Evaluation of Parameter Estimation Methods for Crystallization Processes Modeled via Population Balance Equations

Author of this report:
Maximilian Besenhard
Graz University of Technology
Inffeldgasse 13, Graz, Austria 8010

External Supervisor:
Prof. Rohit Ramachandran
Rutgers-The State University of New Jersey
98 Brett Road, Piscataway, NJ-08854

Internal Supervisor:
Prof. Rohit Ramachandran
Graz University of Technology
Inffeldgasse 13, Graz, Austria 8010

Acknowledgement:

First and foremost, I would like to thank the Marshall Plan foundation for giving me the opportunity to work in Rutgers-The State University of New Jersey USA. This exposure has led to a great enrichment in my research and I am very grateful for this experience. I have grown as a person as well as a researcher. I really appreciate the funding of Marshall Plan foundation encouraging between the universal collaboration between these two countries-United States of America and Austria. I would like to acknowledge my internal supervisor Prof Johannes Khinast for his permanent guidance and the motivation to go abroad. I would like to thank my external supervisor Rohit Ramachandran for all of the invaluable assistance and guidance that he provided during my stay at Rutgers. Furthermore, I would like to thank Anwesha Chaudhury for her assistance and support which was way beyond what someone could expect from a colleague and friend.

Abstract:

Population balance equations (PBE) coupled with mass and energy balance equations represent the common framework for crystallization processes. Often expressions required expressions for crystal growth, nucleation, as well as aggregation and breakage rates contain parameters that need to be estimated from experimental data. To establish a process model, parameter estimation (PE) is applied to determine an optimal set of parameters by minimizing the sum of squared errors between the experimental results and the model output. Inappropriate selection of the objective function, the optimization routine itself and inaccurate or limited experimental data might severely handicap the parameter estimation procedure.

In this study the sensitivity of parameter estimation concepts is investigated. Therefore the limits of multiple optimization algorithms (global and local ones) and the consequence of limited or inaccurate experimental data were analyzed in detail. Furthermore the present work discusses how oversimplified model assumptions affect the interpretation of experimental results and exposes pitfalls in the interpretation of parameter estimation results.

Keywords:

Parameter estimation, crystallization, population balance equation, objective function, erroneous data, inverse problem, simulation

1. Introduction and Objectives

Crystallization processes are crucial for the efficient manufacturing of many solid pharmaceutical drug products. The objective of this step is the formation of a solid product. Yet, in many cases sequential crystallization-dissolution cycles are used as a means of purification. Crystallization has been studied for more than hundred years by a broad scientific community. However, the modeling of crystallization processes is still somewhat complicated by the fact that the crystal properties, such as size, shape, morphology, habit or purity, are typically distributed. Nevertheless, many approaches have been reported in the literature [1,2]. Models are frequently used in the design, optimization and scale-up of a crystallization process. In addition, the increasing demand for model-based control algorithms requires the development of accurate, yet fast models that can predict the impact of system and processing conditions on the quality attributes of the crystals, such as the particle size distribution or shape [3]. Moreover, with increasing global competition there has been a demand for implementing “optimal” process systems engineering approaches to pharmaceutical manufacturing [4]. This also is in line with the Quality by Design approach, introduced by the International Conference on Harmonization (ICH) and increasingly adopted my national or international agencies such as the FDA or EMA.

Owing to the discrete nature of crystals, each crystal being unique in some sense, models that can account for the distribution of properties are required. The ability of population balance equations (PBEs) to capture the discrete and distributed nature of crystals makes their use an appropriate choice for prediction, control and optimization purposes. PBEs are hyperbolic partial differential equations

describe how the populations of specific properties evolve over time. The generalized PBE for spatially homogenous processes as previously proposed by [5] can be written as:

$$\frac{\partial}{\partial t}n(x,t)+\frac{\partial}{\partial x}\left(n(x,t)\frac{\partial x}{\partial t}\right)=B(x,t)-D(x,t) \qquad \text{Equation 1}$$

where n denotes the population of particles, e.g. the number frequency of property x, at the time t. x is the internal coordinate (typically the size) of interest and its derivative with respect to time defines the growth rate $G=\frac{\partial x}{\partial t}$. In case of x being the size G is a given function of supersaturation and other physicochemical properties. The right side of Eq.1 comprises the source terms (e.g., describing aggregation, breakage and nucleation).

The use of PBEs for the modeling of crystallization processes is well-established in the literature [6-11]. Significant work involving 1-D models is due to [6,11-16]. Multi-dimensional PBEs involving the implementation of multidimensional equations were reported as well, with respect to different length scales [17,18] or volume and surface area as the internal coordinates [19,20]. Various solution techniques have been reported including the method of moments [21,22], method of classes [23], high-resolution algorithms [17] or Monte Carlo techniques [24]. In the modeling of crystallization processes, the PBE is coupled with a mass and energy balance providing information about the level of supersaturation and the temperature.

For the aggregation and breakage terms, typically empirical kernels with multiple tunable parameters are used [25,26]. Mechanistic models for defining the kernels are rarely applied. Instead, kernel parameters are optimized to a give best fit of experimental data. This highlights the need for developing effective parameter estimation (PE) techniques for the accurate representation of crystallization processes via PBEs.

Several optimization algorithms for the accurate estimation of empirical parameters have been proposed in the last years. The implementation of a weighted least-square objective function comprising of the estimated and the experimental data was reported in [27-29]**.** Hu et al. considered a maximum likelihood function for the estimation of parameters using the concentration and CSD, and the covariance matrix for testing the robustness of their approach [30]. A similar objective function was considered by [31] for the parameter estimation procedure in combination with the SQP (sequential quadratic programming) algorithm. SQP has also been applied to the parameter estimation of crystallization processes by [32]. A projection method and a Bayesian approach were utilized by [33,34] for the parameter estimation of a multidimensional PBE model for granulation. Gradient-based methods were employed for the fitting of experimental concentration and CSD data using splines by [35]. Optimal control problems in crystallization were previously studied by [36,37] using the particle-swarm optimization (PSO). PSO algorithms were also used for the control and optimization in various other fields (such as fermentation, liquid-liquid extraction, bioinformatics) [38-40]. A comparison of various population-based (e.g. Particle Swarm and Generic Algorithms) and trajectory methods (e.g. Simulated Annealing, Tabu Search and Iterated Local Search) for the optimization of biomass power plants was reported by [41]. Parameter estimation methods by derivate-free and gradient methods within a multi-

dimensional population balance framework were reported by [42]. The authors also discussed the calculation of derivative information and sensitivity analysis to ensure increased accuracy. Dynamic optimization algorithms (such as in gPROMS) for the parameter estimation of PBEs for crystallization processes have been applied as well [43,44].

In this work, we report a detailed study of the estimation of empirical parameters in a PBE models for a crystallization process. Other goals are to test the PE procedure for the amount and quality of the considered experimental data and investigate the consequence of incorrect modeling assumptions. We have used a comparative study with global optimization techniques, including Simulated Annealing Algorithm (SAA), Particle Swarm Optimization (PSO) and the parameter estimation algorithm implemented in gPROMS to investigate the PE procedure for the determination of the growth- and aggregation kinetics.

2. Process Model Development and Simulated Experiments

The concept of this study is to use the inverse problem for the investigation of the PE procedure. Here inverse problem denotes the attempt to estimate the parameters in a PBE model, which have been defined beforehand, from data obtained by calculations of the same PBE model. To begin with, a process model was developed that describes a crystallization process using growth and aggregation kinetics that are as generic as possible. Next, simulations were performed using different process settings to generate a variety of data points. Finally it was tried to obtain the parameters used in the equations of the growth and aggregation rate of the PBE model, based on the results of these simulated experiments.

2.1 Process model

The process model describes a well mixed, seeded batch cooling crystallization process of a model active pharmaceutical ingredient (API) from a model solvent. A schematic draft of this simplified process is shown in Figure 1. Initially the seed crystals are suspended in a saturated solution composed of the solvent and dissolved API. If the solubility of the API goes with the temperature, the cooling of this slurry results in supersaturation, i.e. the solution contains more of the dissolved API than in the saturated state. The level of supersaturation is assumed to remain below the meta-stable limit, i.e. the level that defines the onset of nucleation, throughout the entire process. Hence only the seeded crystals grow.

A Nyvlt model (Eq.2) was used to define the solubility of the model API in the model solvent.

$$\log({X_{API}}^{*}) = N_1 + \frac{N_2}{T} + N_3 \cdot \log(T) \qquad \text{Equation 2}$$

Here $N_1 = 27.769$, $N_2 = -2500.906$, $N_3 = -8.323$ and the temperature T is given in Kelvin. X_{API} is the mole fraction of the solution defined as the ration of dissolved API molecules (n_{API}) to those present in the entire solution ($n_{API} + n_{Solvent}$).

$$X_{API} = \frac{n_{API}[mol]}{n_{API}[mol] + n_{Solvent}[mol]} \qquad \text{Equation 3}$$

The used Nyvlt model was presented by Maia et al. [45] to quantify the temperature dependency of acetylsalicylic acid in ethanol. All model relevant material constants and process settings used to calculate the amount of dissolved and crystalline API are described in

Table **1**.

The supersaturation S is defined as

$$S = \frac{c\ [mol/L]}{c^*[mol/L]} \qquad \text{Equation 4}$$

, where c^* is solubility and c the current concentration of dissolved API. All cooling crystallization processes considered in this work start with a saturated solution, i.e. $S = 1$. Because of this Eq. 3 defines the initial concentration of the dissolved API. The conversion from mole fraction X_{API} to moles per liter, $c\ \left[\frac{mol}{L}\right]$, was conducted by means of Equation 5 to obtain the initial concentration ($\mathrm{c_{API\ initial}}$).

$$\mathrm{c_{API\ initial}\ [mol/L]} = X_{cp}(T_{initial}) \cdot \underbrace{\left(\frac{\varrho_{Solution}}{MW_{API} \cdot (X_{API}(T_{initial})) + \left(1 - X_{API}(T_{initial})\right) \cdot MW_{Solvent}}\right)}_{\text{total amount of moles in one liter saturated solution}} \qquad \text{Equation 5}$$

Besides, only experiments with the same amount of API and solvent (see

Table **1**) were considered. Thus the initial seed mass ($\mathrm{m_{API\_cp\_initial}}$), i.e. the initially non dissolved amount of API being present in crystalline form can be obtained from the total mass of API ($\mathrm{m_{API\_tot}}$) in the system (see Equation 6).

$$\mathrm{m_{API\_cp\_initial}\ [kg]} = \mathrm{m_{API\_tot}} - \underbrace{\mathrm{c_{API\_initial} \cdot MW_{API}}}_{\mathrm{m_{API\_Solution\_initial}}} \qquad \text{Equation 6}$$

Table 1: Material constants and process settings

Material constants	Value	Description
Molecular weight API	$MW_{API} = 180.16\ [kg/mol]$	Molecular weight of acetylsalicylic acid (Aspirin®)
Molecular weight solvent	$MW_{Solvent} = 46.07\ [kg/mol]$	Molecular weight of ethanol
Density crystalline phase	$\varrho_{cp} = 1\ [kg/L]$	For simplification an equal density for the crystalline phase ϱ_{cp}, the pure solvent $\varrho_{Solvent}$ and the solvent containing dissolved species of the API $\varrho_{Solution}$ is assumed. Hence the reactors filling level remains constant.
Density pure solvent	$\varrho_{Solvent} = 1\ [kg/L]$	
Density solvent with dissolved species	$\varrho_{Solution} = 1\ [kg/L]$	
Reactor volume	$V_r = 10^{-3}\ [m^3]$	In the present work we used the same initial mass of solvent and the API (solid & dissolved) for all simulated experiments
Overall mass of solvent in the reactor	$m_{Solvent\_tot} = 0.5\ [kg]$	
Overall mass of API in the reactor	$m_{API\_tot} = 0.5\ [kg]$	
initial mass of the API in crystalline phase	$m_{API\_cp\_initial}$	
initial mass of the API in solution	$m_{API\_Solution\_initial}$	

Crystals were assumed to be cubic. In this work $n(L,t)$ is defined as the number of (cubic) crystals in the reactor volume V_r (see

Table **1**) with an edge length of L at the time t. For all the simulated experiments an equal CSD of seeded crystals was assumed, given by a logarithmic distribution [46] (see Equation 7),

$$n(L,t=0) = N \cdot \frac{1}{\mathrm{L} \cdot \sigma_{\ln} \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{\ln\left(L/L_{50}\right)}{\sigma_{\ln}}\right)^2\right) \qquad \text{Equation 7}$$

with $\sigma_{\ln} = 0.4$, $L_{50} = 100\ \mu m$ and N is used to adjust the total amount of particles. The latter can be obtained from the following constraint:

$$\int_0^{\mathrm{L_{max}}} n(L,t=0) \cdot L^3 \cdot \varrho_{cp} = \mathrm{m_{API\_cp\_initial}} \qquad \text{Equation 8}$$

For the present work a size-independent growth rate G was assumed [12,15,47-49]:

$$G(S,T) = k_{g1} \cdot \exp\left(\frac{-k_{g2}}{R \cdot T}\right) \cdot (S-1)^{k_{g3}} \qquad \text{Equation 9}$$

with parameters $\mathrm{k_{g1}} = 10\ [\frac{m}{s}]$, $\mathrm{k_{g2}} = 10^4\ [\frac{J}{mol}]$, $\mathrm{k_{g3}} = 1\ [-]$. Since the growth rate is size independent, the PBE can be expressed as:

$$\frac{\partial}{\partial t} n(L,t) + G \cdot \frac{\partial}{\partial L}\big(n(L,t)\big) = B(L,t)_{Birth} - D(L,t)_{Death} \qquad \text{Equation 10}$$

Table 2 compares the growth rate of our model API, determined by Equation 1, to growth rates determined for other substances at a level of supersaturation of $S \approx 1.5$. **Why do you use a growth rate that is so different from aspirin when just above you are mentioning that you are looking at aspirin crystallization. → um die model substance "as generic as possible" zu lassen habe ich eine langsamere Wachstumsrate gewählt. Außerdem ergeben sich so für die parameter** k_g "schöne" Werte.

The mass balance equation determining the temporal change in concentration, in $\left[\frac{\mathrm{mol}}{\mathrm{L \cdot s}}\right]$ using the units given in Table 1, G $\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$ and L [m], is given by Equation 11.

$$\frac{\partial c_{API}}{\partial t} = -\frac{3 \cdot \varrho_{cp}}{MW_{API}} \cdot G \cdot \int_0^{\infty} n(L) \cdot L^2 \cdot dL \qquad \text{Equation 11}$$

The aggregation birth and death terms for binary aggregation are [5]:

$$B(L,t) = \frac{L^2}{2} \cdot \int_{\lambda=0}^{L} \frac{\beta_{agg}\left(\lambda, \sqrt[3]{L^3-\lambda^3}\right) \cdot n(\lambda,t) \cdot n\left(\sqrt[3]{L^3-\lambda^3}, t\right)}{\sqrt[3]{(L^3-\lambda^3)^2}}\, d\lambda \qquad \text{Equation 12}$$

$$D(L,t) = n(L,t) \cdot \int_{\lambda=0}^{L} \beta_{agg}(L,\lambda) \cdot n(\lambda,t)\, d\lambda. \qquad \text{Equation 13}$$

The likelihood for the occurrence of aggregation is determined by the aggregation kernel β_{agg} defined in Eq. 14. In this work, a growth rate (and therefore supersaturation) dependent kernel was used. This is the common approach modeling aggregation during crystallization processes [26,50,51]. The size dependence of the used aggregation kernel is similar to the Thomson kernel [52,53] with the difference that $\beta_{agg}(L_1, L_2 = L_1) = 0$ in the original kernel from Thomson.

$$\beta_{agg}(L_1,L_2) = k_{a1} \cdot G^{k_{a2}} \cdot \left(\frac{\left({L_1}^3 - {L_2}^3\right)^2}{\left({L_1}^3 + {L_2}^3\right)} + 2 \cdot \left({L_1}^2 + {L_2}^2\right) \right) \qquad \text{Equation 14}$$

The aggregation model parameters were fixed at $k_{a1} = 3 \cdot 10^{-15} \left[\frac{1}{L}\right]$ and $k_{a2} = 2\,[-]$, which led to a decrease in the total number of particles in the range of 10 % for the described experiments (i.e. simulations). Despite the multitude of reports on aggregation kernels for PBE describing crystallization processes, the number of publications presenting model validation or calibration by means of real experimental data is fairly limited. Generally accepted aggregation kernels do not exist. A summary of reported crystallization models that take into account aggregation is presented in the Appendix.

Table 2: Growth rate at a level of supersaturation value of $\mathbf{S \approx 1.5}$

Substance	Temperature [°C]	Value [m/s]
model API (Eq. 5 & Table 2)	$T = 40$	$G \approx 2 \times 10^{-7}$
L-Glutamic Acid [54]	$T = 70$	$G \approx 3 \times 10^{-8}$
Ibuprofen [47]	$T = 25$	$G \approx 1 \times 10^{-7}$
acetylsalicylic acid [15]	$T = 40$	$G \approx 3 \times 10^{-6}$

2.2 Simulated experiments

Five different cooling crystallization experiments with different initial concentrations ($\mathbf{c_{API\_initial}}$), seed mass ($\mathbf{m_{API\_cp\_initial}}$) and cooling rates ($\mathbf{col.rate}$) are simulated, to generate the data necessary for the PE of used model parameters. The process time ($\boldsymbol{t_{process}}$) was set to end the cooling crystallization process at $\mathbf{25\,°C}$. The process settings of all simulated experiments are listed in Table 3. Note that $\mathbf{c_{API\_initial}}$ and $\mathbf{m_{API\_cp\_initial}}$ depend only on the initial temperature $\boldsymbol{T_{initial}}$ since $\boldsymbol{m_{sol\_tot}}$ and $\boldsymbol{m_{sol\_tot}}$ are equal for each experiment (see Table 1). The supersaturation profile as well as the final and initial CSD of experiment 5 (see Table 3) are shown in Figure 2.

The rate-limiting factor in crystal growth (i.e., surface integration or diffusion to the surface), depends on the molecule itself, the solubility in the used solvent and the level of supersaturation [50]. The growth mechanism and therefore its mathematical representation could switch from surface-integration-limited to diffusion-limited growth if the super saturation increases. Therefore, experiments were designed not to exceed a level of supersaturation of $\boldsymbol{S = 1.5}$, in order to maintain conditions with surface integration being the limiting factor.

Table 3: Process settings of simulated experiments

Experiment	$\mathbf{T_{initial}}$ [°C]	$\mathbf{m_{API\_Solution\_initial}}$ [kg]	col. rate [K/s]	$\mathbf{t_{process}}$ [s]
1	45	0.348	0.005	4000
2	45	0.348	0.025	800
3	55	0.428	0.025	1200
4	55	0.428	0.005	6000
5	50	0.388	0.015	1666

3. Optimization Algorithms

This section briefly describes the global optimization techniques which have been applied. To begin with, boundaries for the parameter space in which these algorithms search for the optimum set of parameters are defined. In the case of the growth rate parameters, the boundaries were set to: $0 < k_{g1} < 500\ \left[\frac{m}{s}\right]$, $0 < k_{g2} < 100000\ \left[\frac{J}{mol}\right]$ and $0 < k_{g3} < 10\ [-]$. For the aggregation kernel parameters, the parameter space was restricted to $10^{-16} < k_{a1}\left[\frac{1}{L}\right] < 10^{-14}$ and $1 < k_{a2}[-] < 5$.

3.1 Decision criterion for growth rate model parameters

To avoid excessive computational effort, unrealistic growth rate parameters have not been evaluated (not accepted in the case of SAA, or not made g_{best} or b_{best} in the case of PSO). To test if a combination of k_{g1}, k_{g2}, k_{g3} is unrealistic, the growth rate G (Eq. 9) for a level of supersaturation of $S = 1.1$ and a temperature of $T = 30\ °C$ was calculated first and compared to an estimation of the growth rate $G_{estimation}$. These numbers values were selected to fairly describe the average process conditions. $G_{estimation}$ was roughly estimated from the average crystal size of the seeds ($\overline{L_{seed}}$) and of the product crystals ($\overline{L_{product}}$) at the end of the process time ($\mathrm{t_{process}}$).

$$G_{estimation} = \frac{\overline{L_{seed}} - \overline{L_{product}}}{\mathrm{t_{process}}} \qquad \text{Equation 15}$$

An estimation of the growth rate from the simulated experiments as described above yields a value

of $G_{estimation} \approx 4 \times 10^{-4}$ $[\frac{\mu m}{s}]$. Before the model was executed by the global optimization algorithms for a newly generated set of parameters ($\boldsymbol{k_{gn}}$), these parameters were tested by Eq. 16.

$$G(S = 1.1, T = 30|\boldsymbol{k_{gn}}) \cdot \frac{1}{5000} < G_{estimation} > G(S = 1.1, T = 30|\boldsymbol{k_{gn}}) \cdot 5000 \qquad \text{Equation 16}$$

The model was not executed if this constraint had been violated.

3.2 Objective function

In case of the stochastic algorithms, i.e. the SAA and PSO, the estimation problem was minimized according to the objective function

$$\Phi(\boldsymbol{x}) = \sum_{i=1}^{a} \sum_{j=1}^{b} \omega_{ij} \cdot \left(\hat{Y}_{ij} - Y_{ij}\right)^2 \qquad \text{Equation 17}$$

where $\hat{Y}_{ij}$ and Y_{ij} are the measurements and model predictions of the j^{th} measured variable at the i^{th} experiment, ω_{ij} is a weighting factor, b is the number of measured variables and a is the number of experiments.

Frequently the concentration trajectory, $c_{API}(t)$, is used to determine growth kinetics, since the measurements are more easily accessible compared to the CSD and numerous analyzers provide online data with no sampling required. Therefore, parameters of the growth function (see Eq. 9) are estimated through concentration data in the objective function, i.e. $\hat{\mathrm{Y}}_{\mathrm{ij}} = \mathrm{c}_{\mathrm{API\ i,j}}\ [\mathrm{mol/L}]$. For each simulated experiment (see Table 3) 200 concentration values were recorded, hence $\mathrm{b} = 200$. All recorded data points were weighted equally $\omega_{\mathrm{ij}} = \frac{1}{\mathrm{a} \cdot \mathrm{b}}$.

To estimate the parameters of the aggregation kernel (see Eq. 14) the CSD data were used exclusively since the concentration trajectory is comparatively hardly affected by aggregation. The product CSD is usually determined most easily. Therefore the normalized product CSD represented by 200 values was used for the objective function, i.e. $\hat{\mathrm{Y}}_{\mathrm{ij}} = \mathrm{n}(\mathrm{L_{i,j}}) \left[\% \frac{\mathrm{particles}}{\mathrm{L}}\right]$ and $\mathrm{b} = 200$. Again, recorded data points were weighted equally $\omega_{\mathrm{ij}} = \frac{1}{\mathrm{a} \cdot \mathrm{b}}$.

3.3 Simulated Annealing Algorithm (SAA)

Simulated annealing is an algorithm that originated in material science engineering, originally introduced to find the equilibrium configuration of a collection of atoms at a given temperature when a liquid freezes and crystallizes during annealing process [55]. Kirkpatrick et al. [56] initially proposed to use this theory for application to optimization problems. For the SAA an artificial temperature T_{SA} and a probability distribution $p_E(\boldsymbol{x}|T_{SA})$ are introduced. The latter is the so called acceptance distribution for

the objective function Φ at the configuration $\boldsymbol{x}$ for a given artificial temperature T_{SA}. According to classical SAA $p_E(\boldsymbol{x}|T_{SA})$ was defined as:

$$p_E(\boldsymbol{x}|T_{SA}) = \frac{1}{Z} \cdot e^{\frac{-\Phi(\boldsymbol{x})}{T_{SA}}} \qquad \text{Equation 18}$$

which corresponds to a Boltzmann distribution with a normalization factor Z. The purpose of p_E is to assign a higher acceptance probability to states with a smaller objective function value, e.g., states $\boldsymbol{x}$ close to the global minimum. The feasibility that a new state $\boldsymbol{x}$ can get accepted, even if the objective function's value is higher than at the previous state, makes the SAA a global optimization routine. The likelihood for accepting inferior states, i.e., to go uphill and to leave a local minimum is dependent on T_{SA}. Classical simulated annealing (CSA) involves three steps.

Generation of states

The new states $\boldsymbol{x_n}$ in the parameter space with the distance $\Delta\boldsymbol{x}$ to current one $\boldsymbol{x_c}$ are generated from the latter according to a Gaussian distribution.

$$p(\Delta\boldsymbol{x}(i) = \boldsymbol{x_c}(i) - \boldsymbol{x_n}(i)) = \frac{1}{\boldsymbol{\sigma}(i) \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\Delta\boldsymbol{x}(i)}{\boldsymbol{\sigma}(i)}\right)^2} \qquad \text{Equation 19}$$

Acceptance of states

The Metropolis probability (see Eq. 20) was chosen to accept or deny the generated new states.

$$p_{accept} = min\left\{1, \frac{p_E(\boldsymbol{x_n}|T_{SA})}{p_E(\boldsymbol{x_c}|T_{SA})}\right\} \qquad \text{Equation 20}$$

Cooling strategy

In the present work the used cooling scheme is

$$T_{SA} = T_{SA0} \cdot q^m \qquad \text{Equation 21}$$

The initial temperature T_{SA0} is of major importance. A high T_{SA0} would lead to high computational effort, since almost every new state can get accepted independent of its objective function value. In the case of a too low T_{SA0} the walker might be trapped in a local minimum. The tunable parameters q, m, σ_i, as well as the initial values for $\boldsymbol{x}_{initial}$, used for the calculations are documented in the results section.

If none of the states $\boldsymbol{x}_n$ are accepted for 200 suggestions, the algorithm is assumed to have reached convergence. If the objective function was observed to be lower during the computation at some other

point in the parameter space, the latter is updated as the new state. A detailed description of SAA in not within the scope of the present work and interested readers might be referred to the literature [56-58].

3.4 Particle Swarm Optimization (PSO) Algorithm

The algorithm was initially proposed by [59], as a methodology for the optimization of nonlinear functions. Various modifications have been made to the original algorithm by researchers in order to make the method more suitable for their system. Although, population based method like the PSO are computationally expensive owing to the multiple function evaluations involved, these function evaluations can be later utilized for statistical analysis.

The method tries to mimic the flocking of birds and is based on the swarming theory. In case of PSO, a bird is an element moving in the parameter space where it calculates the corresponding objective function. The algorithm is based on the synchrony of the flocking behavior and depends significantly on the inter-individual distances between the birds and their neighbors. The "birds" are set to fly in the search domain, i.e. the parameter space, according to

$$\begin{aligned} {\boldsymbol{v}_b}^{t+1} &= w \cdot {\boldsymbol{v}_b}^{t} + c_1 \cdot rand_1 \cdot (\boldsymbol{b}_{best} - {\boldsymbol{x}_b}^{t}) + c_2 \cdot rand_2 \cdot (\boldsymbol{g}_{best} - {\boldsymbol{x}_b}^{t}) \\ {\boldsymbol{x}_b}^{t+1} &= {\boldsymbol{x}_b}^{t} + {\boldsymbol{v}_b}^{t+1} \cdot \Delta t \end{aligned} \qquad \text{Equation 22}$$

Here b labels a bird, t is the number of iterations, v_b and x_b are the velocity and position of the bird respectively, c_1 and c_2 are called the cognitive and social parameters, w is the inertial weight which was an inclusion into the algorithm by [60], $rand_1$ and $rand_2$ are random numbers, $\boldsymbol{b}_{best}$ is the best known position $\boldsymbol{x}$ having the lowest objective function $\Phi(\boldsymbol{x})$ the bird itself and $\boldsymbol{g}_{best}$ is the best known position of the entire particle (bird) swarm.

Initially N birds are distributed randomly in the parameter space with a random initial velocity using a maximum value of a fourth of the parameter spaces elongation per time step ($\Delta t = 1$). Birds leaving the parametric space were updated according to the "mirror back" concept in which the bird is mapped back into the parameter space with a sign inversion of its velocity. After each iteration, $\boldsymbol{b}_{best}$ or $\boldsymbol{g}_{best}$ were updated. The striking characteristic of this algorithm is its aspect of having the birds spread all over in the search domain during the initial iterations and then have the birds more concentrated to the regions which seems more promising. This way the algorithm tends to be more convergent towards the later iterations. An initial value like in the case of the SAA is not required here.

3.5 gEST using gPROMS

The dynamic optimization algorithm implemented in the gPROMS software using SQP is increasingly used due to its user-friendly operability. The choice of using a maximum likelihood function or the least square objective function is available in this parameter estimation platform. If $\hat{Y}_{ijk}$ is the measured variable and Y_{ijk} is the predicted variable, the maximum likelihood function can be defined as

$$\Phi(k,\theta) = \frac{M}{2} \cdot \ln(2\pi) + \frac{1}{2} \cdot min\left\{ \sum_{i=1}^{a} \sum_{j=1}^{b_i} \sum_{k=1}^{c_{i,j}} \left[\ln\left(\sigma_{ijk}^2\right) + \frac{\left(\hat{Y}_{ijk} - Y_{ijk}\right)^2}{\sigma_{ijk}^2} \right] \right\} \qquad \text{Equation 23}$$

Where M is the total number of measurements taken during all experiments, a, b_i, $c_{i,j}$ are the number of experiments, number of variables in the i^{th} variable and number of measurements in the i^{th} experiment of the j^{th} variable respectively. σ_{ijk} is the variance of the k^{th} measurement of the j^{th} variable in the i^{th} experiment and can be defined by many variance models such as a homoscedastic (constant variance) or a heteroscedastic model (variance is a function of the measured and the predicted values). In case of the homoscedastic variance model, the maximum likelihood function is simplified to a weighted least square form whereas when a purely heteroscedastic model is chosen, the maximum likelihood objective function takes a more complicated form with the variance, σ, as a function of the measured value raised to a parameter γ (which lies between 0 and 1). The two asymptotic conditions of the parameter, γ leads to the constant variance ($\gamma = 0$) and the constant relative variance ($\gamma = 1$) cases.

In the implementation any of the abovementioned variance models can be chosen and the other parameters can also be specified by the user. However, the homoscedastic model with a constant or constant variance models was applied in this work. Since our measured/known quantity is generated through simulations, the data are devoid of much variance, thus making it affordable to choose a more simplistic variance model for accurate estimation.

3.6 Quality criterion for PE results

A single value $\hat{G}_{error}$ is introduced to quantify the discrepancy between the original growth rate G_{model} (see Eq. 9) and the ones calculated from parameters determined by the PE routine $\mathrm{G_{PE}}$. $\mathrm{\hat{G}_{error}}$ is defined as the integral of the difference between G_{model} and G_{PE} over temperature ($T[30°C - 50°C]$) and supersaturation ($\mathrm{S}[1.1 - 1.5]$) values that are characteristic for the modeled process. Figure 3 illustrates definition of $\hat{G}_{error}$ formulated in Eq. 24.

$$\hat{G}_{error} = \int_{1.1}^{1.5}\int_{30}^{50} |G_{model}(S,T) - G_{PE}(S,T)| \cdot dT[°C] \cdot dS[-] \qquad \text{Equation 24}$$

Moreover, we define a similar quality criterion for the estimated aggregation kernel parameters as well. This is given by:

$$\hat{\beta}_{error} = \int_{1.1}^{1.5}\int_{30}^{50} \left| k_{a1\,model} \cdot G_{model}(S,T)^{\,k_{a2\,model}} - k_{a1\,PE} \cdot G_{model}(S,T)^{\,k_{a2\,PE}} \right| \cdot dT[°C] \cdot dS[-] \qquad \text{Equation 25}$$

4. Numerical procedure

A large number of solution techniques for PBEs have been reported in the literature [17,22,24,61-65]. Here the method of classes [9,66] was used. It is still one of the most frequently applied solution technique for PBE describing crystallization processes for its ease of coding. Furthermore, growth rates and aggregation/breakage kernels do not have to match mathematical restrictions as it is required for other solution techniques e.g. for the method of moments.

A nonlinear grid of 200 classes ranging from $L_{min} = 1\ \mu m$ to $L_{max} = 1000\ \mu m$ was used for the discretization procedure. For the first 170 classes a linear grid was chosen, i.e., $(L_{i+2} - L_{i+1}) = (L_{i+1} - L_i)$, whereas for the last 30 classes the increment was nonlinear $(L_{i+2} - L_{i+1}) = 1.1 \cdot (L_{i+1} - L_i)$. The cell average algorithm [67-70] was applied to calculate the birth and death terms for aggregation (Eq. 9, 11, 12).

All differential equations, i.e. each class of the discretized PBE and the mass balance, were solved by the MATLAB ode45 solver (utilizing an explicit Runge-Kutta 4th-order method) or by gPROMS (implicit solution).

The SAA (see Eq. 18-21) and PSO (see Eq. 22) were coded within MATLAB using its inbuilt random number generator. MATLAB's simulated annealing optimization routine (*simulannealbnd*, Global Optimization Toolbox) was compared to the described SAA. PE results were the same. A local optimization algorithm was occasionally accompanied with the global optimization algorithms (see 5). In that case the Nelder-Mead method of MATLAB's *fminsearch* was used.

If gPROMS was used for the PE its dynamic optimization routine came to be applied. According to the manual, the latter is based on an SQP algorithm.

5. Results & Discussion

The PE procedure, i.e., minimizing the objective functions Eqs. 17 or 23 using the data of the simulated experiments (see Table 3), was investigated for several cases (section 5.1-5.5). The applicability of different algorithms, including global and local optimization routines, was tested and compared. In order to test the influence of limited experimental data on the PE results, various combinations of the simulated experiments (see Table 3) have been considered: experiments 1, 2, 3 and 4; experiments 1 and 3; and data solely obtained from experiment 5.

5.1 Determining growth function parameters

As described in section 3.2, only the concentration profile ($\widehat{Y}_{ij} = c_{API\ i,j}\ [mol/L]$) was used to estimate the growth function parameters k_{g1}, k_{g2}, k_{g3}.

Table 4 lists the applied settings and initial states for the tested global optimization algorithm and the gPROMS routine. The results of all investigated PE procedures are presented in Table 5 .

Table 4: Constants, initial conditions and settings used within the applied optimization algorithms to determine growth function parameters.

SAA	PSO	gPROMS (SQP)
Initial state: $\boldsymbol{x}_{initial} = [$ $k_{g1} = 1,$ $k_{g2} = 5000,$ $k_{g3} = 2]$ Equation 19 **:** $\sigma_1 = 1$ $\sigma_2 = 1000$ $\sigma_3 = 0.1$ Equation 21: $T_{SA0} = 0.2$ $q = 0.95$ $m = 1$ 100 new states $\boldsymbol{x}_n$ have been evaluated per temperature step. The calculation was stopped after 200 new states $\boldsymbol{x}_n$ were rejected consecutively.	Number of "birds": $N = 20$ Maximum number of iterations if no convergence is not achieved $t_{max} = 100$ Equation $c_1 = 0.5\ ;\ c_2 = 0.5$ $w_{max} = 1\ ;\ w_{min} = 0.1$ $w = w_{max} - \left(\frac{w_{max} - w_{min}}{t_{max}}\right) \cdot t$ The calculation was stopped before maturity if $\boldsymbol{g}_{best}$ had not been updated for 20 steps	Initial state: $\boldsymbol{x}_{initial} = [$ $k_{g1} = 0.1,$ $k_{g2} = 0.1,$ $k_{g3} = 0.1]$ Variance mode: Const and relative variance Range: 0.05-0.15

Table 5: PE results of the growth function parameters using the global optimization algorithms only and accompanied with a Nelder-Mead method as well as the output of gPROMS dynamic optimization routine (original values: $\boldsymbol{k_{g1} = 10}\ [\boldsymbol{m/s}], \boldsymbol{k_{g2} = 10^4}\ [\boldsymbol{J/mol}], \boldsymbol{k_{g3} = 1}\ [-]$).

	SAA		PSO		
	parameters	$\widehat{\mathrm{G}}_{\mathrm{error}}$	parameters	$\widehat{\mathrm{G}}_{\mathrm{error}}$	
Experiment 5	$k_{g1} = 6.4591$ $k_{g2} = 1.4206 \times 10^4$ $k_{g3} = 1.2491$	0.4662	$k_{g1} = 22.9142$ $k_{g2} = 8.8348 \times 10^3$ $k_{g3} = 3.5987$	0.3779	
Experiment 1+3	$k_{g1} = 11.8721$ $k_{g2} = 1.0244 \times 10^4$ $k_{g3} = 1.0701$	0.0107	$k_{g1} = 28.9$ $k_{g2} = 1.8383 \times 10^4$ $k_{g3} = 1.0399$	0.4589	
Experiment 1+2+3+4	$k_{g1} = 12.0201$ $k_{g2} = 1.0421 \times 10^4$ $k_{g3} = 1.0027$	0.0101	$k_{g1} = 43.0210$ $k_{g2} = 1.7491 \times 10^4$ $k_{g3} = 2.0177$	0.4737	
	SAA + Nelder-Mead		SAA + Nelder-Mead		gPROMS (SQP)

	parameters	$\hat{G}_{error} \times 10^{-4}$	parameters	$\hat{G}_{error} \times 10^{-4}$	parameters	$\hat{G}_{error} \times 10^{-4}$
Experiment 5	$k_{g1} = 10.0003$ $k_{g2} = 9.9992 \times 10^{3}$ $k_{3} = 1.0000$	1.7395	$k_{g1} = 10.0002$ $k_{g2} = 9.9992 \times 10^{3}$ $k_{g3} = 1.0002$	0.5895	$\mathrm{k_{g1}} = 16.8921$ $\mathrm{k_{g2}} = 1.1834 \times 10^{3}$ $\mathrm{k_{g3}} = 0.9930$	813.2708
Experiment 1+3	$k_{g1} = 11.0286$ $k_{g2} = 1.0244 \times 10^{3}$ $k_{3} = 1.0001$	21.5971	$k_{g1} = 10.0030$ $k_{g2} = 9.9994 \times 10^{3}$ $k_{3} = 1.0001$	2.5520	$\mathrm{k_{g1}} = 5.4903$ $\mathrm{k_{g2}} = 1.0231 \times 10^{3}$ $\mathrm{k_{g3}} = 0.6988$	1497.1893
Experiment 1+2+3+4	$k_{g1} = 10.0001$ $k_{g2} = 9.9993 \times 10^{3}$ $k_{g3} = 1.0001$	0.8051	$k_{g1} = 10.0030$ $k_{g2} = 9.9993 \times 10^{3}$ $k_{g3} = 1.0001$	2.3538	$\mathrm{k_{g1}} = 9.1317$ $\mathrm{k_{g2}} = 1.0422 \times 10^{3}$ $\mathrm{k_{g3}} = 0.9466$	900.3417

Initially, the parameters were determined using the SAA and PSO only. The growth function parameters obtained by the SAA deviate less from the original parameters. Since the SAA requires an initial value, the lower discrepancies can be attributed to a good initial guess. Results obtained by means of the PSO show a vast difference for all the considered sets of simulated experiments, although their single error values (see Eq. 24) are comparable. This deviation in the results was observed as well for various runs of the PSO, which is not shown in here. The number of considered experiments had a significant effect on the quality of estimated parameters only in the case of the SAA (without Nelder-Mead).

Nevertheless, the results show that none of the applied global optimization algorithms was able to find the original parameters. For this reason both algorithms were accompanied with a local optimization algorithm. Here, the Nelder-Mead method came to be applied, starting with the final output of the global optimization algorithms. As a result, the PE obtained the correct growth function parameters for all for all the considered sets of simulated experiments.

The gPROMS optimization routine was able to determine parameter sets with low $\hat{G}_{error}$ values for all the considered sets of experiments. However, the optimization routine was able to almost reproduce the original growth function parameters only by using the experiments 1-4.

The shape of the objective functions was investigated as well. To do so, the third growth rate parameter was fixed to its optimal value $\mathrm{k_{g3}} = 1\ [-]$ and the objective function was evaluated within a fraction of the investigated parameter space for $\mathrm{k_{g1}}$ and $\mathrm{k_{g2}}$ containing their optimal values. The results are shown in Figure 4a and b. Within the investigated fraction of the parameter space, both evaluated objective functions exhibit a bowl shaped regime with only one local minimum which is as well the global minimum, i.e., the optimum solution since $\Phi\left(\mathrm{k_{g1}} = 10, \mathrm{k_{g2}} = 10^{4}, \mathrm{k_{g3}} = 1\right) \stackrel{\text{def}}{=} 0$. The objective function shows a band of $\left(\mathrm{k_{g1}}, \mathrm{k_{g2}}\right)$ combinations differing only slightly from the global minimum. This explains the vast difference in the PE results obtained by means of the PSO. Because the PSO is a population based optimization algorithm, it can easily converge anywhere in the described band. A logarithmic function was fitted through this band for additional studies (see section 5.2).

Owing to the objective function's smooth shape (Fig. 4a, b) any local optimization method and especially gradient based methods are suitable to find the optimum solution for the considered PE procedure.

5.2 Determining growth function parameters for noisy data

In order to investigate the effect of erroneous data on the PE results, random noise was added to the simulated and therefore ideal data according to Equation 26:

$$\hat{Y}_{ij\,noise} = \hat{Y}_{ij} + \frac{\hat{Y}_{ij}}{100} \cdot randn_{ij} \cdot \sigma_{noise} \qquad \text{Equation 26}$$

Here $\hat{Y}_{ij}$ are the experimental data points (i.e. the concentration profile) and $randn_{ij}$ is a random value from a standard normal distribution. σ_{noise} defines the intensity of noise added. Figure 5 shows an example of the original (smooth) and noisy data. The same set of random numbers has been used for all cases. Hence 200 random numbers were generated once, for the 200 concentration profile values of each experiment.

To identify the extent to which the superimposed error affects the PE results, we evaluated the objective function along the band of $(\mathrm{k_{g1}}, \mathrm{k_{g2}})$ combinations yielding values differing only slightly from the global minimum (here $\phi(\mathrm{k_{g1}} = 10, \mathrm{k_{g2}} = 10^4, \mathrm{k_{g3}} = 1) > 0$ due to the fact that we consider noisy data). Figure 6 compares the objective function along this band for various σ_{noise} levels.

The results reveal that the minimum of the objective function is shifted with an increasing level of noise, i.e., at higher σ_{noise} values. The shift was suppressed by taking into account a larger variety of experimental data, i.e., more experiments. This is in agreement with the PE results shown in Table 6. Only the global optimization algorithms accompanied with a Nelder-Mead method were evaluated, since they obtained the best results in the case of ideal data (see section 5.1). Table 6 demonstrates that noisy data significantly reduce the quality of the PE results. High levels of noise make it impossible to obtain the original growth function parameters used in the simulations. The incorporation of more experiments has been shown to be beneficial for all investigated noise levels. In addition the Nelder-Mead algorithm was applied alone. The obtained parameters exhibited considerably higher $\hat{G}_{error}$ values especially in the case of high noise levels.

Table 6: PE results of the growth function parameters using the global algorithms accompanied with a Nelder-Mead method, for noisy data (original values: $\mathrm{k_{g1}} = 10\ [\mathrm{m/s}], \mathrm{k_{g2}} = 10^4\ [\mathrm{J/mol}], \mathrm{k_{g3}} = 1\ [-]$).

		SAA + Nelder-Mead		**PSO + Nelder-Mead**	
		parameters	$\hat{G}_{error}$	parameters	$\hat{G}_{error}$
$\sigma_{noise} = 3$	Experiment 5	$k_{g1} = 0.2649$ $k_{g2} = 1.2960 \times 10^3$ $k_{g3} = 0.8666$	0.0696	$k_{g1} = 0.2839$ $k_{g2} = 1.4541 \times 10^3$ $k_{g3} = 0.8741$	0.0679
	Experiment 1+3	$k_{g1} = 2.0893$ $k_{g2} = 5.9037 \times 10^3$ $k_{g3} = 1.0143$	0.0136	$k_{g1} = 2.0893$ $k_{g2} = 5.9037 \times 10^3$ $k_{g3} = 1.0143$	0.0136

	Experiment 1+2+3+4	$k_{g1} = 6.4093$ $k_{g2} = 8.8653 \times 10^3$ $k_{g3} = 0.9960$	0.0041	$k_{g1} = 6.4093$ $k_{g2} = 8.8654 \times 10^3$ $k_{g3} = 0.9960$	0.0041
$\sigma_{noise} = 2$	Experiment 5	$k_{g1} = 0.7871$ $k_{g2} = 3.9038 \times 10^3$ $k_{g3} = 0.9102$	0.0502	$k_{g1} = 0.7871$ $k_{g2} = 3.9038 \times 10^3$ $k_{g3} = 0.9102$	0.0502
	Experiment 1+3	$k_{g1} = 3.3688$ $k_{g2} = 7.1510 \times 10^3$ $k_{g3} = 1.0115$	0.0096	$k_{g1} = 3.3688$ $k_{g2} = 7.1510 \times 10^3$ $k_{g3} = 1.0115$	0.0096
	Experiment 1+2+3+4	$k_{g1} = 7.3194$ $k_{g2} = 9.2003 \times 10^3$ $k_{g3} = 0.9990$	0.0030	$k_{g1} = 7.3194$ $k_{g2} = 9.2003 \times 10^3$ $k_{g3} = 0.9990$	0.0030
$\sigma_{noise} = 1$	Experiment 5	$k_{g1} = 2.6265$ $k_{g2} = 6.7938 \times 10^3$ $k_{g3} = 0.9532$	0.0274	$k_{g1} = 2.6265$ $k_{g2} = 6.7938 \times 10^3$ $k_{g3} = 0.9532$	0.0274
	Experiment 1+3	$k_{g1} = 5.6955$ $k_{g2} = 8.5242 \times 10^3$ $k_{g3} = 1.0068$	0.0051	$k_{g1} = 5.6955$ $k_{g2} = 8.5242 \times 10^3$ $k_{g3} = 1.0068$	0.0051
	Experiment 1+2+3+4	$k_{g1} = 8.4804$ $k_{g2} = 9.5758 \times 10^3$ $k_{g3} = 1.0003$	0.0017	$k_{g1} = 8.4804$ $k_{g2} = 9.5758 \times 10^3$ $k_{g3} = 1.0003$	0.0017

5.3 Determining aggregation kernel parameters

To analyze various PE methods for a model considering aggregation, the presented PBE model was evaluated for the experiments listed in Table 3, including the aggregation model at this time. Two parameters occur in the aggregation kernel (see Eq. 14). As described in section 3.2, only the product CSD was used for the objective function ($\widehat{Y}_{ij} = n(L_{i,j}) \left[\% \frac{\text{particles}}{L}\right]$).

Since the SAA and PSO algorithms in combination with a Nelder-Mead method yielded correct results of the growth function parameters, the same algorithms were applied also for the PE of the aggregation kernel parameters. The used settings for PSA were similar to those listed in

. Only the number of "birds" was decreased ($N = 5$) since the parameter space is two dimensional and the computational effort is higher if aggregation is included. The SAA was started from an initial guess $\boldsymbol{x}_{c\ initial} = [k_{a1} = 5 \times 10^{-16}, k_{a2} = 4]$. New states were generated using standard deviations of $[\sigma_1 = 10^{-15}, \sigma_2 = 0.1]$ in Eq. 19 and $[T_{SA0} = 10^{-7}, q = 90, m = 1]$ was used for the cooling scheme (Eq. 21). The PE results are shown in

. As for the PE of the growth functions parameters using ideal data, both investigated algorithms were able to discover the original set of parameters independently how many experiments have been considered for the objective function.

Figure 7 shows the evaluation of the objective function for a fraction of the considered parameter space. Again, the objective functions exhibit a smooth bowl shaped regime with only one local minimum which is as well the global minimum.

Table 7: PE results of the aggregation kernel parameters using the global algorithms accompanied with a Nelder-Mead method (original values: $k_{a1} = 3 \cdot 10^{-15}\ \left[\frac{1}{L}\right]$, $k_{a2} = 2\ [-]$).

	SAA + Nelder-Mead		**PSO + Nelder-Mead**	
	parameters	$\hat{\beta}_{error}$ $\times 10^{-17}$	parameters	$\hat{\beta}_{error}$ $\times 10^{-17}$
Experiment 5	$k_{a1} = 3.0003 \times 10^{-15}$ $k_{a2} = 2.0000$	0.0018	$k_{a1} = 3.0003 \times 10^{-15}$ $k_{a2} = 2.0000$	0.0018
Experiment 1+3	$k_{a1} = 3.0002 \times 10^{-15}$ $k_{a2} = 1.9999$	0.0037	$k_{a1} = 3.0003 \times 10^{-15}$ $k_{a2} = 1.9999$	0.0041
Experiment 1+2+3+4	$k_{a1} = 2.9997 \times 10^{-15}$ $k_{a2} = 2$	0.0012	$k_{a1} = 2.9998 \times 10^{-15}$ $k_{a2} = 2$	0.0012

5.4 Determining aggregation kernel parameters for noisy data

Similar to the growth function parameters, the effect of noisy data on the PE of the aggregation kernel parameters was investigated. Again, random noise was added to the experimental data via Eq. 26. The PE results for the aggregation kernel parameters determined from noisy data are listed in Table 8.

As observed in the case of the PE for the growth function parameters, the correct parameters could not be determined in the case of highly noisy data. Once more, the incorporation of more experiments has been shown to be beneficial for all investigated noise levels. Thus, the number of experiments used was shown to be of major importance, especially when dealing with highly noisy data.

Table 8: PE results of the aggregation kernel parameters using the global algorithms together with a Nelder-Mead method, for noisy data (original values: $k_{a1} = 3 \cdot 10^{-15}\ \left[\frac{1}{L}\right]$, $k_{a2} = 2\ [-]$).

		SAA + Nelder-Mead		**PSO + Nelder-Mead**	
		parameters	$\hat{\beta}_{error}$ $\times 10^{-17}$	parameters	$\hat{\beta}_{error}$ $\times 10^{-17}$
$\sigma_{noise} = 3.$	Experiment 5	$k_{a1} = 4.8699 \times 10^{-15}$ $k_{a2} = 2.1411$	1.5654	$k_{a1} = 4.8691 \times 10^{-15}$ $k_{a2} = 2.1411$	1.5633

	Experiment 1+3	$k_{a1} = 2.9785 \times 10^{-15}$ $k_{a2} = 1.9968$	0.0086	$k_{a1} = 2.9785 \times 10^{-15}$ $k_{a2} = 1.9968$	0.0086
	Experiment 1+2+3+4	$k_{a1} = 3.0001 \times 10^{-15}$ $k_{a2} = 1.9995$	0.0015	$k_{a1} = 3.0001 \times 10^{-15}$ $k_{a2} = 1.9995$	0.0151
$\sigma_{noise} = 2$	Experiment 5	$k_{a1} = 3.7803 \times 10^{-15}$ $k_{a2} = 2.0671$	0.7270	$k_{a1} = 3.7803 \times 10^{-15}$ $k_{a2} = 2.0671$	0.7270
	Experiment 1+3	$k_{a1} = 2.9891 \times 10^{-15}$ $k_{a2} = 1.9983$	0.0079	$k_{a1} = 2.9891 \times 10^{-15}$ $k_{a2} = 1.9983$	0.0079
	Experiment 1+2+3+4	$k_{a1} = 3.0022 \times 10^{-15}$ $k_{a2} = 1.9999$	0.0144	$k_{a1} = 3.0022 \times 10^{-15}$ $k_{a2} = 1.9999$	0.0144
$\sigma_{noise} = 1$	Experiment 5	$k_{a1} = 3.2238 \times 10^{-15}$ $k_{a2} = 2.0270$	0.0874	$k_{a1} = 3.2240 \times 10^{-15}$ $k_{a2} = 2.0270$	0.0877
	Experiment 1+3	$k_{a1} = 2.9966 \times 10^{-15}$ $k_{a2} = 1.9994$	0.0045	$k_{a1} = 2.9966 \times 10^{-15}$ $k_{a2} = 1.9994$	0.0045
	Experiment 1+2+3+4	$k_{a1} = 3.0019 \times 10^{-15}$ $k_{a2} = 2.0001$	0.0044	$k_{a1} = 3.0020 \times 10^{-15}$ $k_{a2} = 2.0001$	0.0048

5.5 Determining growth function parameters in the presence of aggregation

To investigate how incorrect model assumptions can affect the PE, the growth function parameters were estimated from (simulated) data including aggregation whereas aggregation was not considered in the model.

Since the concentration profile data were used for the objective function every effect causing deviations of this data will have an impact on the PE results. Aggregation leads to a decrease in surface area ($\sim L^2$) of the crystalline phase and is therefore affecting the concentration profile (see Eq. 11). Hence, an objective function considering only concentration profile data, as shown here for the estimation of the growth function parameters can be insufficient. The PE results presented in Table 9 support this hypothesis. Moreover, the involvement of data from multiple experiments could not enhance the PE results.

Figure 8 compares the concentration profiles obtained by using the original PBE's growth rate parameters and those obtained from the PE. The discrepancies are minor, even if the parameters (see Table 9) itself vary significantly. A clear discrepancy could be observed only in the case of experiment 3 which exhibits the highest cooling rate and initially dissolved API.

Based on the presented results, we conclude that the negligence of aggregation in the PE leads to a PBE model that overvalues the concentration values at the end of the crystallization process (see e.g. Figure

8, Exp. 2&3). Since this overvaluation can be attributed to the reduced surface area in the presence of aggregation, the negligence of nucleation and breakage events is expected to exhibit reverse features. These conclusions are only valid in the case of an accurate expression of the growth function.

Table 9: PE results of the growth function parameters using the global algorithms accompanied with a Nelder-Mead method, neglecting the presence of aggregation (original values:$\mathrm{k_{g1}} = 10\ [\mathrm{m/s}], \mathrm{k_{g2}} = 10^4\ [\mathrm{J/mol}], \mathrm{k_{g3}} = 1\ [-]$).

	SAA + Nelder-Mead		**PSA + Nelder-Mead**	
	parameters	$\widehat{G}_{error}$	parameters	$\widehat{G}_{error}$
Experiment 5	$k_{g1} = 225.3350$ $k_{g2} = 1.7919 \times 10^4$ $k_{g3} = 1.0504$	0.0277	$k_{g1} = 225.3350$ $k_{g2} = 1.7919 \times 10^4$ $k_{g3} = 1.0504$	0.0277
Experiment 1+3	$k_{g1} = 42.6436$ $k_{g2} = 1.4192 \times 10^4$ $k_{g3} = 0.9399$	0.0444	$k_{g1} = 150.0286$ $k_{g2} = 1.7420 \times 10^4$ $k_{g3} = 0.9467$	0.0396
Experiment 1+2+3+4	$k_{g1} = 22.7542$ $k_{g2} = 1.2517 \times 10^4$ $k_{g3} = 0.9496$	0.0428	$k_{g1} = 22.7542$ $k_{g2} = 1.2517 \times 10^4$ $k_{g3} = 0.9496$	0.0428

6. Summary and Conclusion

Global optimization algorithms (SAA and PSA) have been applied alone and in combination with a local optimization routine, i.e. the Nelder-Mead method, to determine parameters in the PBE's growth rate expression, from the concentration profile recorded during various crystallization experiments. The global optimization algorithms were shown to be inefficient if applied alone. However, in combination with a local optimization algorithm, the original model parameters could be estimated successfully, regardless of the number of considered experimental data. It was revealed that several sets of highly differing growth rate parameters are able to minimize the objective function reasonably. Since the parameters used in the growth function of a PBE for crystallization processes are frequently linked to physical constants, e.g. surface integration energy E_0 with k_{g2} (see Eq. 8) [12,15], the determination of the latter is hardly possible using a PE procedure as described within this work. In the case of noisy and therefore more realistic data, it was illustrated that the number of experiments used in the PE procedure is of major importance for the identification of the growth function's parameters. Furthermore, it was shown that the exclusive use of local optimization algorithms is insufficient considering noisy experimental data.

In addition, PE of the aggregation kernel parameters was performed. The combination of the SAA and PSA with a Nelder-Mead method gave good results. Parameters of the aggregation kernel could be determined accurately assuming exact data of the CSD. Again, a higher variety of experimental data was shown to be beneficial for the PE (only) in the case of erroneous data.

Beside erroneous data, wrong model assumptions were shown to hinder the PE procedure. As shown in the present work, disregarding aggregation in the PBE model, made it impossible to determine the growth function parameters used for the inverse problem. Not even an increase in the multitude of experimental data was beneficial.

The present work demonstrates the need of versatile experimental data for an adequate PE procedure, especially in the case of inaccurate experimental findings.

7. Appendix

Table A1 contains reported aggregation kernels as well as the investigated crystallization systems. It is not known if listed aggregation kernels are material specific. From this confrontation it becomes clear that, in contrast to the growth model, there is no consistent mathematical formulation for aggregation in crystallization processes. Since the derivation of generic aggregation kernels based on first principles is hardly possible at all, even for simple flow patterns, common approaches remain highly phenomenological. Therefore it might be stated that aggregation modeling during crystallization processes is not fully mature yet.

Table A1: Aggregation models for PBEs in the literature

literature	aggregation kernel	parameters	system	crystal sizes
Lim et al. (2002) [71]	$\beta = \beta_1 \cdot (\rho_s \cdot M \cdot \sigma)^{0.36} \cdot (L_i + L_j)^3$	σ ... relative supersaturation $[-]$ M ... magma density $\left[\frac{\text{kg}}{\text{kg}}\right]$ ρ_s ... solution density $\left[\frac{\text{kg}}{\text{m}^3}\right]$ L_i, L_j $[m]$ $\beta \left[\frac{kg_{solvent}}{crytals \cdot min}\right]$ $\beta_1 = 0.55$	K_2SO_4 / H_2O	≈ 1000 µm
Zauner and Jones (2000) [53]	$\beta = \beta_1(1 + \beta_2 \cdot \sqrt{\varepsilon} + \beta_3 \cdot \varepsilon) \cdot S^{2.15} \cdot (L_i + L_j)^3$	ε ... power input per unit vol. $\left[\frac{\text{W}}{\text{kg}}\right]$ S ... supersaturation $[-]$ L_i, L_j $[m]$ $\beta \left[\frac{kg_{solvent}}{crytals \cdot s}\right]$ $\beta_1 = 5.431 \times 10^{-17}$ $\beta_2 = 2.296$ $\beta_3 = -2.429$	CaC_2O_4 / H_2O & $NaCl$	≈ 20 µm
Rohani (1993) [49]	$\beta = \beta_1 \cdot G^{\beta_2} \cdot B^{\beta_3}$	G ... growth rate $\left[\frac{\text{m}}{\text{s}}\right]$ B ... nucleation rate $\left[\frac{crytals\ born}{kg_{solvent} \cdot \text{s}}\right]$ L_i, L_j $[m]$	KCL /H_2O & $NaCl$	≈ 300 µm

		$\beta\left[\frac{kg_{solvent}}{crytals \cdot s}\right]$ $\beta_1 = 1.613 \times 10^{-18}$ $\beta_1 = 0.095$ $\beta_1 = 0.264$		
Ciardha et al. (2012) [72]	$\beta = \beta_1 \cdot G^{\beta_2} \cdot \varepsilon^{\beta_3}$	G ... growth rate $\left[\frac{\mathrm{m}}{s}\right]$ ε ... nucleation rate $\left[\frac{\mathrm{m}^2}{s^3}\right]$ $L_i, L_j\ [m]$ $\beta\left[\frac{m_{solvent}{}^3}{crytals \cdot s}\right]$ $\beta_1 = 0.27 \pm 0.22$ $\beta_1 = 1.34 \pm 0.01$ $\beta_1 = 2.24 \pm 0.37$	$C_8H_9NO_2$ $(paracetamol)$ $/H_2O$ & CH_4O	$\approx 600\ \mu m$
Quintana-Hernandez et al. (2004) [73]	$\alpha = \alpha_1 \cdot \sigma^{\alpha_2} \cdot M_c{}^{\alpha_3} \cdot AR^{\alpha_4}$	σ ... relative supersaturation $[-]$ AR ... agitation rate [rpm] M_c ... total mass of crystals [g] $\alpha\left[\frac{crytals}{cm^{\,3}{}_{slurry} \cdot \mathrm{s} \cdot cm}\right]$ One set of fittet parameters: $\alpha_1 = 1; \alpha_2 = 0.07$ $\alpha_3 = 0.09;\ \alpha_4 = 0.001$ * the number of crystals born to a specific size is not depending on the current CSD→used PBE: $\frac{\partial}{\partial t}n(L,t) + \frac{\partial}{\partial L}\left(G \cdot n(L,t)\right) = B(L,t) + \alpha(L)$	$C_{12}H_{22}O_{11}$ $(sucrose)$ $/H_2O$	$\approx 150\ \mu m$

References:

[1] Z. Nagy, G. Fevotte, H. Kramer, and L. Simon, Chemical Engineering Research and Design **91**, 1903 (2013).

[2] J. Ulrich and M. J. Jones, Chemical Engineering Research and Design **82**, 1567 (2004).

[3] G. V. Reklaitis, J. Khinast, and F. Muzzio, Chemical Engineering Science **65**, iv (2010).

[4] S. Buchholz, Chemical Engineering and Processing **49**, 993 (2010).

[5] D. Ramkrishna, *Population Balances: Theory and Applications to Particulate Systems in Engineering* (Academic Press London, 2000).

[6] H. M. Hulbert, Chemical Engineering Science **9**, 555 (1964).

[7] P. Marchal, R. David, J. P. Klein, and J. Villermaux, Chemical Engineering Science **43**, 59 (1988).

[8] M. D. L., B. Richard, and D. K. Tafti, International Journal of Modern Physics B. **16**, 383 (2002).

[9] H. H. M. and K. S., Chemical Engineering Science **9**, 555 (1964).

[10] R. D, *Population Balances: Theory and Applications to Particulate Systems in Engineering. **2000,** Academic Press, London* (Academic Press, London, 2000).

[11] A. Abbas and J. A. Romagnoli, Separation and Purification Technology **53**, 153 (2007).

[12] Z. K. Nagy, M. Fujiwara, and R. D. Braatz, Journal of Process Control **18**, 856 (2008).

[13] C. B. B. Costa, M. R. W. Maciel, and R. M. Filho, Computers & Chemical Engineering **31**, 206 (2007).

[14] K. L. Choong and R. Smith, Chemical Engineering Science **59** (2004).

[15] C. Lindenberg, M. Kraettli, J. Cornel, M. Mazzotti, and J. Brozio, Crystal Growth & Design **9**, 1124 (2009).

[16] H. Gros, T. Kilpio, and J. Nurmi, Powder Technology **121** (2001).

[17] G. Rudiyanto, F. Irene, and B. R. D., AIChe J. **50**, **2738** (2004).

[18] M. Oullion, F. Puel, G. Févotte, S. Righini, and P. Carvin, Chemical Engineering Science **62**, 833 (2007).

[19] W. Koch and S. K. Friedlander, Journal of Colloid and Interface Science **140**, 419 (1990).

[20] D. E. Rosner, R. McGraw, and P. Tandon, Industrial & Engineering Chemistry Research **42**, 2699 (2003).

[21] S. Qamar, S. Noor, Q. u. Ain, and A. Seidel-Morgenstern, Industrial & Engineering Chemistry Research **49**, 11633 (2010).

[22] D. L. Marchisio and R. O. Fox, Journal of Aerosol Science **36**, 43 (2005).

[23] C. B. B. Costa, A. C. da Costa, and R. M. Filho, Chemical Engineering and Processing: Process Intensification **44**, 737 (2005).

[24] Y. Lin, K. Lee, and T. Matsoukas, Chemical Engineering Science **57**, 2241 (2002).

[25] M. Vanni, Journal of Colloid and Interface Science **221**, 143 (2000).

[26] A. Mersmann, B. Braun, and M. Loffelmann, Chemical Engineering Science **57**, 4267 (2002).

[27] W. R. Witkowski, S. M. Miller, and J. B. Rawlings, in *Crystallization as a Separations Process* (American Chemical Society, 1990), pp. 102.

[28] R. David, J. Villermaux, P. Marchal, and J.-P. Klein, Chemical Engineering Science **46**, 1129 (1991).

[29] A. Tadayon, S. Rohani, and M. K. Bennett, Industrial & Engineering Chemistry Research **41**, 6181 (2002).

[30] Q. Hu, S. Rohani, D. X. Wang, and A. Jutan, AIChE Journal **50**, 1786 (2004).

[31] H. B. Matthews and J. B. Rawlings, AIChE Journal **44**, 1119 (1998).

[32] A. Kalbasenka, A. Huesman, and H. Kramer, Chemical Engineering Science **66**, 4867 (2011).

[33] A. Braumann, P. L. W. Man, and M. Kraft, Industrial & Engineering Chemistry Research **49**, 428 (2010).

[34] A. Braumann, P. L. W. Man, and M. Kraft, Aiche Journal **57**, 3105 (2011).

[35] Y. Qiu and Å. C. Rasmuson, AIChE Journal **37**, **1293** (1991).
[36] H. Seki, N. Furuya, and S. Hoshino, Chemical Engineering Science **77**, 10 (2012).
[37] H. Modares and M.-B. Naghibi Sistani, Engineering Applications of Artificial Intelligence **24**, 476 (2011).
[38] S. A. Schmidt, M. Simon, M. M. Attarakih, L. Lagar G, and H.-J. Bart, Chemical Engineering Science **61**, 246 (2006).
[39] A. P. Mariano, C. B. B. Costa, D. d. F. de Angelis, F. M. Filho, D. I. P. Atala, M. R. Wolf Maciel, and R. Maciel Filho, Journal of Chemical Technology & Biotechnology **85**, 934 (2010).
[40] S. Das, A. Abraham, and A. Konar, *Swarm Intelligence Algorithms in Bioinformatics* (Springer Berlin / Heidelberg, 2008).
[41] P. Reche-López, N. Ruiz-Reyes, S. García Galán, and F. Jurado, Energy Conversion and Management **50**, 2020 (2009).
[42] R. Ramachandran and P. I. Barton, Chemical Engineering Science **65**, 4884 (2010).
[43] S. M. Nowee, A. Abbas, and J. A. Romagnoli, Chemical Engineering and Processing: Process Intensification **46**, 1096 (2007).
[44] D. J. Widenski, A. Abbas, and J. A. Romagnoli, Computers & Chemical Engineering **35**, 2696 (2011).
[45] G. D. Maia and M. Giulietti, Journal of Chemical and Engineering Data **53**, 256 (2008).
[46] M. Stieß, *Mechanische Verfahrenstechnik - Partikeltechnologie 1* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), Vol. 3.
[47] A. Rashid, E. T. White, T. Howes, J. D. Litster, and I. Marziano, Chemical Engineering Research & Design **90** (2012).
[48] R. Gunawan, D. L. Ma, M. Fujiwara, and R. D. Braatz, International Journal of Modern Physics B **16**, 367 (2002).
[49] S. Rohani, Seperations Technology, 1993), pp. 99.
[50] A. B. Mersmann, Klaus. Braun, Björn. Heyer, Axel. Heyer, Christiane, (Wiley-VCH Verlag GmbH, Chemie Ingenieur Technik, 2000).
[51] R. David, P. Marchal, and B. Marcant, Chemical Engineering & Technology **18**, 302 (1995).
[52] P. Thomson, in *Internation Conference on Cloud Physics*Toronto, 1986), pp. 115.
[53] R. Zauner and A. G. Jones, Chemical Engineering Science **55**, 4219 (2000).
[54] X. Z. Wang, J. C. De Anda, and K. J. Roberts, Chemical Engineering Research & Design **85**, 921 (2007).
[55] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Journal of Chemical Physics **21** (1953).
[56] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220** (1983).
[57] W. T. Press, Saul. Vetterling, William. Flannery, Brian., *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, 2007).
[58] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, 2007).
[59] J. Kennedy and R. Eberhart, in *Neural Networks, 1995. Proceedings., IEEE International Conference on*1995), pp. 1942.
[60] Y. Shi and R. Eberhart, in *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*1998), pp. 69.
[61] R. D, *Population Balances: Theory and Applications to Particulate Systems in Engineering* (Academic Press, London, 2000).
[62] F. Févotte and G. Févotte, Chemical Engineering Science **65**, 3191 (2010).
[63] C. B. B. Costa, M. R. W. Maciel, and R. Maciel, Computers & Chemical Engineering **31**, 206 (2007).

[64] D. L. Marchisio, R. D. Vigil, and R. O. Fox, Journal of Colloid and Interface Science **258**, 322 (2003).
[65] S. Mortier, K. V. Gernaey, T. De Beer, and I. Nopens, Computers & Chemical Engineering **50**, 39 (2013).
[66] S. Kumar and D. Ramkrishna, Chemical Engineering Science **51**, 1311 (1996).
[67] J. Kumar, M. Peglow, G. Warnecke, S. Heinrich, and L. Morl, Chemical Engineering Science **61**, 3327 (2006).
[68] J. Kumar, M. Peglow, G. Warnecke, and S. Heinrich, Computers & Chemical Engineering **32**, 1810 (2008).
[69] A. Chaudhury, A. Niziolek, and R. Ramachandran, Advanced Powder Technology **24**, 113 (2013).
[70] A. Chaudhury, A. Kapadia, A. V. Prakash, D. Barrasso, and R. Ramachandran, Advanced Powder Technology, 2013).
[71] Y. I. Lim, J. M. Le Lann, X. M. Meyer, X. Joulia, G. Lee, and E. S. Yoon, Chemical Engineering Science **57**, 3715, Pii s0009-2509(02)00236-1 (2002).
[72] C. T. O'Ciardha, K. W. Hutton, N. A. Mitchell, and P. J. Frawley, Crystal Growth & Design **12**, 5247 (2012).
[73] P. Quintana-Hernandez, E. Bolanos-Reynoso, B. Miranda-Castro, and L. Salcedo-Estrada, Aiche Journal **50**, 1407 (2004).

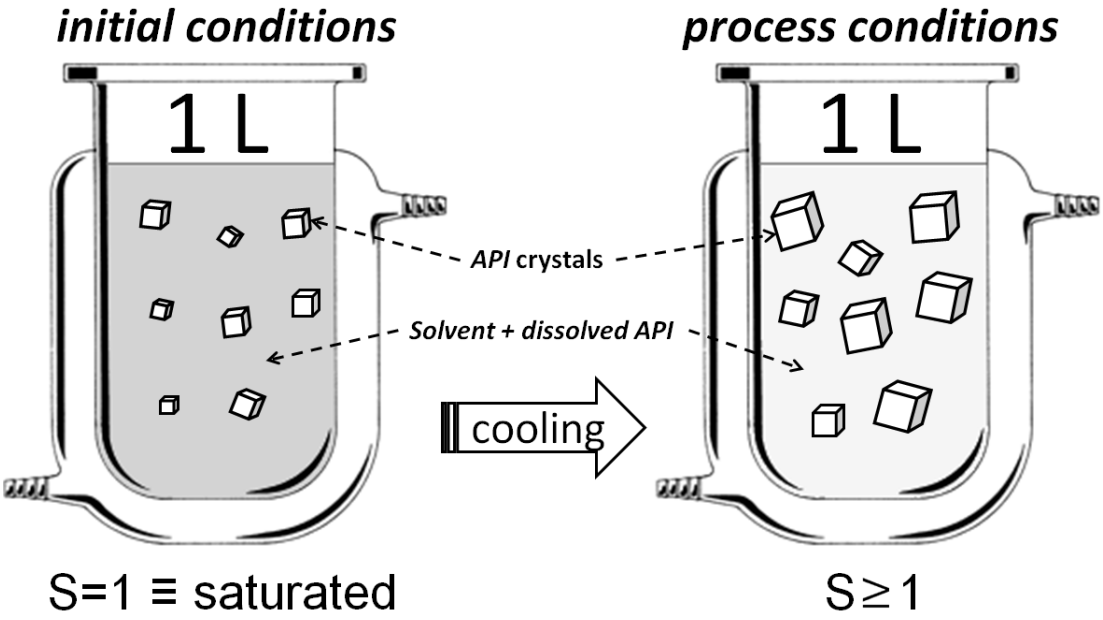
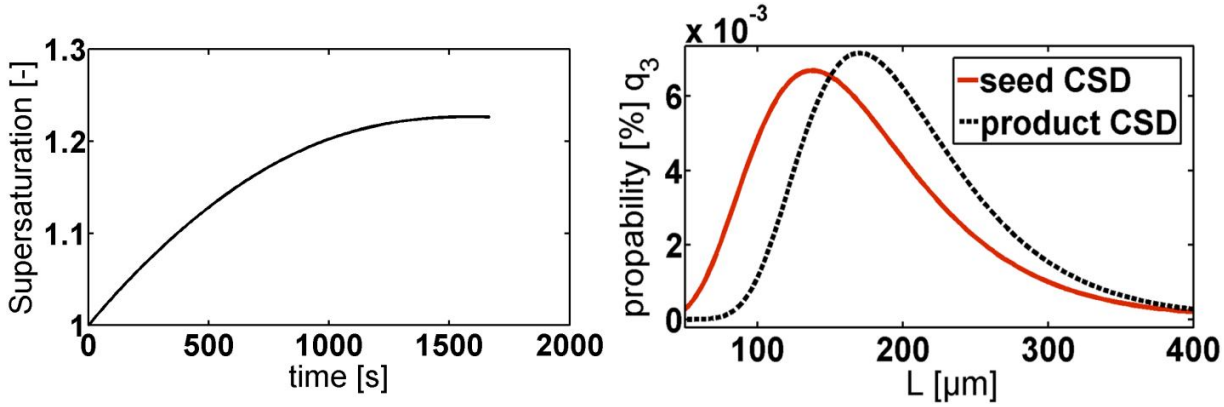


Figure 1: Schematic draft of the modeled well mixed seeded batch cooling crystallization process



a b

Figure 2: Supersaturation profile during the crystallization **(a)** ; simulated change in the CSD caused by crystal growth only **(b)**

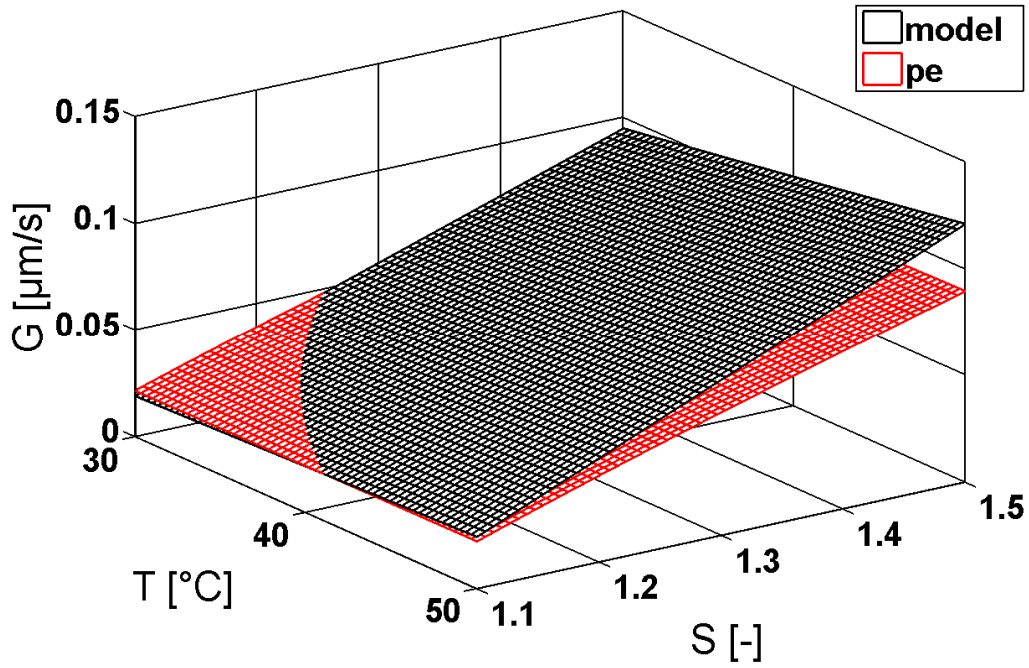


Figure 3: G_{model} and G_{PE} over the considered parameter space for the evaluation of $\hat{G}_{error}$ (PE from noisy data, Table 6, SAA+Nelder-Mead Experiment 5, $\hat{G}_{error} = 0.0696$)

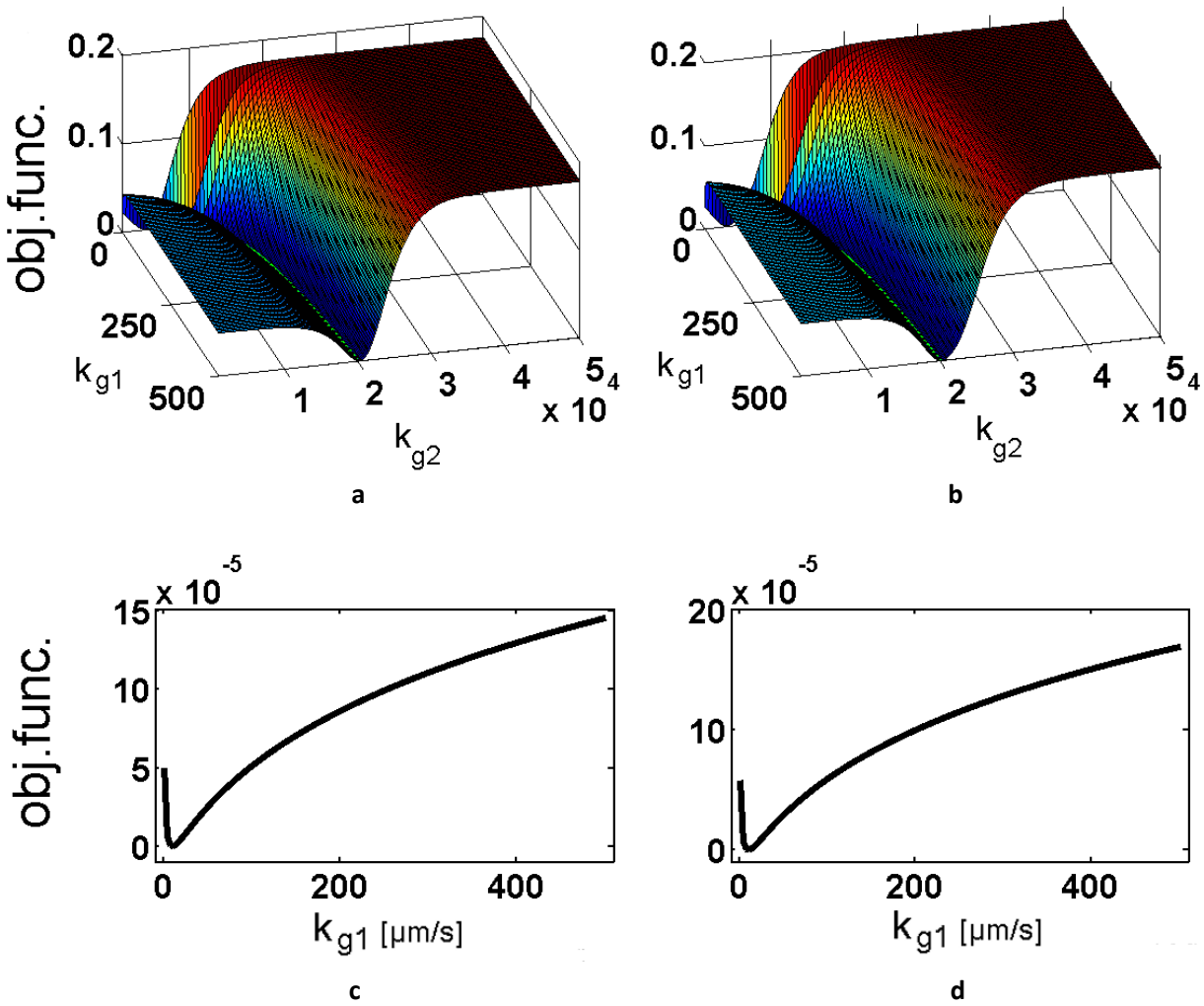


Figure 4: Objective function was evaluated over fraction of the investigated parameter space calculated from data by experiment 5 only **(a)** and experiments 1-4 **(b)**. A logarithmic function was fitted through the band of (k_{g1}, k_{g2}) combinations differing only slightly from the global minimum: $k_{g2} = 2.5432 \cdot e^3 \cdot \log(k_{g1}) + 0.0059 \cdot k_{g1} + 4.1435 \cdot e^3$. The objective function was evaluated along this band separately in the case of experiment 5 only **(c)** and experiments 1-4 **(d)**.

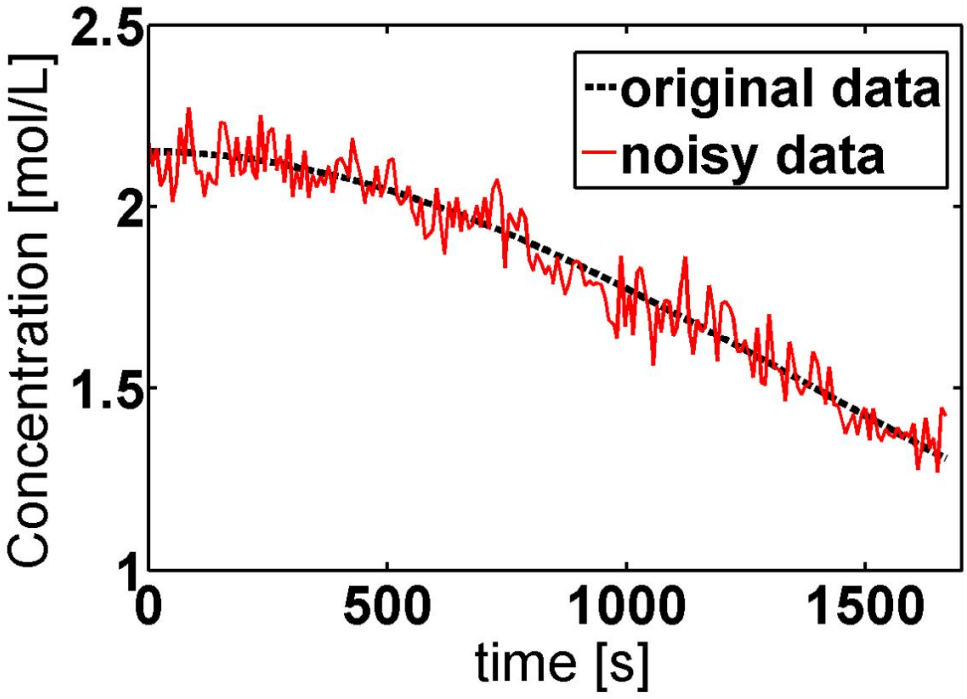


Figure 5: Concentration profile of experiment 5, with the noisy profile generated with $\boldsymbol{\sigma_{noise} = 3}$.

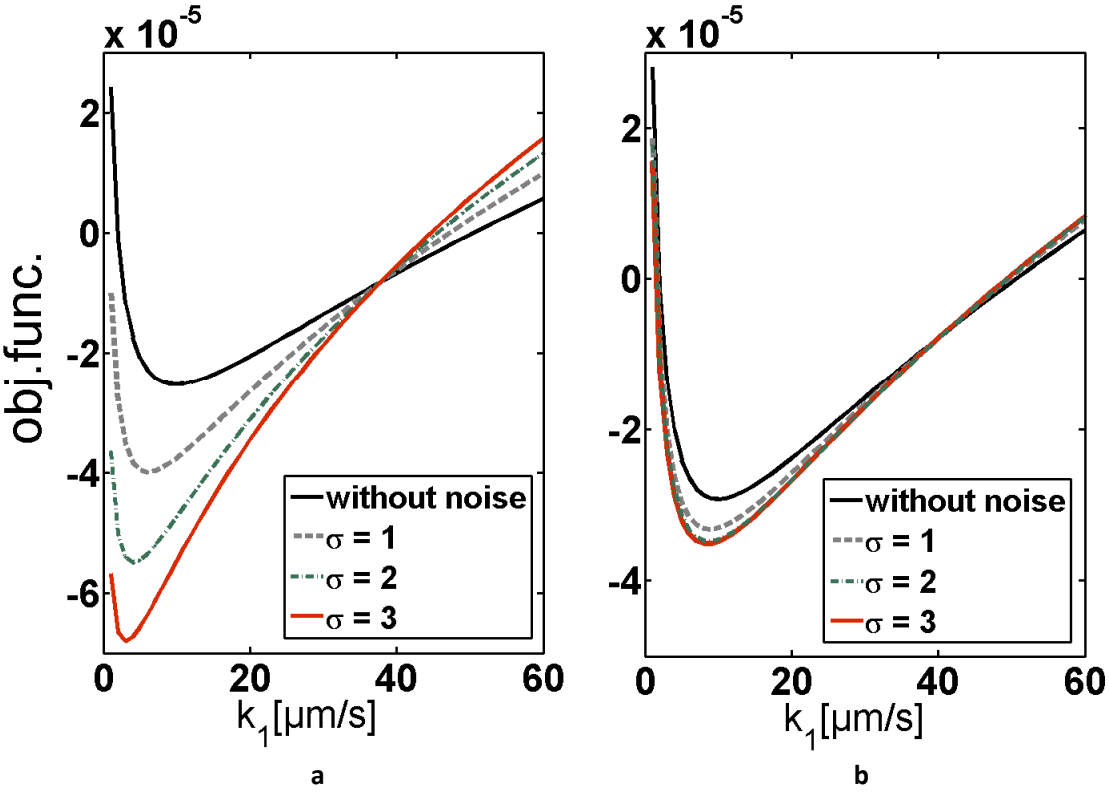


Figure 6: Objective function along the band of $(\mathbf{k_{g1}}, \mathbf{k_{g2}})$ combinations differing only slightly from the global minimum in the case experiment 5 **(a)** and experiments 1-4 **(b)** for different noise levels. The objective function was centered by subtracting its mean value. See the caption of Figure 4 for the assigned $\mathbf{k_{g2}}$ and $\mathbf{k_{g2}}$ values.

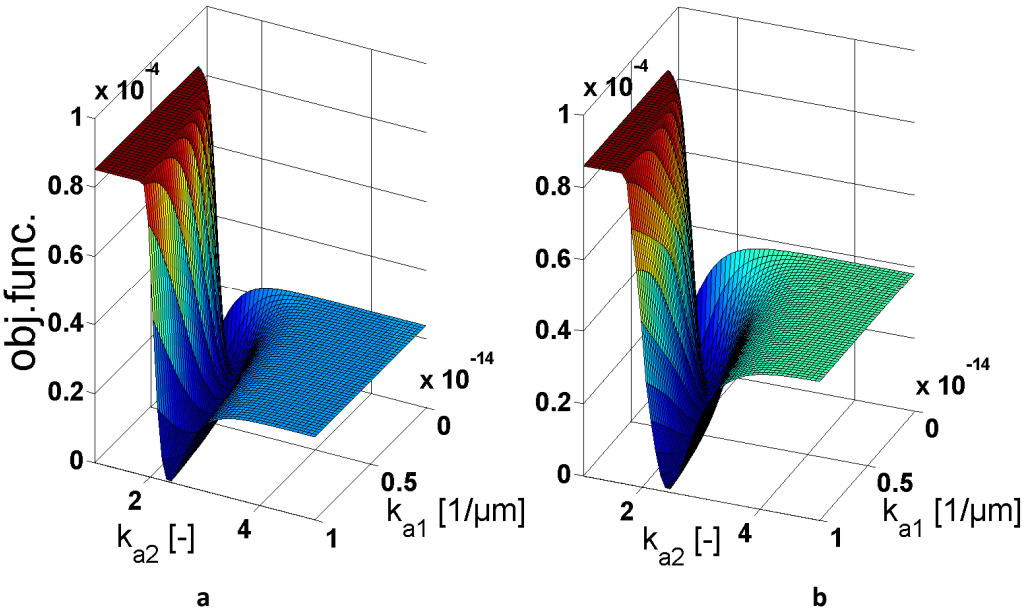


Figure 7: Objective function evaluated in a fraction of the parameter space considered for PE. **(a)** experiment 5 only **(b)**: experiments 1-4

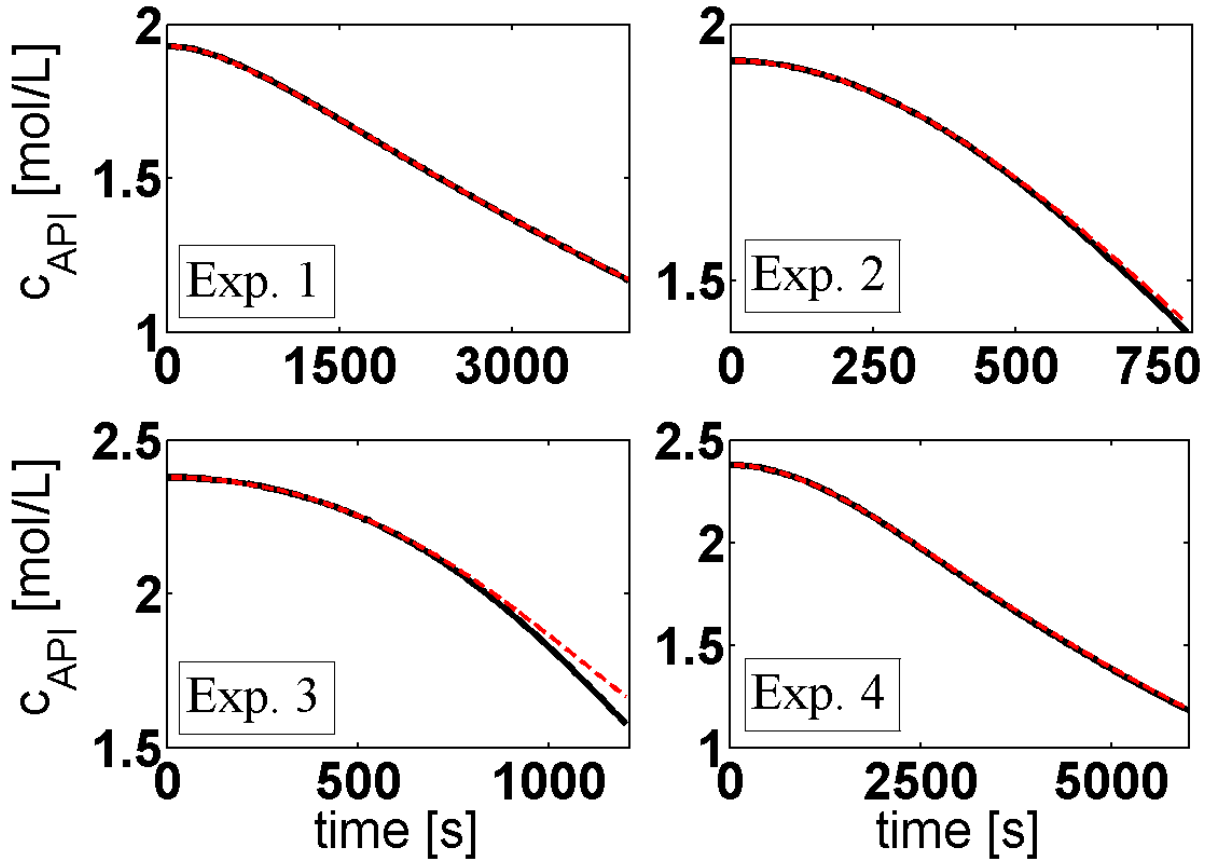


Figure 8: Comparison of concentration profiles of the experiments 1, 2, 3 and 4 (see Table 3). *Solid line:* Process model output using the model parameters; *Dashed line:* Process model output using the growth rate parameters obtained by the PE routines neglecting the presence of aggregation (Experiment 1+2+3+4,
Table 9).