Data-driven modelling of product crystal size distribution and optimal input design for batch cooling crystallization processes

Jingxiang Liu\textsuperscript{a,b,c}, Tao Liu\textsuperscript{a,b,*}, Junhui Chen\textsuperscript{d,*}, Hong Yue\textsuperscript{e}, Fangkun Zhang\textsuperscript{a,b,f}, Feiran Sun\textsuperscript{a,b}

\textsuperscript{a} Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of Ministry of Education, Dalian 116024, China
\textsuperscript{b} Institute of Advanced Control Technology, Dalian University of Technology, Dalian, 116024, China
\textsuperscript{c} School of Marine Electrical Engineering, Dalian Maritime University, Dalian, 116026, China
\textsuperscript{d} Department of Chemical Engineering, Chung-Yuan Christian University, Chung-Li District, Taoyuan, 32023, Taiwan
\textsuperscript{e} Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XL, UK
\textsuperscript{f} College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao, 266061, China

* Corresponding authors. E-mails: liurouter@ieee.org (T. Liu); jason@wavenet.cycu.edu.tw (J. Chen)

Abstract: In this paper, a novel data-driven model building method is proposed for predicting one-dimensional product crystal size distribution (CSD) or chord length distribution (CLD) of batch cooling crystallization processes, based on only batch run data. The proposed model relating the manipulated variable of cooling rate to the product CSD are constructed by two classes of basis functions, one is the wavelet basis function for reshaping the CSD and the other is the polynomial basis function for weighting the chosen wavelet basis functions to reflect the nonlinear relationship between the input and the density of individual crystal size among the product crystals. Correspondingly, a double-layer least-squares algorithm is established to estimate the model parameters, along with an adaptive strategy to determine the location and number of wavelet basis functions. By introducing an objective function that combines the information entropy of product CSD and the sample deviation of product crystals in each batch with respect to the target crystal size, the optimal input design of cooling rate for the desired product CSD is carried out by using a particle swarm optimization (PSO) algorithm to solve the non-convex optimization problem with the established CSD model. Simulation tests on the hen-egg-white lysozyme crystallization process along with experiments on the L-glutamic acid cooling crystallization process are performed to demonstrate the effectiveness and advantage of the proposed method.

Keywords: batch cooling crystallization processes; data-driven model building; double-layer parameter estimation; crystal size distribution; optimal input design.
1. Introduction

The cooling crystallization technology has been widely used to extract or purify industrial products such as fine chemicals and pharmaceuticals. To ensure consistent crystal product properties of physics and/or chemistry from batch to batch, it is generally required to run batch crystallization processes to obtain uniform product crystal size distribution (CSD) or chord length distribution (CLD) [1], [2]. For operating cooling crystallization processes in engineering practice, it has been widely recognized that the cooling temperature profile, which corresponds to the cooling rate, is the most important operating condition [3]. A lot of research efforts have been devoted to control and optimization of batch cooling crystallization processes in the past two decades [4]-[6]. For designing the cooling temperature trajectory, mechanism models of cooling crystallization dynamics such as the population balance model [7]-[9] were studied to determine the optimal temperature trajectory based on the first-principle of crystallization kinetics. However, the existing model-based optimization methods were primarily based on simplified assumptions on the complicated mechanism of cooling crystallization processes while the impact from process operations was underestimated [8], therefore degrading the optimization performance in an intangible way.

Since CSD has been widely recognized as an important quality indicator of crystal products, increasing research efforts were devoted to control the product size distribution in the past years [10]. According to the quality-by-design concept [11], it was proposed to obtain a narrow unimodal shape distribution to increase the efficiency of downstream process operations such as filtration, drying, and washing [12]. The development of process analytical technology (PAT) has generated a paradigm shift from quality-by-test (QbT) to quality-by-design (QbD) [13], opening a way toward more innovative product designs and process engineering methodologies. Various approaches were developed by using PAT tools to design the operating conditions to obtain consistent CSD for batch run. These could be mainly divided into two categories [6],[14], i.e. model-based and model-free (or customarily named data-driven) approaches. Model-based approaches were generally based on the first-principle models of crystal growth kinetics to study optimal operating conditions [15]-[22]. However, it could be very troublesome or even impractical to establish an accurate first-principle
model for describing overall dynamics of a crystallization process in engineering practice [3], [6]. In contrast, data-driven methods have received increasing attentions in the recent years [6], [10], owing to that only batch run data rather than the process mechanism models are needed for QbD. To enumerate a few, design of experiment (DOE) was explored for data-driven design of process optimization [23]-[25]. By prescribing the search range of cooling rate with respect to the cooling crystallization operation constraints, a numerical optimization strategy was presented [26] to determine the cooling rate profile for optimizing the product crystal number and mean size, obtaining good prediction result but with significantly reduced computation effort compared to the previous model-based optimization methods. However, it was left open to optimize the product CSD in terms of the cooling rate. The supersaturation level of crystallization solution measured by using an attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy was referenced to design the optimal cooling curve for batch crystallization of succinic acid [27]. Statistics analysis was conducted based on DOE and experimental results to ascertain fundamental crystallization dynamics for batch operation [28]. Besides, DOE was incorporated into a multi-block partial least-squares (PLS) model for optimizing the batch operation profile [29]. Note that for a nonlinear batch process, it may not be feasible to perform piecewise DOE in terms of dividing the batch operation profile into a number of linearized regions, due to that the operating conditions could not be arbitrarily executed in terms of these linearized regions, e.g., the cooling crystallization process of L-glutamic acid (LGA) [30]. Owing to the complexity and difficulty for modeling nonlinear kinetics of various cooling crystallization processes, it remains open as yet for predicting product CSD with respect to the input conditions in engineering applications.

In this paper, a novel data-driven modelling method is proposed for predicting one-dimensional product CSD (or CLD) with respect to the temperature regulation of batch cooling crystallization processes, based on only batch run data. The proposed model relating the operating condition of cooling rate to product CSD consists of two classes of basis functions, one is the wavelet basis function for reshaping product CSD and the other is the polynomial basis function for weighting the chosen wavelet basis functions. A double-layer least-squares (LS) algorithm is established to estimate the model parameters. **By introducing a comprehensive objective function**
to minimize the information entropy of product CSD with respect to the desired product crystal size based on the above identified model, a particle swarm optimization (PSO) algorithm is given for optimal input design like the typical two-phase cooling strategy. Simulations on the hen-egg-white lysozyme crystallization process and experiments on the L-glutamic acid (LGA) cooling crystallization process are performed to demonstrate the effectiveness of the proposed method and its advantage of improving the crystal yield with the desired size over the existing methods.

For clarity, the paper is organized as follows: The problem description is briefly introduced in Section 2. A double-layer model building method is proposed in Section 3, including an adaptive strategy for choosing the basis functions. An optimal input design for obtaining the desired product CSD is presented in Section 4. Batch cooling crystallization processes of hen-egg-white lysozyme are simulated in Section 5 along with experiments on LGA cooling crystallization in Section 6, to demonstrate the effectiveness and merit of the proposed method. Finally, some conclusions are drawn in Section 7.

2. Problem description

It is well known that the crystal growth rate is mainly affected by the solution supersaturation level during crystallization, which is closely related to product CSD. For running batch cooling crystallization processes, it has been recognized that the cooling rate could determine the solution supersaturation profile and therefore, has the primary impact on CSD during crystallization. For batch operation, it is of paramount importance to regulate the supersaturation level via the cooling rate to guarantee sufficient driving force of crystal growth kinetics, while the crystal nucleation should be suppressed by maintaining the supersaturation level close to the solubility of crystallization solution. Compared to the traditional constant linear cooling rate, two-phase cooling rate strategy is increasingly explored in the recent years, where the cooling rate is a constant in the first half duration and changes to another constant in the latter half duration. However, it remains open to determine the optimal cooling profile for batch crystallization since the crystal growth kinetics and nucleation model are usually not available or roughly estimated in practice. Moreover, even if tentative kinetic models of crystal growth could be established in some specific cases, such a model could be very sensitive to the process operating conditions, and therefore could not give
accurate prediction of product attributes such as CSD with respect to different cooling profiles.

For batch crystallization, CSD or CLD has been primarily used to reflect the product attributes in engineering applications [2], [6], [10], based on one-dimensional measurement or estimation of crystal sizes. Figure 1 shows an illustration of crystal growth in a cooling crystallization process, where the initial CSD has a narrow width with the mean value about 65 \( \mu \text{m} \). When these seed crystals grow larger, the resulting CSD gradually moves to the right side along the time evolution. The final CSD shown by the red line in Figure 1 could have distinct shapes under different cooling rates.

It is therefore desired to establish a quantitative relationship between product CSD and the cooling rate for the convenience of QbD, such that the cooling rate could be optimally designed to obtain the desired product CSD for batch run. Hence, the research task in this study is to establish a feasible model to reflect the above relationship, based on only batch run data rather than estimating the complex crystal growth kinetics during crystallization. Using the identified model relating the cooling rate to the product CSD, an optimal design of the commonly used two-stage cooling strategy is explored to realize the desired product CSD.

3. Data-driven modeling for predicting product CSD

Using the batch run data, a novel double-layer model building method is proposed herein for predicting one-dimensional product CSD (or CLD) with respect to the cooling rate of batch cooling crystallization. The proposed double-layer model consists of two classes of basis functions, one is the wavelet basis function for reshaping the CSD and the other is the polynomial basis function for weighting the chosen wavelet basis functions to reflect the nonlinear relationship between the input and the density of individual crystal size among the product crystals. A double-layer LS algorithm is therefore developed to estimate the model parameters, along with an adaptive choice of these basis functions.

3.1 Double-layer model building

Denote by \( \mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_i, \ldots, \mathbf{u}_b]^T \in \mathbb{R}^{B \times r} \) the input data matrix of cooling rate implemented in batch run of cooling crystallization for model identification, where \( B \) is the number of batches,
\( r \) is the number of input variables (i.e., the tuning parameters of cooling rate) conducted in each batch run. Correspondingly, the measured product crystal sizes are denoted by \( y = [y_1, \ldots, y_k, \ldots, y_K] \in \mathbb{R}^K \) where \( y_k \) is the \( k \)th crystal size and \( K \) is the total number of measured crystal sizes in each batch. An illustration of the measured results on product CSD is shown in Figure 2, where \( \gamma_{i,k} \) denotes the measured density of a specific crystal size denoted by \( y_k \) in the \( i \)-th batch under the input \( u_i \). The size-related crystal density values for all test batches are denoted by

\[
\Gamma = [\gamma_1, \ldots, \gamma_k, \ldots, \gamma_K] = \begin{bmatrix}
\gamma_{1,1} & \cdots & \gamma_{1,k} & \cdots & \gamma_{1,K} \\
\vdots & \cdots & \vdots & \cdots & \vdots \\
\gamma_{B,1} & \cdots & \gamma_{B,k} & \cdots & \gamma_{B,K}
\end{bmatrix}
\]

(1)

where \( \gamma_{i,k} = \gamma(u_i, y_k) \).

To model the relationship between \( \gamma(u_i, y_k) \) and \( u_i \), it is proposed to use a finite number of spatial basis function \( \{\psi_m(y_k)\}_{m=1,\ldots,M} \) in combination with the weight functions \( \{w_m(u_i)\}_{m=1,\ldots,M} \) to approximate \( \gamma(u_i, y_k) \) as

\[
\gamma(u_i, y_k) = \sum_{m=1}^{M} w_m(u_i) \psi_m(y_k) + e_{i,k} \quad \text{(2)}
\]

where \( \psi_m(y_k) \) is the \( m \)th predefined basis function, \( w_m(u_i) \) the \( m \)th weight function, \( M \) the number of basis functions, \( e_{i,k} \) the approximation residual including measurement noise.

Equivalently, equation (2) can be rewritten in the following form,

\[
\gamma(u_i, y_k) = \gamma_{i,k} = w(u_i)^T \psi(y_k) + e_{i,k} \quad \text{(3)}
\]

where \( w(u_i) = [w_1(u_i), \ldots, w_m(u_i), \ldots, w_M(u_i)]^T \) and \( \psi(y_k) = [\psi_1(y_k), \ldots, \psi_m(y_k), \ldots, \psi_M(y_k)]^T \) are the weight function vector and basis function vector, respectively.

An illustration on the proposed numerical approximation of product CSD with respect to the input variable of cooling rate is shown in Figure 3, where the CSD curve under a specific input \( u_i \) is shown in Figure 3 (a), and a two-dimensional (2D) plot of product CSD under different constant-type inputs of cooling rate is shown in Figure 3 (b). It is perceived that the CSD curves could be approximated by using different weight functions denoted by \( w(u_i) \).
To determine proper weight functions denoted by \( \mathbf{w}(\mathbf{u}_i) \) for the above approximation, it is further proposed to express the relationship between \( \mathbf{w}_m(\mathbf{u}_i) \) and \( \mathbf{u}_i \) by another class of basis functions, i.e.,

\[
\mathbf{w}_m(\mathbf{u}_i) = \sum_{n=1}^{N} \varphi_n(\mathbf{u}_i)\mathbf{v}_{m,n} = \varphi(\mathbf{u}_i)^T \mathbf{v}_m
\]

(4)

where \( \varphi(\mathbf{u}_i) = [\varphi_1(\mathbf{u}_i), \ldots, \varphi_j(\mathbf{u}_i), \ldots, \varphi_N(\mathbf{u}_i)]^T \), \( \mathbf{v}_m = [\mathbf{v}_{m,1}, \ldots, \mathbf{v}_{m,j}, \ldots, \mathbf{v}_{m,N}]^T \), \( \varphi_j(\mathbf{u}_i) \) \( (j=1,2,\ldots,N) \) denote another class of pre-specified basis functions, and \( \mathbf{v}_{m,j} \) \( (j=1,2,\ldots,N) \) are the corresponding coefficients to be determined.

It follows from (4) that

\[
\mathbf{w}(\mathbf{u}_i)^T = \varphi(\mathbf{u}_i)^T \mathbf{V}^T
\]

(5)

where

\[
\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_m, \ldots, \mathbf{v}_N]^T = \begin{bmatrix}
\mathbf{v}_{1,1} & \cdots & \mathbf{v}_{M,1} \\
\vdots & \ddots & \vdots \\
\mathbf{v}_{1,N} & \cdots & \mathbf{v}_{M,N}
\end{bmatrix}
\]

(6)

By substituting (4) into (2), a double-layer model relating \( \mathbf{u}_i \) to \( \gamma(\mathbf{u}_i, y_k) \) is established as

\[
\gamma(\mathbf{u}_i, y_k) = f(\mathbf{u}_i, y_k) + \epsilon_i = \sum_{n=1}^{N} \sum_{m=1}^{M} \mathbf{v}_{m,n} \varphi_n(\mathbf{u}_i) \psi_m(y_k) + \epsilon_{i,k}
\]

(7)

which can be rewritten in the following form by substituting (5) into (3),

\[
\gamma(\mathbf{u}_i, y_k) = \mathbf{w}(\mathbf{u}_i)^T \mathbf{V}^T \mathbf{V} \psi(y_k) + \epsilon_{i,k}
\]

\[
= [\varphi_1(\mathbf{u}_i), \ldots, \varphi_N(\mathbf{u}_i)] \begin{bmatrix}
\psi_{1}(y_k) & \cdots & \psi_{M}(y_k)
\end{bmatrix} + \epsilon_{i,k}
\]

(8)

For simplicity, the measured CSD density matrix related to the input matrix of \( \mathbf{u}_i \) is expressed by

\[
\Gamma = \Phi \mathbf{V}^T \Psi + \mathbf{E}
\]

(9)

where

\[
\Phi = [\varphi(\mathbf{u}_1), \ldots, \varphi(\mathbf{u}_N)]^T = \begin{bmatrix}
\varphi_1(\mathbf{u}_1) & \cdots & \varphi_N(\mathbf{u}_1) \\
\vdots & \ddots & \vdots \\
\varphi_1(\mathbf{u}_N) & \cdots & \varphi_N(\mathbf{u}_N)
\end{bmatrix}
\]

(10)
\[ \Psi = [\psi(y_1), \ldots, \psi(y_K)] = \begin{bmatrix} \psi_1(y_1) & \ldots & \psi_1(y_K) \\ \vdots & \ddots & \vdots \\ \psi_M(y_1) & \ldots & \psi_M(y_K) \end{bmatrix} \]  

(11) 

and \( E \) is the approximation residual matrix.

Owing to that the basis functions matrices \( \Phi \) and \( \Psi \) are prespecified, the coefficient matrix \( V \) could be determined by solving the following minimization problem,

\[
\min \left\| \Gamma - \Phi V^\top \Psi \right\|^2_F = \min \text{ tr}((\Gamma - \Phi V^\top \Psi)^\top (\Gamma - \Phi V^\top \Psi))
\]  

(12) 

Denote \( A = \Gamma - \Phi V^\top \Psi \), the above optimization problem becomes

\[
\min \text{ tr}(A^\top A)
\]  

(13) 

Since there stands \( \text{tr}(A^\top A) = \text{tr}(AA^\top) = \text{vec}(A^\top)^\top \text{vec}(A^\top) \), where symbol ‘vec’ denotes a vectorization operator to stretch a matrix into a vector by stacking the columns, the derivative of (13) with respect to \( V \) can be computed as

\[
\frac{\partial \text{vec}(A^\top)^\top \text{vec}(A^\top)}{\partial \text{vec}(V)^\top} = \frac{\partial \text{vec}(A^\top)^\top \text{vec}(A^\top)}{\partial \text{vec}(A^\top)^\top} \frac{\partial \text{vec}(A^\top)}{\partial \text{vec}(V)^\top} = 2 \text{vec}(A^\top)^\top \frac{\partial \text{vec}(A^\top)}{\partial \text{vec}(V)^\top}
\]  

(14) 

Note that

\[
\frac{\partial \text{vec}(A^\top)}{\partial \text{vec}(V)^\top} = \frac{\partial \text{vec}((\Gamma - \Phi V^\top \Psi)^\top)}{\partial \text{vec}(V)^\top} = -\frac{\partial \text{vec}(\Psi^\top V \Phi^\top)}{\partial \text{vec}(V)^\top}
\]  

(15) 

\[
= -(\Phi \otimes \Psi^\top) \frac{\partial \text{vec}(V)}{\partial \text{vec}(V)^\top} = -(\Phi \otimes \Psi^\top)
\]  

where \( \otimes \) denotes the Kronecker product.

By substituting (15) into (14) and letting it be zero, one obtains

\[
\frac{\partial \text{vec}(A^\top)^\top \text{vec}(A^\top)}{\partial \text{vec}(V)^\top} = \text{vec}(A^\top)^\top (\Phi \otimes \Psi^\top) = 0
\]  

(16) 

By matrix transformation, there stands

\[
\text{vec}(A^\top)^\top (\Phi \otimes \Psi^\top) = ((\Phi^\top \otimes \Psi) \text{vec}(A^\top))^\top = \text{vec}(\Psi A^\top \Phi)^\top = 0
\]  

(17) 

\[
\Psi A^\top \Phi = \Psi(\Gamma - \Phi V^\top \Psi)^\top \Phi = 0
\]  

(18) 

It follows from (18) that

\[
\Psi \Gamma^\top \Phi = \Psi \Psi^\top V \Phi^\top \Phi
\]  

(19) 

Hence, a double-layer LS solution is obtained as
\[ V = (\Psi \Psi^T)^{-1} \Psi^T \Phi^T \Phi^{-1} \]  \hspace{1cm} (20)

With the above estimation of model parameters, the product CSD curve can be approximated by the following expression,

\[ \gamma(u) = \varphi(u)^T V^T \Psi \]  \hspace{1cm} (21)

where \( \varphi(u) = [\varphi_1(u), \ldots, \varphi_n(u), \ldots, \varphi_N(u)]^T \), \( V \) and \( \Psi \) have been defined in (6) and (11), respectively. Hence, the product CSD with respect to different inputs of cooling rate can be explicitly predicted in terms of the double-layer model in (21).

### 3.2 Choices of two classes of basis functions

Two different classes of basis functions are considered herein to fit the product CSD curve. One class is the wavelet basis function for reshaping the CSD curve owing to its wonderful properties of orthogonality and decomposition for local approximation. Another class is the polynomial basis function for describing the nonlinear relationship between the process input variable and the density of individual crystal size among the product crystals, owing to its simplicity and continuity for computation of the double-layer LS solution in Eq.(20). Note that these two classes of basis functions correspond to \( \Psi \) and \( V \) in Eq.(21), respectively.

Concerning the use of wavelet basis function for reshaping CSD, the basic idea is that any signal denoted by \( x(t) \) in time domain can be approximated by a set of orthogonal basis functions according to the wavelet analysis theory [32], i.e.,

\[ x(t) = \sum_{j=-\infty}^{\infty} c_j^J f_{J,k}(t) + \sum_{j=1}^{J} \sum_{k=1}^{2^j} d_{k,j}^J g_{J,k}(t) \]  \hspace{1cm} (22)

where \( j \) denotes the \( j \)th decomposition level, and \( k \) is the location of a basis function, \( f_{J,k}(t) \) and \( g_{J,k}(t) \) are the scaling functions at level \( J \) and the wavelet function at level \( j \) with respect to location \( k \), respectively, \( c_j^J \) and \( d_{k,j}^J \) are the corresponding coefficients, \( j = 1, \ldots, J \). In a closed subspace denoted by \( V_0 \), the signal \( x(t) \) can be initially represented by a scaling function with the coefficient \( c_0^0 \). According to the Mallat’s multi-resolution theory [32], \( c_0^0 \) can be decomposed into an approximation component denoted by \( c^1 \) on the subspace \( V_1 \) and a detail component denoted by \( d^1 \) on the orthogonal complement space \( W_1 \), respectively. Then \( V_1 \) can
be further decomposed into a subspace \( V_2 \) and an orthogonal complement space \( W_2 \) in the same way, and so on. Figure 4 gives an illustration on the wavelet decomposition procedure.

There are a few commonly used wavelet functions such as Haar, Daubechies, Biorthogonal, Coiflets and Symlets [31]. However, it remains open as yet for choosing proper wavelet functions for different signal approximation [33]. Herein Daubechies wavelet with four vanishing moments (DB4) is taken as the basis function for reshaping the CSD curve, owing to its properties of good regularity, multiple vanishing moments and approximate symmetry. These properties facilitate describing irregular characteristics in some local ranges such as low-frequency impulses. Accordingly, an adaptive strategy can be easily established to determine the number of scale and wavelet functions of DB4 for wavelet decomposition and reconstruction. The scale and wavelet functions of DB4 are illustrated in Figure 5 for the convenience of understanding.

Concerning the polynomial basis functions in (4) for weighting the chosen wavelet basis functions, each of them denoted by \( \varphi_j(\mathbf{u}) \) \( (j = 1,2,\ldots,N) \) can be generally expressed as

\[
\varphi_j(\mathbf{u}) = u_1^{h_1} u_2^{h_2} \cdots u_r^{h_r}
\]  \hspace{1cm} (23)

where \( h_i \) \( (i = 1,\ldots,r) \) are the exponential orders satisfying \( h_i \leq h \), and \( h \) is a prespecified upper order of these basis functions and thus is related to \( N \). For example, if there are two input variables, i.e., \( r = 2 \), letting \( h = 3 \) turns out the polynomial basis functions with \( N = 6 \) as below,

\[
\{1, u_1, u_2, u_1^2, u_1 u_2, u_2^2\}
\]  \hspace{1cm} (24)

The remaining issue is to determine the numbers of basis functions, denoted by \( M \) and \( N \) in (2) and (4), respectively. Since \( N \) is related to \( h \), the problem becomes how to choose the parameters \( M \) and \( h \). Considering that both \( M \) and \( h \) are integers in a limited range, the well-known brute force method may be used to determine these parameters, which, however, could be time consuming and demand high computation effort. To overcome the deficiency, a hypothesis test in terms of the principle of parsimony is preferred to determine the optimal numbers of these two classes of basis functions.

The historical CSD data denoted by \( \mathbf{\Gamma} \) are fitted by using wavelet functions in an adaptive manner. Since the wavelet functions could be stretched, the first level of basis functions is determined according to the length of samples. The values of scale and wavelet functions of DB4
are user specified, as shown in Figure 5. If the signal length is \( L_s \), then the length of the scale and wavelet functions in the first level should be similar to \( L_s \). Considering that the supported range of DB4 is 7, there are 14 basis functions used in the first level, consisting of 7 shifted scale functions and 7 shifted wavelet functions. For illustration, the scale and wavelet functions in the first level of DB4 are shown in Figure 6 for \( L_s = 800 \), where the lines in different colors and shapes represent different basis functions. The key idea is that the basis functions in the first level are initially specified, as shown by blue lines in Figure 6 (a) and (b). Then they are decomposed into seven levels of basis functions as shown by other lines in Figure 6 (a) and (b). An adaptive choice of these basis functions is given below:

Denote the first level basis functions as \( \mathbf{M}_1 \), the fitting coefficients are computed by using the standard LS method as

\[
\mathbf{v}_w = \Gamma \mathbf{M}_1 \left( \mathbf{M}_1^\dagger \mathbf{M}_1 \right)^{-1}
\]  

(25)

where \( \Gamma \) has been defined in (1). Then the predicted density matrix based on the basis functions denoted by \( \mathbf{M}_1 \) is given by

\[
\hat{\Gamma} = \mathbf{v}_w \mathbf{M}_1^\dagger
\]  

(26)

Correspondingly, the prediction error matrix \( \mathbf{E} \) is computed by

\[
\mathbf{E}_{ij} = (\Gamma_{ij} - \hat{\Gamma}_{ij})^2
\]  

(27)

where \( \mathbf{E}_{ij} \), \( \Gamma_{ij} \) and \( \hat{\Gamma}_{ij} \) are elements in the \( i \)th row and the \( j \)th column of \( \mathbf{E} \), \( \Gamma \) and \( \hat{\Gamma} \), respectively.

Denote the mean fitting error vector by

\[
\mathbf{e}_{LS} = \sum_{i=1}^{B} \mathbf{e}_i
\]  

(28)

where \( \mathbf{e}_i \) is the \( i \)th row of \( \mathbf{E} \) with a column dimension of \( K \).

If the following condition is satisfied,

\[
\sum_{k=1}^{K} \mathbf{e}_{LS,k} < \frac{\theta}{K}
\]  

(29)

where \( \mathbf{e}_{LS,k} \) is the \( k \)th element in \( \mathbf{e}_{LS} \) and \( \theta \) is a predefined threshold, e.g. \( \theta = 10^{-5} \), it could be concluded that the current number of basis functions are sufficient for model fitting. Otherwise,
the specified wavelet functions should be further scaled and then added to the basis function matrix for computing \( e_{ik} \). By locating the maximum in \( e_{ik} \), the corresponding wavelet function is taken as a basis function and meanwhile, those values around the maximum of \( e_{ik} \) is set to zero. Thus, the error vector is updated as \( e'_{ik} \). Then the second maximum error is similarly determined in the updated \( e'_{ik} \) and the corresponding wavelet function is added into the above basis functions. The procedure is iterated until the condition in (29) is satisfied. For ease of understanding, Figure 7 shows an illustration on the adaptive choice of wavelet basis functions.

4. Optimal input design for desired product CSD

To assess the product quality in terms of CSD, it is desired to have the distribution width of product CSD as narrow as possible for uniformity. Considering that the information entropy is a commonly used index to assess flatness or sharpness of a distribution curve [34], i.e., a larger entropy corresponds to a flatter curve and vice versa, the entropy index is therefore introduced to assess the concentration degree of product CSD,

\[
S = -\sum_{k=1}^{K} \gamma(u, y_k) \log(\gamma(u, y_k)) \tag{30}
\]

For a unimodal distribution curve, it follows that the smallest entropy could result in the narrowest distribution width. Hence, the optimal CSD could be determined by finding out the minimum of entropy with respect to different input profiles. The optimization problem is therefore formulated as

\[
\min_u \left( -\sum_{k=1}^{K} \gamma(u, y_k) \log(\gamma(u, y_k)) \right) \tag{31}
\]

However, for a multimodal distribution curve, as often encountered in cooling crystallization processes [10], the above entropy may not be used to determine the optimal CSD curve. An illustrative example is given in Figure 8. It is seen that there are two local modals in the curve, marked by A and B, respectively. Although the entropy of A is obviously smaller than that of B, the crystal distribution of B is more desired than A according to the target crystal size range for production. This fact indicates that only searching the minimum entropy cannot guarantee the optimal product CSD with respect to the target crystal size range. To address the problem, a
synthetic index is proposed for practical application. Define a sample deviation of the final crystals in each batch with respect to the target crystal size by

$$E = \sum_{k=1}^{K} |y_k - d| \gamma(u, y_k)$$  \hspace{1cm} (32)$$

where \(d\) is the target crystal size that is user specified in advance based on a prior knowledge of the process operation. This index means that a smaller \(E\) corresponds to a larger percentage of product sizes in the target crystal size range.

A comprehensive quality index is therefore defined as

$$Q = S + \alpha E$$  \hspace{1cm} (33)$$

where \(\alpha\) is a tuning parameter used to compromise the effect of entropy and the above deviation.

To normalize the magnitudes of \(S\) and \(\alpha E\) to the same level for evaluation, it is suggested to take \(\alpha = \alpha_0 \cdot 10^{(\log_2(S/E))^k} \cdot \alpha_0 \in [0, 1]\), where \([\cdot]\) denotes a rounding operator.

Hence, the optimal product CSD can be determined by minimizing \(Q\) via the input design of cooling rate, i.e.,

$$\min_u (Q) = \min_u (S + \alpha E)$$  \hspace{1cm} (34)$$

For tuning the cooling rate to obtain the desired product CSD, the widely used strategies include constant cooling, linear cooling, and multi-phase cooling in engineering applications [3]. Compared to the former two strategies, the multi-phase cooling strategy have more flexibility to obtain better result, but need to design the cooling rate for each phase, which remains open as yet for batch optimization [2]. A typical two-phase cooling strategy is described below,

$$T(t) = \begin{cases} T_0 - c_1 t, & 0 \leq t < P_1 \\ T_1 - c_2 (t - P_1), & P_1 \leq t < P \end{cases}$$  \hspace{1cm} (35)$$

where \(T_0\) is the initial temperature for cooling crystallization, \(t\) the time index, \(c_1\) a constant cooling rate for the first phase with duration time denoted by \(P_1\), \(T_1\) the solution temperature at the end of the first phase, \(c_2\) another constant cooling rate for the second phase, \(P\) the batch period. For the convenience of implementation, it often takes \(P=2P_1\) in practical applications [3]. Note that if letting \(c_1 = c_2\), it is reduced to the traditional constant cooling strategy.

Based on the above established double-layer model relating the cooling rate to product CSD,
the objective function in (34) can be optimized to determine \( c_1 \) and \( c_2 \) in the above two-phase cooling strategy with a specified \( P_1 \), e.g., \( P_1 = P / 2 \), for obtaining the desired product CSD. Moreover, \( P_1 \) can also be optimized together with \( c_1 \) and \( c_2 \) for obtaining the desired product CSD, by virtue of the general input description defined in (2) for model building. This can be further extended to optimize a multi-phase cooling strategy with multiple input variables, at the cost of a larger number of the polynomial basis functions and training data for model building.

It should be noted that the optimization problem in (34) could become nonlinear or non-convex if a large number of polynomial basis functions are used for model building. To solve such a non-convex optimization problem, a PSO algorithm is herein proposed for implementation. The widely used PSO algorithm [35] is a powerful optimization program based on the evolution of swarm intelligence. The swarm consists of individual particles, and each particle’s position represents a potential solution to the optimization problem. Their positions will be iteratively updated based on the result computed in the previous step (or run) for a predefined objective function. Denote by \( u_i(d) \) the location of the \( i \)th particle in the \( d \)th run, by \( v_i(d) \) the velocity vector, by \( u_i^{\text{opt}}(d) \) the location of the best value for the \( i \)th particle, by \( u_0^{\text{opt}}(d) \) the location of the best value for all particles by the end of the \( d \)th run. In each run, the locations of all particles are updated in order to move forward to \( u_i^{\text{opt}}(d+1) \) and \( u_0^{\text{opt}}(d+1) \) as below:

\[
\begin{align*}
\mathbf{v}_i(d+1) &= w \mathbf{v}_i(d) + c_1 r_1 (u_i^{\text{opt}}(d) - u_i(d)) + c_2 r_2 (u_0^{\text{opt}}(d) - u_i(d)) \\
\mathbf{u}_i(d+1) &= u_i(d) + v_i(d+1)
\end{align*}
\]  

(36)

where \( c_1 \) and \( c_2 \) are the accelerated factors, \( r_1 \) and \( r_2 \) are random numbers set in a range of [0,1] for searching, and \( \mathbf{w} \) is the weight coefficient. For practical application, it is generally suggested to take the accelerated factors as \( c_1 = c_2 = 2 \) to have equal effect on the two searching directions with respect to \( u_i^{\text{opt}}(d) \) and \( u_0^{\text{opt}}(d) \), and \( \mathbf{w} \) in a range of [0.5,1] to tune the effect of the previous velocity on the current one.

The proposed PSO algorithm for determining the optimal input is summarized as follows.

**Step 1.** Initialize the number of particles (e.g., [50, 100]), the maximum number of runs denoted by \( D_{\max} \) (e.g., [200, 500]), the initial locations of all particles with respect to the input
range, the accelerated factors $c_1$ and $c_2$, and the weight coefficient $w$;

**Step 2.** Compute the corresponding CSD by Eq. (21) and the index $Q$ by Eq. (33) for each particle.

**Step 3.** By comparison on the resulting minimum of $Q$ in (34), determine $u_{o,i}^{\text{opt}}(d)$ for the $i$th particle and $u_{o,i}^{\text{opt}}(d)$ for all particles by the end of the $d$th run.

**Step 4.** Update the velocity vector and location vector for searching by Eq. (36).

**Step 5.** Repeat the above steps till a convergence condition, $|u_{o,i}^{\text{opt}}(d) - u_{o,i}^{\text{opt}}(d-1)| < \delta$, is satisfied, where $\delta$ is a predefined threshold, e.g., $\delta = 0.001$, or reach the maximum number of runs $D_{\max}$.

---

5. **Simulation test on hen-egg-white lysozyme crystallization**

In this study, a commercial software named Crystal-MPB (made by Qingdao PharmaVision Ltd) is used for simulating the cooling crystallization processes of hen-egg-white lysozyme. The morphological population balance (MPB) model of crystallization kinetics in the software Crystal-MPB was built from the references [36]-[38]. Figure 9 gives an illustration of this software package. Herein a one-dimensional MPB model is considered for simulating the hen-egg-white lysozyme crystallization process. The process parameters are listed in Figure 9 (b) according to the default values set therein. The meanings of these parameters are explained in Table 1.

Consider the above two-phase cooling strategy with $P_1 = P / 2$ for cooling crystallization, 25 batches are simulated by taking different cooling rates in the range of $c_1, c_2 \in \{1, 2, 3, 4, 5\}$ ($^\circ$C/day). Table 2 lists different combinations of $c_1$ and $c_2$ for test. All the simulation tests are repeated 50 times under the same operations but with random variances of crystal seeds taken in a range of $[3.9, 4.1]$. The corresponding mean and variance of $Q$ under different choices of $c_1$ and $c_2$ are listed in Table 2. By setting $\theta = 10^{-6}$, 44 wavelet functions are chosen as the basis functions by using an adaptive strategy shown in Figure 7. Correspondingly, the number of polynomial basis functions for weighting these wavelet basis functions is chosen as $N = 6$ by using a hypothesis test in terms of the principle of parsimony. The model fitting and prediction results for different cooling rates are shown in Figure 10 and 11, respectively. It is seen that the simulated CSD curves
(blue lines) by Crystal-MPB can be well fitted by the proposed double-layer model (red lines) as shown in Figure 10. Moreover, the prediction results shown in Figure 11 demonstrate that the proposed method well predicts the product CSD under different cooling rates for test.

To maximize the product yield with a desired CSD centered at \( d = 220 \, (\mu m) \), the input variables of cooling rate for two phases are determined as \( c_1 = 1.39 \, (^{\circ}C/day) \) and \( c_2 = 3.40 \, (^{\circ}C/day) \), by using a PSO algorithm with \( \alpha = 100 \) for optimizing the objective function in Eq.(34), resulting in the minimum \( Q = 7.4334 \). From Table 2, it is obvious by an one-dimensional (1-D) search that the minimum is \( Q = 7.4827 \) with variance of 0.0057, which corresponds to the two-phase cooling rates \( c_1 = 1 \) and \( c_2 = 3 \). For comparison, the polynomial interpolation method is used to find out the optimal two-phase cooling rates. It can be intuitively seen from Table 2 that the optimal cooling rates are likely located in a range where \( c_1 = 1 \) and \( c_2 \in [3, 4] \), or another range where \( c_2 = 1 \) and \( c_1 \in [3, 4] \), as shown by two ellipsoid regions in Table 2. Therefore, the 1-D polynomial interpolation method (also named 2-points interpolation) is used to determine the optimal cooling rates in these two local ranges, resulting in \( Q = 7.4808 \) corresponding to \( c_1 = 1 \) and \( c_2 = 3.6 \), and \( Q = 7.5153 \) corresponding to \( c_1 = 3.55 \) and \( c_2 = 1 \), respectively. Moreover, the two-dimensional (2-D) polynomial interpolation method is also used to determine the optimal cooling rates, based on the overall 25 batch tests and the corresponding mean values of \( O \). Accordingly, the minimum is determined as \( Q = 7.4498 \), corresponding to \( c_1 = 1.13 \) and \( c_2 = 3.52 \).

Figure 12 gives an illustration on computing the index \( O \) by the interpolation method and the proposed method, respectively. It is seen that the interpolation method constructs a direct mapping between the input and the index \( O \), which may be involved with serious nonlinearity and non-convex issue, as verified by the above 1-D or 2-D interpolation results. In contrast, the proposed method establishes a relationship between the input and product CSD via a double-layer model, such that the index \( O \) can be quantitively computed for optimization. By using the proposed PSO algorithm for such optimization, the minimum of \( O \) and the corresponding two-phase cooling rates can be easily determined with accuracy. Figure 13 shows the product CSD curve under the optimized input design by the proposed method (red line) in comparison with the input choices.
determined by 1-D search from Table 2 and the polynomial interpolation methods. It is seen that obviously improved product CSD around the desired crystal size of 220 \( \mu m \) (black dash line) is obtained by the proposed method. \textbf{Note that although the computed Q values by the above methods are slightly different from each other, the corresponding CSD curves looks quite different. In particular, the product crystals by the proposed method located around the desired size of 220 \( \mu m \) (\( \pm 5 \mu m \)), as indicated by the green dash-line box in Figure 13, could be relatively 20% more than those results by the other methods.}

6. Experiments on LGA cooling crystallization

The experimental set-up and its schematic for LGA cooling crystallization are shown in Figure 14 (a) and (b), respectively. The crystallizer consists of a 1-liter (L) glass jacketed reactor, a thermostatic circulator (product no. Julabo-CF41), a PT100 thermometer, and a PTFE four-paddle agitator. A diamond ATR immersion probe connected via AgX Fiber as the internal reflectance element attached to the FTIR spectroscopy was used to collect the absorbance spectra of LGA solution. The ATR-FTIR spectra were monitored by a software named ReactIR15 made by Mettler-Toledo Company. A FBRM probe (ParticleTrack G400, made by Mettler-Toledo) coupled with the data acquisition software named iC FBRM was used to measure the chord length distribution (CLD). The probe was set in the fine mode with the laser focused at 0 \( \mu m \) before experiments and set at a scanning speed of 2 m/s during experiments. The CLD histograms consistent of 100 bins from 1 \( \mu m \) to 600 \( \mu m \) and the data were recorded by a sampling period of 30 seconds for each experiment. All data collected by FBRM and ATR-FTIR were displayed in the iC FBRM and ReactIR 15 software packages, respectively. The raw materials for cooling crystallization are \( \beta \)-LGA (C\(_3\)H\(_6\)NO\(_4\), purity: 99%, produced by Sigma) along with distilled water as the solvent. A high-resolution analytical balance (Mettler-Toledo) with a precision of one ten thousandth was used to weigh the \( \beta \)-LGA samples for experiments. \textbf{It should be noted that for a crystallization process only having one-dimensional CSD, e.g., sphere- or needle-type crystals like \( \beta \)-LGA, the measured CLD by FBRM can be effectively transferred into CSD by some numerical computation methods [39],[40], and therefore has been widely used in the crystallization engineering field for analysis of CSD during crystallization. Since CLD is very similar to CSD with high dimension and}
nonlinear characteristics, the measured CLD data by FBRM were straightforwardly used by the proposed method for establishing a double-layer model from the input to product CLD.

In each experiment, 500 ml distilled water was firstly poured into the 1L glass jacketed reactor, and heated by the thermostatic circulator to 75 °C. The stirring speed was kept at 200r/min for all experiments. Then a small amount of solution was discharged from the bottom of the jacketed reactor, which was used to dissolve 10 g LGA solute, and poured back into the reactor. The high temperature of 75 °C was maintained about 60 mins to guarantee that all the solutes were fully dissolved such that the concentration of the LGA solution was 20g/L in the reactor. Subsequently, the solution was cooled down with a cooling rate of 1 °C /min to the temperature of 45 °C, and then the temperature was maintained invariant for 30 mins. After that, 0.5 g LGA seeds were added into the solution, with a size distribution in a range of [112, 125] (μm). Then, a two-phase cooling strategy with $P_1 = P / 2$ for cooling crystallization was conducted. The batch time of the cooling crystallization was set as 60 mins, with the first cooling rate denoted by $c_1$ in the first 30 mins and the second cooling rate denoted by $c_2$ in the last 30 mins. In this study, 16 batch experiments were conducted, under the cooling rates $c_1$ and $c_2$ of $\{0.2, 0.3, 0.4, 0.5\}$ (°C/min), respectively, as listed in Table 3. Each batch experiment were repeated three times under the same input conditions.

Based on the measured product CLD curves and input conditions of $c_1$ and $c_2$ from the above batch experiments, the proposed method is used to establish the product CLD model with respect to the input conditions. For illustration, Figure 15 shows the product CLD model fitting results under different two-phase cooling rates. It is seen that the experimental CLD (blue line) can be well fitted by the proposed double-layer model (red line).

According to a prior knowledge of β-LGA cooling crystallization [30], the product crystal CLD centered at a size of 200 μm is desired for production. To maximize the product yield in terms of the desired crystal size, the input variables of cooling rate for two phases are determined as $c_1 = 0.36$ (°C/min) and $c_2 = 0.44$ (°C/min), by using a PSO algorithm with $\alpha=100$ for optimizing the objective function in (34), resulting in the minimum $Q = 11.4861$. For comparison, the mean values of $Q$ based on three replicated batch experiments for different choices of $c_1$ and $c_2$ are listed in Table 3. It is obvious by 1-D search from Table 3 that the minimum of $Q$ is 11.5545 under
the input conditions of $c_1 = 0.2 \, (^{\circ}\text{C}/\text{min})$ and $c_2 = 0.3 \, (^{\circ}\text{C}/\text{min})$, which is evidently larger than that obtained by the proposed method. Table 3 also indicates that this is a non-convex optimization problem for which the minimum of $Q$ is difficult or even impossible to be found by the above interpolation methods. Figure 16 shows the predicted product CLD under the optimized input design (blue line) in comparison with three other input choices that yield relatively smaller values of $Q$ as shown in Table 3. It is seen that the optimized CLD (blue line) by the proposed method is more closely centered at the desired crystal size $d = 200 \, (\mu m)$ compared to the other results. Note that the product crystals by the proposed method located around the desired size of $200 \, \mu m$ ($\pm 20 \, \mu m$), as indicated by the green dash-line box in Figure 16, could be relatively 10% more than those results by the other cooling conditions. To verify the effectiveness of the proposed input design, the corresponding experiment was conducted for the above batch crystallization process of LGA. The resulting product CLD is shown in Figure 17, well indicating coherence with the predicted optimal design. The corresponding index $Q$ is 11.5007, which is also obviously lower than those indices under other cooling conditions listed in Table 3.

7. Conclusions

A novel double-layer model building method has been proposed for predicting one-dimensional product CSD (or CLD) of batch cooling crystallization processes, based on only batch run data. A salient merit is that the product CSD (or CLD) can be explicitly expressed with respect to the input variable of cooling rate, by using two classes of basis functions, i.e. the wavelet basis functions for reshaping the product CSD and polynomial basis functions for weighting these wavelet basis functions to reflect the nonlinear relationship between the input and the density of individual crystal size among the product crystals. An adaptive strategy is given for choosing these basis functions to guarantee the model fitting accuracy. Based on the established model, a synthetic quality index is proposed for the optimal input design to realize the desired product CSD, which combines the information entropy of product CSD and the sample deviation of product crystals with respect to the target crystal size. The proposed objective function can be used for optimization of multimodal CSD curves and the corresponding input design that are often encountered in engineering applications, by using a PSO algorithm to solve the non-convex optimization problem.
The applications to a simulated hen-egg-white lysozyme crystallization process and a real LGA crystallization process have well demonstrated the effectiveness and merit of the proposed data-driven modelling method and optimal input design in terms of a two-phase cooling strategy, in comparison with the traditional 1-D search or polynomial interpolation methods for the input design.

Acknowledgement

This work was supported in part by the NSF China Grant 61633006, the Talent Project of Revitalizing Liaoning (XLYC1902030), and National Key RD Program of China under Grant 2017YFA0700300.
References


List of Table and Figure Captions

Table 1. List of parameters for a simulated crystallization process
Table 2. Different choices of cooling rate in a two-phase cooling strategy and the corresponding indices of Q for a simulated crystallization process
Table 3. Different choices of cooling rate in a two-phase cooling strategy and the corresponding indices of Q for LGA crystallization

Figure 1. Illustration of crystal growth in a cooling crystallization process
Figure 2. Illustration of product CSD with respect to a specified cooling rate
Figure 3. Illustration of the numerical approximation for (a) the CSD curve under input $u_i$; (b) CSD curves under different inputs
Figure 4. Illustration of the wavelet decomposition procedure
Figure 5. Illustration of DB4 wavelet: (a) the scale function; (b) the wavelet function
Figure 6. DB4 based reshaping basis functions: (a) the scale functions; (b) the wavelet functions
Figure 7. Illustration on the adaptive choice of wavelet basis functions
Figure 8. Illustration of a bimodal distribution curve
Figure 9. Illustration of Crystal-MPB software: (a) the user interface; (b) simulation parameters
Figure 10. Fitting results by the proposed modelling method for a simulated crystallization process under: (a) $c_1=1$, $c_2=2$; (b) $c_1=5$, $c_2=4$
Figure 11. Prediction results by the proposed modelling method for a simulated crystallization process under: (a) $c_1=2.5$, $c_2=3.5$; (b) $c_1=3.5$, $c_2=3.5$
Figure 12. Comparison of interpolation method and the proposed method
Figure 13. Plot of the optimized CSD (blue line) by the proposed method in comparison with polynomial interpolation methods
Figure 14. Experimental set-up for LGA cooling crystallization: (a) external view; (b) schematic diagram
Figure 15. Product CLD model fitting for LGA crystallization under different two-phase cooling rates
Figure 16. Plot of the optimized CLD (blue line) by the proposed method in comparison with three different choices of two-phase cooling rates
Figure 17. Plot of the optimized CLD (blue line) by the proposed method in comparison with the experimental verification
Table 1. List of parameters for a simulated crystallization process

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numx</td>
<td>Number of discrete grid</td>
</tr>
<tr>
<td>Grid size</td>
<td>The size of each grid (μm)</td>
</tr>
<tr>
<td>The grid size of time</td>
<td>The time step for simulation</td>
</tr>
<tr>
<td>Time interval</td>
<td>The total number of simulation steps</td>
</tr>
<tr>
<td>Density of solid</td>
<td>The density of crystals (kg/m³)</td>
</tr>
<tr>
<td>Initial temperature</td>
<td>The initial solution temperature (°C)</td>
</tr>
<tr>
<td>The mean value of x</td>
<td>The initial crystal size (μm)</td>
</tr>
<tr>
<td>The variance of x</td>
<td>The variance of crystal size distribution</td>
</tr>
<tr>
<td>k0</td>
<td>Growth velocity of crystals (m/s)</td>
</tr>
<tr>
<td>n0</td>
<td>Growth velocity exponent of crystals</td>
</tr>
</tbody>
</table>

Table 2. Different choices of cooling rate in a two-phase cooling strategy and the corresponding indices of Q for a simulated crystallization process

<table>
<thead>
<tr>
<th>c2</th>
<th>c1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>9.5959(±0.0234)</td>
<td>8.1092(±0.0057)</td>
<td>7.5221(±0.0031)</td>
<td>7.5509(±0.0059)</td>
<td>8.3516(±0.0133)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>8.3113(±0.010)</td>
<td>7.5046(±0.003)</td>
<td>7.5804(±0.0037)</td>
<td>8.2448(±0.0082)</td>
<td>9.3321(±0.0137)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7.4027(±0.0057)</td>
<td>7.5487(±0.0039)</td>
<td>8.1596(±0.0044)</td>
<td>9.1880(±0.0081)</td>
<td>10.3840(±0.0184)</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7.4838(±0.0031)</td>
<td>8.1113(±0.0029)</td>
<td>9.1210(±0.0050)</td>
<td>10.2787(±0.0055)</td>
<td>11.2945(±0.0147)</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>8.0236(±0.0071)</td>
<td>9.1592(±0.0047)</td>
<td>10.292(±0.0076)</td>
<td>11.2519(±0.0154)</td>
<td>11.8949(±0.0219)</td>
</tr>
</tbody>
</table>

Table 3. Different choices of cooling rate in a two-phase cooling strategy and the corresponding indices of Q for LGA crystallization

<table>
<thead>
<tr>
<th>c2</th>
<th>c1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>11.7689</td>
<td>12.5413</td>
<td>12.1933</td>
<td>12.4879</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>11.5545</td>
<td>12.2269</td>
<td>11.6832</td>
<td>11.8092</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>11.5702</td>
<td>12.0914</td>
<td>11.6715</td>
<td>11.9144</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>11.8911</td>
<td>11.9808</td>
<td>11.6268</td>
<td>12.1656</td>
</tr>
</tbody>
</table>
Figure 1. Illustration of crystal growth in a cooling crystallization process

Figure 2. Illustration of product CSD with respect to a specified cooling rate
Figure 3. Illustration of numerical approximation for (a) the CSD curve under input $u_i$; (b) CSD curves under different inputs

Figure 4. Illustration of the wavelet decomposition procedure
Figure 5. Illustration of DB4 wavelet: (a) the scale function; (b) the wavelet function.
Figure 6. DB4 based reshaping basis functions: (a) the scale functions; (b) the wavelet functions.
Take the first wavelet function and reset the values of $\mathbf{e}_{LS}$ in the range as zeros.

Add the second wavelet function and reset the values of $\mathbf{e}'_{LS}$ in the range as zeros.

Add the third wavelet function and reset the values of $\mathbf{e}''_{LS}$ in the range as zeros.

Obtain a fitting error vector satisfying (29) based on the chosen basis functions.

**Figure 7. Illustration on the adaptive choice of wavelet basis functions**
Figure 8. Illustration of a bimodal distribution curve

Figure 9. Illustration of Crystal-MPB software: (a) the user interface; (b) simulation parameters.
Figure 10. Fitting results by the proposed modelling method for a simulated crystallization process under: (a) c1=1, c2=2; (b) c1=5, c2=4

Figure 11. Prediction results by the proposed modelling method for a simulated crystallization process under: (a) c1=2.5, c2=3.5; (b) c1=3.5, c2=3.5
**Interpolation method:**

Direct mapping

Input: $u$  →  Index: $Q$

Double-layer modeling

The proposed method:

Input: $u$  →  Product CSD  →  Index: $Q$

Quantitative computation

Comprehensive mapping

**Figure 12. Comparison of the interpolation method and the proposed method**

**Figure 13. Plot of the optimized CSD (blue line) by the proposed method in comparison with polynomial interpolation methods**
Figure 14. Experimental set-up for LGA cooling crystallization: (a) external view; (b) schematic diagram.
Figure 15. Product CLD model fitting for LGA crystallization under different two-phase cooling rates.

Figure 16. Plot of the optimized CLD (blue line) by the proposed method in comparison with three different choices of two-phase cooling rates.
Figure 17. Plot of the optimized CLD (blue line) by the proposed method in comparison with the experimental verification.