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

Xi Yu
European Bioenergy Research Inst. (EBRI), School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, U.K.

Dept. of Chemical and Biological Engineering, The University of Sheffield, Sheffield, S1 3JD, U.K.

Michael J Hounslow
Dept. of Chemical and Biological Engineering, The University of Sheffield, Sheffield, S1 3JD, U.K.

Gavin K Reynolds
Pharmaceutical Development, AstraZeneca, Macclesfield, SK10 2NA, U.K.

DOI 10.1002/aic.14837
Published online April 22, 2015 in Wiley Online Library (wileyonlinelibrary.com)

Implementation of a Monte Carlo simulation for the solution of population balance equations (PBEs) 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 Monte Carlo (MC) simulation, and can be related directly to N_0, M, and n. Asymptotic approximations of H^2 are deduced and tested for both one-dimensional (1-D) and 2-D PBEs with coalescence. The central processing unit (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. © 2015 American Institute of Chemical Engineers *AIChE J*, 61: 2394–2402, 2015
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 (e.g., 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 article, the number of particles will decline over time. Therefore, if the size of the simulation box (i.e., 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

Correspondence concerning this article should be addressed to X. Yu at x.yu3@aston.ac.uk.

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[9,10]	–	–	–
Coupled chemical reaction[11]	–	–	–
chemical reaction and coating[12]	3×10^4 (4) 2×10^4 (5)	–	Qualitative comment on accuracy via $1/\sqrt{N}$
Crystallization[13]	6400–25,600 (1-250)	3–620 min in PCs	Good statistical representative is achieved in two ways
Fractal aggregation[14]	3–100 (1000)	–	–
Higher dimensionality problems[15]	–	–	–
Multicomponent aerosols[16]	3×10^6(10)	A few hours(256 Mb RAM and Pentium III 900-MHz pc)	Qualitative comment on accuracy via $1/\sqrt{N}$
Particle aggregation and restructuring[15,17]	10^4(15)	2 h (SPRAC—10 workstations)	Average/ensemble method
	10^4(100)	–	Average procedure
Wet scavenging[18]	3000	25–78 s in PCs	–
Wet granulation[19–24]	1024–4096(64)	–	–

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 personal computer (PC)s 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[12,16] 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 among 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.[13] 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

1. use the Hellinger Distance, a statistical distance between two probability distributions, to measure the accuracy of MCSs.
2. provide guidance on how to select the initial sample number and number of replicates and consider trade-offs between accuracy and computational costs.

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[25–28] 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

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

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

where N_i is number in size bin C_i in the MCS, subscript AS refers to the analytical solution.

Coalescence PBEs

In this study, two cases, one-dimensional (1-D) size-dependent and 2-D 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.

Case 1: 1-D Size-Dependent. Gelbard and Seinfeld[29] 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.

Table 2. Parameters Used in MCS

1-D Size Dependent		2-D 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
I_{agg}	0.99	I_{agg}	0.99
N_0	50–5000	N_0	500–50,000
M	1–20	M	1–20
n	13–13,000	n	7^2-350^2

Case 2: 2-D Size-Independent. A 2-D analytical solution for the size-independent coalescence PBE is revealed by Vale and McKenna,[30] in view of solution proposed by Gelbard and Seinfeld.[29] The initial distribution of particles which have two components in mass mode is

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

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

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

$$\text{where } \theta=4\left(\frac{m_1m_2}{m_{10}m_{20}}\right)^{1/2}\left(\frac{\tau}{\tau+2}\right)^{1/2} \tag{9}$$

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

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 MCS.

Asymptotic approximation of $\mathbf{H^2}$

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, $\hat{N}_k$, in each size range C_k in MC results has a Poisson distribution; (2) A group of $\hat{N}_k$'s are uncorrelated among themselves over the domain.

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

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

where N_k is the number associated with kth internal and $\hat{N}_k$ is an estimate of that number and

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

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

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

Then the expected value of H^2 is

$$\begin{aligned}E[H^2]&=\frac{1}{2MN_0}E\left[\sum_{k=1}^{n}\left(\sqrt{N_k}-\sqrt{\hat{N}_k}\right)^2\right]\\&=\frac{1}{2MN_0}\sum_{k=1}^{n}\left(N_k-2\sqrt{N_k}E\left[\sqrt{\hat{N}_k}\right]+E\left[\hat{N}_k\right]\right)\\&=\frac{1}{N_0}\sum_{k=1}^{n}\left(N_k-\sqrt{N_k}E\left[\sqrt{\hat{N}_k}\right]\right)\end{aligned} \tag{14}$$

However, if N_k is everywhere small $E\left[\sqrt{\hat{N}_k}\right]=N_k$ so

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

Now, the N_k scale with MN_0/n so put $MN_0N_k^1/n$ where the N_k^1 are constants. Therefore

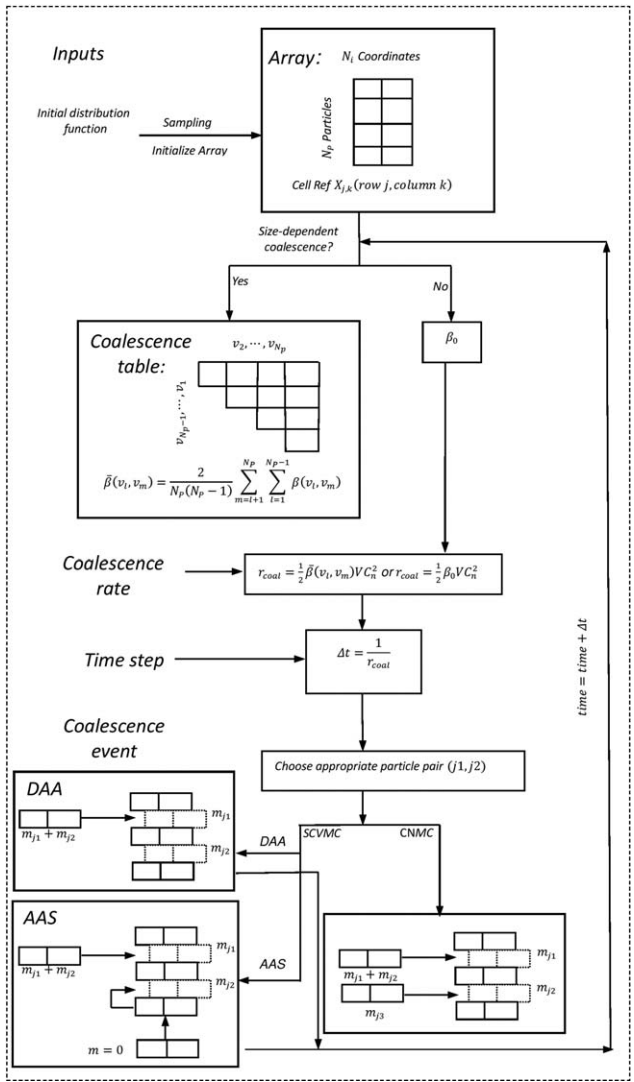

Figure 1. Flowchart of MC solution of coalescence PBEs.

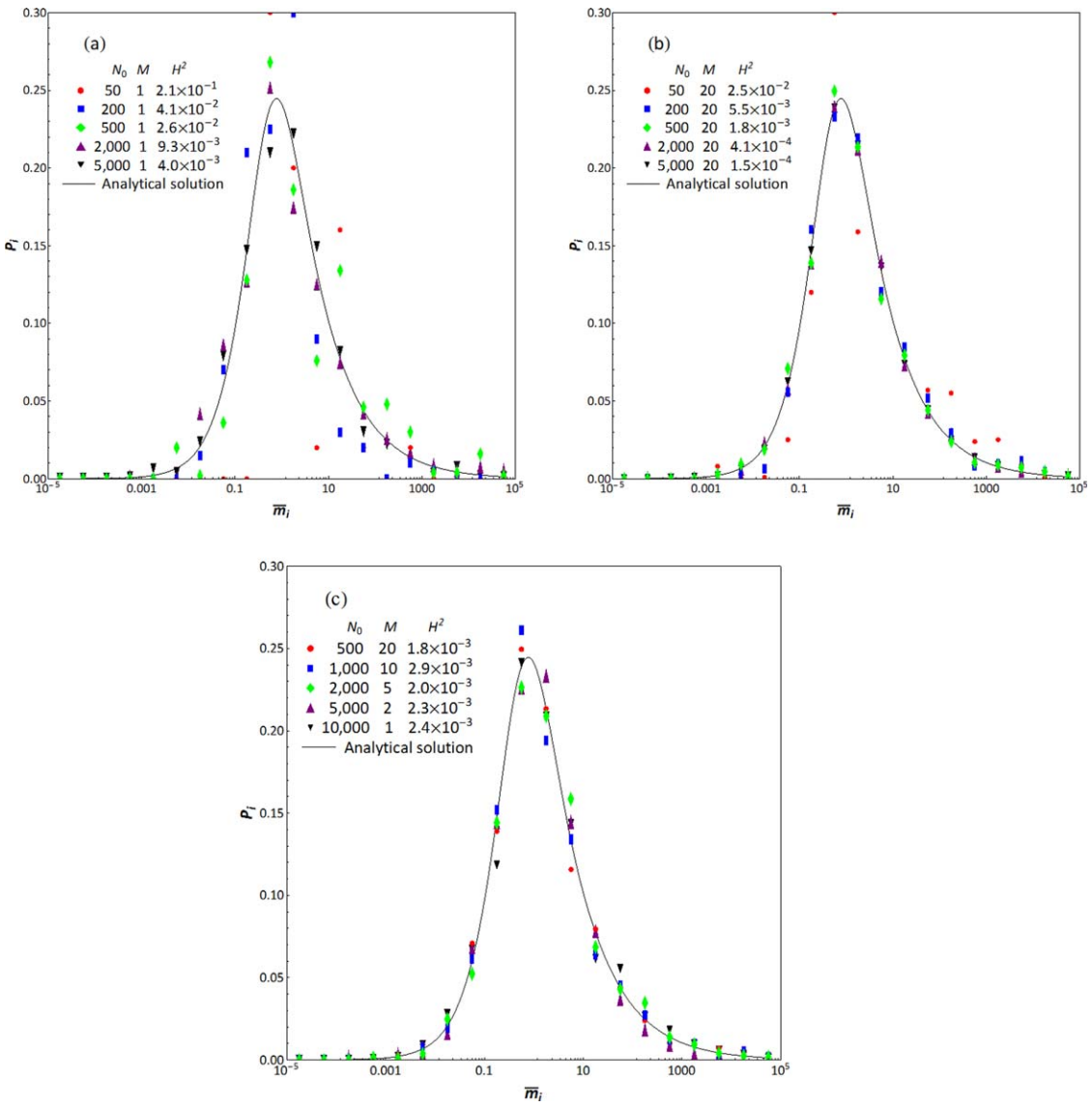

Figure 2. Comparison between particle mass distribution (at constant size bins $n = 21$) obtained by CNMC ($P_i = P(m_i \le m < m_{i+1}), \bar{m}_i = (m_{i+1}+m_i)/2$) and analytical solution $\left(P_i(\bar{m}_i) = \int_{m_i}^{m_{i+1}} n(m)dm\right)$ for 1-D case (size-dependent coalescence).

(a) at constant $M = 1$, (b) at constant $M = 20$, and (c) at constant $MN_0 = 10{,}000$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

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

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

If N_k is everywhere large, the expected value of $\sqrt{\hat{N}_k}$ is given by Kendall et al.[31]

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

Combining Eqs. 14 and 17

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

The expected value of a discrete approximation to the H^2 in Eqs. 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 Figure 1. The particle population is represented in an array with N_p rows to represent each individual particle and Ni columns for each internal coordinate. The shorthand Xj,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[32] in 1-D PBEs and conditional distribution method[32] in 2-D PBEs.

With the array (Figure 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

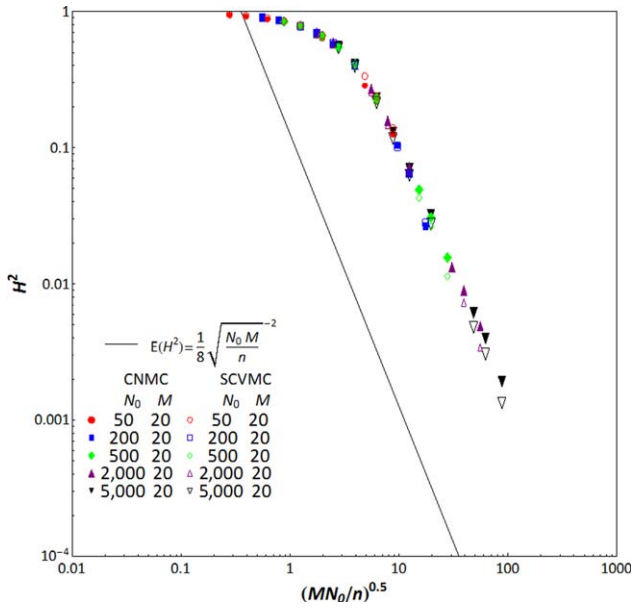

Figure 3. Calculated (CNMC and SCVMC at size bins $n = 13$–$13{,}000$) and asymptotic approximation H^2 (when $(MN_0/n)^{1/2}$ is large) dependence of $(MN_0/n)^{1/2}$ for 1-D case (size-dependent coalescence) at constant $M = 20$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

calculated from the coalescence rate, so that there is a coalescence event per time interval. The time interval is

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

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

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

The rate for size-dependent coalescence is

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

$$C_n = \frac{N_{\text{P}}}{V} \tag{22}$$

The rate for size-independent coalescence is

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

where β_0 is the coalescence rate constant, V is the sample volume in the MCS, C_n is the 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 (Figure 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). ($j3 \neq j1$ 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 (Figure 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)

Results and Discussion

Case 1: 1-D size-dependent

The accuracy of MCS is closely related to N_0 and M. The impact of N_0 and M on accuracy of the MC results for 1-D PBEs with size-dependent coalescence were examined in Figure 2. The predictive distributions of the CNMC approach are compared to the analytical solution (Eq. 5) at a case with $I_{\text{agg}} = 0.99$.[33] In Figure 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 5000. As M increases from 1 to 20 in Figure 2b, both

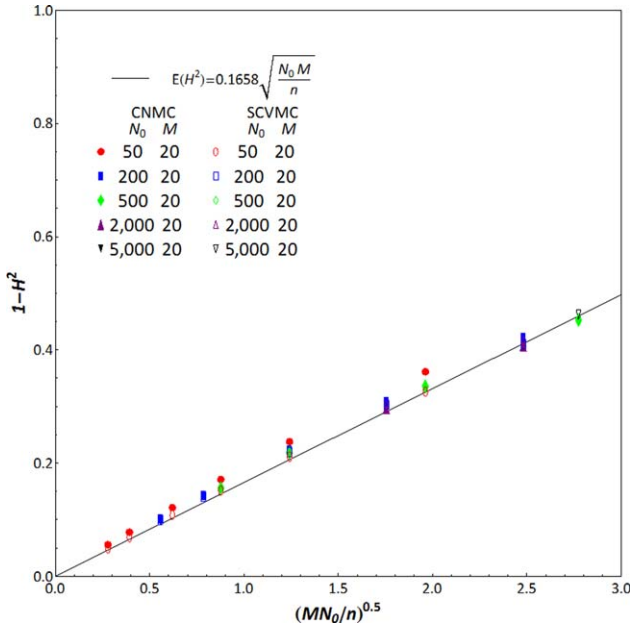

Figure 4. Calculated (CNMC and SCVMC at size bins $n = 13$–$13{,}000$) and asymptotic approximation $1 - H^2$ (when $(MN_0/n)^{1/2}$ is small) dependence of $(MN_0/n)^{1/2}$ for 1-D case (size-dependent coalescence) at constant $M = 20$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

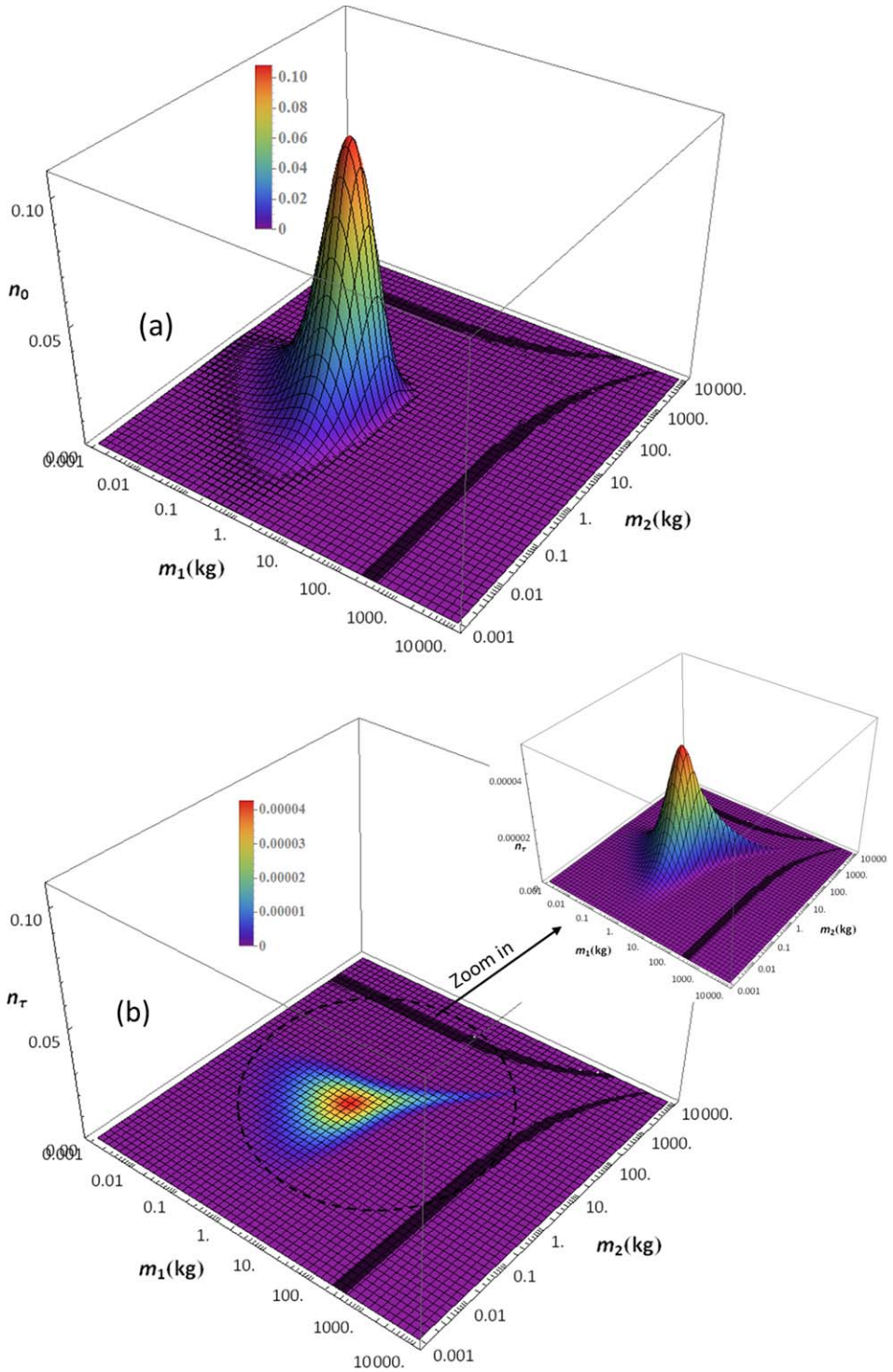

Figure 5. 3-D plot of particle mass (m_1, m_2) distribution of analytical solution for 2-D case (size-independent coalescence).

(a) at $\tau = 0$ ($n_{0,\max} = 0.108$) and (b) at $\tau = 100$ ($n_{\tau,\max} = 4.29 \times 10^{-5}$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the decreases of degree of scatter and H^2, showing a significant increase in the accuracy. Figure 2c shows the consistent accuracy of MC simulation when $M \times N_0$ is kept constant.

The theoretical correlation between H^2 and $(MN_0/n)^{1/2}$ is shown in Figures 3 and 4 for the 1-D 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 Figure 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 $\hat{N}_k$'s are uncorrelated among themselves over the domain and are everywhere large. This latter assumption

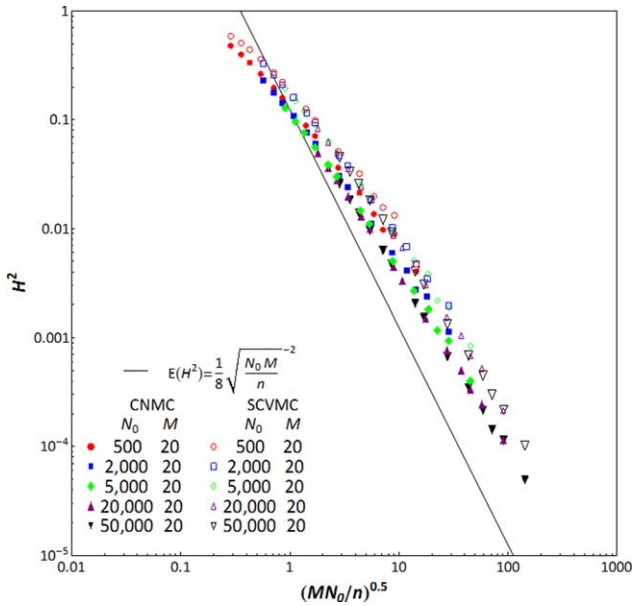

Figure 6. Calculated (CNMC and SCVMC at size bins $n = 7^2-350^2$) and asymptotic approximation H^2 (when $(MN_0/n)^{1/2}$ is large) dependence of $(MN_0/n)^{1/2}$ for 2-D case (size-independent coalescence) at constant $M = 20$.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

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 Figure 3. According to Eq. 16, the expected

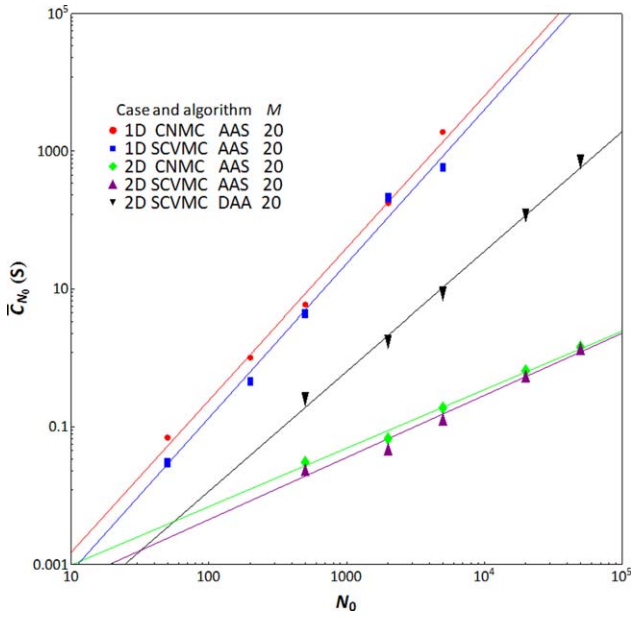

Figure 7. Averaged (constant $M = 20$) CPU cost dependence of N_0 for 1-D case (size-dependent coalescence) using methods (CNMC, SCVMC) and 2-D case (size-independent coalescence) using methods (CNMC, SCVMC [AAS and DAA]), the CPU cost index (b) of each MC algorithm is in Table 3.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 3. Parameters in CPU Cost Correlation

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

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 Figure 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 MCS 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.

Case 2: 2-D size-independent

The theoretical NDF 3-D plots on the m_1 m_2 plane for the analytical solution (Eqs. 7 and 8) of the 2-D size-independent coalescence PBE at $\tau = 0$ and $\tau = 100$ are shown in Figure 5. Particle number density is significantly decreased at $\tau = 100$ due to coalesecence.

The dependence of H^2 on $(MN_0/n)^{1/2}$ is shown in Figure 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.

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

$$C = aN_0{}^bM \tag{24}$$

where a and b are unknown parameters; we term b, the CPU cost index. A series of MCS 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 Figure 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 1-D and 2-D 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 subprocess 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 $(N-1)/N_0$.

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

Table 4. CPU Cost at Different CPU Index

Case No	MC Algorithm	b	N_0	M	C(s)
1-D Size-dependent	CNMC (AAS)	2.212	500	20	173.86
			10,000	1	6309.76
2-D Size-independent	CNMC (AAS)	0.845	5000	20	3.95
			100,000	1	2.48

represent the accuracy (Q_c) of the MC results instead of H^2. It gives

$$Q_c = \frac{MN_0}{n} \tag{25}$$

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

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

$C(M,N_0)$ 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 MCS 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 1-D 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 173.86 s for multiple MCS replicates ($N_0 = 500$, $M = 20$) is far less than the computational time of 6319.76 s for a single MCS ($N_0 = 10{,}000$, $M = 1$). The example of computational time saving at the $0 < b < 1$ condition can be seen in Table 4. The comparison is based on 1-D 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 2.48 s for a single MC simulation ($N_0 = 100{,}000$, $M = 1$) is able to save 37.2% CPU cost than that (3.95 s) of multiple MC simulation replicates ($N_0 = 5000, M = 20$).

Conclusions

Accuracy and optimal sampling strategy in MCS of PBE have been investigated in this study. It is concluded that Squared Hellinger Distance, H^2, is a powerful tool to measure the accuracy of MCS, 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 1-D PBE cases, simulate results for both 1-D and 2-D 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, that is, 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.

Acknowledgments

The authors are grateful to AstraZeneca for funding, and to Professor John Biggins, of the School of Mathematics and Statistics at the University of Sheffield for introducing them to Hellinger's distance and the series approximation to the moments of the Poisson distribution.

Literature Cited

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. Gillespie DT. The stochastic coalescence model for cloud droplet growth. *J Atmos Sci*. 1972;29:1496–1510.
10. Gillespie DT. An exact method for numerically simulating the stochastic coalescence process in a cloud. *J Atmos Sci*. 1975;32:1977–1989.
11. Daniel TG. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys*. 1976;22:403–434.
12. Kruis FE, Maisels A, Fissan H. Direct simulation Monte Carlo method for particle coagulation and aggregation. *AIChE J*. 2000;46:1735–1742.
13. Gooch JRP and Hounslow MJ. Monte Carlo simulation of size-enlargement mechanisms in crystallization. *AIChE J*. 1996;42:1864–1874.
14. Lattuada M, Wu H. A simple model for the structure of fractal aggregates. *J Colloid Interface Sci*. 2003;268:106–120.
15. Rosner DE, Yu S. MC simulation of aerosol aggregation and simultaneous spheroidization. *AIChE J*. 2001;47:545–561.
16. Efendiev Y, Zachariah MR. Hybrid Monte Carlo method for simulation of two-component aerosol coagulation and phase segregation. *J Colloid Interface Sci*. 2002;249:30–43.
17. Tandon P, Rosner DE. Monte Carlo simulation of particle aggregation and simultaneous restructuring. *J Colloid and Interface Sci*. 1999;213:273–286.
18. Zhao H and Zheng C. Monte Carlo solution of wet scavenging of aerosols by precipitation. *Atmos Environ*. 2006;40:1510–1525.

19. 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.
20. Braumann A, Kraft M, Mort PR. Parameter estimation in a multidimensional granulation model. *Powder Tech*.2010;197:196–210.
21. 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.
22. 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.
23. Braumann A, Man PLW, Kraft M. The inverse problem in granulation modeling-two different statistical approaches. *AIChE J*. 2011; 57:3105–3121.
24. 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.
25. Wolfowitz J. The minimum distance method. *Ann Math Stat*. 1957; 28:75–88.
26. Yang S. Minimum hellinger distance estimation of parameter in the random censorship model. *Ann Stat*. 1991;19:579–602.
27. Basu A, Harris IR, Basu S. Tests of hypotheses in discrete models based on the penalized Hellinger distance. *Stat Probab Lett*. 1996; 27:367–373.
28. Karunamuni RJ, Wu J. Title Minimum Hellinger distance estimation in a nonparametric mixture model. *J Stat Plan Inference*. 2009;139: 1118–1133.
29. Gelbard F, Seinfeld JH. Numerical solution of the dynamic equation for particulate systems. *J Comput Phys*. 1978;28:357.
30. 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.
31. Kendall S, Stuart MA, Ord JK. The Advanced Theory of Statistics. High Wycombe: Charles Griffin. 1983
32. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing*, 2nd Edition. Cambridge, England: Cambridge University Press, 1996, ISBN 0-521-57439-0.
33. Hounslow MJ. A discretized population balance for continuous systems at steady state. *AIChE J*. 1990;36:106–116.

Manuscript received Nov. 26, 2014, and revision received Feb. 27, 2015.