Accuracy and Optimal Sampling in Monte Carlo Solution of Population Balance Equations

Xi Yu[1, 2,*], Michael J Hounslow[2], Gavin K Reynolds[3]

1 *European Bioenergy Research Institute (EBRI), School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, UK*
2 *Department of Chemical and Biological Engineering, the University of Sheffield, Sheffield, S1 3JD, UK*
3 *Pharmaceutical Development, AstraZeneca, Macclesfield, SK10 2NA, UK*

[*] Corresponding concerning this article should be addressed to Xi Yu at **x.yu3@aston.ac.uk**

Abstract

Implementation of a Monte Carlo simulation for the solution of population balance equations requires choice of initial sample number (N_0), number of replicates (M) and number of bins for probability distribution reconstruction (n). It is found that Squared Hellinger Distance, H^2, is a useful measurement of the accuracy of MC simulation, and can be related directly to N_0, M and n. Asymptotic approximations of H^2 are deduced and tested for both 1D and 2D PBEs with coalescence. The CPU cost, C, is found in a power-law relationship, $C= aMN_0^b$, with the CPU cost index, b, indicating the weighting of N_0 in the total CPU cost. n must be chosen to balance accuracy and resolution. For fixed n, $M \times N_0$ determines the accuracy of MC prediction; if b>1, then the optimal solution strategy uses multiple replications and small sample size. Conversely if 0<b<1, one replicate and a large initial sample size is preferred.

Keywords: *Monte Carlo, Population Balance Model, Hellinger Distance, Optimal Sampling, Accuracy, Coalescence*

Introduction

Population balance equations (PBEs) describe the evolution of the properties of a collection of particles (eg crystals, agglomerates, soot) in time and perhaps space[1-3]. Such equations usually require numerical solution frequently via a stochastic technique. Monte Carlo simulation (MCS) has been used as such a method over the past few decades[4, 5]. With this approach a large population of particles, perhaps of $O(10^9)$ is represented by a small sample, perhaps $O(10^3)$. Each particle is then simulated by evolving its properties (or internal coordinates, such as size or composition), via mechanisms that involve interaction between particles, selected in some random way, hence the analogy with Monte Carlo methods.

For a coalescence phenomenon in a closed system, as described in this paper, the number of particles will decline over time. Therefore, if the size of the simulation box (ie the apparent size of the space represented by the sample particles) is kept constant, the number of particles used to represent the real system will also decline. Two event-driven methods, stepwise constant volume Monte Carlo (SCVMC)[6, 7] and constant number Monte Carlo (CNMC)[8] have been devised to circumvent this problem and are widely used to solve PBEs. In the SCVMC method, the volume is halved or doubled when the particle number increases or decreases by a factor of two of its initial value, respectively. In the CNMC approach, the volume is continuously adjusted to keep number of particle constant in the virtual simulation box. Maisels et al[7] demonstrated the prediction of the SCVMC method is more accurate than the CNMC method for nucleation and coagulation problems.

The choice of N_0, M and n

The principle of the stochastic MC method for solution of PBEs is that the dynamic evolution of an extremely large population of particles can be represented by monitoring the corresponding discrete events occurring in a smaller number of sample particles. Therefore, sampling a finite number of particles appropriately is crucial to describing population dynamics and prediction of product quality in real systems. Three essential parameters, including initial sample number (N_0), number of replicates (M) and number of size bins (n), need to be chosen if we run an MC simulation and compare with theoretical or experimental data. The number of replicates (M) in this study indicates how many times an MC simulation needs to be run. A review of the published sampling strategies using the MC method from a

range of applications is shown in Table 1, the scope of this survey covers N_0, M, CPU cost and goodness of fit. Almost all the MC applications examine 10^3-10^6 particles at a time on PCs of different ages because of the limitation of CPU speed and memory capacity. Fewer studies consider the computation time of the MC algorithm for some specific application on PCs of different ages. Some researchers[13, 17] comment qualitatively that the accuracy of MC (the relative error between predicted and theoretical values) is proportional to $1/\sqrt{N}$ where N is the number of particles in the system. Smith and Matsoukas[4] quantitatively proved the correlation between MC error, δ and $1/\sqrt{N}$ by fitting simulation results. A successful representative population in accuracy-constrained MC simulation can be achieved in two ways: by running the MC program once with a very large initial sample number, or by combining the results of several runs, each with a smaller sample number[14]. However, none of previous publications specify how to choose N_0, M and n to achieve a specific accuracy with acceptable computational cost. In this study, we aim to

I. use the Hellinger Distance, a statistical distance between two probability distributions, to measure the accuracy of Monte Carlo simulations.

II. provide guidance on how to select the initial sample number and number of replicates and consider tradeoffs between accuracy and computational costs.

Table 1: Summary of application of Monte Carlo method to PBEs in the literatures

Application (Ref. and Authors)	N_0 (M)	Computational time	Goodness of fit
Bipolar charging[7]	—	—	—
Coalescence in a cloud[10,11]	—	—	—
Coupled chemical reaction[12]	—	—	—
chemical reaction and coating[13]	3×10^4(4) 2×10^4(5)	—	Qualitative comment on accuracy via $1/\sqrt{N}$
Crystallization[14]	6400-25600 (1-250)	3-620mins in PCs	Good statistical representative is achieved in two ways
Fractal aggregation[15]	3-100 (1000)	—	—
Higher dimensionality problems[16]	—	—	—
Multi-component aerosols[17]	3×10^6(10)	A few hours(256Mb RAM and Pentium III 900-Mhz pc)	Qualitative comment on accuracy via $1/\sqrt{N}$
Particle Aggregation and restructuring[16,18]	10^4(15) 10^4(100)	2h (SPRAC -ten workstation) —	Average/ensemble method Average procedure
Wet scavenging[19]	3000	25-78s in PCs	—
Wet Granulation[20-25]	1024-4096(64)	—	—

Theory

Squared Hellinger Distance

The Hellinger Distance is used to quantify the similarity between two probability distributions. The most ubiquitous application of Hellinger distance is minimum Hellinger distance estimation[26-29] in statistics.

Squared Hellinger Distance (H^2) between distribution function f and g is

$$H^2(f,g) = \frac{1}{2} \int \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx \tag{1}$$

where $f(x), g(x)$ are probability density functions (PDF), describing the frequency of occurrence at size x (1/m), where x is particle size (m). If *f* and *g* are identical, $H^2 = 0$; if the two distributions do not overlap at all, $H^2 = 1$. In this way H^2 provides a scaled, dimensionless measurement of accuracy that ranges between 0 and 1.

The PDFs in Eq. 1 are both continuous and normalized (i.e. have a zeroth moment of 1) whereas the results from MC simulation are discrete and not normalized. An appropriate modification to Eq. 1 to allow for comparison of discrete MC results with continuous analytical results is:

$$H^2(f_{MC}, f_{AS}) = \frac{1}{2} \sum_i \left(\sqrt{\frac{N_i}{\sum_{C_i} N_i}} - \sqrt{\frac{\int_{C_i} n(\vec{x}, t) d\vec{x}}{\int_{\forall \vec{x}} n(\vec{x}, t) d\vec{x}}} \right)^2 \tag{2}$$

$$f_{MC}(\vec{x}) = \frac{N_i}{\Delta \vec{x}_i \sum_{C_i} N_i}, f_{AS}(\vec{x}) = \frac{n(\vec{x}, t)}{\int \int_{\forall \vec{x}} n(\vec{x}, t) d\vec{x}} \tag{3}$$

Where N_i is number in size bin C_i in the MC simulation, subscript AS refers to the analytical solution.

Coalescence PBEs

In this study, two cases, 1D size-dependent and 2D size-independent coalescence PBEs with analytical solutions are examined. These cases are selected as they present significant differences in algorithm structure and evolving distribution of particle properties, which make a major impact on the correlations of computational time and accuracy with N_0, and *M* in MCS.

Case1: 1D size-dependent

Gelbard and Seinfeld[30] produced a result for the coalescence of an exponential distribution with a kernel given by $\beta(m_1, m_2) = \beta_0(m_1 + m_2)$. Here, m_1 and m_2 are the mass of colliding particles. The initial distribution function of particles volume is

$$n(m, 0) = \frac{n_0}{m_0} Exp\left(-\frac{m}{m_0}\right) \tag{4}$$

The analytical solution for population density function at time t is

$$n(m, t) = \frac{n_0(1 - T(t))}{m\sqrt{T(t)}} Exp\left(-(1 - T(t))\frac{m}{m_0}\right) I_1\left(2\frac{m}{m_0}\sqrt{T(t)}\right) \tag{5}$$

$$T(t) = 1 - Exp(-n_0 b_0 m_0 t) \tag{6}$$

Where I_1 is Bessel function of the first kind of order one. $n(m, 0), n(m, t)$ are number density functions (NDF) at time 0 and t, respectively.

Case2: 2D size-independent

A two dimensional analytical solution for the size independent coalescence PBE is revealed by *Vale and McKenna*[31], in view of solution proposed by *Gelbard and Seinfeld*[30]. The initial distribution of particles which have two components in mass mode is

$$n(m_1, m_2, 0) = \frac{16 n_0}{m_{10} m_{20}}\left(\frac{m_1}{m_{10}}\right)\left(\frac{m_2}{m_{20}}\right) Exp\left(-2\frac{m_1}{m_{10}} - 2\frac{m_2}{m_{20}}\right) \tag{7}$$

Where n_0 is initial number of particles per unit volume ; m_{i0} is initial mean mass of the ith component in a particle.

For a constant coalescence coefficient β_0, the analytical solution is

$$n(m_1, m_2, \tau) = \frac{8 n_0}{m_{10} m_{20}\sqrt{\tau(\tau + 2)^3}} Exp\left(-2\frac{m_1}{m_{10}} - 2\frac{m_2}{m_{20}}\right) (I_0(\theta) - J_0(\theta)) \tag{8}$$

$$where\ \theta = 4\left(\frac{m_1 m_2}{m_{10} m_{20}}\right)^{1/2}\left(\frac{\tau}{\tau + 2}\right)^{1/2} \tag{9}$$

$$\tau = n_0 \beta_0 t \tag{10}$$

Where, $J_0(\theta)$ is Bessel function of the first kind; $I_0(\theta)$ is Modified Bessel function of the first kind.

Table 2 lists the parameters used in the MC simulation.

Table 2: Parameters used in MC simulation

1D Size dependent		2D Size independent	
parameter	Value	parameter	value
β_0	1	β_0	1
n_0	1	n_0	1
b_0	1	m_{10}	1
m_0	1	m_{20}	5
t	6	τ	100
l_{agg}	0.99	l_{agg}	0.99
N_0	50-5,000	N_0	500-50,000
M	1-20	M	1-20
n	13-13,000	n	7^2-350^2

Asymptotic approximation of H²

We consider now two asymptotic cases: that where the number of particles in all size ranges is large, and that when it is small. The expected value of a discrete approximation to the H^2 is based on the following assumptions: (1) the number of particles, $\widehat{N}_k$, in each size range C_k in MC results has a Poisson distribution; (2) A group of $\widehat{N}_k$'s are uncorrelated amongst themselves over the domain.

To obtain the expected value of a discrete approximation to the H^2. Eq. 2 is written as,

$$H^2 = \frac{1}{2MN_0} \sum_{k=1}^{n} \left(\sqrt{N_k} - \sqrt{\widehat{N}_k} \right)^2 \tag{11}$$

Where N_k is the number associated with k^{th} internal and $\widehat{N}_k$ is an estimate of that number and

$$MN_0 = \sum_{k=1}^{n} N_k = \sum_{k=1}^{n} \widehat{N}_k \tag{12}$$

If we assume that the values of $\widehat{N}_k$ have a Poisson distribution of N_k , i.e. the probability that $\widehat{N}_k$ takes on a value x is

$$P_{\widehat{N}_k}(x) = \frac{e^{-N_k} N_k^x}{x!} \tag{13}$$

Then the expected value of H^2 is

$$E[H^2] = \frac{1}{2MN_0} E\left[\sum_{k=1}^{n} \left(\sqrt{N_k} - \sqrt{\widehat{N}_k} \right)^2 \right] = \frac{1}{2MN_0} \sum_{k=1}^{n} \left(N_k - 2\sqrt{N_k} E\left[\sqrt{\widehat{N}_k} \right] + E\left[\widehat{N}_k \right] \right)$$

$$= \frac{1}{N_0} \sum_{k=1}^{n} \left(N_k - \sqrt{N_k} E\left[\sqrt{\widehat{N_k}}\right] \right) \tag{14}$$

However if N_k is everywhere small $E\left[\sqrt{\widehat{N_k}}\right] = N_k$ so

$$E[H^2] = \frac{1}{N_0} \sum_{k=1}^{n} (N_k - N_k\sqrt{N_k}) = \frac{1}{MN_0} \sum_{k=1}^{n} (N_k - N_k\sqrt{N_k}) = 1 - \frac{1}{MN_0} \sum_{k=1}^{n} N_k\sqrt{N_k} \tag{15}$$

Now, the N_k scale with MN_0/n so put $MN_0 N_k^1/n$ where the N_k^1 are constants. Therefore

$$E[H^2] = 1 - \left(\frac{MN_0}{n}\right)^{0.5} \sum_{k=1}^{n} N_k^1 \sqrt{N_k^1} = 1 - a'\left(\frac{MN_0}{n}\right)^{0.5} \tag{16}$$

Where a' is a constant independent of MN_0 and n.

If N_k is everywhere large, the expected value of $\sqrt{\widehat{N_k}}$ is given by Kendall and Stuart[32],

$$E\left[\sqrt{\widehat{N_k}}\right] = \sqrt{N_k} - \frac{1}{8} N_k^{-1/2} - o\left(N_k^{-3/2}\right) \tag{17}$$

Combining Eq. 14 and 17

$$E[H^2] = \frac{1}{N_0} \sum_{k=1}^{n} \left(\frac{1}{8} + o(N_k^{-1}) \right) \cong \frac{n}{8MN_0} = \frac{1}{8}\left(\left(\frac{MN_0}{n}\right)^{1/2}\right)^{-2} \tag{18}$$

The expected value of a discrete approximation to the H^2 in Eq. 16 and 18 shows the relationship of H^2 with $(MN_0/n)^{1/2}$. This suggests a plot of H^2 against $(MN_0/n)^{1/2}$.

Simulation methods

A flowchart of MC solution of coalescence PBEs is shown in in Fig.1. The particle population is represented in an array with N_p rows to represent each individual particle and N_i columns for each internal coordinate. The shorthand $X_{j,\,k}$ is used to refer to the kth internal coordinate of jth the particle (row j column k in the array). To represent the initial particle population, each cell in the array is initialized using the generation procedure, transformation method[9] in one dimensional PBEs and conditional distribution method[9] in two dimensional PBEs.

With the array (Fig.1) initialized, the coalescence rate can be estimated to control the property evolution of the particle population at each time step. The time interval is calculated from the coalescence rate, so that there is a coalescence event per time interval. The time interval is

$$\Delta t = \frac{1}{r_{coal}} \tag{19}$$

The rate for size-dependent coalescence is calculated from a coalescence table. Each cell of coalescence table represents the value of a coalescence kernel e.g. for $\beta(v_l, v_m) = v_l + v_m$, the average coalescence kernel is[4]

$$\bar{\beta}(v_l, v_m) = \frac{2}{N_P(N_P - 1)} \sum_{m=l+1}^{N_P} \sum_{l=1}^{N_P-1} \beta(v_l, v_m) \tag{20}$$

The rate for size-dependent coalescence is

$$r_{coal} = \frac{1}{2} \bar{\beta}(v_l, v_m) V C_n^2 \tag{21}$$

$$C_n = \frac{N_P}{V} \tag{22}$$

The rate for size-independent coalescence is

$$r_{coal} = \frac{1}{2} \beta_0 V C_n^2 \tag{23}$$

Where β_0 is the coalescence rate constant, V is the sample volume in the MC simulation, C_n is total particle number per unit volume in the physical system.

In this algorithm, assumed array size (AAS) and dynamic allocation of array (DAA) are used to store and update the properties of particle population over time. The AAS approach declares an array with a fixed size. The DAA approach dynamically allocates an array of the right size or reallocates an array when it needs to expand.

AAS

In the CNMC algorithm, the array is updated in the case of a coalescence event in three steps (Fig 1):

(1) Replace the property information of particle m_{j1} (row *j1* of the array) with $m_{j1} + m_{j2}$.

(2) Randomly select particle m_{j3} (row *j3* of the array). (*j3* ≠ *j1* or *j2*)

(3) Replace the property information of particle m_{j2} (row *j2* of the array) with m_{j3}

In the SCVMC algorithm, the array is updated for a coalescence event in three steps:

(1) Replace the property information of particle m_{j1} (row *j1* of the array) with $m_{j1} + m_{j2}$.

(2) Replace the property information of particle m_{j2} row *j2* of the array) with m_{jN}

(3) Set property information of particle $m_{jN} = 0$ (row *jN* of the array)

DAA

In the SCVMC algorithm, the array is updated for a coalescence event in two steps (Fig 1):

(1) Replace the property information of particle m_{j1} (row $j1$ of the array) with $m_{j1}+ m_{j2}$.

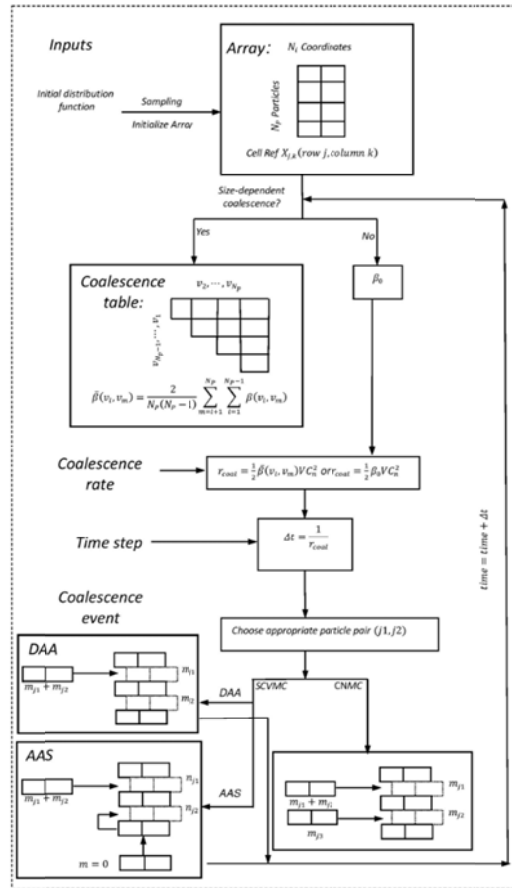(2) Remove the property information of particle m_{j2} (row $j2$ of the array)

Fig 1. Flowchart of MC solution of coalescence PBEs

Results and discussion

Case1: 1D size-dependent

The accuracy of MC simulation is closely related to N_0 and M. The impact of N_0 and M on accuracy of the MC results for 1D PBEs with size-dependent coalescence were examined in Fig. 2. The predictive distributions of the CNMC approach are compared to the analytical solution (Eq.5) at a case with I_{agg}=0.99.[33] In Fig. 2a, some data points are scattered randomly and deviated from the theoretical curve at small N_0 (<500). H^2 declines from 0.21 to 0.004 as N_0 increased from 50 to 5,000. As M increases from 1 to 20 in Fig. 2b, both the decreases of degree of scatter and H^2, showing a significant increase in the accuracy. Fig.2c shows the consistent accuracy of MC simulation when $M_x N_0$ is kept constant.
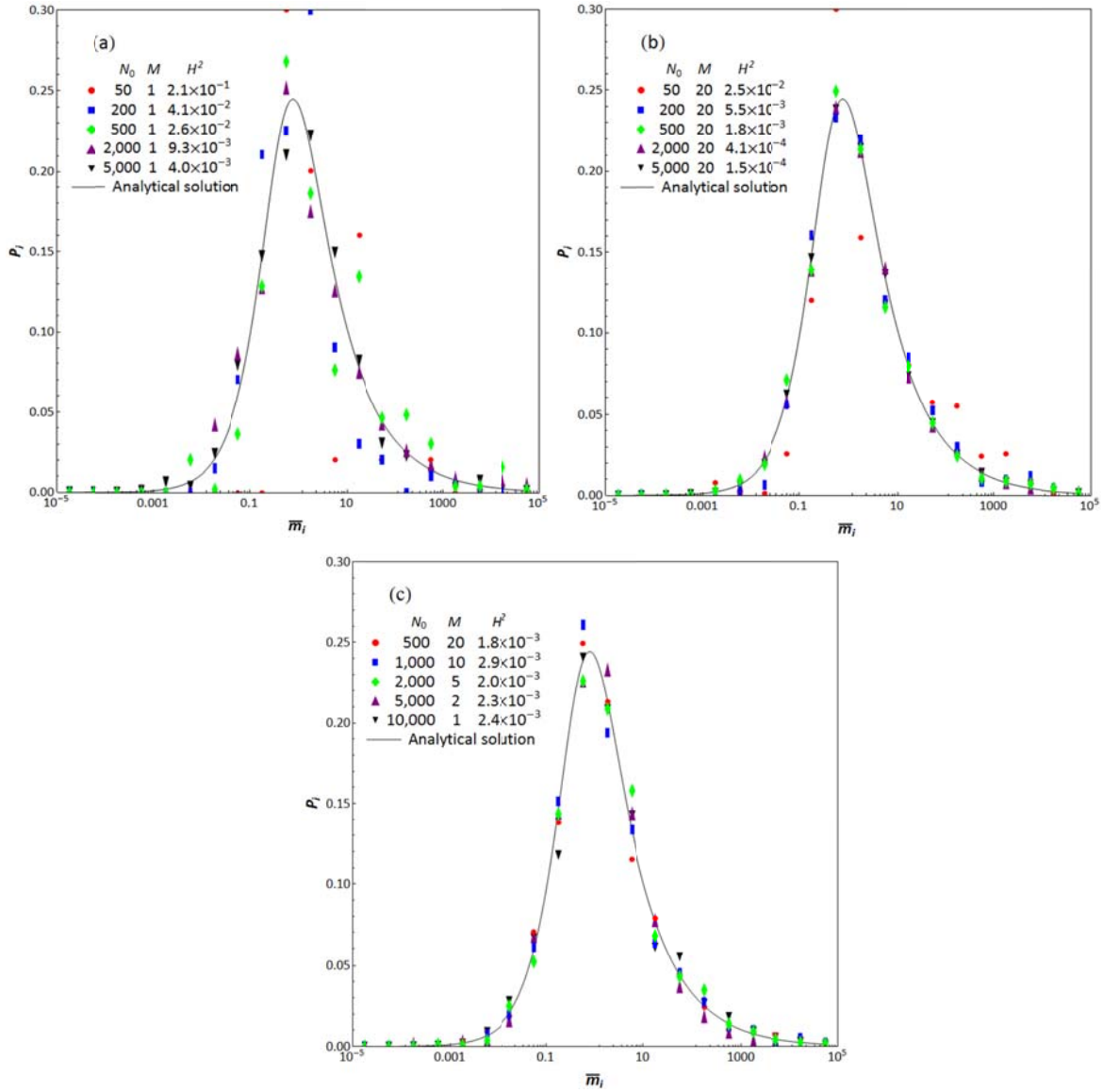
Fig 2. Comparison between particle mass distribution (at constant size bins n=21) obtained by CNMC ($P_i = P(m_i \leq m < m_{i+1}), \overline{m}_i = (m_{i+1} + m_i)/2$) and analytical solution ($P_i(\overline{m}_i) = \int_{m_i}^{m_{i+1}} n(m)dm$) for 1D case (size-dependent coalescence) (a) at constant M=1 (b) at constant M=20 (c) at constant MN_0=10,000
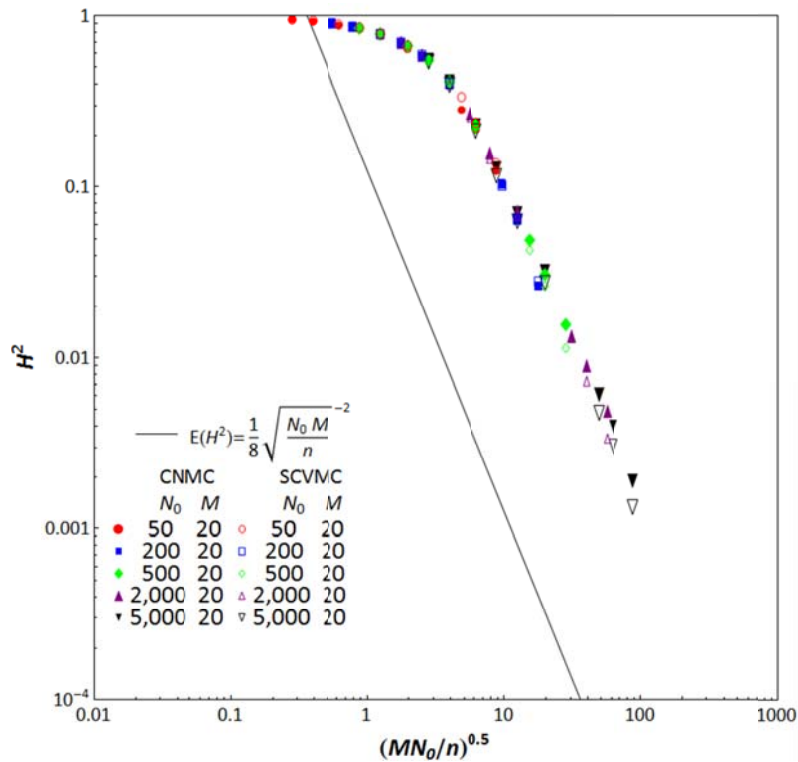
Fig 3.Calculated (CNMC and SCVMC at size bins n=13-13,000) and asymptotic approximation H^2 (when $(MN_0/n)^{1/2}$ is large) dependence of $(MN_0/n)^{1/2}$ for 1D case (size-dependent coalescence) at constant M=20
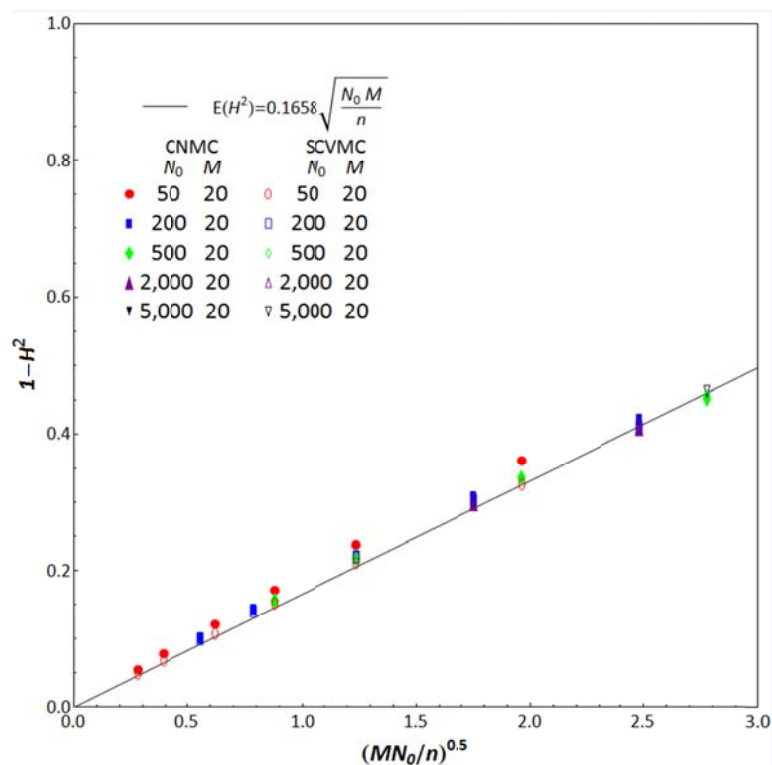

Fig 4. Calculated (CNMC and SCVMC at size bins n=13-13,000) and asymptotic approximation $1-H^2$ (when $(MN_0/n)^{1/2}$ is small) dependence of $(MN_0/n)^{1/2}$ for 1D case (size-dependent coalescence) at constant M=20

The theoretical correlation between H^2 and $(MN_0/n)^{1/2}$ is shown in Fig 3 and 4 for the

1D size-dependent coalescence PBE case using the CNMC and SCVMC approaches. According to Eq. 18, asymptotic approximation of H^2 is a power law of $(MN_0/n)^{1/2}$ with a slope -2 when $(MN_0/n)^{1/2}$ is large. Furthermore, the critical value of H^2 should be 1/8, at $(MN_0/n)^{1/2}$ =1. In Fig 3, the simulated H^2 (CNMC and SCVMC) is over-predicted compared to the asymptotic approximation curve of H^2. The reason for this can be explained by the limitation of our assumption in the derivation process of the asymptotic approximation of H^2. It is assumed that $\widehat{N}_k$'s are uncorrelated among themselves over the domain and are everywhere large. This latter assumption cannot be valid when N_0 is small. However, it is noted that the observed values of H^2 do scale as expected when $(MN_0/n)^{1/2}$ >3 in Fig. 3. According to Eq. 16, the expected value of 1-H^2 is a linear function of $(MN_0/n)^{1/2}$ when $(MN_0/n)^{1/2}$ is small. The prediction trends from both the CNMC and SCVMC approaches are consistent with the theoretical curve when $(MN_0/n)^{1/2}$ <3 in Fig 4. However in this regime, error, or H^2, is always large and so should be avoided. It is noted that n determines the resolution of MC prediction, and the accuracy of MC simulation decreases as n increases. Essentially, small n is always to be avoided (e.g. n=1). It is worth noting that the choice of algorithm has very little impact on accuracy.

Case2: 2D size-independent

The theoretical NDF 3D plots on the m_1 m_2 plane for the analytical solution (Eq. 7 and 8) of the 2D size-independent coalescence PBE at τ=0 and τ=100 are shown in Fig 5. Particle number density is significantly decreased at τ=100 due to coalesecence.
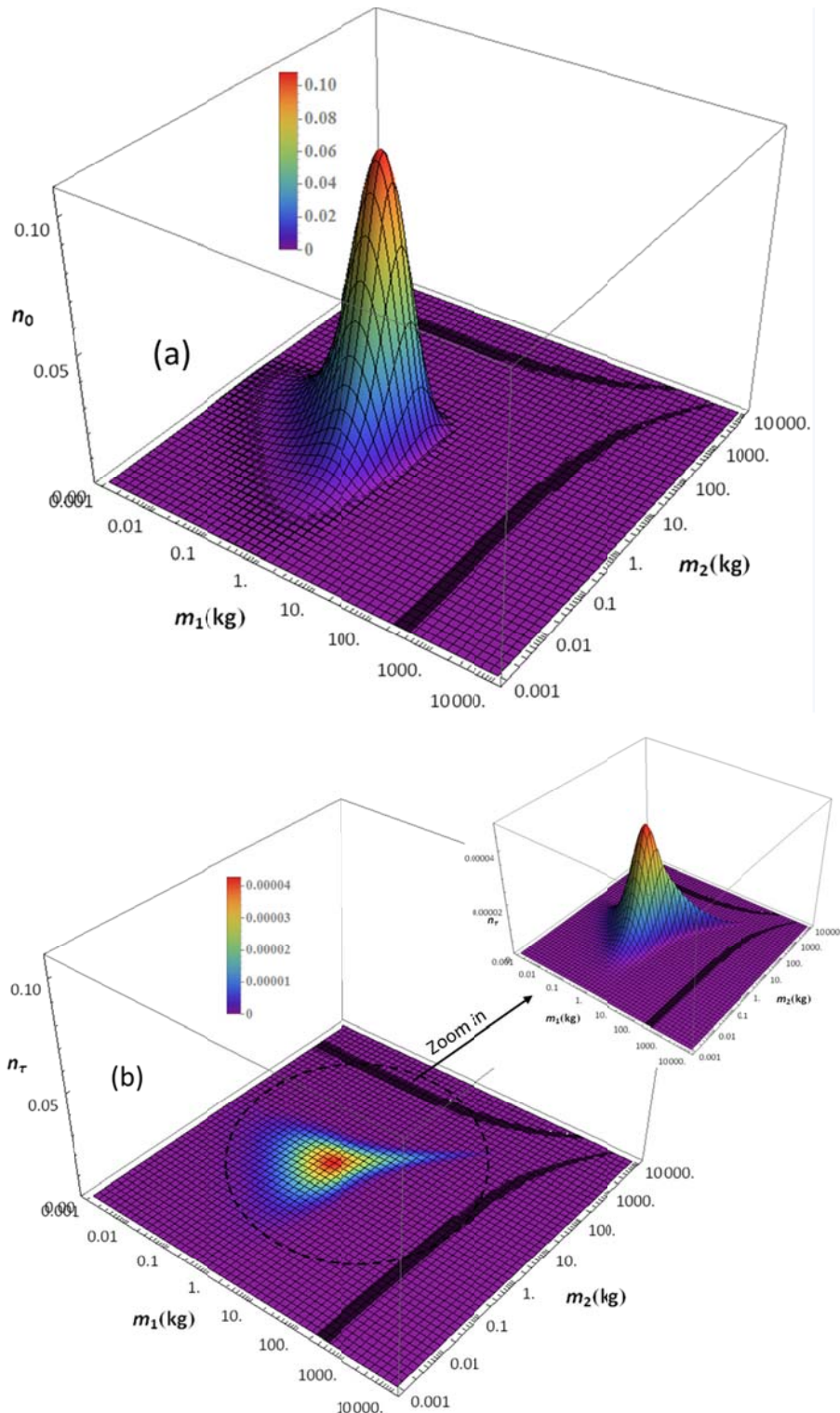
Fig 5. 3D plot of particle mass (m_1, m_2) distribution of analytical solution for 2D case (size-independent coalescence) (a) at $\tau=0$ ($n_{0,max}=0.108$) (b) at $\tau=100$ ($n_{\tau,max}=4.29\times10^{-5}$)

The dependence of H^2 on $(MN_0/n)^{1/2}$ is shown in Fig 6 for this case using the CNMC and SCVMC approaches. Also shown is the asymptotic prediction which is in reasonable agreement with the simulation results. There is some evidence that the CNMC algorithm outperforms SCVMC, presumably because in this case there is

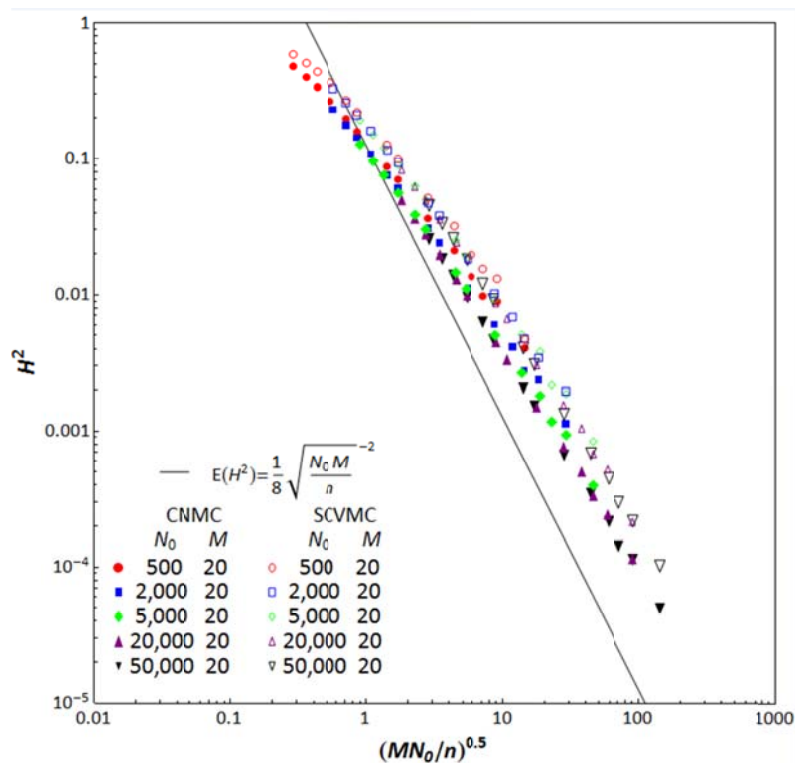considerable reduction in simulated particle numbers.

Fig 6.Calculated (CNMC and SCVMC at size bins $n=7^2$-350^2) and asymptotic approximation H^2 (when $(MN_0/n)^{1/2}$ is large) dependence of $(MN_0/n)^{1/2}$ for 2D case (size-independent coalescence) at constant $M=20$
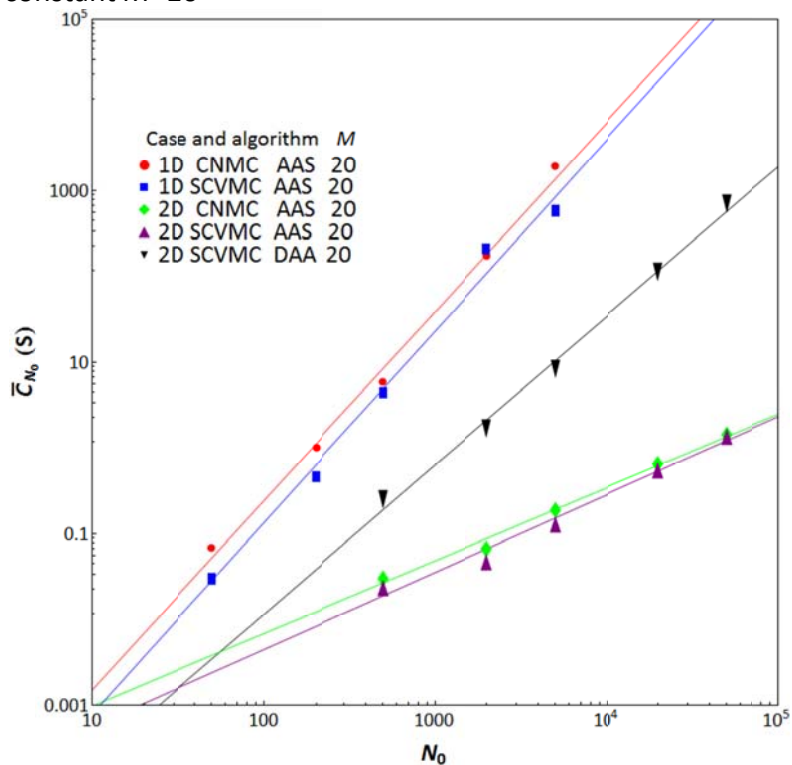

Fig 7. Averaged (constant $M=20$) CPU cost dependence of N_0 for 1D case (size-dependent coalescence) using methods (CNMC, SCVMC) and 2D case (size-independent coalescence) using methods (CNMC, SCVMC (AAS and DAA)), the CPU cost index (b) of each MC algorithm

is in Table 3

CPU cost

In this section, the correlation of computational time (CPU cost) with N_0 and M is examined for the assessment of computational efficiency. The computational time C of a MC simulation is seen to follow a power law relationship in initial sample number N_0 and is a linear function of replication M

$$C = aN_0^b M \qquad (24)$$

Where *a* and *b* are unknown parameters; we term *b* the CPU cost index. A series of MC simulations at different N_0 and constant $M=20$ were implemented and CPU costs recorded. The interrelationship between CPU cost and N_0 is shown in Fig 7. A linear regression approach is used to estimate *a* and CPU cost index *b*, which are shown in Table 3. Both parameters are remarkably sensitive to the computational complexity of the algorithm. The difference in the parameters between 1D and 2D case is due to the coalescence algorithm. The algorithm of size-dependent coalescence used nested DO-Loops to build the coalescence table for calculating coalescence rate. This sub-process needs to recall storage memory in the order of N_0^2. In DAA, $N-1$ storage locations of the previous array with size $N-1$ need to be recalled and replicated into a new array with size $N-1$. This implementation leads to an increment of implement steps of $(N-1)/N_0$

Table 3 : Parameters in CPU cost correlation

Case No	MC algorithm	a	b
1D Size-dependent	CNMC (AAS)	9.312×10^{-6}	2.212
	SCVMC (AAS)	4.368×10^{-6}	2.243
	SCVMC (AAS)	5.51×10^{-5}	0.928
2D Size-independent	SCVMC (DAA)	5.63×10^{-7}	1.794
	CNMC (AAS)	1.478×10^{-4}	0.845

Since H^2 has a correlation with a square root of average number per size bin $(MN_0/n)^{1/2}$, MN_0/n can be used to represent the accuracy (Q_c) of the MC results instead of H^2. It gives

$$Q_c = \frac{MN_0}{n} \qquad (25)$$

If *M* is replaced by Q_c, the CPU cost is obtained as

$$C(M, N_0) = aQ_c nN_0^{b-1} \qquad (26)$$

$C(M,N_0)$ is a monotonic increasing function at $b>1$, so the minimum CPU cost of MC simulation is achieved at $N_0=1$, $M=Q_c n$. Alternatively, $C(M,N_0)$ is a monotonic decreasing function at $0<b<1$, so the minimum CPU cost of MC simulation is achieved at $N_0=Q_c n$, $M=1$. In other words, if the CPU cost index is greater than one, a cost optimal, quality controlled simulation strategy is for a large number of replicates (M large) with small numbers of initial particles (N_0 small). If the cost index is greater than one, the optimal strategy is for a single replicate ($M=1$) and a large number of initial particles (N_0 large)

The example of computational time saving at the $b>1$ condition can be seen in Table 4. The comparison is based on the one dimensional size-dependent coalescence case solved by the CNMC approach, $b=2.212$ (Table 3). Under the same accuracy criterion ($n=433$ and $MN_0=10,000$), the computational time of 173.86s for multiple MC simulation replicates ($N_0=500$, $M=20$) is far less than the computational time of 6319.76s for a single MC simulation ($N_0=10,000$, $M=1$). The example of computational time saving at the $0<b<1$ condition can be seen in Table 4. The comparison is based on one dimensional size-dependent coalescence case solved by the CNMC approach $b=0.845$ (Table 3). Under the same accuracy criterion ($n=70^2$, $MN_0=100,000$), the computational time of 2.48s for a single MC simulation ($N_0=100,000$, $M=1$) is able to save 37.2% CPU cost than that (3.95s) of multiple MC simulation replicates ($N_0=5,000$, $M=20$).

Table 4. CPU cost at different CPU index

Case No	MC algorithm	b	N_0	M	C(s)
1D Size-dependent	CNMC (AAS)	2.212	500	20	173.86
			10,000	1	6309.76
2D Size-independent	CNMC(AAS)	0.845	5,000	20	3.95
			100,000	1	2.48

Conclusions

Accuracy and optimal sampling strategy in Monte Carlo simulation of Population Balance Equations have been investigated in this study. It is concluded that Squared

Hellinger Distance, H^2, is a powerful tool to measure the accuracy of MC simulation, and is related to initial sample number (N_0), number of replicates (M) and Number of bin sizes, (n). The asymptotic approximation of H^2 is derived as $(1/8)(MN_0/n)^{-1/2}$ when $(MN_0/n)^{1/2}$ is large. Although the actual value of H^2 is higher compared to the theoretical trend in the 1D PBE cases, simulate results for both 1D and 2D PBEs with coalescence approximately demonstrated that scaling. A power law relationship, $C = aMN_0^b$ is found to describe the correlation between CPU cost and N_0, and M. The CPU cost index, b, illustrates the weight of N_0 in CPU cost.

Finally, an optimal sampling strategy is given as

1) n determines the resolution of MC prediction and must be chosen by the user trading off the increased resolution available from increased n, with decreased accuracy ie increased H^2.

2) $M \times N_0$ determines the accuracy of MC prediction, and both the accuracy of MC simulation and the CPU cost increase as MN_0 increases. If the CPU cost index $b>1$, the minimum CPU cost is achieved for small numbers of N_0, and large values of M. Alternatively, if the CPU index, $0<b<1$, the minimum CPU cost is achieved at $M=1$ and a large value of N_0.

In this study, an optimal sampling strategy is developed for MC solution of PBEs with coalescence only. However, the approach can be extended to PBEs in any form solved by an MC approach. Substantial savings in computational cost are possible, if an optimal strategy is adopted.

References

1. Hounslow MJ, Ryall RL, Marshall VR, A discretized population balance for nucleation, growth, and aggregation, *AIChE J.* 1988;34: 1821-1832

2. Kumar S, Ramkrishna D, On the solution of population balance equations by discretization—I. A fixed pivot technique, *Chem Eng Sci.* 1996; 51:1311-1332

3. Celnik M, Patterson R, Kraft M, Wagner W, Coupling a stochastic soot population balance to gas-phase chemistry using operator splitting, *Combust.Flame.* 2007; 148:158–176

4. Smith M, Matsoukas T, Constant-number Monte Carlo simulation of population balances, *Chem Eng Sci.* 1998;53:1777-1786

5. Zhao H, Maisels A, Matsoukas T, Zheng C, Analysis of four Monte Carlo methods for the solution of population balances in dispersed systems, *Powder Tech.* 2007;173: 38-50

6. Liffman K, A direct simulation Monte-Carlo method for cluster coagulation, *J. Comput. Phys.* 1992;100:116-127

7. Maisels A, Kruis FE, Fissan H, Direct simulation Monte Carlo for simultaneous nucleation, coagulation, and surface growth in dispersed systems, *Chem Eng Sci.* 2004;59: 2231-2239

8. Lin Y, Lee K, Matsoukas T, Solution of the population balance equation using constant-number Monte Carlo, *Chem Eng Sci.* 2002; 57:2241-2252

9. Press WH, Teukolsky SA, Vetterling WT, Flannery BP, Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing, 2nd Edition, 1996, ISBN 0-521-57439-0.

10. Gillespie DT, The stochastic coalescence model for cloud droplet growth*, J. Atmos. Sci*. 1972;29:1496-1510

11. Gillespie DT, An exact method for numerically simulating the stochastic coalescence process in a cloud, *J. Atmos. Sci.* 1975;32:1977-1989

12. Daniel TG , A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J.Comput. Phys.* 1976; 22:403-434

13. Kruis FE, Maisels A, Fissan H, Direct simulation Monte Carlo method for particle coagulation and aggregation, *AIChE J.* 2000;46:1735-1742

14. Gooch JRP and Hounslow MJ, Monte Carlo simulation of size-enlargement mechanisms in crystallization, *AIChE J.* 1996;42:1864-1874

15. Lattuada, M and Wu H, A simple model for the structure of fractal aggregates, *J. Colloid Interface Sci.* 2003;268:106-120.

16. Rosner DE, Yu S, MC Simulation of Aerosol Aggregation and Simultaneous Spheroidization, *AIChE J.* 2001; 47:545-561

17. Efendiev Y and Zachariah MR, Hybrid Monte Carlo Method for Simulation of Two-Component Aerosol Coagulation and Phase Segregation, *J. Colloid Interface Sci.* 2002;249:30-43

18. Tandon P, Rosner DE,   Monte Carlo Simulation of Particle Aggregation and

Simultaneous Restructuring, *J. Colloid and Interface Sci.*1999*: 213:273-286

19. Zhao H and Zheng C, Monte Carlo solution of wet scavenging of aerosols by precipitation, *Atmos. Environ.* 2006; 40:1510-1525

20. Braumann A, Goodson MJ, Kraft M, Mort PR, Modelling and validation of granulation with heterogeneous binder disperssion and chemical reaction, *Chem Eng Sci.* 2007;62:4717-4728

21. Braumann A, Kraft M, Mort PR, Parameter estimation in a multidimensional granulation model, *Powder Tech.*2010;197:196-210

22. Braumann A, Kraft M, Wagner W, Numerical study of a stochastic particle algorithm solving a multidimensional population balance model for high shear granulation, *J. Comput. Phys.* 2010;229:7672-7691

23. Braumann A, Man PLW, Kraft M, Statistical approximation of the inverse problem in multivariate population balance modeling, *Ind. Eng.Chem. Res.* 2010;49:428-438

24. Braumann A, Man PLW, Kraft M, The inverse problem in granulation modeling—two different statistical appraoches, *AICHE J.2011;57:3105-3121*

25. Rosenboom JG, Antonyuk S, Heinrich S,Kraft M, Characterisation of lactose powder and granules for multivariate wet granulation modelling, *Chem. Eng. Sci.* 2015;123:395-405

26. Wolfowitz J, The minimum distance method, *Ann. Math. Stat.* 1957; 28:75-88

27. Yang S, Minimum hellinger distance estimation of parameter in the random censorship model, *Ann. Stat.* 1991;19:579-602

28. Basu A, Harris IR, Basu S, Tests of hypotheses in discrete models based on the penalized Hellinger distance, *Stat. Probabil. Lett.* 1996; 27:367-373.

29. Karunamuni RJ, Wu J, Title Minimum Hellinger distance estimation in a nonparametric mixture model, *J.Stat. Plan. Inference.* 2009;139: 1118-1133

30. Gelbard F, Seinfeld JH, Numerical solution of the dynamic equation for particulate systems, *J. Comput. Phys.* 1978;28:357.

31. Vale HM, McKenna TF, Solution of the population balance equation for two-component aggregation by an extended fixed pivot technique*, Ind. Eng.Chem. Res.* 2005;44:7885-7891

32. Kendall S, Stuart MA, Ord JK, The advanced theory of statistics. High Wycombe: Charles Griffin. 1983

33. Hounslow MJ, A discretized population balance for continuous systems at steady state, *AIChE J.* 1990;36:106-116

List of Figures

Fig 1. Flowchart of MC solution of coalescence PBEs

Fig 2. Comparison between particle mass distribution (at constant size bins n=21) obtained by CNMC ($P_i = P(m_i \leq m < m_{i+1}), \bar{m}_i = (m_{i+1} + m_i)/2$) and analytical solution ($P_i(\bar{m}_i) = \int_{m_i}^{m_{i+1}} n(m)dm$) for 1D case (size-dependent coalescence) (a) at constant M=1 (b) at constant M=20 (c) at constant MN_0=10,000

Fig 3.Calculated (CNMC and SCVMC at size bins n=13-13,000) and asymptotic approximation H^2 (when $(MN_0/n)^{1/2}$ is large) dependence of $(MN_0/n)^{1/2}$ for 1D case (size-dependent coalescence) at constant M=20

Fig 4. Calculated (CNMC and SCVMC at size bins n=13-13,000) and asymptotic approximation 1-H^2 (when $(MN_0/n)^{1/2}$ is small) dependence of $(MN_0/n)^{1/2}$ for 1D case (size-dependent coalescence) at constant M=20

Fig 5. 3D plot of particle mass (m_1, m_2) distribution of analytical solution for 2D case (size-independent coalescence) (a) at τ=0 ($n_{0,max}$=0.108) (b) at τ=100 ($n_{\tau,max}$=4.29×10^{-5})

Fig 6.Calculated (CNMC and SCVMC at size bins n=7^2-350^2) and asymptotic approximation H^2 (when $(MN_0/n)^{1/2}$ is large) dependence of $(MN_0/n)^{1/2}$ for 2D case (size-independent coalescence) at constant M=20

Fig 7. Averaged (constant M=20) CPU cost dependence of N_0 for 1D case (size-dependent coalescence) using methods (CNMC, SCVMC) and 2D case (size-independent coalescence) using methods (CNMC, SCVMC (AAS and DAA)), the CPU cost index (b) of each MC algorithm is in Table 3

List of Tables

Table 1: Summary of application of Monte Carlo method to PBEs in the literatures

Table 2: Parameters used in MC simulation

Table 3 : Parameters in CPU cost correlation

Table 4. CPU cost at different CPU index