Cambridge Centre for Computational Chemical Engineering

University of Cambridge

Department of Chemical Engineering

Preprint ISSN 1473 – 4273

Resolving conflicting parameter estimates in multivariate population balance models

Peter L. W. Man [1] Andreas Braumann [1] Markus Kraft [1]

released: 27 October 2009

[1] Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge, CB2 3RA
UK
E-mail: mk306@cam.ac.uk

Preprint No. 82

c4e

Key words and phrases: granulation, multimodal parameter distribution, parameter identification, particulate processes, population balance, error propagation

Edited by
Cambridge Centre for Computational Chemical Engineering
Department of Chemical Engineering
University of Cambridge
Cambridge CB2 3RA
United Kingdom.

Fax: + 44 (0)1223 334796
E-Mail: c4e@cheng.cam.ac.uk
World Wide Web: http://www.cheng.cam.ac.uk/c4e/

Abstract

We present an extended methodology for parametric inference in complex population balance models. The aim is twofold. Firstly, it is assumed that the parameter distribution of the model is a multimodal Gaussian rather than a unimodal Gaussian. After projection of experimental data through a response surface approximation, estimates for the parameters and their uncertainties along with the associated weights of each mode are established. Secondly, the methodology is used to ask the following question—if n professors each have a 'best' estimate of a particular parameter, which of these estimates is more likely to be correct? A toy example is used to show the applicability of the methodology, aiding in the discrimination between a bimodal and trimodal parameter distribution. The identification of the 'best' model parameter among two conflicting estimates is demonstrated in an example from granulation modelling.

Contents

1 Introduction

In this article, we present an extended methodology for solving parametric inverse problems for complex population balance models. Furthermore, this methodology helps to resolve conflicting parameter estimates. Population balance models are in widespread use in chemical engineering, for instance for crystallisation processes [15, 34], biological systems [1, 20], liquid-liquid extraction [3, 30, 35], combustion [10, 17, 27, 28], nanoparticle synthesis [14, 18, 19, 24, 33], and granulation [9, 22]. The relevance of the latter is reflected in the variety of equipment in which granulation processes are performed, for example, fluidised beds [12, 29], drum granulators [2] and high shear mixers [5, 11]. Before these models can be used to make predictions about the process behaviour, one is often faced with the inverse problem, i. e., unknown parameters of the models need to be established [23]. When solving this problem, the sensitivities of the process with respect to the unknown parameters are required, whilst special algorithms are available for the sensitivity analysis of coagulation processes [16, 31, 32]. A major difficulty in solving the inverse problem is that often the computational model in question, for predicting outcomes as a function of the unknown parameters, requires much computational effort to evaluate. Faced with this problem, we seek a response surface approximation to the computational model response, and so this surrogate model now replaces the true model response. This approach has for instance been used in the modelling of combustion processes [13]. Given the surrogate model, we then have to solve the problem of considering how any uncertainties in experimental data should inform us on how certain we are about our parameter estimates. This article extends the approach taken in [4, 25], where it was assumed that the parameter distribution is a unimodal Gaussian, when in fact it is an arbitrary distribution in general. We settle with approximating this with a multimodal Gaussian distribution with associated weights. This gives reasonable approximations to the arbitrary distribution, as well as easily interpretable results—the multi-modality allows for different reasonable parameter values to be considered. The multimodal generalisation is mathematically more challenging than the unimodal case.

This feature of the presented methodology is used to look at following question: "If n professors each have a 'best' estimate of a particular model parameter, which of these is more likely to be correct?". The quality and applicability of the methodology is demonstrated for two examples. Besides a toy example, a real life example from wet granulation modelling is presented. The process is modelled with a multidimensional population balance approach, where the coalescence transformation is a function of the collision frequency constant. Under the assumption that this constant has been estimated by two researchers, both using a different method, the current methodology is used to identify which of these values is more likely to be correct.

2 Problem description

In this section, we describe the situation a researcher is in, where he/she is armed with some experimental data as well as a mathematical model of the physical phenomenon in question. The *observed* experimental data are denoted by the vector

$\boldsymbol{\eta}_0^{\text{exp}} = (\eta_{0,1}^{\text{exp}}, \ldots, \eta_{0,N}^{\text{exp}})^\top \in \mathbb{R}^N$ where datum $\eta_{0,i}^{\text{exp}}$ was procured under experimental condition indexed by $i \in 1, \ldots, N$. We assume the researcher also has available the corresponding experimental uncertainties $\boldsymbol{\sigma}^{\text{exp}} = (\sigma_1^{\text{exp}}, \ldots, \sigma_N^{\text{exp}})^\top \in \mathbb{R}_+^N$. This is interpreted in the following way—the '*true*' experimental datum η_i^{exp} has an *uncertainty distribution* of a univariate Gaussian distribution with mean $\eta_{i,0}^{\text{exp}}$ and standard deviation σ_i^{exp}, independently over i, i. e.,

$$\eta_i^{\text{exp}} \sim \mathcal{N}\left(\eta_{0,i}^{\text{exp}}, (\sigma_i^{\text{exp}})^2\right) \quad , \tag{1a}$$

or in multivariate language,

$$\boldsymbol{\eta}^{\text{exp}} \sim \mathcal{N}_N\left(\boldsymbol{\eta}_0^{\text{exp}}, \Sigma^{\text{exp}}\right) \tag{1b}$$

where Σ^{exp} is the $N \times N$ diagonal matrix with diagonal entries $(\sigma_1^{\text{exp}})^2, \ldots, (\sigma_N^{\text{exp}})^2$, and $\mathcal{N}_N$ denotes the N-variate Gaussian distribution.

The mathematical model to estimate the '*true*' datum η_i^{exp} is denoted by $\eta_i(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, \ldots, x_K)^\top \in \mathbb{R}^K$ is some *unknown* vectorial parameter whose value the researcher wishes to infer using the experimental data and the model. In future sections, we test our methodology in a real life example taken from the field of granulation modelling—we take η_i^{exp} to be the average mass of agglomerates, where $i \in 1, \ldots, N$ and in this case, the set of experimental conditions is some combination of impeller speed and composition of the binder added to the granules.

We make a (not unreasonable) assumption that the researcher has an *a priori* assumption that the *most likely* values for x_k lie in $[a_k, b_k]$ for each $k = 1, \ldots, K$ for some a_k, b_k. The x_k are rescaled accordingly so that $a_k = -1$ and $b_k = 1$, i. e., for the rest of this paper, $x_k \in [-1, 1]$ for all k.

2.1 Model response approximation

Suppose that it is computationally expensive to evaluate $\eta_i(\boldsymbol{x})$ for a given $\boldsymbol{x}$. The approach taken to relieve this computational burden is one used in [4] in order to approximate $\eta_i(\boldsymbol{x})$ locally by a second order response surface (dropping the i indices for convenience):

$$\eta(\boldsymbol{x}) \approx \beta_0 + \sum_{k=1}^{K} \beta_k \, x_k + \sum_{k=1}^{K} \sum_{l \geq k}^{K} \beta_{kl} \, x_k \, x_l \quad , \tag{2}$$

with β_0, β_k and β_{kl} being the coefficients of the response surfaces (the i indices have been dropped from these too). For the rest of this paper, we simply replace the true model response by this surrogate version. To further simplify the exposition, we will rewrite eq. (2) as follows:

$$\begin{aligned} \eta(\boldsymbol{x}) &= \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x} + \boldsymbol{x}^\top B \boldsymbol{x} \\ \text{with} \quad \boldsymbol{\beta} &= (\beta_1, \beta_2, \ldots, \beta_K)^\top \quad , \end{aligned} \tag{3}$$

where B is a $K \times K$ matrix with elements $B_{kk} = \beta_{kk}$ and $B_{kl} = B_{lk} = \frac{1}{2}\beta_{kl}$ for $k < l$. Note that B is *symmetric* by construction.

2.2 Parametric inference

A difficulty with parametric inference is that the *true* experimental data $\boldsymbol{\eta}^{\text{exp}}$ is uncertain (eq. (1b)). Any parameter estimate based purely on the observed value $\boldsymbol{\eta}_0^{\text{exp}}$ is potentially highly sensitive to this value. In fact, the uncertainty in $\boldsymbol{\eta}^{\text{exp}}$ induces uncertainties in the parameter values $\boldsymbol{x}$. Thus the approach taken in [4] is followed, where the uncertainty distribution eq. (1b) is 'mapped' through the model response (or the surrogate model eq. (3)) to find the uncertainty distribution of $\boldsymbol{x}$ i. e., $\boldsymbol{x}$ is taken to be a K-variate random variable. In general, the distribution of $\boldsymbol{x}$ can have arbitrary form—however, for the sake of simplicity of computation and interpretability, we restrict the form of this distribution to be multimodal Gaussian. Note that in [4], $\boldsymbol{x}$ was taken to be unimodal Gaussian, and thus this article extends their work.

We now give the definition of a multimodal Gaussian distribution—a scalar random variable y has a scalar multimodal Gaussian distribution if *conditional* on another (discrete) random variable m called the **random mode**, y has a scalar unimodal Gaussian distribution. We express this in mathematical notation as:

$$y|\{m = d\} \sim \mathcal{N}(y_0(d), c(d)^2) \quad \text{where } d \in \{1, \ldots, M\} \text{ and} \tag{4a}$$
$$\mathbb{P}(m = d) = w(d) \quad . \tag{4b}$$

The first line of eq. (4) says that *conditional* on the event that the random mode m is d, y has a Gaussian distribution with mean $y_0(d)$ and standard deviation $c(d)$. The second line gives the distribution of the random mode m, i. e., the probability of being in mode d is $w(d)$. We call $w(d)$ the **weight** of mode d. Note that since d takes values in $1, \ldots, M$, we have M modes, and associated with each mode d, we have the values $y_0(d)$, $c(d)$ and $w(d)$.

We are now in the position to give the form of the $\boldsymbol{x}$ distribution. *Each component* x_k is assumed to have a multimodal Gaussian distribution, independently over the components k. Mathematically, this is:

$$x_k|\{m_k = d_k\} \sim \mathcal{N}(x_{0,k}(d_k), c_k(d_k)^2) \quad \text{where } d_k \in \{1, \ldots, M_k\} \text{ and} \tag{5a}$$
$$\mathbb{P}(m_k = d_k) = w_k(d_k) \quad . \tag{5b}$$

Note that for each component x_k, we have a random mode m_k which takes values in $1, \ldots, M_k$ where the m_k are independently distributed over k. For each value d_k of m_k, we have the values $x_{0,k}(d_k)$, $c_k(d_k)$ and $w_k(d_k)$. For ease of exposition, we rephrase eq. (5) in terms of vectors and matrices:

$$\begin{aligned} \boldsymbol{x} \,|\, \boldsymbol{m} &\sim \mathcal{N}_K(\boldsymbol{x}_0(\boldsymbol{m}), V(\boldsymbol{m})) \\ \text{with} \quad \boldsymbol{m} &= (m_1, m_2, \ldots, m_K)^\top \\ V(\boldsymbol{m}) &= \text{diag}\left[c_1^2(\boldsymbol{m}), c_2^2(\boldsymbol{m}), \ldots, c_K^2(\boldsymbol{m})\right] \ , \end{aligned} \tag{6}$$

where $\mathcal{N}_K(\boldsymbol{x}_0(\boldsymbol{m}), V(\boldsymbol{m}))$ denotes the joint multivariate Gaussian distribution (of dimension K) with mean $\boldsymbol{x}_0(\boldsymbol{m})$ and covariance matrix $V(\boldsymbol{m})$. See that $V(\boldsymbol{m})$ is a diagonal matrix since each of the x_k are assumed to be independent Gaussian random variables given $\boldsymbol{m}$. Note the ease of interpretability of the distributional assumption on $\boldsymbol{x}$—we can

say that for each x_k, there are multiple modes, and the weight of each mode giving the probability that the mean of that mode is the correct value of x_k. Of course, it is non-trivial how to pick the 'best' parameter estimate for x_k, but sometimes giving the full answer is better than forcing a single-point answer—if it really were clear that x_k is almost certainly one particular value, our computation would show that only one mode of x_k has weight of nearly unity. On the other hand, if x_k has two modes of nearly equal weight, then the proper answer is that the means of both modes are equally probable of being the true value of x_k. In this situation, it is clear that more experiments are required, thus providing a mechanism for decision making in terms of performing experiments.

2.3 Objective function

Earlier, the concept of 'mapping' the uncertainty distribution of $\boldsymbol{\eta}^{\text{exp}}$ onto the uncertainty distribution of $\boldsymbol{x}$ was discussed—but since the distribution of $\boldsymbol{x}$ has been restricted to that of a multimodal Gaussian, this 'mapping' reduces down to the problem of suitably choosing the parameters of this multimodal distribution, namely the means $\{\boldsymbol{x}_0(\boldsymbol{m})\}_{\boldsymbol{m}}$, the variances $\{V(\boldsymbol{m})\}_{\boldsymbol{m}}$ and the weights $\{\boldsymbol{w}(\boldsymbol{m})\}_{\boldsymbol{m}}$.

The 'mapping' can be described differently. The randomness of the $\boldsymbol{x}$ induces randomness in the surrogate model response $\boldsymbol{\eta}(\boldsymbol{x})$. Thus, the values of $\{\boldsymbol{x}_0(\boldsymbol{m})\}_{\boldsymbol{m}}$, $\{V(\boldsymbol{m})\}_{\boldsymbol{m}}$ and $\{\boldsymbol{w}(\boldsymbol{m})\}_{\boldsymbol{m}}$ are chosen to minimise the difference between the distributions of $\boldsymbol{\eta}(\boldsymbol{x})$ and the uncertainty distribution of $\boldsymbol{\eta}^{\text{exp}}$. In exactly the same way as in [4], this notion of difference between the distributions is described by some objective function Φ. The value of $\boldsymbol{T} := (\{\boldsymbol{x}_0(\boldsymbol{m})\}_{\boldsymbol{m}}, \{V(\boldsymbol{m})\}_{\boldsymbol{m}}, \{\boldsymbol{w}(\boldsymbol{m})\}_{\boldsymbol{m}})$ which minimises Φ gives us the parametric distribution which best matches the experimental uncertainty distribution.

For this paper, we choose the following objective function (thereby departing from those used in previous works such as [4]):

$$\Phi(\boldsymbol{T}) = \sum_{i=1}^{N} \left[2K_i(\boldsymbol{T}) + \left[\sigma_i^{\text{exp}} - \sigma_i(\boldsymbol{T})\right]^2 \right] + 0.05 \sum_{k=1}^{K} \sum_{d_k=1}^{M_k} \mathbb{1}_{\{3\, M_k\, w_k(d_k)<1\}} \quad [1] \quad , \tag{7}$$

where

$$\begin{aligned} K_i(\boldsymbol{T}) &:= \sum_{d_1=1}^{M_1} \sum_{d_2=1}^{M_2} \cdots \sum_{d_K=1}^{M_K} \left\{ \left(\eta_{0,i}^{\text{exp}} - \mathbb{E}[\eta_i(\boldsymbol{x}|\boldsymbol{m}=\boldsymbol{d})]\right)^2 \prod_{k=1}^{K} w_k(d_k) \right\} \\ \sigma_i(\boldsymbol{T}) &:= \sqrt{\text{Var}\left(\eta_i(\boldsymbol{x})\right)} \quad . \end{aligned} \tag{8}$$

The interpretation of this objective function is as follows—$K_i(\boldsymbol{T})$ represents the weighted squared difference between the experimental data point $\eta_{0,i}$ and the model function η_i evaluated at all the possible mode combinations, whilst $[\sigma_i^{\text{exp}} - \sigma_i(\boldsymbol{T})]^2$ is the squared difference between the standard deviations of the data uncertainty distribution and the random model response. The last term in eq. (7) penalises for weights being too small,

[1] $\mathbb{1}_{\{\text{condition}\}} = \begin{cases} 1 & \text{, if condition is true} \\ 0 & \text{, otherwise} \end{cases}$

and thus discouraging possible redundant modes. Note that this is a constrained optimisation problem—not only do the x_k have to satisfy the bounds $x_k \in [-1, 1]$ for all k, but the weights must sum to unity, i. e., $\sum_{d_k=1}^{M_k} w_k(d_k) = 1$ for all k, since the weights are probabilities. It may also be convenient to constrain the values of $c_k(m_k)$ to be in some region between 0 and some upper bound since we have assumed that the likely values of x_k lie in $[-1, 1]$.

In this paper, the objective function specified in eq. (7) was minimised by using the MATLAB function `fmincon` starting with many randomly generated initial conditions for $\boldsymbol{T}$. The minimising $\boldsymbol{T}$ out of each of these is taken as our best value of $\boldsymbol{T}$.

2.4 Conflicting parameter values

The second part of the problem description is to solve the problem of deciding the 'best' parameter estimate between a set of conflicting estimates (and their uncertainties) as suggested by various experts. In an earlier part of this paper, the issue of interpretability of the multimodality in the distribution of $\boldsymbol{x}$ was discussed, i. e., the usage of the weights as a measurement of probabilities of different modal means being the correct value of $\boldsymbol{x}$. This gives rise to a natural mechanism for deciding the probabilities of conflicting estimates of x_k. The method is as follows: first we choose M_k (i. e., the number of modes) to be equal to the number of conflicting estimates. Then we *fix* the modal means $x_{0,k}(1), \ldots, x_{0,k}(M_k)$ to be the parameter estimates themselves (and $c_k(1), \ldots, c_k(M_k)$ to be the given parameter uncertainties if they are available) in the aforementioned optimisation where previously they were free to vary. After this optimisation has been carried out, the resulting weights $w_k(1), \ldots, w_k(M_k)$ provide the relevant probabilities. As mentioned previously, it is not trivial in general how to decide which mode provides the best estimate for x_k, and one may have to be satisfied with a purely probabilistic statement about x_k, indicating that more experiments need to be carried out.

2.5 Explicit calculations

In this section, the aim is to give a more explicit version of eq. (7), primarily the computation of $K_i(\boldsymbol{T})$ and $\sigma_i(\boldsymbol{T})$ as given in eq. (8). For ease of exposition, we drop all the indices i. It is strongly recommended to read the appendices if the reader wishes to verify these calculations.

Since there are two sources of randomness in the response surface $\eta(\boldsymbol{x})$ (i. e., from $\boldsymbol{m}$ and $\boldsymbol{x}|\boldsymbol{m}$), computation of its mean and variance is made far easier by first considering the conditional distributions *given* $\boldsymbol{m}$, and thus eliminating one source of randomness. Therefore, let us consider the following two quantities (some expectation and variance identities are given in Appendix A):

- $\mathbb{E}(\eta|\boldsymbol{m})$
- $\mathrm{Var}\,(\eta|\boldsymbol{m})$.

To compute the first quantity, $\mathbb{E}(\eta|\boldsymbol{m})$, we note that η is a function of $\boldsymbol{x}$ and the conditional distribution of $\boldsymbol{x}|\boldsymbol{m}$ is described by eq. (6). So:

$$\begin{aligned}\mathbb{E}(\eta|\boldsymbol{m}) &= \mathbb{E}(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x} + \boldsymbol{x}^\top B\boldsymbol{x}|\boldsymbol{m}) \\ &= \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_0(\boldsymbol{m}) + \operatorname{tr}\left[BV(\boldsymbol{m})\right] + \boldsymbol{x}_0(\boldsymbol{m})^\top B\boldsymbol{x}_0(\boldsymbol{m})\end{aligned} \tag{9}$$

where the expression for $\mathbb{E}\left[\boldsymbol{x}^\top B\boldsymbol{x}|\boldsymbol{m}\right]$ follows *immediately* from Theorem 5 in Appendix B since $\mathbb{E}\left[\boldsymbol{x}|\boldsymbol{m}\right] = \boldsymbol{x}_0(\boldsymbol{m})$ and $\operatorname{Var}\left(\boldsymbol{x}|\boldsymbol{m}\right) = V(\boldsymbol{m})$. The other quantity $\operatorname{Var}\left(\eta|\boldsymbol{m}\right)$ can be shown to be:

$$\begin{aligned}\operatorname{Var}\left(\eta|\boldsymbol{m}\right) &= \operatorname{Var}\left(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x} + \boldsymbol{x}^\top B\boldsymbol{x}|\boldsymbol{m}\right) \\ &= \boldsymbol{\beta}^\top V(\boldsymbol{m})\boldsymbol{\beta} + 2\operatorname{tr}\left[B\,V\left(\boldsymbol{m}\right)B\,V\left(\boldsymbol{m}\right)\right] \\ &\quad + 4\,\boldsymbol{x}_0(\boldsymbol{m})^\top B\,V(\boldsymbol{m})B\boldsymbol{x}_0(\boldsymbol{m}) + 4\,\boldsymbol{x}_0^\top\,B\,V\,\boldsymbol{\beta} \quad ,\end{aligned} \tag{10}$$

using Theorem 7 in Appendix B. See that eqs. (9) and (10) show that $\mathbb{E}\left[\eta|\boldsymbol{m}\right]$ and $\operatorname{Var}\left(\eta|\boldsymbol{m}\right)$ are functions of $\boldsymbol{m}$. Now, the objective function demands that the *total* mean μ and total variance σ^2 are evaluated as below:

$$\begin{aligned}\mu(\boldsymbol{T}) &= \mathbb{E}(\eta) \\ &= \mathbb{E}\left[\mathbb{E}(\eta|\boldsymbol{m})\right] \quad \text{(by iterated expectation)}\end{aligned} \tag{11}$$

$$\begin{aligned}\sigma(\boldsymbol{T})^2 &= \operatorname{Var}\left(\eta\right) \\ &= \operatorname{Var}\left(\mathbb{E}(\eta|\boldsymbol{m})\right) + \mathbb{E}\left[\operatorname{Var}\left(\eta|\boldsymbol{m}\right)\right] \quad \text{(by eq. (22))} \quad .\end{aligned} \tag{12}$$

Note that both $\mu(\boldsymbol{T})$ and $\sigma(\boldsymbol{T})$ have now been expressed as expectation and variance of functions of $\boldsymbol{m}$. Computing these explicitly is cumbersome and unnecessary—it is easier to compute these via a brute force method. More explicitly, suppose we wish to compute the expectation and variance of a function $f(\boldsymbol{m})$, of only $\boldsymbol{m}$, then we perform this by:

$$\begin{aligned}\mathbb{E}\left[f(\boldsymbol{m})\right] &= \sum_{d_1=1}^{M_1} \cdots \sum_{d_K=1}^{M_K} \mathbb{P}\left[m_1 = d_1, \ldots, m_K = d_K\right]\, f(d_1, \ldots, d_K) \\ &= \sum_{d_1=1}^{M_1} \cdots \sum_{d_K=1}^{M_K} \left[\prod_{k=1}^{K} w_k(d_k)\right]\, f(d_1, \ldots, d_K)\end{aligned} \tag{13a}$$

and

$$\operatorname{Var}\left(f(\boldsymbol{m})\right) = \mathbb{E}\left[f(\boldsymbol{m})^2\right] - \left(\mathbb{E}\left[f(\boldsymbol{m})\right]\right)^2 \tag{13b}$$

by definition of expectations and variances, and using the fact that the components of $\boldsymbol{m}$ are mutually independent to infer that the joint distribution for $\boldsymbol{m}$ is

$$\mathbb{P}\left[m_1 = d_1, \ldots, m_K = d_K\right] = \prod_{k=1}^{K} \mathbb{P}\left[m_k = d_k\right] = \prod_{k=1}^{K} w_k(d_k)\,. \tag{14}$$

Table 1: *Bimodal test - with 95 % confidence intervals for x*

Mode m	$x_0(m)$	$c(m)$	$w(m)$	CI for $x(m)$
1	-0.241212	0.040472	0.537532	$[-0.3071437, -0.1752803]$
2	0.243928	0.018154	0.462468	$[0.1779963, 0.3098597]$

3 Results and discussion

This section is split into two parts—the first one showing a toy example where we demonstrate the method's ability to find the 'true' parameter value, as well as considering the problem of conflicting information. The second part applies the methodology to a complex example taken from the field of population balance modelling—more explicitly, the modelling of granulation processes.

3.1 Toy example

In this toy example, a 'physical' phenomenon is invented where the 'true' system can be described by a scalar quadratic function of some invented parameter $x \in \mathbb{R}$ whose true value is $x = x_{\text{true}}$. Thus any quadratic surrogate model as described earlier is equivalent to the true model response $\eta(x) \in \mathbb{R}$. The data $\eta^{\text{exp}} \in \mathbb{R}$ are randomly generated from this true system by adding some gaussian noise about $\eta(x_{\text{true}})$ with predetermined standard deviations $\sigma^{\text{exp}} \in \mathbb{R}$. In this particular example, we take the model response to be $\eta(x) = 2x^2 + 3$, $x_{\text{true}} = 0.25$, and $\sigma^{\text{exp}} = 0.03125$, and the 18 generated data points to have mean $\eta_0^{\text{exp}} = 3.11965$.

Given that our researcher is armed *only* with the data η_0^{exp}, uncertainty σ^{exp} and the (presumed) true model response $\eta(x)$, estimation of the parameter distribution proceeds as described in earlier sections. Note that since the model response is symmetric about zero and is quadratic, then given the data, it should be always impossible to distinguish between the parameter values $x = \pm 0.25$. Thus, in the first instance, an analysis is performed by assuming that the parameter distribution is bimodal—the results are shown in **Table 1**. It is clear that we achieve what is expected, that the modal means are approximately $x = \pm 0.25$, the standard deviations are similar, and the weights are almost equal, implying the indistinguishability between the two modes. One may legitimately ask what the purpose of this analysis is, after all, ultimately we seek the true value of the parameter $x_{\text{true}} = 0.25$. However, the analysis shows *what is knowable given the data, their uncertainties, and the model*. When confronted with this situation, it is clear that we need to increase the number of dimensions of the model response as well as obtaining more (and different) data in order to reduce the size of regions of the parameter space which have high probability. As a further check, we perform exactly the same operations, but assuming that the parameter distribution is trimodal rather than bimodal. This is to check that we still produce a bimodal distribution (as we expect), even despite our trimodal assumption. If the distribution is truly bimodal, we expect to have some redundant modes, even despite the penalty term in eq. (7). **Table 2** shows the results—they are mostly as we expect in that the third mode is located near $x = 0.25$ with about 16 % weight, whilst

Table 2: *Trimodal test - with 95 % confidence intervals for x*

Mode m	$x_0(m)$	$c(m)$	$w(m)$	CI for $x(m)$
1	-0.241543	0.038769	0.683315	$[-0.3074747, -0.1756113]$
2	-0.244411	0.000000	0.159319	$[-0.244411, -0.244411]$
3	0.244639	0.000000	0.157366	$[0.244639, 0.244639]$

the other two modes are almost identical with means near $x = -0.25$ and their combined weight being 84 %. The weights are a slight issue, however the fact that two of the modes can be combined is an indication that we have assumed too many modes in the first place. This toy example demonstates the method's ability to produce results that we expect.

3.2 Example from Granulation Modelling

Our second, real life example is taken from granulation modelling. With this we will demonstrate the problem of pre-existing conflicting information of our parameters $\boldsymbol{x}$.

3.2.1 The granulation model

For this example we use a granulation model that has been developed in a previous study [6]. This model is based on a multivariate population balance describing the particles/granules by five independent variables, which are the volumes of original solid, reacted solid, external liquid, internal liquid and pores. During the batch granulation process, the particles are subject to various transformations, namely coalescence, compaction, breakage, penetration and reaction.

Owing to the complexity, the model makes use of some 20 parameters, some of them unknown, so that the inverse problem has to be solved. In an earlier study, [6], the rate constants for the subprocesses coalescence, $\widehat{K}_0$, compaction, k_{porred}, breakage, $\widehat{k}_{\text{att}}$, and reaction, k_{reac}, were estimated, so that we get

$$\boldsymbol{x} = (x_1, x_2, x_3, x_4)^\top = (\widehat{K}_0, k_{\text{porred}}, \widehat{k}_{\text{att}}, k_{\text{reac}})^\top \quad . \tag{15}$$

For convenience we transform the non-dimensionless model parameters into coded, dimensionless variables $\boldsymbol{x}$. The response surfaces are then constructed so that they are expressed as functions of $\boldsymbol{x}$. The surrogate model is attractive as an approximation to the complex granulation model as it leads to smaller computational effort when estimating the unknown parameters distributions.

The model is applied to a batch granulation process of equally sized beads, being mixed with a binder. Such an experimental study has been performed by [26], in which they granulated nonpareils in a bench scale mixer with aqueous polyethylene glycol (PEG) 4000 solutions. These experiments were conducted at different impeller speeds (600, 900, 1200 rpm) leading to different amounts of agglomerates, depending on the impeller speed and binder composition. Second order response surfaces are constructed for all scenarios and applied in the analysis, in a similar way to [7].

Table 3: *Granulation Modelling - unimodal*

	Mode	
	m_1	
Parameter	$x_{0,k}$	c_k
x_1	0.584493	0.210468
x_2	-0.724499	0.000000
x_3	-1.000000	0.616593
x_4	1.000000	1.000000

Table 4: *Granulation Modelling - bimodal*

	Mode					
	m_1			m_2		
Parameter	$x_{0,k}$	c_k	w_k	$x_{0,k}$	c_k	w_k
x_1	0.584492	0.210490	0.728818	0.584508	0.210425	0.271182
x_2	-0.724506	0.000005	0.412621	-0.724504	0.000019	0.587379
x_3	-1.000000	0.616588	0.692392	-1.000000	0.616605	0.307608
x_4	1.000000	1.000000	1.000000	-0.234073	0.877109	0.000000

3.2.2 Parametric analysis

In the first instance we see what is obtained by assuming the parameter distribution is unimodal in *each* component—see **Table 3**. For comparison, we assume that the parameter distribution is bimodal in every component instead—see **Table 4**. In much the same way as with the toy example, notice that the modes of the first three parameters are basically identical in terms of $x_{0,k}$ and c_k values to each other, as well as to the corresponding modes in the unimodal case shown in Table 3. In the fourth variable, we see that the mode with its weight being unity is identical to the corresponding mode in Table 3—the other mode is redundant despite the penalty term in eq. (7). This confirms that the parametric distribution is unimodal for each parameter $x_1, \ldots, x_4$.

Suppose now that two estimates for the parameter x_1, i. e., the collision rate constant, are available from the literature. Assuming these have been derived by two researchers using different methods such as particle image velocimetry (PIV) [21] and positron emission particle tracking (PEPT) [8], it turns out that the estimates for the parameter x_1 are different, for instance $x_1 = -0.25,\ 0.7$. With our set of experiments, we want to test which of the reported constant is more likely.

The same method is used as that produced Table 4, except that the modal means for the first parameter x_1 are *fixed* to be -0.25 and 0.7 respectively, in the optimisation procedure. **Table 5** shows these results—note that the overwhelming preferred mode for x_1 is 0.7 as expected, since the 'true' value of x_1 (obtained by us for the unimodal case) is 0.584493. This means, the collision rate constant $x_1 = 0.7$ should be considered as the right estimate for the current system. The assumption that $x_1 = 0.7$ is the 'right' value for this parameter

Table 5: *Granulation Modelling - conflicting 'known' parameters (in first component)*

	Mode					
	m_1			m_2		
Parameter	$x_{0,k}$	c_k	w_k	$x_{0,k}$	c_k	w_k
x_1	0.700000	0.163131	1.000000	-0.250000	0.659889	0.000000
x_2	-0.886458	0.000000	0.360873	-0.886457	0.000000	0.639127
x_3	-1.000000	0.638586	1.000000	0.960042	0.874635	0.000000
x_4	1.000000	1.000000	0.367740	1.000000	1.000000	0.632260

has a direct impact on the estimation of the compaction rate constant x_2, since we have now $x_2 = -0.89$ compared to about -0.72 as given in Table 4. In contrast, the estimates for the breakage rate constant x_3 and the reaction rate constant x_4 stay unchanged.

4 Conclusions

An extended methodology for solving parametric inverse problems for complex population balance models was presented. This article generalises the work of [4] where the parametric distribution was assumed to be unimodal, when in fact, this could potentially be arbitrary. The parametric distribution, thus, is assumed to be multimodal in nature, with weights for each mode. This provides the user with an indication of which modal mean gives the best parametric value. The extension from the unimodal to the multimodal approach required a substantial expansion of the theoretical framework. We use a new objective function which penalises against modes which have too small a weight, and thus preventing the production of potentially redundant modes—however, what constitutes the 'best' objective function is still very much an open question. Applied to a toy example, the methodology shows that the parametric distribution for this case must be bimodal with equal weight on each mode. In a second example, the problem of conflicting information about a parameter in a multivariate population balance model for a granulation process was studied. For the coalescence subprocess it was assumed that two different estimates for the collision rate constant were available from the literature. The methodology discriminated between the conflicting parameter values, so that the 'right' estimate for the collision rate constant was effectively identified. These examples show the potential of the outlined methodology for future model and process development.

Nomenclature

B	quadratic form matrix	$[\eta]$
Cov	covariance of random variables X and Y	$[X\,Y]$
c	parameter standard deviation	$[x]$
d	mode index	-
$\mathbb{E}$	expectation of the random variable X	$[X]$
K	number of model parameters	-
$\widehat{K}_0$	rate constant for coalescence	m^3
$\widehat{k}_{\mathrm{att}}$	rate constant for breakage	$\mathrm{s\,m}^{-5}$
k_{porred}	rate constant for compaction	$\mathrm{s\,m}^{-1}$
k_{reac}	rate constant for reaction	$\mathrm{m\,s}^{-1}$
M	number of modes	-
m	random mode	-
N	number of experimental observations	-
$\mathbb{P}$	probability	-
$\boldsymbol{T}$	vector of variables to optimise over	-
V	covariance matrix	-
Var	variance of a random variable X	$[X^2]$
w	weight of mode	-
X	random variable	$[X]$
$\boldsymbol{x}$	model parameter (treated as random)	$[x]$
Y	random variable	$[Y]$
Z	random variable	$[Z]$

Greek letters

β	parameter of response surface	$[\eta]$
$\boldsymbol{\gamma}$	a mean quantity	$[\gamma]$
η	model response	$[\eta]$
η^{exp}	experimental response	$[\eta^{\mathrm{exp}}]$
μ	model prediction	$[\eta]$
σ	standard deviation of model response	$[\eta]$
σ^{exp}	experimental uncertainty	$[\eta]$
Σ	covariance matrix	$[\Sigma]$
Φ	objective function	$[\Phi]$

Superscripts

exp	experiment

Subscripts

0	mean value
i	counting variable
k	counting variable
l	counting variable
s	counting variable
t	counting variable

Acknowledgements

The authors thank the EPSRC (grants EP/E01724X/1 and EP/E01772X/1) for funding their research. MK and AB acknowledge the support of Churchill College.

A Some expectation and variance identities

Theorem 1. *Let $\boldsymbol{z} \in \mathbb{R}^K$ be a random variable with mean $\boldsymbol{\gamma}$ and $K \times K$ covariance matrix Σ, and let $\boldsymbol{\beta} \in \mathbb{R}^K$ be a non-random vector. Then:*

$$\mathbb{E}\left[\boldsymbol{\beta}^\top \boldsymbol{z}\right] = \boldsymbol{\beta}^\top \boldsymbol{\gamma} \tag{16a}$$

and

$$\mathrm{Var}\left(\boldsymbol{\beta}^\top \boldsymbol{z}\right) = \boldsymbol{\beta}^\top \Sigma \boldsymbol{\beta} \tag{16b}$$

and letting $\boldsymbol{y} \in \mathbb{R}^L$ be a random variable and $\boldsymbol{\alpha} \in \mathbb{R}^L$ be a non-random vector,

$$\mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{z}, \boldsymbol{\alpha}^\top \boldsymbol{y}\right) = \boldsymbol{\beta}^\top \mathrm{Cov}\left(\boldsymbol{z}, \boldsymbol{y}\right) \boldsymbol{\alpha} \tag{16c}$$

where $\mathrm{Cov}\left(\boldsymbol{z}, \boldsymbol{y}\right)$ is the $K \times L$ covariance matrix between $\boldsymbol{y}$ and $\boldsymbol{z}$.

Theorem 2. *Let $X_1, \ldots, X_n$ be scalar random variables. Then:*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + \sum_{i \neq j} \mathrm{Cov}\left(X_i, X_j\right) \quad . \tag{17}$$

Note that when $n = 2$, this simplifies to

$$\mathrm{Var}\left(X_1 + X_2\right) = \mathrm{Var}\left(X_1\right) + \mathrm{Var}\left(X_2\right) + 2\,\mathrm{Cov}\left(X_1, X_2\right) \tag{18}$$

Theorem 3. *Let X and Y be any random variables. Then:*

$$\mathbb{E}\left[\mathbb{E}\left(X|Y\right)\right] = \mathbb{E}\left[X\right] \tag{19}$$

Theorem 4. *Let X, Y and Z be any random variables. Then:*

$$\mathrm{Cov}\left(X,Y\right) = \mathbb{E}\left[\mathrm{Cov}\left(X,Y|Z\right)\right] + \mathrm{Cov}\left(\mathbb{E}\left[X|Z\right],\mathbb{E}\left[Y|Z\right]\right) \quad . \tag{20}$$

Proof.

$$\begin{aligned}
\mathrm{Cov}\left(X,Y\right) &= \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(XY|Z\right)\right] - \mathbb{E}\left[\mathbb{E}\left(X|Z\right)\right]\mathbb{E}\left[\mathbb{E}\left(Y|Z\right)\right] \quad \text{by Theorem 3 ,} \\
&= \mathbb{E}\left[\mathrm{Cov}\left(X,Y|Z\right) + \mathbb{E}\left(X|Z\right)\mathbb{E}\left(Y|Z\right)\right] - \mathbb{E}\left[\mathbb{E}\left(X|Z\right)\right]\mathbb{E}\left[\mathbb{E}\left(Y|Z\right)\right] \\
&= \mathbb{E}\left[\mathrm{Cov}\left(X,Y|Z\right)\right] + \mathrm{Cov}\left(\mathbb{E}\left[X|Z\right],\mathbb{E}\left[Y|Z\right]\right) \qquad (21)
\end{aligned}$$

NB: Note that setting $X = Y$ results in the conditional variance formula:

$$\mathrm{Var}\left(X\right) = \mathbb{E}\left[\mathrm{Var}\left(X|Z\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[X|Z\right]\right) \tag{22}$$

■

B Expectation and variance of quadratic forms

We derive the identities for the expectation and variance of random quadratic forms.

Theorem 5. *Suppose $\boldsymbol{z} \sim \mathcal{N}_K(\boldsymbol{\gamma}, \Sigma)$ and A is some fully deterministic $K \times K$ matrix. Then*

$$\mathbb{E}\left[\boldsymbol{z}^\top A\,\boldsymbol{z}\right] = \mathrm{tr}\left[A\,\Sigma\right] + \boldsymbol{\gamma}^\top A\,\boldsymbol{\gamma} \quad . \tag{23}$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[\sum_{k,l} z_k\, A_{kl}\, z_l\right] &= \sum_{k,l} A_{kl}\,\mathbb{E}\left[z_k z_l\right] \\
&= \sum_{k,l} A_{kl}\,\left[\mathrm{Cov}\left(z_k,z_l\right) + \mathbb{E}\left(z_k\right)\,\mathbb{E}\left(z_l\right)\right] \\
&= \sum_{k,l} A_{kl}\,\Sigma_{kl} + \sum_{k,l} \gamma_k\, A_{kl}\,\gamma_l \\
&= \mathrm{tr}\left[A\,\Sigma\right] + \boldsymbol{\gamma}^\top A\,\boldsymbol{\gamma} \qquad (24)
\end{aligned}$$

where the last equality follows since Σ is symmetric. ■

Theorem 6. *Suppose $\boldsymbol{z} \sim \mathcal{N}_K(\boldsymbol{\gamma}, \Sigma)$ and A is some fully deterministic* ***symmetric*** *$K \times K$ matrix. Then*

$$\mathrm{Var}\left(\boldsymbol{z}^\top A\,\boldsymbol{z}\right) = 2\,\mathrm{tr}\left[A\,\Sigma\,A\,\Sigma\right] + 4\,\boldsymbol{\gamma}^\top A\Sigma\,A\,\boldsymbol{\gamma} \quad . \tag{25}$$

Proof. Define[2] $\boldsymbol{y} := \Sigma^{-\frac{1}{2}}(\boldsymbol{z} - \boldsymbol{\gamma})$. Now $\boldsymbol{z} \sim \mathcal{N}_K(\boldsymbol{\gamma}, \Sigma)$ implies that $\boldsymbol{y} \sim \mathcal{N}_K(\mathbf{0}, I)$ where I is the K dimensional identity matrix. Thus,

$$\begin{aligned}
&\mathrm{Var}\left(\boldsymbol{z}^\top A\, \boldsymbol{z}\right) \\
&= \mathrm{Var}\left((\Sigma^{\frac{1}{2}}\boldsymbol{y} + \boldsymbol{\gamma})^\top A\, (\Sigma^{\frac{1}{2}}\boldsymbol{y} + \boldsymbol{\gamma})\right) \\
&= \mathrm{Var}\left(\boldsymbol{y}^\top \Sigma^{\frac{1}{2}} A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y} + 2\, \boldsymbol{\gamma}^\top A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) \quad \text{since } \boldsymbol{\gamma}^\top A\, \boldsymbol{\gamma} \text{ is constant} \\
&= \mathrm{Var}\left(\boldsymbol{y}^\top \Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) + 4\, \mathrm{Var}\left(\boldsymbol{\gamma}^\top A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) + 4\, \mathrm{Cov}\left(\boldsymbol{\gamma}^\top A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}, \boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) \\
&= \mathrm{Var}\left(\boldsymbol{y}^\top \Sigma^{\frac{1}{2}} A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) + 4\, \boldsymbol{\gamma}^\top A\, \Sigma\, A\, \boldsymbol{\gamma} + 4\, \boldsymbol{\gamma}^\top A\, \Sigma^{\frac{1}{2}}\, \mathrm{Cov}\left(\boldsymbol{y}, \boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) \quad . \qquad (26)
\end{aligned}$$

where liberal use of eqs. (16b) and (16c) has been made. We now need to evaluate the third term in the last line of eq. (26)—the covariance $\mathrm{Cov}\left(\boldsymbol{y}, \boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right)$ is a K dimensional column vector. Thus, the i^{th} component of this vector is computed (first defining $F := \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}$):

$$\begin{aligned}
\mathrm{Cov}\left(y_i, \boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) &= \mathbb{E}\left[y_i\, \boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right] - \underbrace{\mathbb{E}\left[y_i\right]}_{=0}\, \mathbb{E}\left[\boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right] \\
&= \mathbb{E}\left[y_i\, \boldsymbol{y}^\top\, F\, \boldsymbol{y}\right] \\
&= \mathbb{E}\left[y_i \sum_{k,l} y_k\, F_{kl}\, y_l\right] \\
&= \sum_{k,l} F_{kl}\, \underbrace{\mathbb{E}\left[y_i\, y_k\, y_m\right]}_{=0\,,\ \text{unless } i=k=m} \\
&= \sum_{k} F_{kk}\, \underbrace{\mathbb{E}\left[y_k^3\right]}_{=0} \\
&= 0 \quad . \qquad (27)
\end{aligned}$$

Thus the third term of the last line of eq. (26) evaluates to zero. The first term is now looked at:

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{y}^\top \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) &= \mathrm{Var}\left(\boldsymbol{y}^\top\, F\, \boldsymbol{y}\right) \\
&= \sum_{k,l} F_{kl}^2\, \mathrm{Var}\left(y_k\, y_l\right) + \sum_{(r,s)\neq(k,l)} F_{rs}\, F_{kl}\, \mathrm{Cov}\left(y_r\, y_s, y_k\, y_l\right) \qquad (28)
\end{aligned}$$

where the second equality follows from eq. (17). Note that in the second sum group, the only indices which satisfy $(r, s) \neq (k, l)$ and which give a non-zero covariance

[2]Note that Σ is a positive semi-definite matrix. Thus $\Sigma^{\frac{1}{2}}$ denotes the unique matrix whose square is equal to Σ.

$\mathrm{Cov}\left(y_r\, y_s, y_k\, y_l\right)$ is when $(r,s) = (l,k)$ (where $k \neq l$). Thus we now have:

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{y}^\top\, \Sigma^{\frac{1}{2}}\, A \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) &= \sum_{k,l} F_{kl}^2\, \mathrm{Var}\left(y_k\, y_l\right) + \sum_{k \neq l} F_{lk}\, F_{kl}\, \mathrm{Cov}\left(y_l\, y_k, y_k\, y_l\right) \\
&= \sum_{k,l} F_{kl}^2\, \mathrm{Var}\left(y_k\, y_l\right) + \sum_{k \neq l} F_{kl}^2\, \mathrm{Var}\left(y_k\, y_l\right) \\
&= \sum_{k} F_{kk}^2\, \mathrm{Var}\left(y_k^2\right) + 2 \sum_{k \neq l} F_{kl}^2\, \mathrm{Var}\left(y_k\, y_l\right) \qquad (29)
\end{aligned}$$

where the second equality follows since F is symmetric. Note that

$$\mathrm{Var}\left(y_k\, y_l\right) = \begin{cases} 2 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases} \qquad (30)$$

since $y_k^2 \sim \chi_1^2$ and so we continue the calculation from eq. (29) to find that

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{y}^\top\, \Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\, \boldsymbol{y}\right) &= \sum_{k} 2\, F_{kk}^2 + 2 \sum_{k \neq j} F_{kl}^2 \\
&= 2 \sum_{k,l} F_{kl}^2 \\
&= 2 \sum_{k,l} F_{kl}\, F_{lk} \\
&= 2\, \mathrm{tr}\left[F^2\right] \\
&= 2\, \mathrm{tr}\left[\left(\Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\right)\, \left(\Sigma^{\frac{1}{2}}\, A\, \Sigma^{\frac{1}{2}}\right)\right] \\
&= 2\, \mathrm{tr}\left[\Sigma^{\frac{1}{2}}\, A\, \Sigma\, A\, \Sigma^{\frac{1}{2}}\right] \\
&= 2\, \mathrm{tr}\left[A\, \Sigma\, A\, \Sigma\right] \quad . \qquad (31)
\end{aligned}$$

Thus we plug eq. (31) into eq. (26) to find the result

$$\mathrm{Var}\left(\boldsymbol{z}^\top\, A\, \boldsymbol{z}\right) = 2\, \mathrm{tr}\left[A\, \Sigma\, A\, \Sigma\right] + 4\, \boldsymbol{\gamma}^\top\, A\, \Sigma\, A\, \boldsymbol{\gamma} \quad . \qquad (32)$$

■

Theorem 7. *Suppose the response surface is given by* $S := \beta_0 + \boldsymbol{\beta}^\top\, \boldsymbol{x} + \boldsymbol{x}^\top\, B\, \boldsymbol{x}$*, where* $\boldsymbol{x} \sim \mathcal{N}_K(\boldsymbol{x}_0, V)$*. Then the variance of* S *is:*

$$\mathrm{Var}\left(S\right) = \boldsymbol{\beta}^\top\, V\, \boldsymbol{\beta} + 2\, \mathrm{tr}\left[B\, V\, B\, V\right] + 4\, \boldsymbol{x}_0^\top B\, V\, B\, \boldsymbol{x}_0 + 4\, \boldsymbol{x}_0^\top\, B\, V\, \boldsymbol{\beta} \qquad (33)$$

Proof. First we note that the β_0 term in S makes no difference to its variance. Second, notice that B is symmetric. Now, using Theorem 2, we get:

$$\begin{aligned}
\mathrm{Var}\left(S\right) &= \mathrm{Var}\left(\boldsymbol{\beta}^\top\, \boldsymbol{x}\right) + \mathrm{Var}\left(\boldsymbol{x}^\top\, B\, \boldsymbol{x}\right) + 2\, \mathrm{Cov}\left(\boldsymbol{\beta}^\top\, \boldsymbol{x}, \boldsymbol{x}^\top\, B\, \boldsymbol{x}\right) \\
&= \boldsymbol{\beta}^\top\, V\, \boldsymbol{\beta} + 2\, \mathrm{tr}\left[B\, V\, B\, V\right] + 4\, \boldsymbol{x}_0^\top\, B\, V\, B\, \boldsymbol{x}_0 + 2\, \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}^\top B \boldsymbol{x}\right)
\end{aligned} \qquad (34)$$

where the second equality uses Theorems 1 and 6. This is already quite close to eq. (33)—we just need to prove that $\mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}^\top B\, \boldsymbol{x}\right) = 2\, \boldsymbol{x}_0^\top B\, V\, \boldsymbol{\beta}$. Note that

$$\begin{aligned}
&\mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}^\top B\, \boldsymbol{x}\right) \\
&= \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, (\boldsymbol{x}-\boldsymbol{x}_0)^\top B\, (\boldsymbol{x}-\boldsymbol{x}_0) + 2\, \boldsymbol{x}_0^\top B\, \boldsymbol{x} + \boldsymbol{x}_0^\top B\, \boldsymbol{x}_0\right) \\
&= \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, (\boldsymbol{x}-\boldsymbol{x}_0)^\top B\, (\boldsymbol{x}-\boldsymbol{x}_0) + 2\, \boldsymbol{x}_0^\top B\, \boldsymbol{x}\right) \\
&= \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, (\boldsymbol{x}-\boldsymbol{x}_0)^\top B\, (\boldsymbol{x}-\boldsymbol{x}_0)\right) + 2\, \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}_0^\top B\, \boldsymbol{x}\right) \\
&= \mathrm{Cov}\left(\boldsymbol{\beta}^\top (\boldsymbol{x}-\boldsymbol{x}_0), (\boldsymbol{x}-\boldsymbol{x}_0)^\top B (\boldsymbol{x}-\boldsymbol{x}_0)\right) + 2\, \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}_0^\top B \boldsymbol{x}\right) \\
&= \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{y}, \boldsymbol{y}^\top B\, \boldsymbol{y}\right) + 2\, \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}_0^\top B\, \boldsymbol{x}\right) \qquad (35)
\end{aligned}$$

where $\boldsymbol{y} := \boldsymbol{x} - \boldsymbol{x}_0 \sim \mathcal{N}_K(\mathbf{0}, V)$. Continuing the working, we get that:

$$\begin{aligned}
\mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}^\top B\, \boldsymbol{x}\right) &= \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{y}, \boldsymbol{y}^\top B\, \boldsymbol{y}\right) + 2\, \boldsymbol{\beta}^\top \mathrm{Cov}\left(\boldsymbol{x}, \boldsymbol{x}\right)\, B\, \boldsymbol{x}_0 \\
&= \mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{y}, \boldsymbol{y}^\top B\, \boldsymbol{y}\right) + 2\, \boldsymbol{\beta}^\top V\, B\, \boldsymbol{x}_0 \\
&= \sum_{i,j,k} B_{ij}\, \beta_k\, \mathbb{E}[y_i\, y_j\, y_k] + 2\, \boldsymbol{\beta}^\top V\, B\, \boldsymbol{x}_0 \qquad (36)
\end{aligned}$$

where the last equality follows in much the same way as in eq. (27), noting that $\mathbb{E}[\boldsymbol{y}] = \mathbf{0}$. Note further that the term $\mathbb{E}[y_i\, y_j\, y_k]$ is non-zero only if $i = j = k$ due to the factorisability of expectations for independent random variables (V is diagonal). But $\mathbb{E}[y_i^3] = 0$ is implied by y_i being normally distributed with mean 0. Thus $\mathbb{E}[y_i\, y_j\, y_k] = 0$, and

$$\mathrm{Cov}\left(\boldsymbol{\beta}^\top \boldsymbol{x}, \boldsymbol{x}^\top B\, \boldsymbol{x}\right) = 2\, \boldsymbol{\beta}^\top V\, B\, \boldsymbol{x}_0 \qquad (37)$$

as required. ■

References

[1] T. Abberger, P. Jennings, A. Mirlach, S. Prajczer, and W. Pfaller. Application of a population balance model to a perfusion in vitro toxicity system. *Toxicology in Vitro*, 20:1213–1224, 2006. doi:10.1016/j.tiv.2006.03.006.

[2] A. A. Adetayo, J. D. Litster, S. E. Pratsinis, and B. J. Ennis. Population balance modelling of drum granulation of material with wide size distribution. *Powder Technology*, 82:37–49, 1995. doi:10.1016/0032-5910(94)02896-V.

[3] H.-J. Bart, C. Drumm, and M. M. Attarakih. Process intensification with reactive extraction columns. *Chemical Engineering and Processing: Process Intensification*, 47:745–754, 2008. doi:10.1016/j.cep.2007.11.005.

[4] A. Braumann and M. Kraft. Incorporating experimental uncertainties into multivariate granulation modelling. *Chemical Engineering Science*, 2009. doi:10.1016/j.ces.2009.09.063. in press.

[5] A. Braumann, M. J. Goodson, M. Kraft, and P. R. Mort. Modelling and validation of granulation with heterogeneous binder dispersion and chemical reaction. *Chemical Engineering Science*, 62:4717–4728, 2007. doi:10.1016/j.ces.2007.05.028.

[6] A. Braumann, M. Kraft, and P. R. Mort. Parameter estimation in a multidimensional granulation model. *Powder Technology*, 2009. doi:10.1016/j.powtec.2009.09.014. in press.

[7] A. Braumann, P. L. W. Man, and M. Kraft. Statistical approximation of the inverse problem in multivariate population balance modelling, 2009. Technical Report 78, c4e-Preprint Series. Submitted to Industrial & Engineering Chemistry Research for publication.

[8] J. Bridgwater, S. Forrest, and D. J. Parker. PEPT for agglomeration. *Powder Technology*, 140:187–193, 2004. doi:10.1016/j.powtec.2004.01.009.

[9] I. T. Cameron, F. Y. Wang, C. D. Immanuel, and F. Stepanek. Process systems modelling and applications in granulation: A review. *Chemical Engineering Science*, 60:3723–3750, 2005. doi:10.1016/j.ces.2005.02.004.

[10] M. S. Celnik, M. Sander, A. Raj, R. H. West, and M. Kraft. Modelling soot formation in a premixed flame using an aromatic-site soot model and an improved oxidation rate. *Proceedings of the Combustion Institute*, (32):639–646, 2009. doi:10.1016/j.proci.2008.06.062.

[11] A. Darelius, A. Rasmuson, I. N. Björn, and S. Folestad. High shear wet granulation modelling—a mechanistic approach using population balances. *Powder Technology*, 160:209–218, 2005. doi:10.1016/j.powtec.2005.08.036.

[12] J. Drechsler, M. Peglow, S. Heinrich, M. Ihlow, and L. Mörl. Investigating the dynamic behaviour of fluidized bed spray granulation processes applying numerical simulation tools. *Chemical Engineering Science*, 60:3817–3833, 2005. doi:10.1016/j.ces.2005.02.010.

[13] M. Frenklach, H. Wang, and M. J. Rabinowitz. Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method—combustion of methane. *Progress in Energy and Combustion Science*, 18:47–73, 1992. doi:10.1016/0360-1285(92)90032-V.

[14] M. Heine and S. E. Pratsinis. Polydispersity of primary particles in agglomerates made by coagulation and sintering. *Journal of Aerosol Science*, (38):17–38, 2007. doi:10.1016/j.jaerosci.2006.09.005.

[15] M. Kind and U. Nieken. On the dynamic simulation of mass crystallization with fines removal. *Chemical Engineering and Processing*, 34:323–328, 1995. doi:10.1016/0255-2701(94)04020-6.

[16] P. L. W. Man, M. Kraft, and J. R. Norris. Coupling algorithms for calculating sensitivities of population balances. In T. E. Simos, G. Psihoyios, and C. Tsitouras, editors, *Numerical Analysis and Applied Mathematics*, volume 1048 of *AIP Conference Proceedings*, pages 927–930, 2008.

[17] N. Morgan, M. Kraft, M. Balthasar, D. Wong, M. Frenklach, and P. Mitchell. Numerical simulations of soot aggregation in premixed laminar flames. *Proceedings of the Combustion Institute*, 31:693–700, 2007. doi:10.1016/j.proci.2006.08.021.

[18] N. M. Morgan, C. G. Wells, M. J. Goodson, M. Kraft, and W. Wagner. A new numerical approach for the simulation of the growth of inorganic nanoparticles. *Journal of Computational Physics*, 211:638–658, 2006. doi:10.1016/j.jcp.2005.04.027.

[19] H. Mühlenweg, A. Gutsch, A. Schild, and S. E. Pratsinis. Process simulation of gas-to-particle-synthesis via population balances: Investigation of three models. *Chemical Engineering Science*, 57:2305–2322, 2002. doi:10.1016/S0009-2509(02)00119-7.

[20] T. Müller, J. Schulze-Horsel, Y. Sidorenko, U. Reichl, and A. Kienle. Population balance modelling of influenza virus replication in MDCK cells during vaccine production. *Computer Aided Chemical Engineering*, 25:133–138, 2008. doi:10.1016/S1570-7946(08)80027-2.

[21] A. M. Nilpawar, G. K. Reynolds, A. D. Salman, and M. J. Hounslow. Surface velocity measurement in a high shear mixer. *Chemical Engineering Science*, 61: 4172–4178, 2006. doi:10.1016/j.ces.2005.10.018.

[22] J. M.-H. Poon, C. D. Immanuel, F. J. Doyle III, and J. D. Litster. A three-dimensional population balance model of granulation with a mechanistic representation of the nucleation and aggregation phenomena. *Chemical Engineering Science*, 63:1315–1329, 2008. doi:10.1016/j.ces.2007.07.048.

[23] D. Ramkrishna and A. W. Mahoney. Population balance modeling. Promise for the future. *Chemical Engineering Science*, 57:595–606, 2002. doi:10.1016/S0009-2509(01)00386-4.

[24] M. Sander, R. H. West, M. S. Celnik, and M. Kraft. A detailed model for the sintering of polydispersed nanoparticle agglomerates. *Aerosol Science and Technology*, 43: 978–989, 2009. doi:10.1080/02786820903092416.

[25] D. A. Sheen, X. You, H. Wang, and T. Løvås. Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion. *Proceedings of the Combustion Institute*, 32(1):535–542, 2009. doi:10.1016/j.proci.2008.05.042.

[26] T. Simmons, R. Turton, and P. Mort. An investigation into the effects of time and shear rate on the spreading of liquids in coating and granulation processes. In *Fifth World Congress on Particle Technology*, 2006.

[27] J. Singh, M. Balthasar, M. Kraft, and W. Wagner. Stochastic modelling of soot particle size and age distribution in laminar premixed flames. *Proceedings of the Combustion Institute*, 30:1457–1465, 2005. doi:10.1016/j.proci.2004.08.120.

[28] J. Singh, R. I. A. Patterson, M. Kraft, and H. Wang. Numerical simulation and sensitivity analysis of detailed soot particle size distribution in laminar premixed ethylene flames. *Combustion and Flame*, 145:117–127, 2006. doi:10.1016/j.combustflame.2005.11.003.

[29] H. S. Tan, M. J. V. Goldschmidt, R. Boerefijn, M. J. Hounslow, A. D. Salman, and J. A. M. Kuipers. Building population balance model for fluidized bed melt granulation: lessons from kinetic theory of granular flow. *Powder Technology*, 142: 103–109, 2004. doi:10.1016/j.powtec.2004.04.030.

[30] A. Vikhansky and M. Kraft. Modelling of a RDC using a combined CFD-population balance approach. *Chemical Engineering Science*, 59(13):2597–2606, 2004. doi:10.1016/j.ces.2004.02.016.

[31] A. Vikhansky and M. Kraft. A Monte Carlo methods for identification and sensitivity analysis of coagulation processes. *Journal of Computational Physics*, 200:50–59, 2004. doi:10.1016/j.jcp.2004.03.006.

[32] A. Vikhansky, M. Kraft, M. Simon, S. Schmidt, and H.-J. Bart. Droplets population balance in a rotating disc contactor: An inverse problem approach. *AIChE Journal*, 52(4):1441–1450, 2006. doi:10.1002/aic.10735.

[33] R. H. West, M. S. Celnik, O. R. Inderwildi, M. Kraft, G. J. O. Beran, and W. H. Green. Towards a comprehensive model of the synthesis of TiO_2 particles from $TiCl_4$. *Industrial and Engineering Chemistry Research*, 46(19):6147–6156, 2007. doi:10.1021/ie0706414.

[34] E. J. W. Wynn and M. J. Hounslow. Integral population balance equations for growth. *Chemical Engineering Science*, 52:733–746, 1997. doi:10.1016/S0009-2509(96)00448-4.

[35] G. Zamponi, J. Stichlmaier, A. Gerstlauer, and E. D. Gilles. Simulation of the transient behaviour of a stirred liquid/liquid extraction column. *Computers & Chemical Engineering*, 20(Supplement 2):S963–S968, 1996. doi:10.1016/0098-1354(96)00168-8.