https://doi.org/10.1016/j.apm.2019.11.045

Error analysis in stochastic solutions of population balance equations

Kun Zhou[1], Xiao Jiang[2], and Tat Leung Chan[2*]

[1]The State Key Laboratory of Refractories and Metallurgy,
Wuhan University of Science and Technology, Wuhan, China

[2]Department of Mechanical Engineering, The Hong Kong Polytechnic University,
Kowloon, Hong Kong.

*Corresponding author & Email : mmtlchan@polyu.edu.hk (T. L. Chan)

Abstract

Stochastic simulation of population balance equations (PBEs) is r obust and flexible; however, it exhibits intrinsic stochastic errors which decreases at a very slow rate when increasing the computational resolution. Generally, these stochastic methods can be classified into two groups: (i) the classical Gillespie method and (ii) weighted flow algorithm. An analytical relationship is derived for the first time to connect the variances in these two groups. It also provides a detailed analysis of the resampling process, which has not been given appropriate attention previously. It is found that resampling has a profound effect on the numerical precision. Moreover, by comparing the time evolutions between systematic errors (i.e., errors in the mean value) and stochastic errors (i.e., variances), it is found that the former grows considerably faster than the latter; thus, systematic errors eventually dominate. The present findings facilitate the choice of the most suitable stochastic method for a specific PBE a priori in order to balance numerical precision and efficiency.

Keywords: Population balance equations, Smoluchowski equation, Stochastic methods, Weighted flow algorithm, Stochastic variance, Aerosol dynamics

1. Introduction

The Smoluchowski coagulation equation [1] describes the time evolution of the number density of particles in coagulation. Its extension to a general population balance equation (PBE) [2] has been used in diverse fields, such as aerosol dynam-

ics [3] (which is the main focus of this study) and chemical kinetics [4]. Because of its mathematical complexity (a partial integral-differential equation), analytical solutions to the PBE are only available under limited conditions [5–7]. Numerical techniques are by far the most feasible for solving the PBEs. In general, numerical techniques can be classified into three groups: direct discretization, method of moments, and stochastic methods. Direct discretization may be applicable in the particle size space [8–11] or the functional space [12, 13]. Earlier discretization methods typically encounter the problems of non-positiveness and non-conservation [14]; considerable efforts have been exerted to resolve such problems for better numerical efficiency and flexibility [15, 16]. The method of moments solves a group of moment equations derived from the PBE. It is considerably efficient in providing fundamental statistics pertaining to the particle size distribution (PSD) function (i.e., number density, volume fraction, polydispersity, etc.) without directly solving the PSD. However, closure schemes [17–22] are necessary to deal with the closure problem because of the truncation involved in deriving the moment equations. Apart from the compromise in the numerical precision, the realizability problem (which occurs when the moments sequence fails to find its corresponding distribution function) is also an obstacle [23, 24]. Moreover, the method of moments is not capable of solving the PSD. There have been some research studies conducted to recover the PSD from a group of known moments [25–27]. However, from a pure mathematical point of view, it is a well-known ill-conditioned problem to recover a distribution from its moments [23], and more research studies need to be done.

The stochastic or Monte Carlo (MC) simulation does not solve the PBE; however, it mimics the evolution of the physical particulate system with a cluster of notational particles. All the MC simulations are based on the Marcus-Lushnikov process for modeling coagulation [28] and are extended to include general aerosol dynamic processes through various techniques [29–31]. In an MC simulation, the typical processes of birth (nucleation), growth (condensation and surface reaction), and binary coagulation in the PBE are modeled simply by adding new particles, growing existing

particles, and having two particles collide, respectively. Such type of simulation is considerably robust; however, intrinsic stochasticity and slow convergence rate (with respect to the number of numerical particles, N) are the main deficiencies of stochastic simulation. Increasing numerical efficiency (accelerating the numerical simulation) has been one of the central focuses in the study of stochastic methods. Such efficiency is generally achieved through several directions. (i) Using the MC for multidimensional population balance problems [32–42]. It would be fair to point out that the MC for a multidimensional problem is more computationally expensive than that for a one-dimensional problem; however, the increase in computational cost is considerably smaller than those of the other methods. There are several conditions where a single variate PSD is not sufficient, and a multicomponent or multiparameter description [43] of the joint is necessary. For such multidimensional problems, the MC simulation is superior in tackling "the curse of dimensionality" [44], i.e., the computation grows exponentially with the number of dimensions when the usual deterministic methods are used. On the other hand, the convergence rate in the MC simulation is not affected by the number of dimensions. (ii) Adopting different weights for the different sizes of numerical particles. In an MC simulation, numerical particles represent a large number of physical particles with similar properties. If every numerical particle represents the same number of physical particles (i.e., equal weights), then large particles, whose number density is typically considerably lower than that of small particles, are generally poorly represented (or not represented at all) in the numerical simulation. In order to simulate large particles more precisely, various differentially weighted schemes [37, 40, 45–48] have been developed to increase their resolution. (iii) Introducing proper approximations. Grouping similar particles with similar properties (e.g., size and component) into "bins" [49, 50], firing multiple coagulation events at a time (τ leaping) [51, 52], and approximating complicated coagulation kernels with simple ones by introducing fictitious jumps [53, 54] (it should be noted that the fictitious jump method does not cause approximations in the final results.) are all efficient approaches in accelerating an MC simulation.

(iv) Running numerical simulations in parallel. Parallelization can be implemented on multiple central processing unit processors [31]; it is even more economically implemented on a graphics processing unit [55]. A promising but has yet to be realized approach in the MC simulation of PBE is the introduction of correlations across replications in parallel simulations in order to reduce the simulation variance [56].

In the present study, both systematic and stochastic errors in the stochastic simulations of the generalized Smoluchowski coagulation equation are investigated analytically and numerically for the classical Monte Carlo (CMC) algorithm [28] and the general differentially weighted algorithm (termed weighted flow algorithm, WFA) [37]. Because any MC method is supposed to converge in the mean to the true solution of PBE, the variance can be regarded as a de facto standard to quantify the numerical efficiency of an MC method. The present study is intended to quantify under what conditions the WFA is superior over the CMC in terms of the variance. Another highly relevant but always ignored problem is how resampling affects the PSD [57]. For a general aerosol dynamic system, the number density of particles typically varies over a considerably wide range. Inevitably, the number of numerical particles, N, has to be curved within a proper range in order to avoid extremely high computational costs (large N) or extremely high simulation errors (small N). Resampling refers to the addition or removal of particles from the original distribution in order to obtain a new one with a proper N, while ensuring that the original and new particle size distributions remain statistically the same. Although several ad hoc resampling schemes [30, 58] have been used in MC simulations, the impact of resampling has never been investigated thoroughly. This impact is actually considerably important under several conditions and may even be the decisive factor under some extreme conditions. The present study also provides a proven formula to evaluate the stochastic variance caused by the resampling.

This paper is organized as follows: Section 2 briefly introduces the CMC and WFA methods. Section 3 first reviews the previous theory regarding the Poisson distribution of a specific particle size. Thereafter, the derivation of variance in the

two algorithms for simple coagulation on the premise of the Poisson distribution is introduced. Next, the analysis of variance in the resampling, which is a necessary step in the stochastic simulation, is presented. Section 4 explains the various numerical simulations for verifying the theoretical analyses. An appendix also accompanies this paper to show that although resampling does not introduce any bias to the PSD, it generates variance in the PSD.

2. Smoluchowski Equation and Stochastic Simulations

2.1. Classical Monte Carlo (CMC) method for Smoluchowski equation

The continuous Smoluchowski coagulation equation is

$$\frac{\partial n(v)}{\partial t} = \frac{1}{2}\int_0^v \beta(u, v-u)n(u)n(v-u)\,\mathrm{d}u - \int_0^\infty \beta(v,u)n(v)n(u)\,\mathrm{d}u, \tag{1}$$

where $n(v)$ is the number density of particles with volume v, and $\beta(u,v)$ is the collision kernel function, which describes the coagulation rate of two particles with volumes u and v. Throughout this paper, the time variable in expressions is omitted for conciseness. The continuous coagulation equation is convenient to include more aerosol dynamic processes (i.e., nucleation, condensation, etc.) in order to obtain the general dynamic equation [3]. However, the discrete coagulation equation is a more natural choice in MC simulations

$$\frac{\partial n_l}{\partial t} = \frac{1}{2}\sum_{I+J=l} \beta(I,J)n_I n_J - \sum_{I=1}^{\infty} \beta(I,l)n_I n_J, \tag{2}$$

where n_l is the number density of particles with volume l; l, I, and J are integer indices ($l = 1$ corresponds to the volume of the primary coagulation particle or the smallest particle). In a typical MC method, a group of numerical particles (each group has a specific volume) is used to simulate the evolution of the PSD. These numerical particles are assumed to reside homogeneously in a virtual space with a volume V_0; the number density of particles with size l is then given by

$$n_l = \frac{N_l}{V_0}, \tag{3}$$

where N_l is the number of particles with volume l. The total number of numerical particles is given below

$$N = \sum_{l=1}^{\infty} N_l. \tag{4}$$

In practice, N is always finite. Theoretically, as $N \to \infty$, the discrete MC particles are supposed to represent the true discrete distribution in Eq. (2), which is in turn generally equivalent to the continuous distribution in Eq. (1) [5]. Within the virtual volume, V_0, it is possible for any two particles to coagulate at a rate determined by the coagulation kernel function. After setting a specific N in the numerical simulation, V_0 is determined by the initial condition of the number density of particles; it remains constant except when resampling occurs (Section 3.3). Although V_0 has no direct effect on the precision of the discrete distribution, it can be conveniently adjusted to provide the necessary number density of particles in a numerical simulation.

In the following, the CMC method of Gillespie [28], which has been the corner stone of later developments, is briefly introduced. The basic idea is to randomly choose two numerical particles to coagulate at a random time, τ; note that the random coagulation time satisfies a Poisson distribution

$$P(\tau) = C_0 \exp(-C_0 \tau), \tag{5}$$

where

$$\begin{aligned} C_0 &= \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} C_{ij}, \\ C_{ij} &= \frac{1}{V_0} \beta(v_i, v_j). \end{aligned} \tag{6}$$

In the above, i and j are the indices of N numerical particles, and v_i and v_j are the volume of the corresponding particles, respectively; C_{ij} depicts the coagulation propensity between particles i and j; C_0 represents the entire coagulation propensity, which determines how quick a coagulation event may happen between any two particles. The coagulation pair is selected according to the following probability

$$P(i, j) = \frac{C_{ij}}{C_0}. \tag{7}$$

The detailed implementation of the CMC can be found in the original work of Gillespie [28].

2.2. Weighted Flow Algorithm (WFA)

In the above CMC method, every numerical particle has the same "weight" to represent real particles. It is argued that differentially weighted particles may be more efficient and advantageous under several conditions. For example, the particle size distribution observed in nature and industrial applications usually has a wide spectrum, and the number density of larger particles is usually several orders of magnitude smaller than that of considerably finer particles. Despite their low number density, larger particles are more "important" in determining certain physical quantities, such as the total mass or higher-order moments of the PSD. Hence, it is suggested to introduce the weight function, $w(v)$, into the Smoluchowski equation in order to obtain the equation for the weighted particle number density, $n^{\mathrm{w}}(v) = n(v)/w(v)$, as proposed in [37]

$$\begin{aligned}\frac{\partial n^{\mathrm{w}}(v)}{\partial t} =& \frac{1}{2}\int_0^v R(u, v-u) P_{\mathrm{birth}}(u, v-u) n^{\mathrm{w}}(u) n^{\mathrm{w}}(v-u) du \\ &- \int_0^\infty R(u, v) P_{\mathrm{death}}(u, v) n^{\mathrm{w}}(v) n^{\mathrm{w}}(u) du,\end{aligned} \tag{8}$$

where

$$\begin{aligned}R(v, u) =& \beta(v, u) \frac{w(v) w(u)}{w_{\min}(v, u)}, \\ P_{\mathrm{death}}(v, u) =& \frac{w_{\min}(v, u)}{w(u)}, \\ P_{\mathrm{birth}}(v, u) =& \frac{w_{\min}(v, u)}{w(v+u)}, \\ w_{\min}(v, u) =& \min[w(v), w(u), w(v+u)].\end{aligned} \tag{9}$$

The weight function is suggested [37] to be of the form

$$w(v) = (\frac{v}{v_0})^{\alpha}, \tag{10}$$

where v_0 is a reference volume, and α is the scaling parameter. If $\alpha = 0$ (i.e., the weight function degenerates to the constant, 1), then the WFA coincides with the

CMC method. If $\alpha = -1$ (i.e., the number density $n(v)$ is weighted by particle volume), then the WFA coincides with the mass flow algorithm [45, 53]. Generally, α is negative ($\alpha < 0$) so as to promote the number of larger numerical particles.

Similar to the CMC method, the weighted particle number density is represented by the number of numerical particles, N_l^{w}, as follows

$$n_l^{\mathrm{w}} = \frac{N_l^{\mathrm{w}}}{V_0^{\mathrm{w}}}. \tag{11}$$

The corresponding equation for n_l^{w} is directly obtained from Eq. (8).

The actual implementation of the WFA [37] is still based on the CMC method [28]. However, in the WFA method, the two coagulated particles, u and v, are removed with the probabilities $P_{\mathrm{death}}(v,u)$ and $P_{\mathrm{death}}(u,v)$, respectively. A nascent particle, $u+v$, is produced with probability $P_{\mathrm{birth}}(v,u)$.

3. Variance Analyses for Stochastic Methods

3.1. Poisson distribution for N_l

For the stochastic coalescence model [59], Gillespie [60] showed that function $P(k,l,t)$ (which is the probability of finding the number of k particles with size l at time t) approximates the Poisson distribution with the mean value, $\overline{N_l}$ (ensemble average of the number of particles with size l), as $t \to \infty$. In the absence of correlations, the result is derived. In fact, the stochastic coalescence model turns into the Smoluchowski equation, Eq. (2), when correlations are absent [59]. In the CMC method for simulating the Smoluchowski equation, self-coagulation is excluded; this is exactly the assumption of no correlations in the derivation [60]. Hence, in the repeated MC simulations to obtain ensemble average, N_l fluctuates randomly among independent repetitions, and the corresponding probability function, $P(k,l,t)$, would satisfy the Poisson distribution

$$\lim_{t\to\infty} P(k,l,t) = \frac{\overline{N_l}^k \exp(-\overline{N_l})}{k!}. \tag{12}$$

A notable property of the Poisson distribution is as follows

$$\overline{N_l} = \mathrm{Var}(N_l). \tag{13}$$

In other words, in the CMC method for coagulation with N numerical particles, the actual number of numerical particles with size l is approximately a stochastic variable that satisfies the Poisson distribution (i.e., the variance equals its mean, $\overline{N_l}$).

3.2. Variance in CMC and WFA

3.2.1. variance of distribution

After introducing the probability function $P(k,l,t)$, Gillespie [60] derived the dynamic equation of $P(k,l,t)$ under the local balance condition and thereafter applied the moment transformation of $P(k,l,t)$ to obtain a series of dynamic equations for moments. Finally, Gillespie [60] showed that the obtained moments approximate those of the Poisson distribution as in Eq. (12). The derivation is extremely technical; instead of following the same route to show that the corresponding probability function, $P^{\mathrm{w}}(k,l,t)$, in the WFA also satisfies the Poisson distribution, a simple argument is provided. The WFA is in principle of a CMC method equipped with an additional random birth/death process to "twist" the numerical PSD as desired, and the particles in the PSD are assumed to be uncorrelated during the derivation [60]. Consequently, it is argued that $P^{\mathrm{w}}(k,l,t)$ will also approach the Poisson distribution with the mean $\overline{N_l^{\mathrm{w}}}$ (the foregoing is verified through various numerical examples below). Based on the Poisson distribution property that the mean is equal to its variance, the following is obtained

$$\underset{l}{\mathrm{Var}}(N_l^{\mathrm{w}}) = \overline{N^{\mathrm{w}}} = \overline{n_l}^{\mathrm{w}} V_0^{\mathrm{w}} = \frac{\overline{n_l} V_0^{\mathrm{w}}}{w(l)} = \frac{\overline{N_l} V_0^{\mathrm{w}}}{V_0 w(l)} = \frac{V_0^{\mathrm{w}} \mathrm{Var}(N_l)}{V_0 w(l)}, \tag{14}$$

i.e.,

$$\frac{\mathrm{Var}(N_l)}{\mathrm{Var}(N_l^{\mathrm{w}})} = \frac{V_0}{V_0^{\mathrm{w}}} w(l), \tag{15}$$

then

$$\frac{\mathrm{Var}(n_l)}{\mathrm{Var}(n_l^{\mathrm{w}})} = \frac{\mathrm{Var}(N_l/V_0)}{\mathrm{Var}(N_l^{\mathrm{w}}/V_0^{\mathrm{w}})} = \frac{\mathrm{Var}(N_l)}{\mathrm{Var}(N_l^{\mathrm{w}})} \frac{(V_0^{\mathrm{w}})^2}{V_0^2} = \frac{V_0^{\mathrm{w}}}{V_0} w(l). \tag{16}$$

Hence, the ratio of variances of the number density obtained using the CMC and WFA methods is

$$\frac{\mathrm{Var}(n_l)}{\mathrm{Var}(n_l^{\mathrm{w}} w(l))} = \frac{\mathrm{Var}(n_l)}{w^2(l)\mathrm{Var}(n_l^{\mathrm{w}})} = \frac{V_0^{\mathrm{w}}}{V_0} \frac{1}{w(l)}. \tag{17}$$

The variance relationship expressed in Eq. (17) is one of the major findings of the present study; a detailed interpretation is appropriate. First, note that this result is not trivial. Based on the defined relationship, $n_l^{\mathrm{w}} = n_l/w(l)$ (which is equivalent to $\overline{n_l^{\mathrm{w}}} = \overline{n_l}/w(l)$ in stochastic simulations) in Eq. (8), it is not possible to infer the variance relationship directly. The correctness of this relationship is demonstrated through various numerical examples in Section 4.

Secondly, it is not necessary for the CMC and WFA simulations to use the same number of numerical particles to make the variance comparison possible. Actually, the only requirement is $\overline{n_l^{\mathrm{w}}} = \overline{n_l}/w(l)$, which is presumably satisfied because both the CMC and WFA would predict the same correct results in terms of the mean value. Evidently, the actual number of numerical particles affects the precision in the numerical simulation, as discussed in Section 4.2. The effect of numerical particle number, N, on the variance is reflected by the ratio of virtual volumes, V_0^{w}/V_0, in Eq. (17).

Finally, the simulated number density of particles for a specific size has a lower variance if a smaller weight is used. For instance, in the weight function $w(v) = (v/v_0)^{\alpha}$ in Eq. (10), $\alpha < 0$ is always adopted so as to render a smaller weight for larger particles. If the reference volume, v_0, is taken as the smallest in the PSD, then for particles of all sizes, $w(v) \leq 1$; this means that the variance in the WFA is always no larger than that in the CMC (if we also let $V_0 = V_0^{\mathrm{w}}$). The WFA is superior to the CMC in reducing the variance among larger particles; this is exactly the main objective for developing the WFA [37]. However, the numbers of numerical particles in the CMC and WFA generally evolve at considerably different rates. Our experience on numerical simulations show that the CMC only typically requires a small number of particles (hence, a high numerical efficiency) to achieve comparable a precision comparable to that of the WFA on lower-order statistics (e.g., total number density and total volume fraction). A simple explanation of this phenomenon is that the WFA substantially focuses much on higher-order statistics; accordingly, its resolution for low-order statistics is compromised. The detailed comparisons of the

pros and cons between the CMC and WFA in the simulation results are presented in Section 4.

3.2.2. variance of moments

Moments provide concise but crucial information about the particle size distribution. In this subsection, the analysis of the evolution of mean and variance of the moments in stochastic simulations is presented.

By multiplying both sides of the Smoluchowski equation (Eq. (1)) with v^k, and integrating over the particle size space, the dynamic equation for the kth moment can be derived (A dynamic equation for a general transformation (other than the specific moment transformation) can be found in Drake [61])

$$\frac{\mathrm{d}M_k}{\mathrm{d}t} = \frac{1}{2}\int_0^\infty \int_0^\infty [(v+u)^k - v^k - u^k]\beta(v,u)n(v)n(u)\,\mathrm{d}u\,\mathrm{d}v. \tag{18}$$

One interesting but to a certain extent overlooked conclusion [61] is that the right-hand side of Eq. (18) is negative if $k<1$ and positive if $k>1$; this means that M_k increases with time when $k>1$ and decreases when $k<1$. In particular, M_1 is constant with time because the volume fraction is conserved during the pure coagulation process; on the other hand, M_0, the total number density of particles, decays because of coagulation. If β is a polynomial function of v and u, then the double integral on the right-hand side of Eq. (18) can be directly performed as the polynomial of moments of various orders. In particular, if β is a constant, then only moments lower than the kth order appear on the right-hand side.

Because all stochastic methods proposed for simulating the Smoluchowski equation are supposed to converge to the Smoluchowski equation (Eq. (1)) as $N\to\infty$, the mean discrete moments are expected to evolve approximately as Eq. (18). These moments are defined through

$$M_k = \frac{1}{V_0}\sum_{i=1}^{N} v_i^k, \qquad \text{(CMC)} \tag{19}$$

$$M_k = \frac{1}{V_0^{\mathrm{w}}}\sum_{i=1}^{N} v_i^k w(v_i), \qquad \text{(WFA)} \tag{20}$$

where v_i is the volume of the ith numerical particle.

The evolutions of the moment variance are found as follows (lengthy derivation is not shown here):

$$\frac{\mathrm{d}[\mathrm{Var}(M_k)]}{\mathrm{d}t} = \frac{1}{V_0^2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \overline{C_{ij}} [(v_i + v_j)^k - v_i^k - v_j^k]^2, \quad \text{(CMC)} \quad (21)$$

$$\frac{\mathrm{d}[\mathrm{Var}(M_k)]}{\mathrm{d}t} = \frac{1}{(V_0^{\mathrm{w}})^2} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \overline{C_{ij}} \left\{ w(v_i + v_j)[(v_i + v_j)^k - v_i^k - v_j^k]^2 \right. \quad \text{(WFA)} \quad (22)$$
$$\left. + [w(v_i) - w(v_i + v_j)] v_i^{2k} + [w(v_j) - w(v_i + v_j)] v_j^{2k} \right\},$$

where $\overline{C_{ij}}$ in Eqs. (21) and (22) denotes the mean value of C_{ij} (which is defined in Eq. (6)). If $w(v) = 1$ (for all v) , i.e., no differential weighting, then Eq. (22) degenerates to Eq. (21). For the general case of $w(v) \neq 1$, the preliminary information on the moment variance can be obtained by inspecting the right-hand side of the two equation above. Particularly for $k = 1$ (corresponding to the volume fraction), the right-hand side of Eq. (21) is zero; this means that the volume fraction (or mass) is strictly conserved (no stochasticity) in the CMC method. However, the right hand-side of Eq. (22) is generally non-zero for $k = 1$; this means that stochastic fluctuation is introduced to the volume fraction after the particles are weighted differentially. This might be a crucial problem under certain conditions, such as coupling aerosol dynamics (soot) with sensitive gaseous chemical reactions (i.e., combustion processes), when the strict conservation of total particle mass is necessary [62]. For a general moment, the right-hand side is found to scale approximately with $1/N$. Because C_{ij} is proportional to $1/V_0$ (Eq. (6)), and V_0 is proportional to N for a given number density of particles, the denominator is proportional to $1/N^3$. On the other hand, the double summation in the numerator is proportional to $N(N-1)$; therefore, the whole term is approximately proportional to $1/N$. With a further assumption that N does not affect the PSD, which is true when N is sufficiently large, it can be concluded that the moment variance at a given time is inversely proportional to N, as numerically verified in Section 4.3.2.

3.3. Stochasticity in resampling

Resampling refers to the addition (duplication) or removal of randomly selected particles from existing ones in order to control the number of numerical particles in a specified range. Resampling is explicitly or implicitly adopted in practically all stochastic simulations of aerosol dynamics. However, to the best of the authors' knowledge, it has never been systematically and thoroughly investigated.

For pure coagulation simulation with the WFA, the number of numerical particles remains practically constant when $\alpha = -1$ (statistically corresponding to the mass flow algorithm [45, 53]), where resampling is not necessary. When $\alpha < -1$, the number of particles may increase; hence, downsampling might be necessary to prevent this number from growing extremely large (which would require considerable computational resources). When $\alpha > -1$, the number of particles decreases. In particular, when $\alpha = 0$ (corresponding to the CMC method), there are generally two methods (i.e., constant number [58] and constant volume [47]) that can be employed to prevent this number from becoming extremely small. In the constant number method, a particle randomly selected from existing particles is duplicated after each coagulation event when a particle is depleted (two particles coagulate into one). In the constant volume method, the number of particles is allowed to decease to a certain threshold (typically half of the original particle number), and thereafter increase by duplicating all existing particles. All these operations are actually a certain type of resampling. For general aerosol dynamics, including coagulation and nucleation, resampling is generally inevitable [37].

Resampling is generally presumed to introduce no statistical bias, which is relatively easy to prove mathematically (Appendix A). The variance introduced by resampling has never been investigated. Suppose that a discrete distribution of N particles is given. Through upsampling, m additional particles are added to the existing N while keeping the new distribution statistically equivalent to the old one. A simple way to accomplish this is to select m (assume $m \leq N$) particles from N particles and duplicate those selected; thereafter, put these duplicates into the particle

pool to obtain $N+m$ particles. If $m>N$, then whole N is firstly duplicated as many times as the modulus of $m$mod N (mod denotes the modulo operation); thereafter, the process is repeated for the common remainder as in the case of $m\leq N$. After the addition of $m$particles, the virtual volume has to be scaled up to $V_0(m+N)/N$ in order to render the particle number density distribution statistically unchanged. During downsampling, the virtual volume has to be scaled down to $V_0(N-m)/N$ in order to remove m particles from N. The variance during the upsampling or downsampling is found as below (Eq. (A.12) or Eq. (A.14))

$$\mathrm{Var}(X_+) = \frac{\frac{m}{N}\left(1-\frac{m}{N}\right)}{\left(1+\frac{m}{N}\right)^2}\frac{1}{N}\mathrm{Var_p}[M_k(N)], \tag{23}$$

$$\mathrm{Var}(X_-) = \frac{\frac{m}{N}}{\left(1-\frac{m}{N}\right)}\frac{1}{N}\mathrm{Var_p}[M_k(N)], \tag{24}$$

where

$$X_\pm = \frac{1}{N\pm m}\left(\sum_{i=1}^{N} x_i^k \pm \sum_{i=1}^{m} x_{\xi_i}^k\right), \tag{25}$$

and $\mathrm{Var_p}[M_k(N)]$ is the unbiased estimator of the variance of the kth moment as defined in Eq. (A.13). Equations (23) and (24) indicate the magnitude of variance for the kth-order moment of the particle size distribution introduced in upsampling and downsampling, respectively. The sample variance depends on three parameters: the resampling rate (m/N), the number of numerical particles (N), and the population variance ($\mathrm{Var_p}[M_k(N)]$) of the particle size distribution to be resampled. In the foregoing, with the underlying assumption that a statistically stable variance exists (which reflects the outcome of a large number of repetition), the number of repetitions is not explicitly considered. It is worth pointing out that the population variance (defined in Eq. (A.13)) should be independent of N when N is sufficiently large; this is because the obtained PSD in an MC simulation is expected to converge to the true PSD as N increases. It is evident that the resampling variance is proportional to the population variance; for instance, resampling from a single-peaked distribution (zero variance) will not introduce any variance to the original particle size distribution. However, given that other conditions (resampling rate and population variance) are the same, it is remarkable to see the resampling variance being inversely proportional

to N; for example, under the same resampling rate (asume that half of particles are removed), the sampling variance can still be considerably different when N is varied. This relationship implies that the resampling variance can always be reduced by using a larger N. The resampling rate has different effects on the sampling variance between upsampling and downsampling. In downsampling, the removal of more particles makes the resampling variance sharply increase until the limiting case (only one particle remains; the removal of all particles is meaningless) is reached. With the resampling variance practically equal to the population variance (as indicated by the ratio $(N-1)/N$), downsampling causes enormous stochasticity; this should almost always be avoided. In upsampling, the addition of more particles makes the variance increase first and thereafter drop to zero ($m = N$). At $m/N = 1/3$, the first factor on the right-hand side of Eq. (23) reaches the maximum value of 1/8. Doubling ($m = N$) is commonly adopted in numerical simulations; it is optimal in order to reduce the resampling variance. The above results are derived for a non-replacement selection of m particles. It is also possible to select one particle at a time until m particles are selected (resampling with replacement); however, this results in a variance larger than that in the resampling without replacement. The foregoing can be verified by applying Eq. (23) or (24) for $m = 1$, recursively.

The examples in Section 4.4 provide numerical evidence for the correctness of the above derivation.

4. Simulation Results

4.1. Various types of errors

A brief introduction on the various types of errors is presented here to facilitate the later discussions on numerical simulation results. With regard to the MC simulation errors, there are two fundamental questions to answer. A finite number of numerical particles (N) is used to simulate to a continuous Smoluchowski equation. Accordingly, the natural question is, how large should N be to make the discretization error sufficiently small as desired without considering the stochastic fluctuation

(which can be made arbitrarily small through repetitions in theory)? Although it is known [5] that as $N \to \infty$, the discrete Smoluchowski equation (Eq. (2)) converges to the continuous Smoluchowski equation (Eq. (1)), simple and sound analysis to quantify the discretization error in the moments is scarce. The best-known error estimation states that the discretization error is inversely proportional to the number (N) [63]; this is further discussed in Section 4.2.

On the other hand, a pertinent question related to MC simulations is the stochastic error, which involves two aspects: the stochastic variance lying in a MC simulation (which is closely related to the specific MC method and the number N adopted), and the repetition of the MC simulation. An independent MC simulation is usually repeated N_{rep} times to smooth out the randomness in practice. The stochastic error is generally found as

$$\text{Stochastic Error} = \sqrt{\frac{\text{Variance}}{N_{\text{rep}}}}. \tag{26}$$

The error is well-known to decrease at a rate proportional to $\sqrt{N_{\text{rep}}}$. However, the true variance of the stochastic process is generally unknown. So the unbiased sample variance is used as an estimator for the true variance

$$\text{Var} = \frac{1}{N_{\text{rep}} - 1} \sum_{i=1}^{N_{\text{rep}}} (X_i - \overline{X})^2, \tag{27}$$

where X_i denotes the ith repetition of a random variable, and $\overline{X} = \sum_{i=1}^{N_{\text{rep}}}$ denotes its sample mean. One of the main objectives of the present study is to compare the variances in the CMC and WFA methods.

4.2. Systematic errors

Kolodko and Sabelfeld [63] proved that the discretization error (or systematic error) satisfies the following inequality

$$||\text{E}\left[N_\infty(t)\right] - n(t)|| \leq \frac{1}{2}\left(\frac{3}{4N} + \sup_{\tau \leq t} ||cov(\tau)||\right)\left[\exp(4\beta_{\max} t) - 1\right], \tag{28}$$

where $N_\infty(t)$ and $n(t)$ denotes the simulated and true discrete distribution to the Smoluchowski equation, respectively. The norm is defined as $||n|| = \sum_{I \geq 1} |n_I|$, $\beta_{\max}$

is the maximum of the coagulation kernel; and the vector $cov(t)$ is defined as

$$cov(t) = \left(\sum_{J \geq 1} \mathrm{Cov}[N_i(t), N_j(t)] \right)_{I \geq 1}, \tag{29}$$

where Cov denotes the covariance, and N_I and N_J are the number density of particles with volume I and J, respectively. It is found that $||cov(t)||$ approaches zero as time increases [63]. Moreover, the correlation between any two particles becomes ever weaker as N increases. Using the inequality condition in Eq. (28) for the dis-crete distribution, it is straightforward to derive an inequality for the kth moment according to the definition of Eq. (19). The inequality is multiplied by I^k/V_0 and thereafter summed over I from 1 to N, i.e.,

$$\begin{aligned}
||M_k - m_k|| =& \frac{1}{V_0} \left| \sum_{I=1}^{N} I^k \mathrm{E}\left[N_I(t)\right] - \sum_{I=1}^{\infty} I^k n_I(t) \right| \\
=& \frac{1}{V_0} \left| \sum_{I=1}^{N} I^k \mathrm{E}\left[N_I(t)\right] - \sum_{I=1}^{N} I^k n_I(t) - \sum_{I=N+1}^{\infty} I^k n_I(t) \right| \\
\leq& \frac{1}{V_0} \left(\sum_{I=1}^{N} I^k |E[N_I(t)] - n_I(t)| + \sum_{I=N+1}^{\infty} I^k n_I(t) \right) \\
\leq& \frac{1}{V_0} \left(\sum_{I=1}^{N} I^k ||\mathrm{E}\left[N_\infty(t)\right] - n(t)|| + \sum_{I=N+1}^{\infty} I^k n_I(t) \right) \\
\leq& \frac{1}{2V_0} \sum_{I=1}^{N} I^k \left(\frac{3}{4N} + \sup_{\tau \leq t} ||cov(\tau)|| \right) \left[\exp(4\beta_{\max} t) - 1\right] + \frac{1}{V_0} \sum_{I=N+1}^{\infty} I^k n_I(t),
\end{aligned} \tag{30}$$

where m_k denotes the kth moment of the true discrete distribution. If N is sufficiently large, all n_I/V_0 for $I > N+1$ are practically zero; hence, the second term in the last inequality is negligible compared to the first term, and the systematic error for moments has a form that is similar to that of the PSD.

Although the derivation of Eq. (30) is based on the moment definition in Eq. (19) for the CMC method, the final inequality condition is also applicable to the WFA method. This is because the extra weight factor, $w(v)$, in the moment definition in Eq. (20) for the WFA is canceled with the introduction of factor $1/w(v)$ in

$N_I = N_I^{\mathrm{w}}/w(v)$. Therefore, the systematic error, which is independent of the actual stochastic method, is supposed to be scaled with $1/N$ when N is sufficiently large.

In order to investigate the systematic error numerically, a pure coagulation with a constant kernel, $\beta = 1$, is simulated with the CMC method. Initially, all particles are assumed to have the same size of 1. The analytical solution of the PSD to this problem is well-known [1]

$$n(v,t) = \frac{\left(\frac{t}{t+2}\right)^{v-1}}{\left(1+\frac{t}{2}\right)^2}, \quad v = 1, 2, 3, \ldots. \tag{31}$$

The moments of the PSD are directly available from summing up the weighted PSDs from $v = 1$ to ∞ with the weight v^k (k is the order of moment). The infinite summation can be easily handled with a symbolic calculation to yield $M_0 = 2/(t+2)$, $M_1 = 1$, $M_2 = t + 1$, $M_3 = \frac{3}{2}t^2 + 3t + 1$, etc.; the foregoing can also be obtained alternatively by solving the moment evolution in Eq. (18). These analytical results function as the reference in the calculation of relative errors.

Figure 1 shows the systematic relative error of moments from the numerical simulation with the CMC method at $t = 6$ and 20 that are obtained for different numerical particles, N. For each N, the simulation is repeated 1000 times. The stochastic error defined in Eq. (26) is also shown in the error bar. Only the upper error bar is shown here because the lower error bar is problematic in a logarithmic scale when the value becomes zero or negative. The error for the first moment, M_1 (i.e., volume fraction), is not shown here because the volume fraction is conserved in the CMC method, and the error is zero. In both instances, the systematic relative errors for all moments scale linearly with $1/N$; this agrees with the scaling implied in the inequality in Eq. (30) when the correlation term is neglected. Generally, higher-order moments exhibit larger stochastic errors. At an earlier time (i.e., $t = 6$), the stochastic error of higher-order moments is considerably large; this results in a notable deviation from linearity. At a later time, $t = 20$, the stochastic error becomes substantially smaller when compared with the systematic error, and a considerably good linearity is obtained even for higher-order moments when $N > 200$. This emanates from the

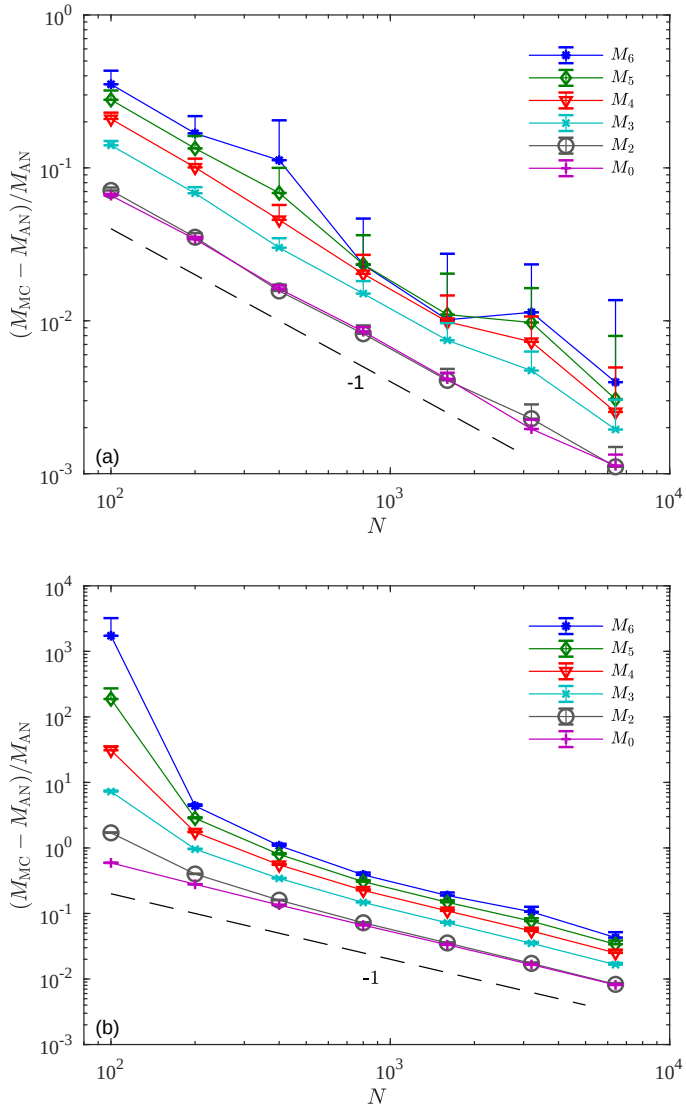

Figure 1: Relative errors of moments with respect to N for constant coagulation using the CMC method; M_{AN} denotes the well-known analytical result. (a) $t = 6$; (b) $t = 20$.

fast exponential growth rate of the systematic error reflected in Eq. (30). In the numerical simulation with $N = 100$ at $t = 20$, the relative error of higher-order moments is exceptionally large; this is presumed to arise from the strong correlation among the numerical particles when N is relatively small. Comparing the magnitude of relative errors between these two instances above, it is evident that the relative error substantially increases with time.

The inequality in Eq. (30) implies that the relative error increases exponentially with t as $\exp(4t)$ (here, $\beta_{\max} = 1$). In Fig. 2, the evolution of relative error is shown for $N = 400$. After the start-up phase, when the correlation is strong, the relative error growth at the rate $\exp(0.15t)$ is considerably more benign than the theoretical prediction. The numerical simulation results suggest that it may be possible to derive a sharper upper bound from the theory.

The same numerical simulations referred to Fig. 1 have also been repeated using the WFA method, which shows the same linear decaying rate with respect to N; accordingly, the results are not shown here.

Although the stochastic simulation of coagulation with a constant kernel is relatively trivial, the information derived here remains of interest. There are three pieces of important information. First, the systematic error decreases linearly with $1/N$; this is not new but is worth reconsidering. Second, the relative error exponentially increases with time; this result has an important implication when simulating coagulation for a relatively long period. In order to obtain good precision at a later time, it may be necessary to increase the number of numerical particles during the numerical simulation; the foregoing involves not only compensating for particles depleted because of coagulation, but also using even a higher number of particles in the later stage than that at the beginning of the numerical simulation. Due to the aforementioned reasons, such numerical simulations with an ever-increasing number of particles over time have never been implemented. Lastly, the stochastic error is significant at the initial period of the numerical simulation when the systematic error is small; further efforts should be expended to reduce the stochastic error during the

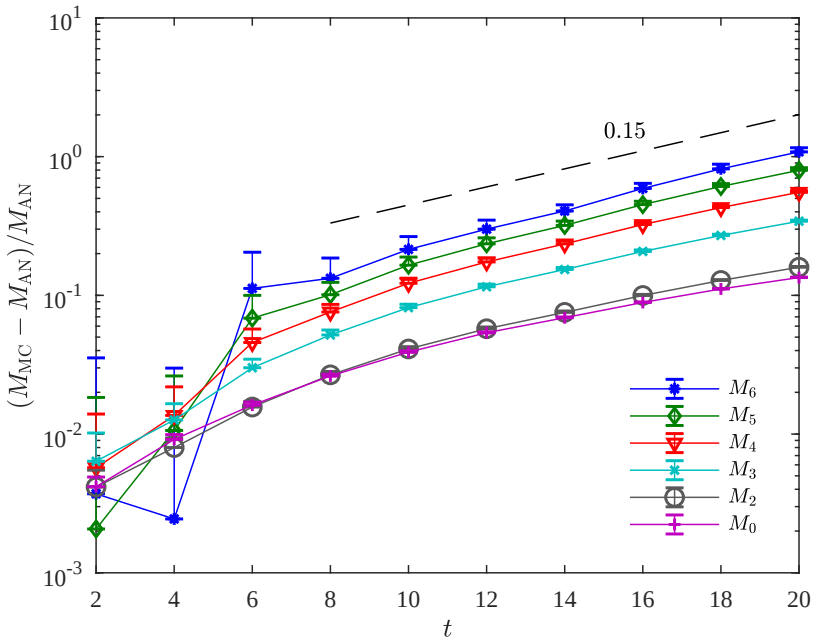

Figure 2: Relative errors of moments with respect to t for a constant coagulation with $N = 400$ using the CMC method.

initial period.

4.3. Variance scaling

4.3.1. Poisson distribution

The variance scaling relation in Eq. (17) between the CMC and WFA is derived from the promise of Poisson distribution for N_l. Numerical simulations of a pure coagulation using both CMC and WFA are carried out to examine the hypothesis of Poisson distribution. In the WFA, the weight function in Eq. (10) is set as $w(v) = 1/v$. The simulations are repeated 400 times to obtain 400 replicas of the discrete distribution of N particles. The number of particles with size l, i.e. N_l, changes randomly during each repetition. Table 1 shows the results from the chi-square goodness-of-fit test ("chi2gof" function in Matlab) with the null hypothesis of a Poisson distribution for each $N_l\,(l = 1, 2, \ldots, N)$ at the 5% significance level. A "0" in Table 1 means that the tested data is highly likely to come from a Poisson distribution, while a "1" means otherwise. Overall, the simulated number N_l for small

Table 1: Poisson distribution test of N_l for a constant kernel coagulation. A "0" means that the tested data is highly likely to come from a Poisson distribution (filled with gray), while a "1" means otherwise. Overall, the ensemble distribution of small particles satisfies the Poisson distribution except for the CMC at $t = 2$ (filled with dark gray), which deviates from the Poisson distribution even for relatively small particles due to the particle doubling operation.

	CMC					WFA				
Size \ Time	1	1.5	2	2.5	3	1	1.5	2	2.5	3
2	0	0	0	0	0	0	0	0	0	0
4	0	1	1	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	0	0
8	1	0	1	1	0	0	0	0	0	0
10	1	1	1	1	0	1	0	1	0	0
12	1	1	1	1	0	1	0	0	0	0
14	1	1	1	1	1	1	1	0	0	0
16	1	1	1	1	1	1	1	1	0	0

particles satisfies the distribution very well from both the CMC and WFA. With the increase of time, the number for ever larger particles also approaches the Poisson distribution. The simulated number N_l from the WFA shows better proximity to the Poisson distribution for large particles. The results from the CMC method at $t = 2$ are clearly outliers, which comes from the particle doubling operation. More details on the doubling operation are provided in Section 4.3.3.

The hypothesis of Poisson distribution for N_l is numerically proved reasonable. Only the number of those under-resolved large particles deviates from the Poisson distribution. In the WFA, the number of larger particles is higher which incurs less deviation from the Poisson distribution.

4.3.2. variance of moments

Figure 3 shows the sample variance of moments in pure coagulation for different values of N at the time instance $t = 2$ from a uniform initial distribution. Both the

constant and the free molecular kernels are considered. The numerical simulations employ the CMC and the WFA methods with the weight function $w(v) = 1/v$. The sample variance is obtained with sufficient repetition for each N in order to reduce the stochastic error in simulating the variance to a negligible level. Hence, the sample variance is presumed to be the true variance in the numerical simulation with a specific N. Despite the fact that higher-order moments have larger variances, the variances in all moments decrease at the same rate, which is linearly proportional to N. This linear scaling is independent of the numerical simulation algorithms for either the CMC or WFA method; it is also independent of the coagulation kernel for either the constant or the free molecular regime. In comparing the magnitude of the variance between the CMC and WFA methods, it is evident that the variance in lower-order moments are comparable, whereas the variance in higher-order moments using the CMC method is considerably higher than that in the WFA method. The WFA method is capable of capturing the larger particle distribution more precisely compared with the CMC method, which renders a lower variance when large particles dominate in the calculation of higher-order moments. One distinct feature is that the variance of M_1 is zero in the CMC method, whereas it is non-zero in the WFA method. These findings considerably conform with the pertinent analysis of the variance evolution in Eqs. (21) and (22).

4.3.3. ratio of variance between CMC and WFA

Equation (17) provides an analytical relationship between the variances obtained by the CMC and WFA methods; such relationship is one of the key findings in the present study. The analytical relationship applies to any type of weight function in the WFA method. Figure 4 compares the variance ratios obtained from numerical simulations and the theory (Eq. (17)) for coagulation with the constant kernel and the free molecular regime kernel (depicted in Eq. (37)). In the WFA simulations, the weight function is chosen as $w(v) = v_0/v$ (i.e., $\alpha = -1$ in Eq. (10)), and the values of reference volume, $v_0 = 1$ and 5, are used separately. Overall, the numerical simulation results sufficiently agree with the theory, except for certain scattering

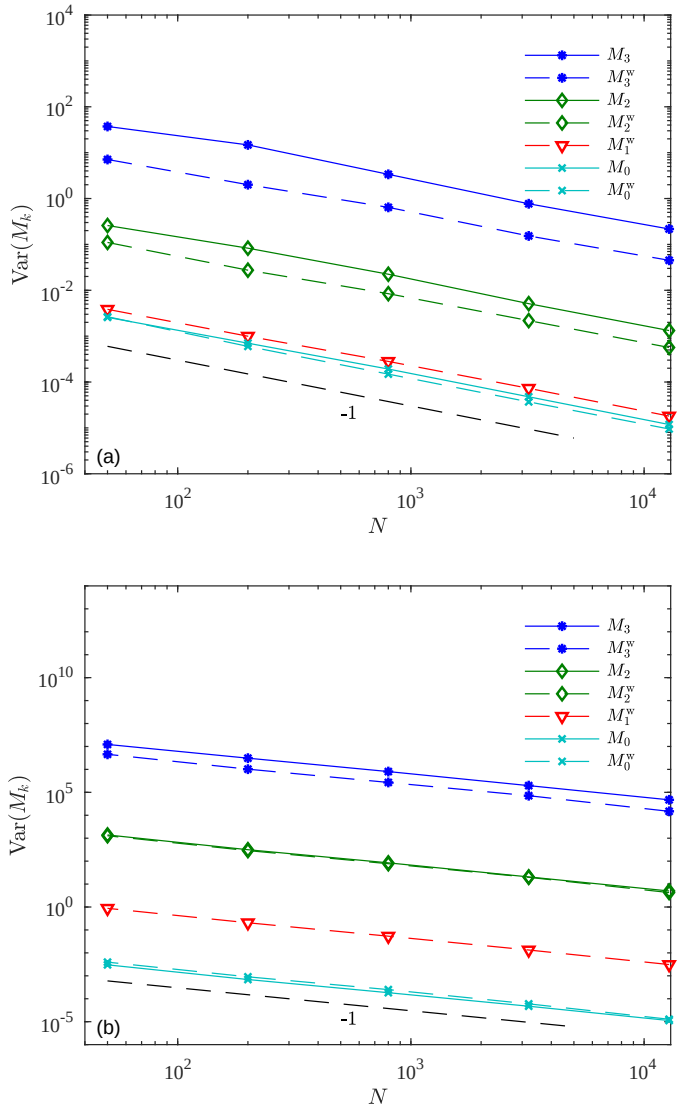

Figure 3: Sample variance of moments in pure coagulation, where the dashed lines represent those obtained from the WFA method, and continuous lines indicate those obtained from the CMC method. (a) constant kernel; (b) free molecular kernel.

caused by the stochasticity. Figure 4(a) shows the results of using two different coagulation kernels; no significant systematic difference is exhibited. The reference volume is $v_0 = 1$; this implies that all particles of different sizes (except for $v = 1$) in the WFA method have lower weights than that those in the CMC method. The corresponding variance in the WFA method is smaller than that in the CMC method. Among larger particles, the discrepancy in variance is even larger. In Fig. 4(b), the reference volume is set as $v_0 = 5$. It is evident that the variance in the CMC method is smaller that that in the WFA method when the particle size is $v < v_0$. When the particle size is $v > v_0$, the variance in the CMC method is larger.

During the coagulation simulation, the number of numerical particles may increase or decrease in different algorithms. In the CMC method, the number of numerical particles always decreases because two coagulating particles are removed, and only one newly coagulated particle is added with each coagulation event. However, in the WFA method, the number of numerical particles is controlled by the power index, α, in the weight function in Eq. (10). When $\alpha = -1$, the number of removed particles is statistically equal to that of the added particles; hence, the total number is unchanged with time. When $\alpha > -1$, the total number decreases. In the extreme case of $\alpha = 0$, the number in the WFA decreases at the same rate as that in the CMC. When $\alpha < -1$, the total number even further increases because the number of particles added is larger than that removed. In order to prevent the number from excessively decreasing during coagulation, a top-up scheme is generally used. The most common one is to replicate all the particles when the number of numerical particles decreases to half of its initial value; at the same time, the simulator volume is doubled as required by Eq. (3).

Figure 5 shows the time evolution of the relative variance between the CMC and WFA methods after the simulator volume is taken into consideration. Initially, the number of numerical particles and the simulator volume are the same in both the CMC and WFA methods. In the WFA method, $\alpha = -1$ is used; hence, the simulator volume, V_0^{w}, does not change with time. In the CMC method, the number

of numerical particles decreases to one-half at $t = 2$; at this instance, the doubling scheme is employed, and the volume, V_0, is doubled. At $t = 6$, another doubling of V_0 occurs in the CMC method. Equation (17) predicts ($v_0 = 1$) the following

$$\frac{1}{v}\frac{\mathrm{Var}(n_l)}{\mathrm{Var}(n_l^{\mathrm{w}}/v)} = \frac{V_0^{\mathrm{w}}}{V_0}. \tag{32}$$

The term on the left-hand side of Eq. (32) depends on the particle volume, v. In Fig. 5, this term is shown after taking the overall average of v; this corresponds to the slope of a straight line obtained by fitting the data points as shown in Fig. 4. It is evident that the simulated curve correctly predicts the jump transitions at $t = 2$ and 6. The curve considerably approaches the dashed line between the transitions during which the simulator volume remains constant.

Through detailed comparisons, as shown in Figs. 4 and 5, it is considerably convincing that the theoretical relationship of the variances between the CMC and WFA methods is correct under general conditions; this provides a simple and convenient guide in choosing the appropriate stochastic simulation method in terms of the stochastic errors.

4.4. Numerical proof of resampling analyses

Equations (23) and (24) indicate the magnitude of variance introduced during the resampling of a given discrete distribution. Numerical simulations are employed to verify the analysis, which is applicable to any type of distribution function. In order to demonstrate the close relevance to a practical PSD of aerosols, a continuous log-normal distribution function is selected; the function has a mean and variance of 1.44 and 0.58, respectively. It is worth pointing out that the actual shape of the distribution is inconsequential to the verification. First, a discrete distribution of N particles is generated according to the log-normal distribution. Thereafter, m particles are randomly selected from N particles; the former particles are duplicated (upsampled) or removed (downsampled). These processes are repeated a thousand times to achieve stable estimations of the variances of random variables defined in Eqs. (A.6) and (A.15). The numerical simulation results along with theoretical

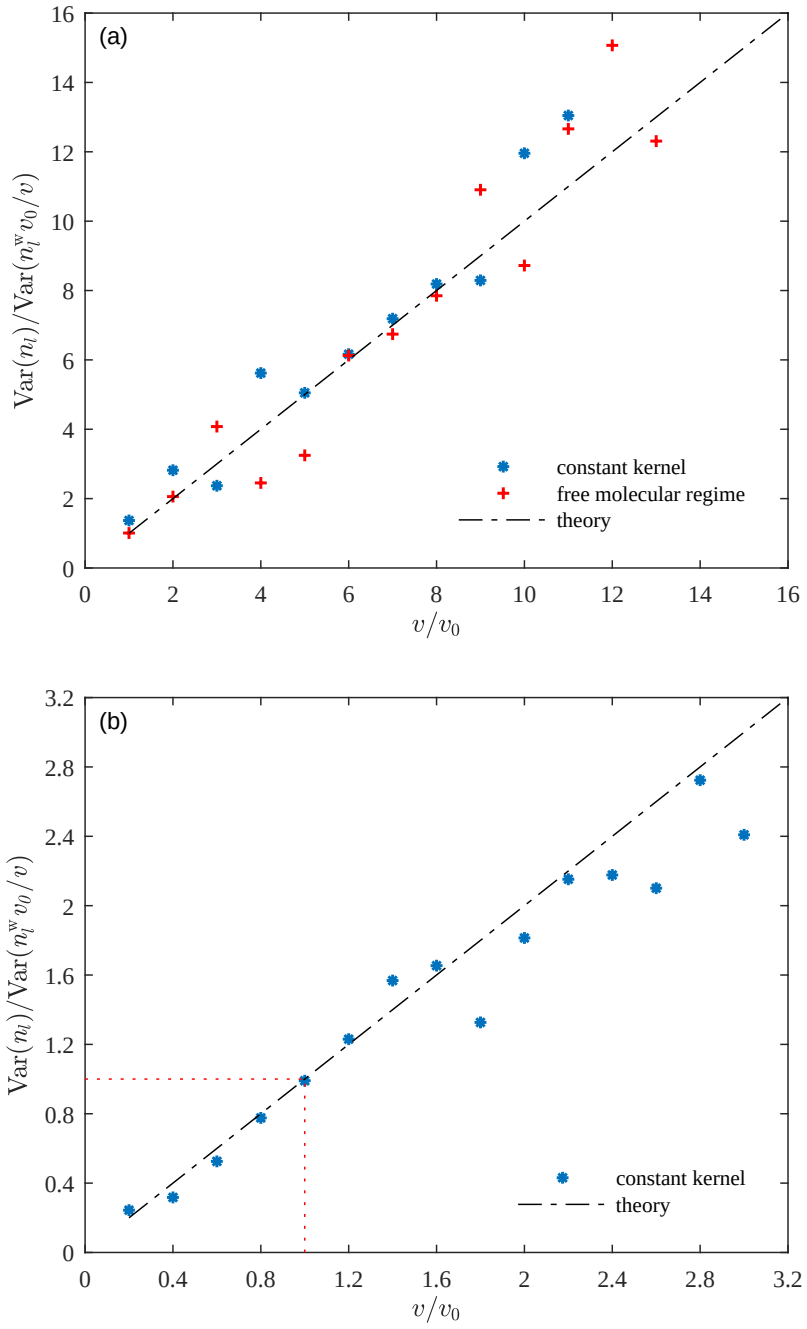

Figure 4: Ratio of variance between the CMC and WFA methods for pure coagulation at $t = 2$. The weight function in the WFA method is set as $w(v) = v_0/v$. Here, "free molecular regime" denotes coagulation with the kernel defined in Eq. (37), and "theory" denotes the numerical prediction from Eq. (17); (a) $v_0 = 1$ and (b) $v_0 = 5$.

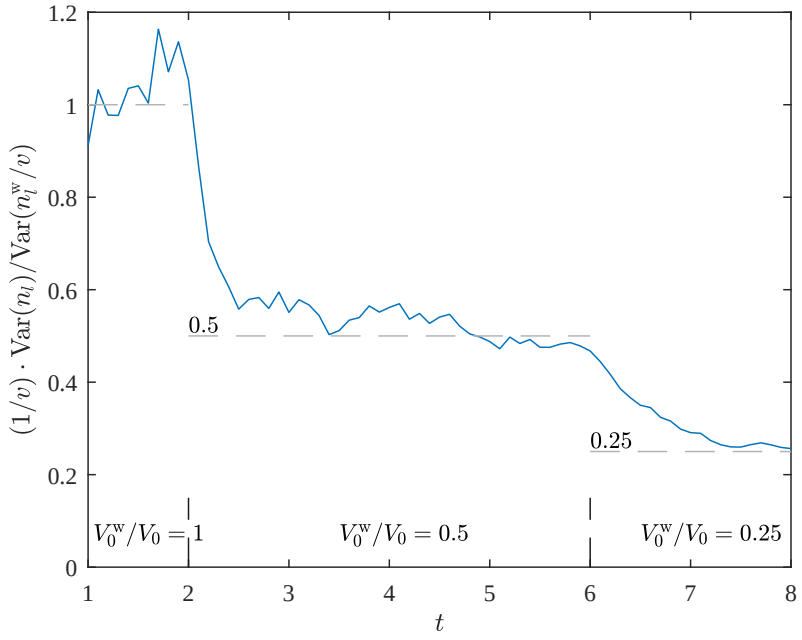

Figure 5: The time evolution of the slope of a straight line is obtained by fitting the data points, as shown in Fig. 4 for the coagulation with a constant kernel. In the WFA method, the simulator volume, V_0^{w}, is constant with time. In the CMC method, the volume, V_0, doubles at $t = 2$ and then doubles again at $t = 6$. This plot tends to verify Eq. (32), where the continuous curve is the average over all values of v obtained from its left-hand side, and the horizontal dashed lines are the values obtained from the right-hand side.

values obtained from Eqs. (23) and (24) are shown in Figure 6; simulation results with $N = 1000$ and 2000 are presented. After scaling the ratio between the sample variance and the population variance with N, the scaled ratio becomes a function only of the resampling rate, m/N, i.e., the first factors on the right hand sides of Eqs. (23) and (24). The numerical simulation results considerably agree with the theory. Increasing N from 1000 to 2000, the resampling variance decreases to one-half. In upsampling, the variance increases as the resampling rate, m/N, increases from 0 to 1/3; thereafter, it decreases to zero when m/N increases to 1. In downsampling, the variance always increases with the sampling rate. Downsampling has more significant effect on the resampling variance than upsampling.

Both theoretical and numerical results demonstrate that the doubling scheme is optimal in regard to the sampling variance minimization. The constant number scheme [58] introduces the resampling variance of the scale $1/N$ at every time step, which may accumulate to a considerable level that may be several orders of magnitude higher than the doubling scheme [30]. Evidently, in doubling N, the computational cost increases accordingly; this might be proportional to N or N^2 depending on how C_{ij} in Eq. (6) is updated to a new time step [28].

4.5. Numerical simulation of general aerosol dynamics

The general aerosol dynamics is described as

$$\frac{\partial n}{\partial t} = I_{\text{nuc}} + G_{\text{cond}} + C_{\text{coag}}, \tag{33}$$

where the source terms of I_{nuc}, G_{cond}, and C_{coag} on the right-hand side denote nucleation, surface growth (condensation), and coagulation, respectively. The aerosol dynamics is then simulated by the operator splitting Monte Carlo method [64, 65], with a second order symmetric Strang splitting as

$$\underbrace{\frac{1}{2}(\text{nuc} \to \text{resampling} \to \text{cond})}_{\text{half step}} \to \underbrace{\text{coagulation}}_{\text{full step}} \to \underbrace{\frac{1}{2}(\text{nuc} \to \text{resampling} \to \text{cond})}_{\text{half step}}$$

where "nuc" and "cond" denote modules for nucleation and condensation, respectively. Resampling is discussed in Section 3.3. The factor 1/2 means a half-time step

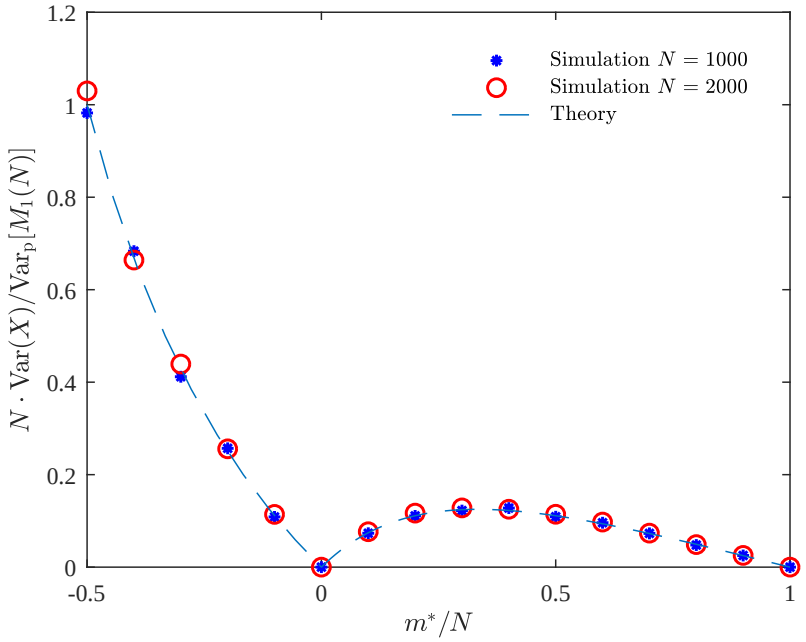

Figure 6: Ratio of the resampling variance to the population variance (for M_1). The number of added particles, m^*, is allowed to be negative to represent the removal of particles. The ratio is scaled by N in order to cancel its dependency on N. Moreover, $\mathrm{Var}(X)$ denotes the variance by adding m^* particles. The theoretical results are obtained from Eq. (23) with $m^* > 0$ and Eq. (24) with $m^* < 0$. The ordinate reflects the magnitude of the variance that is added to the original discrete distribution resulting from the resampling.

integration. The schematic flowchart above shows how a full-time step is carried out.

Detailed aerosol dynamics depends not only on the chemical-physical properties of aerosol particles, but also on fluid transport (e.g., concentration and temperature). Here, an artificial case, including general aerosol dynamics, is constructed to investigate the overall numerical simulation performance of the various stochastic methods. This case is tested to include the essential characteristics of general aerosol evolution to the extent possible. Aerosol particles are supposed to generate from quick nucleation and subsequently undergo condensational growth and coagulation. Nucleated particles are assumed to be uniformly distributed with the smallest unit volume, and nucleation is assumed to occur instantly. Because multiple nucleation events are common in reality [64, 66, 67], two nucleation events are assumed to occur in the testing case: one at the very beginning and the other at the middle of the entire time span of the numerical simulation. Because of the wide particle size spec-trum, three particle size regimes are generally defined: free molecular, transition, and continuum regimes; for each regime, condensation and coagulation are modeled differently. Aerosol particles are considered in the free molecular, continuum, and transition regimes if the particle sizes are considerably smaller, considerably larger, and midsize to the air mean free path, respectively. For the air under the standard conditions, the mean free path is approximately 70 nm [68].

After considering all these factors, the non-dimensional simulation time, t, is set as 100; during this period, the nucleated particles of unit volume may grow to the order of 10^3 (this intends to mimic the particle diameter growth in reality from nanometer to micrometer). In the general dynamic equation (Eq. (33)), the nucleation term is modeled as

$$I_{\text{nuc}} = \iint [\delta(t-0) + \delta(t-50)]\delta(v-1)\,\mathrm{d}t\,\mathrm{d}v, \tag{34}$$

i.e., unit size particles are separately nucleated at $t = 0$ and 50. At each nucleation event, the number density of particles increases by 1.

The condensation term is modeled as follows

$$G(v) = \begin{cases} v^{2/3}, & v \leq 10 \\ v^{1/3}, & v \geq 100 \\ \text{harmonic mean}, & 10 < v < 100 \end{cases} \tag{35}$$

$$G_{\text{cond}} = -\frac{\partial n(v) G(v)}{\partial v}. \tag{36}$$

As commonly adopted in practice, the piecewise function for the condensational growth rate, $G(v)$, has a similar dependency on the particle size, v: $v \leq 10$ for the free molecular regime; $v \geq 100$ for the continuum regime; the harmonic mean of the two aforementioned regimes for the transition regime. It should be pointed out that not all factors (e.g., temperature and saturation pressure), other than the particle size, are considered in the present model. This is because they are not relevant to the main objective of investigating the stochastic methods in the present study.

The coagulation kernel function is given by the following

$$\beta(u,v) = \begin{cases} \left(\frac{1}{u} + \frac{1}{v}\right)^{1/2} \left(u^{1/3} + v^{1/3}\right)^2, & \max(u,v) \leq 10 \\ \left(\frac{1}{u^{1/3}} + \frac{1}{v^{1/3}}\right)(u^{1/3} + v^{1/3}), & \max(u,v) \geq 100 \\ \text{harmonic mean}, & 10 < \max(u,v) < 100 \end{cases}. \tag{37}$$

This parallels the condensation model in the treatment for different regimes.

At $t = 0$, particles of uniform size are generated from nucleation. Thereafter, the particles grow because of condensation and coagulation. When the size of particles are small (in the free molecular regime), the condensational growth rate in diameter is constant [64]; in the continuum regime, this growth rate is inversely proportional to the diameter. Hence, if only condensational growth is considered, then a dispersed distribution of particles tends to become uniform in diameter; if only coagulation is considered, then the PSD finally reaches a self-similar hump shape [69–71].

Figure 7 shows the ensemble averaged PSD at time instances $t = 20, 40, 60$, and 80 with three different scaling parameters $\alpha = 0, -1/2$, and -1 in the weight function. In these numerical simulations, the particle number, N, is 2000, and the repetition

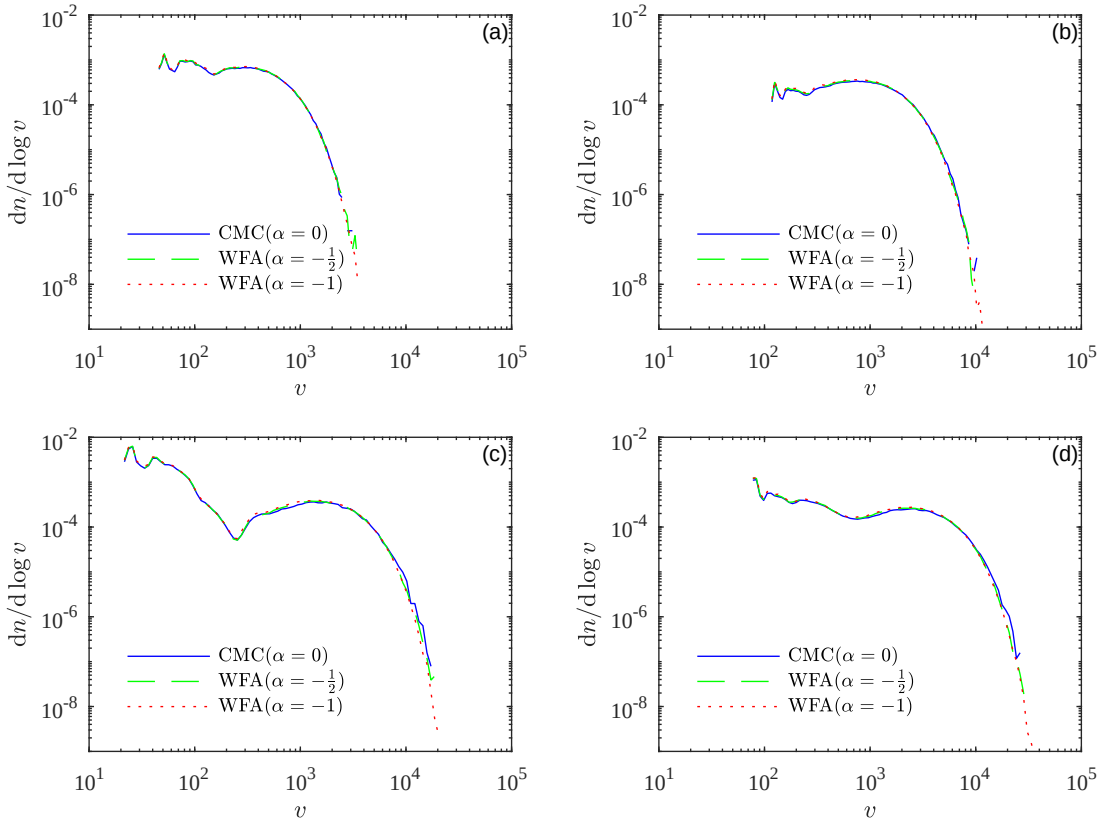

Figure 7: PSD for general aerosol dynamics at various time instances: (a) $t = 20$; (b) $t = 40$; (c) $t = 60$; (d) $t = 80$.

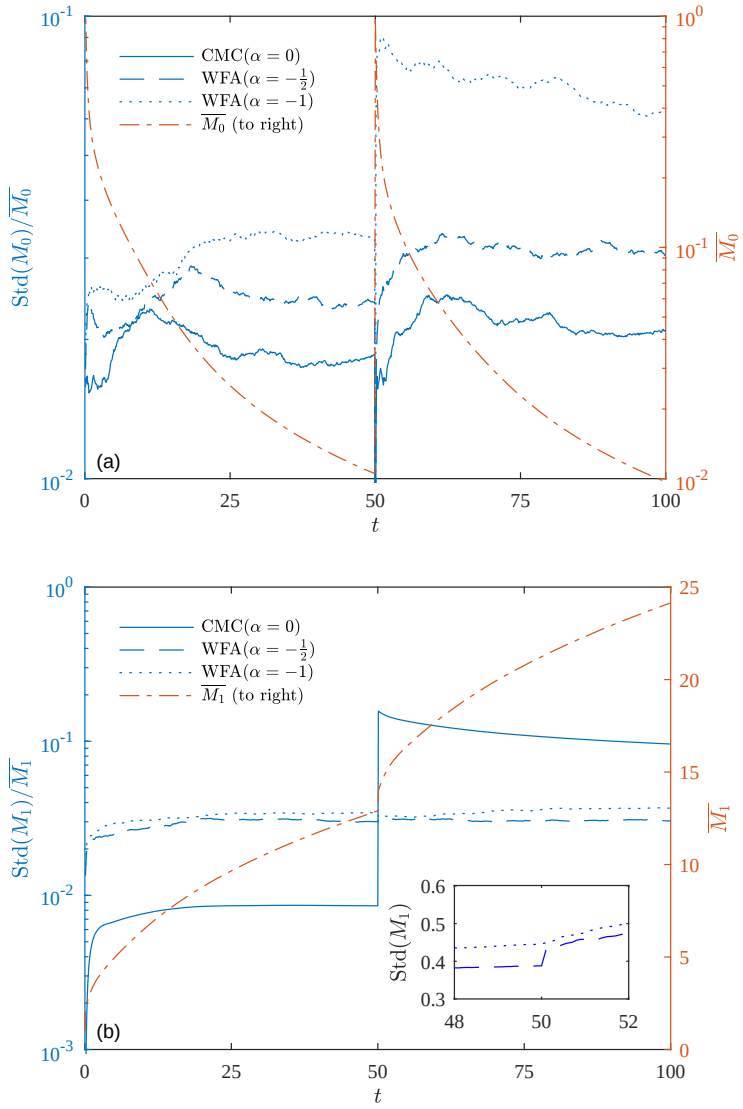

Figure 8: The evolution of the standard deviations of M_0 and M_1 in the general aerosol dynamics, which is normalized by $\overline{M_0}$ and $\overline{M_1}$, respectively. The y-axis on the right is for $\overline{M_0}$ or $\overline{M_1}$. The sharp jump at $t = 50$ corresponds to the second nucleation event, when a large number of particles are generated: (a) M_0; (b) M_1. The inset shows $\mathrm{Std}(M_1)$ at approximately $t = 50$ for $\alpha = -1$ (dotted line) and $\alpha = -1/2$ (dashed line).

number, N_{rep}, is 400. The parameter values $\alpha = 0, -1$, and $-1/2$ correspond to the CMC method (i.e., no differential weight), mass flow algorithm (i.e., differentially weighted by the particle volume), and intermediate weight function, respectively. Overall, the simulated PSDs for $\alpha = 0, -1/2$, and -1 agree extremely well, except that the numerical simulation for $\alpha = -1$ is able to resolve larger particles with a considerably low number density of particles than that for $\alpha = -1/2$ and $\alpha = 0$. The foregoing is the advocated advantage of the WFA method [37]. The PSD for small particles exhibits high fluctuation, which originates from the involved interaction between high coagulation rate and fast condensational growth of small particles. The PSD for large particles is extremely smooth and has a steep front on the right-hand side when the condensational growth tends to render a uniform PSD; this is because smaller particles have a larger growth rate to catch up with larger particles. From $t = 20$ to 40, the PSD gradually evolves to a hump shape because of coagulation; at $t = 50$, the second nucleation event occurs. Hence, in Fig. 7(c), a sudden increase in the number density of small particles is observed at $t = 60$, and a double-peaked PSD, which results from the combined effects of nucleation and coagulation, develops.

Although the ensemble averaged PSDs obtained from different algorithms are practically the same, the variances in these numerical simulations are considerably different. Figure 8(a) shows the evolution of $\text{Std}(M_0)/\overline{M_0}$, where Std denotes the standard deviation. The extent of the stochastic fluctuation of the number density of particles compared to its mean is reflected by $\text{Std}(M_0)/\overline{M_0}$. The evolution of $\overline{M_0}$ (the three methods generate nearly identical values of M_0) is also shown in the same plot (right y-axis). From $t = 0$ to 50, M_0 decreases by practically two orders of magnitude because of coagulation. During this period, the change in $\text{Std}(M_0)/\overline{M_0}$ is not significant and remains in the range of a few percent. The WFA method with $\alpha = -1$ exhibits the highest stochastic error; this error decreases as α varies from -1 to 0. At $t = 50$, $\overline{M_0}$ abruptly increases because of the second nucleation event; concurrently, $\text{Std}(M_0)/\overline{M_0}$ suddenly increases. In fact, the absolute stochastic error, $\text{Std}(M_0)$, is found to increase in these three methods at $t = 50$ (not shown here).

During the entire numerical simulation period, the stochastic error for M_0 in the WFA is persistently larger than that in the CMC method.

Figure 8(b) shows the evolution of $\mathrm{Std}(M_1)/\overline{M_1}$ together with $\overline{M_1}$. The volume fraction, $\overline{M_1}$, increases with time because of condensation with a small jump at $t = 50$ because of nucleation. The most striking difference among the stochastic errors from different methods is that the error in the CMC method is several times smaller than that in the WFA method between $t = 0$ and $t = 50$. On the other hand, a sharp jump occurs in the CMC method at $t = 50$, and the error in the CMC method increases several times larger than that in the WFA method when $t > 50$. Although pure coagulation or condensation does not exhibit stochasticity for M_1 with the CMC method, the combination of coagulation and condensation shows weak stochasticity. This is because the randomness in the PSD causes the volume fraction to grow stochastically. On the other hand, M_1 in pure coagulation with the WFA method is known to fluctuate. Moreover, pure condensational growth also incurs random fluctuation in the WFA method; such fluctuation is inversely proportional to the total number of numerical particles, N [37]. Accordingly, it is reasonable that the stochastic error in the CMC method is smaller than that in the WFA method. It is considerably interesting to observe a sharp jump at $t = 50$ in the CMC method but not in the WFA method; a detailed explanation is appropriate. Because the number density, $\overline{M_0}$, practically increases 100 times at $t = 50$, approximately $100N$ new numerical particles with the critical nucleation size should be generated, and approximately 99% of $(100 + 1)N$ numerical particles should be downsampled to leave only N particles in the numerical simulation. Accordingly, in the CMC method, based on the theoretical analysis in Eq. (24), such a high downsampling rate causes a large stochasticity on M_1 (and all higher-order moments). However, in the WFA method, with $\alpha = -1$, M_1 (Eq. (20)) is calculated as follows

$$M_1 = \frac{1}{V_0^{\mathrm{w}}} \sum_{i=1}^{N} v_i \frac{1}{v_i} = \frac{N}{V_0^{\mathrm{w}}}. \tag{38}$$

This is independent of the PSD. Thus, the resampling in the WFA method with $\alpha = -1$ has totally no effect on M_1. When $\alpha = -1/2$, M_1 is proportional to $\sum_{i=1}^{N} \sqrt{v_i}$

($v_i \geq 1$), which is considerably smaller than $\sum_{i=1}^{N} v_i$ in the CMC method. Hence, a relatively moderate increase is expected in Std(M_1) at $t = 50$ with $\alpha = -1/2$. The inset in Fig. 8(b) clearly shows a small jump when $\alpha - 1/2$; however, there is no observable change when $\alpha - 1$ at $t = 50$. A notable difference in Std(M_1) between the CMC and WFA because of resampling also exists in the kth moment when $k > 1$ (not shown here).

5. Conclusions

Both the systematic and stochastic errors in the stochastic simulations of a population balance equation (PBE) are comprehensively investigated for two typical stochastic methods (i.e., the classical Monte Carlo (CMC) method of Gillespie and the weighted flow algorithm (WFA)).

The systematic error depicts the extent to which a numerical simulation with a finite number of numerical particles, N, deviates from the true solution of the PBE. Over a short period, the correlation among particles becomes considerably small, and the systematic error is inversely proportional to N in all stochastic methods. The systematic error also tends to grow exponentially with time; hence, it is recommended to increase N with time if an extended simulation period is of interest.

The stochastic error, which is depicted by the variance in this study, is intrinsic to all stochastic simulations. A smaller variance is always desirable to reduce the uncertainty in the results. The present theoretical analysis and numerical simulations demonstrate that the stochastic error is inversely proportional to N under all investigated conditions (i.e., different coagulation kernels and numerical simulation algorithms). In order to compare the relative magnitude of the variance between the CMC and WFA methods, a concise relationship of the variances is derived analytically and verified numerically. The WFA method generally provides a smaller variance in the particle size distribution (PSD) when compared with the CMC method, which agrees with the other numerical simulations [72]. However, when considering the particle number density and volume fraction (i.e., lower-order moments of the

PSD), the CMC method exhibits a lower stochastic error than the WFA method.

In the stochastic simulation of general aerosol dynamics including nucleation, condensation, and coagulation, resampling is generally necessary to render an efficient and accurate numerical simulation. Although resampling introduces no bias on the PSD, additional stochasticity is usually added except for the doubling operation. A formula is derived to quantify the stochastic effect of resampling. An unexpected conclusion is that the variance caused by resampling is inversely proportional to N. In upsampling, the addition of more particles causes the variance to increase first and thereafter drop to zero when all particles are doubled. However, in downsampling, the removal of more particles always increases the variance more significantly. The variance introduced in resampling is considerably significant in general aerosol dynamics when nucleation is strong, especially in the CMC method.

Most importantly, this research study provides an explicit relation to connect the stochastic variances in the CMC and WFA. It also provides a quantitative description on the variance caused by the resampling process.

Acknowledgements

To complete this study, KZ and XJ equally contributed. This work was supported by the grants from the General Research Fund, Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 152663/16E), and the Central Research Grant of The Hong Kong Polytechnic University (Project No. B-Q54U).

Appendix A. Statistical Bias and Variance in Resampling

It is important to determine whether resampling introduces statistical bias to the underlying distribution. However, it is not easy to compare quantitatively the distributions before and after resampling. Therefore, the moments of the distributions are investigated instead.

It is generally presumed that the discrete distribution after resampling is statistically equivalent to the distribution prior to resampling (i.e., no statistical bias is introduced). A mathematical proof is easy to provide by considering the corresponding moments of the discrete distribution of N particles

$$M_k(N) = \frac{1}{N}\sum_{i=1}^{N} x_i^k. \tag{A.1}$$

where, x_i denotes the volume of particle i, N is the initial number of particles, and $M_k(N)$ is the kth ($k = 0, 1, 2, \ldots$) moment. It is worth pointing out that the moment definition is slightly different from Eq. (19), where a virtual volume V_0 is used in the prefactor instead of N. Since the virtual volume V_0 is determined by N and the corresponding number density (see context to Eq. (3)), V_0 is linearly proportional to N. The definition in Eq. (A.1) is more natural for a general discrete distribution without special pertinence to the PSD of aerosols. It also renders the following derivation slightly neater.

A particle x_ξ is randomly selected from the N particles (where the index ξ is a random variable taking value from 1, 2, ...N with even chance) and is added to the existing N particles. Then the kth moment of the $N+1$ particles is

$$M_k(N+1) = \frac{1}{N+1}\left(\sum_{i=1}^{N} x_i^k + x_\xi^k\right). \tag{A.2}$$

Take the ensemble average of M_k (i.e., averaged over large number of independent realizations)

$$\overline{M_k(N+1)} = \frac{1}{N+1}\left(\sum_{i=1}^{N} x_i^k + \overline{x_\xi^k}\right). \tag{A.3}$$

Since

$$\overline{x_\xi^k} = \overline{x^k} = \frac{1}{N}\sum_{i=1}^{N} x_i^k, \tag{A.4}$$

substitute it into Eq. (A.3), then

$$\begin{aligned}
\overline{M_k(N+1)} &= \frac{1}{N+1}\left(\sum_{i=1}^{N} x_i^k + \frac{1}{N}\sum_{i=1}^{N-1} x_i^k\right) \\
&= \frac{1}{N+1}\left(1+\frac{1}{N}\right)\sum_{i=1}^{N} x_i^k \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i^k \\
&= M_k(N). \qquad \text{(A.5)}
\end{aligned}$$

Therefore, the process of randomly adding one particle does not introduce statistical bias; the new distribution converges to the old distribution statistically. The above proof can be directly extended to the general case of adding or removing m particles.

In the process of adding m particles, every added particle is chosen randomly from the original N particles. The added particles are denoted as x_{ξ_i}, where ξ_i is a random number from 1 to N corresponding to the ith added particle. Then the kth moment of the new discrete distribution after adding m particles is a random variable, which is defined as

$$X_+ = \frac{1}{N+m}\left(\sum_{i=1}^{N} x_i^k + \sum_{i=1}^{m} x_{\xi_i}^k\right). \qquad \text{(A.6)}$$

Its variance can be evaluated as

$$
\begin{aligned}
\mathrm{Var}(X_+) =& \mathrm{E}\left(X_+^2\right) - \left[\mathrm{E}\left(X_+\right)\right]^2 \\
=& \frac{1}{(N+m)^2}\mathrm{E}\left[\left(\sum_{i=1}^{N} x_i^k + \sum_{i=1}^{m} x_{\xi_i}^k\right)^2\right] - \left[\mathrm{E}\left(X_+\right)\right]^2 \\
=& \frac{1}{(N+m)^2}\mathrm{E}\left[\left(\sum_{i=1}^{N} x_i^k\right)^2 + 2\left(\sum_{i=1}^{m} x_{\xi_i}^k\right)\left(\sum_{i=1}^{N} x_i^k\right) + \left(\sum_{i=1}^{m} x_{\xi_i}^k\right)^2\right] - \left[\mathrm{E}\left(X_+\right)\right]^2 \\
=& \frac{1}{(N+m)^2}\left[\left(\sum_{i=1}^{N} x_i^k\right)^2 + 2\mathrm{E}\left(\sum_{i=1}^{m} x_{\xi_i}^k\right)\left(\sum_{i=1}^{N} x_i^k\right) + \mathrm{E}\left(\left[\sum_{i=1}^{m} x_{\xi_i}^k\right]^2\right)\right] - \left[\mathrm{E}\left(X_+\right)\right]^2 \\
=& \frac{1}{(N+m)^2}\left[N^2[\mathrm{E}\left(X_+\right)]^2 + 2[m\mathrm{E}\left(X_+\right)][N\mathrm{E}\left(X_+\right)] + \mathrm{E}\left(\left[\sum_{i=1}^{m} x_{\xi_i}^k\right]^2\right)\right] - \left[\mathrm{E}\left(X_+\right)\right]^2 \\
=& \frac{1}{(N+m)^2}\left[\mathrm{E}\left(\left[\sum_{i=1}^{m} x_{\xi_i}^k\right]^2\right) - m^2[\mathrm{E}\left(X_+\right)]^2\right] \\
=& \frac{1}{(N+m)^2}\left[\mathrm{E}\left(\left[\sum_{i=1}^{m} x_{\xi_i}^k\right]^2\right) - \left[\mathrm{E}\left(mX_+\right)\right]^2\right] \\
=& \frac{1}{(N+m)^2}\mathrm{Var}\left(\sum_{i=1}^{m} x_{\xi_i}^k\right) \qquad \text{(A.7)}
\end{aligned}
$$

From line three to line four when expanding the first expectation term, the expectation sign for a deterministic term is dropped off. From line four to line five, the conclusion (Eq. (A.5)) and relevant discussion) that adding particles do not change the expectation is applied to obtain the first two terms in the square brackets.

It is known that (page 12 of Lange [73])

$$
\mathrm{Var}\left(\sum_{i=1}^{m} x_{\xi_i}^k\right) = \sum_{i=1}^{m}\mathrm{Var}(x_{\xi_i}^k) + \sum_{i=1}^{m}\sum_{j\neq i}\mathrm{Cov}\left(x_{\xi_i}^k, x_{\xi_j}^k\right), \qquad \text{(A.8)}
$$

where

$$
\mathrm{Cov}\left(x_{\xi_i}^k, x_{\xi_j}^k\right) = \mathrm{E}\left(x_{\xi_i}^k x_{\xi_j}^k\right) - \mathrm{E}\left(x_{\xi_i}^k\right)\mathrm{E}\left(x_{\xi_j}^k\right). \qquad \text{(A.9)}
$$

Evidently, all $\mathrm{Var}(x_{\xi_i}^k)$, $(i = 1 \ldots m)$ have the same value, which is denoted as $\mathrm{Var}(x_\xi^k)$

$$\begin{aligned}\mathrm{Var}(x_\xi^k) =& \mathrm{E}\left[(x_\xi^k)^2\right] - \left[\mathrm{E}\left(x_\xi^k\right)\right]^2 \\ =& \frac{1}{N}\sum_{i=1}^{N}(x_i^k)^2 - (\overline{x^k})^2 \\ =& \frac{1}{N}\sum_{i=1}^{N}(x_i^k - \overline{x^k})^2 \end{aligned} \tag{A.10}$$

Based on symmetry, all $\mathrm{Cov}\left(x_{\xi_i}^k, x_{\xi_j}^k\right)$ for $(i \neq j)$ have the same value, which is found to be [73, page 27]

$$\mathrm{Cov}\left(x_{\xi_i}^k, x_{\xi_j}^k\right) = -\frac{1}{N-1}\mathrm{Var}(x_\xi^k) \tag{A.11}$$

Accordingly, substitute Eqs. (A.11), (A.8), and (A.10) into (A.7); this substitution results in the variance of the kth moment of the discrete distribution after adding m particles:

$$\mathrm{Var}(X_+) = \frac{m(N-m)}{(N+m)^2 N}\mathrm{Var_p}[M_k(N)], \tag{A.12}$$

where $\mathrm{Var_p}[M_k(N)]$ is the unbiased estimator of the variance of the kth moment

$$\mathrm{Var_p}[M_k(N)] = \frac{1}{N-1}\sum_{i=1}^{N}(x_i^k - \overline{x^k})^2. \tag{A.13}$$

Here, the variance is calculated for a given discrete distribution. A subscript, p, is attached to the variance symbol in order to distinguish it from the variance calculated for a stochastic variable. Specifically, if $k = 0$, $\mathrm{Var_p}[M_0(N)]$ is zero. Hence, resampling does not introduce stochasticity to M_0, which originates from the inclusion of the factor $1/(N+m)$ in Eq. (A.6). In other words, the total number density does not change during a stochastic resampling operation. If $k \neq 0$, $\mathrm{Var_p}[M_0(N)]$ is generally non-zero (except for a uniform distribution); resampling introduces stochasticity to the kth moment when $k \neq 0$.

Similarly, the variance of randomly removing m particles is

$$\mathrm{Var}(X_-) = \frac{m}{(N-m)N}\mathrm{Var_p}[M_k(N)], \tag{A.14}$$

where

$$X_- = \frac{1}{N-m}\left(\sum_{i=1}^{N} x_i^k - \sum_{i=1}^{m} x_{\xi_i}^k\right). \tag{A.15}$$

The key conclusions (Eqs. (A.12) and (A.14)) have been verified numerically in the main content.

References

[1] M. Smoluchowski, "Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen," Phys. Z. **17**, 557–585 (1916).

[2] D. Ramkrishna, *Population Balances: Theory and Applications to Particulate Systems in Engineering* (Academic Press, San Diego, 2000).

[3] S. K. Friedlander, *Smoke, Dust, and Haze: Fundamentals of Aerosol Dynamics* (Oxford University Press, New York, 2000), second edition.

[4] D. T. Gillespie, "Stochastic simulation of chemical kinetics," Annu. Rev. Phys. Chem. **58**, 35–55 (2007).

[5] D. J. Aldous, "Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists," Bernoulli **5**, 3–48 (1999).

[6] T. E. Ramabhadran, T. W. Peterson, and J. H. Seinfeld, "Dynamics of aerosol coagulation and condensation," AIChE J. **22**(5), 840–851 (1976).

[7] S. C. Davies, J. R. King, and J. A. D. Wattis, "The Smoluchowski coagulation equations with continuous injection," J. Phys. A: Math. Gen. **32**, 7745–7763 (1999).

[8] F. Gelbard, Y. Tambour, and J. H. Seinfeld, "Sectional representations for simulating aerosol dynamics," J. Colloid Interface Sci. **76**(2), 541–556 (1980).

[9] S. Rigopoulos and A. G. Jones, "Finite-element scheme for solution of the dynamic population balance equation," AIChE J. **49**(5), 1127–1139 (2003).

[10] F. Sewerin and S. Rigopoulos, "An explicit adaptive grid approach for the numerical solution of the population balance equation," Chem. Eng. Sci. **168**, 250–270 (2017).

[11] M. Engh, J. Solsvik, and H. A. Jakobsen, "An hp-adaptive quadrature method for irregular integrands: Application to the population balance equation birth term," Chem. Eng. Sci. **204**, 161–176 (2019).

[12] A. Alexopoulos, A. Roussos, and C. Kiparissides, "Part I: Dynamic evolution of the particle size distribution in particulate processes undergoing combined particle growth and aggregation," Chem. Eng. Sci. **59**(24), 5751–5769 (2004).

[13] J. Solsvik and H. A. Jakobsen, "Evaluation of weighted residual methods for the solution of a population balance model describing bubbly flows: the least-squares, Galerkin, tau, and orthogonal collocation methods," Ind. Eng. Chem. Res. **52**(45), 15988–16013 (2013).

[14] Y. P. Kim and J. H. Seinfeld, "Simulation of multicomponent aerosol dynamics," J. Colloid Interface Sci. **149**(2), 425–449 (1991).

[15] S. Kumar and D. Ramkrishna, "On the solution of population balance equations by discretization-I. A fixed pivot technique," Chem. Eng. Sci. **51**(8), 1311–1332 (1996).

[16] J. Kumar, M. Peglow, G. Warnecke, and S. Heinrich, "An efficient numerical technique for solving population balance equation involving aggregation, breakage, growth and nucleation," Powder Technol. **182**(1), 81–104 (2008).

[17] S. E. Pratsinis, "Simultaneous nucleation, condensation, and coagulation in aerosol reactors," J. Colloid Interface Sci. **124**(2), 416–427 (1988).

[18] R. McGraw, "Description of aerosol dynamics by the quadrature method of moments," Aerosol Sci. Technol. **27**(2), 255–265 (1997).

[19] D. L. Marchisio and R. O. Fox, "Solution of population balance equations using the direct quadrature method of moments," J. Aerosol Sci. **36**(1), 43–73 (2005).

[20] M. Frenklach, "Method of moments with interpolative closure," Chem. Eng. Sci. **57**(12), 2229–2239 (2002).

[21] M. Z. Yu, J. Z. Lin, and T. L. Chan, "A new moment method for solving the coagulation equation for particles in Brownian motion," Aerosol Sci. Technol. **42**(9), 705–713 (2008).

[22] M. Z. Yu and T. L. Chan, "A bimodal moment method model for submicron fractal-like agglomerates undergoing Brownian coagulation," J. Aerosol Sci. **88**, 19–34 (2015).

[23] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (American Mathematical Society, Providence, Rhode Island, 1970), revised edition.

[24] D. L. Wright Jr, "Numerical advection of moments of the particle size distribution in Eulerian models," J. Aerosol Sci. **38**, 352–369 (2007).

[25] G. D. Lin and J. Stoyanov, "On the moment determinacy of the distributions of compound geometric sums," Journal of Applied Probability **39**(3), 545–554 (2002).

[26] R. M. Mnatsakanov and A. S. Hakobyan, "Recovery of distributions via moments," in *Optimality: The Third Erich L. Lehmann Symposium*, pp. 252–265 (Institute of Mathematical Statistics, 2009).

[27] G. D. Lin, "Recent developments on the moment problem," Journal of Statistical Distributions and Applications **4**(1), 5 (2017).

[28] D. T. Gillespie, "An exact method for numerically simulating the stochastic coalescence process in a cloud," J. Atmos. Sci. **32**(10), 1977–1989 (1975).

[29] E. Debry, B. Sportisse, and B. Jourdain, "A stochastic approach for the numerical simulation of the general dynamics equation for aerosols," J. Comput. Phys. **184**, 649–669 (2003).

[30] A. Maisels, F. E. Kruis, and H. Fissan, "Direct simulation Monte Carlo for simultaneous nucleation, coagulation, and surface growth in dispersed systems," Chem. Eng. Sci. **59**, 2231–2239 (2004).

[31] K. Zhou, Z. He, M. Xiao, and Z. Zhang, "Parallel Monte Carlo simulation of aerosol dynamics," Adv. Mech. Eng. **2014**, 435936 (2014).

[32] D. E. Rosner and S. Yu, "MC simulation of aerosol aggregation and simultaneous spheroidization," AIChE J. **47**, 545–561 (2001).

[33] I. J. Laurenzi, J. D. Bartels, and S. L. Diamond, "A general algorithm for exact simulation of multicomponent aggregation processes," J. Comput. Phys. **177**, 418–449 (2002).

[34] L. Alfonso, G. B. Raga, and D. Baumgardner, "Monte Carlo simulations of two-component drop growth by stochastic coalescence," Atmos. Chem. Phys. **9**, 1241–1251 (2009).

[35] M. Sander, R. H. West, M. S. Celnik, and M. Kraft, "A detailed model for the sintering of polydispersed nanoparticle agglomerates," Aerosol Science and Technology **43**(10), 978–989 (2009).

[36] H. Zhao, F. E. Kruis, and C. Zheng, "A differentially weighted Monte Carlo method for two-component coagulation," J. Comput. Phys. **229**, 6931–6945 (2010).

[37] R. DeVille, N. Riemer, and M. West, "Weighted Flow Algorithms (WFA) for stochastic particle coagulation," J. Comput. Phys. **230**(23), 8427–8451 (2011).

[38] S. Shekar, W. J. Menz, A. J. Smith, M. Kraft, and W. Wagner, "On a multivariate population balance model to describe the structure and composition of silica nanoparticles," Comput. Chem. Eng. **43**, 130–147 (2012).

[39] S. A. Matveev, D. A. Zheltkov, E. E. Tyrtyshnikov, and A. P. Smirnov, "Tensor train versus Monte Carlo for the multicomponent Smoluchowski coagulation equation," Journal of Computational Physics **316**, 164–179 (2016).

[40] H. M Liu and T. L. Chan, "Two-component aerosol dynamic s imulation using differentially weighted operator splitting Monte Carlo methods" Appl. Math. Modell. **62**, 237–253 (2018).

[41] G. Kaur, M. Singh, T. Matsoukas, J. Kummar, T. Beer, and I . Nopens, "Two-compartment modeling and dynamics of top-sprayed fluidized bed granulator," Appl. Math. Modell. **68**, 267–280 (2019).

[42] C. S. Lindberg, M. Y. Manuputty, E. K. Y. Yapp, J. Akroyd, R. Xu, and M. Kraft, "A detailed particle model for polydisperse aggregate particles," Journal of Computational Physics **397**, 108799 (2019).

[43] G. Blanquart and H. Pitsch, "Analyzing the effects of temperature on soot formation with a joint volume-surface-hydrogen model," Combust. Flame **156**, 1614–1626 (2009).

[44] R. Bellman, *Dynamic Programming* (Princeton University Press, 1957).

[45] H. Babovsky, "On a Monte Carlo scheme for Smoluchowski's coagulation equation," Monte Carlo Methods Appl. **5**(1), 1–18 (1999).

[46] A. Eibeck and W. Wagner, "Stochastic particle approximations for Smoluchoskis coagulation equation," Ann. Appl. Probab. **11**, 1137–1165 (2001).

[47] H. Zhao and C. Zheng, "A new event-driven constant-volume method for solution of the time evolution of particle size distribution," J. Comput. Phys. **228**, 1412–1428 (2009).

[48] H. M Liu and T. L. Chan, "Differentially weighted operator splitting Monte Carlo method for simulating complex aerosol dynamic processes," Particuology **36**, 114–126 (2018).

[49] R. Irizarry, "Fast Monte Carlo methodology for multivariate particulate systems-I: Point ensemble Monte Carlo," Chem. Eng. Sci. **63**(1), 95–110 (2008).

[50] J. H. Curtis, M. D. Michelotti, N. Riemer, M. T. Heath, and M. West, "Accelerated simulation of stochastic particle removal processes in particle-resolved aerosol models," J. Comput. Phys. **322**, 21–32 (2016).

[51] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," J. Chem. Phys. **115**, 1716–1733 (2001).

[52] R. Irizarry, "Fast Monte Carlo methodology for multivariate particulate systems-II: -PEMC," Chem. Eng. Sci. **63**(1), 111–121 (2008).

[53] A. Eibeck and W Wagner, "An efficient stochastic algorithm for studying coagulation dynamics and gelation phenomena," SIAM J. Sci. Comput. **22**(3), 802–821 (2000).

[54] M. Goodson and M Kraft, "An efficient stochastic algorithm for simulating nanoparticle dynamics," J. Comput. Phys. **183,**(1), 210–232 (2002).

[55] J. Wei and F. Kruis, "GPU-accelerated Monte Carlo simulation of particle coagulation based on the inverse method," J. Comput. Phys. **249**(0), 67–79 (2013).

[56] S. Streltsov and P. Vakili, "Variance reduction algorithms for parallel replicated simulation of uniformized Markov chains," Discrete Event Dynamic Systems: Theory and Applications **6**(2), 159–180 (1996).

[57] G. Kotalczyk and F. E. Kruis, "A Monte Carlo method for the simulation of coagulation and nucleation based on weighted particles and the concepts of stochastic resolution and merging," Journal of Computational Physics **340**, 276–296 (2017).

[58] M. Smith and T. Matsoukas, "Constant-number Monte Carlo simulation of population balance," Chem. Eng. Sci. **53**, 1777–1786 (1998).

[59] M. H. Bayewitz, J. Yerushalmi, S. Katz, and R. Shinnar, "The extent of correlations in a stochastic coalescence process," J. Atmos. Sci. **31**(6), 1604–1614 (1974).

[60] D. Gillespie, "The stochastic coalescence model for cloud droplet growth," J. Atmos. Sci. **29**(8), 1496–1510 (1972).

[61] R. L. Drake, "A general mathematical survey of the coagulation equation," in G. M. Hidy and J. R. Brock, editors, *Topics in Current Aerosol Research*, pp. 201–376 (Pergamon Press, New York, 1972).

[62] X. Jiang, K. Zhou, M. Xiao, K. Sun, and Y. Wang, "Stochastic simulation of soot formation evolution in counterflow diffusion flames," J. Nanotechnology **2018**(9479582), 1 (2018).

[63] A. Kolodko and K. Sabelfeld, "Stochastic particle methods for Smoluchowski coagulation equation: variance reduction and error estimations," Monte Carlo Methods Appl. **9**(4), 315–339 (2003).

[64] K. Zhou, A. Attili, A. Alshaarawi, and F. Bisetti, "Simulation of aerosol nucleation and growth in a turbulent mixing layer," Phys. Fluids **26**, 065106 (2014).

[65] K. Zhou and T. L. Chan, "Operator Splitting Monte Carlo Method for Aerosol Dynamics," in K. Volkov, editor, *Aerosols - Science and Case Studies*, chapter 1 (InTech, Croatia, 2016).

[66] K. Zhou, X. Jiang, K. Sun, and Z. He, "Eulerian-Lagranigan simulation of aerosol evolution in turbulent mixing layer," Appl. Math. Mech. -Engl. Ed. **37**(10), 1305–1314 (2016).

[67] A. Attili, F. Bisetti, M. E. Mueller, and H. Pitsch, "Formation, growth, and transport of soot in a three-dimensional turbulent non-premixed jet flame," Combust. Flame **161**(7), 1849–1865 (2014).

[68] S. Jennings, "The mean free path in air," J. Aerosol Sci. **19**(2), 159–166 (1988).

[69] F. S. Lai, S. K. Friedlander, J. Pich, and G. M. Hidy, "The self-preserving particle size distribution for Brownian coagulation in the free-molecule regime," J. Colloid Interface Sci. **39**(2), 395–405 (1972).

[70] S. Vemury, K. A. Kusters, and S. E. Pratsinis, "Time-lag for attainment of the self-preserving particle size distribution by coagulation," J. Colloid Interface Sci. **165**(1), 53–59 (1994).

[71] S. Vemury and S. E. Pratsinis, "Self-preserving size distributions of agglomerates," J. Aerosol Sci. **26**(2), 175–185 (1995).

[72] R. I. A. Patterson, W. Wagner, and M. Kraft, "Stochastic weighted particle methods for population balance equations," J. Comput. Phys. **230**(19), 7456–7472 (2011).

[73] K. Lange, *Applied Probability* (Springer, New York, 2010), second edition.