

# Robust Bayesian Estimation of Kinetics for the Polymorphic Transformation of L-Glutamic Acid Crystals

Martin Wijaya Hermanto, Nicholas C. Kee, Reginald B. H. Tan, and Min-Sen Chiu

Dept. of Chemical and Biomolecular Engineering, National University of Singapore, Singapore, Singapore 117576

Richard D. Braatz

Dept. of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

DOI 10.1002/aic.11623

Published online November 4, 2008 in Wiley InterScience (www.interscience.wiley.com).

*Polymorphism, in which there exist different crystal forms for the same chemical compound, is an important phenomenon in pharmaceutical manufacturing. In this article, a kinetic model for the crystallization of L-glutamic acid polymorphs is developed from experimental data. This model appears to be the first to include all of the transformation kinetic parameters including dependence on the temperature. The kinetic parameters are estimated by Bayesian inference from batch data collected from two in situ measurements: ATR-FTIR spectroscopy is used to infer the solute concentration, and FBRM that provides crystal size information. Probability distributions of the estimated parameters in addition to their point estimates are obtained by Markov Chain Monte Carlo simulation. The kinetic model can be used to better understand the effects of operating conditions on crystal quality, and the probability distributions can be used to assess the accuracy of model predictions and incorporated into robust control strategies for polymorphic crystallization. © 2008 American Institute of Chemical Engineers AICHE J, 54: 3248–3259, 2008*

*Keywords: pharmaceutical crystallization modeling, polymorphism, Bayesian inference, Markov Chain Monte Carlo*

## Introduction

Polymorphism, in which multiple crystal forms exist for the same chemical compound, is of significant interest to

industry.<sup>1–5</sup> The variation in physical properties such as crystal shape, solubility, hardness, color, melting point, and chemical reactivity makes polymorphism an important issue for the food, specialty chemical, and pharmaceutical industries, where products are specified not only by chemical composition but also by their performance.<sup>2</sup> Controlling polymorphism to ensure consistent production of the desired polymorph is important in those industries, including in drug manufacturing where safety is paramount. With the ultimate goal being to better understand the effects of process conditions on crystal quality and to control the formation of the desired polymorph, this article considers the estimation

Additional Supporting Information may be found in the online version of this article.

N. C. Kee is also affiliated with Dept. of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

R. B. H. Tan is also affiliated with Institute of Chemical and Engineering Sciences Singapore, Singapore 627833.

Correspondence concerning this article should be addressed to M.-S. Chiu at checms@nus.edu.sg

of kinetic parameters in a model for polymorphic crystallization. Such a process model can accelerate the determination of optimal operating conditions and to speed process development, when compared with time-consuming and expensive trial-and-error methods for determining the operating conditions.

In this article, a kinetic model of L-glutamic acid polymorphic crystallization is developed from batch experiments with in situ measurements including ATR-FTIR spectroscopy to infer the solute concentration and FBRM to provide crystal size information. Kinetics of polymorphic transformation have been estimated by various procedures.<sup>6–9</sup> A commonly used method to estimate model parameters in nonlinear process models is weighted least squares,<sup>10–12</sup> which has been applied to polymorphic crystallization.<sup>13–16</sup> Although weighted least squares methods are adequate for many problems, Bayesian inference is able to include prior knowledge in the statistical analysis, which can produce models with higher predictive capability. Although Bayesian inference is not within the standard toolkit for chemical engineers, there have been many applications to chemical engineering problems over the years including the estimation of parameters in chemical reaction,<sup>17</sup> heat transfer in packed beds,<sup>18</sup> microbial systems,<sup>19–21</sup> and microelectronics processes.<sup>22</sup>

Quantifying uncertainties in the parameter estimates is required for assessing the accuracy of model predictions.<sup>23,24</sup> When weighted least squares methods are used for parameter estimation, the widely used approaches to quantify uncertainties in parameter estimates are the linearized statistics and likelihood ratio approaches.<sup>25</sup> In the linearized statistics approach, the model is linearized around the optimal parameter estimates and the parameter uncertainty is represented by a  $\chi^2$  distribution. This model linearization can result in highly inaccurate uncertainty estimates for highly nonlinear models,<sup>25</sup> and this approach ignores physical constraints on the model parameters. The likelihood ratio approach, which is the nonlinear analogue to the well-known F statistic, takes nonlinearity into account but approximates the distribution<sup>25</sup> and ignores constraints on the model parameters. This article applies a Bayesian inference approach that not only avoids making these approximations but also includes prior information during the estimation of parameter uncertainties.

In this article, the parameters in a kinetic model for L-glutamic acid polymorphic crystallization process are determined by Bayesian estimation. The probability distribution over process model parameters is defined through the Bayesian posterior density, from which all parameter estimates of interest (e.g., means, modes, and credible intervals) are calculated. However, the conventional approach to calculate these estimates often involves complicated integrals of the Bayesian posterior density, which are analytically intractable. To overcome this drawback, Markov Chain Monte Carlo (MCMC) integration<sup>26–28</sup> was applied to compute these integrals in an efficient manner. MCMC does not require approximation of the posterior distribution by a Gaussian distribution.<sup>20,29,30</sup> This posterior distribution for the estimated parameters can be used to accurately quantify the accuracy of model predictions and can be incorporated into robust control strategies for crystallization process.<sup>24</sup>

This article is organized as follows. The next section describes the experimental procedure to obtain measurement data for parameter estimation. A short review of Bayesian

**Table 1. Glutamic Acid Aqueous Solutions Used for Calibration**

Concentration (g/g of water)	Temperature Range (°C)
0.00837	35–21
0.01301	48–13
0.01800	57–32
0.02300	64–34
0.02800	64–45

theory and MCMC integration is discussed next. This is followed by the description of the L-glutamic acid crystallization model and the results of the parameter estimation. Finally, conclusions are given.

## Experimental Methods

The crystallization instrument setup used was similar to that described previously.<sup>31</sup> A Dipper-210 ATR immersion probe (Axiom Analytical) with ZnSe as the internal reflectance element attached to a Nicolet Protege 460 FTIR spectrophotometer was used to obtain L-glutamic acid spectra in aqueous solution, with a spectral resolution of 4  $\text{cm}^{-1}$ . The chord length distribution (CLD) for L-glutamic crystals in solution were measured using Lasentec FBRM connected to a Pentium III running version 6.0b12 of the FBRM Control Interface software.

### Calibration for solution concentration

Different solution concentrations of L-glutamic acid (99%, Sigma Aldrich) and degassed deionized water were placed in a 500-ml jacketed round-bottom flask and heated until complete dissolution. The solution was then cooled at 0.5°C/min while the IR spectra were being collected, with continuous stirring in the flask using an overhead mixer at 250 rpm. Table 1 lists the five different solution concentrations used to build the calibration model.

The IR spectra of aqueous L-glutamic acid in the range 1100–1450  $\text{cm}^{-1}$  and the temperature were used to construct the calibration model based on various chemometrics methods such as principal component regression (PCR) and partial least square regression (PLS).<sup>32</sup> The calculations were carried out using inhouse MATLAB 5.3 (The Mathworks) code except for PLS, which was from the PLS Toolbox 2.0. The mean width of the prediction interval was used as the criterion to select the most accurate calibration model. The noise level was selected based on the compatibility of the prediction intervals with the accuracy of the solubility data. The chemometrics method forward selection PCR 2 (FPCR 2)<sup>33</sup> was selected because it gave the smallest prediction interval; using a noise level of 0.001, the prediction interval (0.73 g/kg) was compatible within the accuracy of this model with respect to solubility data reported in the literature.<sup>13</sup>

### Solubility determination and feedback concentration control experiments

The commercially available L-glutamic acid crystals were verified to be pure  $\beta$ -form using powder X-ray diffraction (XRD) and were used for the determination of the  $\beta$ -form solubility curve. Pure  $\alpha$ -form crystals obtained using a rapid

**Table 2. Solubility Data for L-Glutamic Acid Polymorphs**

Temperature (°C)	Solubility of $\alpha$ -Form (g/kg)	Solubility of $\beta$ -Form (g/kg)
25	10.5971	8.5434
30	13.1599	9.7362
35	15.8004	12.4257
40	19.1689	13.7163
45	23.1385	17.0729
50	27.0364	19.8722
55	31.7768	23.3904
60	36.8028	27.7567

cooling method outlined previously<sup>13</sup> were used to determine the  $\alpha$ -form solubility curve in similar fashion as the  $\beta$ -form in a separate experiment. For each polymorph, the IR spectra of L-glutamic acid slurries (saturated, and with excess crystals) were collected at different temperatures ranging from 25 to 60° C. The slurry was equilibrated for 45 min to 1 h at a specified temperature before recording the IR spectra. The solution concentration was then calculated using the aforementioned calibration model. The resulting solubility measurements for L-glutamic acid polymorphs are tabulated in Table 2, and Figure 1 compares the measurements to their quadratic polynomial fitting.

In the seeded batch crystallization experiments, appropriate amounts of L-glutamic acid (99%, Sigma Aldrich) in 400 g of water was heated to about 5° C above the  $\beta$ -form saturation temperature in a 500-mL jacketed round-bottom flask with an overhead mixer at 250 rpm to create an undersaturated solution. The crystallizer was then cooled and seed crystals (either pure  $\alpha$ - or  $\beta$ -form) were added when the solution was supersaturated with respect to the seeded form. Different supersaturation set-point profiles were followed during crystallization based on in situ solution concentration measurement as described previously.<sup>31</sup> The control algorithm was started shortly after seeding.

## Review of Bayesian Inference

### Bayesian posterior

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the results by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations.<sup>28</sup> The fundamental difference between Bayesian and traditional statistical methods is the interpretation of probability. Classical methods, also known as the frequentist methods, perceive probability as the long-run relative frequency of occurrence determined by the repetition of an event. A Bayesian method perceives probability as a quantitative description of the degree of belief in a given proposition.<sup>20,34</sup> With this interpretation of probability, the Bayesian method allows a practitioner to account for prior information in a statistical analysis.

Furthermore, Bayesian inference facilitates a common-sense interpretation of statistical conclusions. For instance, a Bayesian credible interval for an unknown quantity of interest can be directly regarded as having a high probability of containing the unknown quantity, in contrast to a frequentist confidence interval, which may strictly be interpreted only in relation to a sequence of similar inferences that might be made in repeated practice. A brief introduction to Bayesian inference is given later. Interested readers are referred to Refs. 28, 34 and 35 for a thorough discussion.

The main substance of Bayesian inference is Bayes' rule:

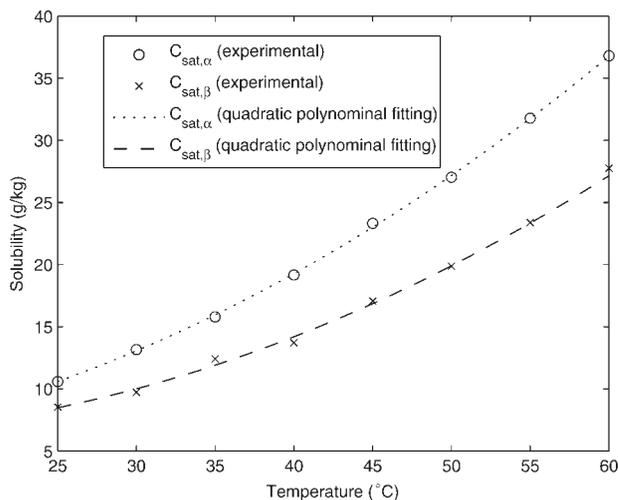
$$\Pr(\theta|\mathbf{y}) = \frac{\Pr(\mathbf{y}|\theta) \Pr(\theta)}{\Pr(\mathbf{y})}, \quad (1)$$

where  $\theta$  is a vector of unknown parameters of interest and  $\mathbf{y}$  represents the collected data which is used to infer  $\theta$ . These data usually consist of observed state variables (e.g., concentration) at different time points.  $\Pr(\theta)$  is the prior distribution of  $\theta$ ,  $\Pr(\mathbf{y}|\theta)$  is referred as the sampling distribution (or data distribution) for fixed parameters  $\theta$ . When the data  $\mathbf{y}$  are known and the parameters  $\theta$  are unknown (i.e., as in parameter estimation), the term  $\Pr(\mathbf{y}|\theta)$  is referred as the likelihood function and denoted as  $L(\theta|\mathbf{y})$ .  $\Pr(\theta|\mathbf{y})$  is referred as the Bayesian posterior distribution of  $\theta$ , and  $\Pr(\mathbf{y}) = \int \Pr(\mathbf{y}|\theta) \Pr(\theta) d\theta$  acts as a normalizing constant to ensure that the Bayesian posterior integrates to unity. This constant is also called marginal likelihood or Bayes factor. For the inference of  $\theta$ , the Bayes factor can be omitted since it does not affect the the resulting posterior distribution of  $\theta$ , which yields the unnormalized posterior distribution:

$$\Pr(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y}) \Pr(\theta). \quad (2)$$

In this article, it is assumed that the model structure is correct, and the measurement noise is distributed normally with zero mean and unknown variance. Then, the likelihood is of the form

$$\begin{aligned} L(\theta|\mathbf{y}) &= L(\theta_{\text{sys}}, \sigma|\mathbf{y}) = \prod_{j=1}^{N_m} \prod_{k=1}^{N_{d_j}} \Pr(y_{jk}|\theta_{\text{sys}}, \sigma) \\ &= \prod_{j=1}^{N_m} \prod_{k=1}^{N_{d_j}} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(y_{jk} - \hat{y}_{jk}(\theta_{\text{sys}}))^2}{2\sigma_j^2}\right) \\ &= \frac{1}{\prod_{j=1}^{N_m} (\sqrt{2\pi}\sigma_j)^{N_{d_j}}} \exp\left(-\sum_{j=1}^{N_m} \sum_{k=1}^{N_{d_j}} \frac{(y_{jk} - \hat{y}_{jk}(\theta_{\text{sys}}))^2}{2\sigma_j^2}\right), \end{aligned} \quad (3)$$



**Figure 1. Solubility curves of L-glutamic acid polymorphs.**

where  $\theta = [\theta_{\text{sys}}, \sigma]^T$  is the vector of parameters of interest, which consist of the system/model ( $\theta_{\text{sys}}$ ) and noise ( $\sigma$ ) parameters,  $y_{jk}$  and  $\hat{y}_{jk}$  are the measurement and predicted value of  $j$ th variable at sampling instance  $k$ , respectively,  $N_m$  is the number of measured variables,  $N_{d_j}$  is the number of time samples of  $j$ th variable, and  $\sigma_j$  is the standard deviation of the measurement noise in the  $j$ th variable.

The prior distribution  $\text{Pr}(\theta)$  can be informative or noninformative depending on the prior knowledge of  $\theta$ . The most commonly used noninformative prior is  $\text{Pr}(\theta) \propto 1$ . However, this is an improper prior distribution, since its integral is infinity, and may lead to an improper posterior distribution. The use of an informative prior distribution is preferred, for example, a prior distribution which specifies the minimum and maximum possible values of  $\theta$  is

$$\text{Pr}(\theta) \propto \begin{cases} 1 & \text{if } \theta_{\min} \leq \theta \leq \theta_{\max} \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

which means that all values of  $\theta$  between  $\theta_{\min}$  and  $\theta_{\max}$  have equal probability. In cases where the prior distribution is available from past parameter estimation studies, the distribution is not uniform.<sup>22</sup> A detailed discussion regarding informative and noninformative priors can be found in the literature.<sup>28,35,36</sup>

The product of the likelihood and prior distribution defines the Bayesian posterior, which is the joint probability distribution for all parameters after data have been observed. Once the Bayesian posterior is defined, it is desirable to determine the mean, mode, and credible intervals associated with each of the parameters. Markov chain simulation, also called MCMC, is employed for that purpose in this article.

### Markov chain simulation

Markov chain simulation draws values of  $\theta$  from approximate distributions and then corrects these values to better approximate the target distribution. In this case, the target distribution is the Bayesian posterior. The samples are drawn sequentially, with the distribution of the sampled values depending on the last value drawn. The Markov chain is a sequence of random variables  $\theta^0, \theta^1, \dots$ , for which, for any  $s$ , the distribution of  $\theta^{s+1}$  given all previous  $\theta^s$  depends only on the most recent value,  $\theta^s$ . The key to the method's success, however, is not the Markov property but rather that the approximate distributions are improved at each step in the simulation, in the sense of converging to the target distribution.\*

In the application of Markov chain simulation, several parallel chains are drawn. Parameters from each chain  $c$ ,  $\theta^{c,s}$ ,  $s = 1, 2, 3, \dots$ , are produced by starting at some point  $\theta^{c,0}$  and then, for each step  $s$ , drawing  $\theta^{c,s+1}$  from a jumping distribution,  $T_s(\theta^{c,s+1}|\theta^{c,s})$  that depends on the previous draw,  $\theta^{c,s}$ . The jumping probability distributions must be constructed so that the Markov chain converges to the target posterior distribution.

The Metropolis algorithm<sup>37</sup> is a simple algorithm to construct a Markov chain, which converges to the posterior distribution. The algorithm is an adaptation of a random walk that uses an acceptance/rejection rule to converge to the specified target distribution. In the Metropolis algorithm, the widely

used approach to create the next step of the chain  $c$ ,  $\theta^{c,s_p}$ , is to perturb the current step of the chain  $\theta^{c,s}$  by adding some amount of noise ( $\theta^{c,s_p} = \theta^{c,s} + \varepsilon$ ), where  $\varepsilon$  is distributed normally with zero mean and covariance matrix  $\Sigma$ . However, specifying the covariance matrix can be challenging. This covariance matrix needs to be chosen in such a way so as to balance progress in each step and a reasonable acceptance rate. A poorly chosen covariance matrix may cause slow convergence. Traditionally, the covariance matrix is estimated from a trial run and much recent research is devoted to ways of doing that efficiently and/or adaptively.<sup>38</sup> If parameters  $\theta$  are highly correlated, special precautions must be taken to avoid singularity of the estimated covariance matrix.

Recently, there has been a development in combining evolutionary algorithms with MCMC.<sup>39-42</sup> Among others, the combination of differential evolution (DE) with MCMC is particularly interesting. DE is an evolutionary algorithm for numerical optimization; its combination with MCMC (shortened as DE-MC<sup>42</sup>) solves an important problem in MCMC, namely that of choosing an appropriate scale and orientation for the jumping distribution (i.e., related to the covariance matrix  $\Sigma$  in the Metropolis algorithm). In DE-MC, the jumps are simply a fixed multiple of two random parameter vectors that are currently in the population, and the selection process of DE-MC works via the usual Metropolis ratio that defines the probability with which a proposal is accepted. Motivated by its efficiency and effectiveness, DE-MC is utilized to construct the Markov chains of  $\theta$  in this article.

Constructing the Markov chains is one step. Next is to monitor the convergence of the chains to decide how many samples need to be collected or when to stop the MCMC simulation. Too few samples will result in an inaccurate distribution of the parameters  $\theta$ . Here, potential scale reduction factors ( $\hat{R}_i$ ) were adopted to monitor the convergence of the Markov chains,<sup>28</sup> which estimate the potential improvement in the Markov chain estimation of the respective  $i$ th parameter  $\theta_i$  if the Markov chain simulation were continued. This potential scale reduction factor is calculated from the following equations:

$$\hat{R}_i = \sqrt{\frac{\hat{\text{var}}^+(\theta|y)_i}{W_i}}, \quad (5)$$

$$\hat{\text{var}}^+(\theta|y)_i = \frac{n-1}{n} W_i + \frac{1}{n} B_i, \quad (6)$$

$$B_i = \frac{n}{m-1} \sum_{s=1}^m (\bar{\theta}_i^s - \bar{\theta}_i)^2, \quad (7)$$

$$W_i = \frac{1}{m} \sum_{s=1}^m (d_i^s)^2, \quad (8)$$

$$\bar{\theta}_i^s = \frac{1}{n} \sum_{c=1}^n \theta_i^{c,s}, \quad (9)$$

$$\bar{\theta}_i = \frac{1}{m} \sum_{s=1}^m \bar{\theta}_i^s, \quad (10)$$

$$(d_i^s)^2 = \frac{1}{n-1} \sum_{c=1}^n (\theta_i^{c,s} - \bar{\theta}_i^s)^2, \quad (11)$$

\* For further information on Markov chains, readers are referred to other literature.<sup>27,28</sup>

where  $\theta_i^{c,s}$  is the simulation draws of parameter  $i$  from step chain  $c$  at step  $s$ ,  $B_i$  and  $W_i$  are the between- and within-sequence variances of parameter  $i$ , respectively,  $m$  is the number of parallel chains, with each chain of length  $n$ . The potential scale reduction factor decreases asymptotically to 1 as  $n \rightarrow \infty$ . Once  $\hat{R}_i$  is near  $1^\dagger$  for all  $i$ , it is safe to stop the simulation.

To summarize, the following is the procedure for constructing Markov chains using DE-MC with the potential scale reduction factor as the stopping criterion:

(1) Draw starting parameters for all chains,  $\theta^{c,0}$  ( $c = 1, \dots, m$ ), from a starting distribution or choose starting parameters from dispersed values around a crude approximation of the parameters.

(2) At each step, create a proposed value  $\theta^{c,s_p}$  according to the jumping rule

$$\theta^{c,s_p} = \theta^{c,s} + \gamma(\theta^{R_{1,s}} - \theta^{R_{2,s}}) + \mathbf{e}, \quad (12)$$

where  $\mathbf{e}$  is drawn from a symmetric distribution with a small variance compared to that of the target, but with unbounded support (e.g.,  $\mathbf{e} \sim N(0, b)^{N_\theta}$  with  $b$  small,  $b = 10^{-4}$  is utilized in this article),  $N_\theta$  is the number of parameters in  $\theta$ ,  $\theta^{R_{1,s}}$  and  $\theta^{R_{2,s}}$  are randomly selected without replacement from all chains at step  $s$ , and  $\gamma$  is a scaling constant with typical values between 0.4 and 1. From the guidelines in the literature,<sup>42</sup> the optimal choice of  $\gamma$  is  $2.38/\sqrt{2N_\theta}$ . This choice of  $\gamma$  is expected to give an acceptance probability of 0.44 for  $N_\theta = 1$ , 0.28 for  $N_\theta = 5$ , and 0.23 for large  $N_\theta$ .

(3) Calculate the ratio of the posterior densities,

$$r = \frac{\Pr(\theta^{c,s_p} | \mathbf{y})}{\Pr(\theta^{c,s} | \mathbf{y})}. \quad (13)$$

and obtain  $\theta^{c,s+1}$  from

$$\theta^{c,s+1} = \begin{cases} \theta^{c,s_p} & \text{with probability } \min\{r, 1\} \\ \theta^{c,s} & \text{otherwise.} \end{cases} \quad (14)$$

(4) For each parameter  $i$ , calculate the potential scale reduction factor  $\hat{R}_i$  by (5) to (11). If  $\hat{R}_i \leq 1.1$  for all  $i = 1, 2, \dots, N_\theta$ , stop the iteration and construct the matrix

$$\Theta = \begin{bmatrix} \theta_1^1 & \dots & \theta_{N_\theta}^1 \\ \vdots & \ddots & \vdots \\ \theta_1^{N_s} & \dots & \theta_{N_\theta}^{N_s} \end{bmatrix}, \quad (15)$$

where  $\Theta$  contains the approximated samples from the target distribution and  $N_s$  is the total number of values drawn from the second halves for all the chains.

Otherwise, if  $\hat{R}_i > 1.1$  for any  $i$ , set  $s = s + 1$  and go to Step 2.

### Monte Carlo integration

In the previous sections, the Bayes posterior was defined and a method for drawing samples from it was described, from which a matrix  $\Theta$  was generated. Here, the significance of this matrix is described through its use for calculating the desired properties of the Bayes posterior.

To calculate any properties of the Bayes posterior, it is necessary to evaluate integral

$$E[f(\theta)] = \int_{\theta_{\min}}^{\theta_{\max}} f(\theta) \Pr(\theta | \mathbf{y}) d\theta, \quad (16)$$

where  $E[\cdot]$  is the expected value and,  $f(\theta)$  is a function for which the expected value is to be estimated. Conventionally, this integration can be performed analytically if the resulting function inside the integral operator is simple. However, the Bayesian posteriors most often have irregular forms such that analytical integrations become infeasible. In such situations, it is suitable to perform Monte Carlo integration,<sup>26–28</sup> which utilizes the matrix  $\Theta$  obtained in the previous section:

$$E[f(\theta)] = \lim_{N_s \rightarrow \infty} \frac{1}{N_s} \sum_{l=1}^{N_s} f(\theta^l) \approx \frac{1}{N_s} \sum_{l=1}^{N_s} f(\theta^l) \text{ for large } N_s, \quad (17)$$

where  $\theta^l = [\theta_1^l, \theta_2^l, \dots, \theta_{N_\theta}^l]$  is a random sample drawn from the Bayesian posterior, which is obtained from the  $l^{\text{th}}$  row of matrix  $\Theta$ . For example, the mean of each parameter  $\theta_i$  is obtained by setting  $f(\theta^l) = \theta^l$  in (17).

It is also desirable to obtain the marginal mode and credible interval for each parameter. Conventionally, this is done by drawing samples from the marginal posterior for each parameter and analyzing their histograms, where the marginal posterior is calculated by integrating the Bayes posterior with respect to all parameters except the desired parameter as follows

$$\Pr(\theta_i | \mathbf{y}) = \int_{\theta_{1,\min}}^{\theta_{1,\max}} \dots \int_{\theta_{j,\min}}^{\theta_{j,\max}} \dots \int_{\theta_{N_\theta,\min}}^{\theta_{N_\theta,\max}} \Pr(\theta_1, \dots, \theta_j, \dots, \theta_{N_\theta} | \mathbf{y}) d\theta_1 \dots d\theta_j \dots d\theta_{N_\theta}, \quad (18)$$

where  $j \neq i$  and  $\Pr(\theta_j | \mathbf{y})$  is the marginal posterior of  $\theta_j$ . By taking advantage of the MCMC approach, this integration is not required since the samples from the marginal posterior of  $\theta_i$  are given by the  $i$ th column of the matrix  $\Theta$ . The marginal mode of  $\theta_i$  was estimated by determining the highest peak in the histograms of the marginal posterior. Finally, the 95% credible interval of  $\theta_i$  was estimated by determining the range of  $\theta_i$ , which have cumulative marginal distribution between 2.5 and 97.5%.

### L-Glutamic Acid Crystallization Model

A kinetic model for the crystallization of metastable  $\alpha$ -form and stable  $\beta$ -form crystals of L-Glutamic acid is developed. This appears to be the first model for polymorphic crystallization that includes all of the kinetic processes and also includes their dependence on the temperature. An earlier model for this system did not include the nucleation and growth kinetics of  $\alpha$ -form crystals.<sup>13</sup> An improved model that includes those kinetics<sup>14</sup> only considered primary hetero-

<sup>†</sup> According to Gelman et al.,<sup>28</sup> a value below 1.1 is acceptable.

geneous nucleation, which only applies when the crystallization is either starved with nuclei or overwhelmed by a burst of new crystals, and hence not applicable to industrial practice.<sup>43</sup> To develop a model amenable for industrial application, secondary nucleation is considered in this article.

### Kinetic model

The mass balance on the crystals is described by a population balance equation<sup>44</sup>

$$\frac{\partial f_i}{\partial t} + \frac{\partial(G_i f_i)}{\partial L} = B_i \delta(L - L_0), i = \alpha, \beta \quad (19)$$

where  $f_i$  is the crystal size distribution of the  $i$ -form crystals ( $\#/m^4$ ) (i.e.,  $\alpha$ - or  $\beta$ -form crystals),  $B_i$  and  $G_i$  are the nucleation ( $\#/m^3s$ ) and growth rate (m/s) of the  $i$ -form crystals, respectively,  $L$  and  $L_0$  are the characteristic size of crystals (m) and nuclei (m), respectively, and  $\delta(\cdot)$  is a Dirac delta function.

For parameter estimation, the method of moments<sup>45</sup> was applied to (19) to give

$$\frac{d\mu_{i,0}}{dt} = B_i, \quad (20)$$

$$\frac{d\mu_{i,n}}{dt} = nG_i \mu_{i,n-1} + B_i L_0^n, \quad n = 1, 2, \dots, \quad (21)$$

where the  $n$ th moment of the  $i$ -form crystals ( $\#m^{n-3}$ ) is given by

$$\mu_{i,n} = \int_0^\infty L^n f_i dL. \quad (22)$$

These equations are augmented by the solute mass balance:

$$\frac{dC}{dt} = -3 \frac{10^3}{\rho_{\text{solv}}} (\rho_\alpha k_{v\alpha} G_\alpha \mu_{\alpha,2} + \rho_\beta k_{v\beta} G_\beta \mu_{\beta,2}), \quad (23)$$

where  $C$  is the solute concentration (g/kg),  $\rho_{\text{solv}}$  is the density of the solvent (kg/m<sup>3</sup>),  $\rho_i$  is the density of the  $i$ -form crystals (kg/m<sup>3</sup>),  $k_{vi}$  is the volumetric shape factor of the  $i$ -form crystals (dimensionless) as defined by  $v_i = k_{vi} L^3$ , where  $v_i$  is the volume of the  $i$ -form crystal (m<sup>3</sup>), and  $10^3$  is a constant (g/kg) to ensure unit consistency. The kinetic expressions are

$$B_\alpha = k_{b\alpha} (S_\alpha - 1) \mu_{\alpha,3} \quad (\alpha\text{-form crystal nucleation rate}), \quad (24)$$

$$G_\alpha = \begin{cases} k_{g\alpha} (S_\alpha - 1)^{g_\alpha} & \text{if } S_\alpha \geq 1 \\ k_{d\alpha} (S_\alpha - 1) & \text{otherwise} \end{cases} \quad (\alpha\text{-form crystal growth/dissolution rate}), \quad (25)$$

<sup>†</sup>The approach applies for the experimental conditions in this study in which data were collected during nucleation and growth. The full population balance Equation (19) is used under conditions in which dissolution occurs.

**Table 3. Values for Densities, Volume Shape Factors, and Saturation Concentration Parameters**

Parameters	Values
$\rho_{\text{solv}}$	990
$\rho_\alpha$	1540
$\rho_\beta$	1540
$k_{v\alpha}$	0.480
$k_{v\beta}$	0.031
$a_{\alpha,1}$	$8.437 \times 10^{-3}$
$a_{\alpha,2}$	0.03032
$a_{\alpha,3}$	4.564
$a_{\beta,1}$	$7.644 \times 10^{-3}$
$a_{\beta,2}$	-0.1165
$a_{\beta,3}$	6.622

$$B_\beta = k_{b\beta,1} (S_\beta - 1) \mu_{\alpha,3} + k_{b\beta,2} (S_\beta - 1) \mu_{\beta,3} \quad (\beta\text{-form crystal nucleation rate}), \quad (26)$$

$$G_\beta = k_{g\beta,1} (S_\beta - 1)^{g_\beta} \exp\left(\frac{-k_{g\beta,2}}{S_\beta - 1}\right) \quad (\beta\text{-form crystal growth rate}), \quad (27)$$

where  $S_i = C/C_{\text{sat},i}$  and  $C_{\text{sat},i} = a_{i,1}T^2 + a_{i,2}T + a_{i,3}$  are the supersaturation and the saturation concentration (g/kg) of the  $i$ -form crystals, respectively, and  $T$  is the solution temperature ( $^\circ\text{C}$ ). The kinetic parameters  $k_{b\alpha}$ ,  $k_{g\alpha}$ , and  $k_{d\alpha}$  correspond to the nucleation ( $\#/m^3s$ ), growth (m/s), and dissolution (m/s) rates of  $\alpha$ -form crystals, respectively, whereas  $k_{b\beta,j}$  and  $k_{g\beta,j}$  correspond to the  $j$ th nucleation ( $\#/m^3s$ ) and growth (m/s) for  $j = 1$  and dimensionless for  $j = 2$  rates of  $\beta$ -form crystals, respectively, and  $g_i$  is the growth exponent of the  $i$ -form crystals, which may have a value between 1 (for diffusion-limited growth) and 2 (for surface integration-limited growth).<sup>45</sup> The Arrhenius equation was used to account for the variability of crystal growth rate with temperature:

$$k_{g\alpha} = k_{g\alpha,0} \exp\left(-\frac{E_{g\alpha}}{8.314(T + 273)}\right), \quad (28)$$

$$k_{g\beta,1} = k_{g\beta,0} \exp\left(-\frac{E_{g\beta}}{8.314(T + 273)}\right), \quad (29)$$

where  $k_{gi,0}$  and  $E_{gi}$  are the pre-exponential factor (m/s) and activation energy (J/mol) for the growth rate of  $i$ -form crystals, respectively. The values for densities, volumetric shape factors, and parameters for the saturation concentration are given in Table 3.

Secondary nucleation is assumed for both  $\alpha$ - and  $\beta$ -form crystals, since it is the dominant nucleation process in seeded crystallization. Primary nucleation is not included in the model since it is negligible compared to the secondary nucleation. The nucleation rate expression (24) and the second term in (26) were adapted from that reported in the literature for  $\beta$  crystals for L-glutamic acid.<sup>13</sup> We have introduced the first term in (26) to model the nucleation of  $\beta$ -form crystals from the surface of  $\alpha$ -form crystals. The growth rate expression for the  $\alpha$ -form crystals includes both growth (positive supersaturation) and dissolution (undersaturation). Dissolution occurs during the polymorphic transformation of  $\alpha$ - to  $\beta$ -form crystals, where  $\alpha$ -form crystals dissolve and  $\beta$ -

**Table 4. Seed Crystal Size Distribution Data and  $\alpha$ -Form Crystal Purity at the End of Batch ( $x_\alpha$ )**

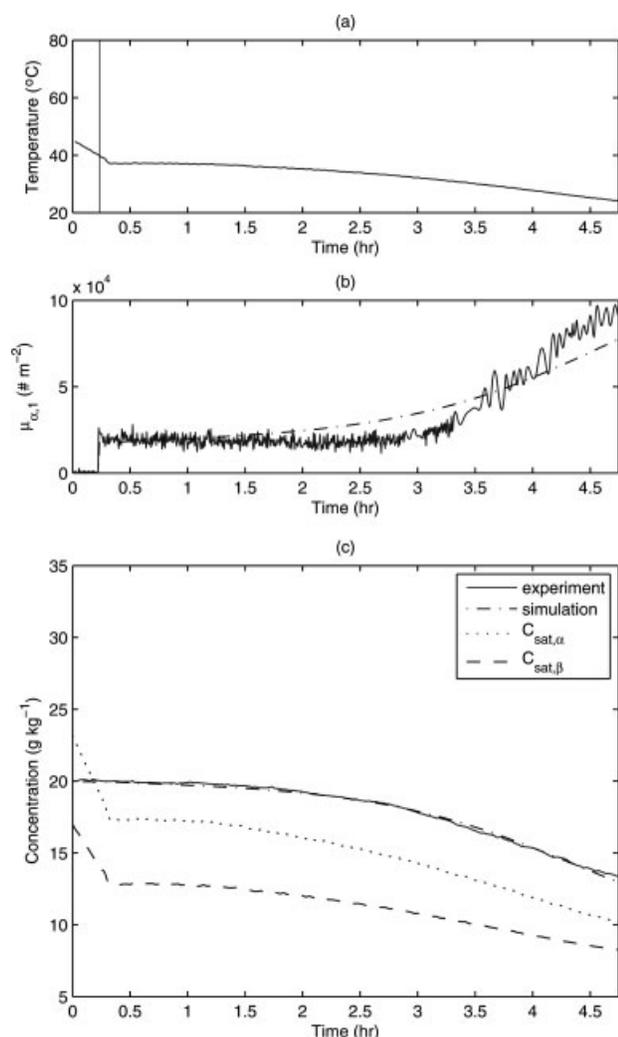
No.	Seed	Size ( $\mu\text{m}$ )	Mass (g/kg)	$\lambda_i$	$\sigma_{\text{seed},i}$ ( $\times 10^6\text{m}$ )	$\mu_{\text{seed},i}$ ( $\times 10^6\text{m}$ )	$x_\alpha$
1	$\alpha$	180–250	0.613	$8.227 \times 10^7$	8.608	214.977	$\approx 1.000$
2	$\alpha$	75–180	0.613	$3.877 \times 10^8$	12.127	127.269	$\approx 1.000$
3	$\alpha$	75–180	0.592	$3.731 \times 10^8$	12.115	127.427	0.924
4	$\beta$	40–270	4.900	$2.483 \times 10^{10}$	27.289	155.069	$\approx 0.000$
5	$\beta$	40–270	3.225	$1.630 \times 10^{10}$	27.989	155.017	$\approx 0.000$
6	$\beta$	40–270	2.972	$1.501 \times 10^{10}$	28.131	155.004	$\approx 0.000$

form crystals nucleate and grow. As reported in the literature,<sup>13,14</sup> the dissolution kinetics cannot be estimated accurately from polymorphic transformation experiments, as the growth rate of  $\beta$ -form crystals is limiting. Thus the simple form of dissolution rate with exponential factor of 1 was used with  $k_{d\alpha}$  determined by a correlation equation based on mass transfer-limited dissolution, as reported in the literature.<sup>14</sup> The growth rate expressions for both  $\alpha$ - and  $\beta$ -form crystals are also adopted from the literature,<sup>46</sup> except that the

exponential term for the  $\alpha$ -form crystals is omitted in this article as it had a negligible effect on the model fitness to the data.

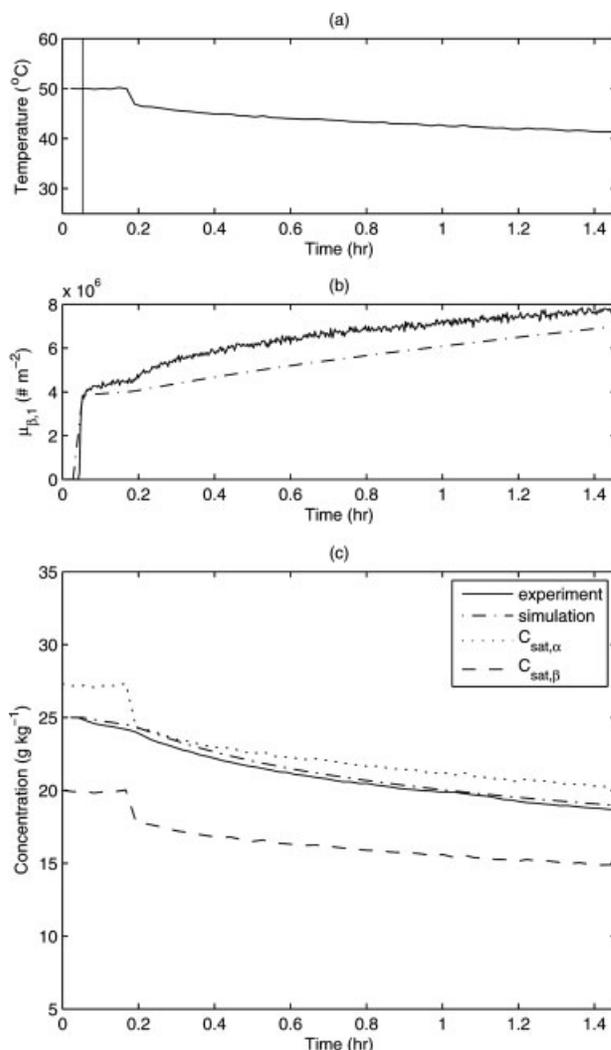
### Parameter estimation

Before parameter estimation is carried out, the measured variables are discussed first. The various in situ sensors that



**Figure 2. Experimental and model trajectories for (a) temperature, (b) the first-order moment of the  $\alpha$ -form crystals, and (c) solute concentration for Experiment 1 of Table 4.**

The vertical line in plot (a) shows the seeding time.



**Figure 3. Experimental and model trajectories for (a) temperature, (b) the first-order moment of the  $\beta$ -form crystals, and (c) solute concentration for Experiment 4 of Table 4.**

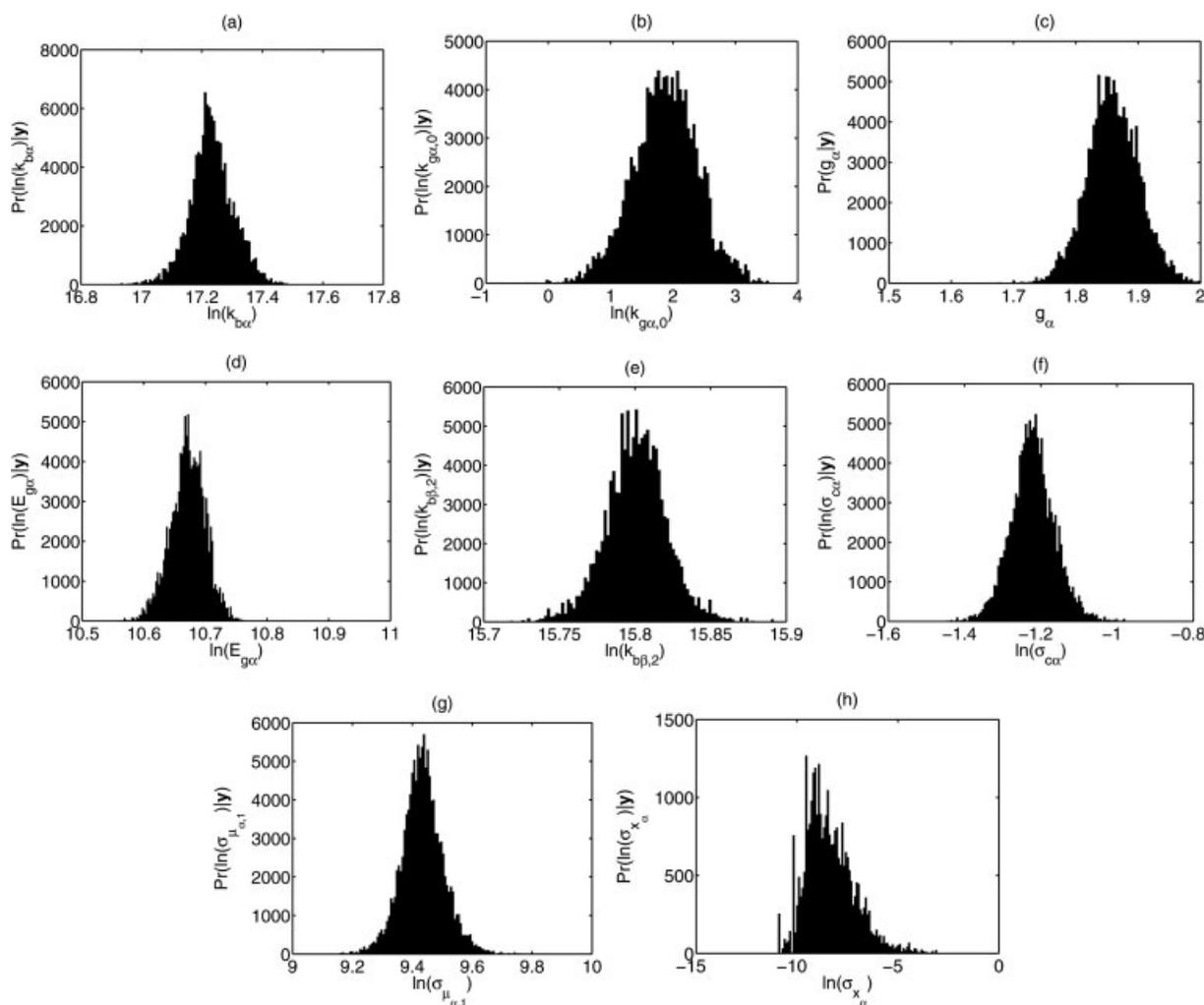
The vertical line in plot (a) shows the seeding time.

**Table 5. Definition of Measured Variables  $y$  and Interested Parameters  $\theta$  for  $\alpha$ - and  $\beta$ -Seeded Experiments**

Seed	$\theta^T$	$y^T$
$\alpha$	$[\ln(k_{bz}), \ln(k_{gz,0}), g_{\alpha}, \ln(E_{gz}), \ln(k_{b\beta,1}), \ln(\sigma_{c\alpha}), \ln(\sigma_{\mu\alpha,1}), \ln(\sigma_{xz})]$	$[C, \mu_{\alpha,1}, x_{\alpha}]$
$\beta$	$[\ln(k_{b\beta,2}), \ln(k_{g\beta,0}), \ln(k_{g\beta,2}), g_{\beta}, \ln(E_{g\beta}), \ln(\sigma_{c\beta}), \ln(\sigma_{\mu\beta,1})]$	$[C, \mu_{\beta,1}]$

have become available for crystallization processes have removed or reduced sampling of the crystal slurry during crystallization and reduced the amount of pharmaceutical needed for each batch experiment. The two in situ measurements utilized in this study were ATR-FTIR spectroscopy, which infers the solute concentration and FBRM that provides crystal size information throughout the batch. Inferential modeling was used to construct a calibration curve to relate the FTIR spectra to the solute concentration, using procedures described elsewhere.<sup>47,48</sup> FBRM measures the chord length distribution (CLD), which is not the same as the crystal size distribution (CSD) that appears in the models in the previous section.

The CSD can be computed from the CLD under certain assumptions.<sup>49-52</sup> For some systems, the square-weighted chord length was found to be comparable to laser diffraction, sieving, and electrical sensing zone analysis over the range of 50–400  $\mu\text{m}$ .<sup>53</sup> Although the aforementioned methods are able to estimate the CSD from CLD successfully for some systems, the theory behind these methods require many assumptions, including that the particles perfectly backscatter light at all angles and that shape of the crystals is known. Although these assumptions are true for many particulate systems (such as round polymer beads with a rough surface in water at low-to-moderate solids densities<sup>52</sup>), the assumptions are not accurate for other particulate systems including the system studied here which has crystals with a similar refractive index as the solution (and hence poor backscattering properties). Because of the limited time and pharmaceutical quantity available in the early stage of batch crystallization design, it is typically not possible to carry out the extensive studies to verify the assumptions and to determine the effects of nonideality of the assumptions on the accuracy of the estimates of the CSD from the CLD. Furthermore, computing the CSD from the CLD when assumptions such as perfect laser backscattering do not hold is still an open problem.<sup>51,54</sup>



**Figure 4. The marginal distributions of parameters  $\theta$  obtained from  $\alpha$ -seeded experiments (Table 5).**

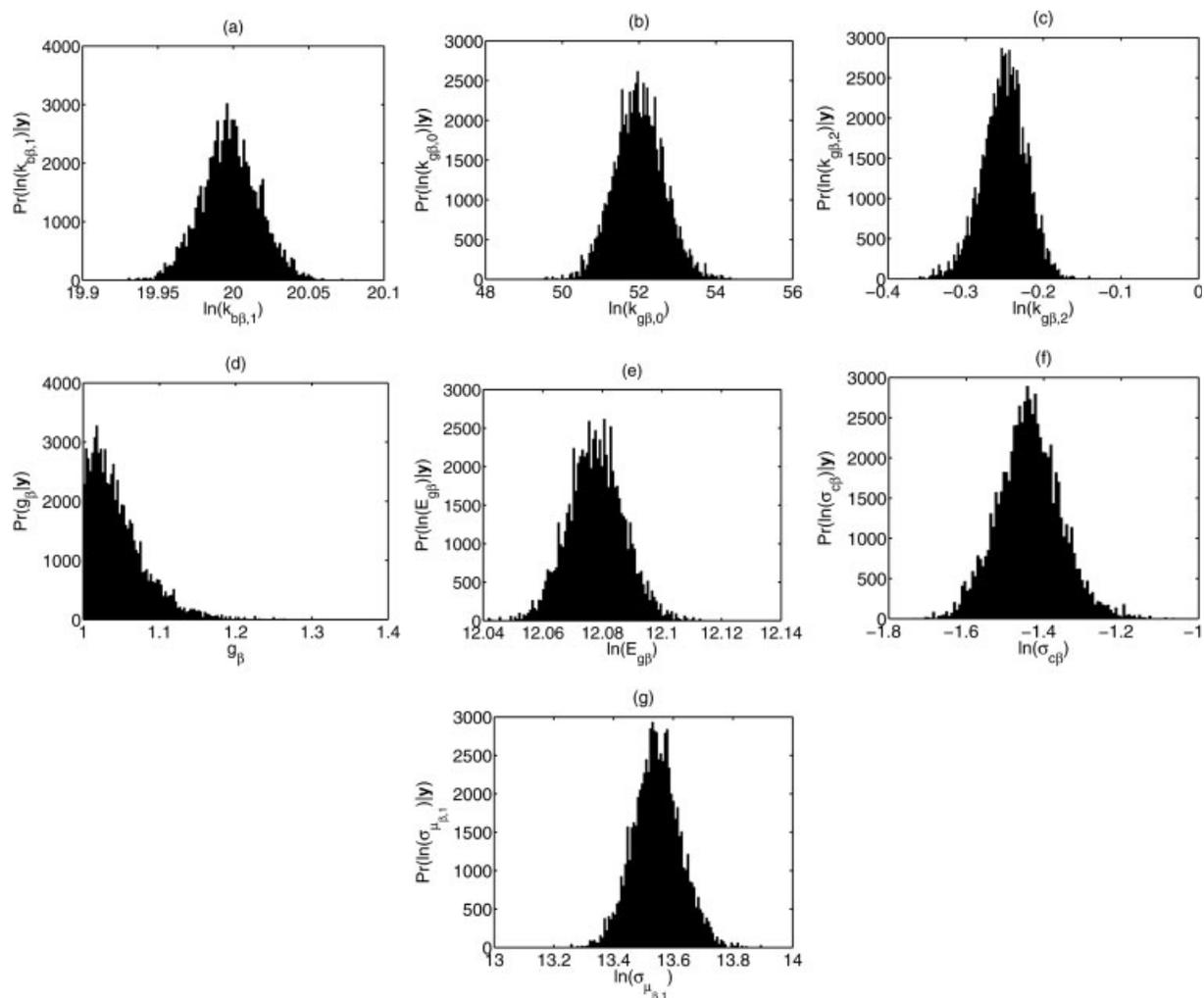


Figure 5. The marginal distributions of parameters  $\theta$  obtained from  $\beta$ -seeded experiments (Table 5).

An alternative approach is to use the low-order moments of the CLD directly,<sup>54,55</sup> without first estimating the CSD from the CLD. This approach replaces the first-principles model for the CSD with a gray-box model for the CLD, in which the structure of the first-principles model for the low-order moments of the CSD is used to parametrize the low-order moments for the CLD.<sup>54</sup> The reasoning behind this particular gray-box model is that the mapping between the CLD and the CSD is static (most of the aforementioned mapping methods assume that the mapping is actually linear), so the low-order moments of the CLD should follow the same dynamic trends as the low-order moments of the CSD. Because of the limitation of the FBRM precision, the zeroth moment was not used because FBRM would undercount the very small crystals. On the other hand, it is not advisable to use moments with order higher than two because higher order moments are sensitive to low-sampling statistics of the large crystals.<sup>55</sup> In this study, the first-order moment was used. As with any model,<sup>25</sup> this article assesses the applicability of this gray-box modeling by quantifying the accuracy of the kinetic parameters and the model's predictions.

The experiments are categorized into two sets, namely,  $\alpha$ -seeded and  $\beta$ -seeded experiments. The seed crystal size

distribution was approximated as a normal distribution

$$f_i(L, 0) = f_{\text{seed},i}(L) = \frac{\lambda_i}{\sqrt{2\pi}\sigma_{\text{seed},i}} \exp\left(-\frac{(L - \mu_{\text{seed},i})^2}{2\sigma_{\text{seed},i}^2}\right), \quad (30)$$

with the parameters ( $\lambda_i$ ,  $\sigma_{\text{seed},i}$ , and  $\mu_{\text{seed},i}$ ) in Table 4. The time series for the temperature, first-order moment of the  $i$ -

Table 6. The Model Parameters Determined from Parameter Estimation

Parameters	Mean	Mode	95% Credible Interval
$\ln(k_{bz})$	17.233	17.213	17.083–17.377
$\ln(k_{gz,0})$	1.878	1.778	0.801–2.912
$g_z$	1.859	1.860	1.775–1.944
$\ln(E_{gz})$	10.671	10.671	10.612–10.725
$\ln(k_{b\beta,1})$	15.801	15.796	15.758–15.842
$\ln(k_{b\beta,2})$	20.000	20.000	19.961–20.036
$\ln(k_{g\beta,0})$	52.002	52.426	50.745–53.322
$\ln(k_{g\beta,2})$	-0.251	-0.251	-0.311–-0.197
$g_\beta$	1.047	1.016	1.002–1.143
$\ln(E_{g\beta})$	12.078	12.076	12.060–12.097

**Table 7. Seed Crystal Size Distribution Data and  $\alpha$ -Form Crystals Purity at the End of Batch ( $x_\alpha$ ) for Model Validation**

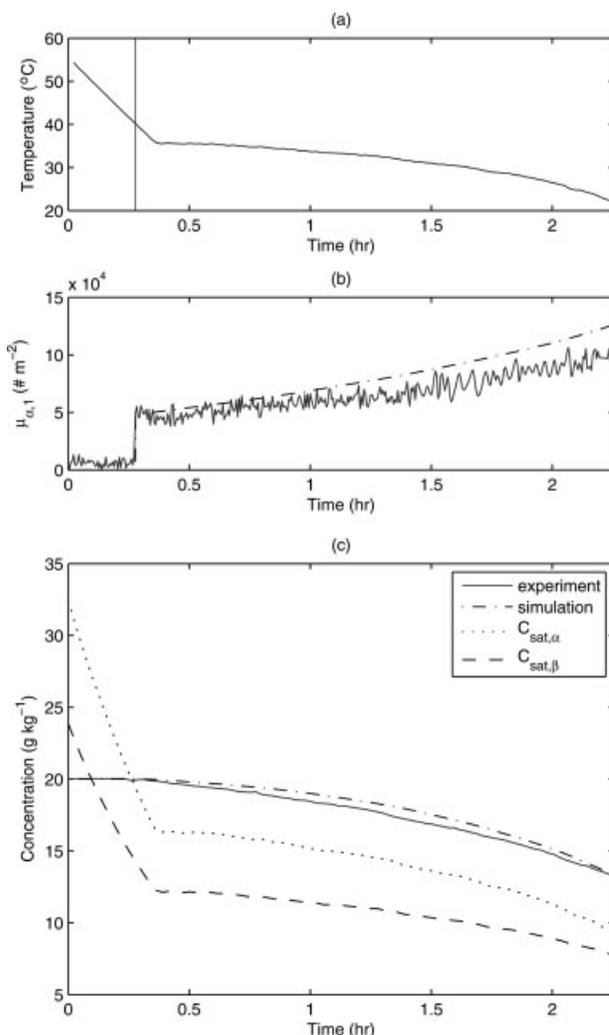
No.	Seed	Size ( $\mu\text{m}$ )	Mass (g/kg)	$\lambda_i$	$\sigma_{\text{seed},i} (\times 10^6\text{m})$	$\mu_{\text{seed},i} (\times 10^6\text{m})$	$x_\alpha$
V1	$\alpha$	75–180	0.613	$3.877 \times 10^8$	12.127	127.269	$\approx 1.000$
V2	$\beta$	40–270	3.060	$1.547 \times 10^{10}$	28.081	154.978	$\approx 0.000$

form crystals, and solute concentration for Experiments 1 and 4 are shown by the solid lines in Figures 2 and 3.<sup>8</sup> For all the  $\beta$ -seeded experiments, there is no apparent formation of  $\alpha$ -form crystals at the end of all batches (Table 4).<sup>1</sup> As a result, the kinetic parameters for  $\beta$ -form crystals were independently obtained from the  $\beta$ -seeded experiments, except for  $k_{b\beta,1}$ , which accounts for the nucleation of  $\beta$ -form  $\alpha$ -form crystals. One  $\alpha$ -seeded experiment was operated at a high enough temperature that a measurable quantity of  $\beta$ -form crystals nucleated and grew (Experiment 3 in Table 4), so there would be enough information content in the data for  $k_{b\beta,1}$  to be estimated. This experimental design enabled the kinetic parameters for  $\beta$ -form crystals to be obtained before determining the kinetic parameters for  $\alpha$ -form crystals.

The nucleation and growth kinetics of  $\alpha$  and  $\beta$ -form crystals have 10 parameters to be estimated, four ( $k_{b\alpha}$ ,  $k_{g\alpha,0}$ ,  $g_\alpha$ ,  $E_{g\alpha}$ ) corresponding to the kinetics of  $\alpha$ -form crystals and six ( $k_{b\beta,1}$ ,  $k_{b\beta,2}$ ,  $k_{g\beta,0}$ ,  $k_{g\beta,2}$ ,  $g_\beta$ ,  $E_{g\beta}$ ) corresponding to the kinetics of  $\beta$ -form crystals. In relation to the notation defined in the Review of Bayesian Inference section, the measured variables  $y$  and parameters of interest  $\theta$  for each set of experiments are defined in Table 5, where  $\sigma_{c_i}$ ,  $\sigma_{\mu_{i,1}}$ ,  $\sigma_{x_i}$  are the noise parameters for the  $i$ -form crystals. The prior distribution  $\text{Pr}(\theta)$  came from a preliminary parameter estimation that was carried out using maximum likelihood techniques as described in Miller and Rawlings,<sup>23</sup> which resulted in a normal distribution for each parameter. These were modified for  $g_\alpha$  and  $g_\beta$  according to (4) to limit their values between 1 and 2. The resulting marginal probability distributions of  $\theta$  from  $\alpha$ - and  $\beta$ -seeded experiments are in Figures 4 and 5, respectively. While some of the marginal probability distributions could be approximated by a normal distribution, others are not. These distributions can be directly inserted into those model predictive control and other control algorithms that have been designed to ensure robustness to stochastic parameter uncertainties.<sup>24</sup> The means, modes, and 95% credible intervals for the model parameters based on their marginal probability distributions are in Table 6. Figures 2 and 3 compare the temperature, first-order moment of the  $\alpha$ -form crystals, and solute concentration trajectories obtained from experimental data and those predicted through simulation using the aforementioned mean values as the model parameters.

It is well known that concentration data alone are not sufficient to characterize nucleation.<sup>23</sup> The small uncertainties in the nucleation kinetic parameters indicate that the first-order moment of the FBRM provided enough information to characterize the nucleation kinetics. The small range in the uncertainties for the activation energies indicates that the temperature range from 24 to 55°C in the experiments was large enough to enable activation energies to be estimated. The

rather large uncertainty in  $k_{g\alpha,0}$  is mainly due to the large correlation coefficient of 0.993 between  $k_{g\alpha,0}$  and  $E_{g\alpha}$ , where a small change in  $E_{g\alpha}$  necessitates a larger change in  $k_{g\alpha,0}$  to ensure the resulting  $k_{g\alpha}$  in (28) is of the same order of magnitude. Similar reasoning explains the large uncertainty in  $k_{g\beta,0}$ , with the correlation coefficient between  $k_{g\beta,0}$  and  $E_{g\beta}$  equal to 0.997. The growth exponent for the  $\alpha$ -form is near 2, which indicates that the  $\alpha$ -form growth rate is surface integration-limited, whereas that for the  $\beta$ -form is near 1, suggesting that the  $\beta$ -form growth rate is diffusion-limited. Unlike past studies that quantified uncertainties in the kinetic

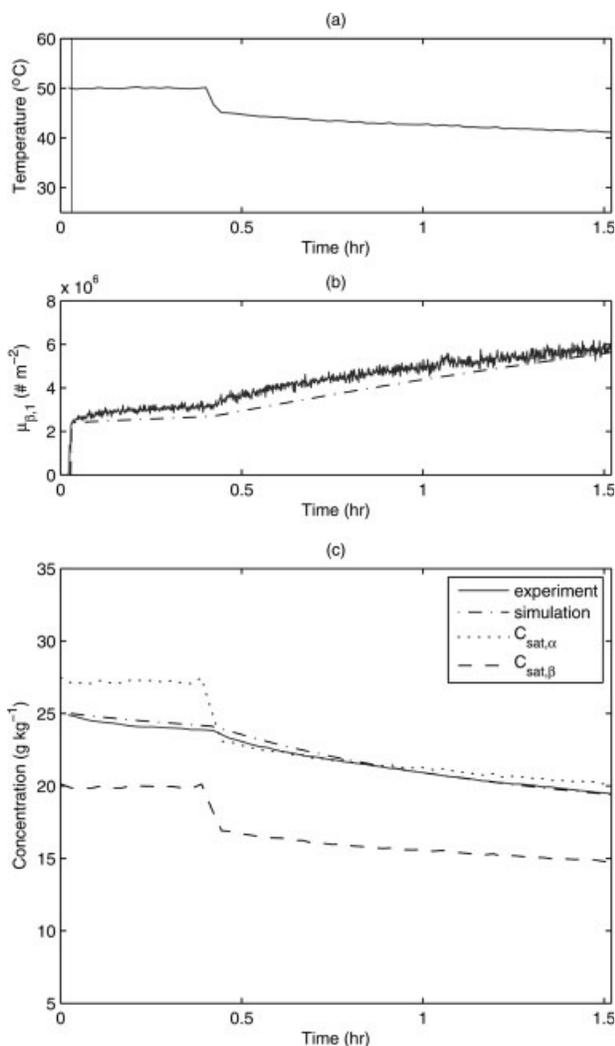


**Figure 6. Experimental and predictive trajectories of (a) temperature, (b) the first-order moment of the  $\alpha$ -form crystals, and (c) solute concentration for Experiment V1 of Table 7.**

The vertical line in plot (a) shows the seeding time.

<sup>8</sup>The time series for the other moments and other experiments are in Supporting Information.

<sup>1</sup>Samples were taken at the end of all batches and XRD was used to determine the crystal form purity.



**Figure 7. Experimental and predictive trajectories of (a) temperature, (b) the first-order moment of the  $\beta$ -form crystals, and (c) solute concentration for Experiment V2 of Table 7.**

The vertical line in plot (a) shows the seeding time.

parameters for crystallization processes,<sup>23,54</sup> the analysis in this article explicitly takes into account hard theoretical bounds on the values for the parameters. In particular, the application of the linearized analyses used in past papers would have resulted in a confidence interval that included values of  $g_{\beta} < 1$ , whereas the Markov Chain simulation approach takes the lower bound of 1 into account during the statistical analysis (see Figure 5d).

To assess the predictive capability of the resulting model, another pair of experiments (i.e., one  $\alpha$ - and one  $\beta$ -seeded experiment) were carried out with the seed distributions in Table 7. The trajectories of the temperature, first-order moment of the  $i$ -form crystals, and solute concentration trajectories obtained from experimental data and those predicted through simulation are plotted in Figures 6 and 7. As can be seen from these figures, the predictive capability of the model is sufficiently accurate for use in process design and control. The solute concentration predicted by the model are

quite close to the measured solute concentration in both validation experiments, with the differences between the predicted and experimental first-order moment being comparable to or smaller than the differences in the model and experimental first-order moments in the experiments used for parameter estimation (compare Figures 6 and 7 with Figures 2 and 3). The biases observed in the model predictions for the first-order moment of the  $i$ -form crystals could be due to the FBRM undercounting very small and large crystals, which would cause a different time-varying bias in different experiments.

## Conclusions

A model of polymorphic crystallization of L-glutamic acid, which consist of  $\alpha$ - and  $\beta$ -form crystallization, has been developed. The detailed kinetics model takes into account the temperature dependence of the crystals growth kinetic parameters when compared with the past studies on the modeling of L-glutamic acid crystallization.<sup>13,14</sup> In addition to providing point estimates of the kinetic parameters, a Bayesian inference approach is used to determine a detailed marginal probability distribution for each parameter. The marginal probability distributions of the parameters can give practitioners insight regarding the parameter uncertainties and are of significant value to develop robust control strategies for the crystallization process.<sup>24</sup>

Although this article considers a specific polymorphic crystallization, the same parameter estimation method can be applied for crystallizations in which many nucleation and growth rates occur simultaneously, or when there are no prior literature data or estimates for the model parameters. The details of the nucleation and growth rate expressions may be different, depending on the particular solute–solvent system. With multiple polymorphs in the crystallizer, improved parameter estimates would be obtained by including polymorph ratio measurements obtained from in situ Raman spectroscopy in (3).<sup>56</sup>

## Literature Cited

- De Anda JC, Wang XZ, Lai X, Roberts KJ. Classifying organic crystals via in-process image analysis and the use of monitoring charts to follow polymorphic and morphological changes. *J Process Control*. 2005;15:785–797.
- Blagden N, Davey R. Polymorphs take shape. *Chem Britain*. 1999;35:44–47.
- Fujiwara M, Nagy ZK, Chew JW, Braatz RD. First-principles and direct design approaches for the control of pharmaceutical crystallization. *J Process Control*. 2005;15:493–504.
- Rohani S, Horne S, Murthy K. Control of product quality in batch crystallization of pharmaceuticals and fine chemicals, Part 1: Design of the crystallization process and the effect of solvent. *Org Process Res Dev*. 2005;9:858–872.
- Yu LX, Lionberger RA, Raw AS, D’Costa R, Wu HQ, Hussain AS. Applications of process analytical technology to crystallization processes. *Adv Drug Delivery Rev*. 2004;56:349–369.
- Cardew PT, Davey RJ. The kinetics of solvent-mediated phase transformations. *Proc R Soc Lond A*. 1985;398:415–428.
- Sakai H, Hosogai H, Kawakita T, Onuma K, Tsukamoto K. Transformation of  $\alpha$ -glycine to  $\gamma$ -glycine. *J Cryst Growth*. 1992;116:421–426.
- Saranteas K, Bakale R, Hong YP, Luong H, Foroughi R, Wald S. Process design and scale-up elements for solvent mediated polymorphic controlled tecastemizole crystallization. *Org Process Res Dev*. 2005;9:911–922.

9. Yu L. Survival of the fittest polymorph: how fast nucleator can lose to fast grower. *CrystEngComm*. 2007;9:841–851.
10. Bard Y. *Nonlinear Parameter Estimation*. New York: Academic Press, 1974.
11. Bates DM, Watts DG. *Nonlinear Regression Analysis and Its Applications*. New York: Wiley, 1988.
12. Mendes P, Kell DB. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*. 1998;14:869–883.
13. Ono T, Kramer HJM, Ter Horst JH, Jansens PJ. Process modeling of the polymorphic transformation of L-glutamic acid. *Cryst Growth Des*. 2004;4:1161–1167.
14. Scholl J, Bonalumi D, Vicum L, Mazzotti M. In situ monitoring and modeling of the solvent-mediated polymorphic transformation of L-Glutamic acid. *Cryst Growth Des*. 2006;6:881–891.
15. Cailliet A, Sheibat-Othman N, Fevotte G. Crystallization of monohydrate citric acid. II. Modeling through population balance equations. *Cryst Growth Des*. 2007;7:2088–2095.
16. Fevotte G, Alexandre C, Nida SO. A population balance model of the solution-mediated phase transition of citric acid. *AIChE J*. 2007;53:2578–2589.
17. Box GEP, Draper NR. The Bayesian estimation of common parameters from several responses. *Biometrika*. 1965;52:355–365.
18. Duran MA, White BS. Bayesian estimation applied to effective heat transfer coefficients in a packed bed. *Chem Eng Sci*. 1995;50:495–510.
19. Bois FY, Fahmy T, Block JC, Gatel D. Dynamic modeling of bacteria in a pilot drinking-water distribution system. *Water Res*. 1997;31:3146–3456.
20. Coleman MC, Block DE. Bayesian parameter estimation with informative priors for nonlinear systems. *AIChE J*. 2006;52:651–667.
21. Pouillot R, Albert I, Cornu M, Denis JB. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to listeria monocytogenes. *Int J Food Microbiol*. 2003;81:87–104.
22. Gunawan R, Jung MY, Seebauer EG, Braatz RD. Maximum a posteriori estimation of transient enhanced diffusion energetics. *AIChE J*. 2003;49:2114–2122.
23. Miller SM, Rawlings JB. Model identification and control strategies for batch cooling crystallizers. *AIChE J*. 1994;40:1312–1327.
24. Nagy ZK, Braatz RD. Worst-case and distributional robustness analysis of finite-time control trajectories for nonlinear distributed parameter systems. *IEEE Trans Control Syst Technol*. 2003;11:694–704.
25. Beck JV, Arnold KJ. *Parameter Estimation in Engineering and Science*. New York: Wiley, 1977.
26. Tierney L. Markov Chains for exploring posterior distributions. *Ann Stat*. 1994;22:1701–1728.
27. Liu JS. *Monte Carlo Strategies in Scientific Computing*. New York: Springer, 2001.
28. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. New York: Chapman & Hall/CRC, 2004.
29. Chen WS, Bakshi BR, Goel PK, Ungarala S. Bayesian estimation via sequential Monte Carlo sampling: unconstrained nonlinear dynamic systems. *Ind Eng Chem Res*. 2004;43:4012–4025.
30. Lang L, Chen WS, Bakshi BR, Goel PK, Ungarala S. Bayesian estimation via sequential Monte Carlo sampling - constrained dynamic systems. *Automatica*. 2007;43:1615–1622.
31. Fujiwara M, Chow PS, Ma DL, Braatz RD. Paracetamol crystallization using laser backscattering and ATR-FTIR spectroscopy: metastability, agglomeration and control. *Cryst Growth Des*. 2002;2:363–370.
32. Togkalidou T, Fujiwara M, Patel S, Braatz RD. Solute concentration prediction using chemometrics and ATR-FTIR spectroscopy. *J Cryst Growth*. 2001;231:534–543.
33. Xie YL, Kalivas JH. Evaluation of principal component selection methods to form a global prediction model by principal component regression. *Anal Chim Acta*. 1997;348:19–27.
34. Bretthorst GL. An introduction to parameter estimation using Bayesian probability theory. In: Fougere PF, editor. *Maximum Entropy and Bayesian Methods*. Dordrecht: Kluwer Academic Publishers, 1990:53–79.
35. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC, 2000.
36. Box GEP, Tiao GC. *Bayesian Inference in Statistical Analysis*. Reading, Mass: Addison-Wesley, 1973.
37. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH. Equation of state calculations by fast computing machines. *J Chem Phys*. 1953;21:1087–1092.
38. Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli*. 2001;7:223–242.
39. Liang FM, Wong WH. Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models. *J Am Stat Assoc*. 2001;96:653–666.
40. Liang F. Dynamically weighted importance sampling in Monte Carlo computation. *J Am Stat Assoc*. 2002;97:807–821.
41. Laskey KB, Myers JW. Population Markov Chain Monte Carlo. *Machine Learn*. 2003;50:175–196.
42. Braak CJ. A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. *Stat Comput*. 2006;16:239–249.
43. Clontz NA, McCabe WL. Contact nucleation of magnesium sulfate heptahydrate. *AIChE Symp Ser*. 1971;67:6–17.
44. Hulburt HM, Katz S. Some problems in particle technology: a statistical mechanical formulation. *Chem Eng Sci*. 1964;19:555–574.
45. Mersmann A. *Crystallization Technology Handbook*, 2nd ed. Boca Raton: CRC Press, 2001.
46. Kitamura M, Ishizu T. Growth kinetics and morphological change of polymorphs of L-glutamic acid. *J Cryst Growth*. 2000;209:138–145.
47. Togkalidou T, Braatz RD, Johnson B, Davidson O, Andrews A. Experimental design and inferential modeling in pharmaceutical crystallization. *AIChE J*. 2001;47:160–168.
48. Togkalidou T, Tung HH, Sun Y, Andrews A, Braatz RD. Solution concentration prediction for pharmaceutical crystallization processes using robust chemometrics and ATR FTIR spectroscopy. *Org Process Res Dev*. 2002;6:317–322.
49. Tadayyon A, Rohani S. Determination of particle size distribution by Par-Tec 100: modeling and experimental results. *Part Part Syst Charact*. 1998;15:127–135.
50. Simmons M, Langston P, Burbidge A. Particle and droplet size analysis from chord distributions. *Powder Technol*. 1999;102:75–83.
51. Ruf A, Worlitschek J, Mazzotti M. Modeling and experimental analysis of PSD measurements through FBRM. *Part Part Syst Charact*. 2000;17:167–179.
52. Hukkanen EJ, Braatz RD. Measurement of particle size distribution in suspension polymerization using in situ laser backscattering. *Sens Actuators B*. 2003;96:451–459.
53. Heath AR, Fawell PD, Bahri PA, Swift JD. Estimating average particle size by focused beam reflectance measurement (FBRM). *Part Part Syst Charact*. 2002;19:84–95.
54. Togkalidou T, Tung HH, Sun Y, Andrews AT, Braatz RD. Parameter estimation and optimization of a loosely-bound aggregating pharmaceutical crystallization using in-situ infrared and laser backscattering measurements. *Ind Eng Chem Res*. 2004;43:6168–6181.
55. Gunawan R, Ma DL, Fujiwara M, Braatz RD. Identification of kinetic parameters in a multidimensional crystallization process. *Int J Mod Phys B*. 2002;16:367–374.
56. Starbuck C, Spartalis A, Wai L, Wang J, Fernandez P, Lindemann CM, Zhou GX, Ge ZH. Process optimization of a complex pharmaceutical polymorphic system via in situ Raman spectroscopy. *Cryst Growth Des*. 2002;2:515–522.

Manuscript received Jan. 14, 2008, and revision received July 8, 2008.