Chemical Engineering Science 65 (2010) 4038–4045

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Resolving conflicting parameter estimates in multivariate population balance models

Peter L.W. Man, Andreas Braumann, Markus Kraft *

Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK

ARTICLE INFO

Article history:
Received 28 October 2009
Accepted 25 March 2010
Available online 30 March 2010

Keywords:
Granulation
Multimodal parameter distribution
Parameter identification
Particulate processes
Population balance
Error propagation

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.

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 (Wynn and Hounslow, 1997; Kind and Nieken, 1995), biological systems (Abberger et al., 2006; Müller et al., 2008), liquid–liquid extraction (Zamponi et al., 1996; Vikhansky and Kraft, 2004a; Bart et al., 2008), combustion (Singh et al., 2005, 2006; Morgan et al., 2007; Celnik et al., 2009), nanoparticle synthesis (Mühlenweg et al., 2002; Morgan et al., 2006; West et al., 2007; Heine and Pratsinis, 2007; Sander et al., 2009), and granulation (Cameron et al., 2005; Poon et al., 2008). The relevance of the latter is reflected in the variety of equipment in which granulation processes are performed, for example, fluidised beds (Tan et al., 2004; Drechsler et al., 2005), drum granulators (Adetayo et al., 1995) and high shear mixers (Darelius et al., 2005; Braumann et al., 2007). 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 (Ramkrishna and Mahoney, 2002). 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 (Vikhansky and Kraft, 2004b; Vikhansky et al., 2006; Man et al., 2010). 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(Frenklach et al., 1992). 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 by Sheen et al. (2009) and Braumann and Kraft (2010), 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

* Corresponding author. Tel.: +44 1223 762784; fax: +44 1223 334796.
E-mail address: mk306@cam.ac.uk (M. Kraft).

0009-2509/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2010.03.042

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

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

or in multivariate language,

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

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

The mathematical model to estimate the '*true*' datum $\eta_i^{\exp}$ 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 $\eta_i^{\exp}$ 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 Braumann and Kraft (2010) in order to approximate $\eta_i(\boldsymbol{x})$ locally by a second order response surface (dropping the i indices for convenience):

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

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:

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

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 $\boldsymbol{\eta}^{\exp}$ are uncertain (Eq. (1b)). Any parameter estimate based purely on the observed value $\boldsymbol{\eta}_0^{\exp}$ is potentially highly sensitive to this value. In fact, the uncertainty in $\boldsymbol{\eta}^{\exp}$ induces uncertainties in the parameter values $\boldsymbol{x}$. Thus the approach taken in Braumann and Kraft (2010) 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 Braumann and Kraft (2010), $\boldsymbol{x}$ 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

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

$$\mathbb{P}(m = d) = w(d). \tag{4b}$$

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

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

$$\mathbb{P}(m_k = d_k) = w_k(d_k). \tag{5b}$$

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:

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

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