AIChE

Morphological Population Balance Model in Principal Component Space

Xue Z. Wang and Cai Y. Ma
Institute of Particle Science and Engineering, School of Process, Environmental and Materials Engineering, University of Leeds, Leeds LS2 9JT, U.K.

DOI 10.1002/aic.11860
Published online July 23, 2009 in Wiley InterScience (www.interscience.wiley.com).

Multidimensional and morphological population balance (PB) models for crystallization processes have been proposed in literature, which can be used to simulate the dynamic evolution of particle shape as well as particle size distribution. These models, however, can become computationally expensive when the crystal has a large number of independent faces, and are not applicable to noncrystalline, irregularly shaped particles such as those encountered in granulation and milling. This article addresses these challenges by introducing principal component analysis (PCA) into morphological PB modeling. PCA transforms the shape description of a particle from a high-dimensional domain to a lower dimensional, principal component (PC) space. Morphological PB models can then be built in this latent variable space, greatly reducing the computational complexity. It also makes it possible to model noncrystalline irregularly shaped particles. The original particle shape at any time can be reconstructed from the PCs. © 2009 American Institute of Chemical Engineers *AIChE J,* 55: 2370–2381, 2009

Keywords: principal component analysis, morphological population balance, multidimensional population balance, crystal shape, model reduction

Introduction

Population balance (PB) models have been widely used in engineering for modeling the dynamic evolution of particle size distribution (PSD) in crystallization, granulation, agglomeration, and milling processes, as well as in biological systems for simulating the growth of microbial and cell populations. In these models, the size of a particle is often defined as the diameter of a sphere having the same volume as the particle. This is an over-simplified treatment because it definitely misses important information about the particle shape, a very important property that affects not only the processability of particles such as flowability but also the end-use property such as bioavailability for drug particles. Motivated by this observation, several researchers have carried out research on introducing particle shape information into PB simulation and proposed multidimensional PB models for crystallization processes. The models can predict the evolution of PSD in more than one size dimension,[1–11] for example, the length and width distributions for plate-like crystals. However, the majority of efforts have been limited to using two characteristic size dimensions. Ma et al.[12,13] recently proposed a new multidimensional PB model for crystallization processes, which is called morphological PB model. It defines the shape and size of a crystal as a function of the normal distances of all its faces to the geometrical center of the crystal, and predicts the dynamic evolution of size distribution of every face for the population of crystals. The growth rate for every face is defined as a function of supersaturation and size. Morphological PB model emphasizes the capability of full reconstruction of the particle shape in high fidelity at any moment of the simulation. A morphological PB model is obviously a multidimensional PB model, but not always vice versa. Multidimensional PB models proposed in literature are so far restricted to use two characteristic size dimensions, and are not always able to

Correspondence concerning this article should be addressed to X. Z. Wang at x.z.wang@leeds.ac.uk

© 2009 American Institute of Chemical Engineers

fully reconstruct the particle shape unless for very specific situations such as plate-like crystals. In addition, the term "multidimensional PB" has also been used in literature to refer to PB simulation that uses only one size dimension (sphere diameter), with other dimensions being variables not related to particle size or shape. Because a morphological PB model is able to retain the particle shape, it provides an effective multiscale modeling tool, linking single particle simulation, e.g., single crystal morphology prediction, with a process model about the behavior of the particle population in a processing vessel.

In principle, the morphological PB model has no limit on the number of faces for the modeled crystals. In other words, it has no limit on the number of size dimensions. But with the increase in the number of size dimensions, solution of the PB equations becomes computationally too expensive, resulting in unacceptably long CPU time. In applying the morphological PB model to the case studies,[12,13] symmetrical faces were treated as having identical growth behavior, which effectively reduced the size dimensions, and so the computational time.

Multidimensional PB models have only been investigated for crystallization processes. Their application to processes such as granulation and milling that produce irregularly shaped particles has not been attempted. This does not imply that particle shape is not important for these processes. It is most probably because it is much more difficult to accurately define the shape of irregularly shaped particles than for crystals, and therefore, how the multidimensional PB models can be applied to these processes is not obvious.

The purpose of this study is to develop a new methodology for carrying out particle shape-based PB modeling that uses only a small number of size dimensions, but misses little particle shape information. The method is based on principal component analysis (PCA). It uses principal components (PCs) to describe the particle shape so that PB modeling can be carried out in a transformed, latent variable domain, or PC space. The PCs are calculated directly from the original multiple size dimensions of particles, but the number of PCs can be much smaller than that of the original size dimensions used for shape representation. At any time, the real particle shape can be reconstructed from the latent PCs using the PCA model.

Shape Description in PC Space

Principal component analysis

Given a data set consisting of m observations for n variables, x_1, x_2, … x_n, PCA calculates a new variable y_1 that accounts for the variation in the n original variables. The first PC, y_1, is the linear combination of the n original variables,

$$y_1 = w_{11}x_1 + w_{12}x_2 + \ldots + w_{1n}x_n, \tag{1}$$

where the sample variance is greatest for all of the coefficients, w_{11}, w_{12}, … w_{1n}, conveniently written as a vector $\boldsymbol{w_1}$. w_{11}, w_{12}, …w_{1n} need to satisfy the constraint that the sum-of-squares of the coefficients, i.e., $\boldsymbol{w'_1}\,\boldsymbol{w_1}$, should be unity. The second PC, y_2, is also given by the linear combination of the n original variables as,

$$y_2 = w_{21}x_1 + w_{22}x_2 + \ldots + w_{2n}x_n \tag{2}$$

or

$$\boldsymbol{y_2} = \boldsymbol{w'_2}\mathrm{x}$$

which has the greatest variance subject to two conditions,

$$\boldsymbol{w'_2 w_2} = 1 \text{ and } \boldsymbol{w'_2 w_1} = 0 \text{ (so that } y_1 \text{ and } y_2 \text{ are uncorrelated)}. \tag{3}$$

Similarly, the jth PC can be calculated as

$$y_j = \boldsymbol{w'_j}\mathrm{x}, \tag{4}$$

which has greatest variance subject to $\boldsymbol{w'_j w_j} = 1$, $\boldsymbol{w'_j w_i} = 0$ $(i < j)$.

The coefficients defining the first PC, i.e., the elements of $\boldsymbol{w_1}$ can be determined by maximizing the variance of y_1 subject to the constraint, $\boldsymbol{w'_1 w_1} = 1$. The variance of y_1 is then given by

$$\mathrm{Var}(y_1) = \mathrm{Var}(\boldsymbol{w'_1}\mathrm{x}) = \boldsymbol{w'_1}\mathbf{S}\,\boldsymbol{w_1}, \tag{5}$$

where $\mathbf{S}$ is the variance-covariance matrix of the original variables. The solution of w_{11}, w_{12}, … w_{1n} to maximize the variance y_1 is the eigenvector of $\mathbf{S}$ corresponding to the largest eigenvalue. The eigenvalues of $\mathbf{S}$ are roots of the equation,

$$|\mathbf{S} - \lambda\mathbf{I}| = 0. \tag{6}$$

The eigenvalues, $\lambda_1,\lambda_2,\ldots\lambda_n$ can be arranged from the largest to the smallest. The first d eigenvectors, $d \ll n$, are the PCs that capture most of the variance of the original data, whereas the remaining PCs can be considered as mainly representing noise, therefore, can be removed. Therefore, the data are reduced from dimension of n to a much smaller dimension of d, with loss of mainly the noise information. PCA is scale dependent, and so the data must be scaled in some meaningful way before PCA analysis. The most usual way of scaling is to scale each variable to unit variance.

The PCs are also called as scores, whereas the weights are often known as loadings. As every PC is the linear addition of the original variables, it can be interesting to know which original variables make more contribution than others to a specific PC. This can be analyzed by plotting the loadings. As the original variables are scaled before PCA is applied, the larger the loading, the more contribution that original variable makes.

The operation can be described in matrix algebra as follows:

$$[y]_{m\times n} = [x]_{m\times n}[w]_{m\times n} \tag{7}$$

$$[y]_{m\times d} \approx [x]_{m\times n}[w]_{m\times d}. \tag{8}$$

Then, the original variables can be reconstructed by the following operation:

$$[x]_{m\times n} \approx [y]_{m\times d}[w_{m\times d}]^T, \tag{9}$$

where T is the transpose. Clearly, the old data cannot be exactly reconstructed because the PCA operation has lost some information. PCA is the most important technique behind the advances in recent years in multivariate statistical process control,[14,15] and has also been successfully applied to model reduction of complex chemical mechanisms such as for combustion.[16,17]

Shape and size description in PC space

For a crystalline particle, such as the one shown in Figure 1a that has 26 faces, its shape and size can be represented by the normal distance of each face to the geometrical center.[12,13] Using the same method to define the shape and size of an noncrystalline, irregularly shaped particle in two dimension (2D), such as the one shown in Figure 1b,[18,19] the distance of every pixel at the edge to the geometric center of the object is required. To describe the shape and size of the same object in three dimension (3D), Figure 1c, it requires the distances to the geometric center of the particle from all the points on the particle surface. Here, it needs to point out that the discussion is about simultaneous description of both the shape and size of an object. To explain this more clearly, considering a plate object with zero thickness, length L, width W, its simultaneous description of both shape and size needs to include both L and W, but to purely describe its shape, only the aspect ratio, L/W is needed.

In summary, the shape and size of a single particulate object in 2D or 3D, either a structured crystal or an irregularly shaped particle, can be described by a finite number of sizes, namely, x_1, x_2, …, x_n. To consider a population of m particles, each with n shape and size descriptors, a matrix is formed as,

	x_1	x_2	----	x_n
particle - 1				
⋮				
particle - m				

(E-1)

Applying PCA to this data matrix and selecting d number of PCs, $d < n$, leads to:

	PC1	PC2	----	PCd
particle - 1				
⋮				
particle - m				

(E-2)

where PC1, PC2, … , PCd represent the new latent size dimensions, each is calculated by the linear combination of all the original sizes x_1, x_2, … , x_n. To conduct multidimensional PB modeling in the original size dimensional space, one has to consider all the n dimensions. The idea of multidimensional PB modeling in a PC space is to model the size distribution evolution in each of the PC directions, PC1, PC2, … , PCd. As d is expected to be much smaller than n, the number of size dimensions that need to be modeled can be significantly reduced. In an extreme situation, if $d = 1$, it turns out to be a one size dimensional PB that has particle shape information.

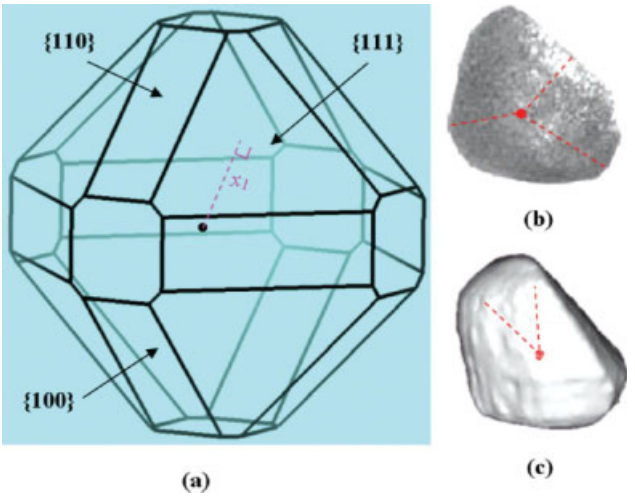

Figure 1. (a) Potash alum crystal with 26 faces, and (b, c) noncrystalline irregularly shaped particle in 2D and 3D.

In morphological PB model, the shape and size of the crystal in (a) is defined by the normal distances of individuals faces to the geometric center of the crystal. To apply the same approach to the objects in (b) and (c) will need the distance to the center of every pixel on the edge of the 2D object in (b), or every pixel on the surface of the 3D object in (c). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

As the $(n - d)$ components neglected are considered mainly noise, the dimension reduction process does not cause much loss of the shape information. At any time of simulation, the shape and size of particles represented by x_1, x_2, …, x_n can be reconstructed from the selected PCs, using Eq. 9.

In the case study to be presented in the next section for a crystallization process of potash alum, it will be demonstrated that 26 dimensions, defined as the normal distances of the 26 faces to the geometrical center, can be transformed to three PCs. For each latent, PC direction, growth rate can be estimated, and the dynamic changes of size distributions for the PCs are modeled. At any time of the simulation, the original 26 sizes can be reconstructed from three PCs. Therefore, it represents a reduction in size dimensions from 26 to 3. The principle demonstrated is applicable to noncrystalline irregularly shaped particles processing, for which shape-based multidimensional PB modeling has been considered by many people as having no obvious solutions.

Morphological PB Modeling of Potash Alum Crystallization in PC Space

In this section, a case study of a seeded potash alum crystallization process is presented to demonstrate that morphological PB modeling can be carried out in PC space. Potassium (potash) alum ($KAl(SO_4)_2{\cdot}12H_2O$) crystals can be easily crystallized from aqueous solutions. The crystal morphology, Figure 1a, is dominated by the large octahedron face {111} and two essential but considerably smaller faces, the cubic face {100} and the rhomb-dodecahedron face

{110}.[20,21] Three minor faces, {221}, {112}, and {012}, can only exist at early stages of the crystallization process. The three primary growth forms are manifested in the external morphology through the multiplicities associated with cubic symmetry, which yields a total of 26 crystal growth surfaces, i.e., eight {111}, six {100}, and 12 {110} crystallographically equivalent faces, respectively.

The operating condition for the seeded cooling process was deliberately maintained at low supersaturation level, so that the dominant process is crystal growth, and the effect of secondary nucleation is negligible, as already confirmed in an earlier study.[13] Crystal breakage and aggregation are not included either because for potash alum crystal, the complex inorganic hydrate compound displays no known cleavage planes for particle fracture. Strongly bonded equant particle morphology is not particularly known for attrition. The crystal chemistry for all crystal habit faces displays an intimate mixture of cation, anion, and hydrating water molecules in such a manner that significant particle/particle agglomeration is not expected. Under these assumptions, to be able to carry out morphological PB modeling in the PC space, all needed are the initial size distributions of the selected PC size dimensions, and the growth rate in each PC direction as a function of process operating conditions. In the following, we will first introduce how the original simulation data for the case study were generated, and then describe how the data were processed using PCA to estimate the initial size distributions of the selected PC dimensions, and the growth rate in each PC direction.

Generation of original data

The morphological PB model developed for seeded cooling crystallization of potash alum in our previous study[13] is used to generate the original data. Ideally, the model should provide the dynamic evolution data of size distributions in all the 26 crystal faces. Although such morphological PB equations can be easily formulated, it proved to be computationally too expensive to complete the simulation within a reasonable computational time. As the potash alum crystal has symmetrical faces, the 12 {110}, six {100}, and eight {111} faces, and the symmetrical faces were assumed to have identical growth behavior, 3D PB simulation was actually conducted.[13] Under the original assumption, i.e., symmetrical faces being identical, the simulation based on three independent faces can be used to generate data for the other 23 faces. But instead of using the data directly, we have introduced 10% white noise to the data so that even for symmetrical faces the data have variations. Figure 2 shows the time-varying mean distances to the particle center for the 26 faces. It needs to be emphasized that the purpose of the case study here is to demonstrate that morphological PB modeling can be carried out in the PC space. Once this is proved, the ways that PCA-based morphological PB modeling can be used will be discussed in a later section.

Calculation of initial PC size distributions and growth rates in PC directions

The generated data are arranged in a matrix format of expression E-3. The columns, x_1, x_2, … , x_{26}, represent the

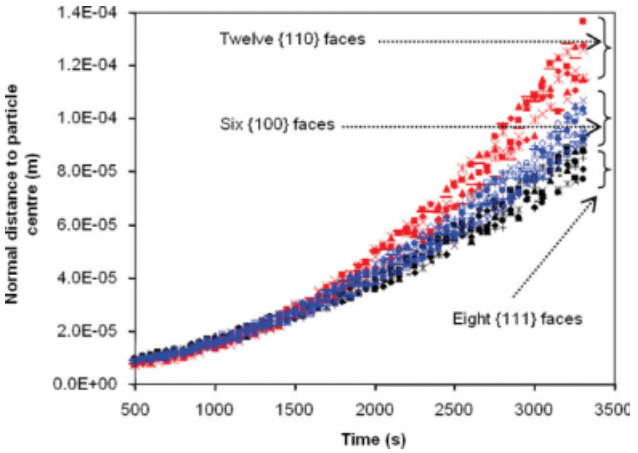

Figure 2. The time-varying mean distances to particle center for the 26 faces.

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

sizes of the 26 faces, i.e., the distances of the faces to the geometrical center of the crystal. The rows are divided into sections, t_1, t_2, … , t_T, representing the time points of simulation. At a given time point, there are m rows, representing m particles, particle-1, … , particle-m. Therefore, the data matrix shown in E-3 is of the size $(m \times T) \times 26$. It is expected that after PCA analysis of this matrix, the columns, x_1, x_2, …, x_{26}, will be reduced to PC1, PC2, … , PCd, $d \ll 26$, whereas there is no change in the number of rows.

time	particles	x_1	x_2	----	x_{26}
t_1	particle - 1				
	⋮				
	particle - m				
⋮	⋮				
t_T	particle - 1				
	⋮				
	particle - m				

(E-3)

For each time point of t_1, t_2, … , t_T of expression E-3, however, there would be too many rows because there are so many particles. In this case study, $m = 2.2 \times 10^7$. As a result, only a limited number of representative particles of different shapes and size were selected. The selection was based on the faceted size distributions of the seed crystals at the start of crystallization. As there are symmetrical faces, i.e., six {100}, 12 {100}, and eight {111} faces, the selection was based on the size distributions of three independent faces. For each of the three faceted Gaussian size distributions, 15 sizes were chosen as the mean μ, $\mu \pm 1/4\sigma$, $\mu \pm 2/4\sigma$, $\mu \pm 3/4\sigma$, $\mu \pm 4/4\sigma$, $\mu \pm 5/4\sigma$, $\mu \pm 6/4\sigma$, and $\mu \pm 7/4\sigma$, where μ is the mean and σ is the standard deviation of a Gaussian distribution. So, the total number of selected particles at each time point was $m = 15 \times 15 \times 15 = 3375$. The simulated time span was 3300 s, which was simulated over 57 time steps. So, the matrix in expression E-3 is $(m \times T) \times 26 = (3375 \times 57) \times 26$.

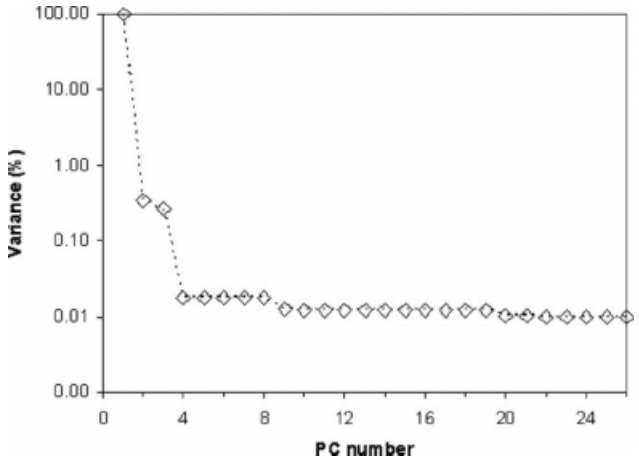

Figure 3. Variance of 26 principal components.

The curve becomes flat from the fourth principal component indicates that only three principal components are needed.

Please note that there are only two purposes here: to determine how many PCs are needed for carrying out PCA-based morphological PB modeling, and to estimate the growth rates in each PC direction. So, using only a selected number of particles that are representatives of different shapes and sizes in the matrix of E-3 is acceptable. Afterward, in the PCA-based PB modeling, the whole population of particles is to be used for the simulation instead of only the selected particles.

PCA is applied to the matrix of E-3. The variances of the 26 PCs are plotted in Figure 3, based on which three PCs were selected. Although various approaches were proposed for determining how many PCs should be retained, there is no definite conclusion on which method is the best. In the current case study, three PCs were selected based on the slope of plot of variance vs. PC number in Figure 3. Three PCs were selected because from the fourth PC there is a sharp change in the slope. It is interesting to note that the number of selected PCs equals the number of original independent crystal habit faces, representing {100}, {110}, and {111}. This might not be purely coincident: it could be the case that to fully reconstruct the original crystal shape of 26 faces, three size dimensions are needed, either three PC faces or three original independent faces representing the six {100}, 12 {110}, and eight {111} faces. Therefore, it naturally leads to a speculation whether the three PCs actually correspond to the three independent faces. It was found that they are not exactly the three independent faces, but some kind of relationships can be found using contribution plots. Figure 4 shows the contribution plots of the 26 faces to PC1, PC2, and PC3, plotted over time. It can be seen that for PC1, the contributions of the six {100} faces are larger than that by the 12 {110} and eight {100} faces. This means that PC1 represents more information of the six {100} faces than other faces. However, the contributions by 12 {110} and eight {100} faces are not zero, so PC1 cannot be considered as only representing the six {100} faces. Similar observations can be made for PC2 and PC3.

After PCA analysis and the selection of three PCs, the matrix of E-3 becomes a new matrix, which has three columns, but with no change in the number of rows. This new matrix is used to estimate the growth rates in each PC direction. At each time point of the time span of simulation, t_1, t_2, … ,t_T, for each PC size dimension, the m particles

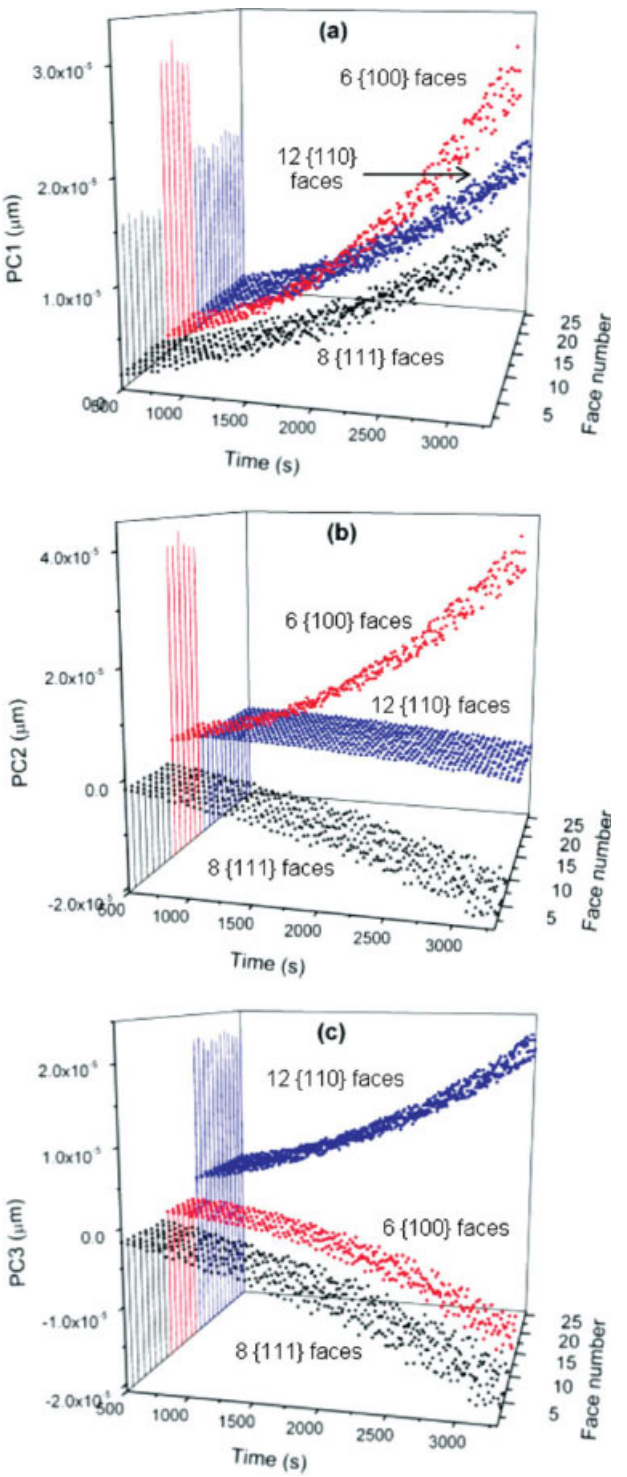

Figure 4. Contribution plot of every face to each selected principal component, plotted over time.

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

have their size distributions. The mean is considered as the size of the particles at that time point. Instead of directly calculating the growth rate data from these mean sizes, we first correlated these mean sizes with supersaturation. The sizes of the three PC directions are correlated with the relative supersaturation as follows,

$$\text{Size}_{\text{PC1}} = 0.0483\ \sigma^2 - 0.00464\ \sigma + 1.2118 \times 10^{-4} \quad (10)$$

$$\text{Size}_{\text{PC2}} = 0.01058\ \sigma^2 - 0.00197\ \sigma + 7.6469 \times 10^{-4} \quad (11)$$

$$\text{Size}_{\text{PC3}} = -2.5299 \times 10^{-4}\ \sigma^2 + 4.122 \times 10^{-5}\ \sigma + 1.5302 \times 10^{-5}, \quad (12)$$

where σ is the relative supersaturation defined as the ratio between crystallization driving force (C-C*) and solubility (C*). Equations 10–12 were then used to calculate the growth rates. Equations 10–12 are used to generate the growth rate data for each PC face, which was then correlated with supersaturation and size using Eqs. 13–15. Here, it was assumed that each selected PC face is a function of both supersaturation and the size of the face.

$$G_1 = 1.05 \times 10^{-5}\sigma^{1.24} \times \text{Size}_{\text{PC1}}{}^{0.16} \quad (13)$$

$$G_2 = 6.28 \times 10^{-7}\sigma - 5.84 \times 10^{-8} \quad (14)$$

$$G_3 = -1.50 \times 10^{-8}\sigma + 1.22 \times 10^{-9}. \quad (15)$$

It turned out that only the growth at the PC1 direction is size dependent.

PB model in the PC space

The PCA-based 3D PB equation can be written as follows:

$$\frac{1}{V_{\text{T}}(t)}\frac{\partial}{\partial t}[\psi(\text{PC1},\text{PC2},\text{PC3},t)V_{\text{T}}(t)] + \frac{\partial}{\partial \text{PC1}}[G_1(\text{PC1},t)\psi(\text{PC1},\text{PC2},\text{PC3},t)] + \frac{\partial}{\partial \text{PC2}}[G_2(\text{PC2},t)\psi(\text{PC1},\text{PC2},\text{PC3},t)] + \frac{\partial}{\partial \text{PC3}}[G_3(\text{PC3},t)\psi(\text{PC1},\text{PC2},\text{PC3},t)] = 0, \quad (16)$$

where ψ is the number population density function at time t, G_1, G_2, and G_3 are the growth rates in three PC directions, which can be estimated using Eqs. 13–15, and V_{T} is the total volume of the suspension in the reactor. The first term on the left-hand side in Eq. 16 is the accumulation of population. The second, third, and fourth terms are the population changes for the three PCs, respectively, because of the crystal growth in the corresponding directions.

To solve Eq. 16, the initial Gaussian distributions of crystal population in 26 dimensions (Figure 5) were converted to new distributions in the three PC dimensions. The new distributions were curve-fitted using a Gaussian function with the mean values of 43.5, −6.1, and −1.0 μm for PC1, PC2, and PC3, respectively, in the latent variable space (please note, in the latent variable space, negative values are allowed). The 3D PCA-based PB equation (Eq. 16) can be solved using a discretization method, moment of classes, which is an extension of the numerical solution procedure for a 2D PB equation described in David et al.[22] and Puel et al.[1] The method is only briefly described here because its detailed description can be found in literature.[1,8,12,22] It is worth noting that there are many other algorithms developed in literature,[4,23–27] but moment of classes has proved to be effective for the current process as demonstrated previously,[12,13] so is also used here.

The 3D size domain was equally discretized into n_1, n_2, n_3 classes. The corresponding size and characteristic dimensional length of each class in each dimension are ΔPC1_i =

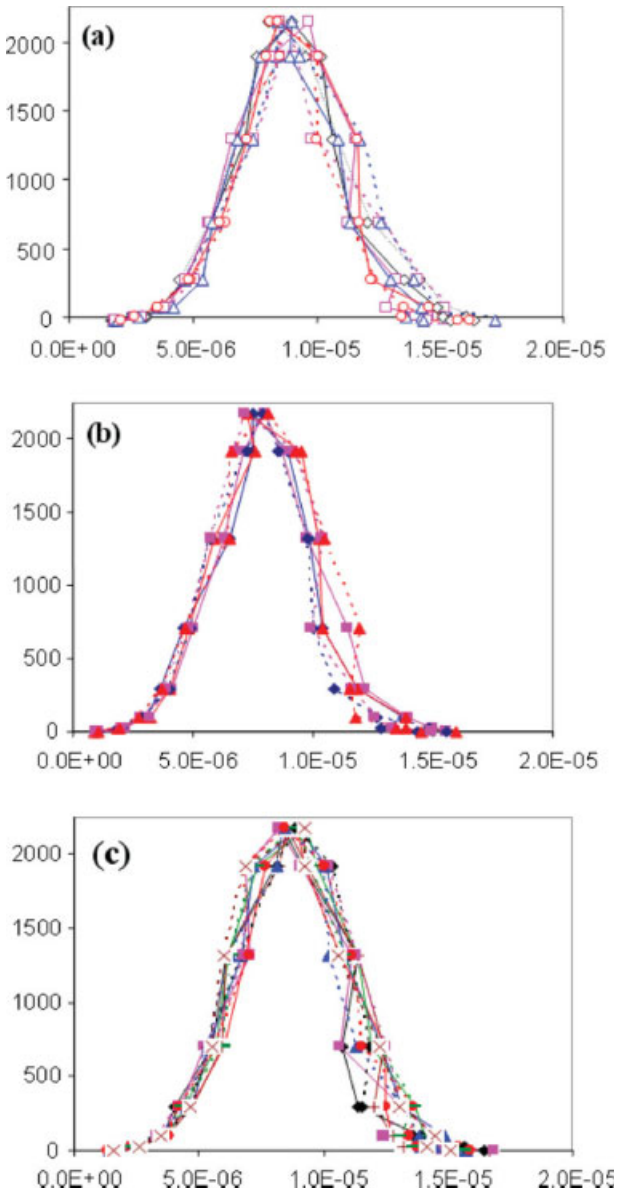

Figure 5. Initial size distributions for 26 faces.

(a) eight {111} faces, (b) six {100} faces, and (c) 12 {110} faces. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$PC1_i - PC1_{i-1}$, $\Delta PC2_j = PC2_j - PC2_{j-1}$, $\Delta PC3_k = PC3_k - PC3_{k-1}$ and $\overline{PC1}_i = (PC1_{i-1} + PC1_i)/2$, $\overline{PC2}_i = (PC2_{i-1} + PC2_i)/2$, $\overline{PC3}_i = (PC3_{i-1} + PC3_i)/2$, respectively. After discretization over the size domain, a set of $n_1 \times n_2 \times n_3$ ordinary differential equations can be formed by integrating Eq. 16 over class $Cl_{i,j,k}$ of PC1, PC2, and PC3:

$$\frac{1}{V_T(t)}\frac{d}{dt}[N_{i,j,k}(t)V_T(t)] + [FX^o_{i,j,k}(t) - FX^I_{i,j,k}(t)] + [FY^o_{i,j,k}(t) - FY^I_{i,j,k}(t)] + [FZ^o_{i,j,k}(t) - FZ^I_{i,j,k}(t)] = 0, \quad (17)$$

where $N_{i,j,k}(t)$ is the number of crystals in the class $Cl_{i,j,k}$:

$$N_{i,j,k}(t) = \int_{PC1_{i-1}}^{PC1_i}\int_{PC2_{j-1}}^{PC2_j}\int_{PC3_{k-1}}^{PC3_k} \psi(PC1, PC2, PC3, t)\, dPC1\, dPC2\, dPC3. \quad (18)$$

The first term on the left-hand side of Eq. 17 is the accumulation term and the second term is the net flow of crystals in class $Cl_{i,j,k}$ in PC1 direction with the superscripts, O and I, denoting the crystal flowing outletting from and inletting into the $Cl_{i,j,k}$ class:

$$FX^o_{i,j,k}(t) - FX^I_{i,j,k}(t) = G_1(\overline{PC}1_i, t)[a_i N_{i,j,k}(t) + b_i N_{i+1,j,k}(t)] - G_1(\overline{PC}1_{i-1}, t)[a_{i-1}N_{i-1,j,k}(t) + b_{i-1}N_{i,j,k}(t)], \quad (19)$$

where $a_i = \Delta PC1_{i+1}/[\Delta PC1_i(\Delta PC1_{i+1} + \Delta PC1_i)]$ and $b_i = \Delta PC1_i/[\Delta PC1_{i+1}(\Delta PC1_{i+1} + \Delta PC1_i)]$. Similarly, the third and fourth terms on the left hand side of Eq. 17 are the net flows of crystals in class $Cl_{i,j,k}$ in PC2 and PC3 directions, which can be estimated in the similar way. The boundary conditions for crystal flow fluxes are as follows:

$$FX^I_{1,j,k}(t) = FY^I_{i,1,k}(t) = FZ^I_{i,j,1}(t) = 0 \quad (20)$$

$$FX^o_{1,j,k}(t) = FY^o_{i,1,k}(t) = FZ^o_{i,j,1}(t) = 0. \quad (21)$$

The solid concentration, $C_s(t)$, in unit volume of suspension in a well-mixed batch crystallizer can be calculated by

$$C_s(t) = \frac{\rho_s}{M_s}\int_{x_1}\int_{x_2}\cdots\int_{x_{26}} V_c\psi(x_1, x_2, \cdots, x_{26}, t)dx_1 dx_2 \cdots dx_{26}. \quad (22)$$

In the current case, because of their similarity of eight {111} faces, six {100} faces, and 12 {110} faces, a simpler formulation with their corresponding average normal distances, $\overline{x}, \overline{y}, \overline{z}$ for {111}, {100}, and {110} faces, respectively, was used to calculate the solid concentration as

$$C_s(t) = \frac{\rho_s}{M_s}\int_{\overline{x}}\int_{\overline{y}}\int_{\overline{z}} V_c \cdot \psi(\overline{x}, \overline{y}, \overline{z}, t)d\overline{x}d\overline{y}d\overline{z} \quad (23)$$

to simplify the calculation and save computational time. With negligible effect of crystallization and temperature variation on total volume, the volume of suspension, $V_T(t)$, can be calculated by $V(0)/[1 - M_s/\rho_s\ C_s(t)]$, and the solute concentration, $C(t)$, can be estimated with $C(0) - C_S(t)/[1 - M_s/\rho_s\ C_s(t)]$.

The discretized PCA-based PB equations (Eq. 17), together with a Gaussian-type initial population distribution and the boundary conditions (Eqs. 20 and 21), were solved in the mesh ranges of 22.5–113.8, −23.5 to 50.7, and −25.5 to 78.8 μm in PC1, PC2, and PC3 directions, respectively, with the number of the size classes being $100 \times 100 \times 100$ and the corresponding mesh width being $0.91 \times 0.74 \times 1.0$ μm. The generated 1 million discretized PB equations in ordinary differential form were solved simultaneously using the Runge–Kutta–Fehlbergh fourth/fifth order solver[28] with automatic time-step control and the criteria of the absolute and relative tolerance being 10^{-6} and 10^{-4}, respectively. A moving mesh technique[24] was applied during the calculation when the mean values of the PCs (PC1, PC2, and PC3) were propagated a few meshes from their starting locations. Test runs with either larger number of size classes, hence smaller corresponding mesh width, or smaller absolute and relative tolerance produced similar prediction.

For a crystal with known values for PC1, PC2, and PC3, the reconstruction of its values of x_1, x_2, $\ldots x_{26}$ can be easily performed using the loading matrix and Eq. 9.

Results

Figure 6 shows examples of size distributions of three PCs at 500th, 2000th, and 3300th s. It can be seen that the most significant PC, PC1, grew much faster with time (Figure 6a). The second significant PC, PC2, was first decreasing with time, then increasing after 2150th s. The third PC, PC3, did not show obvious change in size during crystallization, but with slightly negative growth because of the small negative growth rate. Using the PCA reconstruction technique, the size distributions of the original 26 faces can be obtained and three typical distributions in x_1, x_9, and x_{15} face directions, representing {111}, {100}, and {110} faces, respectively, at different crystallization stages were illustrated in Figure 7 and, for comparison, also plotted in Figure 7 are the size distributions of {111}, {100}, and {110} face directions obtained from the original morphological PB model.[13] Figure 7 indicates that the PCA reconstruction produced very similar size distributions to that of the original simulation. The small discrepancy on the shape of size distribution and mean values of the three face directions might have been caused by the truncation of the less important, 23 PCs and/or the curve-fitting errors during the estimation of growth rates of the three PCs. The differences of mean values between the two simulations at the crystallization time of the 3300th s are 0.5, 2, and 1.7 μm for the {111}, {100}, and {110} faces or 0.6, 1.6, and 1.7% in relative errors.

As the sizes for the 26 faces can be reconstructed from PCs, the shape of crystals can be drawn at any time of the crystallization process. Figure 8 shows the shape evolution of a crystal. In Figure 8, at 3300 s, it shows the comparison of crystal shape reconstructed from PCs and the original shape from the original simulation.

The crystal surface areas of the three independent faces, {111}, {100}, and {110}, reconstructed after PCA-based PB simulation and from original morphological PB model, show good agreement. Figure 9 depicts the evolution of the {110}

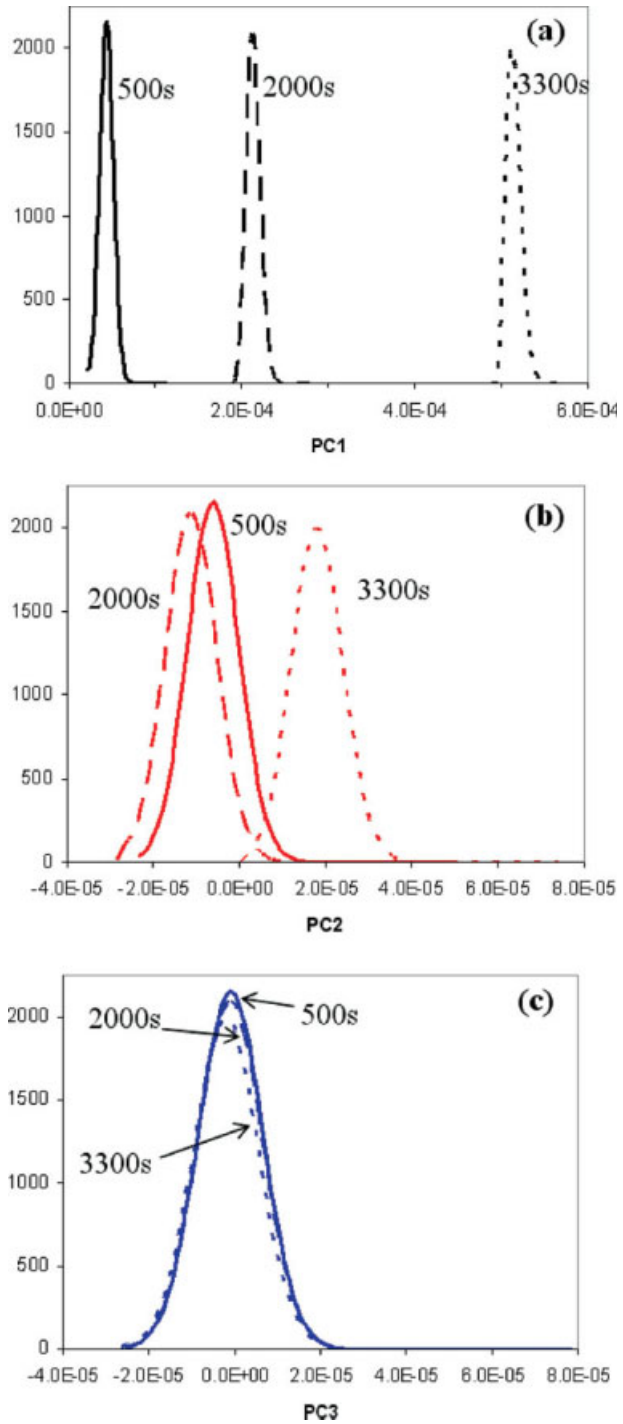

Figure 6. Typical size distributions for individual principal components.

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

boundary sector obtained in the current study and from the previous simulation.[13] In Figure 9, the rings refer to different steps of the simulation, i.e., the growth history. Agreement between the two simulations is clearly good. The comparisons for the two characteristic habit faces, {100} and {110}, are plotted in Figure 10. It can be seen from Figures 10c, d that for face {110}, the growth history predicted by the PCA method matches the original simulation very well. For face {100}, i.e., Figures 10a, b, PCA reconstruction misses some details in comparison with the growth history predicted by the original simulation. Nevertheless, even for face {100}, the major features such as the growth trend and the surface area are well preserved by the PCA-based simulation.

The computational time saved by carrying out a 3D PB simulation, in comparison with otherwise carrying out a 26-dimensional PB simulation, cannot be directly estimated. This is because it has not been able to implement the later because of the huge computational resource required. However, on the basis of the 3D PB calculation, we made an order of magnitude estimation for the time that could be needed to carry out a 26-dimensional PB simulation, which turns out to be about 46 orders of magnitude longer than the 3D PB modeling using PCA.

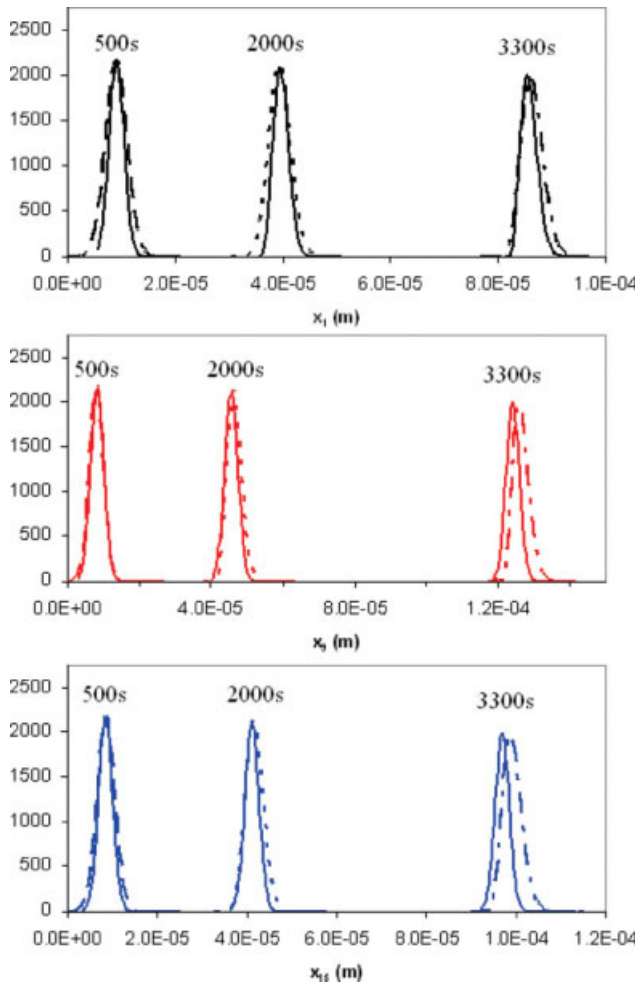

Figure 7. Size distributions of individual faces showing good agreement between solid curves (reconstructed from principal components) and dotted curves (original simulation without using PCA).

x_1 one of the eight {111} faces; x_9 one of the six {100} faces; and x_{15} one of the 12 {110} faces (c). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

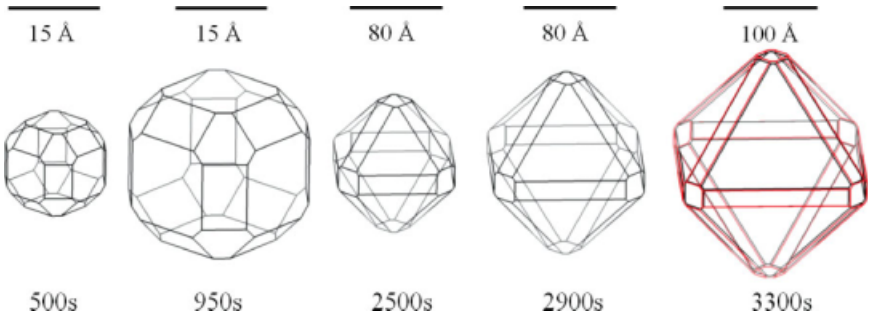

Figure 8. Shape evolution for a crystal, reconstructed from principal components.

Please note: (1) The crystal at 2500 s is larger than the crystal at 950 s because the scale 80 Å is larger than the scale 15 Å; (2) At 3300 s, black color is reconstructed from principal components and red color is the original shape. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Discussion

The results demonstrated that morphological PB modeling can be effectively performed in the transformed, latent variable space. The complicated real particle shape at any time of simulation can be reconstructed with high fidelity from only a few PCs.

The case study based on which the above conclusion is reached also brings about a question: now the principle is proved, but how can the methodology be used in practice? The case study used original data of crystal size distributions of all 26 faces over the time span studied; however, it cannot be expected that such original data are always available for all other processes. As a matter of fact, as stated earlier in the article, that even the full morphological PB model using all the original size dimensions can be built, there is great difficulty in obtaining a solution within an acceptable computational time.

The PCA-based morphological PB modeling technique can be used in various ways and is open for future exploration. Below, we look at one way of its application to show that it can be used in a much simpler way than in the case study, where the purpose was to demonstrate a principle. We still consider a seeded cooling crystallization as an example. In order not to complicate the discussion, we also assume that secondary nucleation can be ignored, and there is no crystal aggregation or breakage, i.e., the total number of crystals does not change. Only the size distributions of the growing crystals vary with time.

First, the seed particles should be characterized to get data of the morphology and size. The obtained data are arranged into a matrix format of expression (E-1), with the m rows as the particles that have been analyzed, and the n columns as the descriptors characterizing the shape and size. PCA is then applied to the matrix. After the selection of the number of PCs, a new matrix as shown in the form of expression (E-2) is obtained. Based on the matrix of (E-2), the size distribution for each of the selected PCs can be calculated. The aforementioned procedure produces: (1) the number of PCs, d, for which PB simulation will be carried out; (2) the initial size distributions for the selected d PCs; and (3) a transformation matrix containing the loadings that can be used at any time of the simulation to reconstruct the original n variables (original shape descriptors) from the values of the d PCs.

The only missing item prior to PCA-based morphological PB simulation is the growth rate of each PC as a function of supersaturation (and its size if the crystal shows size-dependent growth):

$$G_{\mathrm{PC}_i} = f(\text{supersaturation } \sigma, \text{size of the PC}_i, \text{parameters}). \quad (21)$$

Suppose the structure of above equation is known, the PCA-based PB equations can be formulated with the values of the parameters in Eq. 21 to be determined. The values of the parameters can be obtained via model identification using data from a real crystallization experiment.

A real cooling crystallization experiment can be carried out using the seed crystals characterized, with operating conditions, e.g., supersaturation and temperature continuously recorded during the process. The product at the end should be characterized in the same way as for the seed crystals. The product characterization data will be used as the targets in model identification for determining the parameters in Eq. 21. PCA-based morphological PB models thus built can be used to simulate the crystallization process of the same chemical in other operational conditions and varied seed crystal shape and size. Determining growth rate equations for the PC faces in PCA-based morphological PB modeling probably is the most practical solution to the problem. Model identification has proved to be an effective approach for determining the length and width growth rates for plate-like crystals in a 2D PB model developed by Oullion et al.[6]

The aforementioned procedure for seeded cooling crystallization could be applied to processes producing irregularly shaped particles. As far as PCA is concerned, the methodology for dealing with particles whose shape is represented by 26 size dimensions can be directly applied to particles for which the shape needs to be described by hundreds of size descriptors. The original particle shape and size can be characterized first. The characterization data can be processed using PCA to obtain the number of PCs to be selected and the size distributions for the PCs at time equal to zero. It also provides the loading matrix to be used for shape reconstruction in future. We can use one batch run of the process

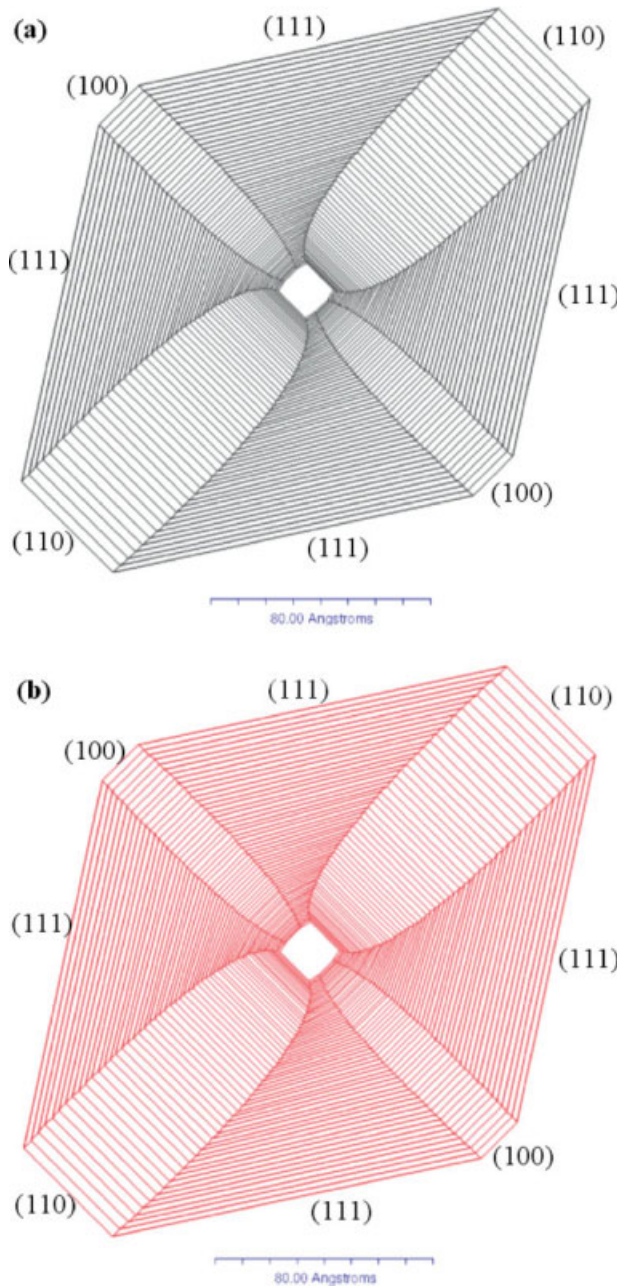

Figure 9. Evolution of {110} boundary sector, (a) reconstructed from principal components and (b) original simulation.

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

to get the final product. The product is also characterized in the same way as the original raw materials. The data can be used to determine the parameters used in the PCA-based models, such as in particle breakage or aggregation models. Clearly, more research is needed to consider how to deal with breakage and aggregation. At least the proposed idea in this article provides a methodology that can potentially allow us to carry out shape-based PB modeling for irregularly shaped particles, which there has been no obvious solution to date.

The aforementioned discussion describes examples of applying PCA-based morphological PB modeling without the need of relying on the original morphological PB model. However, even for a process that the original morphological PB model can be built and solved within an acceptable computational time, the PCA-based approach can still offer benefits. For example, in the context of optimization, the expensive simulation is run only once to enable the PCA; afterward, the PCA-based PB model replaces the original model and is then used in the subsequent optimization. In this case, however, one has to ensure that the PCA model remains valid in all the process scenarios examined by the optimizer, which may be quite different from the case where the initial dataset was generated.

Concluding Remarks

The work has provided a proof of concept that morphological PB models can be built in a transformed, dimensionally reduced, latent variable domain, i.e., the PC space. In principle, by defining the morphology of a crystal using the distances of all its faces to its geometrical center, the morphological PB model proposed previously[12,13] can model the population growth behavior of crystal of any complicated shape with no limitation on the number of size dimensions. In practice, the enormous computational load needed means that if the number of crystals faces is high, the morphological PB equations cannot be solved within an acceptable time. The proposed PCA morphological PB modeling approach has effectively addressed this problem because it can significantly reduce the number of dimensions without compromising much on the particle shape information retained. In addition, for irregularly shaped particles, such as those produced in milling and granulation, no attempts have been made to introduce particle shape information into PB simulation. The multidimensional PB models proposed in literature for crystallization are not applicable to such particulate processes. For example, many have used two characteristic sizes to describe crystal shape in 2D PB for crystallization, e.g., the length and width of a plate crystal. It is clear that the same approach cannot be adopted for irregularly shaped particles. The proposed PCA-based morphological PB approach in this article has made it potentially realistic to carry out particle shape-based PB for particulate processes with irregularly shaped particles. In contrast to the case study presented in this article that used 26 size dimensions to define the shape and size of a crystal, for an irregularly shaped particle, several hundred size dimensions might be required to faithfully represent its shape and size. But the latter can still be reduced to a significantly smaller number of PCs, in the same way as reducing from 26 size dimensions to three PCs. Of course, there will be anticipated and unexpected new issues arising that have to be resolved to apply the proposed method in this article to processes handling irregularly shaped particles. An example of these issues is that the method requires the number of original descriptors representing shape and size for all particles, large or small, to be the same. These issues will need new research but the new PB modeling methodology proposed in his article has laid the foundation forward for future investigations.

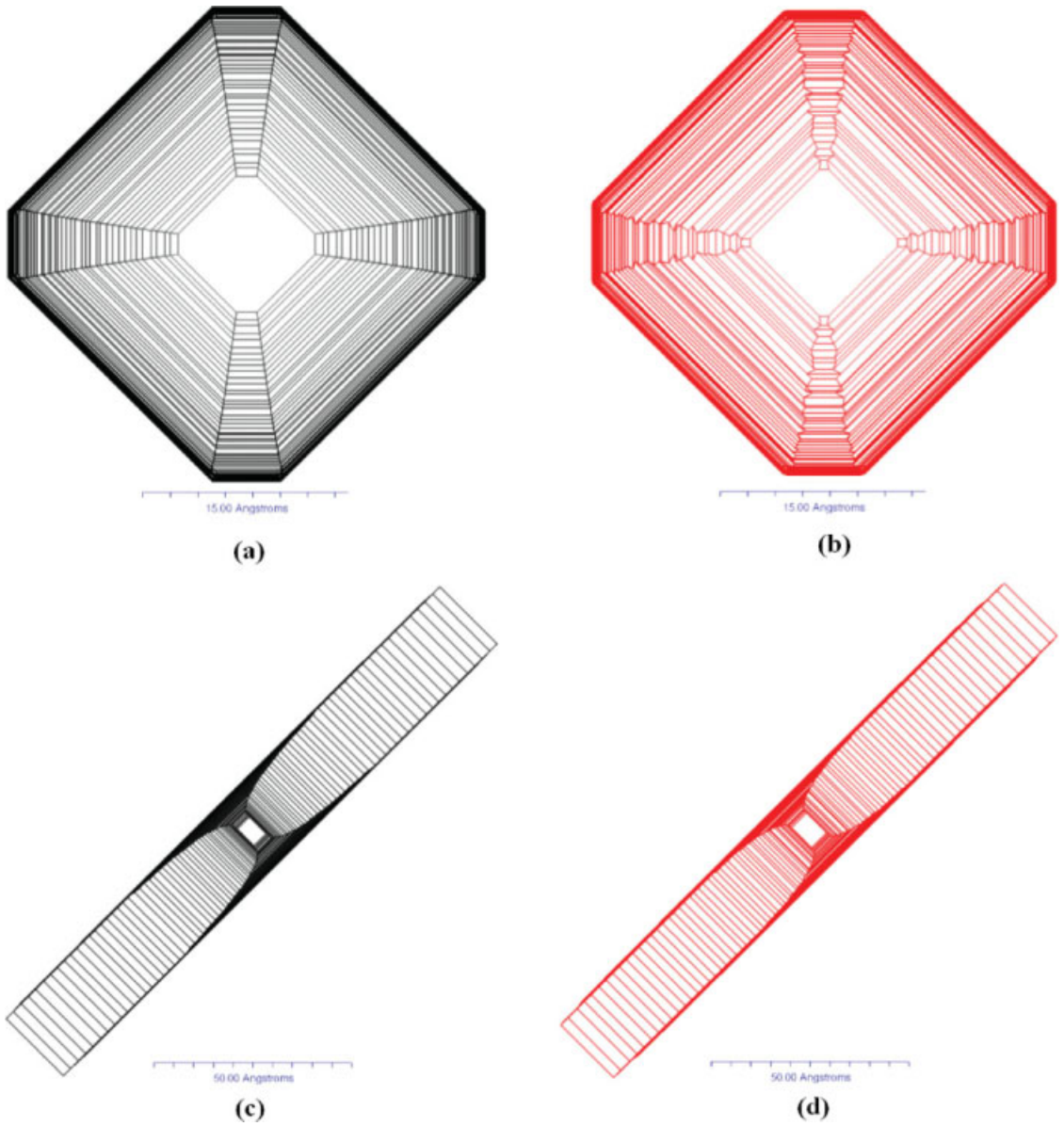

Figure 10. Comparison of the reconstructed faces from principal components and the original simulation results.

(a) Reconstructed {100} face, (b) original {100} face, (c) reconstructed {110} face, and (d) original {110} face. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Acknowledgments

The work presented in this article is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) in two projects (Grant references: EP/C009541 and EP/E045707). The authors thank the industrial participants of the two projects: AstraZeneca (Dr. Gerry Steele and Dr. Mike J. Quayle), Malvern Instruments (Mr. Fraiser-McNeil Watson, Mr. Andy Prior, Mr. David Watson, and Mr. Duncan Roberts), National Nuclear Laboratory (Dr. Dominic Rhodes), Pfizer (Dr. Ivan Marziano and Prof. Robert Docherty), Syngeta (Dr. Neil George), and 3M Health Care (Dr. Chris Blatchford). They also thank Dr. Ruifa Li, Dr. Jian Wan, and Prof. Kevin J. Roberts for useful discussions. The sponsorship of Malvern Instruments for the University of Leeds–Malvern Partnership IntelliSense (www.intellisense.org.uk) is also gratefully acknowledged.

Notation

a_i, b_i = coefficient, m^{-1}
C = solute concentration, mol m^{-3}
$\boldsymbol{C}$ = covariance of a vector $\boldsymbol{X}$
C^* = solubility, mol m^{-3}
$Cl_{i,j,k}$ = three-dimensional class
C_S = concentration of solid in the suspension, mol m^{-3}
d = number of eigenvectors selected
dPC1, dPC2, dPC3 = differential distances in PC1, PC2, PC3 directions, m
dx_1, dx_2,⋯, dx_{26} = differential distances in x_i (i = 1,26) directions
$d\bar{x}$, $d\bar{y}$, $d\bar{z}$ = differential distances in $\bar{x}, \bar{y}, \bar{z}$ directions, m
$FX_{i,j,k}$, $FY_{i,j,k}$, $FZ_{i,j,k}$ = net flow rate of crystals in class $Cl_{i,j,k}$ in PC1, PC2, PC3 directions, [nb] m^{-3} s^{-1}
G_1, G_2, G_3 = growth rate of three principal components, m s^{-1}
L = length, m
$\boldsymbol{M}$ = mean vector of a vector $\boldsymbol{X}$
M_s = molecular weight of solid, kg mol^{-1}
m = total number of particles or representative particles
m_1, m_2, ⋯, m_n = n dimensional values of a mean vector $\boldsymbol{M}$

$N_{i,j,k}$ = number of crystals in the class $Cl_{i,j,k,}$ [nb] m^{-3}
n = size dimensions
n_1, n_2, n_3 = number of classes in PC1, PC2, PC3 directions
PC1, PC2, PC3 = principal component, m
$PC1_i$, $PC2_j$, $PC3_k$ = discretized characteristic parameters of crystal, m
PC1, PC2, …, PCn = eigenvectors of n dimensions
PC1, PC2, …, PCd = eigenvectors of d dimensions
$\overline{PC1_i}$,$\overline{PC2_j}$,$\overline{PC3_k}$ = characteristic parameters of class Cl_i, Cl_j, and Cl_k
$\boldsymbol{S}$ = eigenvector
t = time, s
t_1, t_2,…, t_T = time instant, s
T = total number of time sections
V = volume of solution, m^3
V_c = volume of crystals, m^3
V_T = total volume of suspension, m^3
W = width, m
$\boldsymbol{w_1}$, $\boldsymbol{w_2}$, …, $\boldsymbol{w_n}$ = vectors of PCA coefficients
$\boldsymbol{X}$ = a vector of n dimensions
x_1, x_2, ⋯, x_{26} = normal distances in multiple face directions, m
x_1, x_2, ⋯, x_n = n dimensional values of a vector $\boldsymbol{X}$
$\bar{x}, \bar{y}, \bar{z}$ = average normal distances ($\bar{x} = \sum_{i=1}^{8} x_i/8$, $\bar{y} = \sum_{i=9}^{14} x_i/6, \bar{z} = \sum_{i=15}^{26} x_i/12$), m
y_1, y_2, …, y_n = eigenvectors of n dimensions

Greek letters

$\Delta PC1_i$, $\Delta PC2_j$, $\Delta PC3_k$ = extent of classes Cl_i, Cl_j, and Cl_k, m
λ_1, λ_2, …, λ_n = n dimensional eigenvalues
μ = mean of Gaussian distribution
ρ_s = density of solid, kg m^{-3}
σ = relative supersaturation ($=C/C^* - 1$)
σ = standard deviation
ψ = number population density function in the suspension, [nb] m^{-3} m^{-3}

Literature Cited

1. Puel F, Fevotte G, Klein JP. Simulation and analysis of industrial crystallization processes through multidimensional population balance equations. I. A resolution algorithm based on the method of classes. *Chem Eng Sci*. 2003;58:3715–3727.
2. Puel F, Fevotte G, Klein JP. Simulation and analysis of industrial crystallization processes through multidimensional population balance equations. II. A study of semi-batch crystallization. *Chem Eng Sci*. 2003;58:3729–3740.
3. Puel F, Marchal P, Klein J. Habit transient analysis in industrial crystallization using two dimensional crystal sizing technique. *Chem Eng Res Des*. 1997;75:193–205.
4. Ma DL, Tafti DK, Braatz RD. High-resolution simulation of multidimensional crystal growth. *Ind Eng Chem Res*. 2002;41:6217–6223.
5. Oullion M, Puel F, Fevotte G, Righini S, Carvin P. Industrial batch crystallization of a plate-like organic product. In situ monitoring and 2D-CSD modelling. I. Experimental study. *Chem Eng Sci*. 2007; 62:820–832.
6. Oullion M, Puel F, Fevotte G, Righini S, Carvin P. Industrial batch crystallization of a plate-like organic product. In situ monitoring and 2D-CSD modelling. II. Kinetic modelling and identification. *Chem Eng Sci*. 2007;62:833–845.
7. Briesen H. Simulation of crystal size and shape by means of a reduced two-dimensional population balance model. *Chem Eng Sci*. 2006;61:104–112.
8. Ma CY, Wang XZ, Roberts KJ. Multi-dimensional population balance modelling of the growth of rod-like L-glutamic acid crystals using growth rates estimated from in-process imaging. *Adv Powder Technol*. 2007;18:707–723.
9. Lovette MA, Browning AR, Griffin DW, Sizemore JP, Snyder RC, Doherty MF. Crystal shape engineering. *Ind Eng Chem Res*. 2008;47:9812–9833.
10. Wan J, Wang XZ, Ma CY. Particle shape manipulation and optimization in cooling crystallization involving multiple crystal morphological forms. *AIChE J*. 2009;55:2049–2061.
11. Sato K, Nagaib H, Hasegawab K, Tomorib K, Kramer HJM, Jansensa PJ. Two-dimensional population balance model with breakage of high aspectr atio crystals for batch crystallization. *Chem Eng Sci*. 2008;63:3271–3278.
12. Ma CY, Wang XZ, Roberts KJ. Morphological population balance for modelling crystal growth in individual face directions. *AIChE J*. 2008;54:209–222.
13. Ma CY, Wang XZ. Crystal growth rate dispersion modelling using morphological population balance. *AIChE J*. 2008;54:2321–2334.
14. Nomikos P, MacGregor JF. Monitoring batch processes using multi-way principal component analysis. *AIChE J*. 1994;40:1361–1375.
15. Chiang LH, Russell EL, Braatz RD. *Fault Detection and Diagnosis in Industrial Systems*. London: Springer, 2001.
16. Gokulakrishnan P, Lawrence AD, McLellan PJ, Grandmaison EW. A functional-PCA approach for analyzing and reducing complex chemical mechanisms. *Comput Chem Eng*. 2007;30:1093–1101.
17. Brown NJ, Li GP, Koszykowski ML. Mechanism reduction via principal component analysis. *Int J Chem Kinet*. 1997;29:393–414.
18. Bottlinger M, Bujak B, Bruecher M. *Three-Dimensional Measurement of Particle Shapes and Simulation of Bulk Solid Properties. Control of Particulate Processes VII*. BC, Canada: Harrison Hot Springs, 2006.
19. Bujak B, Bottlinger M. Three-dimensional measurement of particle shape. *Part Part Syst Charact*. 2008;25:293–297.
20. Klapper H, Becker RA, Schmiemann D, Faber A. Growth-sector boundaries and growth-rate dispersion in potassium alum crystals. *Cryst Res Technol*. 2002;37:747–757.
21. Amara N, Ratsimba B, Wilhelm A, Delmas H. Growth rate of potash alum crystals: comparison of silent and ultrasonic conditions. *Ultrason Sonochem*. 2004;11:17–21.
22. David R, Marchal P, Marcant B. Modeling of agglomeration in industrial crystallization from solution. *Chem Eng Technol*. 1995;18: 302–309.
23. Gerstlauer A, Gahn C, Zhou H, Rauls M, Schreiber M. Application of population balances in the chemical industry—current status and future needs. *Chem Eng Sci*., 2006;61:205–217.
24. Kumar S, Ramkrishna D. On the solution of population balance equations by discretization. III. Nucleation, growth and aggregation of particles. *Chem Eng Sci*., 1997;52:4659–4679.
25. Henson MA. Cell ensemble modeling of metabolic oscillations in continuous yeast cultures. *Comput Chem Eng*. 2005;29:645–661.
26. Pinto MA, Immanuel CD, Doyle FJ. A feasible solution technique for higher-dimensional population balance models. *Comput Chem Eng*. 2007;31:1242–1256.
27. Sotowa KI, Naito K, Kano M, Hasebe S, Hashimoto I. Application of the method of characteristics to crystallizer simulation and comparison with finite difference for controller performance evaluation. *J Process Cont*. 2000;10:203–208.
28. Shampine LF, Watts HA. *Subroutine RKF45*. In: Forsythe GE, Malcolm MA, Moler CB, editors. *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice-Hall, 1977:135–147.

Manuscript received Oct. 5, 2008, and revision received Jan. 12, 2009.