

A Bayesian Approach to Pharmaceutical Product Quality Risk Quantification

Linas MOCKUS¹, José Miguel LAÍNEZ¹, Gintaras REKLAITIS¹,
Lee KIRSCH²

¹*School of Chemical Engineering, Purdue University
480 Stadium Mall Drive, West Lafayette, IN 47907-2100, USA*

²*Division of Pharmaceutics, University of Iowa
Iowa City, IA 52242, USA*

e-mail: lmockus@purdue.edu, jlainez@purdue.edu, reklaiti@purdue.edu, lee-kirsch@uiowa.edu

Received: July 2011; accepted: September 2011

Abstract. The FDA's Quality by Design initiative and associated design space construct (ICH, 2009), have stimulated the use of quantitative methods, mathematical and statistical models, and designed experiments in the process of drug development and manufacture. For a given drug product, the design space may be interpreted as the constrained region of the manufacturing operating variable space within which assurance can be provided that drug product quality specifications will be met. It is now understood, at least conceptually, that this assurance is not deterministic, rather it must be stated in probabilistic terms. In this paper, we report on the use of Bayesian methods to develop a suitable risk metric based on both mathematical and statistical models of the manufacturing processes and product properties. The Bayesian estimation is carried out to determine the joint posterior distribution of the probability of the product meeting quality specifications. The computations are executed using a novel Variational Bayes approximation. In this paper the direct computational approach using this approximation is compared to the widely used but computationally very intensive Markov Chain Monte Carlo method. The approach is illustrated using experimental data and models drawn from a recent QbD study on the drug gabapentin in which the authors were participants.

Keywords: Bayesian inference, risk metrics, Markov Chain Monte Carlo, Variational Bayes, pharmaceutical manufacturing.

1. Introduction

Bayesian inference is applied in multiple disciplines: econometrics (Chib *et al.*, 2003), ecology (Bolker, 2007), astronomy (Gregory, 2008), social sciences (Jackman, 2004), pharmacokinetics (Lainez *et al.*, 2011), design of experiments (Chaloner and Verdinelli, 1995) including clinical trials (Ding *et al.*, 2008) and toxicokinetic (Bois *et al.*, 1999) studies. The most commonly applied method to address this class of problems is Markov Chain Monte Carlo (MCMC). A key limitation of this family of methods is the high computational burden when model complexity and/or dimensionality grow. In this paper we discuss and demonstrate an alternative approach for Bayesian inference which addresses

these limitations. While the application presented deals with risk quantification in the context of pharmaceutical QbD, the Variational Bayes approach is generally applicable to a wide range of problems including those in the references cited above.

Figure 1 shows the spectrum of methods available for statistical inference. At the low computational expense end of the spectrum, point based approximations such as *maximum likelihood* and *maximum a posteriori* (MAP) estimates are available and offer very tractable procedures in terms of their associated computational burden; however they are not able to directly account for the uncertainty. At the other end of the spectrum are the family of MCMC sampling methods. Generally, this family of sampling methods starts from a proposal of what the posterior distribution might be and systematically draws samples from this proposal distribution to find the true posterior distribution. Note that there is no need for a proposal if the full conditionals are known (i.e., the special case of Gibbs sampling). One of the advantages of the MCMC method is that is simple to implement and, once convergence is achieved, the samples are actually being taken from the posterior distribution. In addition, sampling methods make few assumptions about the posterior distribution which results in better accuracy for those cases in which the posterior distribution takes a complex shape (Minka, 2001). However, the number of samples required to achieve convergence may be large, and thus the computation time may be excessive for complex models. Additionally it is important to point out that although theoretically the convergence of a Markov chain to the true distribution is guaranteed, this does not imply that a *finite* sample from such a chain yields a good approximation. The tools available to diagnose whether a finite chain has reached a stationary regime (i.e., trace plots, cumulative means, and Kolmogorov–Smirnov statistic) cannot guarantee its convergence (Cosma and Evers, 2010; Gilks *et al.*, 1996).

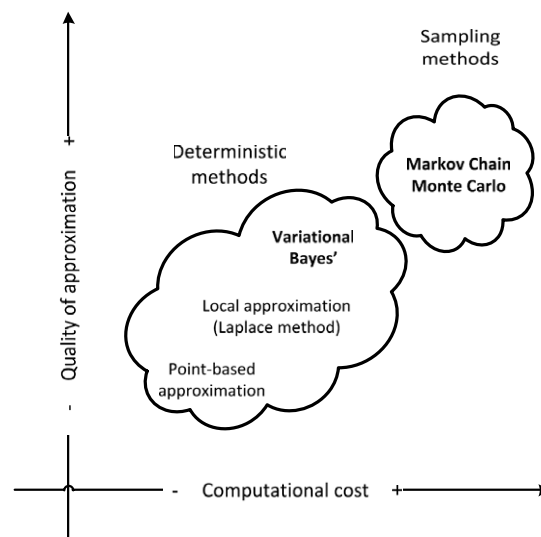


Fig. 1. Schematic of statistical inference approaches, adapted from Smidl and Quinn (2006).

The Variational Bayes (VB) approach presented here lies between these two extremes. It has the potential to offer an attractive trade-off between computational cost and level of accuracy, while also not suffering from convergence issues. However, it too has drawbacks as will be reviewed in Section 4.

2. The Variational Bayes Approach

The idea behind the VB is that the data log-marginal distribution can be decomposed into two elements as shown in (1) to (3). Here, $p(\phi)$, $p(\phi|\mathbf{D})$ and q represent the prior distribution, the true posterior distribution and its approximation, respectively. If q , the *approximate posterior distribution*, is free to be any function, then the maximum lower bound (\mathcal{L}) is obtained when the Kullback–Leibler (KL) divergence is zero. This occurs when q is exactly equal to the posterior distribution $p(\phi | \mathbf{D})$ (Bishop, 2006). Note that minimizing the KL divergence is equivalent to maximizing the lower bound (\mathcal{L}).

$$\ln p(\mathbf{D}) = \mathcal{L}(q) + \text{KL}(q||p), \tag{1}$$

$$\mathcal{L}(q) = - \int q(\phi) \ln \left\{ \frac{q(\phi)}{L(\mathbf{D}|\phi)p(\phi)} \right\} d\phi, \tag{2}$$

$$\text{KL}(q||p) = \int q(\phi) \ln \left\{ \frac{q(\phi)}{p(\phi|\mathbf{D})} \right\} d\phi. \tag{3}$$

If it is assumed that the residuals are normally and independently distributed then the likelihood function ($L(\mathbf{D}|\phi)$) can be expressed as in (4). Equation (5) defines the predictive model for each measurement n of the response variable y as a function of the unknown parameters ϕ and the control variables \mathbf{u} .

$$L(\mathbf{D}|\phi) = \prod_{i \in \{1 \dots N\}} \left\{ \frac{1}{(2\pi)^{1/2}\sigma_n} \exp \left(- \frac{(y_n - \hat{y}_n)^2}{2\sigma_n^2} \right) \right\}, \tag{4}$$

$$\hat{y}_n = \mathbf{g}(\mathbf{u}_n, \phi) \quad \forall n. \tag{5}$$

Then, the VB inference problem can be posed as a general optimization problem of the following form:

$$\max_q \mathcal{L}(q); \quad \text{s.t.} \quad (2), (4) \text{ and } (5). \tag{6}$$

A deterministic approximate inference results by constraining the choices for the function q . In this work, q is assumed to be a probability distribution from the exponential family (e.g., the Gaussian distribution). No assumptions are necessary regarding the shape of the prior distribution. It can be described using any probability function (e.g., a mixture of normal distributions).

Instead of optimizing \mathcal{L} , another approach, Expectation Propagation (EP; Minka, 2001) consists in optimizing a variant of the function defined by (2), which is derived

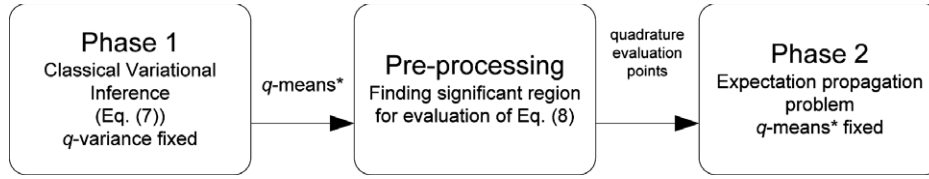


Fig. 2. Proposed sequential variational inference approach.

from the KL divergence between $L(\mathbf{D}|\phi)p(\phi)$ and the function q . The advantage of this approach is that the optimal solution to this problem is given by moment matching. This property can be exploited for the case in which q is assumed to be a fully factorized distribution from the exponential family (Bishop, 2006).

$$\int L(\mathbf{D}|\phi)p(\phi) \ln q(\phi) d\phi. \quad (7)$$

We propose the sequential approach, shown in Fig. 2. First, the optimization problem (6) is solved by approximating the integral evaluation using a Gauss–Hermite quadrature rule. The q -covariances are fixed to a very small value and the q -means are optimized. The Phase 2 problem results from applying the EP approach. The q -means resulting from Phase 1 allow us to find the region within which the values of the integrand are significant and thus are to be included in Phase 2. This is done in the “pre-processing” step. In Phase 2 the integral evaluation is approximated using a Gauss–Legendre quadrature rule. The q -means are given as parameters to the Phase 2 problem and the q -covariances are optimized. By following this decomposition scheme the complexity and computational load of the optimization problem is considerably reduced.

2.1. Phase 1: Classical Variational Inference

As aforementioned, in order to compute the integrals defining in the optimization problem, we follow a quadrature rule approach. Generally, posterior distributions have low density areas what causes the quadrature points to be located in low significance regions. This fact requires an increase in the number of quadrature points required to obtain a good interpolation. For this Phase 1, we utilize a Hermite–Gauss quadrature rule which is suitable for the numerical integration of probability distributions from the exponential family. In addition, one could leave the quadrature points as variables as a function of the unknown parameters (θ) of the parametric approximate distribution q in order to reduce the effect of the true posterior distribution sparsity.

$$\mathcal{L}(\theta) \approx \sum_{\phi} q(\phi|\theta) \ln \{L(\mathbf{D}|\phi)p(\phi)\} - \sum_{\phi} q(\phi|\theta) \ln q(\phi|\theta). \quad (8)$$

Notice that the values of q are fixed given the quadrature points. Subsequently, the maximization of the approximate value of \mathcal{L} is equivalent to determining the *maximum a posteriori* parameters. One of the challenges of this phase is to guarantee that the global optimal has been found when the predictive model (5) is non-linear.

2.2. The Pre-Processing Stage

This stage is intended to establish a region within which the significant values of the posterior distribution lie. This is done by taking into account that the *maximum a posteriori* point (ϕ^*) and its corresponding non-normalized posterior density ($L(\mathbf{D}|\phi^*)p(\phi^*)$) are known from the previous step. It is assumed that the *significant* region is located around this point. A threshold (τ) for the log-non-normalized posterior density is defined in (9).

$$\tau = \ln (L(\mathbf{D}|\phi^*)p(\phi^*)) - k. \tag{9}$$

For the sake of tractability, this threshold is used to determine the *significant* range of each parameter. The end points of each interval correspond to the value of the studied parameter that gives a log-non-normalized posterior density ($\ln L(\mathbf{D}|\phi^*)p(\phi^*)$) equal to the threshold (τ) while the other parameters are fixed in the *maximum a posteriori* value. Figure 3 shows this procedure graphically. If we assume that a Gaussian distribution is a good shape approximation to the true posterior distribution, then this delimitation works well for those cases in which there is no correlation among the parameters (Fig. 3a). Nevertheless, the delimitation may be poor when correlation among parameters exists as depicted in Fig. 3b.

In order to deal with correlation, a second step is included which exploits the relationship between the spectral decomposition of a covariance matrix and the estimate of its eigenvectors. This will allow us to delimit the significant region approximately along the principal axes of the true posterior distribution. An estimate of the eigenvectors (λ) is obtained from a covariance matrix computed from a sample taken from the region defined in the previous step (Fig. 4a). Then, the threshold is used to determine a *significant* range for each variable in this new rotated system as shown in Fig. 4b.

The quadrature points for the Phase 2 are determined in the new system of coordinates (ϕ') and within its corresponding significant region and afterwards transformed to the original system.

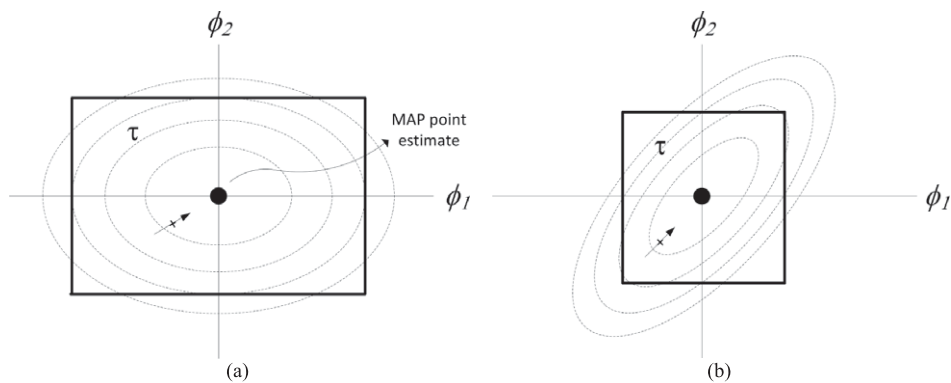


Fig. 3. The significant region definition in the original system of coordinates.

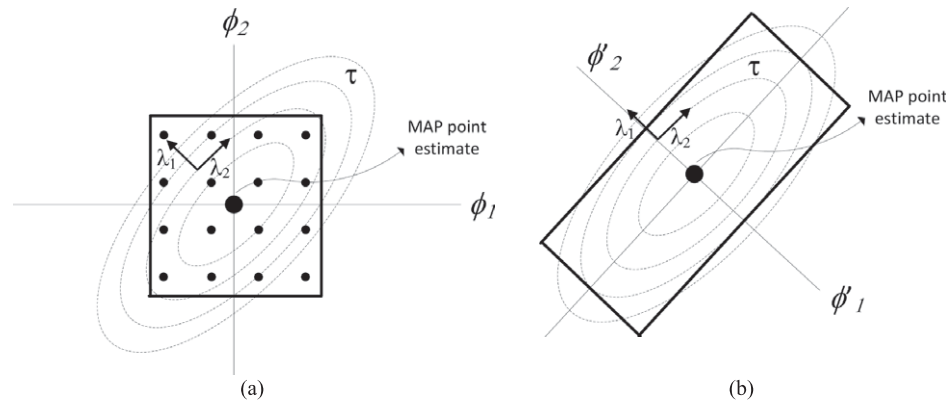


Fig. 4. The significant region definition in the principal axes of the posterior distribution.

2.3. Phase 2: The Expectation Propagation Approach

In the Phase 1 an estimate of the means of the approximate distribution q have been obtained. The purpose of the Phase 2 is to estimate the variance-covariance matrix of q by maximizing the objective function given in (7). The integral will be computed using a quadrature rule whose quadrature points will be given by the results obtained from the pre-processing stage. The processing stage provides a significant region so that quadrature points are not wasted and a good approximation is achieved with the minimum number of points. Notice that the corners of the significant region shown in Fig. 4b do not contribute enough to the integral value. Therefore, the computational burden in the optimization can be saved by pruning those quadrature points whose non-normalized posterior distribution value is lower than τ .

Finally, we would like to emphasize that Variational Inference methods try to approximate the posterior using a given shape and their performance depend on how well this assumption holds. On the other hand, the computational time will depend highly of how fast the optimization problem formulated in the Phase 1 can be solved to global optimality.

2.4. An Illustrative Example

In order to illustrate this methodology, we will use the nonlinear regression model for the temporal evolution of the dry weight of onion bulbs data, Example 7.3 in Gamerman and Lopes (2006). The predictive model is:

$$\hat{y}_n = \phi_1 - \phi_2 \phi_3^x. \quad (10)$$

It is assumed that error is homoscedastic and normally distributed (11).

$$\varepsilon \sim \mathcal{N}(0, \phi_4^2). \quad (11)$$

The prior distributions for the model parameters $(\phi_1, \phi_2, \phi_3, \phi_4)$ and the error variance are described in (12).

$$\begin{aligned}
 p(\phi_1) &= \mathcal{U}(0, 20.0), \\
 p(\phi_2) &= \mathcal{U}(0, 20.0), \\
 p(\phi_3) &= \mathcal{U}(0, 1), \\
 p(\phi_4) &= \mathcal{U}(0, 5.0 \times 10^2), \\
 p(\phi) &= \prod_i p(\phi_i).
 \end{aligned}
 \tag{12}$$

Stage I. Next, the optimization problem for Stage I is formulated for the illustrative example. The log-likelihood and the prior density given a set of parameters ϕ are presented in (13) and (14).

$$\ln L(\mathbf{D}|\phi) = -\frac{1}{2} \left[27 \ln(2\pi) + 54 \ln(\phi_4) + \frac{1}{\phi_4^2} \sum_{n=1}^{27} (y_n - \phi_1 + \phi_2 \phi_3^{x_n})^2 \right], \tag{13}$$

$$\ln p(\phi) = -12.21. \tag{14}$$

Then, the *maximum a posteriori* parameters can be computed by using (15). The optimum parameters obtained using GAMS-CONOPT3 are $\phi_1^* = 2.66$, $\phi_2^* = 0.96$, $\phi_3^* = 0.87$, and $\phi_4^* = 0.09$ which result in an objective function value equal to 14.56.

$$\arg \max_{\phi} \{ \ln L(\mathbf{D}|\phi) + \ln p(\phi) \}. \tag{15}$$

The *pre-processing step*. Following (9), we defined k equal to -275 which results in a threshold (τ) equal to -260.44 . The *significant* range for each parameter is shown in Table 1. Henceforth, we will work with the log-scale values of the parameters. This is common practice in MCMC methods in order to streamline the sampling procedure when positive values for the raw value parameters are required. Now, an estimated covariance is computed using a few points from the significant region. We will use the solutions of a 5 point-Legendre polynomial (V) for each

Table 1
Significant range for each parameter $\phi([a, b])$

Parameter	a	b
$\ln(\phi_1)$	0.81	1.12
$\ln(\phi_2)$	-4.19	0.65
$\ln(\phi_3)$	-2.94	-0.03
$\ln(\phi_4)$	-4.01	8.28

parameter evaluation. Then, the evaluation points for the model are given from all the possible combinations of the parameters evaluation points ($V_1 \otimes V_2 \otimes V_3 \otimes V_4$). A weighted covariance is estimated using as weight for each evaluation point the corresponding value of $L(\mathbf{D}|\phi)p(\phi)$. The covariance matrix (Σ) can be factorized using an spectral decomposition as shown in (16).

$$\Sigma = \Phi \Lambda^{-1} \Phi^{-1}, \quad (16)$$

where Φ is the square matrix whose columns are the covariance eigenvectors and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues of Σ . The Φ matrix obtained following this procedure is shown in Table 2. These eigenvectors are used to rotate the original system of coordinates by using (17). The significant range corresponding to this rotated system of coordinates is given in Table 3.

$$\phi' = \Phi^T(\phi - \phi^*). \quad (17)$$

A 6 point trapezoidal rule will be used to determine the points where the integral included in Stage II is evaluated. The trapezoidal rule is defined within the region established by the ranges presented in Table 3. Let us designate such points (vectors) as \mathbf{f}_k . Then the value of $\ln L(\mathbf{D}|\mathbf{f}_k) + \ln p(\mathbf{f}_k)$ can be directly computed.

Stage II. Let us defined the approximate posterior q as

$$q(\ln \phi) = \mathcal{N}(\ln \bar{\phi}, \Sigma^q). \quad (18)$$

Table 2
Matrix Φ for the illustrative example

$\ln(\phi_1)$	-0.051	-0.149	0.411	0.898
$\ln(\phi_2)$	-0.189	-0.969	-0.058	-0.146
$\ln(\phi_3)$	0.055	-0.003	0.909	-0.414
$\ln(\phi_4)$	-0.979	0.195	0.040	-0.042

Table 3
Significant range for each parameter ϕ' in the new rotated system of coordinates ($[c, d]$)

Parameter	c	d
$\ln(\phi'_1)$	-0.70	0.32
$\ln(\phi'_2)$	-0.31	0.43
$\ln(\phi'_3)$	-0.23	0.07
$\ln(\phi'_4)$	-0.03	0.03

Table 4
Optimal covariance matrix Σ^q for the illustrative example

5.819E-04	6.723E-04	7.993E-04	-8.331E-04
6.723E-04	6.073E-03	-1.412E-04	-5.746E-04
7.993E-04	-1.412E-04	1.531E-03	-1.908E-03
-8.331E-04	-5.746E-04	-1.908E-03	2.596E-02

We already know from (16) that

$$\Sigma^q = \Phi^q(\Lambda^q)^{-1}(\Phi^q)^{-1}. \tag{19}$$

Equation (20) constraints the eigenvectors to be an orthonormal set so that Σ^q results in a real and symmetric matrix. In this equation, \mathbf{I} represents the identity matrix.

$$\Phi^q(\Phi^q)^T = \mathbf{I}. \tag{20}$$

Similarly to (17), we can define \mathbf{Z} as a new set of variables in (21) which will be helpful in the evaluation of the approximate posterior distribution q .

$$\mathbf{Z}_k = (\Phi^q)^T(\mathbf{f}_k - \ln \bar{\phi}) \quad \forall k. \tag{21}$$

Then, the density of the approximate posterior distribution corresponding to each evaluation point k can be computed as follows:

$$\ln q(\mathbf{f}_k) = - \sum_{j=1}^4 \left[\frac{1}{2} \ln(2\pi\Lambda_{jj}) + \frac{Z_{jk}^2}{2\Lambda_{jj}} \right] \quad \forall k. \tag{22}$$

The optimization problem related to Stage II can be posed as defined in (23). The integral expressed in (7) can then be approximated by the objective function of problem (23). Here, w_k is the weight associated with the evaluation point k according to the trapezoidal rule.

$$\max_{\ln \bar{\phi}, \Sigma^q} \sum_k w_k L(\mathbf{D}|\mathbf{f}_k)p(\mathbf{f}_k) \ln q(\mathbf{f}_k); \quad \text{s.t. Eqs. (19)–(22)}. \tag{23}$$

The optimal values for the mean vector are $\ln(\bar{\phi}_1) = 0.98$, $\ln(\bar{\phi}_2) = -0.04$, $\ln(\bar{\phi}_3) = -0.14$, and $\ln(\bar{\phi}_4) = -2.41$. The optimal covariance matrix Σ^q is given in Table 4.

In order to demonstrate the accuracy of the sequential Variational Bayes approach proposed in this work, the estimation of the posterior distribution using an MCMC approach has been also carried out. The comparison between the marginal probability distributions of each parameter is depicted in Fig. 5.

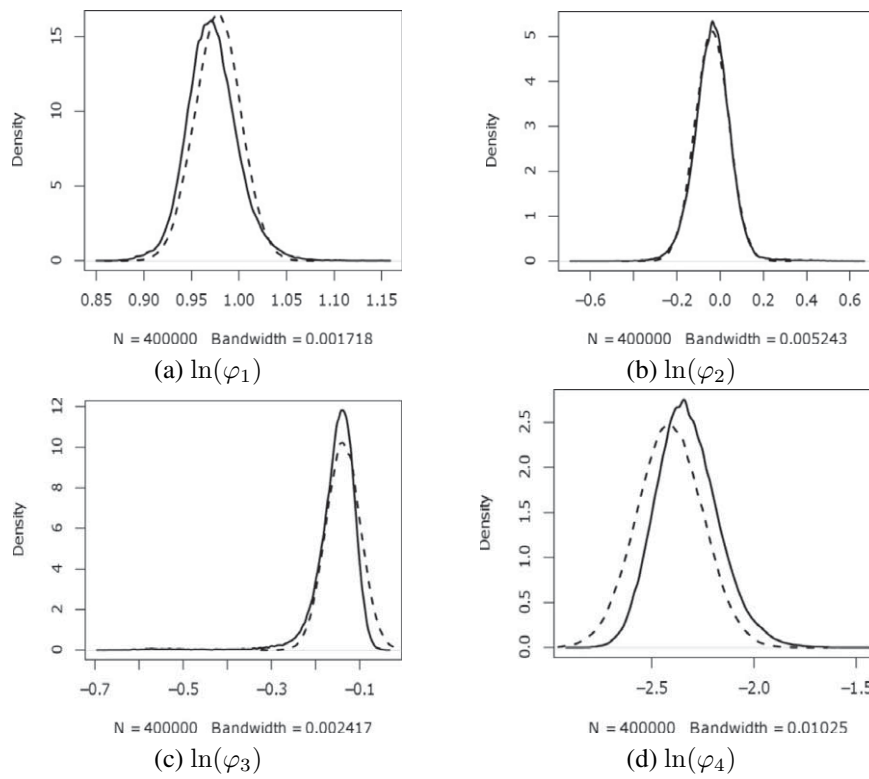


Fig. 5. Marginal probability distributions for the parameters of illustrative example (MCMC: continuous line; variational Bayes: dashed line).

3. QbD Case Study

As stipulated by ICH Q8 R2 (ICH, 2009), the prediction of critical process parameters based on process modeling is a part of an enhanced, quality by design approach to product development. It defines the design space as the “multidimensional combination and interaction of input variables (e.g., material variables) and process parameters that have been demonstrated to provide assurance of quality”. A design space is thus a model that predicts the level of quality to be expected at any feasible combination of input variable values within the multidimensional space. Feasibility is defined in terms of attainment of acceptable level of quality. In the context of pharmaceutical products, quality can be taken to include aspects of safety (e.g., toxicity), efficacy, potency, and other product characteristics such as tablet hardness.

The process to identify critical quality variables to be used in the construction of design space is quite well established on the basis of the ICH guidances as well as recent discussions at industry conferences (Moore, 2007). However, the best procedure for constructing a design space from data composed of such variables is still not well defined. Typically, a set of designed experiments for a specific process operation, which spans the range of the critical quality variables, is developed and executed to generate values of the

critical quality attributes (CQAs). This data is then used to construct a response surface for that unit operation. The sets of response surfaces from the successive operations constituting the entire production train are then combined to generate the combined response and then the subset of variables that yield desired quality levels define the associated process design space.

However, response surfaces determined using traditional statistical tools typically represent a “mean” response surface and do not quantify the assurance (probability) that product critical quality attributes (CQAs) will be met. ICH Q9 on Quality Risk Management defines risk as “the probability of occurrence of harm” and recommends that risks to quality be assessed. Clearly, in order to provide assurance of quality, it is important for a design space model to predict the probability of safety, efficacy, and other critical attributes as well as their levels. Moreover, traditional design space boundaries based on single-point estimates provide insufficient information to a formulation scientist because they are not probabilistic. That is, they do not consider the uncertainties present in data or estimated parameters, do not take into account correlations among multiple responses, and convey no information about the risk of exceeding the design space boundary. Bayesian treatments, on the other hand, can readily incorporate all these considerations. In addition, the Bayesian paradigm provides an effective algorithm for incorporating prior knowledge from theory or experience and naturally leads to a structured process of knowledge building. Bayesian approaches thus fully support the spirit of FDA’s goal of reducing “uncertainty about product performance throughout the product life cycle through scientific research” (Woodcock, 2009).

Bayesian approaches using linear models have been applied successfully to solid oral dosage forms (Miró-Quesada *et al.*, 2004; Peterson, 2004; Peterson and Yahyah, 2009; Peterson, 2007, 2008). Application of Bayesian methodology in the pharmaceutical domain, namely prediction of lyophilization cycle duration, may be found in Mockus *et al.* (2011). The present work extends the Bayesian treatment to a multiple unit operation process for a specific solid oral dosage form, and illustrates the use of a nonlinear model for shelf life stability.

3.1. Case Study Description

In order to illustrate benefits provided by the Variational Bayes approach for risk based design space definition, the set of data from a four step drug manufacturing process (Fig. 6) for the production of gabapentin tablets was utilized. We first report the Bayesian results obtained via conventional MCMC calculations and then compare to the results obtained using VB.

Gabapentin is an anticonvulsant that is used for preventing seizures and for treating postherpetic neuralgia. The stability of gabapentin is complicated by the formation of a toxic degradation product, gabapentin-lactam, which must be maintained at sufficiently low levels. The generation of the lactam is dependent on manufacturing conditions, particularly stresses induced by compression or heating. Additionally, lactam will be formed while the finished product is in storage. The rate of degradation does depend on temperature and relative humidity. Furthermore, for the product to be acceptable, the level of

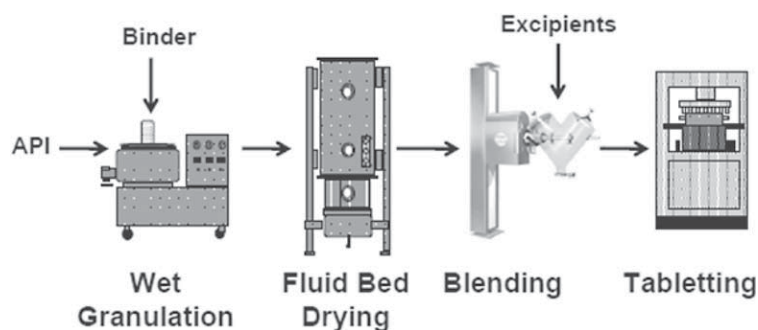


Fig. 6. Diagram of manufacturing process.

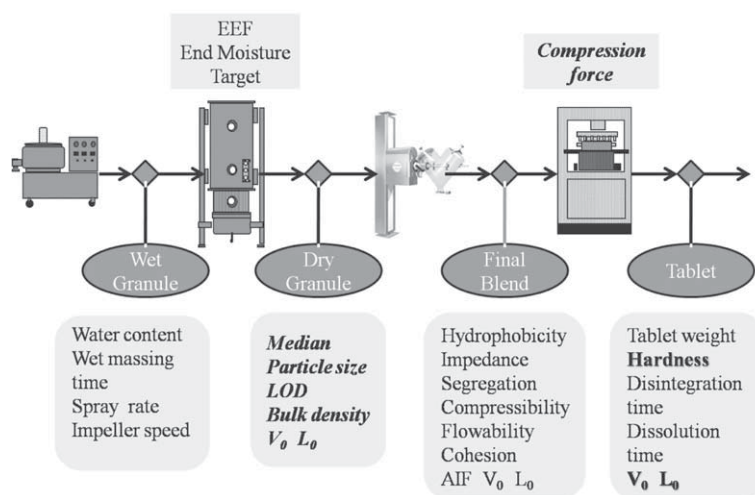


Fig. 7. Key process variables and critical quality attributes.

lactam at the end of the stated product shelf life must be within allowed limit. Prediction of the lactam level at the end of a nominal two year shelf life requires a suitable semi-empirically derived shelf-life or long term degradation model. Such a model is incorporated into the present study.

For the details of the specific experimental plan employed, the reader is referred to a series of papers (Kayrak-Talay *et al.*, 2011; Zacour *et al.*, 2011; Hoag, 2011; Kirsch, 2011). The data is drawn from the experiment conducted with one kilogram scale batches. The key process and quality variables associated with each processing operation and the material produced by that operation are shown in Fig. 7. Not all of the variables shown proved to be significant enough to be used in design space construction: these less significant variables are shown in black. Predictor variables such as level of dryness (LOD) of dry granulate, median particle size of dry granulate, and bulk density of dry granulate were experimentally measured after the completion of fluid bed drying operation (see Fig. 7 in bold italic). The compression force was varied during the tableting operation.

Responses such as tablet hardness, initial level of lactam at the end of manufacture (L_0), and initial rate of lactam formation (V_0) were experimentally measured at the completion of the tableting operation (Zong *et al.*, 2011); see Fig. 7 in bold. The lactam level (L_t) was experimentally measured using replicate stability samples at each of four different time points (initial, one, three, and six months). In order to further simplify the degradation model (Kirsch, 2011), only the data for stability samples maintained at room temperature and 5% relative humidity was used for the purpose of this case. The LOD, median particle size, bulk density, initial lactam level (L_0), and rate of lactam formation (V_0) data was collected for fourteen granulation batches. Lactam level (L_t) data was collected for each of four batches.

It was found that the compression force, median particle size, and bulk density are strong predictors of tablet hardness and initial rate of lactam formation (V_0), while LOD, median particle size, and bulk density are strong predictors of initial lactam level (L_0). Therefore those predictors were used in defining the design space.

3.2. Risk Based Design Space Definition

In this section we develop the design space by using traditional Markov Chain Monte Carlo sampling. This will serve as our baseline. Then we show that for models that are nonlinear in the parameters (see (25)) Variational Bayes provides very similar results with a significant gain in computational time, i.e., baseline is improved. For models that are linear in the parameters (see (24) and (26)) the VB computational benefit is not as significant and therefore is not presented.

3.2.1. Baseline Results Provided by MCMC

We consider risk of failure to be defined in terms of the failure to meet two requirements for the final drug product: tablet hardness and level of degradation. Tablet hardness imposes boundaries on the design space because if the tablet is too soft it may not withstand stresses during handling and transportation. The level of degradation also imposes boundaries on the design space because the degradation product gabapentin lactam is a toxic substance and its level should not exceed a specified threshold after 2 year shelf life under normal storage conditions, defined as room temperature and 5% relative humidity.

The model utilized for this case study combines both an empirical model for tablet hardness and a mechanistic model for degradation.

$$\begin{aligned} \log(\text{TabletHardness}) \\ = a_0 + a_1 * \text{CompressionForce} + a_2 * \text{MedianPS} \\ + a_3 * \text{BulkDensity} + a_4 * \text{MedianPS} * \text{BulkDensity}, \end{aligned} \quad (24)$$

where *CompressionForce* is the mean compression force during tableting, *MedianPS* is the median particle size of dry granulate, and *BulkDensity* is the bulk density of dry granulate. Estimates of parameter posteriors after 10,000 burn-in and 1000 sampling iterations by WinBUGS (Cowles, 2004) are shown in Fig. 8a through Fig. 8f.

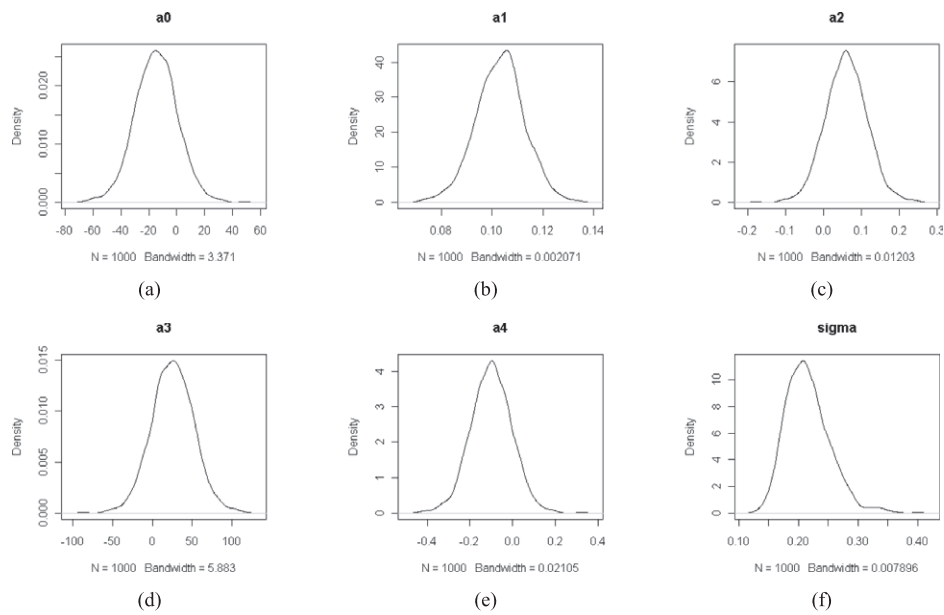
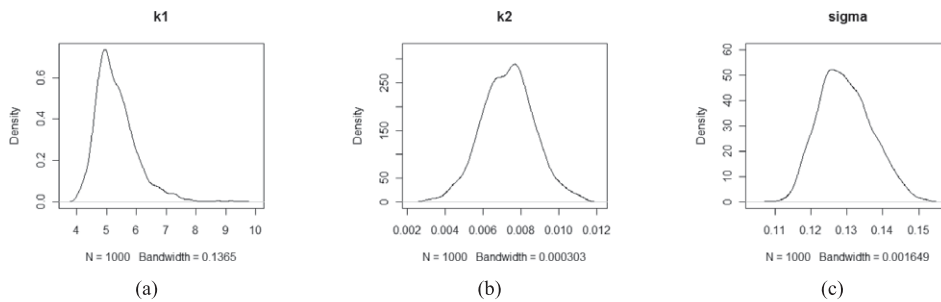


Fig. 8. Parameter posteriors of tablet hardness model.

Fig. 9. Parameter posteriors of L_t model.

The development and parameterization of a simplified version of the general degradation model for gabapentin has been reported (Radaduen *et al.*, 2011):

$$L_t = L_0 + k_1 * V_0 * (1 - e^{-k_2 * t}), \quad (25)$$

where L_t is level of degradant at time t , L_0 is level of degradant at the end of manufacture, and V_0 is experimentally determined rate of degradation at stressed conditions (Zong *et al.*, 2011). Estimates of parameter posteriors after 99,999 burn-in and 1000 sampling iterations by WinBUGS are shown in Fig. 9a through Fig. 9c. Because of a significant autocorrelation, a thin interval of 100 was used.

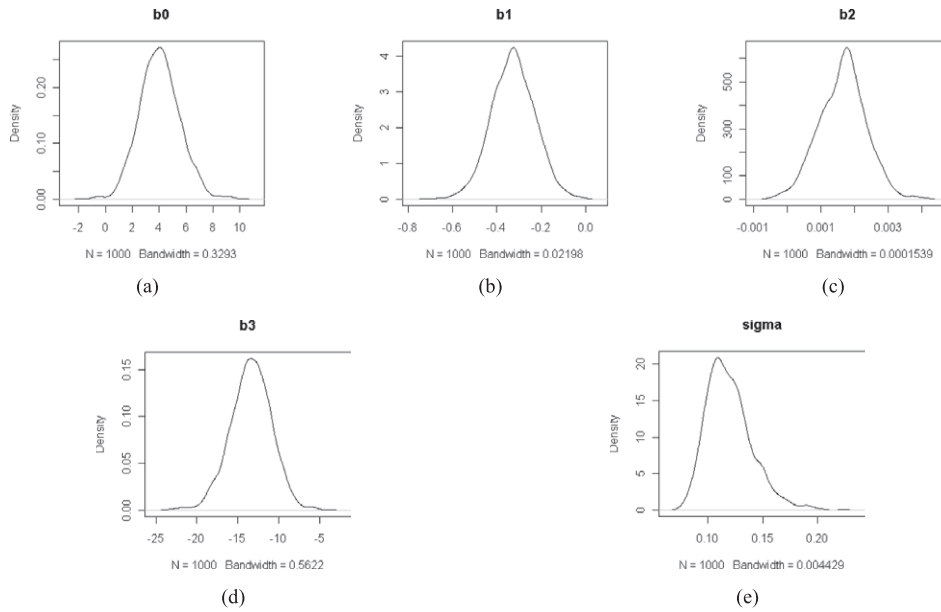


Fig. 10. Parameter posteriors of L_0 model.

The empirical model for L_0 and V_0 was utilized to link manufacturing data to the shelf life stability data.

$$\begin{aligned} \log(L_0) &= b_0 + b_1 * LOD + b_2 * MedianPS + b_3 * BulkDensity, \\ \log(V_0) &= c_0 + c_1 * CompressionForce + c_2 * MedianPS \\ &\quad + c_3 * BulkDensity, \end{aligned} \tag{26}$$

where LOD is level of dryness (LOD) of dry granulate. Estimates of parameter posteriors after 10,000 burn-in and 1000 sampling iterations by WinBUGS are reported in Fig. 10a through Fig. 10e for the L_0 model and in Fig. 11a through Fig. 11e for the V_0 model.

A log normal error model was utilized to address experimental error in the tablet hardness, L_0 , V_0 , and L_t measurements. Such an error model is commonly used for analytical measurements.

$$\log(\text{Measured}) \sim N(\log(\text{Predicted}), \sigma), \tag{27}$$

where *Measured* is the measured value, *Predicted* is the predicted value, and σ is the standard deviation of measurement error.

In this case, failure is defined as the probability that either the target hardness is less than 6 kP or the level of lactam is more than 0.21 mol%. Those requirements for final product quality (acceptance limits) are chosen in order to demonstrate the conflicting nature of the effect of the compression force on tablet hardness and lactam level (both in-

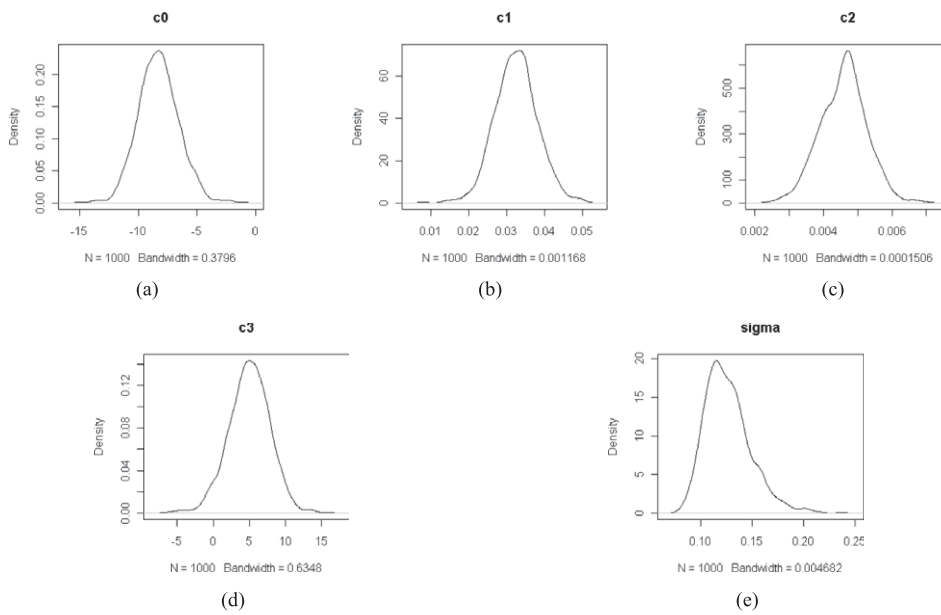


Fig. 11. Parameter posteriors of V_0 model.

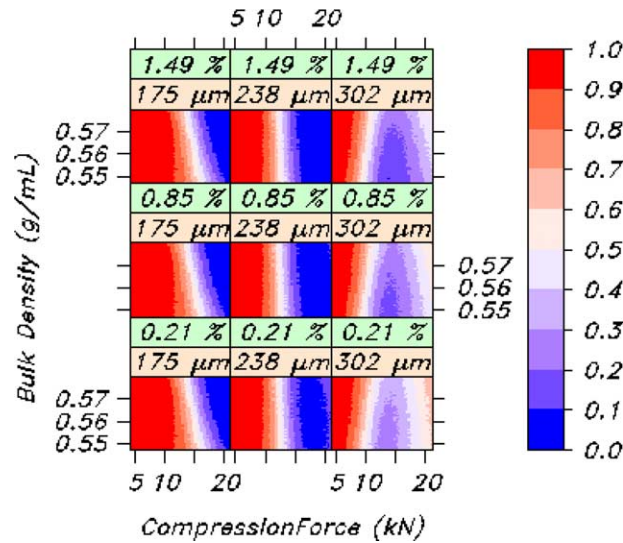


Fig. 12. Risk based design space.

crease with compression force): effectively the objective is to “maximize” tablet hardness and “minimize” lactam level.

The trellis chart (Fig. 12) depicts the risk of failure for various levels of dry granulate LOD (level of dryness), dry granulate median particle size, dry granulate bulk density, and compression force.

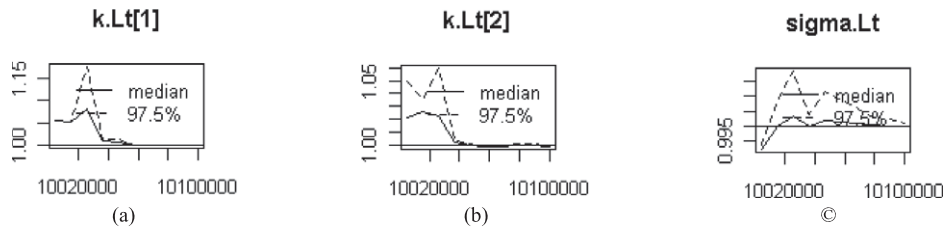


Fig. 13. Evolution of Gelman and Rubin's shrink factor for parameters k_1 , k_2 , and σ .

In Fig. 12, the upper strip represents LOD of dry granulate in %, the lower strip represents median particle size of dry granulate in μm , the x axis represents compression force in kN , and the y axis represents bulk density of dry granulate in g/mL . The bar on the right represents the probability of failure. Figure 9 is a very useful risk assessment tool. It provides a visual display of the probability of exceeding these limits over the various combinations of input variables: LOD, particle size, compression force, and bulk density. Based on this design space representation, we would choose to operate the manufacturing process within the region indicated in blue (low probability of failure). For instance we could require that the process only be used with combinations of input variables for which we predict less than some low probability of failure. The choice of an acceptable probability level limit may be product dependent and include considerations of the severity or cost associated with failure. Once this choice is made, and depending on the particular process under study, a control strategy should be established to assure that the manufacturing process will operate only within the combinations of input variables associated with this probability limit.

The Fig. 13 indicates the convergence of WinBUGS (for a thinning interval of 100) for the L_t model. These plots show the evolution of the Gelman and Rubin's shrink factor (Gelman and Rubin, 1992) as the number of iterations increases. By calculating the shrink factor at several points in time, one can determine whether the shrink factor has really converged. The black dashed line indicates the 97.5% confidence interval while the black solid line indicates the evolution of the median. The plots are generated using two parallel chains.

3.2.2. Computational Improvement of Variational Bayes

The estimates of the parameter distribution posteriors were also determined using the VB methodology. The parameter posterior estimates provided by the Variational Bayes approach were very close to the corresponding estimates provided by WinBUGS (Gibbs sampling). By way of example, in Fig. 14, we superimpose the posteriors obtained from the two methods. The black solid line indicates the estimates of parameter posteriors provided by WinBUGS while the black dashed line indicates estimates provided by Variational Bayes.

The logarithm of likelihood is correspondingly 103,329.1 and 103,334.1: these values indicate that the Variational Bayes provides slightly superior results. However there is a significant difference in the computational effort required. The simulation time using

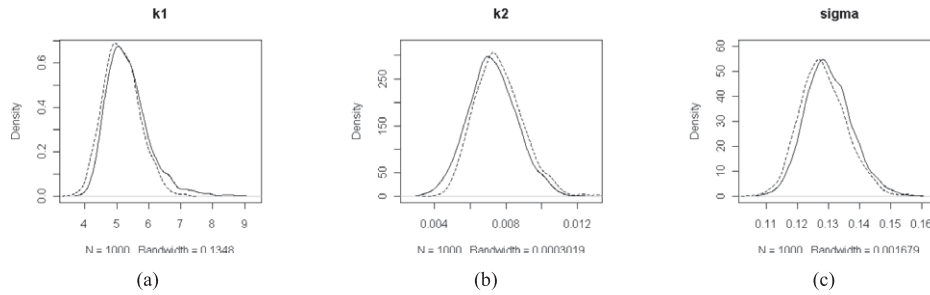


Fig. 14. Parameter posteriors of L_t model provided by MCMC and VB.

WinBUGS on a 64 bit machine under Windows 7 takes around 38 minutes while it is a matter of fraction of a second for Variational Bayes on 32 bit machine running under Windows XP. Both WinBUGS and Variational Bayes were run on an Intel i7 2.66 GHz processor. It is worth noting that by decreasing the thinning interval to 50, the computational time is reduced by one half while the logarithm of likelihood becomes 103,187.4.

4. Discussion

This work has presented the Variational Bayes approach as an alternative to the commonly used MCMC method for computing approximations for the posterior distribution and applied this approach to the risk based design space formulation using the data from a comprehensive experimental study. As noted, the MCMC approach can be quite computational expensive, can present problems of autocorrelation in the sampling as well as possible difficulties in evaluating the convergence to a stationary process (see Section 3.2.1). The Variational Bayes approach can significantly reduce the computational effort required to obtain the posterior approximations. While one ought not generalize from this case alone, our general experience is that the computational advantage of VB vs. MCMC grows with the dimension of the estimation problem, thus, suggesting that it is a promising tool for large dimensionality applications. In addition, the Variational Bayes approach does not suffer from the convergence evaluation problems nor the autocorrelation sampling problem since it is an optimization based-procedure which approximates the posterior distribution to a given parametric distribution. The Variational Bayes problem is then focused on the determination of the parameters defining the approximating distribution q by minimizing its KL divergence with the true posterior distribution. The quality of the results of the VB approach depends on how similar the shape of the approximating distribution q is to the true posterior distribution. Here, it is assumed that q belongs to the family of Gaussian distributions. This assumption works well for cases in which the posterior distribution is unimodal. When approximating multi-modal distributions, q must be replaced by a *mixture* of parametric distributions. It is also important to mention that non-convex predictive models may possess multiple optimal points; in such cases a global optimizer or a meta-heuristic/evolutionary algorithm should be used to solve the

Phase 1 of the proposed sequential VB approach (see Section 2.1). The former solution may significantly reduce the savings in computational effort, while the latter one carries the problem that there might be cases in which meta-heuristics tend to converge towards a local optima. The task of finding multiple optima in reasonable computational times for case specific applications and the consideration of a mixture of parametric distributions for q constitutes part of our ongoing work and current work.

Acknowledgements. We are grateful to the National Institute for Pharmaceutical Technology and Education (NIPTE) and the U.S. Food and Drug Administration (FDA) for providing funds for this research. This study was funded by the FDA-sponsored contract “Development of Quality by Design (QbD) Guidance Elements on Design Specifications across Scales with Stability Considerations” (contract number HHSF223200819929C). We would like to acknowledge Dr. Tzuchi Rob Ju of Abbott Laboratories and his team for allowing the execution of experiments at one kilogram scale on their pilot plant. Dr. David LeBlond of Abbot Laboratories was instrumental in providing critical and insightful comments as well as providing industrial perspective on usefulness of diagrams that define risk based design space (Fig. 6).

References

- Bishop, C. (2006). In *Pattern Recognition and Machine Learning*. Springer, London.
- Bois, F.Y., Smith, T.J., Gelman, A., Chang, H.Y., Smith, A.E. (1999). Optimal design for a study of butadiene toxicokinetics in humans. *Toxicological Sciences*, 49(2), 213–224.
- Bolker, B. (2007). *Ecological Models and Data in R*. Princeton University Press, Princeton.
- Chaloner, K., Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science*, 10(3), 273–304.
- Chib, S., Koop, G., Griffiths, B., Terrell, D. (2003). *Bayesian Econometrics (Advances in Econometrics)*. Wiley-Interscience, New-York.
- Cosma, I.A., Evers, L. (2010). *Markov Chains and Monte Carlo Methods*. African Institute for Mathematical Sciences, Cape Town.
- Cowles, M. (2004). Review of WinBUGS 1.4. *The American Statistician*, 58(4), 330–336.
- Ding, M., Rosner, G.L., Muller, P. (2008). Bayesian optimal design for phase ii screening trials. *Biometrics*, 64, 886–894.
- Gamerman, D., Lopes, F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2nd edn. Chapman & Hall, London.
- Gelman, A., Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. CRC/Chapman & Hall, London.
- Gregory, P.C. (2008). Detecting extra-solar planets with a Bayesian hybrid MCMC Kepler periodogram. In: *JSM Proceedings*, Denver, CO.
- Hoag, S. (2011). NIPTE-FDA collaborative case study on model-based design space development across scales & with stability considerations, impact of manufacturing variations on degradant formation. In: *Critical Path Research for Process Scale-up and Stability*, Washington, DC.
- ICH. (2009). *ICH Harmonised Tripartite Guideline, Pharmaceutical Development: Q8(R2)*.
- Jackman, S. (2004). Bayesian analysis for political research. *Annual Review of Political Science*, 7, 483–505.
- Kayrak-Talay, D., Litster, J., Bondi, R., Zacour, B.M., Drennen, J.K., Spence, K. *et al.* (2011). NIPTE-FDA collaborative case study on model-based design space development across scales & with stability considerations, wet granulation. In: *Critical Path Research for Process Scale-up and Stability*, Washington, DC.

- Kirsch, L.E. (2011). NIPTE-FDA collaborative case study on model-based design space development across scales & with stability considerations, connecting manufacturing output to shelf-life prediction. In: *Critical Path Research for Process Scale-up and Stability*, Washington, DC.
- Lainez, J., Blau, G., Mockus, L., Orcun, S., Reklaitis, G. (2011). Pharmacokinetic based design of individualized dosage regimens using a Bayesian approach. *Industrial and Engineering Chemistry Research*, 50, 5114–5130.
- Minka, T.P. (2001). Expectation propagation for approximate Bayesian inference. In: *Uncertainty in Artificial Intelligence*.
- Miró-Quesada, G., del Castillo, E., Peterson, J. (2004). A Bayesian approach for multiple response surface optimization in the presence of noise variables. *Journal of Applied Statistics*, 31, 251–270.
- Mockus, L., LeBlond, D., Basu, P.K., Shah, R.B., Khan, M.A. (2011). A QbD case study: Bayesian prediction of lyophilization cycle parameters. *AAPS PharmSciTech*, 12(1), 442–448.
- Moore, C. (2007). Design space – an FDA perspective. In: *43rd Annual Meeting of the Drug Information Association*, Atlanta.
- Peterson, J.J. (2004). A posterior predictive approach to multiple response surface optimization. *Journal of Quality Technology*, 36, 139–153.
- Peterson, J.J. (2007). A review of Bayesian reliability approaches to multiple response surface optimization in Bayesian process monitoring, control, and optimization. In: Colosimo, B.M., Castillo, E. (Eds.). Chapman and Hall, London.
- Peterson, J.J. (2008). A Bayesian approach to the ich q8 definition of design space. *Journal of Biopharmaceutical Statistics*, 18(5), 959–975.
- Peterson, J.J., Yahyah, M. (2009). A Bayesian design space approach to robustness and system suitability for pharmaceutical assays and other processes. *Statistics in Biopharmaceutical Research*.
- Radaduen, T., Stamatis, S., Nguyen, H.Q., Zong, Z., Kirsch, L.E. (2011). Formulation of a gabapentin drug degradation model that combines manufacturing and storage stress variables. In: *Critical Path Research for Process Scale-up and Stability*, Washington, DC.
- Smidl, V., Quinn, A. (2006). *The Variational Bayes Method in Signal Processing*. Springer, Berlin.
- Woodcock, J. (2009). *Dr. Woodcock Comments on The Critical Path Initiative*.
- Zacour, B.M., Anderson, C.A., Drennen, J.K. (2011). NIPTE-FDA Collaborative case study on model-based design space development across scales & with stability considerations, fluid bed drying. In: *Critical Path Research for Process Scale-up and Stability*, Washington, DC.
- Zong, Z., Desai, S., Kaushal, A.M., Barich, D., Huang, H.-S., Munson, E. et al. (2011). The stabilizing effect of moisture on the solid-state degradation of gabapentin. *AAPS PharmSciTech*.

L. Mockus is a senior research scientist at Purdue University. He received BSc in applied mathematics from Moscow Institute of Physics and Technology in 1984. In 1991 Mockus earned MSc in computer science from Moscow Institute of Physics and Technology and PhD in chemical engineering from Purdue University in 1996. Currently Mockus is managing diverse research projects on the application of Quality by Design (QbD) principles to product and process development and scale up. Projects include small molecule and biologics, lyophilized parenteral and oral solid dosage forms, integrated multi-unit operation process. His previous industrial experience includes managing validation projects at Allergan and supporting product and process development activities at Monsanto/Pharmacia. Mockus authored/co-authored more than 30 publications in peer reviewed journals and co-authored one monograph on global Bayesian optimization.

J.M. Laínez is a postdoctoral research associate at Purdue University. He received a BSc in industrial engineering from Universidad Centro Americana “Jose Simeon Cañas”, El Salvador in 2002. He obtained his PhD in process systems engineering at Universitat Politècnica de Catalunya in 2010. Currently his research is focused on the application of Bayesian inference to problems arising in the pharmaceutical industry.

G.V. Reklaitis is Burton and Kathryn Gedge professor of chemical engineering at Purdue University. He received BS ChE from Illinois Institute of Technology and PhD ChE from Stanford University. He served as long term editor-in-chief of *Computers & Chemical Engineering*, member of the board of directors of the American Institute of Chemical Engineers, past chair of the NIPTE faculty committee and currently is deputy director of the NSF Engineering Research Center on Structured Organic Particulate Systems. He is a fellow of the AIChE and member of the US National Academy of Engineering. His research interests lie in mathematical modeling, optimization and simulation methodologies and their application to operational and design problems arising in the chemical, pharmaceutical and energy systems sectors.

L. Kirsch is a professor of pharmaceutics and chemical and biochemical engineering at the University of Iowa. He received a BS in pharmacy from Purdue University and a PhD in pharmaceutical chemistry from the Ohio State University. He was the editor of the PDA *Journal of Pharmaceutical Science and Technology* from 2000 to 2008 and is currently the editor-in-chief of the *AAPS PharmSciTech* journal. He is the past chair of the faculty committee for the National Institute for Pharmaceutical Technology and Education (NIPTE), AAPS Sterile Products Focus Group and Arden House conference.

Bajesinė rizikos metrika kiekybiškai įvertinti vaistų kokybę

Linas MOCKUS, José Miguel LAÍNEZ, Gintaras REKLAITIS, Lee KIRSCH

FDA kokybės per dizainą iniciatyva ir su ja susijusi dizaino erdvės konstrukcija (ICH, 2009) skatina naudoti matematinius modelius bei eksperimentų planavimą vaistų kūrimui ir gamybai. Dizaino erdvė gali būti interpretuota kaip regionas, esantis gamybinių kintamųjų erdvėje, kuriai duoto vaisto specifikacijos yra patenkintos. Dabar, bent jau konceptualiai, suprantama, kad ši garantija iš tikrųjų yra stochastinė. Šis straipsnis skirtas rizikos metrikai, kuri yra paremta Bajesiniais metodais bei vaistų gamybos matematiniais modeliais. A posteriori paskirstymas, kad vaistas atitinka specifikacijas, surandamas naudojant variacinį Bayeso algoritmą. Skaičiavimai sulgyinti su plačiai paplitusiu Markovo grandinės Monte Karlo metodu. Algoritmas yra pailiustruotas naudojantis kokybės pagal dizainą tyrimo, skirto gabapentin vaisto gamybai, eksperimentiniais duomenimis bei modeliais.