min} = 1$, $x_{max} = 101$ and $N = 10^4$.

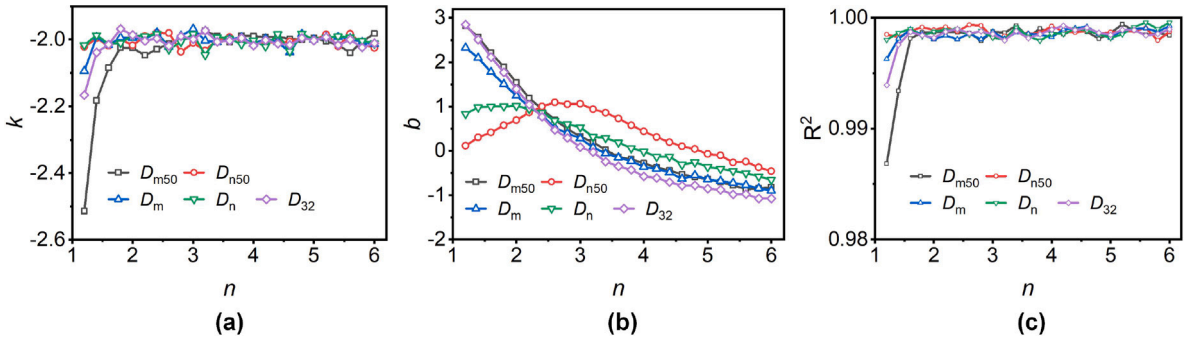

Fig. 12. Linear regressions for GGS distribution. (a) The coefficient k, (b) the coefficient b, (c) the correlation coefficient R^2.

b' and σ_m. The coefficient b is decided by n, indicating that b can be expressed as a function of σ_m. Let $b' = P(\sigma_m)$ denote the function.

Thus, an empirical model can be derived according to the simulations to evaluate the relationship between sample size, relative error, and span ratio at a confidence level of 95%. The empirical formulate is expressed as

$$\log N = -2 \cdot \log \delta + P(\sigma_m) \tag{20}$$

Through testing and comparison, we found that using a cubic polynomial can accurately describe the functional relationship between b' and σ_m for D_{m50}, D_m, D_n, and D_{32}

$$P(\sigma_m) = p_1\sigma_m^3 + p_2\sigma_m^2 + p_3\sigma_m + p_4 \tag{21}$$

For D_{n50}, a rational polynomial showed a better fitting effect

$$P(\sigma_m) = \frac{p_1\sigma_m^2 + p_2\sigma_m + p_3}{\sigma_m^2 + q_1\sigma_m + q_2} \tag{22}$$

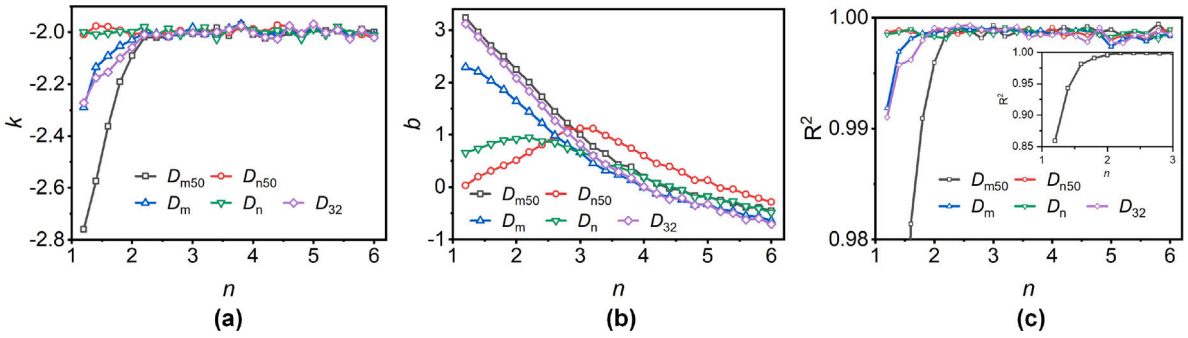

Fig. 13. Linear regressions for RR distribution. (a) The coefficient k, (b) the coefficient b, (c) the correlation coefficient R^2.

Table 2
Fit type and coefficients of $P(\sigma_m)$ for GGS distribution.

Diameter type	Fit type	Adjusted R^2	Fit coefficients
D_{m50}	Poly3	0.9996	[−65.75, 35.32, 8.322, −1.805]
D_{n50}	Rat22	0.9987	[−0.3541, 0.2342, −0.02097, −0.2705, 0.02528]
D_m	Poly3	0.9997	[−30.53, 7.857, 12.44, −1.978]
D_n	Poly3	0.9987	[58.57, −75.36, 29.75, −2.679]
D_{32}	Poly3	0.9995	[−79.52, 46.92, 5.618, −1.84]

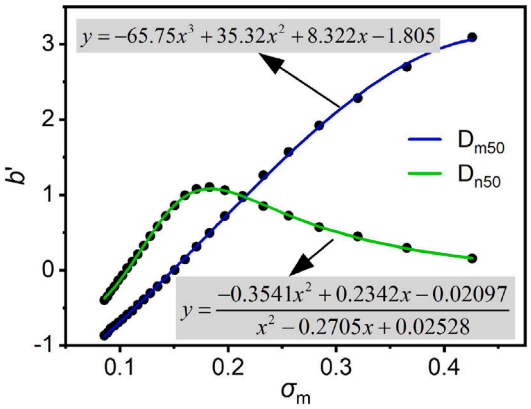

Fig. 14. Fitting the function relationship between b' and σ_m. D_{m50} and D_{n50} are taken as two examples and the distribution is GGS.

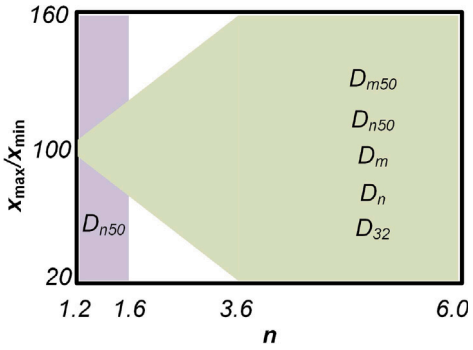

Fig. 15. Applicable ranges of the empirical equations.

where p_1, p_2, p_3, p_4, q_1, and q_2 are fitting coefficients. As shown in Fig. 14, twenty-five points were used to perform curve fitting. All fitting results are summarized in Tables 2 and 3 for GGS and RR distributions. The adjusted correlation coefficients R^2 are larger than 0.995 for all of the polynomial fitting.

It should be emphasized that these empirical formulas are not universally applicable to all scenarios. According to the simulation results in Fig. 11, boundary sizes also affect the errors, whereas the empirical formulas are derived based on $x_{max}/x_{min} \approx 100$. Therefore, when x_{max}/x_{min} deviates from 100, the applicability needs to be judged based on the value of n. Fig. 15 provides the approximate applicable ranges. When n is greater than 3.6, the empirical formulas proposed in this paper are generally applicable to all mean particle sizes. When n is less than 1.6, they are generally applicable to D_{n50}. For n in the range of 1.2 to 3.6, the formulas are applicable near $x_{max}/x_{min} = 100$. Moreover, as n increases, the applicable range of x_{max}/x_{min} expands, provided a certain accuracy requirement is met. The blank areas are inapplicable regions where using the empirical formulas may lead to misleading error assessments.

4. Experimental verification

PSD data of pulverized coal powder whose mass-weighted CDF is commonly described by RR distribution, is analyzed to verify the proposed method in this work. Unfortunately, there is no practical data of GGS distribution. The size data of coal powder were obtained on a coal-fired power plant by an on-line instrument, Digital Holographic Particle Analyzer based on holographic imaging. Each measurement data has been stored locally such that we can easily retrieve the raw data for extensive analysis. Details of the measurement technique and instrument can be found in the literature [7]. In online particle size measurement, an important issue to consider is how many particles need to be measured to obtain a reliable data point.

Three coal particle populations were utilized for verification. The first sample contained about 2×10^6 particles while the other two samples contained about 3×10^6 particles. The raw mass-weighted CDFs of the coal samples were first fitted based on the doubly-truncated RR distribution with *fitnlm* function in Matlab. The distribution parameters of the coal samples are summarized in Table 4. The spread parameter n of the three samples were 1.91, 1.76, and 1.59, respectively. Among the three samples, Coal 3 had the largest σ_m but the smallest σ_n. Fig. 16 shows the mass CDFs of the raw experimental data and fitted data. The raw CDFs match quite well with the fitted curves.

Then artificial populations for each coal samples were generated according to the parameters in Table 4. The theoretical errors for the five types of mean diameters were evaluated with the numerical method introduced in this study. Meanwhile, the experimental errors of the raw coal particle populations were also evaluated with the same method, which represent the uncertainty region in practical measurements.

Table 3
Fit type and coefficients of $P(\sigma_m)$ for RR distribution.

Diameter type	Fit type	Adjusted R^2	Fit coefficients
D_{m50}	Poly3	0.9996	[−17.02, 15.51, 4.652, −1.565]
D_{n50}	Rat22	0.9990	[−0.3894, 0.3697, −0.05594, −0.5386, 0.09024]
D_m	Poly3	0.9996	[−10.79, 7.594, 6.044, −1.76]
D_n	Poly3	0.9968	[15.12, −30.98, 19.31, −2.881]
D_{32}	Poly3	0.9994	[−16.57, 14.29, 5.324, −1.854]

Table 4
Distribution parameters of three coal samples. Particle size unit: μm.

Coal	Quantity	x_{min}	x_{max}	λ	n	σ_m	σ_n
1	1,946,812			97.37	1.91	1.66	2.24
2	3,148,998	6	350	74.51	1.76	1.73	2.04
3	3,187,563			93.03	1.59	1.83	1.99

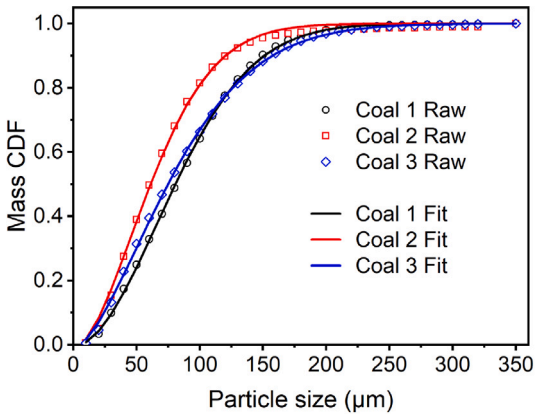

Fig. 16. Mass-weighted CDFs of the three pulverized coal samples measured by holographic imaging and corresponding distribution curves fitted using doubly-truncated RR function (Eq. (5)).

Fig. 17 depicts comparison between the theoretical errors and experimental errors at 95% confidence level for various sample sizes. The black and red solid lines represent the theoretical and experimental results, respectively. Each gray dot represents the result of a single random sampling from a raw coal particle population. It can be observed from Fig. 17 that the experimentally obtained errors are either smaller than or close to the theoretical error lines for D_{m50}, D_m, D_n, and D_{32}. This indicates that the simulation method proposed in this paper can predict the relationship between relative errors at 95% confidence level and sample size well for these several mean diameters. However, for D_{n50}, the opposite phenomenon, that the experimental errors are 1.5 to 2 times larger than the theoretical errors, occurred. This could be due to the measurement principle of the imaging method. The imaging method calculates particle size based on the number of pixels occupied by particle images, resulting in a series of particle size results with large intervals in the small particle size region. For example, for the instrument used in this study, the imaging pixel size is 5.5 μm, resulting in particle size measurements such as 6.21, 8.78, 10.75, 12.41, 13.88, and so on. This makes it difficult to accurately obtain the number-weighted CDF, leading to an increase in the error of D_{n50}, while relatively continuous size values can be produced in simulations. However, this has a smaller impact on other mean diameters.

5. Conclusions

For GGS and RR, two mass-weighted cumulative distribution functions, a simulation method has been developed to estimate the particle sample size required for a reliable mean diameter measurement at a specific confidence level. Five types of mean diameter including mass median diameter D_{m50}, number median diameter D_{n50}, mass mean diameter D_m, number mean diameter D_n, and Sauter diameter D_{32}, were studied in this work. Effects of sample size, span ratio, and boundary sizes on the relative errors have been investigated. Moreover, the proposed method was verified by experimental size data of pulverized coal powder. The main conclusions are as follows:

(1) A strong linear relationships between the common logarithms of sample size $\log N$ and relative error $\log \delta$ have been observed. It was also found that the mass-based span ratio σ_m may be more appropriate to establish the error model than the number-based span ratio σ_n since same σ_n may lead to different error results. The boundary sizes have a significant impact on the errors when the distribution parameter n is small, while the influence can be neglected when $n > 3.6$.

(2) An empirical formula $\log N = -2 \cdot \log \delta + P(\sigma_m)$ has been proposed to describe the relationship between the sample size N, span ratio σ_m, and relative error δ. The expressions of $P(\sigma_m)$ were determined by polynomial fitting. The applicable regions of the empirical equations were recommended.

(3) The experimental verification showed that the proposed method can provide reliable prediction on the relative errors of D_{m50}, D_m, D_n, and D_{32}. The relative errors of D_{n50} could be overestimated due to the principle of imaging-base sizing technique.

CRediT authorship contribution statement

Qiwen Jin: Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Zhiming Lin:** Software, Writing – review & editing. **Yingchun Wu:** Conceptualization, Software, Writing – review & editing. **Xuecheng Wu:** Conceptualization, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge the support from the Zhejiang Provincial Natural Science Foundation of China (No. ZCLQ24E0601), the National Natural Science Foundation of China (No. 52276167) and the Fundamental Research Funds for the Central Universities, China (2022ZFJH004).

Appendix A. Properties of span ratios

Let $h(n)$ denote D_{m84}/D_{m50} as a function of n while $x_{max} = m_1 x_{min}$ and $\lambda = m_2 x_{min}$ with $1 < m_2 < m_1$. For GGS, D_{m84}/D_{m50} can be expressed as following by solving Eq. (1)

$$h_{GGS}(n) = \frac{D_{m84}}{D_{m50}} = \frac{0.84^{1/n}(m_1 - 1) + 1}{0.50^{1/n}(m_1 - 1) + 1} \tag{A.1}$$

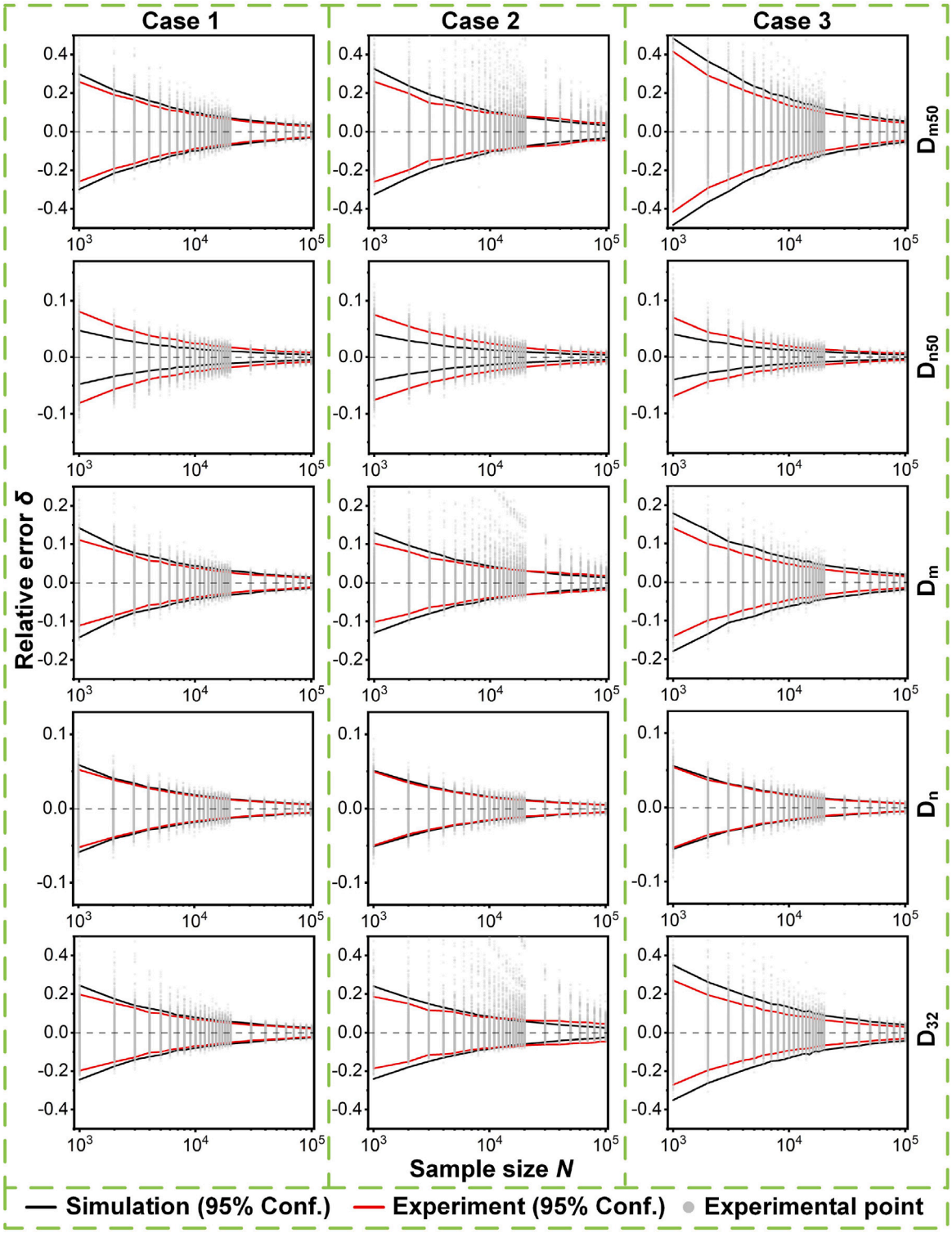

Fig. 17. Comparison between the theoretical errors obtained through simulation and the experimental values. Black solid lines: theoretical errors at 95% confidence level. Red solid lines: experimental errors at 95% confidence level. Gray dots: error values obtained from each sampling of the experimental particle data.

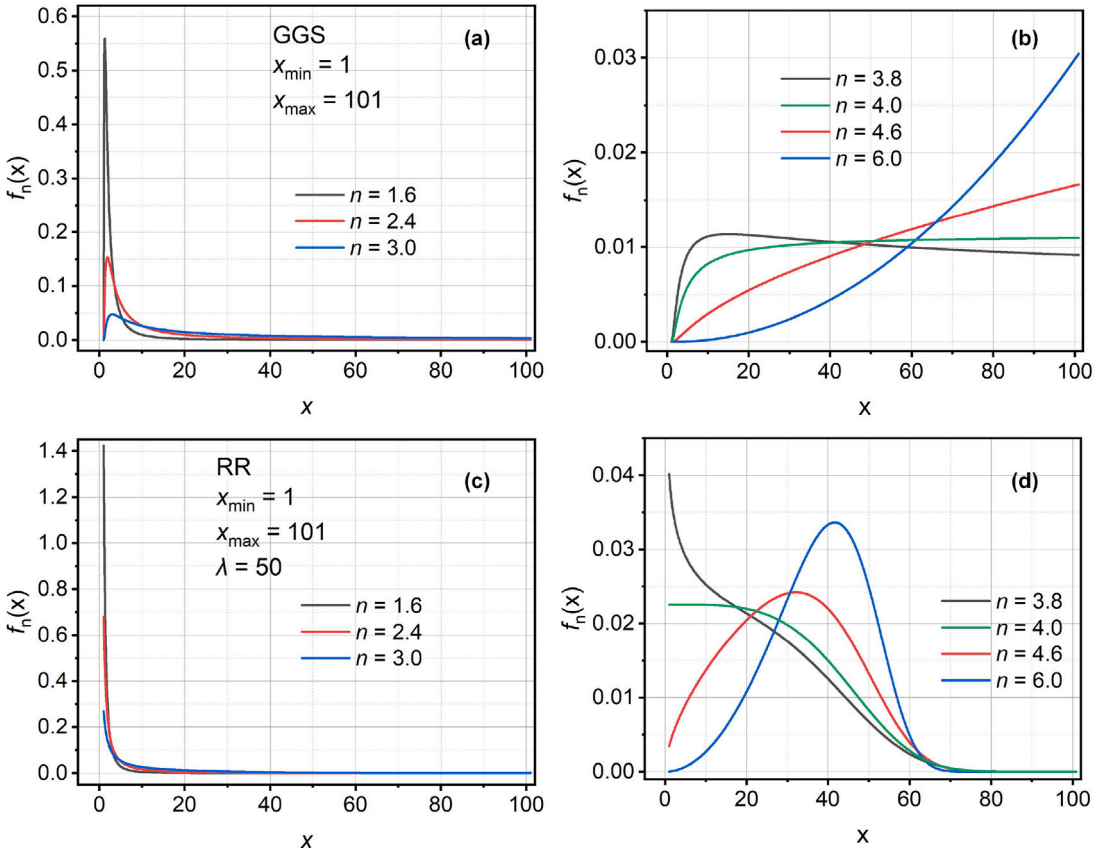

Fig. A.1. Number-weighted PDFs of GGS and RR distribution as n increases from a small to a large value. (a) and (b) GGS. (c) and (d) RR.

Similarly, for RR, D_{m84}/D_{m50} is expressed as

$$h_{RR}(n) = \frac{D_{m84}}{D_{m50}} = \left[\frac{\ln\left(0.16e_1 + 0.84e_2\right)}{\ln\left(0.50e_1 + 0.50e_2\right)}\right]^{1/n} \tag{A.2}$$

where $e_1 = \exp\left[-\left(\frac{1}{m_2}\right)^n\right]$, $e_2 = \exp\left[-\left(\frac{m_1}{m_2}\right)^n\right]$. $h_{GGS}(n)$ and $h_{RR}(n)$ are monotonically decreasing functions. Thus σ_m decreases as n increases for GGS and RR, as shown in Fig. 3(a).

However, the mathematical expressions of number-based D_{n84}/D_{n50} cannot be provided since the inverse function of the hypergeometric function in Eq. (3) and the upper incomplete gamma function in Eq. (7) are difficult to solve. D_{n84} and D_{n50} are solved numerically in our code using the Matlab function *vpasolve*.

The reason for the span ratio based on number distribution initially increasing and then decreasing with n can possibly be explained by the characteristics of number weighted PDFs of GGS and RR distributions.

For GGS, the number-weighted PDF is written as Eq. (4). The monotonicity of the functions $g(x) = \left(x - x_{min}\right)^{n-1} x^{-3}$ and $f_n(x)$ is consistent since the other terms are constants. The derivative of $g(x)$ is expressed as

$$g'(x) = \frac{\left[(n-4)\,x + 3x_{min}\right]\left(x - x_{min}\right)^{n-2}}{x^4} \tag{A.3}$$

The monotonicity of the function $f_n(x)$ can be determined by the sign of $g'(x)$. When $n \geq 4$, $f_n(x)$ is monotonically increasing. When $n < 4$, $f_n(x)$ increases in the range $\left[x_{min}, \frac{3x_{min}}{4-n}\right]$ and decreases in the range $\left[\frac{3x_{min}}{4-n}, x_{max}\right]$, indicating the presence of a maximum of $f_n(x)$ at $x = \frac{3x_{min}}{4-n}$.

For the number-weighted PDF of RR (Eq. (8)), let $g(x) = x^{n-4}\exp\left[-\left(\frac{x}{\lambda}\right)^n\right]$. The derivative of $g(x)$ is expressed as

$$g'(x) = \left(n - 4 - \frac{n}{\lambda^n}\right) x^{n-5} \exp\left[-\left(\frac{x}{\lambda}\right)^n\right] \tag{A.4}$$

Similarly, $f_n(x)$ is monotonically decreasing when $n \leq 4$. When $n > 4$, $f_n(x)$ increases in the range $\left[x_{min}, \sqrt[n]{\frac{n-4}{n}}\lambda\right]$ and decreases in the range $\left[\sqrt[n]{\frac{n-4}{n}}\lambda, x_{max}\right]$, indicating the presence of a maximum of $f_n(x)$ at $x = \sqrt[n]{\frac{n-4}{n}}\lambda$.

Fig. A.1 shows the number-weighted PDFs of GGS and RR with various n which are consistent with monotonicity analysis. The above analysis indicates that when n is small, the number distribution is concentrated in the smaller particle sizes, and when n is large, the number distribution is concentrated in the larger particle sizes. However, when n is at an intermediate value, the number distribution is more evenly spread across the entire range, suggesting a larger σ_n. It should be noted that $n = 4$ serves as the boundary between different monotonicities, but it does not correspond to the maximum σ_n.

Appendix B. Evaluating sampling error with Gy's theory

Theory of sampling, developed by Pierre Gy [29], holds a pivotal and influential place in the field of sampling, especially in industries such as mining, soil science, and agriculture. Gy's sampling theory introduced a comprehensive mathematical framework to understand and manage sampling errors. Gy has systematically studied the sampling of complex particle mixtures and derived an equation that describe the relationship between the relative variance of the fundamental sampling error S^2_{FSE} and the physico-chemical properties of the particles [30].

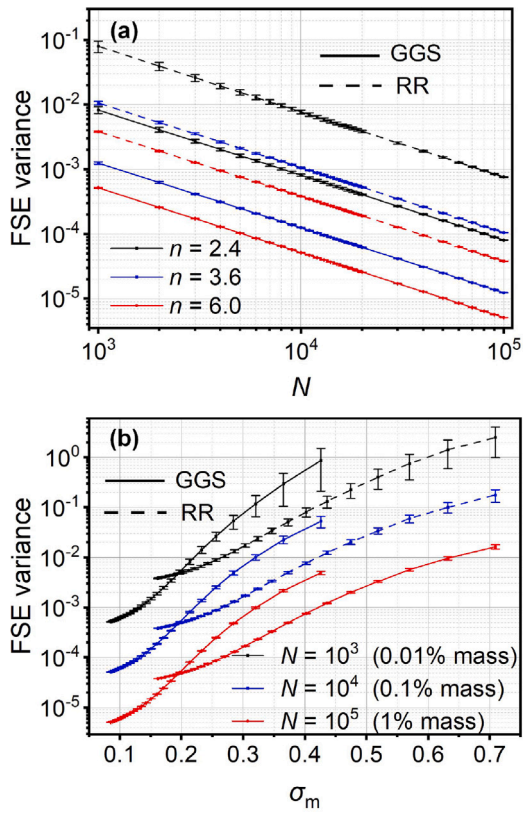

Fig. B.1. Effects of (a) sample size N and (b) span ratio σ_m on the fundamental sampling error for GGS and RR distribution based on Gy's sampling theory. $x_{min} = 1$, $x_{max} = 101$ and $\lambda = 50$.

The famous Gy's formula is written as

$$S_{FSE}^2 = fg\beta cd^3\left(\frac{1}{M_S} - \frac{1}{M_L}\right) \tag{B.1}$$

where d is the characteristic particle size = 95% limit of the size distribution $(d = D_{m95})$. M_S is the sample mass and M_L is the lot mass. f is the shape factor describing the deviation from the ideal shape of a square. For spherical particles $f = 0.5$, which often can be used as the default value. g is the size distribution factor describing the span of particle sizes in the lot. $g = 0.25$ for wide-size distribution, and $g = 1$ for uniform particle sizes. β is the liberation factor describing the degree of liberation of the critical component from the matrix. $\beta = 1$ for totally liberated particles. c is the constitution factor which can be calculated from the following formula

$$c = \frac{\left(1 - \frac{a_L}{\alpha}\right)^2}{\frac{a_L}{\alpha}}\rho_c + \left(1 - \frac{a_L}{\alpha}\right)\rho_m \tag{B.2}$$

where a_L represents the average concentration of the lot. α is the concentration of the analyte in the critical particles. ρ_c is the density of the critical particles. ρ_m is the density of the matrix or diluent particles. In this work, particles are assumed to be spherical and totally liberated. The densities of the critical particles and the matrix are 1.5 g/cm^3. The concentration of the lot is assumed to be 0.1 and the concentration of the analyte in the critical particles is 1. Note that those factors may require experimental estimation in practical applications.

Simulations were conducted based on Gy's sampling theory under consistent conditions using the method proposed in this paper. In each simulation, the sampling times was set to 2000. Fig. B.1 shows the effects of sample size N and span ratio σ_m on S_{FSE}^2 for GGS and RR distribution. In Fig. B.1(a), logarithms of FSE variance and sample size exhibit a perfect linear relationship. If straight lines are fitted to the data, the slopes are found to be 1, which is consistent with the relationship described in Eq. (B.1). Fig. B.1(b) illustrates the variation of the FSE variance under different span ratios (i.e., different distribution modulus n). The results indicate that the wider the distribution, the greater the FSE variance, necessitating a larger sample size (in terms of mass or quantity).

References

[1] Henk G. Merkus, Particle Size Measurements: Fundamentals, Practice, Quality, Springer, 2009.
[2] Stanley J. Dapkunas, A. Jillavenkatesa, L.H. Lum, Particle Size Characterization, NIST Recommended Practice Guide, 2001.
[3] Henk G. Merkus, Sampling errors in particle size analysis? Part. Part. Syst. Charact. 24 (1) (2007) 34–39.
[4] Mahmut Camalan, The use of non-parametric tests between subsamples and particle population for the assessment of minimum number of particles in microscopic analysis, Particul. Sci. Technol. 38 (6) (2020) 703–710.
[5] James Clarke, John F Gamble, John W Jones, Mike Tobyn, Richard Greenwood, Andy Ingram, Alternative approach for defining the particle population requirements for static image analysis based particle characterization methods, Adv. Powder Technol. 30 (5) (2019) 920–929.
[6] Lajos Madarász, Ákos Köte, Bence Hambalkó, Kristóf Csorba, Viktor Kovács, László Lengyel, György Marosi, Attila Farkas, Zsombor Kristóf Nagy, András Domokos, In-line particle size measurement based on image analysis in a fully continuous granule manufacturing line for rapid process understanding and development, Int. J. Pharm. 612 (2022) 121280.
[7] Qiwen Jin, Lei Zeng, Xijiong Chen, Pei Li, Haiyuan Fu, Yonggang Zhou, Yingchun Wu, Xuecheng Wu, Portable digital holographic particle analyzer (DHPA) for pneumatically conveyed fuel monitoring: Design and validation, Powder Technol. 430 (2023) 119030.
[8] Hiroaki Masuda, Koichi Iinoya, Theoretical study of the scatter of experimental data due to particle-size-distribution, J. Chem. Eng. Japan 4 (1) (1971) 60–66.
[9] Hiroaki Masuda, Kuniaki Gotoh, Study on the sample size required for the estimation of mean particle diameter, Adv. Powder Technol. 10 (2) (1999) 159–173.
[10] Organización Internacional de Normalización, ISO 13322-1: 2014, Particle Size Analysis, Image Analysis Methods. Static Image Analysis Methods, ISO, 2014.
[11] Procedure for predicting a minimum volume or mass of sample to provide a given size parameter precision, Part. Part. Syst. Charact. 18 (3) (2001) 109–113.
[12] A Macıas-Garcıa, Eduardo M Cuerda-Correa, MA Dıaz-Dıez, Application of the Rosin–Rammler and Gates–Gaudin–Schuhmann models to the particle size distribution analysis of agglomerated cork, Mater. Charact. 52 (2) (2004) 159–164.
[13] Maarten Alderliesten, Mean particle diameters. Part VII. The Rosin-Rammler size distribution: Physical and mathematical properties and relationships to moment-ratio defined mean particle diameters, Part. Part. Syst. Charact. 30 (3) (2013) 244–257.
[14] Hideto Yoshida, Tatsuo Igushi, Tetsuya Yamamoto, Hiroaki Masuda, Theoretical calculation of fundamental uncertainty region based on the maximum and/or the minimum size in the preparation of standard reference particles for particle size measurement, Adv. Powder Technol. 22 (1) (2011) 43–49.
[15] Hideto Yoshida, Tetsuya Yamamoto, Kunihiro Fukui, Hiroaki Masuda, Theoretical calculation of uncertainty region based on the general size distribution in the preparation of standard reference particles for particle size measurement, Adv. Powder Technol. 23 (2) (2012) 185–190.
[16] Hideto Yoshida, Hirosuke Sugasawa, Funihiro Fukui, Tetsuya Yamamoto, Hiroaki Masuda, Theoretical calculation of uncertainty region for spherical particles based on a picket fence, quasi-monodisperse particles, Adv. Powder Technol. 25 (2) (2014) 524–529.
[17] Hideto Yoshida, Yasushige Mori, Hiroaki Masuda, Tetsuya Yamamoto, Particle size measurement of standard reference particle candidates and theoretical estimation of uncertainty region, Adv. Powder Technol. 20 (2) (2009) 145–149.
[18] Yoshiyuki Endo, Estimate of confidence intervals for geometric mean diameter and geometric standard deviation of lognormal size distribution, Powder Technol. 193 (2) (2009) 154–161.
[19] Tatsushi Matsuyama, Estimation of uncertainty of percentile values in particle size distribution analysis as a function of number of particles, Adv. Powder Technol. 30 (11) (2019) 2616–2619.
[20] Anthony J. Paine, Error estimates in the sampling from particle size distributions, Part. Part. Syst. Charact. 10 (1) (1993) 26–32.
[21] Mahmut Camalan, Simulating probabilistic sampling on particle populations to assess the threshold sample sizes for particle size distributions, Particul. Sci. Technol. 39 (4) (2021) 511–520.
[22] E. Vigneau, C. Loisel, M.F. Devaux, P. Cantoni, Number of particles for the determination of size distribution from microscopic images, Powder Technol. 107 (3) (2000) 243–250.
[23] Catherine L. Evans, Timothy J. Napier-Munn, Estimating error in measurements of mineral grain size distribution, Miner. Eng. 52 (2013) 198–203.

[24] Tatsushi Matsuyama, An application of bootstrap method for analysis of particle size distribution, Adv. Powder Technol. 29 (6) (2018) 1404–1408.
[25] Mahmut Camalan, Estimating the number-weighted equivalents of the mass-weighted size distribution functions, Powder Technol. 369 (2020) 106–113.
[26] Bernard D. Flury, Acceptance–rejection sampling made easy, Siam Rev. 32 (3) (1990) 474–476.
[27] Macaulay Okwuokenye, Karl E. Peace, A comparison of inverse transform and composition methods of data simulation from the lindley distribution, Commun. Stat. Appl. Methods 23 (6) (2016) 517–529.
[28] Maarten Alderliesten, Mean particle diameters. Part VI: Fundamental distinction between statistics based (ISO/DIN) and physics based (moment-ratio) definition systems, Part. Part. Syst. Charact. 27 (1–2) (2010) 7–20.
[29] Pierre Gy, Sampling of Particulate Materials Theory and Practice, Elsevier, 2012.
[30] Lars Petersen, Pentti Minkkinen, Kim H. Esbensen, Representative sampling for reliable data analysis: theory of sampling, Chemometr. Intell. Lab. Syst. 77 (1–2) (2005) 261–277.