The Inverse Problem in Granulation Modeling—Two Different Statistical Approaches

Andreas Braumann, Peter L. W. Man, and Markus Kraft
Dept. of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, U.K.

DOI 10.1002/aic.12526
Published online February 15, 2011 in Wiley Online Library (wileyonlinelibrary.com).

This article is concerned with parameter estimation for a multidimensional population balance model for granulation. Experimental results were obtained by running a laboratory mixer with sodium carbonate and aqueous polyethylene glycol solutions. Subsequently, a prescan of suitable parameter combinations utilising the experimental results is performed, and a local surrogate model constructed around the best combination. For the actual estimation of the parameters and their uncertainties two different approaches are applied—a projection method and a Bayesian approach. It is found that the model predictions with the parameters obtained through both methods are similar. Furthermore, the uncertainties in the model predictions increase as the experimental uncertainties are increased. Studies of the marginal densities of two-parameter combinations obtained through the Bayesian approach show a correlation between the collision and breakage rate constant, giving potential hints for further model development. Furthermore, a bimodal distribution of the compaction rate constant is observed.
 AIChE J, 57: 3105–3121, 2011
Keywords: mathematical modeling, process simulation, particle Technology, statistical analysis

Introduction

Granulation, or agglomeration, is a well established unit operation in the chemical industry, in which small particles are transformed into bigger entities, called granules.[1] The widespread use of granulation across industries, e.g., pharmaceutical, chemical, and food can be attributed to the range of benefits from the application of this unit operation. Granules usually feature enhanced handling and application properties compared to the precursors. This means, they are not only better for storage and transport (flowability, dustiness), they also show favorable features in the application (controlled release, reduced segregation).[2]

If the assembly of fine powders into granules is aided by the addition of liquid, the process is called wet granulation. This process is performed in a variety of apparatus such as drums, fluidized beds and mixers.[3–6] Systematic (experimental) studies of wet granulation were performed as early as in the 1950s by Capes and Danckwerts,[7,8] Newitt and Conway-Jones,[9] and Rumpf.[10,11] Over the decades numerous other studies have been conducted, examining the influence of the material properties such as particle size,[12] binder properties[13] and binder addition method,[14,15] to name but a few.

Besides experimental studies, mathematical descriptions of the granulation process have also been pursued. Such a kind of description is eventually needed for the control of the

Correspondence concerning this article should be addressed to M. Kraft at mk306@cam.ac.uk.

manufacturing process.[16,17] Amongst others, population balance models are the most widespread for the mathematical description of granulation processes, but can also be found in modeling of other processes such as nanoparticle synthesis.[18–20] Population balance models take account of the properties of the individual particles making up the bulk. Early models accounted for only one granule property, namely the volume,[21,22] with only coalescence as a transformation. Advances in these models saw an increase in the number of internal coordinates to three and more as well as the incorporation of transformations such as liquid addition, breakage, compaction, and reaction.[23–26]

Irrespective of the kind of model used, one is faced with the problem that some model parameters are unknown. To establish their values, experiments have to be conducted and the inverse problem solved. For relatively simple models such as the coalescence of droplets in a rotating disc contactor, Monte Carlo based algorithms have been applied for identification and sensitivity analysis.[27–29] A more fundamental algorithmic development for the sensitivity of coagulation only processes has been presented by Man et al.[30] and Bailleul et al.[31] In more general terms, the inverse problem for population balance models has been the subject of research for many years, for instance for the determination of rate constants in agglomeration problems,[32] crystallization[33] and sludge flocculation.[34] More recently, the parameter estimation for a three-dimensional population balance model framework was reported by Ramachandran and Barton.[35] They developed a framework that addresses the numerical stability of the optimization, being challenged by the solving of the discretized population balance model. The decision-based approach makes use of the Fisher-Information matrix to decide about identifiability of the parameters and attend also to the way the derivatives of the objective function is obtained for the used gradient based optimization routines. Braumann et al.[36] presented an approach using a surrogate model to solve the inverse problem for a multivariate population balance model for granulation. Given that experimental findings possess uncertainties, the methodology was developed further,[37,38] to estimate the unknown parameters along with their uncertainties, building on findings by Sheen et al.[39] These uncertainties then carry on through the model response, enabling further investigation such as model discrimination. The existing methodology is somewhat restrictive in the sense that the scientist concerned with the problem must have a good initial guess for the values of the unknown parameters. However, in the absence of this knowledge, a more global approach should be pursued. Secondly, the inverse problem might not have a unique solution, i.e., it is not identifiable. In addition, there might be correlations between the different model parameters.

The purpose of this article is to present an integrated approach for estimating unknown parameters in a multidimensional population balance model for granulation, based on experimental data. This inverse problem is solved using a three-stage approach. Firstly, a global search over the parameter space using quasi-random sequences is performed. Secondly, the model is approximated locally with response surfaces around the best candidate for the unknown parameters. Thirdly, the unknown parameters and their uncertainties are estimated using both the existing methodology of Braumann et al.[38] and a Bayesian approach, allowing for a direct comparison of both methods.

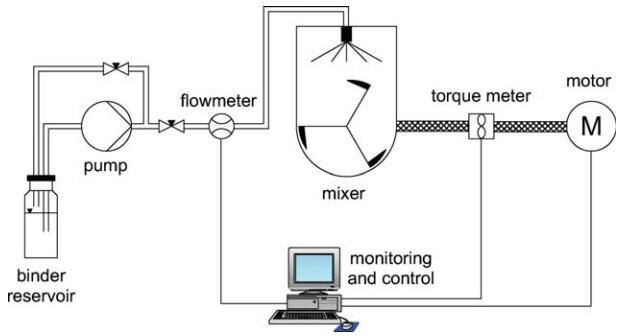

Figure 1. Experimental setup.

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

This article is organized into five sections. In the first section we introduce a process setup along with some experimental results which were obtained by using a bench scale mixer. This is followed in the second section by brief comments about the population balance model that is applied to this process. In the next section the theoretical foundation to remove restrictions of the existing methodology is laid out. The global search method based on low discrepancy series is introduced and combined with the existing approach for parameter estimation. Subsequently, we lift the Gaussian assumption on the parameter distribution by applying the Bayesian approach. This kind of statistical approach has previously found application in parameter estimation problems, for instance for heat transfer[40] and crystallization.[41] Amalgamating the model, experimental results and the extended methodology, results are presented and discussed in the results section. Finally, conclusions are drawn in the last section.

Experiments

The subject of this study is a wet granulation process. In such a process, powder(s) and liquid(s) are mixed under certain conditions, resulting in particles consisting of different materials. Here, we study the granulation of anhydrous soda carbonate with aqueous polyethylene glycol (PEG) mixtures in a bench scale mixer.

Setup

The experiments were carried out in a 5 litre laboratory ploughshare mixer (Kemutec), with further details about its design given by Jones and Bridgwater.[42] The mixer with its horizontal shaft is driven by a DC motor and can be operated at variable speed. A torque meter with integrated tacho (DRBK-n, ETH Messtechnik, Germany) is mounted between the shaft and the motor (Figure 1).

Liquid (binder) is added to the powder in the mixing chamber via a single fluid nozzle (Model 121, Düsen-Schlick, Germany), after being taken out of the binder reservoir with a diaphragm pump (AD4/90, Totton, UK). The binder flowrate is measured with a flowmeter (OG1, Nixon Flowmeters, UK). For data recording and control purposes a

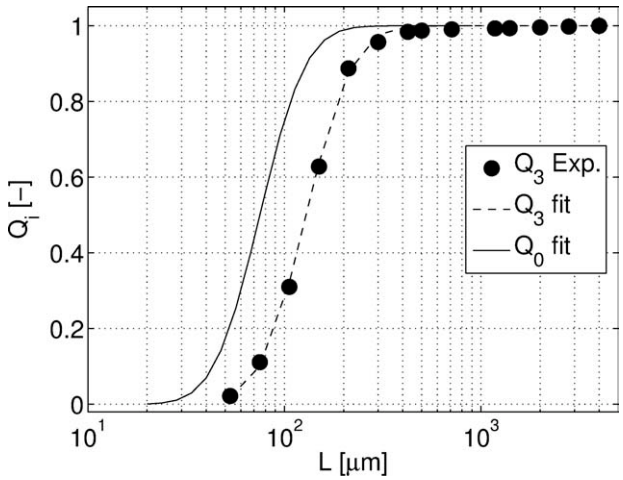

Figure 2. Normalized cumulative mass and number size distributions of anhydrous sodium carbonate.

LabView 7.1 application has been written, and the respective communication with the devices was facilitated with data recording cards (6009 and 6601, National Instruments). To control the speed of the motor driving the mixer, the speed signal (from the torque meter) was put through a LabView application facilitating the output of the required voltage to the motor by a DC motor controller (Model 506, Parker SSD).

Materials

The experiments were performed with anhydrous sodium carbonate and three mixtures of deionized water and polyethylene glycol 4000 (Breckland Scientific). Sieve analysis of the sodium carbonate yields the cumulative mass distribution Q_3 shown in Figure 2.

The cumulative number size distribution Q_0 is needed for the initialization of the population balance model. This can be estimated by firstly fitting a lognormal distribution to the experimental data, and then carrying out the conversion into the cumulative number size distribution Q_0,

$$Q_0(L) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{\ln(L/\mu_{\mathrm{geo},0})}{\sqrt{2}\,\ln(\sigma_{\mathrm{geo}})}\right) + 1\right] \qquad (L > 0) \qquad (1)$$

where $\mu_{\mathrm{geo},0} = 75\ \mu$m and $\sigma_{\mathrm{geo}} = 1.53$.

Varying the composition of the aqueous PEG4000 mixtures results in a change of density and viscosity. These two properties have been determined for different mixtures with a PEG4000 fraction between 10 and 50 wt%, and clear dependencies of the density and viscosity on the composition can be observed (Figures 3 and 4).

The viscosity measurements were performed at shear rates between 11 and 700 s^{-1} and a Newtonian behavior for all examined solutions was observed.

Procedure

Upon loading the mixer with sodium carbonate powder, the mixer was run for roughly one minute to aerate the powder. At the same time the binder was pumped from the reservoir through the nozzle into a cap to establish a constant spray pattern. The cap was then removed and binder added for 5 minutes, while keeping the shaft speed constant at 120, 150 or 180 rpm. The spray rate was chosen such as to give 25 wt% binder in the final mixture. After the 5 minutes, the spraying and mixing was stopped and the material withdrawn from the mixing chamber. The material was then split with a sample splitter and about 400 g of the product was analysed by sieving with a standard $\sqrt{2}$ sieve stack (BS410).

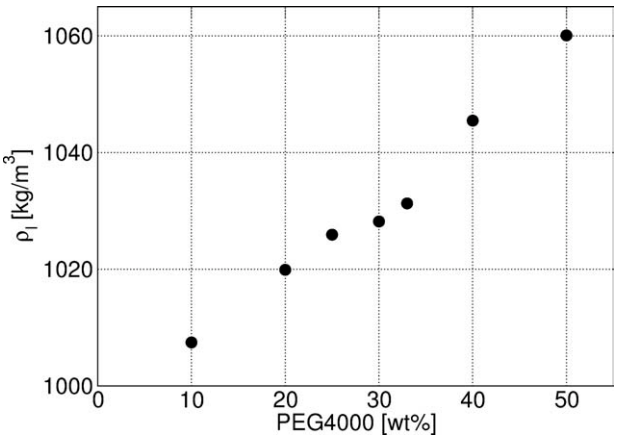

Figure 3. Measured density of aqueous PEG4000 solutions.

Results

A mass size distribution was deduced from the sieving results. To ease comparison between the different scenarios, each result is represented by the geometric mass mean size $L^{\mathrm{exp}}_{\mathrm{m,geo},3}$,

$$L^{\mathrm{exp}}_{\mathrm{m,geo},3} = \exp\left(\frac{\sum_{i=1}^{n} m_i \ln L_i}{\sum_{i=1}^{n} m_i}\right), \qquad (2)$$

obtained from n classes, where m_i is the mass of particles in the class i and L_i the arithmetic mean of the class i (calculated from mesh size of current and upper sieve). The resulting

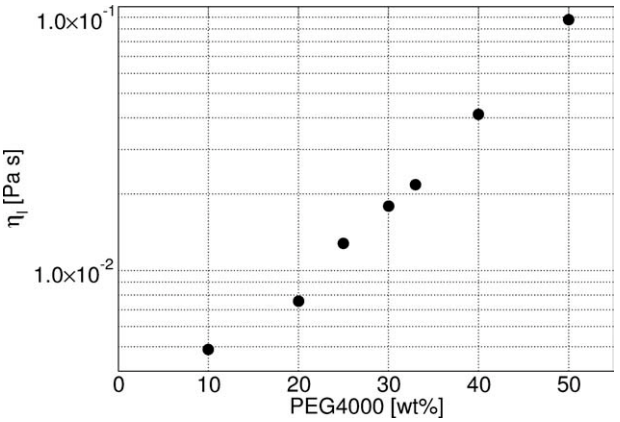

Figure 4. Measured viscosity of aqueous PEG4000 solutions.

Table 1. Experimental Geometric Mass Mean Size $L^{exp}_{m,geo,3}$ in μm for Varying Impeller Speed and Binder Composition at $t = 300$ s

Impeller Speed [rpm]	PEG 4000 Fraction [wt%]		
	10	25	33
120	190.9	191.3	173.3
150	196.2	190.1	173.5
180	197.7	229.5	186.0

values for the different operating conditions are summarized in Table 1.

Modeling of the Granulation Process

The granulation process shall be modelled with a multidimensional population balance model for a concentrated system. In this, the temporal evolution of properties of entities (here granules) of a population is tracked. The volumes of the five different components of a particle are chosen as properties, i.e., as internal coordinates, thereby tracking the particle composition. For this system, following transformations (physical and chemical processes that move a particle in type space) are taken into account,

- Addition of liquid
- Coalescence of particles, with the collision rate constant $\hat{K}_0$
- Compaction of particles, with the compaction rate constant k_{porred}
- Breakage, with the kinetic constant $\hat{k}_{att}$
- Penetration, with the penetration rate constant $\hat{k}_{pen}$
- Chemical reaction, with the rate constant k_{reac},

with a summary of the particle model and transformations being given in Figure 5.

For conciseness, no further details of the population balance model are provided here, however a detailed mathematical description of the model is given by Braumann et al.,[43] and a fuller explanation of the underlying physics of the submodels is given by Braumann et al.[23,36] The population balance model is solved numerically using the LPDA algorithm.[44]

Parameter Estimation Theory

Given the mathematical model and the experimental data, we wish to estimate unknown model parameters. In this study, the rate constants for the coalescence, compaction, breakage, penetration and reaction are unknown. These constants should be universal for the studied system such as the process conditions are incorporated in the mathematical descriptions of the transformations. Owing to experimental uncertainties, the estimates for the model parameters will have uncertainties as well.

The solving of the inverse problem for a multidimensional population balance model for wet granulation has been done in previous studies,[37,38] where the model parameters were assumed to be normally distributed. Their defining parameters are then estimated by linking the experimental observations and the model responses in an appropriate objective function. The model responses are obtained from surrogate models, in this case second order response surfaces constructed over ranges where the values of the unknown model parameters are assumed to be found. These response surfaces allow for a much quicker computation of the model response compared to an evaluation of the complex population balance model. However, one needs to keep in mind that response surfaces are only local approximations, so we propose an extension to the methodology. In a first step, we perform a search over a wide range of possible parameter values, identifying the "best" parameter combination by means of an appropriate objective function. In the second step, response surfaces are constructed around this point, so that the final estimates for the unknown model parameters and their uncertainties can be made. For this estimation we use a projection method and a Bayesian approach.

Prescan of Suitable Parameter Combinations

When searching for a candidate for a "good" parameter combination, one is faced with the problem of considering which points in the parameter space and how many of them should be evaluated. Ideally these points are well spread out, and as dense as possible. However, the evaluation of more points comes with additional computational cost.

A fixed grid for the points to be evaluated might be an obvious approach, but if additional points should be added to the scheme their placement is not obvious, because the grid size would change. Alternatively to this, one may want to spread the evaluation points randomly across the parameter space. This would guarantee a uniform distribution of the points, but only if the number of samples is sufficiently high. In cases where the number of evaluation points is rather small, i.e., the parameter space would be sampled in a sparse manner, both approaches fail. Hence, a third approach is chosen in this study—quasi-random sequences.[45] More specifically, Halton sequences[46] are used in this study.

Each element x_k of the sample point $\boldsymbol{x}$ is bound by

$$x_k \in [0, 1] \tag{3}$$

with $k = 1, \ldots, K$ and K being the number of dimensions of the design. The connection between these normalized/coded variables $\boldsymbol{x}$ and the actual physical parameters y is made via an appropriate transformation. To do this, limits for the variables have to be set and shall be denoted by $x_{k,low}$, $x_{k,up}$,

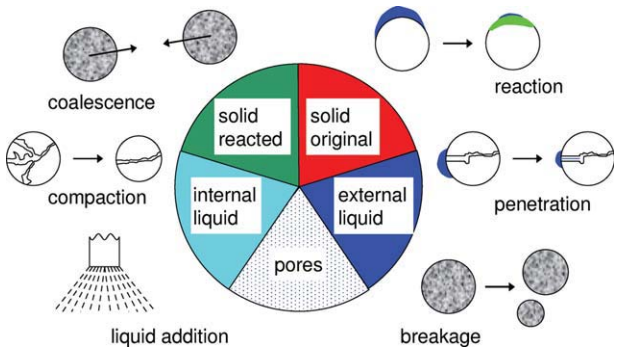

Figure 5. Particle model and transformations in granulation process.

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

$y_{k,\mathrm{low}}$, and $y_{k,\mathrm{up}}$. A logarithmic transformation shall be used for the sampling of the parameter combinations, so that

$$y_k = y_{k,\mathrm{up}} \left(\frac{y_{k,\mathrm{low}}}{y_{k,\mathrm{up}}} \right)^{\frac{x_{k,\mathrm{up}} - x_k}{x_{k,\mathrm{up}} - x_{k,\mathrm{low}}}} . \tag{4}$$

Further details about this transformation as well as for the linear form can be found in the Appendix. For each parameter combination/sample point $\boldsymbol{y}$ the objective function Φ_{LD} is then evaluated,

$$\Phi_{\mathrm{LD}}(\boldsymbol{y}) = \sum_{j=1}^{J} \left(L_{\mathrm{m,geo},3,j}(\boldsymbol{y}) - L^{\mathrm{exp}}_{\mathrm{m,geo},3,j} \right)^2 , \tag{5}$$

with J as the number of scenarios. Among the set of sample points, the best one $\boldsymbol{y}_{\mathrm{best}}$ is chosen according to

$$\boldsymbol{y}_{\mathrm{best}} = \underset{\boldsymbol{y}}{\mathrm{argmin}} \{ \Phi_{\mathrm{LD}}(\boldsymbol{y}) \} . \tag{6}$$

Construction of Local Surrogate Model

The parameter combination $\boldsymbol{y}_{\mathrm{best}}$ serves as initial guess for the estimation of the unknown model parameters and their uncertainties. For this, we need to construct a surrogate model of the complex granulation model around $\boldsymbol{y}_{\mathrm{best}}$. The surrogate model is a second order response surface, an approach that has already been used in previous work.[38]

Choice of boundaries

The parameter combination in the uncoded space, $\boldsymbol{y}_{\mathrm{best}}$ has a corresponding vector in the coded space, $\boldsymbol{x}_{\mathrm{best}}$. We define a distance r_i of the sample point i such as

$$r_i = \sqrt{\sum_{k=1}^{K} \left(x_{\mathrm{best},k} - x_{i,k} \right)^2} , \tag{7}$$

with K being the number of dimensions of the parameter vector $\boldsymbol{x}$. The shortest distance is given by

$$r_{\mathrm{min}} = \min_{i=1,\ldots,N_{\mathrm{SP}}} \{ r_i \} , \tag{8}$$

where N_{SP} is the total number of sample points.

Centred around $\boldsymbol{x}_{\mathrm{best}}$, a hypercube with edge length a is set up. A possible choice for the edge length is $a = r_{\mathrm{min}}$, but there is scope for a less arbitrary choice of the parameter space. For instance, gradient information could be taken into account to choose the region for the surrogate model. Such a procedure would be iterative and shall neither be discussed nor used in this article. However, the edge length a is subject to constraints. Firstly, no parameter combination of previously examined sample points shall be included in the region of the surrogate model, i.e.,

$$a \leq 2 \min_i \left(\max_k \{ |x_{\mathrm{best},k} - x_{i,k}| \} \right) . \tag{9}$$

Secondly, no parameter combination of the surrogate model should lie outside the region of parameter combinations in which the Halton sequence points where sampled,

$$\frac{a}{2} \leq \min_k \{ x_{\mathrm{best},k} - x_{\mathrm{low},k}, x_{\mathrm{up},k} - x_{\mathrm{best},k} \} \quad \forall k = 1, \ldots, K . \tag{10}$$

Choice of points

New coded variables $\tilde{\boldsymbol{x}}$ are introduced for the creation of the response surfaces, so that

$$\tilde{x}_k \in [-1, 1] \, (k = 1, \ldots, K) \quad \text{and} \quad \tilde{x}(x_{\mathrm{best}}) = 0 , \tag{11}$$

where the transition from x to $\tilde{x}$ is a simple linear transformation. We shall denote this hypercube in which $\tilde{\boldsymbol{x}}$ lies in as $\mathscr{X}$. An uncoded variable y is then calculated by

$$y = y_{\mathrm{up}} \left(\frac{y_{\mathrm{low}}}{y_{\mathrm{up}}} \right)^{1 - \frac{a}{2}\tilde{x} - x_{\mathrm{best}}} . \tag{12}$$

In addition to the 2^K corner points, selected points within the bounds of the parameter space shall be evaluated and used for the construction of the response surfaces. These are the centre point $\boldsymbol{x}_{\mathrm{best}}$ ($\tilde{\boldsymbol{x}} = \boldsymbol{0}$) and

$$\begin{aligned} &\tilde{\boldsymbol{x}}_{\mathrm{axis},1} = (0.5, 0, 0, 0, 0), \quad \tilde{\boldsymbol{x}}_{\mathrm{axis},2} = (0, 0.5, 0, 0, 0), \\ &\tilde{\boldsymbol{x}}_{\mathrm{axis},3} = (0, 0, 0.5, 0, 0), \quad \tilde{\boldsymbol{x}}_{\mathrm{axis},4} = (0, 0, 0, 0.5, 0), \\ &\tilde{\boldsymbol{x}}_{\mathrm{axis},5} = (0, 0, 0, 0, 0.5). \end{aligned} \tag{13}$$

The evaluations of the complex granulation model at these points are used to construct the surrogate model—second order response surfaces with the model response $\eta(\tilde{\boldsymbol{x}})$,

$$\eta(\tilde{\boldsymbol{x}}) = \beta_0 + \sum_{k=1}^{K} \beta_k \tilde{x}_k + \sum_{k=1}^{K} \sum_{l \geq k}^{K} \beta_{kl} \tilde{x}_k \tilde{x}_l . \tag{14}$$

The coefficients β_0, β_k, β_{kl} of the response surface are obtained by fitting the surrogate model to the evaluations of the population balance model f_{sim},

$$\eta(\tilde{\boldsymbol{x}}) = f_{\mathrm{sim}}(\tilde{\boldsymbol{x}}) + \epsilon , \tag{15}$$

with ϵ being the approximation error.

Parameter Estimation with the Projection Method

For the estimation of the unknown model parameters $\tilde{\boldsymbol{x}}^*$ and their uncertainties $\boldsymbol{c}^*$ the experimental results and the surrogate model are brought together in the projection method,[37] so that the parameters are obtained through

$$\{ \tilde{\boldsymbol{x}}^*, \boldsymbol{c}^* \} = \underset{\tilde{\boldsymbol{x}}, \boldsymbol{c}}{\mathrm{argmin}} \left(\Phi(\tilde{\boldsymbol{x}}, \boldsymbol{c}) \right) , \tag{16}$$

where Φ is the moment-matching objective function, identical to the one used by Braumann et al.[38]

Parameter Estimation with the Bayesian Approach

In this section, we explain the Bayesian philosophy of modeling uncertainty.[47–49] The approach is not new and was already used for parameter estimation in other fields such as structural mechanics.[50] First consider the following notation:

- x is the unknown parameter
- η^{exp} is a scalar experimental datum.
- $\eta(x)$ is the model response which is dependent upon the unknown parameter x.

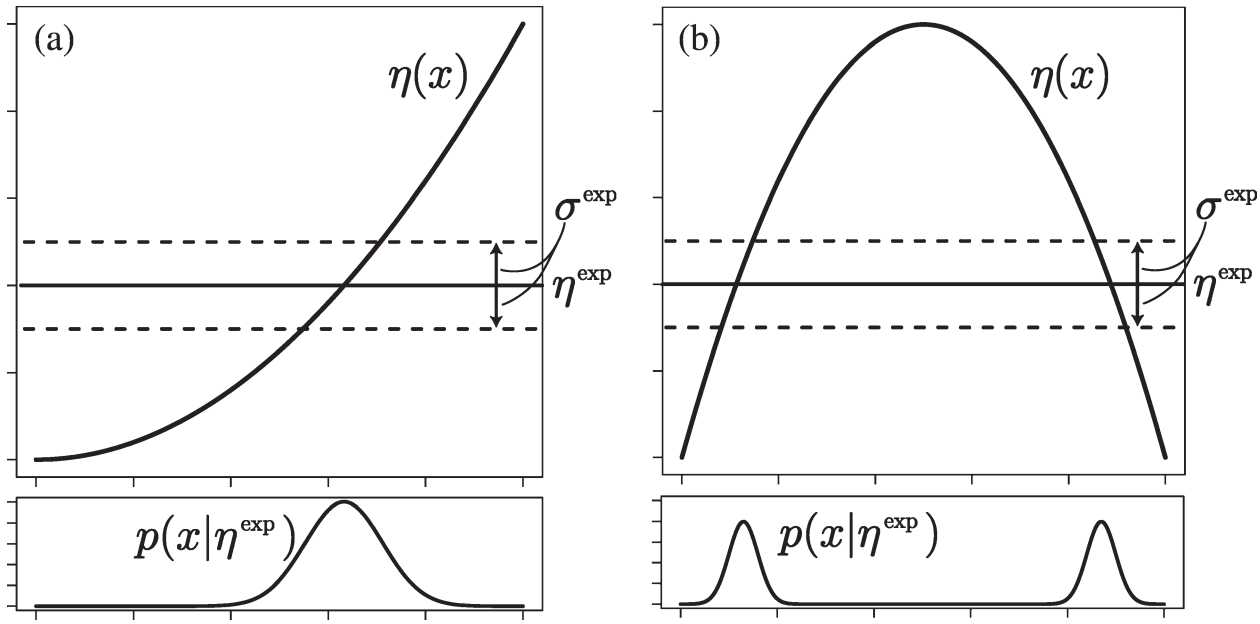

Figure 6. Unimodal and bimodal belief distributions for unknown parameters.

Earlier, it was explained that often one minimises an objective function such as that given in Eq. 5 to choose the unknown x which best fits the data $\eta^{\exp}$, given the model $\eta(x)$. However, there are serious issues with this method. Consider Figure 6.

In Figure 6a, we have scalar data $\eta^{\exp}$, $\eta(x)$ and x. The solution to the least squares problem given by Eq. 5 is solved at the value of x for which the lines of $\eta(x)$ and $\eta^{\exp}$ coincide. The data $\eta^{\exp}$ have uncertainties $\sigma^{\exp}$ attached to them. In current methodologies, it is possible to "map" these uncertainties of the data through the model response $\eta(x)$ to give some indication of uncertainties in the parameter x due to the uncertainty in $\eta^{\exp}$. However, in that example, it so happened that there was a unique solution to the minimization of the objective function. However, it is easy to see that if there is a lack of identifiability of the parameters given the data and model response, then there may be multiple solutions to the minimization problem, or at least many parameter values which give similar values to the objective function, and thus the optimal solution could be extremely sensitive to slight changes in the data values. An example of the lack of identifiability of the parameter value x is given in Figure 6b. Since the model function is symmetric, there are exactly two solutions to the minimization problem—neither should be favored over the other. One could argue that we simply report both values of x. But what if the model function were slightly asymmetric? One would expect in this situation that the optimal values are slightly unequal in "weight" in this situation due to the uncertainty in the data. But how to weight the solutions? Man et al.[51] make an attempt at this by generalising the assumed parameter distribution from a Gaussian distribution to a "multimodal" Gaussian. Meanwhile, in both the examples above, it was assumed that the uncertainties in the data are exactly known. What if they are unknown? What if we have some information about the uncertainties, but not exact information about their values? What if something is already known about x from previous experiments—the question would be how to incorporate this existing knowledge, rather than starting from scratch.

To answer all these questions, we appeal to the Bayesian philosophy of statistics. In the first place, the data are modelled as "noisy," and thus attached with some uncertainty. Typically, it is assumed that the data are realizations from a Gaussian distribution with the mean being the unknown "true" data. Arguably, this assumed randomness is not true randomness, but purely our way of modeling uncertainty in the data, given that we are unable to control every aspect of our experimental procedures. If this is the case, one could argue that uncertainties of any kind can be modelled in a similar way. This is precisely the Bayesian view. Any unknown quantity is assumed to be random (in exactly the same way the data are assumed to be random), where the probabilities of the quantity taking certain values models our belief in the quantity taking those values. Thus, one sees that the uncertainties in the data and the uncertainties in unknown quantities are put on an equal footing—it is then easy to use the highly developed mathematical framework of probability theory to make inference about our unknown quantities. Thus all the information about our belief about the unknown quantities are encoded in the belief distributions, including the means, variances and even correlations, etc. In other words, the notion of optimization is deemed unnecessary since it gives little information. However, it is also possible, at the point of a gun, to provide an optimal estimate for the unknown quantities. We shall attempt to answer some of the questions made above in the following subsections.

Bayesian preliminaries

Given that we are modeling uncertainties via randomness, we must attach belief probability distributions to each unknown quantity. This necessitates that we have an

initial belief distribution—we call this the prior distribution. Thus, any a priori information known about our unknown quantities gained from previous experiments or experts can in principle be incorporated into our data analysis, rather than wasting this useful information. In keeping with our notation, the unknown quantities denoted by (the possibly vectorial) x. The probability density function corresponding to the prior distribution shall be denoted by $p(x)$.

Given we have this prior belief in x, how can we use the experimental data to update our belief distribution of x? As stated earlier, we assume that the data are noisy, and thus "random," and so we must state the distribution of the data. This is given by the probability density function $p(\eta^{\exp}|x)$, which is typically a Gaussian density with the mean being the true data. Note that this distribution depends on the unknown parameter x. The examples above show this—it was assumed that the true mean of the data is given by the model response $\eta(x)$ which depends on x.

Armed with this information, it is now possible to use probability theory to form our new belief distribution for the parameter x—we call this the posterior distribution whose density is denoted by $p(x|\eta^{\exp})$. To find this, simply apply the famous Bayes' Theorem:

$$p(x|\eta^{\exp}) = \frac{p(\eta^{\exp}|x)\,p(x)}{\int p(\eta^{\exp}|x')\,p(x')\,\mathrm{d}x'} \propto p(\eta^{\exp}|x)\,p(x)\,, \tag{17}$$

noting that we have removed the denominator as a constant of proportionality since it is independent of x. Now that the posterior $p(x|\eta^{\exp})$ has been obtained, it is possible to extract modes, means, variances, etc., thus providing the user complete information about x. Given the posterior distribution of x, we may sensibly ask what our "best" estimate for the parameters are. This question is intrinsically at odds with the notion of denoting belief by whole distributions. However, we can appeal to decision theory, which is the mathematical theory of optimal decisions under uncertainty. We are still uncertain about our parameters, even after obtaining data. To find an optimal choice of parameters (i.e., best decision), we need to define our notion of "loss." As a default setting, we may wish to choose the parameter value x^* which minimises the averaged squared distance between our parameter choice and the true parameter value, where the averaging is performed using our posterior distribution for x. For this notion of loss, our "best" parameter value is the posterior mean of x.[49]

Application to granulation process

In the granulation example, we are given the following information:

- $\widetilde{\boldsymbol{x}}$ are the K-dimensional unknown (coded) parameters of interest.
- $\eta_j^{\exp} := L^{\exp}_{\mathrm{m,geo},3,j}$ for $j = 1, \ldots, J$ are the experimental data.
- $(\sigma_j^{\exp})^2$ are the variances (known or unknown) of the noise of the data $\eta_j^{\exp}$.
- $\eta_j(\widetilde{\boldsymbol{x}})$ is the response from the surrogate model (response surfaces) for the j^{th} scenario.

Likelihood. As stated in the Bayesian preliminaries section, we first need the data distribution, which can be taken to be Gaussian with mean being our model response. Thus, we have that

$$\eta_j^{\exp} \sim \mathcal{N}\left(\eta_j(\widetilde{\boldsymbol{x}}), (\sigma_j^{\exp})^2\right) \quad \text{independently } \forall j = 1, \ldots, J\,, \tag{18}$$

which provides us with the following probability density for the data (given the unknown parameter $\widetilde{\boldsymbol{x}}$ and the possibly unknown $(\sigma_j^{\exp})^2$ values):

$$\begin{aligned} p(\eta^{\exp}|\widetilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2) &= \prod_{j=1}^{J} \left\{ (2\pi(\sigma_j^{\exp})^2)^{-\frac{1}{2}} \right. \\ &\quad \left. \times \exp\left[-\frac{1}{2(\sigma_j^{\exp})^2} \left(\eta_j^{\exp} - \eta_j(\widetilde{\boldsymbol{x}})\right)^2 \right] \right\} = (2\pi)^{-\frac{J}{2}} \\ &\quad \times \left(\prod_{j=1}^{J} (\sigma_j^{\exp})^2 \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{j=1}^{J} \frac{1}{(\sigma_j^{\exp})^2} \left(\eta_j^{\exp} - \eta_j(\widetilde{\boldsymbol{x}})\right)^2 \right]. \end{aligned} \tag{19}$$

Prior distribution—known $(\sigma_j^{exp})^2$. The next component required for a Bayesian analysis of the unknown values are the prior distributions. This is a non trivial choice, and one must be careful how this is chosen as it can affect the posterior distribution greatly. However, for our granulation problem, it may be reasonable to consider a uniform distribution over our hypercube $\mathscr{X}$ (recalling that this is the region in K-dimensional space such that $\tilde{x}_k \in [-1, 1]$ for all $k = 1,\ldots,K$). In the first instance, we shall deal with the case that the $(\sigma_j^{\exp})^2$ are known a priori. Thus, we have that the prior probability density for $\widetilde{\boldsymbol{x}}$ is

$$p(\widetilde{\boldsymbol{x}}) = \frac{1}{|\mathscr{X}|} \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \tag{20}$$

where $|.|$ denotes the size/volume of a set and $\mathbb{1}_{\{\text{condition}\}}$ is the indicator function (i.e., it is equal to unity if "condition" is true, and zero otherwise).

Posterior distribution—known $(\sigma_j^{exp})^2$. Using Eq. 17, the posterior density is now easy to compute (up to a constant factor) as follows:

$$\begin{aligned} p(\widetilde{\boldsymbol{x}}|\eta^{\exp}) &\propto p(\eta^{\exp}|\widetilde{\boldsymbol{x}})\,p(\widetilde{\boldsymbol{x}}) \quad \text{by Eq. 17} \\ &\propto (2\pi)^{-\frac{J}{2}} \left(\prod_{j=1}^{J} (\sigma_j^{\exp})^2 \right)^{-\frac{1}{2}} \\ &\quad \times \exp\left[-\frac{1}{2} \sum_{j=1}^{J} \frac{1}{(\sigma_j^{\exp})^2} \left(\eta_j^{\exp} - \eta_j(\widetilde{\boldsymbol{x}})\right)^2 \right] \cdot \frac{1}{|\mathscr{X}|} \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \\ &\propto \exp\left[-\frac{1}{2} \sum_{j=1}^{J} \frac{1}{(\sigma_j^{\exp})^2} \left(\eta_j^{\exp} - \eta_j(\widetilde{\boldsymbol{x}})\right)^2 \right] \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \end{aligned} \tag{21}$$

Prior distribution—unknown $(\sigma_j^{exp})^2$. We make the assum-ption that our uncertainty in $\widetilde{\boldsymbol{x}}$ is independent from our uncertainty in the $(\sigma_j^{\exp})^2$, which are also assumed to be mutually independent from each other. This gives us the following expression for the prior distribution:

$$p(\tilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2) = p(\tilde{\boldsymbol{x}}) \prod_{j=1}^{J} p((\sigma_j^{\exp})^2) \tag{22}$$

where the $p((\sigma_j^{\exp})^2)$ are our prior densities for the $(\sigma_j^{\exp})^2$. In the case of having *no* prior information about these, we use the well-used Inverse Gamma non informative prior,[52] given by

$$p((\sigma_j^{\exp})^2) = \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \left(\frac{1}{(\sigma_j^{\exp})^2}\right)^{\alpha_j+1} \exp\left[-\frac{\beta_j}{(\sigma_j^{\exp})^2}\right], \tag{23}$$

where α_j and β_j are usually taken be small but positive values such as 0.001. Prior specification remains very much an open question.

Posterior distribution—unknown $(\boldsymbol{\sigma}_j^{exp})^2$. The posterior density (up to a constant factor) is computed as follows:

$$\begin{aligned}
&p(\tilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2 \mid \eta^{\exp}) \\
&\quad \propto p(\eta^{\exp} \mid \tilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2)\, p(\tilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2) \\
&\quad \propto p(\eta^{\exp} \mid \tilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2)\, p(\tilde{\boldsymbol{x}}) p((\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2) \\
&\quad \propto (2\pi)^{-\frac{J}{2}} \left(\prod_{j=1}^{J} (\sigma_j^{\exp})^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{J} \frac{1}{(\sigma_j^{\exp})^2}\left(\eta_j^{\exp} - \eta_j(\tilde{\boldsymbol{x}})\right)^2\right] \\
&\qquad \times \frac{1}{|\mathscr{X}|} \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \cdot \prod_{j=1}^{J} \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \left(\frac{1}{(\sigma_j^{\exp})^2}\right)^{\alpha_j+1} \exp\left[-\frac{\beta_j}{(\sigma_j^{\exp})^2}\right] \\
&\quad \propto \exp\left[-\sum_{j=1}^{J} \frac{1}{(\sigma_j^{\exp})^2}\left\{\beta_j + \frac{1}{2}\left(\eta_j^{\exp} - \eta_j(\tilde{\boldsymbol{x}})\right)^2\right\}\right] \\
&\qquad \times \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \cdot \prod_{j=1}^{J} \left(\frac{1}{(\sigma_j^{\exp})^2}\right)^{\alpha_j+\frac{3}{2}}.
\end{aligned} \tag{24}$$

Furthermore, the $(\sigma_j^{\exp})^2$ are nuisance parameters—they are not of primary interest to the user. Thus we can consider the marginal posterior for $\tilde{\boldsymbol{x}}$, which is computed as follows:

$$\begin{aligned}
p(\tilde{\boldsymbol{x}} \mid \eta^{\exp}) &:= \int_0^{\infty} \cdots \int_0^{\infty} p(\tilde{\boldsymbol{x}}, (\sigma_1^{\exp})^2, \ldots, (\sigma_J^{\exp})^2 \mid \eta^{\exp})\, \mathrm{d}(\sigma_1^{\exp})^2\, \mathrm{d}(\sigma_2^{\exp})^2 \ldots \mathrm{d}(\sigma_J^{\exp})^2 \\
&\propto \int_0^{\infty} \cdots \int_0^{\infty} \exp\left[-\sum_{j=1}^{J} \frac{1}{(\sigma_j^{\exp})^2}\left\{\beta_j + \frac{1}{2}\left(\eta_j^{\exp} - \eta_j(\tilde{\boldsymbol{x}})\right)^2\right\}\right] \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \cdot \prod_{j=1}^{J} \left(\frac{1}{(\sigma_j^{\exp})^2}\right)^{\alpha_j+\frac{3}{2}} \mathrm{d}(\sigma_1^{\exp})^2 \mathrm{d}(\sigma_2^{\exp})^2 \ldots \mathrm{d}(\sigma_J^{\exp})^2 \\
&\propto \prod_{j=1}^{J} \frac{\Gamma(\alpha_j)}{\left[\beta_j + \frac{1}{2}\left(\eta_j^{\exp} - \eta_j(\tilde{\boldsymbol{x}})\right)^2\right]^{\alpha_j+\frac{1}{2}}} \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}} \propto \frac{1}{\prod_{j=1}^{J} \left[\beta_j + \frac{1}{2}\left(\eta_j^{\exp} - \eta_j(\tilde{\boldsymbol{x}})\right)^2\right]^{\alpha_j+\frac{1}{2}}} \mathbb{1}_{\{\tilde{\boldsymbol{x}} \in \mathscr{X}\}}.
\end{aligned} \tag{25}$$

Next step. What do we do with these expressions for the posteriors? In the first instance, notice that we only have the expressions for the posteriors up to a constant positive factor—these factors can in principle be computed (numerically or perhaps analytically) using the fact that the posterior densities are probability density functions which must integrate to unity. However, for higher dimensional $\tilde{\boldsymbol{x}}$, the normalising factor may be hard to compute, and furthermore, it is hard to visualise the scalar posterior function as a function of high-dimensional space. Thus we opt to sample from the posterior distribution. From the sample, it is possible to estimate any number of quantities of interest such as the means, variances or even correlations between unknown quantities. Furthermore, visualization of posterior is made easier since we can easily estimate* all possible joint posterior densities of pairs of unknown quantities. In the above granulation examples, this translates to estimating $p(\tilde{x}_{k_1}, \tilde{x}_{k_2} \mid \eta^{\exp})$ for all $k_1, k_2 \in \{1, \ldots, K\}$.

Sampling from the density. Remembering that we (only) have access to an analytic expression for the posterior density but only up to a constant positive normalization factor, we seek an algorithm which can sample from this density given this information. The most famous algorithm for this task is the Metropolis-Hastings algorithm.[53,54] Although this algorithm shall be used in this study, it should be noted that more sophisticated (and complicated) approaches have been developed over the recent years, e.g., Cheung and Beck.[55] The Metropolis-Hastings algorithm works by forming a continuous space discrete time Markov Chain which has a stationary distribution which is identical to the distribution that we want to sample from—call the density $\pi(x)$. In simple terms, this means sequentially choosing values of x (these values of x are called *states* in Markov Chain language) to move in such a way as to visit the state x with frequency density $\pi(x)$. Skipping all the details of why it works, we give the algorithm at its simplest and most generic form in Algorithm 1. Notice that Eq. 27 has the factor $\pi(x')/\pi(x)$ implying that we need only the density $\pi(x)$ up to a constant normalization factor. Hence, if we wish to sample from the posteriors given in Eqs. 21 or 25, we need only substitute $\pi(x)$ for those expressions.

In the case of Eq. 25, Algorithm 1 generates the following sequence: $(\tilde{\boldsymbol{x}}^{(t)})_{t=1,\ldots,T}$. Since this is a (albeit correlated) sample from the marginal posterior $p(\tilde{\boldsymbol{x}} \mid \eta^{\exp})$, we can estimate any quantity of interest. For example, the posterior mean of $\tilde{x}_k$ is estimated by:

$$\mathbb{E}[\tilde{x}_k \mid \eta^{\exp}] \approx \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_k^{(t)} \quad . \tag{26}$$

Furthermore, estimation of the joint marginal posterior density $p(\tilde{x}_{k_1}, \tilde{x}_{k_2} \mid \eta^{\exp})$ can be estimated using kernel density

*Using kernel density estimation.

Table 2. Limits of Coded and Uncoded Variables

Transformation	Parameter		Limit	Coded	Uncoded	Unit
	Uncoded	Coded				
Coalescence	$\hat{K}_{coag}$	x_1	Low	0	1.0×10^{-11}	m^3
			Up	1	1.0×10^{-8}	m^3
Compaction	k_{porred}	x_2	Low	0	0.01	s/m
			Up	1	0.6	s/m
Breakage	$\hat{k}_{att}$	x_3	Low	0	1.0×10^{9}	s/m^5
			Up	1	1.0×10^{12}	s/m^5
Penetration	$\hat{k}_{pen}$	x_4	Low	0	1.0×10^{8}	$kg^{1/2}s^{-3/2}m^{-7/2}$
			Up	1	1.0×10^{12}	$kg^{1/2}s^{-3/2}m^{-7/2}$
Reaction	k_{reac}	x_5	Low	0	1.0×10^{-10}	m/s
			Up	1	1.0×10^{-6}	m/s

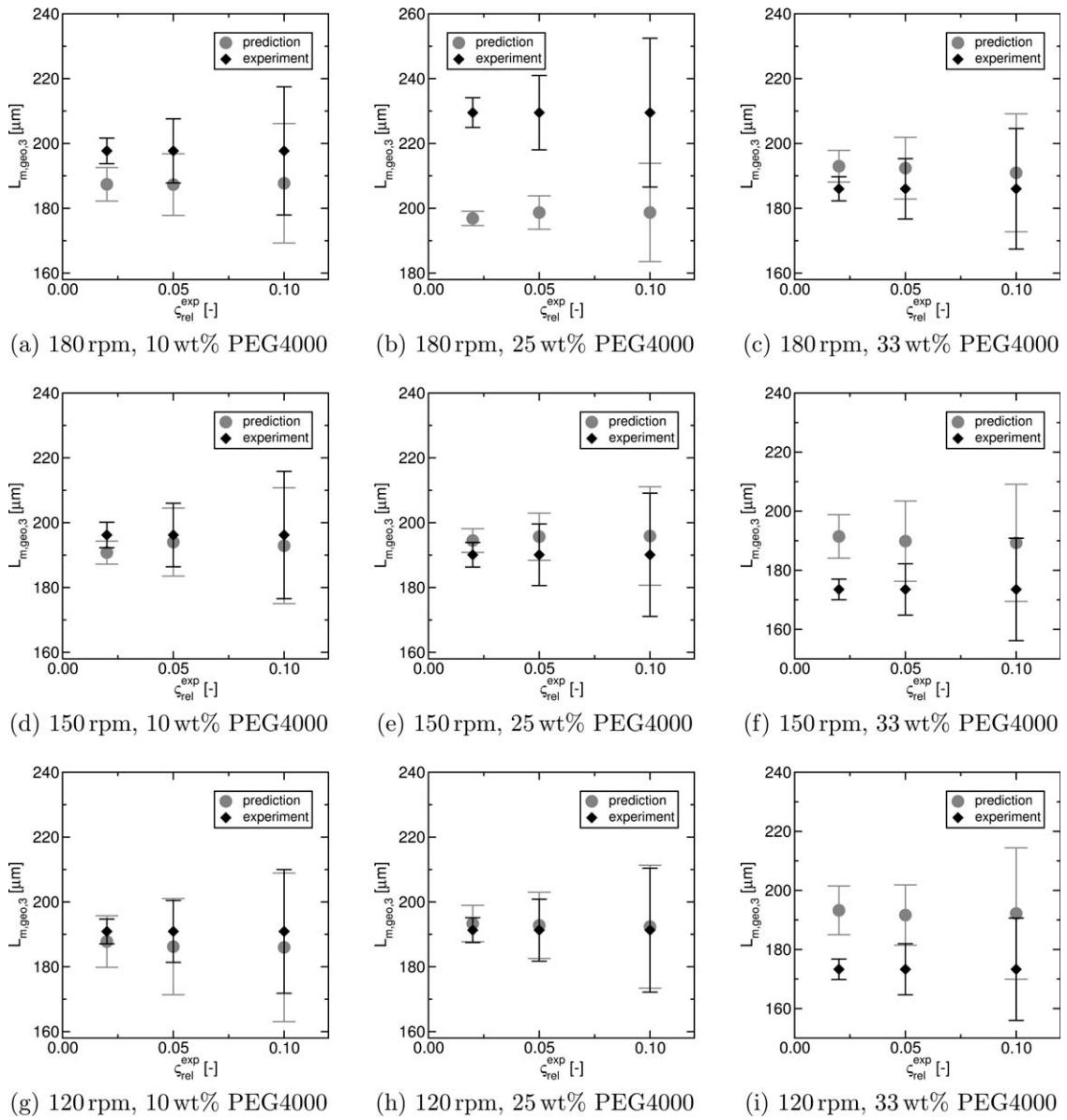

Figure 7. Experiments vs. predictions with parameters from projection method for different impeller speed and binder composition.

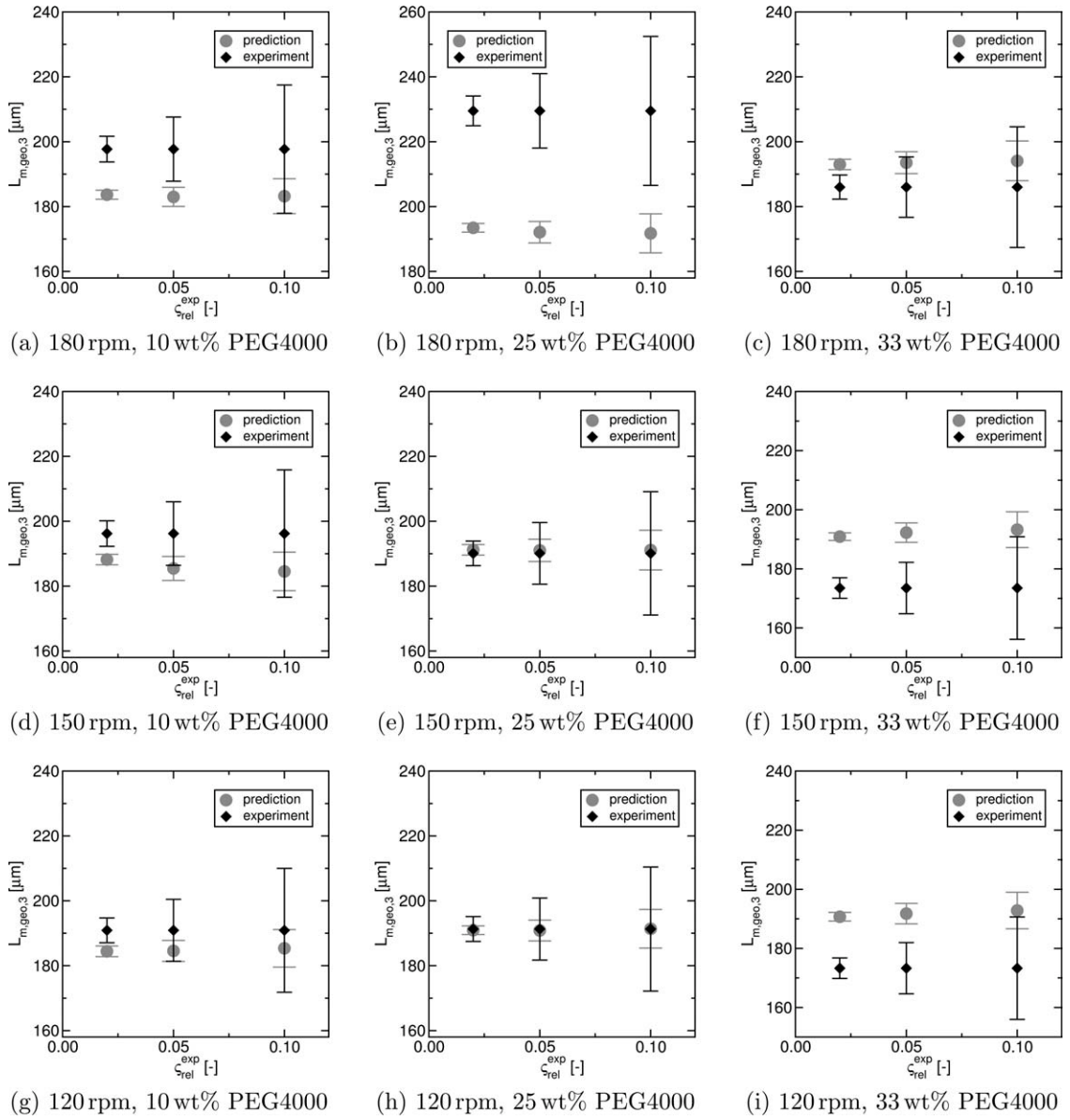

Figure 8. Experiments vs. predictions with parameters from Bayesian approach for different impeller speed and binder composition.

estimation on the restricted sample $((\tilde{x}_{k_1}^{(t)}, \tilde{x}_{k_2}^{(t)})^\top)_{t=1,\ldots,T}$, i.e., the sample $(\tilde{\boldsymbol{x}}^{(t)})_{t=1,\ldots,T}$, but ignoring all components of $\tilde{\boldsymbol{x}}^{(t)}$ except the components k_1 and k_2.

We have just discussed how to sample from some arbitrary density $\pi(x)$ for some arbitrary states x. To apply Algorithm 1 to the granulation example, we need only set the states x to be $\tilde{\boldsymbol{x}}$ and the density of interest to be the posterior density $\boldsymbol{p}(\tilde{\boldsymbol{x}} \,|\, \eta^{\exp})$. In the results in the next sections, we use the proposal density $q(x \to x')$ (as explained in Step 1 of Algorithm 1) to be $\boldsymbol{q}(\boldsymbol{x} \to \boldsymbol{x}') = \boldsymbol{q}(\tilde{\boldsymbol{x}} \to \tilde{\boldsymbol{x}}') = \mathbf{0.2}\, \frac{\mathbf{1}}{|\mathscr{X}|}\, \mathbb{1}_{\{\tilde{\boldsymbol{x}}' \in \mathscr{X}\}} + \mathbf{0.8}\, \frac{\mathbf{1}}{|\mathscr{X}(\tilde{\boldsymbol{x}})|}\, \mathbb{1}_{\{\tilde{\boldsymbol{x}}' \in \mathscr{X}(\tilde{\boldsymbol{x}})\}}$, where $\mathscr{X}(\tilde{\boldsymbol{x}})$ is the intersection of the hypercube $\mathscr{X}$ and the hypercube with edgelength 0.1 centred at $\tilde{\boldsymbol{x}}$. The idea behind this choice was to jump (with small probability 0.2) completely uniformly in $\mathscr{X}$ so that long jumps are made, but jumping far can often result in a rejection, so to ensure that not too many rejections are made, we jump (with large probability 0.8) somewhere nearby to the current position in $\mathscr{X}$.

Algorithm 1: Metropolis-Hastings algorithm

1 Choose a proposal density $q(x \to x')$ which is a probability density function for choosing a new state x' given that we are currently positioned in the state x.
2 Set $t = 0$. Start with any initial state $x^{(0)}$ for x.
while $t < T$ do
3 Propose a new state x' sampled from (any) *proposal density* $q(x \to x^{(t)})$.
4 Compute the quantity

$$\alpha_{\text{accept}} := \frac{\pi(x')\, q(x' \to x^{(t)})}{\pi(x^{(t)})\, q(x^{(t)} \to x')} \tag{27}$$

5 Perform a rejection step, i.e., with probability $r := \min\{1, \alpha_{\text{accept}}\}$, *accept* the proposed state x', i.e., set $x^{(t+1)} = x'$, otherwise, set $x^{(t+1)} = x^{(t)}$.
6 Set $t \leftarrow t + 1$.
7 STOP.

Results and Discussion

The theory outlined in the previous section is now applied to the granulation process.

Prescan

200 parameter combinations based on Halton sequences and the boundaries given in Table 2 were evaluated.

Through evaluation of the objective function (5) the parameter combination of sample point 69 was identified to be most applicable to the studied problem,

$$\boldsymbol{x}_{\text{best}} = \boldsymbol{x}_{\text{SP}\,69} = (0.633, 0.284, 0.936, 0.901, 0.322)\,, \tag{28}$$

which corresponds to the parameter combination $\boldsymbol{y}_{\text{best}}$,

$$\boldsymbol{y}_{\text{best}} = \begin{pmatrix} 7.925 \cdot 10^{-10}\,\text{m}^3 \\ 0.0320\,\text{s/m} \\ 6.427 \cdot 10^{11}\,\text{s/m}^5 \\ 4.018 \cdot 10^{11}\,\text{kg}^{1/2}\,\text{s}^{-3/2}\,\text{m}^{-7/2} \\ 1.941 \cdot 10^{-9}\,\text{m/s} \end{pmatrix}. \tag{29}$$

Surrogate model construction

Response surfaces around sample point 69 were constructed. The corner points of the hypercubes and additional points for these experimental designs and their corresponding parameter values were chosen according to Eqs. 11–13.

The distance between the sample point $\boldsymbol{x}_{\text{best}}$ and the nearest other sample point is

$$r_{\min} = 0.3540\,. \tag{30}$$

The edge length of the hypercube a may be chosen as $a = r_{\min}$. However, this is not possible in the current case as the sample point 69 is close to the boundaries of the design, so that a is limited by constraint (10), leading to

$$a = 0.1280\,. \tag{31}$$

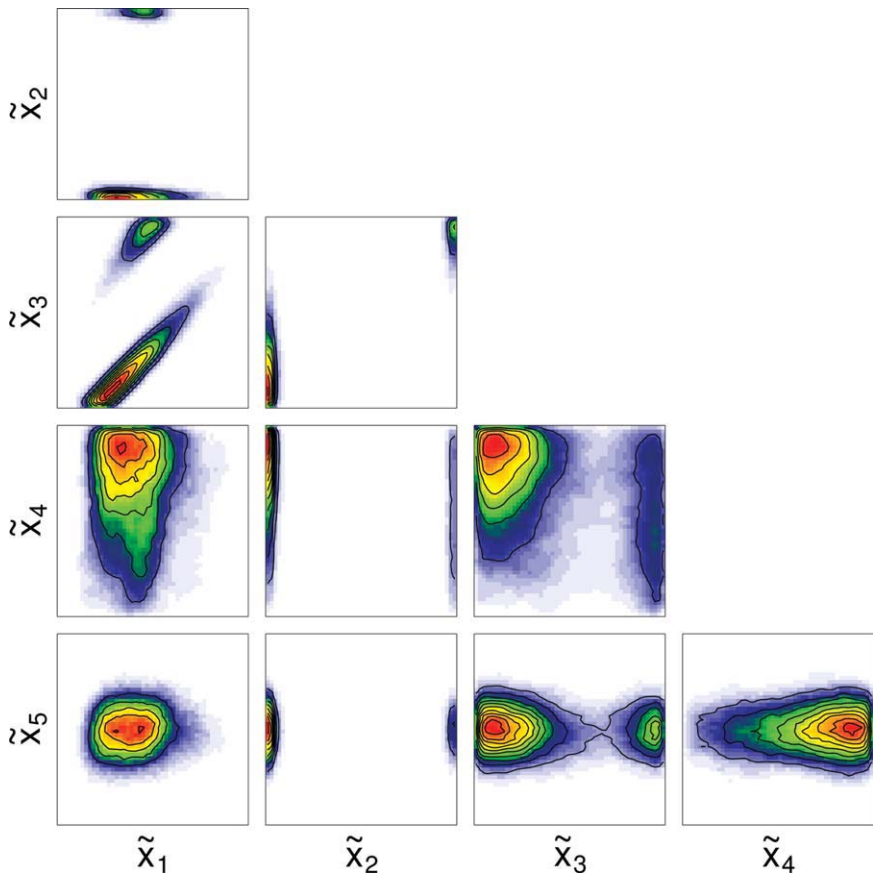

Figure 9. Estimates of joint marginal posterior densities $p(\tilde{x}_{k_1}, \tilde{x}_{k_2} \mid \eta^{\text{exp}})$—known $(\sigma_j^{\text{exp}})^2$ values, set with relative errors $\varsigma_{\text{rel}}^{\text{exp}} = 0.02$: Pairs of parameters.

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

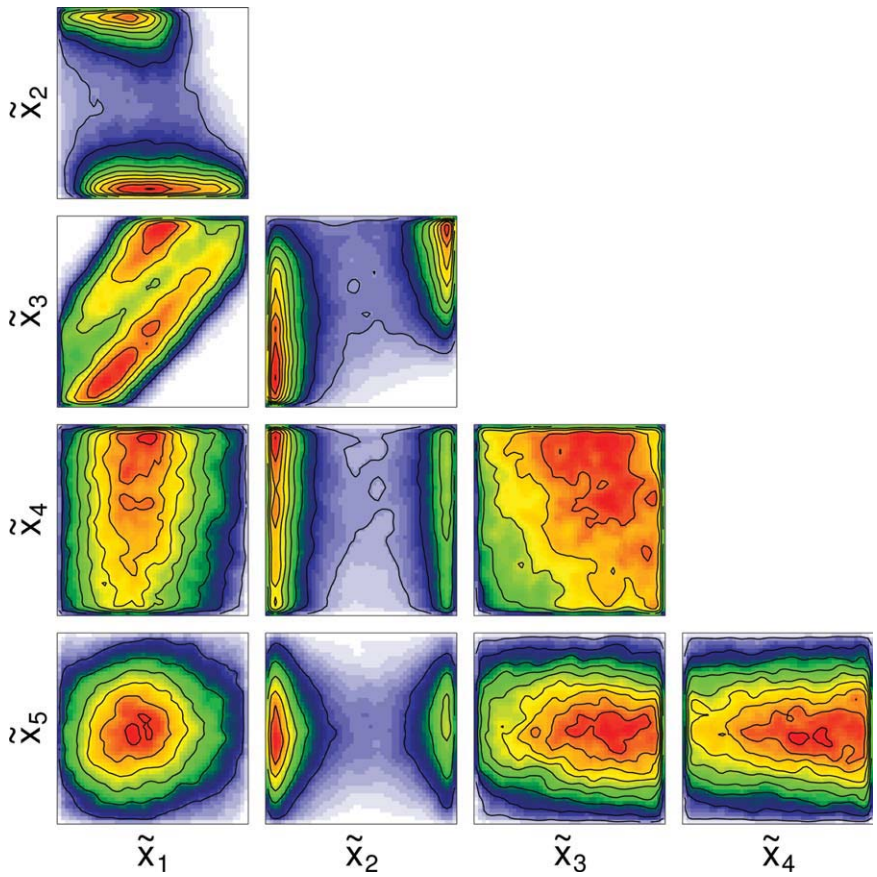

Figure 10. Estimates of joint marginal posterior densities $p(\tilde{x}_{k_1}, \tilde{x}_{k_2} \mid \eta^{exp})$—known $(\sigma_j^{exp})^2$ values, set with relative errors $\varsigma_{rel}^{exp} = 0.05$: Pairs of parameters.

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

Parameter estimation results

In this section we estimate the unknown parameters and their uncertainties using the projection method and the Bayesian approach. For the former one we apply Eq. 16 along with the moment-matching objective function. To do so, the uncertainty in the experimental results has to be known. Due to a lack of repetitions, we assume the experimental uncertainties to be,

$$\sigma_j^{exp} = \varsigma_{rel}^{exp} L_{m,geo,3,j}^{exp} \qquad (j \in [1, \ldots, J]) . \tag{32}$$

The unknown parameters are estimated using a relative experimental uncertainty of $\varsigma_{rel}^{exp} = 0.02, 0.05, 0.10$. Model predictions from the surrogate model with these parameter estimates are computed and compared to the experimental results for the different operating conditions in Figure 7.

Although all nine scenarios (operating conditions) are used for the parameter estimation without any weighting between the scenarios, differences in the agreement between predictions and experimental results for the different scenarios can be observed. It appears that better agreement with the estimated parameter set is achieved for lower impeller speeds and less PEG in the binder. Irrespective of this finding, it is noticeable that the predictions for the geometric mean sizes are very similar for estimation of different experimental uncertainties.

If the Bayesian approach is applied to estimate the unknown rate constants of the granulation model, predictions after parameter estimation are computed and compared to the experimental results too (Figure 8).

Figure 8 was generated in the following way – we take the posterior sample generated using the Metropolis-Hastings algorithm, and for each of the samples, we evaluate the model function, which gives us a posterior sample of model predictions. The sample mean and standard deviations are computed from this sample. The agreement between the predictions and the experimental results looks similar to those for the projection method. However, the uncertainties in the predictions with the parameters obtained through the Bayesian approach are smaller than those for the projection method. Although this might be surprising in the first instance, it can be explained by the fact that we try to match the experimental uncertainty and the model uncertainty in the moment-matching objective function used for the parameter estimation in the projection function. We also note the apparent discrepancy between model prediction and experimental data for both methodologies. It is possible that this discrepancy arises due to the mathematical model being a not so good predictor (for certain process conditions) of the data. In this case, the user of either methodology should consider obtaining more data to reduce uncertainties in the model prediction, and also possibly modifying the model

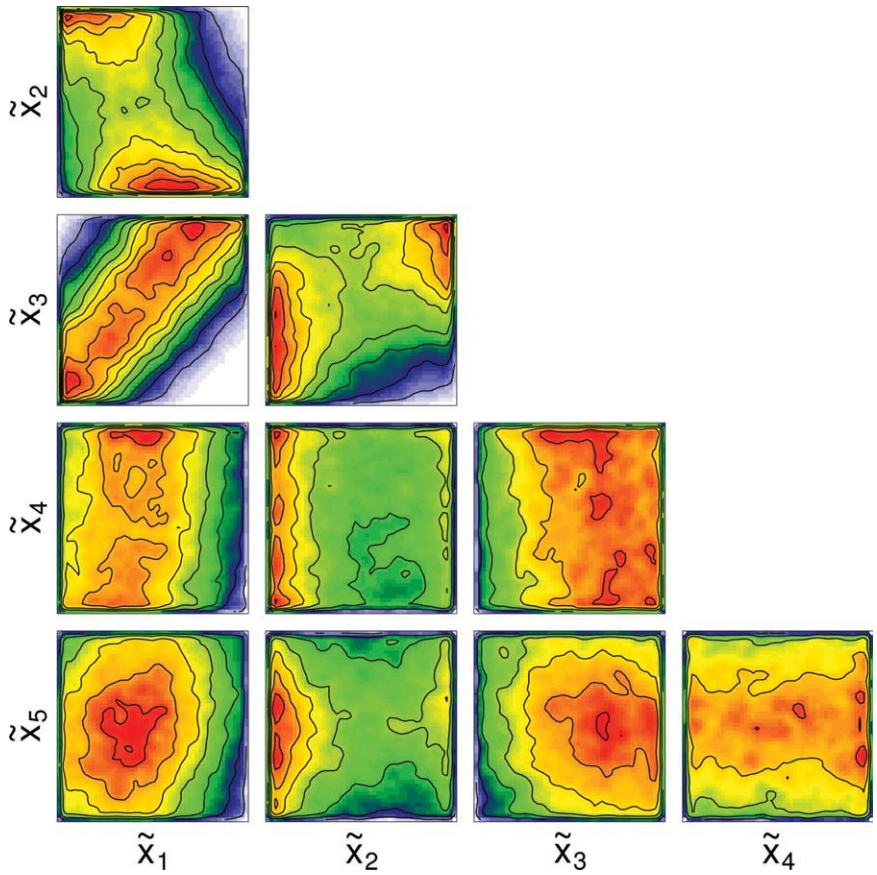

Figure 11. Estimates of joint marginal posterior densities $p(\tilde{x}_{k_1}, \tilde{x}_{k_2} \mid \eta^{exp})$—known $(\sigma_j^{exp})^2$ values, set with relative errors ς_{rel}^{exp} = 0.10: Pairs of parameters.

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

altogether if the model prediction uncertainties are already small relative to the apparent discrepancy.

From the parameter distribution obtained through the Bayesian approach, it is possible to deduce two-dimensional (marginal) distributions of any combination of parameters. These distributions are plotted in Figures 9–11 for relative experimental uncertainties of $\varsigma_{rel}^{exp} = 0.02, 0.05, 0.10$.

Firstly it can be observed that the distributions become more spread out as the experimental uncertainty increases (as one would naturally expect). Secondly, most distributions are unimodal with respect to one parameter, except the distributions in $\tilde{x}_2$, the compaction rate constant, for which a bimodal behavior can be observed. This means, essentially two solutions can be obtained for the inverse problem. A very simplistic physical explanation for this behavior could be the following—if the system is in a state with a low compaction rate (constant), internal liquid (inside the pores) is less likely to be squeezed out onto the external surface. Hence, not so much binder is present on the external surface, and coalescence between particles is not so likely. In contrast, a higher compaction rate (constant) will lead to an increased squeezing of internal liquid out onto the external surface of the particle, thereby increasing the binder amount that will promote coalescence, i.e., particle growth. However, the transfer of the binder from the interior of the granules to the external surface increases the likelihood of particle breakage, which counteracts the increased coalescence. These bimodal distributions indicate a lack of identifiability for the given model and the experimental data. In addition to bimodal distributions, a positive correlation between the collision rate constant $\tilde{x}_1$ and the breakage rate constant $\tilde{x}_3$ can be spotted. Such behavior can be somewhat expected as coalescence and breakage are antagonistic processes. However, the frequency of these events happening also determines how the liquid (binder) is spread within the particle ensemble. Interestingly, such a correlation between the parameters can also be observed in the marginal density when the parameters are estimated without any experimental uncertainties known (Figure 12).

From the marginal densities for all four settings, it is apparent that parameters such as the reaction rate constant $\tilde{x}_5$ are neither multimodal nor correlated with other parameters. Hence, it is likely that unique estimates for its values can be found, but it has to be kept in mind that any marginalization is an averaging operation that may remove higher-order correlations, etc.

To allow further comparison of the results obtained through the projection method and the Bayesian approach, the mean and the uncertainty for each single parameter $\tilde{x}_i$ are summarized in Table 3.

Looking at the results from the projection method first (Table 3), it is clear that the parameter estimates $\tilde{x}_i^*$ are sensitive to the experimental uncertainties, in particular for $\tilde{x}_1^*$ and

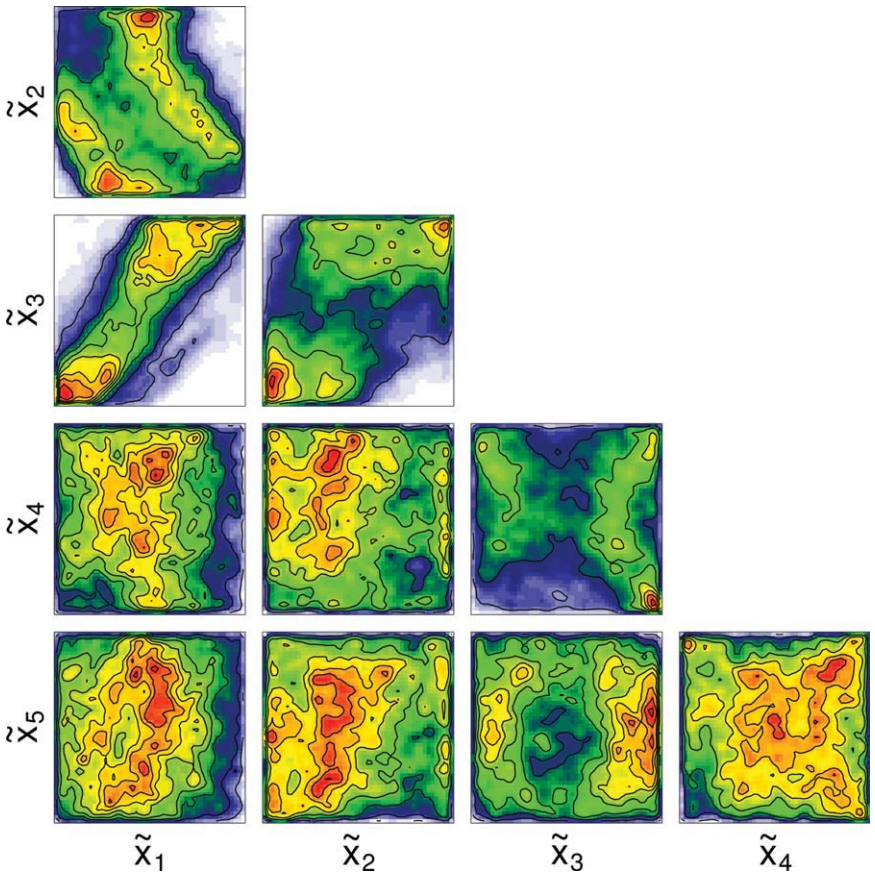

Figure 12. Estimates of joint marginal posterior densities $p(\tilde{x}_{k_1}, \tilde{x}_{k_2} \mid \eta^{exp})$—unknown $(\sigma_j^{exp})^2$ values, set with $\alpha_j = \beta_j = 0.001$: Pairs of parameters.

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

$\tilde{x}_3^*$. The uncertainties in the parameter estimates vary with the experimental uncertainty too, becoming bigger as the experimental uncertainty increases (as one would expect). With respect to the results from the Bayesian approach, we notice that the estimates for all but the experimental uncertainty of $\varsigma_{rel}^{exp} = 0.02$ are not too dissimilar (Table 3). An increase in parameter uncertainty with increasing experimental uncertainty can actually only be made out for the coalescence rate constant $\tilde{x}_1$ and the reaction rate constant $\tilde{x}_5$. However, the marginal densities in Figures 9–11 clearly show an increase in the uncertainties for all parameters. A comparison between the values for the projection method and the Bayesian approach reveals distinct differences. This clearly shows that there is a lack of identifiability for the current system. However, the differences between the estimates are not surprising. Take for instance the bimodal distribution in the compaction rate constant $\tilde{x}_2$ predicted by the Bayesian approach. Due to the fact that the projection method assumes a unimodal parameter distribution, it favors one peak of the bimodal distribution. In contrast, the mean for that parameter when sampled from the bimodal distribution is distinctively different from the projection method result.

Conclusions

A systematic approach of parameter estimation for a multidimensional population balance model for granulation using

Table 3. Estimated Parameters (Coded Form) for Different Experimental Uncertainties ς_{rel}^{exp}

ς_{rel}^{exp}	$\tilde{x}_1^*$[-]	$\tilde{x}_2^*$[-]	$\tilde{x}_3^*$[-]	$\tilde{x}_4^*$[-]	$\tilde{x}_5^*$[-]	c_1^* [-]	c_2^* [-]	c_3^* [-]	c_4^* [-]	c_5^* [-]
Projection method										
0.02	−0.3647	−1.0000	−0.9398	1.0000	0.0044	0.0000	0.5499	0.0000	0.0000	0.0000
0.05	−0.4448	−1.0000	−1.0000	1.0000	0.0084	0.0000	1.0000	0.0000	0.4918	0.0000
0.10	0.0778	−1.0000	−0.3600	1.0000	−0.0174	0.2238	1.0000	1.0000	0.7545	0.0000
Bayesian approach										
0.02	−0.2021	−0.5494	−0.2922	0.3425	−0.0066	0.2786	0.7879	0.6246	0.4808	0.2101
0.05	−0.0887	−0.1645	0.0891	0.0660	−0.0128	0.4762	0.7315	0.5498	0.5702	0.4473
0.10	−0.1047	−0.0756	0.1006	0.0155	0.0098	0.5190	0.6156	0.5456	0.5860	0.5505
none	−0.0883	−0.0738	0.0666	0.0540	0.0136	0.5028	0.5719	0.6138	0.5604	0.5487

a projection method and a Bayesian approach has been studied. Following granulation experiments with sodium carbonate and different aqueous PEG mixture in a laboratory mixer, the process was simulated with a population balance model in which the particles are described by five internal coordinates. These particles are subject to the transformations coalescence, compaction, breakage, penetration and reaction, each of them characterized by a rate constant. These rate constants were unknown and needed to be estimated.

As a first step, a screening over a large range of parameter combinations was performed to find a parameter combination for which the model matches the experiment results best. A low discrepancy approach was used to distribute the 200 used parameter combinations across the parameter space. In a second step, surrogate models (2nd order response surfaces) were constructed around the previously identified "best" parameter combination, providing a local approximation of the complex granulation model. These surrogate models were then used together with the experimental results in a projection method as well as a Bayesian approach to estimate the values of the unknown model parameters and their uncertainties.

It was found that the model predictions with the parameters obtained through both methods are similar. Furthermore, the uncertainties in the model predictions increased as the experimental uncertainties were increased. Studies of the marginal densities of two-parameter combinations obtained through the Bayesian approach showed a correlation between the collision and breakage rate constant, giving potential hints for further model development. Furthermore, a bimodality in the distribution of the compaction rate constant was observed, raising the question about the identifiability of the studied system. Further studies on this topic seem to be necessary. As part of the system, second order response surfaces were used as local surrogate models to estimate the unknown parameters. However, although it was beyond the scope of this study, it is clear that a response surface that is more global in nature is required. Furthermore, it is necessary to gauge how our uncertainties in our response surface (as an approximation to the model response) influences our uncertainties in the model parameters. With respect to parameter identifiability issues or discrepancy between model prediction and experimental data, the user of the presented methodologies should either obtain more data of the same kind to reduce uncertainties in the model predictions, or more data of a different kind to reduce the possibility of identifiability issues. Both of these problems come under the problem of experimental design, which will be covered in future work.

Acknowledgments

The authors thank the EPRSC (grant EP/E01772X/1), Churchill and Girton Colleges for financial support.

Literature Cited

1. Mort P, Tardos G. Scale-up of agglomeration processes using transformations. *KONA*. 1999;17:64–75.
2. Iveson SM, Litster JD, Hapgood K, Ennis BJ. Nucleation, growth and breakage phenomena in agitated wet granulation processes: a review. *Powder Technol*. 2001;117:3–39.
3. Tardos GI, Khan MI, Mort PR. Critical parameters and limiting conditions in binder granulation of fine powders. *Powder Technol*. 1997;94:245–258.
4. Heinrich S, Peglow M, Mörl L. Unsteady and steady-state particle size distributions in batch and continuous fluidized bed granulation systems. *Chem Eng J*. 2002;86:223–231.
5. Tan HS, Salman AD, Hounslow MJ. Kinetics of fluidised bed melt granulation I: the effect of process variables. *Chem Eng Sci*. 2006;61:1585–1601.
6. Wauters PAL, van de Water R, Litster JD, Meesters GMH, Scarlett B. Growth and compaction behaviour of copper concentrate granules in a rotating drum. *Powder Technol*. 2002;124:230–237.
7. Capes CE, Danckwerts PV. Granule formation by the agglomeration of damp powders—part I: the mechanism of granule growth. *Trans Inst Chem Eng*. 1965;43:T116–T124.
8. Capes CE, Danckwerts PV. Granule formation by the agglomeration of damp powders—part II: the distribution of granules sizes. *Trans Inst Chem Eng*. 1965;43:T125–T130.
9. Newitt DM, Conway-Jones JM. A contribution to the theory and practice of granulation. *Trans Inst Chem Eng*. 1958;36:422–440.
10. Rumpf H. Grundlagen und Methoden des Granulierens. *Chem Ing Tech*. 1958;30:144–158.
11. Rumpf H. Grundlagen und Methoden des Granulierens. 3. Teil: Überblick über die technischen Granulierverfahren. *Chem Ing Tech*. 1958;30:329–336.
12. Realpe A, Velázquez C. Growth kinetics and mechanism of wet granulation in a laboratory-scale high shear mixer: effect of initial polydispersity of particle size. *Chem Eng Sci*. 2008;63:1602–1611.
13. Rajniak P, Mancinelli C, Chern RT, Stepanek F, Farber L, Hill BT. Experimental study of wet granulation in fluidized bed: impact of the binder properties on the granule morphology. *Int J Pharm*. 2007;334:92–102.
14. Knight PC, Instone T, Pearson JMK, Hounslow MJ. An investigation into the kinetics of liquid distribution and growth in high shear mixer agglomeration. *Powder Technol*. 1998;97:246–257.
15. Ax K, Feise H, Sochon R, Hounslow M, Salman A. Influence of liquid binder dispersion on agglomeration in an intensive mixer. *Powder Technol*. 2008;179:190–194.
16. Gatzke EP, Doyle FJ III. Model predictive control of a granulation system using soft constraints and prioritized control objectives. *Powder Technol*. 2001;121:149–158.
17. Mort PR, Capeci SW, Holder JW. Control of agglomerate attributes in a continuous binder-agglomeration process. *Powder Technol*. 2001;117:173–176.
18. Morgan NM, Wells CG, Goodson MJ, Kraft M, Wagner W. A new numerical approach for the simulation of the growth of inorganic nanoparticles. *J Comput Phys*. 2006;211:638–658.
19. West RH, Celnik MS, Inderwildi OR, Kraft M, Beran GJO, Green WH. Towards a comprehensive model of the synthesis of TiO_2 particles from $TiCl_4$. *Ind Eng Chem Res*. 2007;46:6147–6156.
20. Celnik MS, Sander M, Raj A, West RH, Kraft M. Modelling soot formation in a premixed flame using an aromatic-site soot model and an improved oxidation rate. *Proc Combust Inst*. 2009;32:639–646.
21. Kapur PC, Fuerstenau DW. A coalescence model for granulation. *I&EC Process Design Dev*. 1969;8:56–62.
22. Sastry KVS. Similarity size distribution of agglomerates during their growth by coalescence in granulation or green pelletization. *Int J Miner Process*. 1975;2:187–203.
23. Braumann A, Goodson MJ, Kraft M, Mort PR. Modelling and validation of granulation with heterogeneous binder dispersion and chemical reaction. *Chem Eng Sci*. 2007;62:4717–4728.
24. Darelius A, Rasmuson A, Björn IN, Folestad S. High shear wet granulation modelling—a mechanistic approach using population balances. *Powder Technol*. 2005;160:209–218.
25. Verkoeijen D, Pouw GA, Meesters GMH, Scarlett B. Population balances for particulate processes–a volume approach. *Chem Eng Sci*. 2002;57:2287–2303.
26. Poon JM-H, Immanuel CD, Doyle FJ III, Litster JD. A three-dimensional population balance model of granulation with a mechanistic representation of the nucleation and aggregation phenomena. *Chem Eng Sci*. 2008;63:1315–1329.
27. Vikhansky A, Kraft M. A Monte Carlo methods for identification and sensitivity analysis of coagulation processes. *J Comput Phys*. 2004;200:50–59.
28. Vikhansky A, Kraft M, Simon M, Schmidt S, Bart H-J. Droplets population balance in a rotating disc contactor: an inverse problem approach. *AIChE J*. 2006;52:1441–1450.

29. Vikhansky A, Kraft M. Two methods for sensitivity analysis of coagulation processes in population balances by a Monte Carlo method. *Chem Eng Sci*. 2006;61:4966–4972.
30. Man PLW, Norris JR, Bailleul IF, Kraft M. Coupling algorithms for calculating sensitivities of Smoluchowski's coagulation equation. *SIAM J Sci Comput*. 2010;32:635–655.
31. Bailleul IF, Man PLW, Kraft M. A stochastic algorithm for parametric sensitivity in Smoluchowski's coagulation equation. *SIAM J Numer Anal*. 2010;48:1064–1086.
32. Muralidar R, Ramkrishna D. An inverse problem in agglomeration kinetics. *J Colloid Interface Sci*. 1986;112:348–361.
33. Zauner R, Jones AG. Determination of nucleation, growth, agglomeration and disruption kinetics from experimental precipitation data: the calcium oxalate system. *Chem Eng Sci*. 2000;55:4219–4232.
34. Nopens I, Koegst T, Mahieu K, Vanrolleghem PA. PBM and activated sludge flocculation: from experimental data to calibrated model. *AIChE J*. 2005;51:1548–1557.
35. Ramachandran R, Barton PI. Effective parameter estimation within a multi-dimensional population balance model framework. *Chem Eng Sci*. 2010;65:4884–4893.
36. Braumann A, Kraft M, Mort PR. Parameter estimation in a multidimensional granulation model. *Powder Technol*. 2010;197:196–210.
37. Braumann A, Kraft M. Incorporating experimental uncertainties into multivariate granulation modelling. *Chem Eng Sci*. 2010;65:1088–1100.
38. Braumann A, Man PLW, Kraft M. Statistical approximation of the inverse problem in multivariate population balance modeling. *Ind Eng Chem Res*. 2010;49:428–438.
39. Sheen DA, You X, Wang H, Løvås T. Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion. *Proc Combust Inst*. 2009;32:535–542.
40. Parthasarathy S, Balaji C. Estimation of parameters in multi-mode heat transfer problems using Bayesian inference—Effect of noise and a priori. *Int J Heat Mass Transfer*. 2008;51:2313–2334.
41. Hermanto MW, Kee NC, Tan RBH, Chiu M-S, Braatz RD. Robust Bayesian estimation of kinetics for polymorphic transformation of L-glumatic acid crystals. *AIChE J*. 2008;54:3248–3259.
42. Jones JR, Bridgwater J. A case study of particle mixing in a ploughshare mixer using positron emission particle tracking. *Int J Miner Process*. 1998;53:29–38.
43. Braumann A, Kraft M, Wagner W. Numerical study of a stochastic particle algorithm solving a multidimensional population balance model for high shear granulation. *J Comput Phys*. 2010;229:7672–7691.
44. Patterson RIA, Singh J, Balthasar M, Kraft M, Norris JR. The linear process deferment algorithm: a new technique for solving population balance equations. *SIAM J Sci Comput*. 2006;28:303–320.
45. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in Fortran 77: The Art of Scientific Computing;1:299–306*. Cambridge: Cambridge University Press, 1992.
46. Halton JH. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer Math*. 1960;2:84–90.
47. Lindley DV. The philosophy of statistics. *J Roy Stat Soc D-Sta*. 2000;49:293–337.
48. Beers KJ. *Numerical Methods for Chemical Engineering*. Cambridge: Cambridge University Press, 2007.
49. Bernardo JM. *Bayesian methodology in statistics*. In: Brown SD, Tauler R, Walczak B, editor. *Comprehensive Chemometrics, Vol. 1*. Oxford: Elsevier, 2009:213–245.
50. Beck JL, Katafygiotis LS. Updating models and their uncertainties. I: Bayesian statistical framework. *J Eng Mech*. 1998;124:455–461.
51. Man PLW, Braumann A, Kraft M. Resolving conflicting parameter estimates in multivariate population balance models. *Chem Eng Sci*. 2010;65:4038–4045.
52. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC, 2004.
53. Hastings WK. Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*. 1970;57:97–109.
54. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. 1953;21:1087–1091.
55. Cheung SH, Beck JL. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *J Eng Mech*. 2009;135:243–255.

Appendix: Transformation Between Coded and Uncoded Variables

As the model parameters are usually dimensional and in a different size range other than [0, 1], the sampling points need to be transformed into the dimensional model parameters. A linear and logarithmic transformation are presented below.

The normalized variables obtained from a sequence shall be denoted as coded variables, x, and the dimensional model parameters shall be called uncoded variables, y, (Figure A1).

For each uncoded variable exists a corresponding coded variable,

$$y_{\text{low}} \rightarrow x_{\text{low}}, \quad y \rightarrow x, \quad y_{\text{up}} \rightarrow x_{\text{up}}. \tag{A1}$$

Note that this definition is one-dimensional. For quantities with more than one dimension, we simply apply the transformation rules to each dimension separately.

Linear transformation

In case of a linear transformation between the coded and uncoded variables, following relationship exists,

$$\frac{y - y_{\text{low}}}{x - x_{\text{low}}} = \frac{y_{\text{up}} - y_{\text{low}}}{x_{\text{up}} - x_{\text{low}}}, \tag{A2}$$

so that

$$y = \frac{y_{\text{up}} - y_{\text{low}}}{x_{\text{up}} - x_{\text{low}}} (x - x_{\text{low}}) + y_{\text{low}}, \tag{A3}$$

and after introduction of $S = \frac{y_{\text{up}} - y_{\text{low}}}{x_{\text{up}} - x_{\text{low}}}$ we get

$$y = S (x - x_{\text{low}}) + y_{\text{low}}. \tag{A4}$$

Let the coded variable x be expressed by

$$x = x_0 + c\,\xi, \tag{A5}$$

with ξ being standard normally distributed. Hence

$$x \sim \mathcal{N}\left(x_0, c^2\right). \tag{A6}$$

Together with Eq. A4 this leads to

$$y \sim \mathcal{N}\left(S (x_0 - x_{\text{low}}) + y_{\text{low}}, S^2 c^2\right). \tag{A7}$$

This means the model parameter y with its mean y_0 and its uncertainty δy can be expressed by

$$y = y_0 \pm \delta y, \quad = S (x_0 - x_{\text{low}}) + y_{\text{low}} \pm S\,c. \tag{A8}$$

Logarithmic transformation

In this case, the following relationship shall be used,

$$y = y_{\text{base}} f^x. \tag{A9}$$

If the problem presents itself as in Figure A1, then it is possible to express y_{base} and f in terms of x_{low}, x_{up}, y_{low}, and y_{up}, so that one obtains,

$$y_{\text{base}} = \frac{y_{\text{low}}}{f^{x_{\text{low}}}} = \frac{y_{\text{up}}}{f^{x_{\text{up}}}}, \tag{A10}$$

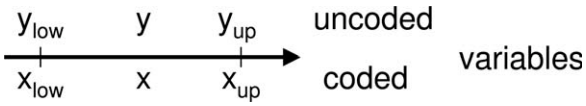

Figure A1. Uncoded (*y*) and coded (*x*) variables.

yielding

$$f = \left(\frac{y_{up}}{y_{low}}\right)^{\frac{1}{x_{up}-x_{low}}}. \tag{A11}$$

The combination of Eqs. A9–A11 provides y as a function of x,

$$y = y_{up}\left(\frac{y_{low}}{y_{up}}\right)^{\frac{x_{up}-x}{x_{up}-x_{low}}}. \tag{A12}$$

For the calculation of the uncertainties we perform two substitutions in Eq. A12,

$$S = \frac{y_{up}}{y_{low}}, \tag{A13}$$

$$\tilde{x} = \frac{x - x_{up}}{x_{up} - x_{low}}, \tag{A14}$$

leading to

$$y = y_{up}\, S^{\tilde{x}}, \tag{A15}$$

which can be transformed into

$$y = y_{up}\, \exp(\tilde{x}\, \ln S). \tag{A16}$$

The most likely values for the model parameter y fall within

$$y = [y^-(x^-); y^+(x^+)]. \tag{A17}$$

Since

$$x \sim \mathcal{N}(x_0, c^2) \tag{A18}$$

we have

$$x^- = x_0 - \delta x^-, \tag{A19}$$

$$x^+ = x_0 + \delta x^+. \tag{A20}$$

If δx^- and δx^+ are the one σ uncertainty in x then,

$$\delta x^- = \delta x^+ = \delta x = c. \tag{A21}$$

This means the lower and upper bound for y take following forms,

$$y^- = y_{up}\, \exp\left(\frac{x_0 - c - x_{up}}{x_{up} - x_{low}} \cdot \ln\left(\frac{y_{up}}{y_{low}}\right)\right), \tag{A22}$$

$$y^+ = y_{up}\, \exp\left(\frac{x_0 + c - x_{up}}{x_{up} - x_{low}} \cdot \ln\left(\frac{y_{up}}{y_{low}}\right)\right). \tag{A23}$$

Manuscript received Apr. 13, 2010, revision received Sept. 15, 2010, and final revision received Dec. 16, 2010.