Chemical Engineering Science 65 (2010) 4884–4893

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Effective parameter estimation within a multi-dimensional population balance model framework

Rohit Ramachandran [a,*], Paul I. Barton [b]

[a] Department of Chemical and Biochemical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[b] Process Systems Engineering Laboratory, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 13 January 2010
Received in revised form
12 May 2010
Accepted 21 May 2010
Available online 1 June 2010

Keywords:
Granulation
Multi-dimensional population balance model
Kernels
Parameter estimation
Optimization
Automatic differentiation

ABSTRACT

This study considers optimization problems with multi-dimensional population balance models embedded. The objective function is formulated as a least-squares problem, minimizing the difference between target data and simulated model output and the goal is to find model parameter values that best fit the data. Results show that derivative-free methods, such as the Nelder–Mead simplex method, fail to converge to an optimal solution. A similar result was obtained with gradient-based methods such as BFGS, quasi-Newton, Newton, Gauss–Newton, Levenberg–Marquardt and SQP, and with a stochastic genetic algorithm. It was hypothesized that three main issues could contribute to these convergence failures: (1) gradients were calculated based on finite differences, and as a result of improper step size determination, the numerical error could be prohibitive resulting in inaccurate derivative information, (2) the parameters may not be identifiable and (3) numerical instability could occur during the course of optimization. To circumvent these issues, this work addresses the calculation of derivative information based on automatic differentiation and sensitivity analysis to ensure increased accuracy. Issues such as parameter identifiability are also dealt with by analyzing an accurate Fisher information matrix. Given the computational burden in calculating accurate Jacobians and Hessians, compounded by the potential nonsmoothness introduced into the objective function as a result of granule nucleation, other optimization strategies may be warranted and this work addresses those accordingly. Overall, by systematically assessing the problem formulation and mechanisms, the results show that substantial improvements in convergence can be achieved by utilizing appropriate optimization techniques, thus leading to more successful and optimal parameter estimation.

1. Introduction

Multi-dimensional and multi-scale population balance models are often required for the modeling and analysis of particulate systems that are distributed with respect to two or more of their internal coordinates (e.g., granulation) (Iveson, 2002). The population balance model is essentially a meso-scale framework that utilizes microscopic information (e.g., material properties, kinetic and thermodynamic parameters) in the form of kernels, to predict macroscopic particle traits (e.g., size distribution of crystals, porosity distribution of granules, etc.). These kernels can be mechanistic (i.e., formulated based on the physics and chemistry of the process) or empirical (i.e., formulated based on power-law expressions with unknown fitting parameters). Given that most particulate processes are complex and intricate, most of the kernels in the literature are of the latter form. An important step in the formulation of a model for a given process is estimation of values for model parameters. The accuracy of the estimates for these parameters is crucial to the applicability of the formulated model for process design, control and optimization. Parameter estimation can sometimes be straightforward, for instance when the process and formulated model are relatively simple and sufficient data are available. However, these conditions are not always met and this can result in difficulties in determining accurate parameter values. The focus of this study is on systems described by population balances and the nature of these systems results in tremendous difficulties for optimization and control and conventional optimization routines can often fail when attempting to determine the optimal parameter values (Wang and Cameron, 2007). In many instances, due to the complexity of these model structures, systematic optimization techniques are abandoned and parameters are estimated based on trial and error, often resulting in sub-optimal parameter values. The multiple dimensions, nonlinearity, nonconvexity, hyperbolic and integral terms associated with the population balance partial–differential

* Corresponding author. Fax: +1 732 445 2581.
E-mail address: rohit.r@rutgers.edu (R. Ramachandran).

0009-2509/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2010.05.039

equation are potential causes, amongst others, for convergence failure (Wang and Cameron, 2007).

2. Background

This section details the generalized population balance equation, and the multi-dimensional population balance models and kernels used in this study.

2.1. Population balance equation

In its most general form, the population balance equation (PBE) is written as shown below (Ramkrishna, 2000):

$$\frac{\partial F}{\partial t}(\mathbf{x},t)+\frac{\partial}{\partial \mathbf{x}}\left[F\frac{d\mathbf{x}}{dt}\right](\mathbf{x},t)=\Re_{formation}(\mathbf{x},t)-\Re_{depletion}(\mathbf{x},t). \quad (1)$$

Here $\mathbf{x}$ is the vector of state variables used to characterize the distribution and $F(\mathbf{x},t)$ is the population distribution function. In granulation for example, a single state variable, size, is often employed; and the resulting distribution is called the granule size distribution. The term $(\partial/\partial\mathbf{x})[F(d\mathbf{x}/dt)](\mathbf{x},t)$ accounts for growth processes. Thus in granulation, $(\partial/\partial\mathbf{x})[F(d\mathbf{x}/dt)](\mathbf{x},t)$ would then account for the rate at which the distribution evolves due to granule consolidation. $\Re_{formation}(\mathbf{x},t)$ accounts for formation of particles due to nucleation, aggregation and breakage phenomena, and $\Re_{depletion}(\mathbf{x},t)$ accounts for depletion of particles by the same phenomena.

2.2. Multi-dimensional population balance models

One-dimensional models have been well studied in particulate processes for model-based analysis (Christofides, 2001; Ma et al., 2002; Immanuel and Doyle III, 2003). They are often adequate for processes where the effect of key mechanisms (such as nucleation, aggregation and breakage) on the process dynamics can be accounted for through the consideration of a single particle characteristic (e.g., size distribution of emulsion polymers). However, the accurate modeling of granulation processes warrants the consideration of multiple particle characteristics. Aside from granule size, binder content has an important effect on granule growth with many studies concurring that the amount of binder correlates directly with the rate of granule growth, due to a larger availability of surface-wet granules (Knight, 1993; Kristensen, 1996). As a direct result of this, induction time decreases thus promoting steady growth (Iveson and Litster, 1998; Wauters, 2000; Knight et al., 1998; Hoornaert et al., 1998). Intuitively as well, it is clear that improper liquid distribution (which often occurs in granulation) will promote different granule growth rates (Reynolds et al., 2004). Granule porosity is an essential parameter that has a profound effect on granule growth behaviour. Above a certain critical porosity, pore saturation is low and binder liquid remains within the granule resulting in surface-dry granules. Below the critical porosity, pores are saturated and liquid is squeezed on the surface resulting in surface-wet granules and higher growth rates. Porosity also has an effect on deformability and strength and as a result it can have a pronounced effect on growth and breakage. For instance, in the work of Annapragada and Neilly (1996) it was observed that porous granules when initially formed, go onto break quickly (resulting in higher breakage rates) and thereafter aggregate to form stronger denser granules (resulting in lower breakage rates).

As a result of the above discussion, the granulation process should be described by a 3-D population balance model, which accounts for granule size, binder content and porosity. The potential heterogeneity of the population distribution with respect to all three attributes is such that lumping in any of these dimensions would result in considerable model errors (Iveson, 2002). In formulating the 3-D population balance, the three particle attributes are re-cast in terms of their individual volumes of solid (s), liquid (l) and gas (g) (Verkoeijen et al., 2002). The purpose of modeling the granulation process in terms of its individual volumes enables decoupling of the individual mesoscopic processes (i.e., aggregation, consolidation, etc.). This essentially offers two advantages. Firstly, one can model a single rate process at a time. Secondly, the mutually exclusive character of the internal coordinates substantially improves the numerical solution of the model as rate processes with differing time constants are segregated. The resulting 3-D population balance equation is then given by

$$\frac{\partial F}{\partial t}(s,l,g,t)+\frac{\partial}{\partial g}\left(F(s,l,g,t)\frac{dg}{dt}\right)+\frac{\partial}{\partial s}\left(F(s,l,g,t)\frac{ds}{dt}\right)+\frac{\partial}{\partial l}\left(F(s,l,g,t)\frac{dl}{dt}\right)$$
$$=\Re_{nuc}(s,l,g,t)+\Re_{agg}(s,l,g,t)+\Re_{break}(s,l,g,t), \quad (2)$$

where $F(s,l,g,t)$ represents the population density function such that $F(s,l,g,t)\,ds\,dl\,dg$ is the moles of granules (adopting a number-based population balance instead of mass or volume-based) with solid volume between s and $s+ds$, liquid volume between l and $l+dl$ and gas volume between g and $g+dg$. The partial derivative term with respect to s accounts for the layering of fines onto the granule surfaces; the partial derivative term with respect to l accounts for the drying of the binder and the re-wetting of granules; the partial derivative with respect to g accounts for consolidation which, due to compaction of the granules, results in an increase of pore saturation and decrease in porosity. In this study, layering and drying are neglected and the simplified equation is given by

$$\frac{\partial F}{\partial t}(s,l,g,t)+\frac{\partial}{\partial g}\left(F(s,l,g,t)\frac{dg}{dt}\right)=\Re_{nuc}(s,l,g,t)+\Re_{agg}(s,l,g,t)+\Re_{break}(s,l,g,t). \quad (3)$$

The formation and depletion terms associated with the aggregation phenomenon ($\Re_{agg}$) are defined in Eqs. (4)–(6) (Ramkrishna, 2000; Immanuel and Doyle III, 2005). In these equations, s_{nuc} is the solid volume of nuclei (assumed constant), and $\beta(s',s-s',l',l-l',g',g-g')$ is the size-dependent aggregation kernel that signifies the rate constant for aggregation of two granules of internal coordinates (s',l',g') and $(s-s',l-l',g-g')$. β is essentially a measure of how successful collisions between two particles resulting in a larger granule are

$$\Re_{agg}(s,l,g,t)=\Re_{agg}^{formation}-\Re_{agg}^{depletion}, \quad (4)$$

where

$$\Re_{agg}^{formation}=\frac{1}{2}\int_{s_{nuc}}^{s-s_{nuc}}\int_0^{l_{max}}\int_0^{g_{max}}\beta(s',s-s',l',l-l',g',g-g')\times F(s',l',g',t)F(s-s',l-l',g-g',t)\,ds'\,dl'\,dg', \quad (5)$$

$$\Re_{agg}^{depletion}=F(s,l,g,t)\int_{s_{nuc}}^{s_{max}}\int_0^{l_{max}}\int_0^{g_{max}}\beta(s',s,l',l,g',g)\times F(s',l',g',t)\,ds'\,dl'\,dg'. \quad (6)$$

The breakage term ($\Re_{break}$) is described mathematically in two parts: the breakage kernel (k_{break}) and the breakage function (b). The former describes the rate at which a particle of size s, l and g breaks into a fragment of size s', l' and g'. The latter describes the sizes of these fragments formed and is based on the work of Pinto et al. (2007, 2008) and Pinto (2008). Based on the probabilities of particles in a particular finite volume breaking to form daughter particles in one or more smaller finite volumes in the 3-D space,

the authors performed a numerical operation that was able to describe the distribution of these fragments. $\Re_{break}$ is defined in

$$\Re_{break}(s,l,g,t) = \Re_{break}^{formation} - \Re_{break}^{depletion}, \tag{7}$$

$$\Re_{break}^{formation} = \int_0^{s_{max}} \int_0^{l_{max}} \int_0^{g_{max}} k_{break}(s',l',g') b(s,l,g,s',l',g') \times F(s',l',g',t)\, ds'\, dl'\, dg', \tag{8}$$

$$\Re_{break}^{depletion} = k_{break}(s,l,g) F(s,l,g,t). \tag{9}$$

The nucleation model is represented mathematically as shown in

$$\Re_{nuc} = B_0 \delta(V - V_0), \tag{10}$$

where B_0 represents the nucleation rate constant, δ the Dirac–Delta function, V the size of the particles which is defined to be $V = s+l+g$ and V_0 the size of the nuclei which is defined to be $V_0 = s_0+l_0+g_0$, where the subscript 0 indicates the critical lower limit, above which nucleation occurs.

The consolidation model is represented by an empirical exponential decay relation and is shown in Eq. (12). The porosity of granules is defined by Eq. (11) and substituting Eq. (11) into Eq. (12) gives a formal expression explicitly in terms of the three independent internal coordinates (see Eq. (13)), which can then be used in Eq. (3):

$$\varepsilon = \frac{l+g}{s+l+g}, \tag{11}$$

$$\frac{d\varepsilon}{dt} = -c(\varepsilon - \varepsilon_{min}), \tag{12}$$

$$\frac{dg}{dt} = \frac{-c(s+l+g)(1-\varepsilon_{min})}{s}\left[l - \frac{\varepsilon_{min} s}{1-\varepsilon_{min}} + g\right]. \tag{13}$$

Here ε_{min} is the minimum porosity of the granules (arbitrarily set at $\varepsilon_{min} = 0.2$) and c is the compaction rate constant.

2.3. Identification of kernels

A primary challenge in the development of population balance models is the identification of appropriate kernels that describe the individual mechanisms. While the development of a multi-dimensional population balance model is motivated by the physics of the problem, it is a tougher task to obtain 3-D kernels that account for the dependence of the rates on the particle traits (i.e., size, binder content and porosity).

2.3.1. Aggregation kernel

The aggregation kernel is essentially a measure of how successful is a binary collision of two particles v and u of volumes (s',l',g') and $(s-s',l-l',g-g')$, assuming that $F(s',l',g')$ $F(s-s',l-l',g-g')$ is a measure of the total number density of all possible binary collisions that can occur. This is a reasonable assumption if the system is assumed to be free-in-space, implying that any given particle can collide with all other particles as the paths to other particles are not hindered (Sastry and Fuerstenau, 1973; Wauters, 2000). In the event that the system is restricted-in-space (i.e., particles can only collide with neighbouring particles), the number of possible binary collisions is significantly less than $F(s',l',g')F(s-s',l-l',g-g')$ (Sastry and Fuerstenau, 1973; Wauters, 2000). To compensate for this, Sastry and Fuerstenau (1973) introduced a means to correct this error by stating that the number of possible binary collisions would reduce to $F(s',l',g')F(s-s',l-l',g-g')/N_{tot}$, where N_{tot} is the total number of particles (in terms of number density) in the system. For granulation, the decision to ascertain if the system is free-in-space or restricted-in-space is fairly arbitrary with some studies adopting the correction factor (Liu and Litster, 2002; Adetayo and Ennis, 2000; Wauters, 2000) and others choosing not to (Hounslow, 1998; Biggs et al., 2003; Madec et al., 2003; Darelius et al., 2005; Tan et al., 2005).

The aggregation kernel ($\beta(s',s-s',l',l-l',g',g-g')$) used in this study is based on the equi-partition of kinetic energy (EKE model) first described by Hounslow (1998), which assumes that particles collide as a consequence of their random component of velocity and that the random components result in equal distribution of the particles kinetic energy and is described in

$$\beta(s',s-s',l',l-l',g',g-g') = \beta_0 (D(s',l',g') + D(s-s',l-l',g-g'))^2 \times \sqrt{\frac{1}{D^3(s',l',g')} + \frac{1}{D^3(s-s',l-l',g-g')}}, \tag{14}$$

where D is the particle diameter in size classes s',l',g' and $s-s',l-l',g-g'$ and β_0 is the aggregation rate constant. $D(s,l,g)$ can be obtained from the relation $D(s,l,g) = [6(s+l+g)/\pi]^{1/3}$.

2.3.2. Breakage kernel

As seen from Eqs. (8) and (9), breakage is described by the breakage function (b) and the breakage kernel (K_{break}). The breakage function describes how fragments, resulting from the breakup of granules, are distributed in terms of their volume. There are several possible functional forms for this distribution, given either by continuous (e.g., normal or lognormal distribution) or discrete (e.g., binary) distributions. The breakage function used in this study is based on the work of Pinto et al. (2007, 2008) and Pinto (2008) (described in Section 2.2). Using these probabilities, semi-analytical solutions were derived for the triple integrals present in $\Re_{break}^{formation}$ (see Eq. (8)), thus eliminating the integrals altogether and aiding the numerical solution. The breakage kernel used in this study is a power law expression based on experimental observations proposed by Pandya and Spielman (1983) and is defined by

$$K_{break}(s,l,g) = \frac{P_1 G (D(s,l,g))^{P_2}}{2}, \tag{15}$$

where K_{break} is the breakage kernel, P_1 and P_2 are adjustable constants and G is the shear rate.

3. Theory

3.1. Numerical methods

As seen from the above discussions, multi-dimensional population balances and kernels are warranted for modeling the granulation process. As a result, the challenges lie not only in the development of these models/kernels but also in their applications which require efficient techniques for their numerical solution. The population balance equation used to model the granulation process, results in complicated hyperbolic integro-partial differential equations. Furthermore, the multiple time scales in the process render the model equations with a large stiffness, substantially complicating the numerical integration. The multi-scale nature of the spatial coordinates in the partial differential equations makes it a challenge to develop a suitable discretization of the spatial coordinates. The different states of the models (the particle versus the lumped quantities) also complicate numerical integration. Lastly, the multi-dimensional distributions (along multiple spatial coordinates) add to the computational complexity as well as to the memory needs for the computers. Hence, it is of interest to develop robust, realistic and

efficient solution techniques for multi-dimensional population balance equations (Ma et al., 2002; Mantzaris et al., 2001).

In this study, the particle population is first discretized into sub-populations and the population balance is formulated for each of these semi-lumped sub-populations. This is obtained by the integration of the population balance equation (see Eq. (3)) over the domain of the sub-populations and re-casting the population into finite volumes. In this finite volume scheme, Eq. (3) may be re-written in a discrete form as shown in

$$\frac{dF'_{i,j,k}}{dt}+\frac{F'_{i,j,k}}{\Delta g_k}\frac{dg}{dt}\bigg|_{g_k}-\frac{F'_{i,j,k+1}}{\Delta g_{k+1}}\frac{dg}{dt}\bigg|_{g_{k+1}}=\Re_{nuc}(s_i,l_j,g_k)+\Re_{agg}(s_i,l_j,g_k)+\Re_{break}(s_i,l_j,g_k). \tag{16}$$

Here $F'_{i,j,k}=\int_{s_i}^{s_{i+1}}\int_{l_j}^{l_{j+1}}\int_{g_k}^{g_{k+1}}F(s,l,g)\,ds\,dl\,dg$, s_i is the value of the solid volume at the upper end of the ith bin along the solid volume axis, l_j is the value of the liquid volume at the upper end of the jth bin along the liquid volume axis, g_k is the value of the gas volume at the upper end of the kth bin along the gas volume axis. Δg_k is the size of the kth gas bin with respect to the gas volume axis. The particle population is assumed to be uniform within each of the finite volumes. Thus, by this technique, the integro partial–differential equation as represented by the population balance equation is reduced to a system of ordinary differential equations in terms of the rates of nucleation ($\Re_{nuc}(s_i,l_j,g_k)$), aggregation ($\Re_{agg}(s_i,l_j,g_k)$) and breakage ($\Re_{break}(s_i,l_j,g_k)$) (see Eqs. (4)–(10)). Off-line semi-analytical solutions are proposed for $\Re_{agg}(s_i,l_j,g_k)$ and $\Re_{break}(s_i,l_j,g_k)$ (Immanuel and Doyle III, 2003, 2005; Pinto et al., 2007). This results in casting the complex triple integrals in simpler addition and multiplication terms, major portions of which are computed once a priori to the start of the simulation. $\Re_{nuc}$ being much less computationally intensive is updated at every time-step.

Two numerical methods are used in the course of this study. The first is a hard-coded first-order explicit predictor–corrector Euler method with error control and the second is a multi-step implicit method known as the backward differentiation formula (BDF) method which is the default method coded in DSL48S (Feehery et al., 1997). DSL48S is a code for numerical solution and sensitivity analysis of large-scale sparse differential-algebraic equations (DAEs). DSL48S implements a variable step size, variable order version of the BDF method.

3.2. Parameter estimation

This work is concerned with parameter estimation for dynamic models of population balance systems that are formulated (after discretization) as systems of nonlinear ordinary differential equations (ODEs). The dynamical system is written as shown in

$$\dot{\mathbf{y}}(t,\mathbf{p})=\mathbf{f}(\mathbf{y}(t,\mathbf{p}),\mathbf{p}),\quad \forall t\in(0,t_f], \tag{17}$$

$$\mathbf{y}(0,\mathbf{p})=\mathbf{y}_0, \tag{18}$$

where $\mathbf{y}$ are the discretized states, or particle population densities, and $\mathbf{p}$ are the model parameters, which in this case is the column vector of the kernel constants as shown in

$$\mathbf{p}=[\beta_0\ \ P_1\ \ P_2\ \ c\ \ B_0]^{\mathrm{T}}. \tag{19}$$

3.2.1. Methods for parameter estimation

A method often used for parameter estimation in such models is least-squares (LS) fitting. It involves the minimization of the squared error between target data, $\mathbf{x}_{data}$ and model-derived data, $\mathbf{x}_{fit}$. The data are usually mass or volume frequencies against size classes which can be obtained via algebraic manipulation of the states such that $\mathbf{x}(t,\mathbf{p})=\mathbf{g}(\mathbf{y}(t,\mathbf{p}))$. Subsequently, LS fitting involves minimizing an objective function as shown in

$$\Omega(\mathbf{p})=\sum_{z=1}^{n_z}\|\mathbf{x}_{fit}(t_z,\mathbf{p})-\mathbf{x}_{data}(t_z)\|_2^2. \tag{20}$$

The aim of any parameter estimation method is to minimize Ω and find a unique optimum point. The quality of fit is heavily influenced by the number of experimental measurements, n_z, and the dimensionality of the vector $\mathbf{x}$. Often, gradient-based optimization techniques are used to minimize Ω, where the gradient information can be obtained by integrating the model sensitivities $\partial y_j/\partial p_k$ along with the model states $\mathbf{y}$ over the time horizon, which are given by

$$\frac{\partial\dot{\mathbf{y}}}{\partial\mathbf{p}}(t,\mathbf{p})=\frac{\partial\mathbf{f}}{\partial\mathbf{y}}(t,\mathbf{p})\frac{\partial\mathbf{y}}{\partial\mathbf{p}}(t,\mathbf{p})+\frac{\partial\mathbf{f}}{\partial\mathbf{p}}(t,\mathbf{p}),\quad \forall t\in(0,t_f], \tag{21}$$

$$\frac{\partial\mathbf{y}}{\partial\mathbf{p}}(0,\mathbf{p})=\mathbf{0}. \tag{22}$$

Eq. (23) shows the sensitivity ODEs combined with the chain rule to obtain $\partial\Omega/\partial\mathbf{p}$:

$$\frac{\partial\Omega}{\partial\mathbf{p}}(\mathbf{p})=\sum_{z=1}^{n_z}\sum_{m=1}^{n_m}2(g_m(\mathbf{y}(t_z,\mathbf{p}))-x_{data,m}(t_z))\frac{\partial g_m}{\partial\mathbf{y}}(t_z,\mathbf{p})\frac{\partial\mathbf{y}}{\partial\mathbf{p}}(t_z,\mathbf{p}). \tag{23}$$

3.3. Formulation of the objective function

In the course of granulation experiments, three-phase data should be gathered to obtain a realistic validated model as demonstrated in the experimental work of Ramachandran et al. (2008). The primary data are typically in the form of 1-D normalized volume frequencies (x_{fit}) as a function of particle diameter (D). This can be obtained from a mapping of 3-D finite volumes (or bins) to 1-D size classes. (Note: a bin is defined to be a 3-D grid space of coordinates s_i, l_j and g_k and a size class is defined to be a 1-D grid space of coordinate D_m.) In this study, 1000 finite volumes are mapped into 15 size classes. During the mapping, the algorithm checks to see which finite volumes of discretized coordinates s_i, l_j and g_k map into a size class (diameter), using the relation $D_{ijk}\equiv[(6(s_i+l_j+g_k))/\pi]^{1/3}$ such that $L_m\le D_{ijk}<L_{m+1}$, where m represents each 1-D size class (diameter) and L_m is the lower bound of these size classes. Since it is entirely possible that several finite volumes may translate into one size class, the 1-D x_{fit} is obtained from the 3-D y as shown in

$$x_{fit,m}(t)=\sum_{L_m\le D_{ijk}<L_{m+1}}\left(\frac{y(s_i,l_j,g_k,t)}{\sum_{i=1}^{n_s}\sum_{j=1}^{n_l}\sum_{k=1}^{n_g}y(s_i,l_j,g_k,t)}\right)\frac{V_m}{\sum_{m=1}^{n_m}V_m}, \tag{24}$$

where V_m is the particle volume in an mth size class, y is the particle population density, n_s is the number of solid bins, n_l is the number of liquid bins, n_g is the number of gaseous bins and n_m is the number of 1-D size classes. The secondary data obtained are fractional binder content ($\overline{\phi}$) which represents an averaged binder content for each individual particle size class as indicated in

$$\overline{\phi}_{fit,m}(t)=\sum_{L_m\le D_{ijk}<L_{m+1}}\left(\frac{\phi_{ijk}y(s_i,l_j,g_k,t)}{\sum_{i=1}^{n_s}\sum_{j=1}^{n_l}\sum_{k=1}^{n_g}y(s_i,l_j,g_k,t)}\right), \tag{25}$$

where ϕ_{ijk} is the fractional binder content within each finite volume and is obtained from the relation $\phi_{ijk}=l_j/(s_i+l_j+g_k)$ and $\overline{\phi}_{fit,m}$ is the average fractional binder content across all the size classes. In a similar fashion, the third type of data, porosity ($\overline{\varepsilon}_{fit,m}$), is defined to be the averaged porosity of all the granules in each individual size class as shown in

$$\overline{\varepsilon}_{fit,m}(t)=\sum_{L_m\le D_{ijk}<L_{m+1}}\left(\frac{\varepsilon_{ijk}y(s_i,l_j,g_k,t)}{\sum_{i=1}^{n_s}\sum_{j=1}^{n_l}\sum_{k=1}^{n_g}y(s_i,l_j,g_k,t)}\right), \tag{26}$$

where ε_{ijk} is defined to be the porosity within each finite volume and is obtained from the relation $\varepsilon_{ijk}=(l_j+g_k)/(s_i+l_j+g_k)$. Individual objective functions can then be formulated for each of the data attributes to minimize the difference between target data and the simulated model data and are described in

$$\Omega^{size}(\mathbf{p})=\sum_{z=1}^{n_z}\sum_{m=1}^{n_m}(x_{fit,m}(t_z,\mathbf{p})-x_{data,m}(t_z))^2, \tag{27}$$

$$\Omega^{binder}(\mathbf{p})=\sum_{z=1}^{n_z}\sum_{m=1}^{n_m}(\overline{\phi}_{fit,m}(t_z,\mathbf{p})-\overline{\phi}_{data,m}(t_z))^2, \tag{28}$$

$$\Omega^{porosity}(\mathbf{p})=\sum_{z=1}^{n_z}\sum_{m=1}^{n_m}(\overline{\varepsilon}_{fit,m}(t_z,\mathbf{p})-\overline{\varepsilon}_{data,m}(t_z))^2, \tag{29}$$

where *data* denote measured data, *fit* denotes simulated model data, m denotes the mth size class and z denotes the zth time. An overall objective function can then be formulated as described in

$$\Omega^{tot}(\mathbf{p})=\Omega^{size}(\mathbf{p})+\Omega^{binder}(\mathbf{p})+\Omega^{porosity}(\mathbf{p}). \tag{30}$$

4. Results and discussion

4.1. Existing optimization methods

In most studies that deal with multi-dimensional population balance models, the method of parameter estimation has been to fit the experimental/target data by means of systematically varying one parameter at a time until a point where there is a reasonable visual overlap between the data (Darelius et al., 2006; Poon et al., 2009; Ramachandran et al., 2009). This is mainly because of the complex nature of the population balance models which render most of the existing optimization techniques ineffective as will be demonstrated in this section. While the focus of those studies was not on detailed parameter estimation, nevertheless it is important to obtain accurate validated models so that further analysis of the model is precise. Braumann et al. (2010) went one step further and estimated four parameters using a conventional optimization algorithm (Fmincon in *Matlab*®). It must be noted that in their optimization routine, the model response did not originate from the full population balance model but from a first order response surface created from the full model. Using the full PBM in a conventional optimization routine could potentially result in convergence failures and Table 1 shows the various optimization routines (available in *Matlab*®) that can be used in minimizing the overall objective function (see Eq. (30)).

To evaluate the efficacy of the local optimization routines, a simpler model with only breakage and consolidation was considered. A $10\times10\times10$ grid was chosen resulting in 1000 states/bins and three parameters, i.e., $\mathbf{p}=[P_1\ \ P_2\ \ c]^T$. ε_{min} was set at 0.2 and G was set at 200 s^{-1}. P_1, P_2 and c were set at 12, 1.08 and 1.2, respectively. Furthermore, only the sum of squared error between the actual and simulated granule size distributions was considered as the objective function (see Eq. (27)). The 1000 bins were further discretized into 15 size classes (see Eq. (24)) to simulate a realistic scenario of size classification based on sieve analysis. The mapping from bins to size classes is shown in Eq. (24). Subsequently, a volume frequency ($x_{fit,m}$) for each of the size classes was obtained for particular parameter values (as mentioned above). This simulated granule size distribution (GSD) was then chosen as the target GSD. Pursuant to that, the optimization methods were tested by perturbing the parameter values within $\pm20\%$ of their optimal values (which are known a priori since the simulated data first time around is the optimal target distribution). Actual experimental data were not used as the goal was to determine if the optimization routines could converge to optimal parameter values, firstly when one of the three parameters were perturbed, and secondly when all parameters were perturbed at the same time. Using actual experimental data would imply that the optimal parameter set may not be known.

Table 1
Optimization routines used in *Matlab*® version 7.0.0.

No.	Command	Method
1	Fminsearch	Nelder–Mead simplex
2	Fminunc	(a) BFGS quasi-Newton (line search) (b) Newton (trust region)
3	Lsqnonlin	(a) Gauss–Newton (b) Newton (c) Levenberg–Marquardt
4	Fmincon	SQP
5	GA	Genetic algorithm

Fig. 1 depicts the simplest case where the consolidation constant (c) and the one of the breakage rate constants (P_2) was fixed and the other breakage constant (P_1) was perturbed. The final time (t_f) is the only time point depicted. The optimization tolerance for the objective function was set at 1×10^{-3} (using the *Matlab*® function "optimset") to ensure accurate convergence. Results show that for all the optimization routines (as shown in Table 1), convergence was poor and sub-optimal profiles were attained. End-point objective function values were in the region of 0.2–0.3, indicating that the different methods performed equally poorly. Since the local optimization methods (with the exception of the genetic algorithm) are initial guess dependent, a series of different initial guesses were chosen but sub-optimal profiles still resulted (figures not shown).

All model simulations (in this section and future sections) were carried out on a 2.4 GHz Intel Dual Core processor desktop computer with 1 GB RAM using a LINUX platform and a gFORTRAN compiler. The total discretized domain for the simulation comprised $10\times10\times10$ bins along the solid, liquid and gas internal coordinate axes, respectively. The width of each bin along the solid volume is $\Delta s=3.35\times10^{-5}$ m, the width of each bin along the liquid volume is $\Delta l=1.05\times10^{-5}$ m and the width of each bin along the gas volume is $\Delta g=1.5\times10^{-6}$ m. Lower bounds for each coordinate (i.e., s_0, l_0 and g_0) are 0 m^3. Simulation time in all cases was 10 min, with measurements

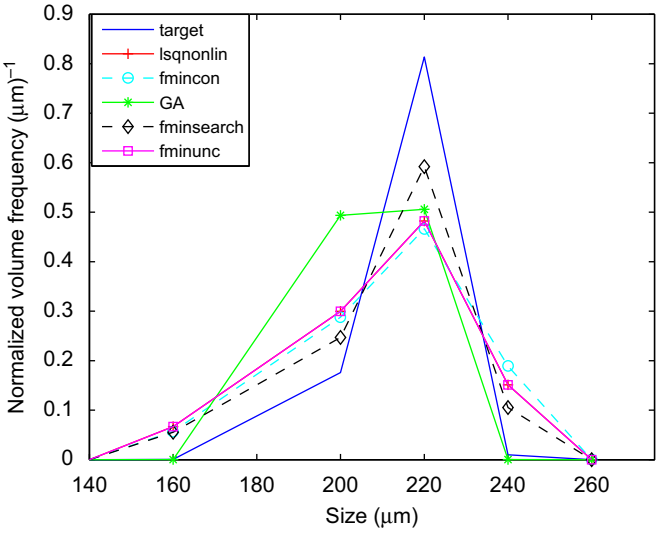

Fig. 1. Target (x_{data}) and simulated (x_{fit}) GSD profiles when P_1 is perturbed by +20% for the purpose of estimating P_1 and matching up with the target GSD, taken at $t_f=10$ min.

being taken at 3, 6 and 10 min (i.e., $n_z=3$). Initial particle size was 260 μm and initial particle population density was 1×10^{12} particles resulting in a 3-D population of distribution of $F'_{8,8,8}=1\times10^{12}$, where the indices indicate the location of the finite volume (i.e., $D_{888}\equiv[(6(s_8+l_8+g_8))/\pi]^{1/3}=260\,\mu m$). In future sections, Eq. (30) was used as the objective function.

4.2. Novel optimization framework

It was hypothesized that the optimization routines failed due to one or more of several reasons. The first reason could be numerical instability during integration. Given that the breakage-consolidation model was solved using the explicit method, it is possible that the Courant–Friedrichs–Lewis (CFL) condition (LeVeque, 2002), which is a necessary condition for numerical stability, was violated. The CFL condition for a 3-D population balance model with a hyperbolic term (advection) is described in

$$\frac{GR\times\Delta t}{\Delta s}+\frac{GR\times\Delta t}{\Delta l}+\frac{GR\times\Delta t}{\Delta g}=CFL<1, \tag{31}$$

where GR is the growth rate (wave velocity), which is the consolidation term in this case. Qualitatively speaking, if a wave (indicated by the arrow) is crossing a discrete grid (see Fig. 2), then the time step must be less than the time taken for the wave to travel to adjacent grid points. Therefore, when the grid point separation is reduced, the upper limit for the time step also decreases.

In the case of the explicit method, the step size control algorithm may not satisfy this necessary condition for stability. As a result spurious oscillations can occur in the simulation as depicted in Figs. 3a and b. This occurs as a result of a violation of the CFL condition during the course of optimization. Implicit methods, on the other hand, are not susceptible to violation of the CFL condition due to the CFL condition not being an issue when the hyperbolic terms are evaluated at time $t=t+1$ (implicit scheme).

The second reason why the optimization might fail is due to the issue of parameter identifiability. When an optimization problem such as parameter estimation is solved, the question of uniqueness of the optimal solution is raised. Given local optimization methods, and the stochastic nature of some global optimization methods such as the genetic algorithm used in this study, the solution that is obtained need not be a global solution. Alternatively, the global solution may not be unique, and a set of alternate parameter values may yield the same or similar fit. Often, the optimization algorithm may terminate in a higher dimensional valley instead of at an isolated global or local minimum. This implies that there exists a subset of the parameter space which does not appear to affect the fit. As a result all parameter values cannot be determined uniquely, given the current data and formulation.

Assuming that the model is accurate and no error is present in the target data, the value of the least-squares objective (see Eq. (20)) at the true solution is zero. Then, the second-order term goes to zero and the first-order sensitivity matrix

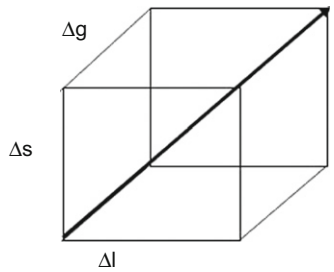

Fig. 2. Schematic of the CFL condition wave propagation.

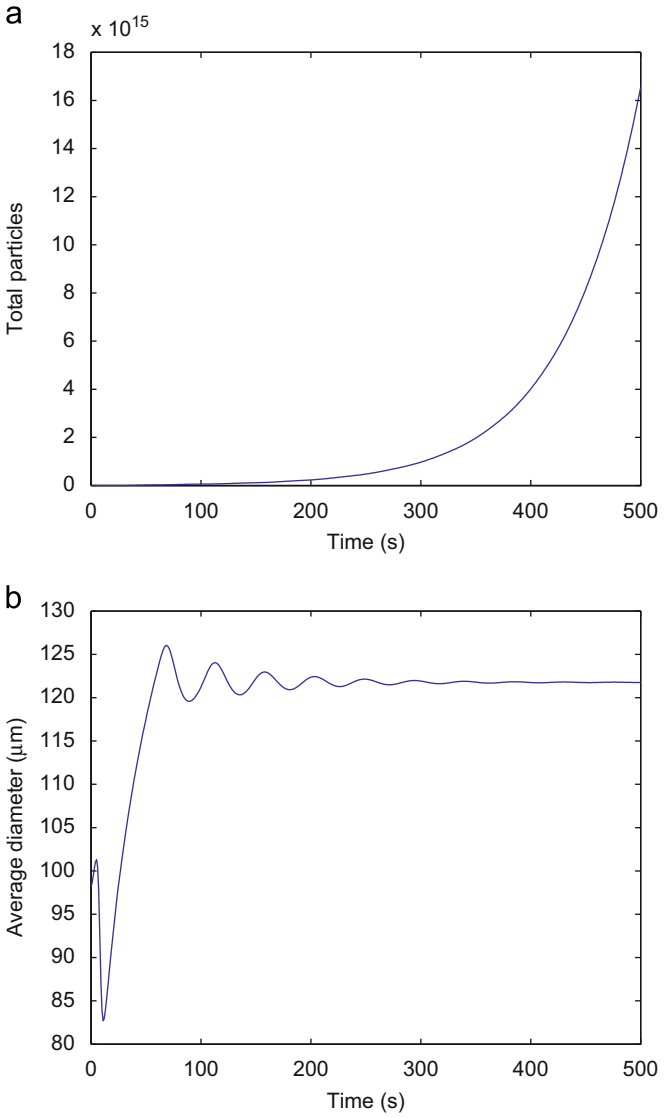

Fig. 3. Evolution of (a) total particles and (b) average diameter, when the CFL condition is violated during the course of optimization.

determines the properties of the Hessian of the objective function at that point as seen in

$$\Omega_{\mathbf{pp}}(\mathbf{p})=\sum_{m=1}^{n_m}\sum_{z=1}^{n_z}2\left(\frac{\partial x_{fit,m}}{\partial\mathbf{p}}(t_z,\mathbf{p})\right)^{\mathrm{T}}\frac{\partial x_{fit,m}}{\partial\mathbf{p}}(t_z,\mathbf{p})-\sum_{m=1}^{n_m}\sum_{z=1}^{n_z}2(x_{data,m}(t_z)-x_{fit,m}(t_z,\mathbf{p}))\frac{\partial^2 x_{fit,m}}{\partial\mathbf{p}^2}(t_z,\mathbf{p}). \tag{32}$$

The first term in Eq. (32) is the Fisher information matrix (FIM) when the measurement errors are mutually independent and follow a normal distribution ($N(0,1)$):

$$\mathbf{FIM}(\mathbf{p})=\sum_{m=1}^{n_m}\left(\frac{\partial\mathbf{x}_m}{\partial\mathbf{p}}(\mathbf{p})\right)^{\mathrm{T}}\frac{\partial\mathbf{x}_m}{\partial\mathbf{p}}(\mathbf{p}). \tag{33}$$

The FIM matrix can be used to evaluate local identifiability (Jacquez and Perry, 1990) and the eigenvalue spectrum is informative and indicates the shape of the objective function surface in the vicinity of a local minimum. If the FIM is nonsingular (i.e., does not have a zero eigenvalue), the whole system is locally identifiable. A parameter variation in any direction results in a change to the objective function $\Omega(\mathbf{p})$. For the breakage-consolidation model described earlier, there was no

zero eigenvalue. This implied that parameter identifiability was not an issue for this problem.

The third cause for optimization failure could be inaccurate derivatives leading to improper gradient calculations. Most of the gradient-based methods such as those described earlier calculate derivatives based on numerical finite differences. Finite differences are very simple to implement in principle, but a central issue is the determination of an appropriate step-size. Too large a step-size would result in a large truncation error and the reverse would result in a large round-off error. Either way, the numerical error can be prohibitively large resulting in inefficient and inaccurate derivatives which further serve to exacerbate the optimization routines. To circumvent the issue of inaccurate derivative via finite differences, one can manually code analytical derivatives, which would not suffer from the inaccuracy caused by finite differences. However, more lines of code are needed and it is harder to maintain. Furthermore, symbolically differentiating certain functions is impractical and nearly impossible; and altogether it is difficult to get accurate and efficient code.

Automatic differentiation (AD) (Griewank, 2000) is a means to evaluate numerically computer code to compute exact values of derivative information which is accurate to machine precision (Tolsma and Barton, 1998). The basic principle of AD is to decompose the computer program into a finite sequence of elementary (unary and binary) operations. Thereafter, each operation if differentiated according to the basic rules of calculus. Finally, the elementary derivatives are accumulated according to the chain rule to compute the desired derivatives. In this study, AD is performed by means of the software DAEPACK (Tolsma and Barton, 2000). DAEPACK is a software library for general numerical calculations. DAEPACK is divided into two major libraries: symbolic analysis and transformation, and numerical calculation. The symbolic analysis and transformation library consists of components for analyzing general Fortran-90 codes and automatically generating codes evaluating symbolic information required when using modern numerical algorithms. One component is that for evaluating analytical derivatives of the code using automatic differentiation (AD). In this study, with the model in question, the overall chain rule is shown in Eq. (34). In conventional optimization routines, the left-hand-side of Eq. (34) is evaluated by finite differences. In this paper, AD is used in two different ways. For the explicit integration method, the entire "algorithm" (including the time integration) for evaluating the objective function is differentiated by AD. For the implicit integration method, the sensitivities+chain rule+AD of the objective function routine are used (see Eq. (34)). AD is also used to set up sensitivity function evaluations (see Eq. (19)) for DSL48S:

$$\frac{\partial \Omega}{\partial \mathbf{p}}(\mathbf{p}) = \sum_{z=1}^{n_z} w_z \sum_{m=1}^{n_m} 2(g_m(\mathbf{y}(t_z,\mathbf{p})) - x_{data,m}(t_z)) \frac{\partial g_m}{\partial \mathbf{y}}(t_z,\mathbf{p}) \frac{\partial \mathbf{y}}{\partial \mathbf{p}}(t_z,\mathbf{p}). \tag{34}$$

A novel optimization decision-based framework is proposed which combines the features of numerical stability, parameter identifiability and automatic differentiation. The rationale is to ensure that any optimization method is performing to the best of its ability, and if not, to ascertain the means of its failure to converge. Fig. 4 illustrates this framework.

In Fig. 4, it can be seen that once the population balance model (PBM) is formulated and discretized, the ODEs may be integrated either by an explicit or implicit method. If the former is used, the code checks to see if the CFL condition is violated. If it is, the implicit method should be used. If the CFL condition is not violated or the implicit method is used, optimization is carried out to attempt to attain convergence. If convergence is attained, the optimal solution is deemed to have been found and the optimization is terminated. If convergence is never attained, the issue of parameter identifiability is investigated. If the FIM matrix is singular, this could imply that for this set of data and formulation, the parameters are not identifiable and further convergence is not possible. If the parameters are identifiable, and convergence is still not attained, this implies that the lack of convergence could be caused by other factors such as nonsmooth functions, discontinuities or being trapped in a sub-optimal local minimum. In all cases, automatic differentiation is preferred to finite differences due to increased accuracy without compromising on computational speed. In the next subsections, we present various cases to test the proposed optimization framework to see if any improvement in convergence is achieved; to see when convergence may not be achieved and to suggest alternative means of optimization to enable convergence.

4.2.1. Cases 1–4

Table 2 summarizes the various models used to test the optimization framework and within each model: (1) a different subset of parameters are optimized, (2) a combination of explicit/implicit methods and finite difference/automatic differentiation was used and (3) the initial, final and optimal values for the parameters are shown. BR signifies breakage, CON signifies consolidation, AGG signifies aggregation and NUC signifies nucleation. These cases realistically cover what one would encounter when dealing with parameter estimation related to population balances, specifically granulation processes.

The first case replicates the model used in Section 4.1. In Section 4.1, it could be seen that poor convergence was obtained using the different conventional optimization routines. Using the optimization framework described in Fig. 4, it was seen that when the explicit method was used, the CFL condition was violated (see Figs. 3a and b). As a result, the implicit method was chosen which circumvented this issue. When finite differences were used in combination with the implicit method, the expected convergence was not obtained although marginal improvement was reported (see Case 3). Therefore, parameter identifiability was performed to check if convergence could be achieved in principle. Since the matrix was not singular, the hypothesis was that inaccurate derivatives due to finite differences could be the cause of poor convergence. Another possible reason for failed convergence could be that the objective function was nonsmooth. However, this was shown not to be the case as seen in Fig. 5 where it can be seen that the function is smooth and convex near the minimum. In Fig. 5, it can also be observed that there is a very small slope formed as P_1 increases. The presence of this small slope results in ill-conditioning which is a potential cause of convergence failure when derivative information is not calculated accurately. However, when AD was then used to determine the gradients to be used in the optimization method (SQP), which was performed by SNOPT (Gill et al., 2005), convergence was observed satisfying the optimization tolerance (see Case 4).

4.2.2. Cases 5–32

To further test the proposed optimization framework, various other cases were considered involving combinations of the different rate processes and parameters (see Table 2). Results show that in Cases 5–32, when AD was used in combination with the implicit method, the expected convergence was achieved. This is in contrast to the poor convergence observed when using conventional optimization routines based on finite differences in combination with either explicit or implicit methods. Alternatively, the implicit method in combination with AD could have used in every case to circumvent the issue of CFL condition

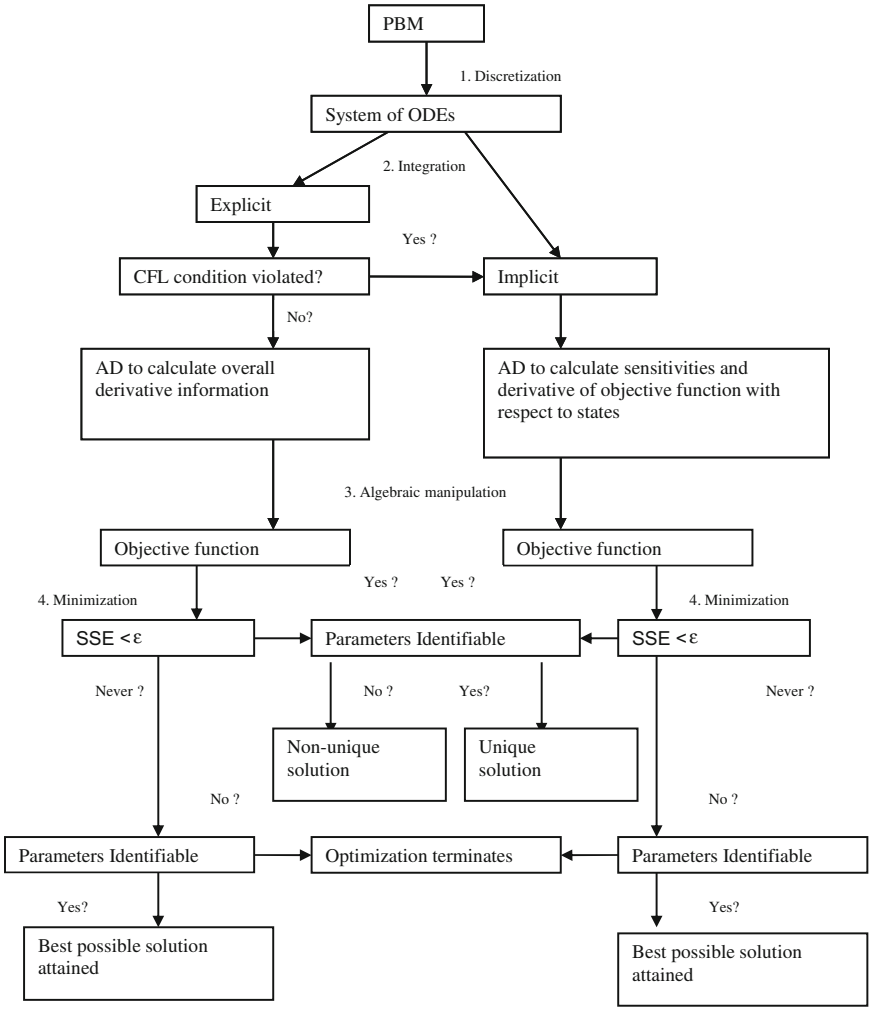

Fig. 4. Decision-based optimization framework.

violation. The tradeoff is that if the CFL condition is not prone to being violated, the explicit method may yield savings in computational time as the time step can be cheaper. Similarly, FD may be slightly computationally cheaper than AD if the explicit method is used. However, should the CFL condition be rate limiting, the time step is forced (by the CFL condition) to be smaller and as a result computational time increases. In Cases 5–28, finite differences performed poorly although they were effective in Cases 29–32. This could be attributed to the nature of the consolidation model which is simpler in structure than the breakage/aggregation/nucleation models. It is interesting to note that the explicit method in combination with AD also performed poorly. Since AD is now used to differentiate the entire algorithm to evaluate the objective function, accurate derivative information may be lost while differentiating the user-built error control during the integration. This has also been verified by Eberhard and Bischof (1999).

The other cause of failed convergence is parameter identifiability. In this study, the numerical experiments were based on perturbing the parameter values and ascertaining if they converge to the target solution (which is known a priori). As a result, parameter identifiability can be performed on this known solution to establish if the parameters are locally identifiable. In a more realistic scenario, the optimal solution set is not known and hence parameter identifiability cannot be determined a priori. If the optimization has stopped converging and the optimization tolerance is not met, local parameter identifiability is investigated. If the parameters are not locally identifiable, the optimization is terminated as a better solution cannot be obtained. This increases the efficiency of the method. In Cases 5–32, the parameters were identifiable and therefore the issue of parameters not being identifiable did not apply. But, a good code should converge reliably to one of the nonunique solutions, so there is still a need to check identifiability even if converged.

Based on the results so far and given that we are not dealing with nonsmooth functions, the most likely cause of poor convergence was due to improper and inaccurate first-order derivatives, so a series of recommendations may be made from the results of the numerical experiments. If the finite differences are computationally cheaper, it can be used as a first choice and switches to AD should be made if finite differences yield poor convergence. If both methods are equally cheap, then AD should be used. In the cases depicted here, there were no savings between the methods, hence AD was used for all the cases. However, finite differences were still used to compare the efficacy of AD over finite differences. A similar comparison can be made between the explicit and implicit methods. Should there be no CFL condition violation, explicit methods may be used. Should the CFL condition be violated, implicit methods should be used to ensure better convergence. Ideally, barring no significant increase in computational time as seen in this study, a combination of AD and implicit methods yield significant improvements in convergence.

Table 2
Optimization results from cases 1–32 for parameter(s) perturbation of +20%.

Case	Model	Parameters	Method	Initial set	Final set	Optimal set	Function value
1	BR-CON	$[P_1]$	Explicit/FD	[12]	[11.51]	[10]	0.29
2	BR-CON	$[P_1]$	Explicit/AD	[12]	[10.92]	[10]	0.20
3	BR-CON	$[P_1]$	Implicit/FD	[12]	[10.53]	[10]	0.12
4	BR-CON	$[P_1]$	Implicit/AD	[12]	[10.03]	[10]	0.0004
5	BR-CON	$[P_1\ c]^T$	Explicit/FD	$[12\ 1.2]^T$	$[11.53\ 1.16]^T$	$[10\ 1]^T$	0.34
6	BR-CON	$[P_1\ c]^T$	Explicit/AD	$[12\ 1.2]^T$	$[11.36\ 1.13]^T$	$[10\ 1]^T$	0.25
7	BR-CON	$[P_1\ c]^T$	Implicit/FD	$[12\ 1.2]^T$	$[11.04\ 1.08]^T$	$[10\ 1]^T$	0.18
8	BR-CON	$[P_1\ c]^T$	Implicit/AD	$[12\ 1.2]^T$	$[10.03\ 1.01]^T$	$[10\ 1]^T$	0.0005
9	BR-CON	$[P_2\ c]^T$	Explicit/FD	$[1.08\ 1.2]^T$	$[0.97\ 1.15]^T$	$[0.90\ 1]^T$	0.51
10	BR-CON	$[P_2\ c]^T$	Explicit/AD	$[1.08\ 1.2]^T$	$[0.96\ 1.14]^T$	$[0.90\ 1]^T$	0.49
11	BR-CON	$[P_2\ c]^T$	Implicit/FD	$[1.08\ 1.2]^T$	$[0.95\ 1.14]^T$	$[0.90\ 1]^T$	0.48
12	BR-CON	$[P_2\ c]^T$	Implicit/AD	$[1.08\ 1.2]^T$	$[0.89\ 1.02]^T$	$[0.90\ 1]^T$	0.0006
13	BR-CON	$[P_1\ P_2]^T$	Explicit/FD	$[12\ 1.08]^T$	$[11.21\ 0.99]^T$	$[10\ 0.90]^T$	0.86
14	BR-CON	$[P_1\ P_2]^T$	Explicit/AD	$[12\ 1.08]^T$	$[11.12\ 0.97]^T$	$[10\ 0.90]^T$	0.72
15	BR-CON	$[P_1\ P_2]^T$	Implicit/FD	$[12\ 1.08]^T$	$[11.04\ 0.95]^T$	$[10\ 0.90]^T$	0.59
16	BR-CON	$[P_1\ P_2]^T$	Implicit/AD	$[12\ 1.08]^T$	$[10.04\ 0.89]^T$	$[10\ 0.90]^T$	0.0007
17	BR-CON	$[P_1\ P_2\ c]^T$	Explicit/FD	$[12\ 1.08\ 1.2]^T$	$[11.20\ 0.97\ 1.13]^T$	$[10\ 0.90\ 1]^T$	0.75
18	BR-CON	$[P_1\ P_2\ c]^T$	Explicit/AD	$[12\ 1.08\ 1.2]^T$	$[11.17\ 0.96\ 1.10]^T$	$[10\ 0.90\ 1]^T$	0.71
19	BR-CON	$[P_1\ P_2\ c]^T$	Implicit/FD	$[12\ 1.08\ 1.2]^T$	$[11.15\ 0.95\ 1.07]^T$	$[10\ 0.90\ 1]^T$	0.69
20	BR-CON	$[P_1\ P_2\ c]^T$	Implicit/AD	$[12\ 1.08\ 1.2]^T$	$[10.05\ 0.89\ 1.02]^T$	$[10\ 0.90\ 1]^T$	0.0008
21	BR-CON-AGG	$[P_1\ P_2\ c\ \beta_0]^T$	Explicit/FD	$[12\ 1.08\ 1.2\ 1.2]^T$	$[11.22\ 0.97\ 1.15\ 1.18$	$[10\ 0.90\ 1\ 1]^T$	0.81
22	BR-CON-AGG	$[P_1\ P_2\ c\ \beta_0]^T$	Explicit/AD	$[12\ 1.08\ 1.2\ 1.2]^T$	$[11.15\ 0.95\ 1.12\ 1.16$	$[10\ 0.90\ 1\ 1]^T$	0.69
23	BR-CON-AGG	$[P_1\ P_2\ c\ \beta_0]^T$	Implicit/FD	$[12\ 1.08\ 1.2\ 1.2]^T$	$[11.12\ 0.94\ 1.10\ 1.14]^T$	$[10\ 0.90\ 1\ 1]^T$	0.60
24	BR-CON-AGG	$[P_1\ P_2\ c\ \beta_0]^T$	Implicit/AD	$[12\ 1.08\ 1.2\ 1.2]^T$	$[10.04\ 0.92\ 1.05\ 1.04]^T$	$[10\ 0.90\ 1\ 1]^T$	0.0009
25	BR-CON-AGG-NUC	$[P_1\ P_2\ c\ \beta_0\ B_0]^T$	Explicit/FD	$[12\ 1.08\ 1.2\ 1.2\ 1.2]^T$	$[11.27\ 0.97\ 1.16\ 1.17\ 1.18]^T$	$[10\ 0.90\ 1\ 1\ 1]^T$	0.94
26	BR-CON-AGG-NUC	$[P_1\ P_2\ c\ \beta_0\ B_0]^T$	Explicit/AD	$[12\ 1.08\ 1.2\ 1.2\ 1.2]^T$	$[11.22\ 0.96\ 1.14\ 1.15\ 1.16]^T$	$[10\ 0.90\ 1\ 1\ 1]^T$	0.78
27	BR-CON-AGG-NUC	$[P_1\ P_2\ c\ \beta_0\ B_0]^T$	Implicit/FD	$[12\ 1.08\ 1.2\ 1.2\ 1.2]^T$	$[11.17\ 0.94\ 1.10\ 1.11\ 1.15]^T$	$[10\ 0.90\ 1\ 1\ 1]^T$	0.67
28	BR-CON-AGG-NUC	$[P_1\ P_2\ c\ \beta_0\ B_0]^T$	Implicit/AD	$[12\ 1.08\ 1.2\ 1.2\ 1.2]^T$	$[10.11\ 0.92\ 1.04\ 1.06\ 1.04]^T$	$[10\ 0.90\ 1\ 1\ 1]^T$	0.02
29	CON	[c]	Explicit/FD	[1.2]	[1.02]	[1]	0.0003
30	CON	[c]	Explicit/AD	[1.2]	[1.02]	[1]	0.0003
31	CON	[c]	Implicit/FD	[1.2]	[1.02]	[1]	0.0003
32	CON	[c]	Implicit/AD	[1.2]	[1.01]	[1]	0.0002

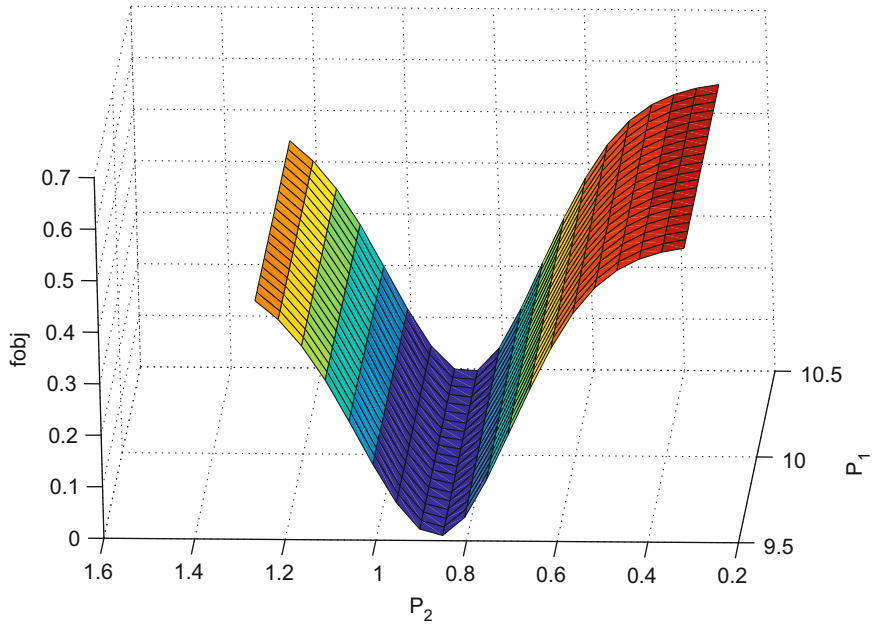

Fig. 5. Shape of the objective function with respect to parameters P_1 and P_2. Interior point (10,0.90) is the optimal point.

It should be noted, however, that the explicit integration method performed poorly in almost all cases thus corroborating the need for other integration and optimization approaches. In all cases (1–32), parameters were perturbed +20% as a representative of a typical perturbation. Perturbations of −20%, +10% and −10% were also considered and results obtained were very similar and thus not shown.

5. Conclusions

Effective parameter estimation is crucial in mathematical modeling, such that the validated model realistically mimics the actual process. In doing so, the model can then be used for further analysis to arrive at important structural and parametric decisions. Particulate processes are ubiquitous in many industries

(e.g., pharmaceutical) and should be represented realistically by multi-dimensional population balance models. However, the complex nature of such models often render existing parameter estimation and optimization methods inadequate. In this study, using granulation as an example, a multi-dimensional population balance is formulated along with an objective function. Results then showed that conventional optimization routines (local and global; gradient-based and nongradient based) failed to yield suitable convergence. Numerical experiments were then proposed which utilize a decision-based approach to make a series of recommendations to ensure that any existing optimization method is performing to the best of its ability and if not, to understand why. Results confirmed significant improvements in attaining the optimal solution. The idea of the framework was to also ensure that the different issues that pertain to convergence are evaluated at each decision point. This is to ensure that unnecessary time is not spent to attain an optimal solution that could have otherwise been attained at a reduced time and cost. The framework discussed so far is primarily suited to smooth functions. It is entirely possible that other models can result in nonsmooth functions. In this case, alternative methods can be incorporated for the purpose of optimization, as seen in the work of Yunt and Barton (2009).

It should also be noted that all methods described in this paper locate local solutions of the parameter estimation problem. Deterministic global optimization methods for parameter estimation (e.g., Singer et al., 2006) have begun to emerge in recent years, and the application of these methods to population balance models is an interesting avenue of further research.

Acknowledgement

The authors would like to acknowledge financial and technical support provided by Novartis.

References

Adetayo, A.A., Ennis, B.J., 2000. A new approach to modeling granulation processes for simulation and control purposes. Powder Technology 108, 202–209.
Annapragada, A., Neilly, J., 1996. On the modelling of granulation processes: a short note. Powder Technology 89, 83–84.
Biggs, C.A., Sanders, C., Scott, A.C., Willemse, A.W., Hoffman, A.C., Instone, T., Salman, A.D., Hounslow, M.J., 2003. Coupling granule properties and granulation rates in high shear granulation. Powder Technology 130, 162–168.
Braumann, A., Kraft, M., Mort, P.R., 2010. Parameter estimation in a multi-dimensional granulation model. Powder Technology 197, 196–210.
Christofides, P.D., 2001. Control of nonlinear distributed process systems: recent development and challenges. A.I.Ch.E. Journal 47, 514–518.
Darelius, A., Brage, H., Rasmuson, A., Bjorn, I.N., Folestad, S., 2006. A volume based multi-dimensional population balance approach for modelling high shear granulation. Chemical Engineering Science 61, 2482–2493.
Darelius, A., Rasmuson, A., Bjorn, I.N., Folestad, S., 2005. High shear wet granulation modelling—a mechanistic approach using population balances. Powder Technology 160, 209–218.
Eberhard, P., Bischof, C.H., 1999. Automatic differentiation of numerical integration algorithms. Mathematics of Computation 68, 717–731.
Feehery, W.F., Tolsma, J.E., Barton, P.I., 1997. Efficient sensitivity analysis of large scale differential-algebraic systems. Applied Numerical Mathematics 25, 41–54.
Gill, P.E., Murray, W., Saunders, M.A., 2005. SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Review 47, 99–131.
Griewank, A., 2000. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. SIAM.
Hoornaert, F., Wauters, P.A.L., Meesters, G.M.H., Pratsinis, S.E., Scarlett, B., 1998. Agglomeration behaviour of powders in a Lodige mixer granulator. Powder Technology 96, 116–128.
Hounslow, M.J., 1998. The population balance as a tool for understanding particle rate processes. KONA 16, 179–193.
Immanuel, C.D., Doyle III, F.J., 2003. Computationally efficient solution of population balance models incorporating nucleation, growth and coagulation: application to emulsion polymerization. Chemical Engineering Science 52, 3681–3698.
Immanuel, C.D., Doyle III, F.J., 2005. Solution technique for a multi-dimensional population balance model describing granulation processes. Powder Technology 156, 213–225.
Iveson, S.M., 2002. Limitations of one-dimensional population balance models of wet granulation processes. Powder Technology 124, 219–229.
Iveson, S.M., Litster, J.D., 1998. Growth regime map for liquid-bound granules. A.I.Ch.E. Journal 113, 1510–1518.
Jacquez, J.A., Perry, T., 1990. Parameter estimation: local identifiability of parameters. American Physiological Society 42, 19–30.
Knight, P.C., 1993. An investigation into the kinetics of granulation using a high shear mixer. Powder Technology 77, 159–169.
Knight, P.C., Instone, T., Pearson, J.M.K., Hounslow, M.J., 1998. An investigation into the kinetics of liquid distribution and growth in high shear mixer agglomeration. Powder Technology 97, 246–257.
Kristensen, H.G., 1996. Particle agglomeration in high shear mixers. Powder Technology 88, 197–202.
LeVeque, R.J., 2002. Finite Volume Methods for Hyperbolic Problems. Cambridge University Press.
Liu, L.X., Litster, J.D., 2002. Population balance modelling of granulation with a physically based coalescence kernel. Chemical Engineering Science 57, 2183–2191.
Ma, D.L., Tafti, D.K., Braatz, R.D., 2002. High resolution simulation of multi-dimensional crystallization. Industrial and Engineering Chemistry Research 41, 6217–6223.
Madec, L., Falk, L., Plasari, E., 2003. Modelling of the agglomeration in suspension process with multidimensional kernels. Powder Technology 130, 147–153.
Mantzaris, N.V., Daoutidis, P., Srienc, F., 2001. Numerical solution of multi variable cell population balance models: 3. Finite difference methods. Computers and Chemical Engineering 25, 1463–1481.
Pandya, J.D., Spielman, L.A., 1983. Floc breakage in agitated suspensions: effect of agitation rate. Chemical Engineering Science 38, 1983–1992.
Pinto, M.A., 2008. Modelling and control of biological systems. Ph.D. Thesis, Imperial College London.
Pinto, M.A., Immanuel, C.D., Doyle III, F.J., 2007. A feasible solution technique for higher-dimensional population balance models. Computers and Chemical Engineering 31, 1242–1256.
Pinto, M.A., Immanuel, C.D., Doyle III, F.J., 2008. A two-level discretisation algorithm for the efficient solution of higher-dimensional population balance models. Chemical Engineering Science 63, 1304–1314.
Poon, J.M.H., Ramachandran, R., Sanders, C.F.W., Glaser, T., Immanuel, C.D., Doyle III, F.J., Litster, J.D., Stepanek, F., Wang, F.Y., Cameron, I.T., 2009. Experimental validation studies on a multi-scale and multi-dimensional population balance model of batch granulation. Chemical Engineering Science 64, 775–786.
Ramachandran, R., Immanuel, C.D., Stepanek, F., Litster, J.D., Doyle III, F.J., 2009. A mechanistic model for granule breakage in population balances of granulation: theoretical kernel development and experimental validation. Chemical Engineering Research and Design 87, 598–614.
Ramachandran, R., Poon, J.M.H., Sanders, C.F.W., Glaser, T., Immanuel, C.D., Doyle III, F.J., Litster, J.D., Stepanek, F., Wang, F.Y., Cameron, I.T., 2008. Experimental studies on distributions on granule size, binder content and porosity in batch drum granulation: inferences on process modelling requirements and process sensitivities. Powder Technology 188, 89–101.
Ramkrishna, D., 2000. Population Balances. Academic Press, San Diego.
Reynolds, G.K., Biggs, C.A., Salman, A.D., Hounslow, M.J., 2004. Non-uniformity of binder distribution in high-shear granulation. Powder Technology 140, 203–208.
Sastry, K.V.S., Fuerstenau, D.W., 1973. Mechanisms of agglomerate growth in green pelletization. Powder Technology 7, 97–105.
Singer, A.B., Taylor, J.W., Barton, P.I., Green, W.H., 2006. Global dynamic optimization for parameter estimation in chemical kinetics. Journal of Physical Chemistry A 110 (3), 971–976.
Tan, H.S., Salman, A.D., Hounslow, M.J., 2005. Kinetics of fluidised bed granulation V: simultaneous modelling of aggregation and breakage. Chemical Engineering Science 60, 3847–3866.
Tolsma, J.E., Barton, P.I., 1998. On computational differentiation. Computers and Chemical Engineering 22, 475–490.
Tolsma, J.E., Barton, P.I., 2000. DAEPACK: an open modeling environment for legacy models. Industrial Engineering and Chemistry Research 39, 1826–1839.
Verkoeijen, D., Pouw, G.A., Meesters, G.M.H., Scarlett, B., 2002. Population balances for particulate processes—a volume approach. Chemical Engineering Science 57, 2287–2303.
Wang, F.Y., Cameron, I.T., 2007. A multi-form modelling approach to the dynamics and control of drum granulation processes. Powder Technology 179, 2–11.
Wauters, P.A.L., 2000. Modelling and mechanisms of granulation. Ph.D. Thesis, Delft University of Technology.
Yunt, M., Barton, P.I., 2009. Nonsmooth optimization of systems with varying structure. In: Proceedings of IFAC Symposium on Advanced Control of Chemical Processes, Istanbul, Turkey, pp. 537–544.