

Bayesian Hierarchical Modeling for Integrating Low-accuracy and High-accuracy Experiments

Peter Z. G. Qian and C. F. Jeff Wu

*Department of Statistics,
University of Wisconsin, Madison, WI 53706
(zhiguang@stat.wisc.edu)*

*School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, GA 30332-0205
(jeffwu@isye.gatech.edu)*

August 31, 2007; 2nd Revised version

Abstract

Standard practice in analyzing data from different types of experiments is to treat data from each type separately. By borrowing strength across multiple sources, an integrated analysis can produce better results. Careful adjustments need to be made to incorporate the systematic differences among various experiments. To this end, some Bayesian hierarchical Gaussian process models (BHGP) are proposed. The heterogeneity among different sources is accounted for by performing flexible location and scale adjustments. The approach tends to produce prediction closer to that from the high-accuracy experiment. The Bayesian computations are aided by the use of Markov chain Monte Carlo and Sample Average Approximation algorithms. The proposed method is illustrated with two examples: one with detailed and approximate finite elements simulations for mechanical material design and the other with physical and computer experiments for modeling a food processor.

KEY WORDS: computer experiments; kriging; Gaussian process; Markov chain Monte Carlo; stochastic programming.

1 Introduction

A challenging and fascinating problem in design and analysis of experiments is the synthesis of data from different types of experiments. With the advances in computing and experimentation, scientists can quickly access data from different sources. Complex mathematical models, implemented in large computer codes, are widely used to study real systems. Doing the corresponding physical experimentation would be more time-consuming and costly. For example, each physical run of the fluidized bed process (to be discussed in Section 4) can take days or even weeks to finish while running the associated computer code only takes minutes per run. Furthermore, a large computer program can often be run at different levels of sophistication with vastly varying computational times. Consider, for example, two codes that simulate linear cellular alloys for electronic cooling systems (to be discussed in Section 3). One code uses finite element analysis while the other is based on finite difference method. The two codes differ in the numerical method and the resolution of the grid, resulting in an accurate but slow version and a crude but fast approximation. In this paper, we consider a generic situation in which two sources (or experiments) are available and one source is generally more accurate than the other but also more expensive to run. The two experiments considered are called *low-accuracy experiment* and *high-accuracy experiment* and referred to as LE and HE respectively. The pair can be physical vs. computer experiments or detailed vs. approximate computer experiments. Experimenters are often faced with the problem of how to integrate these multiple data sources efficiently. There is a recent surge of interests in this problem. Related work includes Goldstein and Rougier (2004), Higdon et al. (2004), Kennedy and O’Hagan (2000, 2001), Qian et al. (2006) and Reese et al. (2004).

The purpose of this paper is to introduce Bayesian hierarchical Gaussian process (BHGP) models to integrate multiple data sources. The heterogeneity among different sources is accounted for by performing flexible location and scale adjustments. The article is organized as follows. The BHGP models are developed in Section 2. Sections 3 and 4 illustrate the method with two real examples: one with detailed and approximate computer experiments and

the other with physical and computer experiments. Concluding remarks and extensions are given in Section 5. Some computational details are included in the Appendix.

2 Bayesian Hierarchical Gaussian Process Models

Standard approaches to the synthesis of low-accuracy and high-accuracy experiments analyze data from each type separately. By borrowing strength across multiple sources, an integrated analysis can produce better results. Qian et al. (2006) introduces a two-step approach to integrate results from detailed and approximate computer experiments. It starts with fitting a Gaussian process model for the approximate experiment data. In the second step, the fitted model is adjusted by incorporating the more accurate data from the detailed experiment. The present work can be viewed as an extension of theirs. The essential differences between the two approaches are two-fold. First, new hierarchical Gaussian process models are introduced to carry out location and scale adjustments more flexibly. Second, the present approach adopts the Bayesian formulation and can absorb uncertainty in the model parameters in the prediction. Reese et al. (2004) proposes another hierarchical method by using linear models to integrate data from physical and computer experiments. Although this approach has advantages such as the ease of computation and interpretation, the linear models cannot serve as interpolators whereas the Gaussian process models have this feature when modeling deterministic computer experiments. Also the linear models are not as flexible as the Gaussian process models in representing complex nonlinear relationships.

Suppose that the LE and HE involve the same k factors $\mathbf{x} = (x_1, \dots, x_k)$. Denote by $D_l = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ the design set for the LE with n runs, and $\mathbf{y}_l = (y_l(\mathbf{x}_1), \dots, y_l(\mathbf{x}_n))^t$ the corresponding LE data. Because an HE run requires more computational effort to generate than an LE run, usually there are fewer HE runs available. Without loss of generality, we assume that D_h , the design set of the HE, consists of the first n_1 ($n_1 < n$) runs of D_l . The outputs from the HE are denoted by $\mathbf{y}_h = (y_h(\mathbf{x}_1), \dots, y_h(\mathbf{x}_{n_1}))^t$. Note that the subscripts h and l denote “high” and “low”. The main goal of the proposed method is to predict y_h at some *untried* points (i.e., these

points outside D_h). Central to the method are Bayesian hierarchical Gaussian process (BHGP) models, which consist of the following two parts:

1. A smooth model for the LE data.
2. A flexible model to “link” the LE and the HE data.

Detailed descriptions of these two models are given in Sections 2.2 and 2.3.

2.1 Bayesian Gaussian process model

In this section, we present the basics of Bayesian analysis of a Gaussian process model as the basis for later development. A good reference for Gaussian process models is Santner, Williams and Notz (2003) (hereafter abbreviated as SWN, 2003). For simplicity, throughout the paper a Gaussian process with mean μ and variance σ^2 is denoted by $GP(\mu, \sigma^2, \phi)$, where ϕ will be defined below. Suppose $y(\mathbf{x})$ is a real-valued stationary Gaussian process on the real line with mean $E\{y(\mathbf{x})\} = \mathbf{f}(\mathbf{x})^t \boldsymbol{\beta}$, where $\mathbf{x} = (x_1, \dots, x_k)$, $\mathbf{f}(\mathbf{x}) = \{f_1(\mathbf{x}), \dots, f_q(\mathbf{x})\}^t$ is a known vector-valued function and $\boldsymbol{\beta}$ is a vector of unknown regression coefficients. Furthermore, the covariance function for two input values \mathbf{x}_1 and \mathbf{x}_2 is represented by $\text{cov}(y(\mathbf{x}_1), y(\mathbf{x}_2)) = \sigma^2 K_\phi(\mathbf{x}_1, \mathbf{x}_2)$, where σ^2 is the variance, and $K_\phi(\cdot, \cdot)$ is the correlation function and depends on the unknown *correlation parameters* ϕ . Although the proposed method works for general correlation functions, we specifically apply it to the following *Gaussian correlation function* (SWN, 2003)

$$K_\phi(\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^k \exp\{-\phi_{i1}(x_{1i} - x_{2i})^2\}. \quad (1)$$

Here, the scale correlation parameters ϕ_{i1} are positive. The power correlation parameters are fixed at 2 (SWN, 2003), thus reducing the complication of estimating the correlation parameters. In addition, the sample path of the Gaussian process with this assumption is infinitely differentiable, which is a reasonable assumption for many applications including the examples in Sections 3 and 4. As a result, this correlation is often adopted in the computer experiments literature (Welch et al., 1992; SWN, 2003). In general, we observe $\mathbf{y} = \{y(\mathbf{x}_1), \dots, y(\mathbf{x}_n)\}$ and are interested in predicting y at a given point \mathbf{x}_0 .

The priors for the model parameters $\boldsymbol{\beta}$, σ^2 , $\boldsymbol{\phi}$ take the following structure

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) = p(\boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\phi}) = p(\boldsymbol{\beta}|\sigma^2)p(\sigma^2)p(\boldsymbol{\phi}). \quad (2)$$

The choice of priors requires some care. As pointed out in Berger et al. (2001), improper priors chosen for $\boldsymbol{\phi}$ may lead to improper posteriors as well. To avoid this problem, proper priors are adopted as follows:

$$\begin{aligned} p(\sigma^2) &\sim IG(\alpha, \gamma), \\ p(\boldsymbol{\beta}|\sigma^2) &\sim N(\mathbf{u}, v\mathbf{I}_{q \times q}\sigma^2), \end{aligned}$$

and

$$\phi_i \sim G(a, b), \text{ for } i = 1, \dots, k, \quad (3)$$

where $IG(\alpha, \gamma)$ denotes the inverse gamma distribution with density function

$$p(z) = \frac{\gamma^\alpha}{\Gamma(\alpha)} z^{-(\alpha+1)} \exp\left\{-\frac{\gamma}{z}\right\}, \quad z > 0,$$

$G(a, b)$ is the gamma distribution with density function

$$p(z) = \frac{b^a}{\Gamma(a)} z^{a-1} e^{-bz}, \quad z > 0,$$

$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ and $\mathbf{I}_{q \times q}$ is the $q \times q$ identity matrix.

It can be shown (SWN, 2003) that the conditional distribution of y at \mathbf{x}_0 , given the observed \mathbf{y} the correlation parameters $\boldsymbol{\phi}$, is the non-central t distribution

$$T_1(n + \nu_0, \mu_1, \sigma_1^2), \quad (4)$$

where

$$\begin{aligned} \mu_1 &= \mathbf{f}_0^t \boldsymbol{\mu}_{\beta|n} + \mathbf{r}_0^t \mathbf{R}^{-1}(\mathbf{y} - \mathbf{F} \boldsymbol{\mu}_{\beta|n}), \\ \boldsymbol{\mu}_{\beta|n} &= (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F} + v^{-1} \mathbf{I}_{q \times q})^{-1} (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{y} + \mathbf{u} v^{-1} \mathbf{I}_{q \times q}), \\ \widehat{\boldsymbol{\beta}} &= (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{y}), \\ \sigma_1^2 &= \frac{Q_1^2}{\nu_1} \left\{ 1 - (\mathbf{f}_0^t, \mathbf{r}_0^t) \begin{bmatrix} -v^{-1} \mathbf{I}_{q \times q} & \mathbf{F}^t \\ \mathbf{F}^t & \mathbf{R} \end{bmatrix}^{-1} \begin{pmatrix} \mathbf{f}_0 \\ \mathbf{r}_0 \end{pmatrix} \right\}, \\ Q_1^2 &= c_0 + \mathbf{y}^t [\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{F} (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}^{-1}] \mathbf{y} \\ &\quad + (\mathbf{u} - \widehat{\boldsymbol{\beta}})^t [v \mathbf{I}_{q \times q} + (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1}]^{-1} (\mathbf{u} - \widehat{\boldsymbol{\beta}}), \end{aligned}$$

$\nu_0 = 2a$, $\nu_1 = n + 2a$, $c_0 = \sqrt{\frac{b}{a}}$, $\mathbf{f}_0 = \mathbf{f}(\mathbf{x}_0)$, $\mathbf{r}_0 = (R(\mathbf{x}_0, \mathbf{x}_1), \dots, R(\mathbf{x}_0, \mathbf{x}_n))^t$, \mathbf{R} is the correlation matrix with entry $R(\mathbf{x}_i, \mathbf{x}_j)$ for $i, j = 1, \dots, n$, and $\mathbf{F} = [\mathbf{f}(\mathbf{x}_1)^t, \dots, \mathbf{f}(\mathbf{x}_n)^t]^t$ is the regressor matrix.

2.2 Low-accuracy experiment data

We assume that $y_l(\mathbf{x}_i)$ can be described by

$$y_l(\mathbf{x}_i) = f_l^t(\mathbf{x}_i)\boldsymbol{\beta}_l + \epsilon_l(\mathbf{x}_i), \quad i = 1, \dots, n, \quad (5)$$

where $f_l(\mathbf{x}_i) = (1, x_{i1}, \dots, x_{ik})^t$, $\boldsymbol{\beta}_l = (\beta_{l0}, \beta_{l1}, \dots, \beta_{lk})^t$ and $\epsilon_l(\cdot)$ is assumed to be $GP(0, \sigma_l^2, \boldsymbol{\phi}_l)$. Here, the mean function includes linear effects, because in many circumstances (including the two examples given later) it is reasonable to assume the factors considered in the experiments have linear effects on the outputs (Handcock and Wallis, 1994; Joseph, Hung and Sudjianto, 2007; Qian et al., 2006). In addition, inclusion of “weak” main effects in the mean of a Gaussian process can bring additional numerical benefits for estimating the correlation parameters. Suppose, instead, the mean in (5) includes only a constant μ , the likelihood of \mathbf{y}_l will be

$$\propto \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\{-(\mathbf{y}_l - \mu\mathbf{1}_n)^t \boldsymbol{\Sigma}^{-1} (\mathbf{y}_l - \mu\mathbf{1}_n)\}, \quad (6)$$

where the covariance matrix $\boldsymbol{\Sigma}$ depends on the unknown correlation parameters $\boldsymbol{\phi}_l$ and $\mathbf{1}_n$ represents the n -unity column vector. For a large number of observations, (6) can be extremely small regardless of the values of $\boldsymbol{\phi}_l$. As a result, $\boldsymbol{\phi}_l$ cannot be accurately estimated. The inclusion of some weak main effects in the mean can partially mitigate this problem by “dampening” the Mahalabonis distance between $\mu\mathbf{1}_n$ and \mathbf{y}_l .

If LE were the only source considered, then this model would be fitted using the Bayesian Gaussian process model discussed in Section 2.1. Because the LE data are not very accurate, the HE data need to be incorporated to improve the quality of the fitted model.

2.3 High-accuracy experiment data

Because LE and HE are conducted by using different mechanisms (physical or computational) or distinct numerical methods with different mesh sizes, orders of elements, or other important aspects, their outputs can be different.

In general we can classify the relationship between y_l and y_h into three broad categories:

1. LE produces outputs almost as good as HE;
2. No similarities can be found (or defined) between y_l and y_h ;
3. LE and HE give different outputs but share similar trends.

For category (1), the differences between y_l and y_h can be largely ignored, and using a single model for both data sources will suffice. Furthermore, the HE runs can be replaced by the LE runs, resulting in huge computational savings. However, these scenarios do not occur often in practice. The second category consists of cases, where LE and HE are “oranges” and “apples”. No sensible methods can be used to adjust the LE results and to integrate the LE and HE data. In such situations, the experimenters need to scrutinize the underlying assumptions or the set-ups of the LE and try to make improvements by better understanding the differences between LE and HE. Many problems in practice fall in category (3), which is the focus of the paper.

In order to “link” the HE data with the LE data, we consider the following *adjustment model*

$$y_h(\mathbf{x}_i) = \rho(\mathbf{x}_i)y_l(\mathbf{x}_i) + \delta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i), \quad i = 1, \dots, n_1. \quad (7)$$

Here $\rho(\cdot)$, assumed to be $GP(\rho_0, \sigma_\rho^2, \boldsymbol{\phi}_\rho)$, accounts from *scale change* from LE to HE. We assume $\delta(\cdot)$ to be $GP(\delta_0, \sigma_\delta^2, \boldsymbol{\phi}_\delta)$ and represent *location adjustment*. The measurement error $\epsilon(\cdot)$ is assumed to be $N(0, \sigma_\epsilon^2)$. Furthermore, $y_l(\cdot)$, $\delta(\cdot)$, $\rho(\cdot)$ and $\epsilon(\cdot)$ are assumed to be independent. The proposed model can be viewed as an extension of that used in Kennedy and O’Hagan (2000), where an autoregressive model is used as an adjustment model with a constant chosen for scale adjustment. Discussions and comparisons between the proposed method and some existing methods (including the Kennedy-O’Hagan method) will be given in Section 2.8. Similar to the discussion on identifiability in Kennedy and O’Hagan (2001), the hierarchical model in (5) and (7) are identifiable in the classical sense when the correlation parameters are given (Koopmans and Reiersøl, 1950; Rothenberg, 1971).

The unknown parameters $\boldsymbol{\theta}$ involved in models (5) and (7) can be collected into three groups: mean parameters $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_l, \rho_0, \delta_0)$, variance parameters $\boldsymbol{\theta}_2 = (\sigma_l^2, \sigma_\rho^2, \sigma_\delta^2, \sigma_\epsilon^2)$ and correlation parameters $\boldsymbol{\theta}_3 = (\boldsymbol{\phi}_l, \boldsymbol{\phi}_\rho, \boldsymbol{\phi}_\delta)$. The description of the hierarchical models in (5) and (7) is complete with the

specification of priors. It is similar to that of the Bayesian Gaussian process model in Section 2.1. The chosen priors take the following form

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)p(\boldsymbol{\theta}_3) = p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_3), \quad (8)$$

where

$$\begin{aligned} p(\sigma_l^2) &\sim IG(\alpha_l, \gamma_l), \\ p(\sigma_\rho^2) &\sim IG(\alpha_\rho, \gamma_\rho), \\ p(\sigma_\delta^2) &\sim IG(\alpha_\delta, \gamma_\delta), \\ p(\sigma_\epsilon^2) &\sim IG(\alpha_\epsilon, \gamma_\epsilon), \\ p(\boldsymbol{\beta}_l|\sigma_l^2) &\sim N(\mathbf{u}_l, v_l\mathbf{I}_{(k+1)\times(k+1)}\sigma_l^2), \\ p(\rho_0|\sigma_\rho^2) &\sim N(u_\rho, v_\rho\sigma_\rho^2), \\ p(\delta_0|\sigma_\delta^2) &\sim N(u_\delta, v_\delta\sigma_\delta^2), \\ \phi_{l_i} &\sim G(a_l, b_l), \phi_{\rho_i} \sim G(a_\rho, b_\rho), \phi_{\delta_i} \sim G(a_\delta, b_\delta), \text{ for } i = 1, \dots, k. \end{aligned} \quad (9)$$

2.4 Bayesian prediction and Markov chain Monte Carlo sampling

Now we discuss the prediction of y_h at an untried point \mathbf{x}_0 . An empirical Bayes approach is taken here by estimating the correlation parameters for computational convenience. This approach is popular for fitting Gaussian process (GP) models (Bayarri et al., 2007; Lu, Sun and Zidek, 1997; Williams, Santner and Notz, 2000). An intuitive explanation for taking this approach is “one expects modest variations in correlation parameters to have little effect on the Gaussian process predictors because they are interpolators” (Bayarri et al., 2007). In addition, Nagy (2006) recently reported simulations for problems with 1-10 input variables indicating reasonably close accuracy between the predictions using this approach and the predictions using the fully Bayesian approach, which uses posterior samples of the correlation parameters. The fully Bayesian analysis would be computationally more intensive since it needs to draw posterior samples for $\boldsymbol{\theta}_3$. Sampling from $\boldsymbol{\theta}_3$ can be difficult because the conditional distribution $p(\boldsymbol{\theta}_3|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ has a very

irregular form

$$p(\boldsymbol{\theta}_3) \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} \exp \left\{ -\frac{(\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})}{2\sigma_\rho^2} \right\} \\ \cdot \frac{1}{|\mathbf{R}_l|^{\frac{1}{2}}} \exp \left\{ -\frac{(\mathbf{y}_l - \mathbf{F}_l \boldsymbol{\beta}_l)^t \mathbf{R}_l^{-1} (\mathbf{y}_l - \mathbf{F}_l \boldsymbol{\beta}_l)}{2\sigma_l^{2n}} \right\}, \quad (10)$$

where $\boldsymbol{\theta}_3$ appear in $p(\boldsymbol{\theta}_3)$ and all elements of four complex matrix inversions and determinants $|\mathbf{M}|^{-\frac{1}{2}}$, \mathbf{M}^{-1} , $|\mathbf{R}_l|^{-\frac{1}{2}}$ and \mathbf{R}_l^{-1} (with their definitions given in (14)).

In this section, we develop a prediction procedure by assuming the value of $\boldsymbol{\theta}_3$ is given. In Section 2.5, we shall discuss the fitting of $\boldsymbol{\theta}_3$. For the ease of methodological development, we first assume that the untried point \mathbf{x}_0 belongs to D_l but is not a point in D_h (otherwise $y_h(\mathbf{x}_0)$ is readily available). This assumption shall be relaxed later. The prediction is based on the *Bayesian predictive density function*

$$p[y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l] = \int_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} p[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3] p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2. \quad (11)$$

In this approach, uncertainty in the model parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is naturally absorbed in the prediction.

The integration of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in (11) needs to be done numerically. A Markov chain Monte Carlo (MCMC) (Liu, 2001) algorithm to approximate $p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$ is given as follows:

1. Generate $[\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}], \dots, [\boldsymbol{\theta}_1^{(M)}, \boldsymbol{\theta}_2^{(M)}]$ from $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3)$.
2. Approximate $p[y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l]$ by

$$\hat{p}_m[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3] = \frac{1}{M} \sum_{i=1}^M p[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3]. \quad (12)$$

For the ease of posterior sampling in step 1, we introduce new parameters $\tau_1 = \frac{\sigma_\delta^2}{\sigma_\rho^2}$ and $\tau_2 = \frac{\sigma_\epsilon^2}{\sigma_\rho^2}$ and use $(\sigma_\rho^2, \tau_1, \tau_2)$ instead of $(\sigma_\rho^2, \sigma_\delta^2, \sigma_\epsilon^2)$ in the model. With some abuse of notation, we shall still use $\boldsymbol{\theta}_2$ to denote $(\sigma_l^2, \sigma_\rho^2, \tau_1, \tau_2)$.

From (9), the prior for σ_ρ^2 , τ_1 and τ_2 is easily shown to be

$$\begin{aligned}
p(\sigma_\rho^2, \tau_1, \tau_2) &= \frac{\gamma_\rho^{\alpha_\rho}}{\Gamma(\alpha_\rho)} (\sigma_\rho^2)^{-(\alpha_\rho+1)} \exp\{-\gamma_\rho/\sigma_\rho^2\} \\
&\quad \cdot \frac{(\gamma_\delta)^{\alpha_\delta}}{\Gamma(\alpha_\delta)} (\sigma_\rho^2 \tau_1)^{-(\alpha_\delta+1)} \exp\{-\gamma_\delta/(\sigma_\rho^2 \tau_1)\} \\
&\quad \cdot \frac{(\gamma_\epsilon)^{\alpha_\epsilon}}{\Gamma(\alpha_\epsilon)} (\sigma_\rho^2 \tau_2)^{-(\alpha_\epsilon+1)} \exp\{-\gamma_\epsilon/(\sigma_\rho^2 \tau_2)\} \sigma_\rho^4. \quad (13)
\end{aligned}$$

The key for deriving the full conditional distributions of $(\boldsymbol{\beta}_l, \delta_0, \rho_0, \sigma_l^2, \sigma_\rho^2, \tau_1, \tau_2)$, which is used in the MCMC, is to note that by conditioning on $\boldsymbol{\theta}_3$, these parameters can be viewed as coming from some general linear models. Given $\boldsymbol{\theta}_3$, the full conditional distributions for $(\boldsymbol{\beta}_l, \delta_0, \rho_0, \sigma_l^2, \sigma_\rho^2, \tau_1, \tau_2)$ can be shown to be

$$\begin{aligned}
p(\boldsymbol{\beta}_l | \mathbf{y}_l, \mathbf{y}_h, \bar{\boldsymbol{\beta}}_l) &\sim N \left(\left[\frac{1}{v_l} \mathbf{I}_{(k+1) \times (k+1)} + \mathbf{F}_l^t \mathbf{R}_l^{-1} \mathbf{F}_l \right]^{-1} \left(\frac{\mathbf{u}_l}{v_l} + \mathbf{F}_l^t \mathbf{R}_l^{-1} \mathbf{y}_l \right), \right. \\
&\quad \left. \left[\frac{1}{v_l} \mathbf{I}_{(k+1) \times (k+1)} + \mathbf{F}_l^t \mathbf{R}_l^{-1} \mathbf{F}_l \right]^{-1} \sigma_l^2 \right), \\
p(\rho_0 | \mathbf{y}_l, \mathbf{y}_h, \bar{\rho}_0) &\sim N \left(\frac{\frac{u_\rho}{v_\rho} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} (\mathbf{y}_h - \delta_0 \mathbf{1}_{n_1})}{\frac{1}{v_\rho} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_{l_1}}, \frac{\sigma_\rho^2}{\frac{1}{v_\rho} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_{l_1}} \right), \\
p(\delta_0 | \mathbf{y}_l, \mathbf{y}_h, \bar{\delta}_0) &\sim N \left(\frac{\frac{u_\delta}{v_\delta \tau_1} + \mathbf{1}_{n_1}^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1})}{\frac{1}{v_\delta \tau_1} + \mathbf{1}_{n_1}^t \mathbf{M}^{-1} \mathbf{1}_{n_1}}, \frac{\sigma_\rho^2}{\frac{1}{v_\delta \tau_1} + \mathbf{1}_{n_1}^t \mathbf{M}^{-1} \mathbf{1}_{n_1}} \right), \\
p(\sigma_l^2 | \mathbf{y}_l, \mathbf{y}_h, \bar{\sigma}_l^2) &\sim IG \left(\frac{n}{2} + \frac{k+1}{2} + \alpha_l, \right. \\
&\quad \left. \frac{1}{2} \frac{(\boldsymbol{\beta}_l - \mathbf{u}_l)^t (\boldsymbol{\beta}_l - \mathbf{u}_l)}{v_l} + \frac{1}{2} (\mathbf{y}_l - \mathbf{F}_l \boldsymbol{\beta}_l)^t \mathbf{R}_l^{-1} (\mathbf{y}_l - \mathbf{F}_l \boldsymbol{\beta}_l) + \gamma_l \right), \\
p(\sigma_\rho^2 | \mathbf{y}_l, \mathbf{y}_h, \bar{\sigma}_\rho^2) &\sim IG \left(\frac{n_1}{2} + \frac{1}{2} + \alpha_\rho + \alpha_\delta + \alpha_\epsilon, \frac{(\rho_0 - u_\rho)^2}{2v_\rho} + \gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\epsilon}{\tau_2} \right. \\
&\quad \left. + \frac{(\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})}{2} \right),
\end{aligned}$$

$$p(\tau_1, \tau_2 | \mathbf{y}_l, \mathbf{y}_h, \overline{\tau_1}, \overline{\tau_2}) \propto \frac{1}{\tau_1^{\alpha_\delta + \frac{3}{2}}} \frac{1}{\tau_2^{\alpha_\epsilon + 1}} \exp \left\{ -\frac{1}{\tau_1} \left(\frac{\gamma_\delta}{\sigma_\rho^2} + \frac{(\delta_0 - u_\delta)^2}{2v_\delta \sigma_\rho^2} \right) - \frac{\gamma_\epsilon}{\tau_2 \sigma_\rho^2} \right\} \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} \\ \cdot \exp \left\{ -\frac{(\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})}{2\sigma_\rho^2} \right\}, \quad (14)$$

where $\overline{\tau_1}$ represents all the components of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ except for ω , $\mathbf{M} = \mathbf{W}_\rho + \tau_1 \mathbf{R}_\delta + \tau_2 \mathbf{I}_{n_1 \times n_1}$ and depends on $\boldsymbol{\phi}_\rho$, $\boldsymbol{\phi}_\delta$, τ_1 and τ_2 , $\mathbf{y}_{l_1} = (y_l(\mathbf{x}_1), \dots, y_l(\mathbf{x}_{n_1}))^t$, $\mathbf{W}_\rho = \mathbf{A}_1 \mathbf{R}_\rho \mathbf{A}_1$, $\mathbf{A}_1 = \text{diag}\{y_l(\mathbf{x}_1), \dots, y_l(\mathbf{x}_{n_1})\}$ and \mathbf{R}_ρ and \mathbf{R}_δ are the correlation matrices of $\boldsymbol{\rho} = (\rho(\mathbf{x}_1), \dots, \rho(\mathbf{x}_{n_1}))^t$ and $\boldsymbol{\delta} = (\delta(\mathbf{x}_1), \dots, \delta(\mathbf{x}_{n_1}))^t$ respectively.

The Metropolis-within-Gibbs algorithm (Liu, 2001; Gelman et al., 2004) can be used to sample from this full conditional distribution, where a Metropolis draw is included for sampling τ_1 and τ_2 within the usual Gibbs loop.

The second step of the approximation in (12) is straightforward. The analytic form of $p[y_h(\mathbf{x}_0) | \mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3]$ can be obtained by rewriting it as

$$\frac{p[y_h(\mathbf{x}_0), \mathbf{y}_h | \mathbf{y}_l, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3]}{p(\mathbf{y}_h | \mathbf{y}_l, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)}. \quad (15)$$

From the assumption that $\rho(\cdot)$, $\delta(\cdot)$ and $\epsilon(\cdot)$ are independent of \mathbf{y}_l in (7), the distributions of the numerator and the denominator in (15) are as follows:

$$p(y_h[\mathbf{x}_0], \mathbf{y}_h | \mathbf{y}_l, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \sim N(\rho_0 \mathbf{y}_{l_1}^* + \delta_0 \mathbf{1}_{n_1+1}, \sigma_\rho^2 \mathbf{W}_\rho^* + \sigma_\delta^2 \mathbf{R}_\delta^* + \sigma_\epsilon^2 \mathbf{I}_{(n_1+1) \times (n_1+1)}), \\ p(\mathbf{y}_h | \mathbf{y}_l, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \sim N(\rho_0 \mathbf{y}_{l_1} + \delta_0 \mathbf{1}_{n_1}, \sigma_\rho^2 \mathbf{W}_\rho + \sigma_\delta^2 \mathbf{R}_\delta + \sigma_\epsilon^2 \mathbf{I}_{n_1 \times n_1}), \quad (16)$$

where

$$\mathbf{y}_{l_1}^* = [y_l(\mathbf{x}_0), y_l(\mathbf{x}_1), \dots, y_l(\mathbf{x}_{n_1})]^t, \\ \mathbf{W}_\rho^* = \mathbf{A}_1^* \mathbf{R}_\rho^* \mathbf{A}_1^*, \\ \mathbf{A}_1^* = \text{diag}\{y_l(\mathbf{x}_0), y_l(\mathbf{x}_1), \dots, y_l(\mathbf{x}_{n_1})\},$$

and \mathbf{R}_ρ^* and \mathbf{R}_δ^* are the correlation matrices of $\boldsymbol{\rho}^* = [\rho(\mathbf{x}_0), \rho(\mathbf{x}_1), \dots, \rho(\mathbf{x}_{n_1})]^t$ and $\boldsymbol{\delta}^* = [\delta(\mathbf{x}_0), \delta(\mathbf{x}_1), \dots, \delta(\mathbf{x}_{n_1})]^t$ respectively. Equivalently, this posterior density can be computed directly as a conditional Multivariate normal density. Since it only depends on ρ_0 , δ_0 , σ_ρ^2 and (τ_1, τ_2) , an alternative is to draw posterior samples in the first step from the joint posterior density for these parameters.

We can use the approximated predictive density in (12) to compute the posterior expectation

$$\widehat{y}_h(\mathbf{x}_0) = E[y_h(\mathbf{x}_0) | \mathbf{y}_l, \mathbf{y}_h] \quad (17)$$

as the *predictor* for $y_h(\mathbf{x}_0)$ and $\text{Var}[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h]$ as the *prediction variance*.

Next, we relax the assumption $\mathbf{x}_0 \in D_l/D_h$ and consider the prediction when \mathbf{x}_0 does not belong to D_l . The additional difficulty is that the value of $y_l(\mathbf{x}_0)$ is not observed. In the Bayesian framework, we can fit the Bayesian Gaussian process model as described in Section 2.1 and impute $y_l(\mathbf{x}_0)$ by $\hat{y}_l = E[y_l(\mathbf{x}_0)|\mathbf{y}_l]$ (the mean of a non-central t distribution). Then we can add \hat{y}_l to the set of \mathbf{y}_l so that \mathbf{x}_0 belongs to the expanded set $D_l \cup \{\mathbf{x}_0\}$. As suggested by one referee, we can also predict $y_h(\mathbf{x}_0)$ through the following approximation of $p[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3]$ (Gelman et al., 2004):

$$\hat{p}[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3] = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N p[y_h(\mathbf{x}_0)|\mathbf{y}_{l,j}^*, \mathbf{y}_h, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3],$$

where $\mathbf{y}_{l,j}^* = (y_{l,j}(\mathbf{x}_0), y_l(\mathbf{x}_1), \dots, y_l(\mathbf{x}_{n_1}))^t$ and $y_{l,1}(\mathbf{x}_0), \dots, y_{l,N}(\mathbf{x}_0)$ are N independent draws from $p(y_l(\mathbf{x}_0)|\mathbf{y}_l, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \boldsymbol{\theta}_3)$.

2.5 Estimation of correlation parameters

As discussed in Section 2.4, the proposed empirical Bayes approach fixes the correlation parameters $\boldsymbol{\theta}_3$ at their posterior modes. In this section, we discuss the fitting of $\boldsymbol{\theta}_3$.

Note that

$$p(\boldsymbol{\theta}_3|\mathbf{y}_h, \mathbf{y}_l) \propto p(\boldsymbol{\theta}_3) \int_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) p(\mathbf{y}_l, \mathbf{y}_h|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2, \quad (18)$$

which can be shown (see Appendix) to be proportional to

$$\begin{aligned} L_1 = p(\boldsymbol{\theta}_3) \int_{\tau_1, \tau_2} \tau_1^{-(\alpha_\delta + \frac{3}{2})} \tau_2^{-(\alpha_\epsilon + 1)} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} (a_2 a_3)^{-\frac{1}{2}} \\ \cdot \left(\gamma_l + \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8} \right)^{-(\alpha_l + \frac{n}{2})} \left(\gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\epsilon}{\tau_2} + \frac{4a_3 c_3 - b_3^2}{8a_3} \right)^{-(\alpha_\rho + \alpha_\delta + \alpha_\epsilon + \frac{n_1}{2})} d\tau_1 d\tau_2. \end{aligned} \quad (19)$$

The posterior mode estimator $\hat{\boldsymbol{\theta}}_3$ is given by an optimal solution to the optimization problem

$$\max_{\phi_l, \phi_\rho, \phi_\delta} L_1,$$

which is equivalent to solving the following two separable problems

$$\max_{\phi_l} p(\phi_l) |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} \left(\gamma_l + \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8} \right)^{-(\alpha_l + \frac{n}{2})}$$

and

$$\begin{aligned} \max_{\phi_\rho, \phi_\delta} \int_{\tau_1, \tau_2} p(\phi_\rho) p(\phi_\delta) \tau_1^{-(\alpha_\delta + \frac{3}{2})} \tau_2^{-(\alpha_\epsilon + 1)} |\mathbf{M}|^{-\frac{1}{2}} (a_2 a_3)^{-\frac{1}{2}} \\ \cdot \left(\gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\epsilon}{\tau_2} + \frac{4a_3 c_3 - b_3^2}{8a_3} \right)^{-(\alpha_\rho + \alpha_\delta + \alpha_\epsilon + \frac{n_1}{2})} d\tau_1 d\tau_2. \end{aligned}$$

The first optimization problem can be solved by using standard non-linear optimization algorithms like the quasi-Newton method. Solving the second problem is more elaborate because its objective function involves integration. This problem can be recast as

$$\max_{\phi_\rho, \phi_\delta} \{L_2 = E_{\tau_1, \tau_2} f(\tau_1, \tau_2)\}, \quad (20)$$

where

$$f(\tau_1, \tau_2) = \frac{p(\phi_\rho) p(\phi_\delta) \exp(\frac{2}{\tau_1}) \exp(\frac{2}{\tau_2})}{|\mathbf{M}|^{\frac{1}{2}} (a_2 a_3)^{\frac{1}{2}} \left(\gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\epsilon}{\tau_2} + \frac{4a_3 c_3 - b_3^2}{8a_3} \right)^{\alpha_\rho + \alpha_\delta + \alpha_\epsilon + \frac{n_1}{2}}},$$

$p(\tau_1) \sim IG(\alpha_\delta + \frac{1}{2}, 2)$, $p(\tau_2) \sim IG(\alpha_\epsilon, 2)$, and $p(\tau_1)$ and $p(\tau_2)$ are independent.

The problem in (20) can be viewed as a *stochastic program* in mathematical programming and solved by using the *Sample Average Approximation* (SAA) method (Ruszczynski and Shapiro, 2003). Generate Monte Carlo samples (τ_1^s, τ_2^s) from $p(\tau_1, \tau_2)$, $s = 1, \dots, S$, and estimate L_2 by

$$\widehat{L}_2 = \frac{1}{S} \sum_{s=1}^S f(\tau_1^s, \tau_2^s). \quad (21)$$

We refer to an optimal solution $\tilde{\phi}_\rho$ and $\tilde{\phi}_\delta$ of the problem

$$\max_{\phi_\rho, \phi_\delta} \widehat{L}_2 \quad (22)$$

as the *simulated posterior mode*. When S is large, the simulated posterior mode will be close to the true posterior mode (Ruszczynski and Shapiro,

2003). In the optimization literature, the SAA method is widely regarded as one of the most efficient algorithms for solving stochastic programs. It is known to be capable of solving large-scale problems with many variables. For discussions on theoretical properties of this algorithm, see Shapiro and Homem-de-Mello (2000) and Shapiro and Nemirovski (2005). Recent successful applications of this algorithm include Linderoth, Shapiro and Wright (2006) and Kleywegt, Shapiro and Homem-de-Mello (2001).

2.6 Simplifications when y_h is deterministic

Suppose y_h is deterministic (i.e., $\epsilon(\cdot) = 0$ in (7)), which is the case for the problem of detailed vs. approximate computer experiments. Some parts of the aforementioned procedure can be simplified as follows:

(a) Sampling from $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3)$.

Because $\tau_2 = 0$ in the model, $p(\sigma_\rho^2, \tau_1, \tau_2)$ in (13) is simplified by dropping its parts involving $\alpha_\epsilon, \gamma_\epsilon$ and τ_2 ; similarly, $p(\tau_1, \tau_2 | \mathbf{y}_l, \mathbf{y}_h, \bar{\tau}_1, \bar{\tau}_2)$ and \mathbf{M} are simplified by removing the parts involving $\alpha_\epsilon, \gamma_\epsilon$ and τ_2 .

(b) Bayesian prediction and interpolation.

If y_h is deterministic, the predictor $E[y_h(\mathbf{x}_i) | \mathbf{y}_l, \mathbf{y}_h]$ is $y_h(\mathbf{x}_i)$ and its prediction variance $Var[y_h(\mathbf{x}_i) | \mathbf{y}_l, \mathbf{y}_h]$ is 0, for $\mathbf{x}_i \in D_h$. This property implies that the predictor from the integrated analysis smoothly *interpolates* all the HE data points.

(c) Estimation of correlation parameters.

Because $\tau_2 = 0$ in the model, L_1 in (19) is simplified by dropping its parts involving $\alpha_\epsilon, \gamma_\epsilon$ and τ_2 , and becomes a one-dimensional integral; similarly, L_2 in (20) is simplified by removing the part involving $\alpha_\epsilon, \gamma_\epsilon$ and τ_2 , and becomes a stochastic program with one random variable.

2.7 Extension to the general situation with $D_h \not\subseteq D_l$

Here we discuss extending the proposed method to the general situation with $D_h \not\subseteq D_l$. Let R denote the design space for HE and LE with D_l and D_h as its two subsets. For the present situation, the main obstacle in predicting y_h is that the HE and LE data are not available at the same set of input values. To mitigate this difficulty, we can “combine” the observed HE and LE data with some carefully chosen *missing data* (Liu, 2001) such that after the data

combination, for every point in $D_l \cup D_h$, the responses for both LE and HE are available. To illustrate this idea, first partition the space R into four mutually exclusive components: R_1, R_2, R_3 and R_4 , where $R_1 = D_h \cap D_l$, $R_2 = D_h \cap D_l^c$, $R_3 = D_h^c \cap D_l$, $R_4 = D_h^c \cap D_l^c$ with “c” denoting the complement set. This partition implies that, for $\mathbf{x} \in R_1$, both y_l and y_h are available; for $\mathbf{x} \in R_2$, y_h is available but not y_l ; for $\mathbf{x} \in R_3$, y_l is available but not y_h ; and for $\mathbf{x} \in R_4$, neither y_l nor y_h is available. It is necessary to predict y_h for $\mathbf{x} \in R_3$ or R_4 , where y_h is not observed. Below we discuss prediction of y_h for $\mathbf{x}_0 \in R_4$. The prediction with $\mathbf{x}_0 \in R_3$ is similar and simpler, to which after some simplifications the proposed method is also applicable. The missing data here, denoted by y_{mis} , has three constituents. The first is the missing HE data, denoted by y_{hmis} . Since $(D_h \cup D_l) \cap D_h^c = R_3$, $y_{hmis} = \{y_h(\mathbf{x}), \mathbf{x} \in R_3\}$. The second is the missing LE data, denoted by y_{lmis} . Since $(D_h \cup D_l) \cap D_l^c = R_2$, $y_{lmis} = \{y_l(\mathbf{x}), \mathbf{x} \in R_2\}$. In the spirit of the prediction procedure in Section 2.4, y_{mis} includes $y_l(\mathbf{x}_0)$ as its third constituent.

To accommodate the inherent uncertainty in y_{mis} , we treat them as a latent variable. Another source of uncertainty in the prediction comes from the model parameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$. To incorporate the uncertainties in both $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and y_{mis} , a *data-augmentation* approach (Tanner and Wong, 1987) is taken here by augmenting these two sets of “parameters” together in the prediction. As in Section 2.4, prediction of $y_h(\mathbf{x}_0)$ is through the Bayesian predictive density $p[y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l]$, but with a modified form:

$$p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l) = \int_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, y_{mis}} p[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, y_{mis}] p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3, y_{mis}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 dy_{mis}.$$

For approximating this function, we modify the two-step procedure in (12) of Section 2.4 to:

1. Generate $[\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}, y_{mis}^{(1)}], \dots, [\boldsymbol{\theta}_1^{(M)}, \boldsymbol{\theta}_2^{(M)}, y_{mis}^{(M)}]$ from $p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, y_{mis} | \mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3)$.
2. Approximate $p(y_h(\mathbf{x}_0)|\mathbf{y}_h, \mathbf{y}_l)$ by

$$\hat{p}_M[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_3] = \frac{1}{M} \sum_{i=1}^M p[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, y_{mis}^{(i)}, \boldsymbol{\theta}_3].$$

For the posterior sampling in the first step, we modify the conditional sampling method in Section 2.4 as follows:

- (i) Expand the six components of the full conditional distributions in (14) by incorporating y_{mis} as an additional parameter.
- (ii) Include $p(y_{mis}|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ as a new component (i.e., the seventh) in the full conditional distributions in (14). Sampling from it is relatively easy since it is proportional to the ratio of two normal densities: $p(y_{mis}, \mathbf{y}_l, \mathbf{y}_h|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_2)$ and $p(\mathbf{y}_l, \mathbf{y}_h|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_2)$.

Similar to the situation with $D_h \subseteq D_l$, the calculation in the second step is straightforward because $p[y_h(\mathbf{x}_0)|\mathbf{y}_l, \mathbf{y}_h, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, y_{mis}, \boldsymbol{\theta}_3]$ is easily obtainable.

2.8 Comparison with existing methods

There are major differences between the proposed method and those in Kennedy and O’Hagan (2000) and Qian et al. (2006). The latter two consider integrating data from two deterministic experiments, while ours is also applicable to situations where HE has measurement errors. Ours is also more flexible in the modeling strategy. Kennedy and O’Hagan uses an autoregressive model as an adjustment model with a constant chosen for scale adjustment, which cannot handle complex scale change from LE to HE. Qian et al. uses a regression model for the scale component, which captures linear change. By utilizing a Gaussian process model, the scale adjustment in (7) can account for non-linear and complex changes as evidenced in the analysis of two examples given later. Qian et al. adopts a frequentist formulation and thus cannot account for uncertainties in the model parameters. The Kennedy-O’Hagan method uses a plug-in estimate $\hat{\rho}$ in the prediction, which cannot account for the variation in $\hat{\rho}$. The prediction in our approach is based on the Bayesian predictive density function in (11) so that the uncertainties in the model parameters are reflected. Furthermore, different from these methods, ours can be applied to situations with $D_h \not\subseteq D_l$.

3 Example 1: Designing Linear Cellular Alloys

We consider the data used in Qian et al. (2006), which consists of the outputs from computer simulations for a heat exchanger used in an electronic cooling application. As illustrated in Figure 1, the device is used to dissipate

heat generated by a heat source such as a microprocessor. The response y of interest is the total rate of steady state heat transfer of the device, which depends on the mass flow rate of entry air \dot{m} , the temperature of entry air T_{in} , the temperature of the heat source T_{wall} and the solid material thermal conductivity k . The device is assumed to have fixed overall width (W), depth (D), and height (H) of 9, 25, and 17.4 millimeters, respectively. This study uses two types of experiments: a detailed but slow simulation based on FLU-ENT finite element analysis (HE) and an approximate but fast simulation using finite difference method (LE). The responses of the two experiments are denoted by y_h and y_l respectively. Each HE run requires two to three orders of magnitude more computing time than the corresponding LE run. For example, the first run in Table 1 requires 1.75 hours and 2 seconds for HE and LE respectively on a 2.0 GHz Pentium 4 PC with 1 GB of RAM. Details on the engineering background can be found in Qian et al. (2006).

(Figure 1)

Table 1 gives the data consisting of 36 LE runs and 36 HE runs. The values of design variables are given in columns 1-4 of the table and the responses from the two experiments are given in columns 5 and 6. The data is divided into a training set and a testing set. We fit BHGP models using a training set consisting of 24 randomly selected HE runs and all 36 LE runs. The remaining 12 HE runs (i.e., run no. 1, 4, 9, 11, 13, 17, 18, 21, 26, 28, 30 and 31 as in the table) are left to form the testing set for model validation. Column 7 in the table gives the status of each HE run as training or testing.

(Table 1)

3.1 The Analysis

Table 1 shows that the four design variables have different scales and are thus standardized. The values of hyper-parameters used in this example are as follows: $(\alpha_l, \gamma_l, \alpha_\rho, \gamma_\rho, \alpha_\delta, \gamma_\delta) = (2, 1, 2, 1, 2, 1)$, $\mathbf{u}_l = (0, 0, 0, 0, 0, 0)^t$, $v_l = 1$, $(u_\rho, v_\rho, u_\delta, v_\delta) = (1, 1, 0, 1)$, $(a_l, b_l, a_\rho, b_\rho, a_\delta, b_\delta) = (2, 0.1, 2, 0.1, 2, 0.1)$. They are chosen to reflect our understanding of the model parameters. The “vague” prior $IG(2, 1)$ is chosen for $\sigma_l^2, \sigma_\rho^2$ and σ_δ^2 . The “location-flat” priors $N(0, \mathbf{I}_{7 \times 7} \sigma_l^2)$ and $N(0, \sigma_\delta^2)$ are chosen for β_l and δ_0 , and the “scale-flat” prior

$N(1, \sigma_\rho^2)$ for ρ_0 . The prior for each correlation parameters in θ_3 is $G(2, 0.1)$, having high variance of 200.

Calculation of $\hat{\phi}_\rho$ and $\hat{\phi}_\delta$ needs solving a stochastic program in (20) with one random variable. To achieve good approximation to the one-dimensional integration, the Monte Carlo sample size S in (21) is fixed at 100. Estimated posterior modes of the correlation parameters are given as: $\phi_l = (0.47, 0.56, 0.04, 0.07)$, $\phi_\rho = (0.20, 0.06, 0.04, 0.07)$, $\phi_\delta = (1.69, 1.82, 0.31, 0.08)$.

The intensive Bayesian computation is implemented in *WinBugs*, a general-purpose Bayesian computing environment. It is found that convergence of Markov Chain is achieved after the first 50000 burn-in iterations, which is assessed visually and with the method in Geweke (1992). Additional 100000 runs are then generated for posterior calculations.

Figures 2(a)-(d) show the posteriors of the adjustment parameters ρ_0 , σ_ρ^2 , δ_0 and σ_δ^2 . Their means and 95% credible HPD intervals are given in Table 2. Several interesting observations have emerged. First, the plot for ρ_0 is multi-modal, indicating complex scale change from LE to HE. Second, σ_ρ^2 and σ_δ^2 are relatively small, indicating a good fit of the adjustment model. Third, the average response 21.39 for 24 HE runs is a bit larger than 19.80, the average for the corresponding 24 LE runs. Table 1 shows no consistent pattern in comparing the HE and LE values. This is different from the example in Section 4, where one experiment consistently gives higher values than the other. For the current example, a simple mean comparison analysis will yield little information, whereas the proposed method can unveil complex relationships between LE and HE.

(Figure 2)

(Table 2)

Finally, we compare predictions on twelve untried runs using BHGP models with those from the separate analysis as well as those using the model of Kennedy and O'Hagan (2000) but with assuming the independence of $\delta(\cdot)$ and y_l as in (7), and the method in Qian et al. (2006). The separate analysis builds a Bayesian Gaussian process model using 24 HE runs, while the other three methods fit both the LE and HE data. Note that the HE response for run 18 is very small (4.55), deviating significantly from the others in the data. Judging by the engineering background of this example, it is conceivable that an unsuccessful run was made due to immature termination of the

finite element code. Its prediction from the BHGP model is 10.52, which could be close to the true HE response. Suppressing the results for run 18, the predictions of the remaining 11 runs are given in Table 3. In the table, column 1 gives the corresponding run numbers in Table 1; column 2 gives the y_h values from HE; columns 3-4 give the values of \hat{y}_h of the integrated analysis, the separate analysis, the Qian et al. method, the Kennedy-O'Hagan method, respectively. The SRMSEs (standardized-root-mean-square-errors)

$$\sqrt{\frac{\sum_{i=1}^{11} \{[\hat{y}_h(\mathbf{x}_i) - y_h(\mathbf{x}_i)]/y_h(x_i)\}^2}{11}}$$

for the four methods (in the same order) are 0.08, 0.15, 0.09 and 0.07, respectively. The three methods that fit both LE and HE runs all perform well and give better prediction results than the separate analysis.

4 Example 2: Fluidized Bed Processes

Dewettinck et al. (1999) reported a physical experiment and several associated computer models for predicting the steady-state thermodynamic operation point of a GlattGPC-1 fluidized-bed unit. The base of the unit consists of a screen and an air jump, with coating sprayers at the side of the unit. Reese et al. (2004) proposed a linear model approach to analyze a sample example in Dewettinck et al. The same data will be analyzed using the proposed BHGP models.

Several variables that can potentially affect the steady-state thermodynamic operating point are: fluid velocity of the fluidization air (V_f), temperature of the air from the pump (T_a), flow rate of the coating solution (R_f), temperature of the coating solution (T_s), coating solution dry matter content (M_d), pressure of atomized air (P_a), temperature (T_r) and humidity (H_r).

Dewettinck et al. (1999) considered 28 different process conditions with coating solution used for distilled water (i.e., $M_d = 0$) and the room temperature set at 20°C. As a result, six factors ($H_r, T_r, T_a, R_f, P_a, V_f$) with different values are considered in the analysis. These values are given in Table 4.

(Table 4)

For each factor combination, one physical run ($T_{2,\text{exp}}$) and three computer runs ($T_{2,1}, T_{2,2}$ and $T_{2,3}$) were conducted. The results are given in Table 5.

(Table 5)

There are major differences among the three computational models (see Dewettinck et al. 1999 for details). In summary, $T_{2,3}$, which includes adjustments for heat losses and inlet airflow, is the most accurate (i.e., producing the closest response to $T_{2,\text{exp}}$). The computer model $T_{2,2}$ includes only the adjustment for heat losses. The model $T_{2,1}$ does not adjust for heat losses or inlet airflow and is thus the least accurate.

For illustration, we only synthesize data from the physical experiment and the second computer model $T_{2,2}$, which has medium accuracy. The responses from $T_{2,\text{exp}}$ and $T_{2,2}$ are denoted by $y_{2,\text{exp}}$ and $y_{2,2}$ respectively.

4.1 The Analysis

It is clear from Table 4 that the six process variables have different scales and should be standardized. The data set is divided into a training set and a testing set. The training set, used to build BHGP models, consists of 20 randomly sampled $T_{2,\text{exp}}$ runs and all 28 $T_{2,2}$ runs. The remaining eight $T_{2,\text{exp}}$ runs (i.e., run no. 4, 15, 17, 21, 23, 25, 26 and 28 as in Table 4) are left to form the testing set for model validation.

As stated in Reese et al., a full second-order linear model is saturated for the data from $T_{2,2}$, given its relatively small run size. As a solution, they implemented a Bayesian variable selection procedure (Wu and Hamada 2000) to find several “most likely” sub-models. Instead of relying on linear models, the proposed method fits a Gaussian process model including all model parameters (mean and correlation parameters) at once, thus avoiding the complex sub-model selection procedure.

The values of the hyper-parameters used in this example are chosen as follows: $(\alpha_l, \gamma_l, \alpha_\rho, \gamma_\rho, \alpha_\delta, \gamma_\delta, \alpha_\epsilon, \gamma_\epsilon) = (2, 1, 2, 1, 2, 1)$, $\mathbf{u}_l = (0, 0, 0, 0, 0, 0, 0)^t$, $v_l = 1$, $(u_\rho, v_\rho, u_\delta, v_\delta) = (1, 1, 0, 1)$, $(a_l, b_l, a_\rho, b_\rho, a_\delta, b_\delta) = (2, 0.1, 2, 0.1, 2, 0.1)$.

Because little knowledge about model parameters is known beforehand, “vague” priors are chosen. The priors for $\sigma_l^2, \sigma_\rho^2, \sigma_\delta^2$ and σ_ϵ^2 are $IG(2, 1)$. The “location-flat” priors $N(0, \mathbf{I}_{7 \times 7} \sigma_l^2)$ and $N(0, \sigma_\delta^2)$ are chosen for β_l and δ_0 , and “scale-flat” prior $N(1, \sigma_\rho^2)$ for ρ_0 . The prior for each correlation parameter is $G(2, 0.1)$, having variance as high as 200.

Calculation of $\hat{\phi}_\rho$ and $\hat{\phi}_\delta$ needs solving a stochastic program in (20) with two random variables. To achieve good approximation for the two-dimensional integration, the Monte Carlo sample size S in (21) is fixed at

200. Estimated correlation parameters are given as:

$$\phi_l = (0.29, 0.16, 0.26, 0.17, 0.22, 0.18), \phi_\rho = (0.32, 0.25, 0.31, 0.09, 0.08, 0.11), \\ \phi_\delta = (0.09, 0.19, 0.11, 0.14, 0.15, 0.05).$$

The intensive Bayesian computation is implemented in *WinBugs*. Convergence of Markov Chain is achieved after the first 50000 burn-in iterations which is assessed visually and with the method in Geweke (1992). Additional 100000 runs are then generated for posterior calculations.

The posterior mean of β_0 is 35.59. The means and 95% credible HPD intervals are shown in Table 6. These intervals are relatively large and contain zero. If these results were obtained from a linear model, we would suspect that some of these effects may not be statistically significant and further analysis is needed to remove insignificant ones from the model. Because a Gaussian process model has a simple mean structure (i.e., including linear effects only), any further simplification of the mean part will yield little benefit. Furthermore, for a Gaussian process model the complex relationship between the inputs and the response is primarily explained by the correlation structure rather than the mean structure. Therefore, all the linear coefficients are retained in the model. The posterior mean 0.34 of the measurement error σ_ϵ^2 is relatively small with standard deviation 0.19.

(Table 6)

The posteriors of ρ_0 , σ_ρ^2 , δ_0 and σ_δ^2 , associated with the adjustments are shown in Figures 3(a)-(d). The means and 95% credible HPD intervals are given in Table 7. The results indicate several important and appealing aspects of the integrated analysis. First, the density plot of ρ_0 has three modes, implying an *intricate scale change* from $y_{2,2}$ to $y_{2,\text{exp}}$. Any attempt to simplify the scale term to a constant will fail to model this change adequately. The capability of modeling complex scale change comes as a benefit of the Bayesian formulation. A frequentist's analysis can only produce a point estimate for ρ_0 unless a complicated mixture model is correctly employed and asymptotic distributions are obtained. Second, σ_ρ^2 and σ_δ^2 in Table 7 are relatively small, in part indicating that the two data sources are well integrated in the analysis. Third, the average response 42.33 for 20 $T_{2,\text{exp}}$ runs is lower than 44.22, the average for the corresponding 20 $T_{2,2}$ runs. This observation comes as no surprise, as for each run in Table 5, $T_{2,\text{exp}}$ consistently produces a lower response than $T_{2,2}$. On average, the difference is -1.89. However, the posterior mean for δ_0 is -0.03, which is much smaller than

-1.89 in magnitude. This is due to the inclusion of scale adjustment in the model. The scale change may be of significant interest to the experimenters and this treasured information is uncovered by the proposed analysis.

(Figure 3)

(Table 7)

Finally, we assess the prediction accuracy of the proposed method by comparing it with that of a separate analysis. For the latter, 20 T_{exp} runs are used to fit a Bayesian Gaussian process model. Table 8 lists the prediction results on eight untried points with column 1 giving the run no's of the testing runs and columns 2, 3, 4 giving the values of $\hat{y}_{2,\text{exp}}$ from the integrated analysis, $\hat{y}_{2,\text{exp}}$ from the separate analysis, the values of $y_{2,\text{exp}}$, respectively. In general, the integrated analysis produces better results. The integrated analysis significantly improves prediction over the separate analysis since the SRMSE (standardized-root-mean-square-error) (0.02) of the integrated analysis is much smaller than the SRMSE (0.21) of the separate analysis.

Table 8

5 Concluding Remarks and Extensions

This paper has developed some hierarchical Gaussian process models for modeling and integrating LE and HE data. Use of the adjustment model in (7) allows a flexible location and scale adjustment of the more abundant but less accurate LE data to be closer to the HE data. Use of MCMC and Sample Average Approximation algorithms makes it feasible to carry out the Bayesian computation. By using the Bayesian predictive density in (11), the prediction can incorporate uncertainties in the model parameters. As demonstrated by the results in Sections 3 and 4, the proposed method can better account for the heterogeneity between the LE and HE data.

Extensions of the present work can be made in several directions. First, the Bayesian prediction in (11) uses a point estimate of the correlation parameters θ_3 , which is developed in Section 2.5. Following Higdon et al. (2004), a fully Bayesian approach can be developed that computes the posterior of θ_3 for the prediction. One possibility is to modify the conditional sampling method in (14) by including a step to sample from $p(\theta_3|y_l, \mathbf{y}_h, \bar{\theta}_3)$ given in

(10). Since this function has a complicated form, efficient MCMC algorithms need to be developed. Second, it is possible to extend the proposed methodology to incorporate calibration parameters. Denote by $\mathbf{x} = (x_1, \dots, x_k)$ the input variables of the LE that can be measured or changed in the HE. Denote by $\mathbf{t} = (t_1, \dots, t_q)$ the input variables of the LE that are related to calibration and cannot be observed or manipulated in the HE. For calibration, the variables $\mathbf{t} = (t_1, \dots, t_q)$ correspond to q -dimensional variables, $\boldsymbol{\tau}$, in the HE. Following Kennedy and O’Hagan (2001), we can extend the HE and LE models in (5) and (7) to accommodate the calibration parameters as follows:

$$\begin{aligned} y_l(\mathbf{x}, \mathbf{t}) &= \kappa(\mathbf{x}, \mathbf{t}), \\ y_h(\mathbf{x}) &= \rho(\mathbf{x})\kappa(\mathbf{x}, \boldsymbol{\tau}) + \delta(\mathbf{x}) + \epsilon(\mathbf{x}), \end{aligned}$$

where $\kappa(\mathbf{x}, \mathbf{t})$ is assumed to be a Gaussian process in the space of \mathbf{x} and \mathbf{t} . For this model, we can modify the estimation procedure in Section 2.5 to estimate $\boldsymbol{\theta}_3$ and $\boldsymbol{\tau}$, and then, conditional on these estimates, use a modified version of the prediction procedure in Section 2.4 to predict the response of the HE at untried design points. Following Goldstein and Rougier (2004), we can further extend our method to situations where the input variables for the LE and HE are in different spaces. Third, the proposed method can be extended to more than two sources like low-, medium- and high-accuracy experiments in a relatively straightforward way.

Acknowledgments

Qian was supported by NSF grant DMS 0705206. Wu was supported by US ARO grant W911NF-05-1-0264 and NSF grants DMI 0620259 and DMS 0705261. We are grateful to C. C. Seepersad and J. K. Allen for their assistance and insight on the linear cellular alloy example, and to Kwok Tsui, Shuchun Wang and Tom Santer for useful comments. We are also grateful to the editor, the associate editor and two referees for very insightful comments, which led to significant improvements of the paper.

References

- [1] Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J. and Walsh, D.

- (2007), “Computer Model Validation with Functional Output,” *Annals of Statistics*, to appear.
- [2] Berger, J. O., Oliveira, V. D. and Sansó, B. (2001), “Objective Bayesian Analysis of Spatially Correlated Data,” *Journal of the American Statistical Association*, **96**, 1361-1374.
- [3] Dewettinck, K., Visscher, A. D., Deroo, L. and Huyghbaet, A. (1999), “Modeling the Steady-State Thermodynamic Operation Point of Top-spray Fluidized Bed Processing,” *Journal of Food Engineering*, **39**, 131-143.
- [4] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. R. (2004), *Bayesian Data Analysis*, Boca Raton: Chapman & Hall/CRC.
- [5] Geweke, J. (1992), “Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments,” *Bayesian Statistics 4*, Oxford, U.K.: Oxford University Press, 169-193.
- [6] Goldstein, M. and Rougier, J. C. (2004), “Probabilistic Formulations for Transferring Inferences from Mathematical Models to Physical Systems,” *SIAM Journal on Scientific Computing*, **26**, 467-487.
- [7] Handcock, M. S. and Stein, M. L. (1993), “A Bayesian Analysis of Kriging,” *Technometrics*, **35**, 403-410.
- [8] Handcock, M. S. and Wallis, J. R. (1994), “An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields,” *Journal of the American Statistical Association*, **89**, 368-378.
- [9] Higdon, D., Kennedy, M. C., Cavendish, J. C., Cafoe, J. A. and Ryne, R. D. (2004), “Combining Field Data and Computer Simulations for Calibration and Prediction,” *SIAM Journal on Scientific Computing*, **26**, 448-466.
- [10] Joseph, R. V., Hung, Y. and Sudjianto, A. (2007), “Blind Kriging: A New Method for Developing Metamodels,” *ASME Journal of Mechanical Design*, to appear.
- [11] Kennedy, M. C. and O’Hagan (2000), “Predicting the Output from a Complex Computer Code When Fast Approximations are Available,” *Biometrika*, **87**, 1133-1152.

- [12] Kennedy, M. C. and O'Hagan, A. (2001) "Bayesian Calibration of Computer Models (with discussion)," *Journal of the Royal Statistical Society, Series B*, **63**, 425-464.
- [13] Kleywegt, A. J., Shapiro, A. and Homem-de-Mello, T. (2001), "The Sample Average Approximation Method for Stochastic Discrete Optimization," *SIAM Journal on Optimization*, **12**, 479-502.
- [14] Koopmans, T. C. and Reiersøl, O. (1950), "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, **21**, 165-181.
- [15] Linderoth, J., Shapiro, A. and Wright, S. (2006), "The Empirical Behavior of Sampling Methods for Stochastic programming," *Annals of Operations Research*, **142**, 215-241.
- [16] Liu, J. S. (2001), *Monte Carlo Strategies in Scientific Computing*, New York: Springer.
- [17] Lu, N. D., Sun, W. and Zidek, J. V. (1997), "Bayesian Multivariate Spatial Interpolation with Data Missing by Design," *Journal of Royal Statistical Society Series B*, **59**, 501-510.
- [18] Nagy, B. (2006), "Fast Bayesian Implementation (FBI) of Gaussian Process Regression," Presentation at Workshop on Complex Computer Models, Simon Fraser University, August 11-16, 2006.
- [19] Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K. and Wu, C. F. J. (2006), "Building Surrogate Models Based on Detailed and Approximate Simulations," *ASME Journal of Mechanical Design*, **128**, 668-677.
- [20] Reese, C. S., Wilson, A. G, Hamada, M., Martz, H. F. and Ryan, K. J. (2004), "Integrated Analysis of Computer and Physical Experiments," *Technometrics*, **46**, 153-164.
- [21] Rothenberg, T. J. (1971), "Identification in Parametric Models," *Econometrica*, **39**, 577-591.
- [22] Ruszczyński, A. and Shapiro A.(eds) (2003), *Stochastic Programming. Handbooks in Operations Research and Management Science*, **10**, Elsevier.

- [23] Santner, T. J., Williams, B. J. and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer.
- [24] Shapiro, A. and Homem-de-Mello, T. (2000), “On Rate of Convergence of Monte Carlo Approximations of Stochastic Programs,” *SIAM Journal on Optimization*, **11**, 70-86.
- [25] Shapiro, A. and Nemirovski, A. (2005), “On Complexity of Stochastic Programming Problems,” *Continuous Optimization: Current Trends and Applications*, 111-144, V. Jeyakumar and A.M. Rubinov (Eds.), Springer.
- [26] Tanner, M. and Wong. W. (1987) “The Calculation of Posterior Distributions by Data Augmentation,” *Journal of the American Statistical Association*, **82**, 528-540.
- [27] Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992), “Screening, Predicting and Computer Experiments,” *Technometrics*, **34**, 15-25.
- [28] Williams, B. J., Santner, T. J. and Notz, W. I. (2000), “Sequential Design of Computer Experiments to Minimize Integrated Response Functions,” *Statistica Sinica*, **10**, 1133-1152.
- [29] Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis and Parameter Design Optimization*, New York: John Wiley and Son.

Appendix

Proof of (19)

Recall from (18) that

$$p(\boldsymbol{\theta}_3 | \mathbf{y}_l, \mathbf{y}_h) \propto \int_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2} p(\boldsymbol{\theta}_3) p(\boldsymbol{\theta}_2) p(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2) p(\mathbf{y}_l, \mathbf{y}_h | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2. \quad (\text{A1})$$

Perform the integration in (A1) in the following two steps:

1. Integrate out β_l , ρ_0 and δ_0 ;
2. Integrate out σ_l^2 and σ_ρ^2 .

After perform the two steps, (A1) can be simplified to an integral involving τ_1 and τ_2 only.

Step 1: Integrate out β_l , ρ_0 and δ_0 .

Perform the integration

$$\begin{aligned} & \int_{\beta_l, \rho_0, \delta_0} p(\boldsymbol{\theta}_3)p(\boldsymbol{\theta}_2)p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p(\mathbf{y}_l, \mathbf{y}_h|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)d\beta_l d\rho_0 d\delta_0 \\ &= p(\boldsymbol{\theta}_3)p(\boldsymbol{\theta}_2) \int_{\beta_l, \rho_0, \delta_0} p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p(\mathbf{y}_l, \mathbf{y}_h|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)d\beta_l d\rho_0 d\delta_0 \end{aligned} \quad (\text{A2})$$

by integrating out β_l , ρ_0 and δ_0 one by one.

(a) Integrate out β_l .

$$\begin{aligned} & \int_{\beta_l} p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p(\mathbf{y}_l, \mathbf{y}_h|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)d\beta_l \\ & \propto (\sigma_l^2)^{-\frac{n}{2}} (\sigma_\rho^2)^{-\frac{n_1+2}{2}} \tau_1^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\rho_0 - u_\rho)^2}{2v_\rho \sigma_\rho^2} - \frac{(\delta_0 - u_\delta)^2}{2v_\delta \tau_1 \sigma_\rho^2} \right\} \\ & \quad \cdot \exp \left\{ -\frac{(\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})}{2\sigma_\rho^2} - \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8\sigma_l^2} \right\}, \end{aligned} \quad (\text{A3})$$

where

$$\mathbf{a}_1 = v_l^{-1} \mathbf{I}_{(k+1) \times (k+1)} + \mathbf{F}_l^t \mathbf{R}_l^{-1} \mathbf{F}_l, \quad \mathbf{b}_1 = -2v_l^{-1} \mathbf{u}_l - 2\mathbf{F}_l^t \mathbf{R}_l^{-1} \mathbf{y}_l, \quad c_1 = v_l^{-1} (\mathbf{u}_l^t \mathbf{u}_l) + \mathbf{y}_l^t \mathbf{R}_l^{-1} \mathbf{y}_l.$$

(b) Integrate out ρ_0 .

$$\begin{aligned} & \int_{\rho_0} (\sigma_l^2)^{-\frac{n}{2}} (\sigma_\rho^2)^{-\frac{n_1+2}{2}} \tau_1^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\rho_0 - u_\rho)^2}{2v_\rho \sigma_\rho^2} - \frac{(\delta_0 - u_\delta)^2}{2v_\delta \tau_1 \sigma_\rho^2} \right\} \\ & \quad \cdot \exp \left\{ -\frac{(\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \rho_0 \mathbf{y}_{l_1} - \delta_0 \mathbf{1}_{n_1})}{2\sigma_\rho^2} - \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8\sigma_l^2} \right\} d\rho_0 \\ & \propto (a_2)^{-\frac{1}{2}} (\sigma_l^2)^{-\frac{n}{2}} (\sigma_\rho^2)^{-\frac{n_1+1}{2}} \tau_1^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\delta_0 - u_\delta)^2}{2v_\delta \tau_1 \sigma_\rho^2} - \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8\sigma_l^2} \right\} \\ & \quad \cdot \exp \left\{ -\frac{t_1 \delta_0^2 + t_2 \delta_0 + t_3}{2a_2 \sigma_\rho^2} \right\}, \end{aligned} \quad (\text{A4})$$

where

$$\begin{aligned}
a_2 &= v_\rho^{-1} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_{l_1}, \quad b_2 = -2u_\rho v_\rho^{-1} - 2\mathbf{y}_{l_1}^t \mathbf{M}^{-1} (\mathbf{y}_h - \delta_0 \mathbf{1}_{n_1}), \\
c_2 &= u_\rho^2 v_\rho^{-1} + (\mathbf{y}_h - \delta_0 \mathbf{1}_{n_1})^t \mathbf{M}^{-1} (\mathbf{y}_h - \delta_0 \mathbf{1}_{n_1}), \\
t_1 &= (v_\rho^{-1} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_{l_1}) (\mathbf{1}_{n_1}^t \mathbf{M}^{-1} \mathbf{1}_{n_1}) - (\mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{1}_{n_1})^2, \\
t_2 &= -2[(v_\rho^{-1} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_{l_1}) (\mathbf{1}_{n_1}^t \mathbf{M}^{-1} \mathbf{y}_h) - (u_\rho v_\rho^{-1} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_h) (\mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{1}_{n_1})], \\
t_3 &= (v_\rho^{-1} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_{l_1}) (u_\rho^2 v_\rho^{-1} + \mathbf{y}_h^t \mathbf{M}^{-1} \mathbf{y}_h) - (u_\rho v_\rho^{-1} + \mathbf{y}_{l_1}^t \mathbf{M}^{-1} \mathbf{y}_h)^2.
\end{aligned}$$

(c) Integrate out δ_0 .

$$\begin{aligned}
& \int_{\delta_0} (a_2)^{-\frac{1}{2}} (\sigma_l^2)^{-\frac{n}{2}} (\sigma_\rho^2)^{-\frac{n_1+1}{2}} \tau_1^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{(\delta_0 - u_\delta)^2}{2v_\delta \tau_1 \sigma_\rho^2} - \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8\sigma_l^2} \right\} \\
& \quad \cdot \exp \left\{ -\frac{t_1 \delta_0^2 + t_2 \delta_0 + t_3}{2a_2 \sigma_\rho^2} \right\} d\delta_0 \\
& \propto (a_2 a_3)^{-\frac{1}{2}} (\sigma_l^2)^{-\frac{n}{2}} (\sigma_\rho^2)^{-\frac{n_1}{2}} \tau_1^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} \exp \left\{ -\frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8\sigma_l^2} - \frac{4a_3 c_3 - b_3^2}{8\sigma_\rho^2 a_3} \right\}, \tag{A5}
\end{aligned}$$

where

$$a_3 = (v_\delta \tau_1)^{-1} + t_1 a_2^{-1}, \quad b_3 = -2u_\delta (v_\delta \tau_1)^{-1} + t_2 a_2^{-1}, \quad c_3 = u_\delta^2 (v_\delta \tau_1)^{-1} + t_3 a_2^{-1}.$$

Step 2: Integrate out σ_l^2 and σ_ρ^2 .

$$\begin{aligned}
& \int_{\sigma_l^2, \sigma_\rho^2} p(\boldsymbol{\theta}_2) (a_2 a_3)^{-\frac{1}{2}} (\sigma_l^2)^{-\frac{n}{2}} (\sigma_\rho^2)^{-\frac{n_1}{2}} \tau_1^{-\frac{1}{2}} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} \\
& \quad \cdot \exp \left\{ -\frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8\sigma_l^2} - \frac{4a_3 c_3 - b_3^2}{8\sigma_\rho^2 a_3} \right\} d\sigma_l^2 d\sigma_\rho^2 \\
& \propto \tau_1^{-(\alpha_\delta + \frac{3}{2})} \tau_2^{-(\alpha_\epsilon + 1)} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} (a_2 a_3)^{-\frac{1}{2}} \left(\gamma_l + \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8} \right)^{-(\alpha_l + \frac{n}{2})} \\
& \quad \cdot \left(\gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\epsilon}{\tau_2} + \frac{4a_3 c_3 - b_3^2}{8a_3} \right)^{-(\alpha_\rho + \alpha_\delta + \alpha_\epsilon + \frac{n_1}{2})}. \tag{A6}
\end{aligned}$$

Therefore, (A1) is proportional to

$$\begin{aligned}
 p(\boldsymbol{\theta}_3) & \int_{\tau_1, \tau_2} \tau_1^{-(\alpha_\delta + \frac{3}{2})} \tau_2^{-(\alpha_\epsilon + 1)} |\mathbf{a}_1|^{-\frac{1}{2}} |\mathbf{R}_l|^{-\frac{1}{2}} |\mathbf{M}|^{-\frac{1}{2}} (a_2 a_3)^{-\frac{1}{2}} \\
 & \cdot \left(\gamma_l + \frac{4c_1 - \mathbf{b}_1^t \mathbf{a}_1^{-1} \mathbf{b}_1}{8} \right)^{-(\alpha_l + \frac{n}{2})} \left(\gamma_\rho + \frac{\gamma_\delta}{\tau_1} + \frac{\gamma_\epsilon}{\tau_2} + \frac{4a_3 c_3 - b_3^2}{8a_3} \right)^{-(\alpha_\rho + \alpha_\delta + \alpha_\epsilon + \frac{n+1}{2})} d\tau_1 d\tau_2.
 \end{aligned} \tag{A7}$$

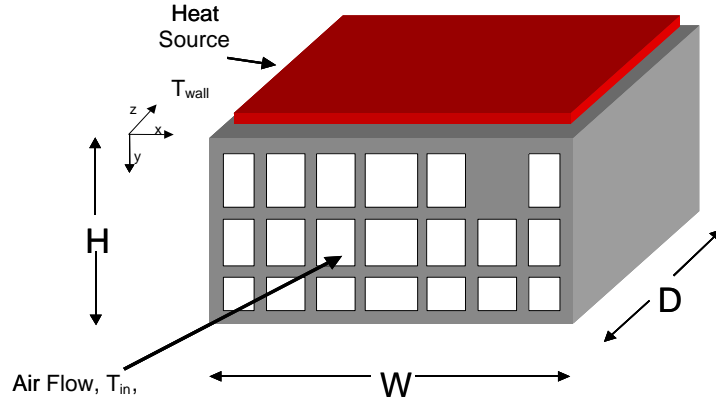


Figure 1: Linear Cellular Alloy in Qian et al. (2006).

Run #	$\dot{m}(kg/s)$	$T_{in}(K)$	$k(W/mk)$	$T_{wall}(K)$	y_l	y_h	Status
1	0.000500	293.15	362.73	393.15	27.24	25.82	Test
2	0.000550	315.00	310.00	365.00	7.02	7.48	Train
3	0.000552	293.53	318.63	388.29	25.61	23.54	Train
4	0.000560	277.01	354.98	374.00	25.53	19.77	Test
5	0.000566	285.77	266.71	367.27	21.23	20.15	Train
6	0.000578	302.17	358.13	343.72	11.44	10.17	Train
7	0.000580	272.26	211.71	333.65	15.03	15.29	Train
8	0.000589	278.16	225.78	351.83	18.55	18.39	Train
9	0.000594	279.54	258.51	360.13	20.74	20.52	Test
10	0.000612	280.83	291.53	394.72	30.22	30.12	Train
11	0.000620	275.00	225.00	340.00	16.40	18.78	Test
12	0.000626	284.89	350.46	352.29	18.13	18.17	Train
13	0.000627	287.60	243.96	382.54	25.02	24.68	Test
14	0.000639	270.45	241.21	341.81	17.92	19.05	Train
15	0.000643	276.17	216.99	371.60	24.20	24.96	Train
16	0.000652	298.04	303.96	361.58	17.47	16.95	Train
17	0.000657	294.24	330.63	375.53	22.48	22.30	Test
18	0.000680	313.28	259.12	350.00	10.23	4.55	Test
19	0.000700	288.15	300.00	400.00	30.90	34.45	Train
20	0.000751	287.99	326.02	354.08	18.17	19.57	Train
21	0.000763	292.82	254.84	373.38	21.96	23.33	Test
22	0.000780	292.73	267.84	369.00	20.92	21.97	Train
23	0.000800	303.15	250.00	350.00	13.08	14.83	Train
24	0.000814	286.39	339.92	332.40	12.68	14.36	Train
25	0.000842	294.39	203.45	346.05	13.75	15.12	Train
26	0.000850	270.00	325.00	385.00	31.14	32.85	Test
27	0.000850	301.31	317.85	341.00	11.30	11.92	Train
28	0.000851	273.71	315.27	381.14	29.08	34.80	Test
29	0.000857	282.12	262.30	350.10	18.25	21.31	Train
30	0.000874	282.50	253.25	396.36	30.90	36.11	Test
31	0.000882	299.22	288.45	385.07	24.45	27.36	Test
32	0.000903	284.25	290.90	364.99	22.22	25.37	Train
33	0.000910	248.87	206.74	398.00	36.56	47.05	Train
34	0.000940	271.32	362.73	400.00	35.53	42.93	Train
35	0.000950	280.00	270.00	330.00	13.54	17.41	Train
36	0.001000	293.15	202.40	373.15	21.60	22.89	Train

Table 1: Data from Linear Cellular Alloy Experiment

Parameter	Posterior mean	Lower Bound	Upper Bound
ρ_0	0.95	0.70	1.18
σ_ρ^2	0.40	0.14	1.05
δ_0	0.81	-0.43	2.22
σ_δ^2	1.70	0.72	3.38

Table 2: Posterior Means and 95% Credible HPD Intervals for ρ_0 , σ_ρ^2 , δ_0 and σ_δ^2 in Linear Cellular Alloy Experiment

Run	y_h	$\hat{y}_{h,BHGP}$	$\hat{y}_{h,SEP}$	$\hat{y}_{h,QIAN}$	$\hat{y}_{h,KO}$
1	25.82	23.57	23.22	24.20	28.66
4	19.77	23.61	26.67	25.00	22.97
9	20.52	20.20	22.26	20.46	20.24
11	18.78	16.81	16.32	17.29	17.34
13	24.68	26.01	23.76	24.82	25.86
17	22.30	22.76	20.32	21.90	21.76
21	23.33	23.18	21.65	22.67	22.94
26	32.85	37.07	34.38	35.80	33.85
28	34.80	34.33	31.75	33.28	31.89
30	36.11	35.90	31.10	35.86	34.87
31	27.36	26.04	21.36	26.15	25.49

Table 3: Prediction Results on 11 Untried Points for Linear Cellular Alloy Experiment

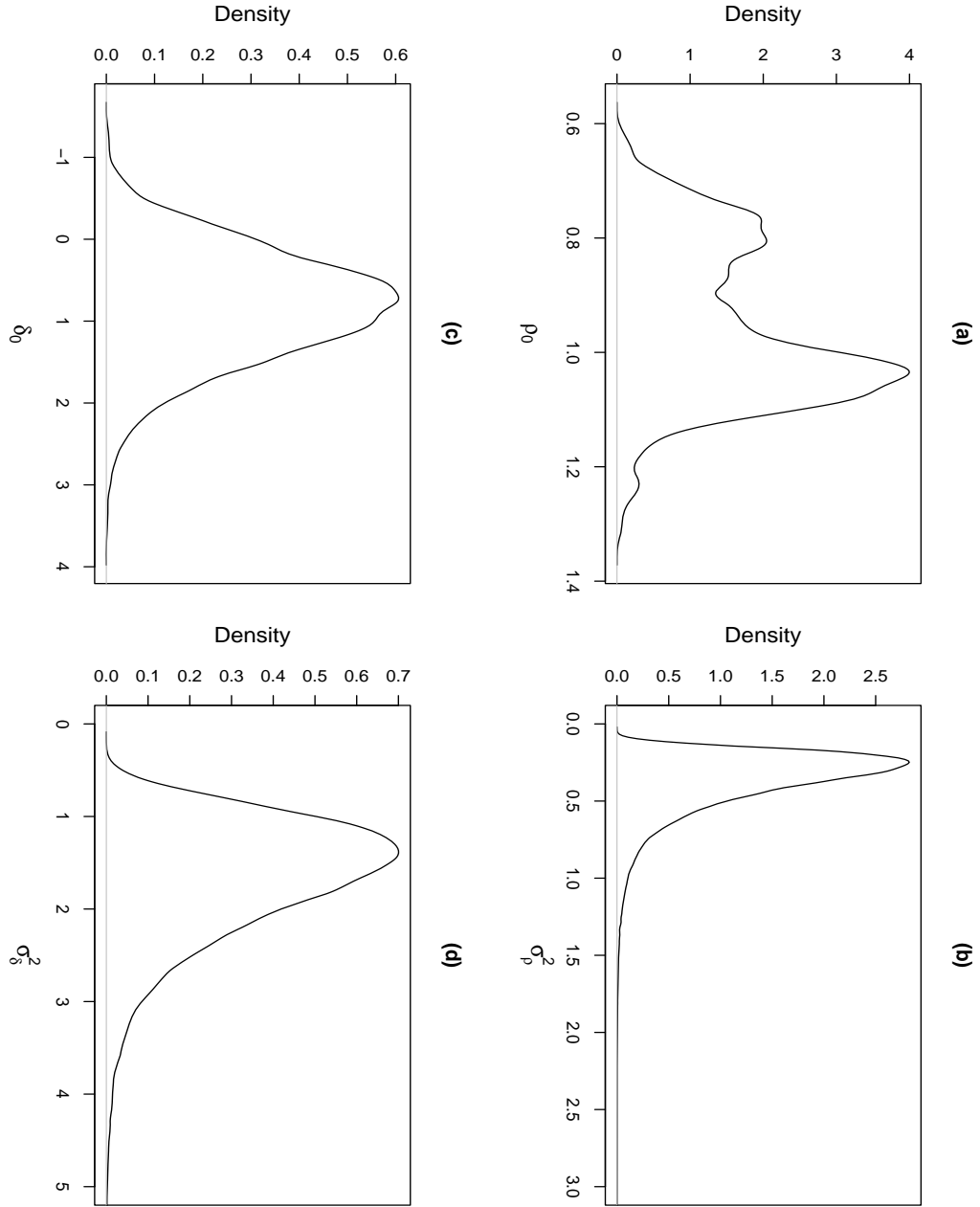


Figure 2: Posteriors of ρ_0 , σ_ρ^2 , δ_0 and σ_δ^2 for Linear Cellular Alloy Experiment.

Run #	$H_r(\%)$	$T_r(C)$	$T_a(C)$	$R_f(g/min)$	$P_a(bar)$	$V_f(m/s)$
1	51.00	20.70	50.00	5.52	2.50	3.00
2	46.40	21.30	60.00	5.53	2.50	3.00
3	46.60	19.20	70.00	5.53	2.50	3.00
4	53.10	21.10	80.00	5.51	2.50	3.00
5	52.00	20.40	90.00	5.21	2.50	3.00
6	45.60	21.40	60.00	7.25	2.50	3.00
7	47.30	19.50	70.00	7.23	2.50	3.00
8	53.30	21.40	80.00	7.23	2.50	3.00
9	44.00	20.10	70.00	8.93	2.50	3.00
10	52.30	21.60	80.00	8.91	2.50	3.00
11	55.00	20.20	80.00	7.57	1.00	3.00
12	54.00	20.60	80.00	7.58	1.50	3.00
13	50.80	21.10	80.00	7.40	2.00	3.00
14	48.00	21.20	80.00	7.43	2.50	3.00
15	42.80	22.40	80.00	7.51	3.00	3.00
16	55.70	20.80	50.00	3.17	1.00	3.00
17	55.20	20.70	50.00	3.18	1.50	3.00
18	54.40	20.70	50.00	3.19	2.00	3.00
19	55.40	19.80	50.00	3.20	2.50	3.00
20	52.90	20.00	50.00	3.19	3.00	3.00
21	28.50	18.30	80.00	7.66	2.50	3.00
22	26.10	19.00	80.00	7.69	2.50	4.00
23	24.20	18.90	80.00	7.69	2.50	4.50
24	25.40	18.50	80.00	7.70	2.50	5.00
25	45.10	19.60	50.00	3.20	2.50	3.00
26	43.10	20.30	50.00	3.23	2.50	4.00
27	42.70	20.40	50.00	3.20	2.50	4.50
28	38.70	21.60	50.00	3.22	2.50	5.00

Table 4: Six Process Variables for Fluidized Bed Process Experiment

Run#	$T_{2,exp}$	$T_{2,1}$	$T_{2,2}$	$T_{2,3}$
1	30.40	32.40	31.50	30.20
2	37.60	39.50	38.50	37.00
3	45.10	46.80	45.50	43.70
4	50.20	53.80	52.60	51.00
5	57.90	61.70	59.90	58.20
6	32.90	35.20	34.60	32.60
7	39.50	42.40	41.00	39.10
8	45.60	49.50	48.50	46.40
9	34.20	37.50	36.60	34.80
10	41.10	45.50	44.30	42.00
11	45.70	50.50	49.00	47.00
12	44.60	49.80	48.40	46.30
13	44.70	49.80	48.40	46.30
14	44.00	49.20	48.00	45.70
15	43.30	48.60	47.50	45.40
16	37.00	39.50	38.00	37.70
17	37.20	39.50	38.50	37.10
18	37.10	39.50	37.50	36.70
19	36.90	39.50	38.50	36.10
20	36.80	37.70	37.20	36.20
21	46.00	48.70	47.30	45.10
22	54.70	57.70	56.20	54.20
23	57.00	60.10	58.70	57.00
24	58.90	62.00	60.50	58.70
25	35.90	37.90	37.10	36.10
26	40.30	41.70	40.80	40.10
27	41.90	43.00	42.30	41.40
28	43.10	43.90	43.30	42.60

Table 5: Results from Fluidized Bed Process Experiment

	Posterior mean	Lower Bound	Upper Bound
β_{l_1}	0.09	-0.31	0.50
β_{l_2}	0.28	-0.10	0.654
β_{l_3}	-0.08	-0.40	0.23
β_{l_4}	-0.13	-0.59	0.34
β_{l_5}	-0.02	-0.69	0.65
β_{l_6}	0.15	-0.32	0.62

Table 6: Posterior Means and 95% Credible Intervals for β_l in Fluidized Bed Process Experiment

Parameter	Posterior Mean	Lower Bound	Upper Bound
ρ_0	1.07	0.74	1.40
σ_ρ^2	0.12	0.06	0.22
δ_0	-0.03	-1.95	1.83
σ_δ^2	2.05	0.28	5.67

Table 7: Posterior Means and 95% Credible HPD Intervals for ρ_0 , σ_ρ^2 , δ_0 and σ_δ^2 in Fluidized Bed Process Experiment

Run #	\hat{y}_h (integrated analysis)	\hat{y}_h (separate analysis)	y_h
4	50.19	38.34	50.20
15	43.10	26.42	43.30
17	37.64	30.00	37.20
21	47.51	48.41	46.00
23	56.76	50.24	57.00
25	37.00	34.65	35.90
26	40.67	34.26	40.30
28	42.47	31.72	43.10

Table 8: Prediction Results on Eight Untried Points for Fluidized Bed Process Experiment

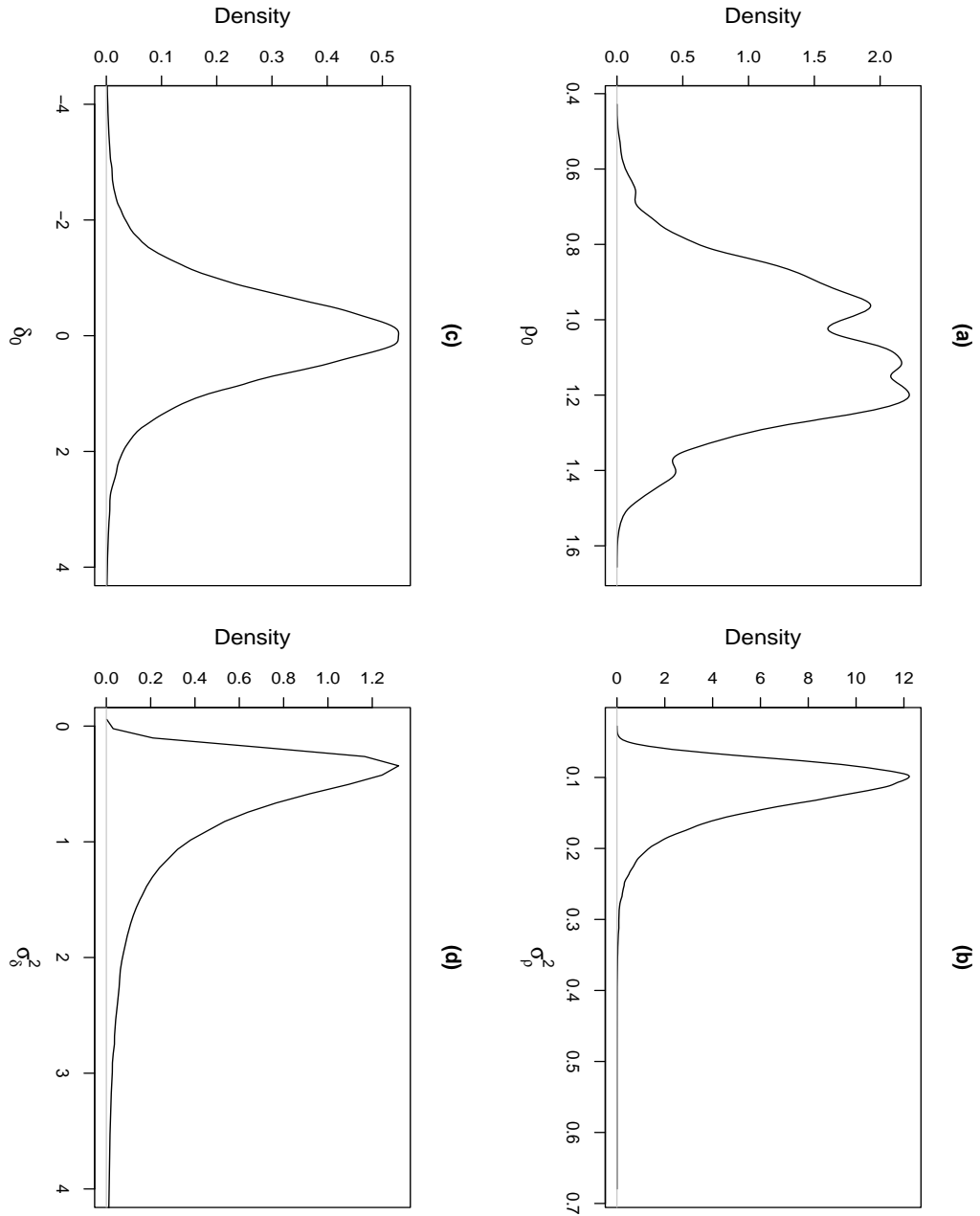


Figure 3: Posterior of ρ_0 , σ_ρ^2 , δ_0 and σ_δ^2 for Fluidized Bed Process Experiment.