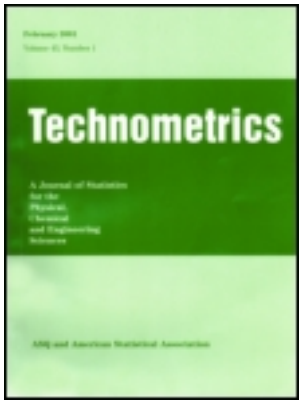

Technometrics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utch20

Integrated Analysis of Computer and Physical Experiments

C. Shane Reese[a], Alyson G Wilson[b], Michael Hamada[b], Harry F Martz[b] & Kenneth J Ryan[c]

[a] Department of Statistics, Brigham Young University, Provo, UT 84602

[b] Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, NM 87545

[c] Department of Statistics, University of Illinois, Chicago, IL 60208

Published online: 01 Jan 2012.

To cite this article: C. Shane Reese, Alyson G Wilson, Michael Hamada, Harry F Martz & Kenneth J Ryan (2004) Integrated Analysis of Computer and Physical Experiments, Technometrics, 46:2, 153-164, DOI: 10.1198/004017004000000211

To link to this article: http://dx.doi.org/10.1198/004017004000000211

Integrated Analysis of Computer and Physical Experiments

C. Shane REESE

Department of Statistics
Brigham Young University
Provo, UT 84602
(*reese@stat.byu.edu*)

Alyson G. WILSON, Michael HAMADA, and Harry F. MARTZ

Statistical Sciences Group
Los Alamos National Laboratory
Los Alamos, NM 87545

Kenneth J. RYAN

Department of Statistics
University of Illinois
Chicago, IL 60208

Scientific investigations frequently involve data from computer experiment(s) as well as related physical experimental data on the same factors and related response variable(s). There may also be one or more expert opinions regarding the response of interest. Traditional statistical approaches consider each of these datasets separately with corresponding separate analyses and fitted statistical models. A compelling argument can be made that better, more precise statistical models can be obtained if the combined data are analyzed simultaneously using a hierarchical Bayesian integrated modeling approach. However, such an integrated approach must recognize important differences, such as possible biases, in these experiments and expert opinions. We illustrate our proposed integrated methodology by using it to model the thermodynamic operation point of a top-spray fluidized bed microencapsulation processing unit. Such units are used in the food industry to tune the effect of functional ingredients and additives. An important thermodynamic response variable of interest, **Y**, is the steady-state outlet air temperature. In addition to a set of physical experimental observations involving six factors used to predict **Y**, similar results from three different computer models are also available. The integrated data from the physical experiment and the three computer models are used to fit an appropriate response surface (regression) model for predicting **Y**.

KEY WORDS: Bayesian hierarchical models; Calibration; Regression.

1. INTRODUCTION

Computer models are often used to perform experiments before expensive physical experiments are undertaken. The computer models attempt to reproduce the physical properties of a process by mathematically representing the individual physical subprocesses. For example, in the food industry, fluidized-bed (or air-suspension) processes are increasingly used to coat food particles with preservatives and flavor enhancers. Some of the physical principles that govern the operation of fluidized beds are fairly well understood (e.g., heat transfer and fluid flow), but others are less well characterized. As a result, computer models based on these thermodynamic principles of physics are constructed that resemble and simulate the actual physical process. In this article we analyze data collected from three such computer models (with each model accounting for different effects), as well as data collected from a corresponding physical experiment. We consider this example further in Section 3.

It is statistically efficient and desirable to fit a single common response surface model that combines the physical experimental data and the computer model output data to express the relationship between the factors and the response variable. Although the response variables of interest in the computer and physical experiments may not be the same, we assume that they can be related by a known transfer function. Thus we effectively consider the same response variable in both types of experiments. However, the computed (or measured) value of the response variable need not be considered at the same factor values in both experiments. We require only that there exist some common set of factors (either all or at least some) for both experiments (see Sec. 2.3). For example, a broad (screening) computer experiment may be performed first, followed by a physical experiment in a smaller region of particular interest (perhaps a corner) of the overall computer experiment design space.

In addition, one or more expert opinions may be available regarding the response variable of interest. Traditional statistical approaches consider each of these datasets separately with corresponding separate designs, analyses, and results. A compelling argument can be made that better, more powerful statistical results can be obtained if we simultaneously analyze the combined data using a recursive Bayesian hierarchical model (RBHM) that we propose in Section 2. As we illustrate, the simultaneous analysis of such combined data permits the unknown coefficients in an assumed overall regression (or response surface) model to be estimated more precisely, thereby producing a better-fitting response surface.

In Section 2 we present the methodology, including our implementation of the RBHM. In Section 3 we describe the mechanics and process variables involved in the fluidized bed example and the experiment from which the data arise. We apply the RBHM methodology to the fluidized bed study and present the resulting response surface in Section 3.3. We discuss sensitivity to prior specification in Section 4 and the results and methodology in Section 5.

TECHNOMETRICS, MAY 2004, VOL. 46, NO. 2
DOI 10.1198/004017004000000211

2. DATA INTEGRATION MODEL AND ANALYSIS

Fundamental to Bayesian estimation is the notion and use of prior and posterior distributions. A good elementary discussion of prior and posterior probabilities and distributions was given by Berry (1996). An RBHM provides a convenient way to sequentially combine the data as follows. Initial informative, but diffuse, prior distributions are defined, one for each unknown parameter. Any available expert opinion data that exist are then used to update these priors to form corresponding posterior distributions. This represents stage 1 of the combined analysis. These posteriors then become the prior distributions for the second stage, in which the computer experimental data are used to update these priors to form stage 2 posterior distributions. At stage 2, the posteriors thus represent the combined use of only the expert opinion and computer data. Finally, these posteriors become the priors for stage 3, in which the physical experimental data are used to construct the final desired posteriors. In this way, all available data are used recursively within the context of the model to successively (and more precisely) estimate all of the desired parameters of interest.

The design and analysis of computer experiments has evolved as the power of computers has grown (although it has certainly not kept pace). Sacks, Welch, Mitchell, and Wynn (1989) provided a review of techniques used in the analysis of output from complex computer codes, as well as issues for design. Latin hypercube sampling had its genesis in the design of computer experiments (McKay, Beckman, and Conover 1979). A Bayesian treatment of the design and analysis of computer experiments was presented by Currin, Mitchell, Morris, and Ylvisaker (1991). These authors were concerned primarily with issues when the only source of information is the output from a complex computer model.

Combining multiple sources of information had its genesis in the meta-analytic literature. Zeckhauser (1971) provided an early treatment of meta-analysis, and Hedges and Olkin (1987) provided a nice review of meta-analytic techniques. Meta-analysis has not been viewed without strong criticism (Shapiro 1994). Müller, Parmigiani, Schildkrout, and Tardella (1999) presented a Bayesian hierarchical modeling approach for combining case-control and prospective studies, where effects due to different studies as well as different centers are allowed.

Craig, Goldstein, Rougier, and Scheult (2001) presented an approach to forecasting from computer models that explicitly incorporates two of the data sources that we consider, expert opinion and computer experiments. They considered the possibility of multivariate responses on the computer model (which they called computer simulators). Physical data in the form of historical measurements are included by using this information in prior (expert opinion specification). Their approach is based on a Bayesian treatment with no hierarchical modeling and inventive ways of including several types of expert opinion. The primary concern is improving prediction of the computer code.

Kennedy and O'Hagan (2001) considered the three sources of data that we consider in this article. Their approach uses a general Gaussian process model for the computer model as a function of inputs. They used physical data to calibrate the computer experimental data and to estimate unknown parameters of that model. They also found Bayesian hierarchical models to be a useful tool in implementing their models. Their framework is flexible and, in the context of trying to improve computer models, the appropriate approach. The essential difference between their work and our proposed approach is that we are trying to use computer model outputs and expert opinion to improve estimation and prediction of the physical process, and Kennedy and O'Hagan were trying to use physical experimental data and expert opinion to improve the computer model.

The statistical notion of pooling data (sometimes also known as "borrowing strength") underlies our discussion of the RBHM and analysis. A commonly used and extremely powerful method for borrowing strength is hierarchical Bayesian modeling; a nice introduction to both hierarchical Bayesian modeling and borrowing strength was given by Draper et al. (1992). The basic idea involves the notion that when information concerning some response of interest arises from several independent but not identical data sources, a hierarchical model is often useful to describe relationships involving the observed data and unobserved parameters of interest. For example, unobserved parameters might be the coefficients and error variance in an assumed response surface model, as well as unknown biases. Each source of data provides perhaps biased information about these parameters, in which case methods that borrow strength will be useful. We illustrate the practical advantages of borrowing strength for estimating the unknown parameters in Section 3.2.

We propose fitting models using information from three distinct sources: expert opinion, computer experiments, and physical experiments. The problem is difficult, because the information sources are not necessarily all available at each of the design points. For example, physical experiments may be performed according to a statistically designed experiment, whereas computer experiments may be collected at (possibly) different design points. In addition, expert opinions may be available at only a very limited set of design points, such as the center or corners of the statistical design region. Our goal is to combine these sources of information using an appropriately flexible integration methodology that considers (and automatically adjusts for) the uncertainties and possible biases in each of these three data sources.

Thus we begin by considering regression models of the form

$$\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon},$$

where $\mathbf{X}$ is a design matrix, $\boldsymbol{\beta}$ is a vector of unknown coefficients, and $\boldsymbol{\varepsilon}$ is a vector of unobserved errors. Note that although this formulation can accommodate a general class of models, $f(\cdot)$, that includes both linear and nonlinear regression models, here we consider only linear models [i.e., $f(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$]. Although the strategy that we use is quite general, the model and mathematics that we develop is applied to a normal linear model. In addition, we consider only quantitative variables, although qualitative variables coded with indicator variables fit naturally into this framework.

2.1 Physical Experimental Data

We assume that we are interested in estimating the parameters of a model that describes a physical experiment. For this example, assume that the physical experimental data can be described using the familiar model

$$\mathbf{Y}_p \sim \mathrm{N}(\mathbf{X}_p\boldsymbol{\beta}, \sigma^2\mathbf{I}),$$

where the subscript p denotes the "physical experiment." Thus the physical experimental data are assumed to be normally distributed with mean $\mathbf{X}_p\boldsymbol{\beta}$, where $\mathbf{X}_p$ is a model matrix and $\boldsymbol{\beta}$ is a vector of parameters that need to be estimated. We see that each physical observation is independent of the others and has common (homoscedastic) variance σ^2, which also must be estimated.

If physical experimental data were the only information source considered, then this model would typically be fit using either standard least squares regression methods (Draper and Smith 1998) or standard Bayesian linear model methods (Gelman, Carlin, Stern, and Rubin 1995). However, we want to incorporate information both from experts and computer experimental data to "improve" our estimates of $\boldsymbol{\beta}$ and σ^2.

2.2 Expert Opinion

Suppose that there are e expert opinions. These opinions do not have to be from distinct experts. The ith expert opinion ($i = 1, \ldots, e$) is elicited at design point x_i. Some points in the design space will have exactly one elicited expert opinion, whereas others will have many or none. Each expert observation contains the following information:

- The expected response, y_{o_i}.
- A subjective coverage probability on the physical response y_i, ξ_i, and the quantile associated with that probability, q_{ξ_i}; that is, $\Pr(y_i \le q_{\xi_i}) = \xi_i$.

Typically, the analyst elicits a quantile of interest; that is, ξ_i is specified. However, the expert may indicate which quantile he or she is most interested in specifying. The methods developed here do not depend on which approach is taken. In addition, we consider the elicited "worth" of the opinion in units of equivalent physical experimental data observations, $m_{o_i}^{(e)}$. In other words, suppose that a physical experiment could be conducted at x_i that would yield one observation; if the expert's opinion should be weighted half of that observation, then $m_{o_i}^{(e)} = .5$. At times, the elicited values (y_{o_i}, ξ_i, q_{ξ_i}, $m_{o_i}^{(e)}$) may be obtained simply by requesting them from the expert. However, it may be difficult for the expert to provide information directly on these values (especially q_{ξ_i} and $m_{o_i}^{(e)}$), and other elicitation techniques may be useful (Meyer and Booker 1990).

To use these data, we need to transform these individual pieces of information into probability distributions that provide information about $\boldsymbol{\beta}$ and σ^2. Assume for the moment that the three aforementioned quantities can be used to create "data" with the following model:

$$\mathbf{Y}_o \sim \mathrm{N}(\mathbf{X}_o\boldsymbol{\beta} + \boldsymbol{\delta}_o, \sigma^2\boldsymbol{\Sigma}_o).$$

Like the physical experimental data, the expert data are assumed to be normally distributed. However, the mean is $\mathbf{X}_o\boldsymbol{\beta} + \boldsymbol{\delta}_o$, where $\boldsymbol{\delta}_o$ is a vector of location biases that are expert-specific. The variances are also biased, and the matrix $\boldsymbol{\Sigma}_o$ contains the scale biases for each expert. Besides location biases, in which an expert's average value is high or low relative to the true mean, scale biases often occur due to information overvaluation and are well documented in the elicitation literature. For example, an expert may be asked to provide what he or she thinks is a .90 quantile, but responds with what is actually only a .60 quantile (Meyer and Booker 1990). Although responses from experts can be correlated by having nondiagonal elements in $\boldsymbol{\Sigma}_o$, we consider uncorrelated responses; thus

$$\boldsymbol{\Sigma}_o = \begin{bmatrix} 1/k_{o_1} & 0 & \cdots & 0 \\ 0 & 1/k_{o_2} & 0 & \cdots \\ \vdots & 0 & \ddots & \cdots \\ 0 & \cdots & \cdots & 1/k_{o_e} \end{bmatrix}.$$

In addition, we assume the following prior distributions for the unknown parameters $\boldsymbol{\beta}$ and σ^2:

$$\boldsymbol{\beta}|\sigma^2 \sim \mathrm{N}(\boldsymbol{\mu}_o, \sigma^2\mathbf{C}_o)$$

and

$$\sigma^2 \sim IG(\alpha_o, \gamma_o),$$

where $IG(a, b)$ is the inverse gamma distribution with density function

$$f(z|a, b) \propto z^{-(a+1)} \exp\left\{-\frac{b}{z}\right\}, \qquad z > 0.$$

Assume for the moment that we know $\boldsymbol{\delta}_o$ and $\mathbf{m}_o$, where $\mathbf{m}_o$ is a vector denoting the "worth" of the expert opinions. Continue to assume that we have created "data" $\mathbf{y}_o$ from the expert opinions, and write out the likelihood for the data model,

$$\left(\frac{1}{|\sigma^2\boldsymbol{\Sigma}_o|^{.5}}\right) \times \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\mathbf{y}_o - (\mathbf{X}_o\boldsymbol{\beta} + \boldsymbol{\delta}_o)\right)'\boldsymbol{\Sigma}_o^{-1}\left(\mathbf{y}_o - (\mathbf{X}_o\boldsymbol{\beta} + \boldsymbol{\delta}_o)\right)\right]\right\}.$$

Using Bayes's theorem, we can use the data provided by the expert opinions to update the prior distributions for $\boldsymbol{\beta}$ and σ^2. The resulting stage 1 posterior/updated prior distribution for $(\boldsymbol{\beta}, \sigma^2)$, conditional on $\boldsymbol{\eta} = (\boldsymbol{\delta}_o, \boldsymbol{\Sigma}_o, \mathbf{m}_o, \mathbf{C}_o, \boldsymbol{\mu}_o, \alpha_o, \gamma_o)$, is

$$\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\eta}, \mathbf{y}_o) \sim \mathrm{N}\left((\mathbf{X}_o'\boldsymbol{\Sigma}_o^{-1}\mathbf{X}_o + \mathbf{C}_o^{-1})^{-1}\mathbf{z}, \sigma^2(\mathbf{X}_o'\boldsymbol{\Sigma}_o^{-1}\mathbf{X}_o + \mathbf{C}_o^{-1})^{-1}\right)$$

and

$$\pi(\sigma^2|\boldsymbol{\eta}, \mathbf{y}_o) \sim IG\Bigg(\alpha_o + \frac{\sum_{i=1}^{e} m_{o_i}}{2}, \gamma_o + .5\big[(\mathbf{y}_o - \boldsymbol{\delta}_o)'\boldsymbol{\Sigma}_o^{-1}(\mathbf{y}_o - \boldsymbol{\delta}_o) + \boldsymbol{\mu}_o'\mathbf{C}_o^{-1}\boldsymbol{\mu}_o - \mathbf{z}'(\mathbf{X}_o'\boldsymbol{\Sigma}_o^{-1}\mathbf{X}_o + \mathbf{C}_o^{-1})^{-1}\mathbf{z}\big]\Bigg),$$

where $\mathbf{z} = \mathbf{X}_o'\boldsymbol{\Sigma}_o^{-1}(\mathbf{y}_o - \boldsymbol{\delta}_o) + \mathbf{C}_o^{-1}\boldsymbol{\mu}_o$.

Given that the full vector of observations $\mathbf{y}_o$ was not elicited (only sufficient statistics were), we cannot immediately evaluate any term in these expressions. We instead reexpress the components in these posterior distributions in terms of the elicited values, so they can be evaluated. Suppose that m_{o_i} observations were elicited as $\mathbf{y}_{o_i}$ from the ith expert opinion. Then

$$\begin{aligned}\left(\mathbf{X}_o'\boldsymbol{\Sigma}_o^{-1}(\mathbf{y}_o - \boldsymbol{\delta}_o)\right)_j &= k_{o_1}x_{1j}\left(\sum_{n=1}^{m_{o_1}}\left(y_{o_{jn}} - \delta_{o_1}\right)\right) + \cdots + k_{o_e}x_{ej}\left(\sum_{n=1}^{m_{o_e}}\left(y_{o_{jn}} - \delta_{o_e}\right)\right) \\ &= k_{o_1}m_{o_1}x_{1j}\left(y_{o_1} - \delta_{o_1}\right) + \cdots + k_{o_e}m_{o_e}x_{ej}\left(y_{o_e} - \delta_{o_e}\right),\end{aligned}$$

because y_{o_i} is the expected or average response for the design point.

Using a similar argument, we can show that

$$(\mathbf{y}_o - \boldsymbol{\delta}_o)'\boldsymbol{\Sigma}_o^{-1}(\mathbf{y}_o - \boldsymbol{\delta}_o)$$
$$= k_{o_1}\left(\sum_{n=1}^{m_{o_1}}\left(y_{o_{jn}} - \delta_{o_1}\right)^2\right) + \cdots + k_{o_e}\left(\sum_{n=1}^{m_{oe}}\left(y_{o_{jn}} - \delta_{o_e}\right)^2\right)$$
$$= \sum_{i=1}^{e} k_{o_i} m_{o_i}\left(s_i^2 + \left(y_{o_i} - \delta_{o_i}\right)^2\right), \tag{1}$$

where $s_i^2 = (y_{o_i} - q_{\xi_i})^2/Z_\xi^2$, which is the variance approximation implicitly elicited from expert i. Equation (1) follows from the identity $\text{var}(Y) = E[Y^2] - E[Y]^2$.

By a similar argument,

$$(\mathbf{X}_o'\boldsymbol{\Sigma}_o^{-1}\mathbf{X}_o)_{ij} = \sum_{n=1}^{e} k_n m_{o_n} x_{ni} x_{nj}.$$

These representations allow calculation of the quantities in the posterior distributions based on the elicited values rather than on the actual observations.

For the unknown parameters $\boldsymbol{\eta} = (\boldsymbol{\delta}_o, \boldsymbol{\Sigma}_o, \mathbf{m}_o, \mathbf{C}_o, \boldsymbol{\mu}_o, \alpha_o, \gamma_o)$, we propose the following prior distributions:

$$\boldsymbol{\mu}_o = \mathbf{a}_{\boldsymbol{\mu}_o},$$
$$\mathbf{C}_o = a_{\mathbf{C}_o}\mathcal{I},$$
$$\alpha_o = a_{\alpha_o},$$
$$\gamma_o = a_{\gamma_o},$$
$$m_{o_i} \sim \text{uniform}\left(.5m_{o_i}^{(e)}, 2.0m_{o_i}^{(e)}\right),$$
$$\delta_{o_i} \overset{\text{iid}}{\sim} \text{N}(\theta_o, \xi_o^2),$$
$$\theta_o \sim \text{N}\left(m_{\theta_o}, s_{\theta_o}^2\right),$$
$$\xi_o^2 \sim IG\left(a_{\xi_o^2}, b_{\xi_o^2}\right),$$
$$k_{o_i} \overset{\text{iid}}{\sim} G(\phi_o, \omega_o),$$
$$\phi_o \sim G\left(a_{\phi_o}, b_{\phi_o}\right),$$

and

$$\omega_o \sim G\left(a_{\omega_o}, b_{\omega_o}\right),$$

where a and b subscripted indicate constants, and $G(a, b)$ indicates a gamma distribution with mean ab and variance ab^2. These highly parametric specifications suggest that sensitivity may result from choices of distributional form as well as hyperparameter choices. As with any analysis, increasing the degree of assumption increases the *potential* for sensitivity to those assumptions. For example, inadequate sample sizes will certainly exacerbate these sensitivities. In Section 4 we consider a sensitivity study to examine the degree to which our results depend on the foregoing hyperparameter choices.

There are similarities between this approach to the quantification of expert opinion and Zellner's approach using g-prior distributions (Zellner 1986; Agliari and Parisetti 1988). Both approaches rely on the natural conjugate prior for $(\boldsymbol{\beta}, \sigma^2)$. However, Zellner (1986) elicited posterior means for $\boldsymbol{\beta}$ and σ^2, whereas we elicit predicted observations $\mathbf{y}_o$. Agliari and Parisetti (1988) extended Zellner's methods to include a different design matrix, $\mathbf{XA}$; similarly, we do not require that the factor levels where the expert elicitation occurs correspond to the levels where the physical or computer experimental data are collected.

2.3 Computer Experimental Data

We have used the expert opinion data to develop stage 1 posterior distributions for $\boldsymbol{\beta}$ and σ^2. We continue to update our knowledge about these parameters using data from computer experiments. Let the computer data and associated model parameters be indexed by c, where the jth element of the response vector $\mathbf{Y}_c$ is y_{c_j}. Consider the following model:

$$\mathbf{Y}_c \sim \text{N}(\mathbf{X}_c\boldsymbol{\beta} + \boldsymbol{\delta}_c, \sigma^2\boldsymbol{\Sigma}_c),$$
$$\boldsymbol{\beta}|\sigma^2 \sim \text{N}(\boldsymbol{\mu}_c, \sigma^2\mathbf{C}_c),$$

and

$$\sigma^2 \sim IG(\alpha_c, \gamma_c).$$

For this development, assume that $\boldsymbol{\Sigma}_c$ and $\mathbf{C}_c$ have the same diagonal form as $\boldsymbol{\Sigma}_o$ and $\mathbf{C}_o$. The "prior" distributions for $\boldsymbol{\beta}|\sigma^2$ and σ^2 are the stage 1 posterior distributions given the expert opinion data. The only other unspecified prior distributions are

$$\delta_{c_j} \overset{\text{iid}}{\sim} \text{N}(\theta_c, \xi_c^2),$$
$$\theta_c \sim \text{N}\left(m_{\theta_c}, s_{\theta_c}^2\right),$$
$$\xi_c^2 \sim IG\left(a_{\xi_c^2}, b_{\xi_c^2}\right),$$
$$k_{c_j} \overset{\text{iid}}{\sim} G(\phi_c, \omega_c),$$
$$\phi_c \sim G\left(a_{\phi_c}, b_{\phi_c}\right),$$

and

$$\omega_c \sim G\left(a_{\omega_c}, b_{\omega_c}\right).$$

Although assuming a diagonal structure for $\boldsymbol{\Sigma}_c$ yields a model for the computer experiment where the observations are *conditionally* independent given $\boldsymbol{\beta}$, $\boldsymbol{\delta}_c$, σ^2, and $\mathbf{k}_c$, the observations are not *unconditionally* independent once the uncertainty in the unknown parameters is integrated out. For example, Broemeling (1985) derived the distribution for $\mathbf{Y}_c$ for the conjugate Bayesian linear model. Our model for the correlation structure differs from those proposed by Currin et al. (1991) and Welch et al. (1992), who assumed a distance-based parametric form for $\boldsymbol{\Sigma}_c$ with the parameters selected using cross-validation or maximum likelihood estimation. Although these forms of prior distribution could be incorporated into our analysis, we have chosen to induce correlation through the hierarchical structure of the prior.

Computer models, especially when the physical processes are not well known, often produce estimates that are biased with respect to the physical data. These biases may be in the mean structure (location bias) or in the variance (scale bias). Computer experimental data are especially likely to have scale biases, because these data usually tend to be less variable than physical experimental data; in fact, there is often no stochastic

variability for given values of the factors, because a computer code is often deterministic. The variability occurs relative to the assumed model. Another reason for the reduced variability relative to physical experimental data is that we know that not all factors generating the physical experimental data are incorporated into the computer code—perhaps all of the factors causing variability are unknown. Because the location bias addresses only differences in the intercept term (β_0) between the computer and physical data, more general bias structures for the parameters also can be modeled. In Section 3 we motivate these ideas by introducing the operation of fluidized beds and the computer models for that process.

Because the location biases are additive (instead of multiplicative), the model only requires that data exist for a subset of the full set of factors. That is, if only one data source includes information on a factor, then only that source is used in estimating that effect. The precision with which those effects are estimated will be affected by the differing amounts of data used in estimation. However, distributions can be calculated. If the model is to be chosen based on the physical data only (as in our example), then all of the factors would need to be present in the physical experimental data. Thus the framework is quite general and does not require that all factors be present in each data source.

Other approaches that might be considered for modeling the computer experimental data often use a Gaussian process (GP) model (Santner, Williams, and Notz 2003). Although the GP approach is commonly (and appropriately) used for many problems, the RBHM provides an alternative that is useful and easily interpreted for certain classes of problems. The benefits of a linear models approach as outlined in the RBHM are that it is computationallytractable, easily interpretable, and easy to visualize. Disadvantages of using a linear models approach include that they cannot act as an interpolator (whereas GP models have this feature, which explicitly acknowledges the deterministic nature of computer experiments), they are not as flexible as GP models, and they require more observations when higher-order terms are needed in the model. Considering these benefits and limitations, researchers must ascertain the suitability of this formulation (or any other modeling approach) when combining data sources that are diverse, such as computer experiments and physical experiments.

In Section 3 we illustrate a problem that is well suited for our proposed modeling approach (RBHM).

2.4 Incorporating Physical Experimental Data

Recall from Section 2.1 that the model for the physical experimental data is

$$\mathbf{Y}_p \sim \mathbf{N}(\mathbf{X}_p\boldsymbol{\beta}, \sigma^2\mathbf{I}).$$

After incorporating the computer experimental data into the analysis, we have a stage 2 posterior that is used as the prior for $(\boldsymbol{\beta}, \sigma^2)$ in the stage 3 analysis.

The stage 3 analysis calculates the final distributions for the parameters of interest. These calculations cannot be done in closed form, but are carried out using Markov chain Monte Carlo (MCMC). The Appendix provides general information on MCMC and the Metropolis–Hastings algorithm.

3. APPLICATION OF RECURSIVE BAYESIAN HIERARCHICAL MODELS TO FLUIDIZED BED PROCESSES

Fluidized-bed microencapsulation processes are used in the food industry to coat certain food products with additives. Dewettinck, Visscher, Deroo, and Huyghebaert (1999) described a physical experiment and several corresponding thermodynamic computer models developed for predicting the steady-state thermodynamic operation point of a Glatt GPCG-1 fluidized-bed unit in the top-spray configuration. Figure 1 illustrates the simple geometry of this unit, which is essentially an upside-down truncated cone. The base of the unit contains a screen, below which is an air pump. Also, there are coating sprayers at the side of the unit.

To use the unit, a batch of uncoated food product is placed inside the "cone," and the air pump and coating sprayers are turned on. This "fluidizes" the product in the unit and coats the product as it passes by the sprayer. This is continued until the desired coating thickness is achieved.

When room conditions and process conditions are constant, a fluidized-bed process will attain its steady-state thermodynamic operation point. This state can be described in terms of the temperature and humidity inside the unit. The importance of the steady-state operation point is that product characteristics, such as coating evenness and efficiency, are directly related to it.

Several variables potentially affect the steady-state thermodynamic operating point:

- V_f, fluid velocity of the fluidization air
- T_a, temperature of the air from the pump
- R_f, flow rate of the coating solution
- T_s, temperature of the coating solution
- M_d, coating solution dry matter content
- P_a, pressure of atomization air.

The ambient room conditions inside the plant, such as temperature (T_r) and humidity (H_r), may also have an effect on the steady-state process conditions.

3.1 The Data

Dewettinck et al. (1999) considered 28 process conditions of particular interest (settings) for a GPCG-1 fluidized-bed process. In the experiment, distilled water was used as the coating solution. Thus M_d was 0 (no dry matter content) for all 28 runs. Also, T_s was at room temperature (about 20°C) for

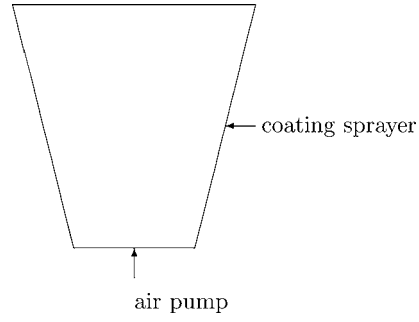

Figure 1. A Glatt GPCG-1 Fluidized Bed Unit.

Table 1. Process Variables

H_r(%)	T_r(°C)	T_a(°C)	R_f(g/min)	P_a(bar)	V_f(m/s)
51.0	20.7	50	5.52	2.5	3.0
46.4	21.3	60	5.53	2.5	3.0
46.6	19.2	70	5.53	2.5	3.0
53.1	21.1	80	5.51	2.5	3.0
52.0	20.4	90	5.21	2.5	3.0
45.6	21.4	60	7.25	2.5	3.0
47.3	19.5	70	7.23	2.5	3.0
53.3	21.4	80	7.23	2.5	3.0
44.0	20.1	70	8.93	2.5	3.0
52.3	21.6	80	8.91	2.5	3.0
55.0	20.2	80	7.57	1.0	3.0
54.0	20.6	80	7.58	1.5	3.0
50.8	21.1	80	7.40	2.0	3.0
48.0	21.2	80	7.43	2.5	3.0
42.8	22.4	80	7.51	3.0	3.0
55.7	20.8	50	3.17	1.0	3.0
55.2	20.7	50	3.18	1.5	3.0
54.4	20.7	50	3.19	2.0	3.0
55.4	19.8	50	3.20	2.5	3.0
52.9	20.0	50	3.19	3.0	3.0
28.5	18.3	80	7.66	2.5	3.0
26.1	19.0	80	7.69	2.5	4.0
24.2	18.9	80	7.69	2.5	4.5
25.4	18.5	80	7.70	2.5	5.0
45.1	19.6	50	3.20	2.5	3.0
43.1	20.3	50	3.23	2.5	4.0
42.7	20.4	50	3.20	2.5	4.5
38.7	21.6	50	3.22	2.5	5.0

all 28 runs. Table 1 gives the room conditions (i.e., T_r and H_r) and settings for the remaining four process variables (i.e., T_a, R_f, P_a, and V_f). Thus the six factors actually studied are T_r, H_r, T_a, R_f, P_a, and V_f.

For each factor combination, glass beads were put in the unit, and the process was run for 15 minutes to attain steady state. Then temperature inside the unit was measured at 20, 25, and 30 minutes and their average was recorded. The average outlet air temperature (the steady-state response of interest), $T_{2,\exp}$, is reported in Table 2. Dewettinck et al. (1999) also considered three unique computer models to predict the steady-state outlet air temperature for each run. These computational responses are also given in Table 2, denoted by $T_{2,1}$, $T_{2,2}$, and $T_{2,3}$.

Table 2. Experimental and Computer Model Steady-State Temperatures

$T_{2,exp}$ (°C)	$T_{2,1}$ (°C)	$T_{2,2}$ (°C)	$T_{2,3}$ (°C)
30.4	32.4	31.5	30.2
37.6	39.5	38.5	37.0
45.1	46.8	45.5	43.7
50.2	53.8	52.6	51.0
57.9	61.7	59.9	58.2
32.9	35.2	34.6	32.6
39.5	42.4	41.0	39.1
45.6	49.5	48.5	46.4
34.2	37.5	36.6	34.8
41.1	45.5	44.3	42.0
45.7	50.5	49.0	47.0
44.6	49.8	48.4	46.3
44.7	49.8	48.4	46.3
44.0	49.2	48.0	45.7
43.3	48.6	47.5	45.4
37.0	39.5	38.0	37.7
37.2	39.5	38.5	37.1
37.1	39.5	37.5	36.7
36.9	39.5	38.5	36.1
36.8	37.7	37.2	36.2
46.0	48.7	47.3	45.1
54.7	57.7	56.2	54.2
57.0	60.1	58.7	57.0
58.9	62.0	60.5	58.7
35.9	37.9	37.1	36.1
40.3	41.7	40.8	40.1
41.9	43.0	42.3	41.4
43.1	43.9	43.3	42.6

Table 3. Correlation Matrix

	H_r	T_r	T_a	R_f	P_a	V_f	$T_{2,exp}$
H_r	1.00	.57	−.26	−.33	−.39	−.69	−.53
T_r	.57	1.00	−.09	−.07	−.04	−.28	−.37
T_a	−.26	−.09	1.00	.82	.06	−.08	.73
R_f	−.33	−.07	.82	1.00	.09	−.10	.35
P_a	−.39	−.04	.06	.09	1.00	.18	.08
V_f	−.69	−.28	−.08	−.10	.18	1.00	.47
$T_{2,\exp}$	−.53	−.37	.73	.35	.08	.47	1.00

There are important differences among the three computational models described in detail by Dewettinck et al. (1999). In summary, the first computer model does not include adjustments for heat losses in the process. The second computer model takes those heat losses into account. A further adjustment for the inlet airflow represents the fundamental difference between the second and third computer models.

3.2 Modeling $T_{2,\exp}$ in Terms of Room and Process Conditions

Table 3 shows the correlation matrix for the room conditions, process conditions, and observed steady-state temperature $T_{2,\exp}$. Figure 2 is a matrix plot of these seven variables. Note that T_a has the highest correlation with $T_{2,\exp}$ ($r = .73$).

Model choice is complicated by the fact that the underlying design is not at all clear. The covariance matrix reveals that some of the covariates are highly correlated (as high as .82) indicating possible collinearity. We also note that the full second-order model is fully saturated.

Chipman, Hamada, and Wu (1997) described a Bayesian variable selection procedure that places hierarchical prior distributions on second-order effects. In their approach, higher prior probability is given to interactions if one of the main effects is in the model, and an even higher probability is placed on interactions when both main effects are in the model. Using their approach on the physical data, we obtain the variable selection results displayed in Table 4, which provide the most likely models and their respective posterior probabilities.

To illustrate the RBHM approach, we use the most likely model from Table 4 to form $\mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is composed of a column of 1's (for the intercept) and columns corresponding to T_a, R_f, V_f, and $R_f \times V_f$, whose respective regression parameters

Table 4. Bayesian Variable Selection Results

Model	Pr(model|data)
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 R_f \times V_f$	.1169
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 H_r \times T_r + \beta_5 R_f \times V_f$	.0349
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 H_r^2$	.0155
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 T_r \times T_a + \beta_5 R_f \times V_f$	.0141
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 R_f \times V_f + \beta_5 V_f^2$	.0136
$\beta_0 + \beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 H_r^2$	.0132
$\beta_1 T_r + \beta_2 T_a + \beta_3 R_f + \beta_4 V_f + \beta_5 R_f \times V_f$	.0130

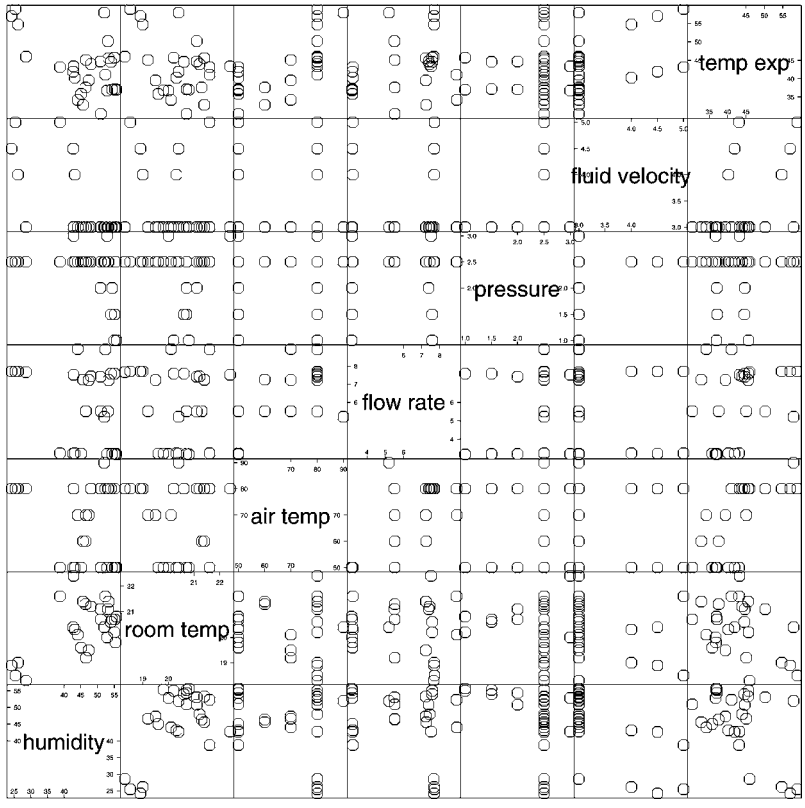

Figure 2. Scatterplot Matrix of the Experimental Response With Each of the Six Covariates.

are $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_4)'$. Table 5 contains the OLS fit of the most likely model given in Table 4.

The hyperparameter values used in our example are given in Table 6. Note that we used the same hyperparameters for all three computer experiments. Because we have no prior knowledge as to the sign of the location bias, we center the distribution of δ_{c_i} at 0 (i.e., unbiased in location) and allow the mean of that distribution to have a standard deviation of 10. Although we believe that the computer models are all reasonably good approximations of the physical model, we do not have a good idea about the degree of separation, and thus allow a generous variability for the location biases [$a_{\xi_c^2} = 2{,}000$ and $b_{\xi_c^2} = 3.0$ suggest a mean for the variance distribution of $2{,}000/(3-1) = 1{,}000$ and a standard deviation of $2{,}000/((3-1)\cdot(3-2)) = 1{,}000$]. The distribution of scale biases is also somewhat unknown. With little or no prior knowledge, we would allow the mean of the scale biases to be unity (unbiased in scale). Further, we believe that the standard deviation of the scale biases should be no greater than 15, and thus we let the mean of the scale bias distribution be 1 and the standard deviation to be 15. This allows a generous range for the scale biases.

Table 5. OLS Fit for $T_{2,exp} = \beta_0 + \beta_1(T_a - \bar{T}_a) + \beta_2(R_f - \bar{R}_f) + \beta_3(V_f - \bar{V}_f) + \beta_3((R_f - \bar{R}_f) \times (V_f - \bar{V}_f)) + \epsilon$

Variable	*DF*	*Parameter estimate*	*Standard error*	*T for H_0: Parameter = 0*	*Pr > |T|*
Intercept	1	42.9769	.1714	250.7352	0
T_a	1	9.4756	.3056	31.0076	0
R_f	1	−4.9048	.3035	−16.1626	0
V_f	1	3.9345	.1761	22.3445	0
$R_f \times V_f$	1	1.4263	.1671	8.5336	0
$adj - R^2 = .9855$ on 23 df					

3.3 Recursive Bayesian Hierarchical Model Analysis Results

Figures 3(a)–(e) show the posterior for $\boldsymbol{\beta}$ with only the physical experimental data, the physical data with the computer experimental data taken separately, and the final posterior distribution for $\boldsymbol{\beta}$ after incorporating all sources of information. Figure 3(f) shows the corresponding posteriors for σ^2. The figures indicate two important and appealing aspects of our RBHM approach. First, the additional sources of information reduce uncertainty in the distribution of the parameters, thus making our estimates more precise. Second, the additional data

Table 6. Hyperparameter Values for Parameters in Computer Experiments

Hyperparameter	*Value*
C_c	1.0×10^{-4}
α_C	3.0
β_C	3.0
m_{θ_c}	0
$s^2_{\theta_c}$	100.0
$a_{\xi_c^2}$	2,000.0
$b_{\xi_c^2}$	3.0
a_{ϕ_c}	1.0×10^{-3}
b_{ϕ_c}	1.0×10^{-3}
a_{ω_c}	1.0×10^{-3}
b_{ω_c}	1.0×10^{-3}

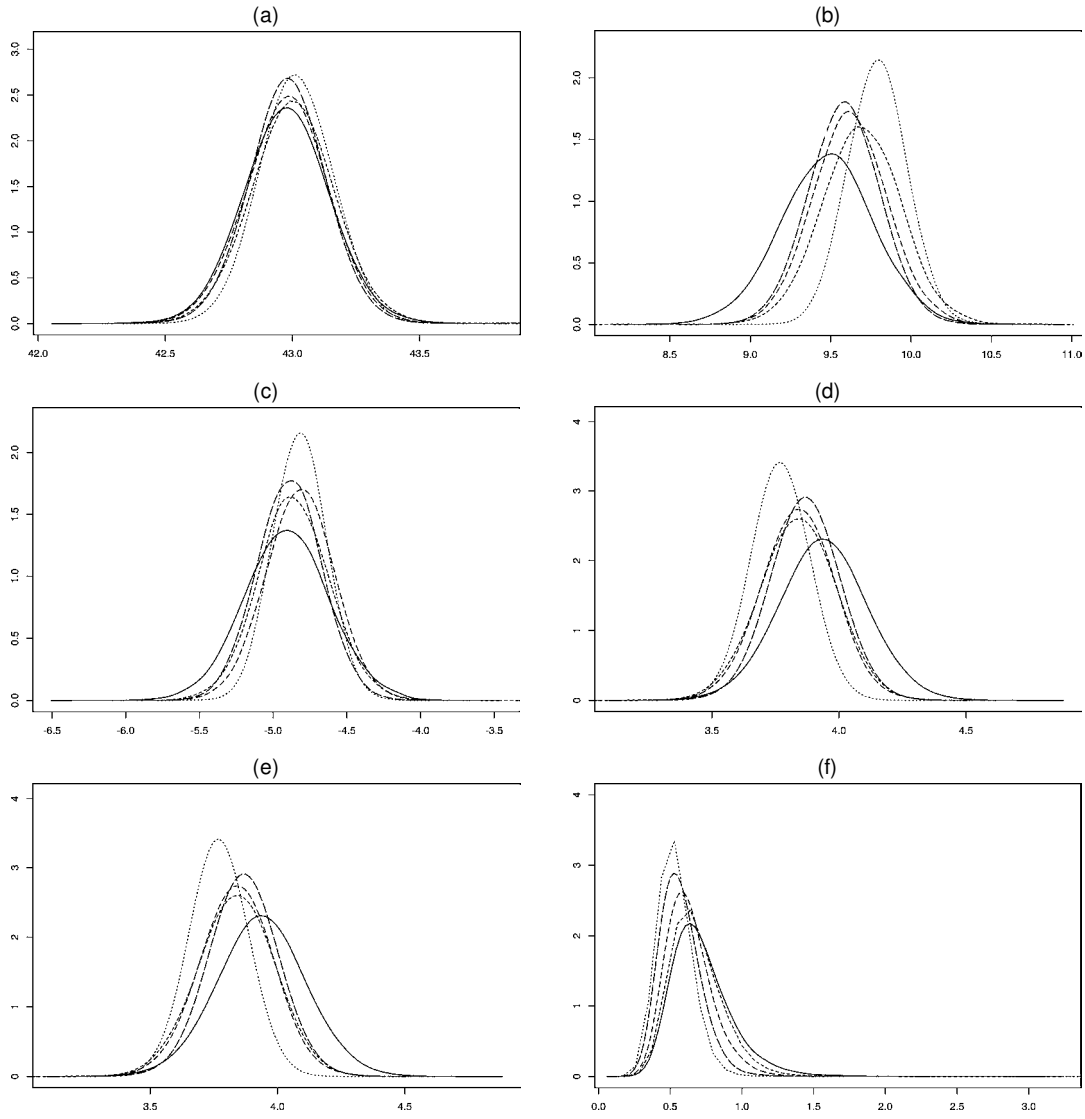

Figure 3. Comparison of Posterior Distributions Conditional on Different Sources of Information: (a) β_0, Intercept; (b) β_1, Air Temperature; (c) β_2, Flow Rate; (d) β_3, Fluid Velocity; (e) β_4, Interaction Between Flow Rate and Fluid Velocity; and (f) σ^2. The different lines indicate inclusion of different data sources. (—— physical only; ····· physical + 3 computers; ----- physical + computer 1; - - - physical + computer 2; —·— physical + computer 3.)

sources do not necessarily contain the same amount of information (although in our example they do have the same number of observations).

In addition to posterior distributions for $\boldsymbol{\beta}$ and σ^2, our modeling approach allows us to estimate the bias terms. As an illustration, Figures 4(a) and 4(b) present the location and scale

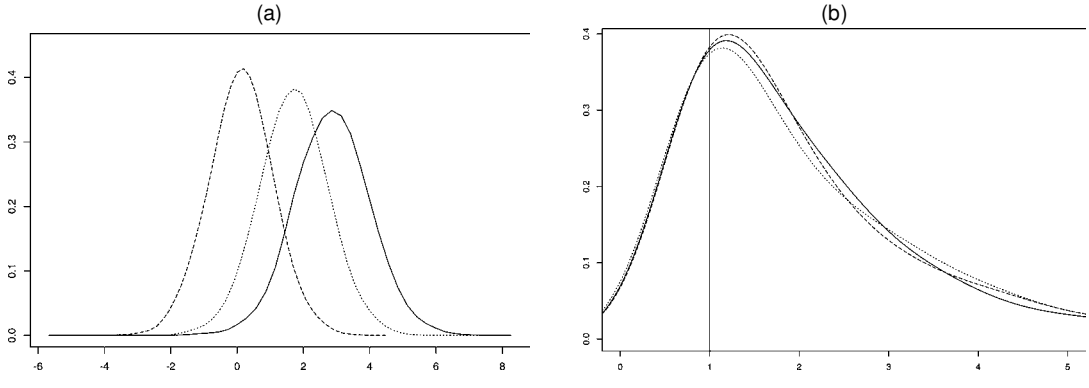

Figure 4. Comparison of (a) Location Bias and (b) Scale Bias Predictive Distributions for Three Different Computer Models of the Fluidized Bed Process. [(a) —— computer 1; ····· computer 2; ----- computer 3; (b) ····· computer 1; ----- computer 2; —— computer 3.]

Table 7. Comparison of Confidence and Credible Intervals

		95% confidence interval		Posterior	95% credible HPD interval	
	MLE	Lower	Upper	mean	Lower	Upper
σ^2	.81	.49	1.60	.53	.36	.77
β_0	42.97	42.62	43.33	43.01	42.75	43.28
β_1	9.47	8.84	10.10	9.79	9.44	10.13
β_2	−4.90	−5.53	−4.27	−4.82	−5.15	−4.48
β_3	3.93	3.57	4.29	3.76	3.56	3.96
β_4	1.42	1.08	1.77	1.35	1.17	1.53

bias predictive distributions for each of the computer models. Note that these distributions are integrated over the distribution of individual-specific locations and scale bias terms. One appealing feature of these plots is that they indicate a new approach to computer model validation relative to the physical observations. Those models that have most mass over 0 are less location biased for the physical experimental data. For example, the bias is more concentrated around 0 for the third computer model than for the other two computer models. These plots also reveal the uncertainty associated with the bias terms, a feature that cannot easily be inferred from a casual examination of the data. Note that the third model is the computer model that attempts to account for more phenomena. Figure 4(b) reveals that all three computer models tend to underestimate the variability in the physical experimental data. Scale bias terms greater than 1 (because the scale bias is parameterized as $1/k_{C_i}$) indicate underestimation of variability.

Table 7 contains the maximum likelihood estimates (MLEs), 95% confidence intervals (calculated from only the physical experimental data), and the posterior mean and 95% highest posterior density (HPD) intervals calculated using the integrated computer and physical experimental data for $\boldsymbol{\beta}$ and σ^2. Recall that an HPD interval is the shortest interval in the posterior distribution containing 95% of the posterior probability. Notice that the HPD intervals are shorter (sometimes significantly so) than the 95% confidence intervals, reflecting the additional information that has been incorporated into the analysis.

3.3.1 Expert Opinion Data. Although no expert opinions were available for use in the fluidized bed example, it is interesting to observe the impact of such data on the results. For purely illustrative purposes, suppose that eight expert opinions were elicited for use in the fluidized bed example. The expert opinions are given in Table 8, where $T_{2,o}$ denotes the expected steady-state outlet air temperature, $q_{.9}$ is the corresponding subjective .9 quantile on the outlet air temperature, and $m_o^{(e)}$ is the equivalent "worth" of the opinion (see Sec. 2.2).

Table 8. Example Expert Opinion Data

T_a (°C)	R_f (g/min)	V_f (m/s)	$T_{2,o}$ (°C)	$q_{.9}$ (°C)	$m_o^{(e)}$
50	3	3	37	39	.5
90	3	3	68	70	.5
50	9	3	23	25	.5
90	9	3	51	53	.75
50	3	5	49	53	1.0
90	3	5	75	77	.5
50	9	5	42	43	.75
90	9	5	69	72	.5

Figure 5 contains two posterior distributions, one distribution for the regression coefficient for flow rate (β_2) and one for the error variance (σ^2). The solid line is the posterior distribution conditional on the artificial expert opinion with one computer model and the physical experimental data. The dotted line is the posterior distribution with only the physical experimental data and one computer model. Due to estimation of location and scale biases for both the computer data and the artificial expert opinion data, only a small gain in information results from adding the expert opinion data. No inference from these posterior distributions should be made, because the expert opinion data were generated for illustration purposes only.

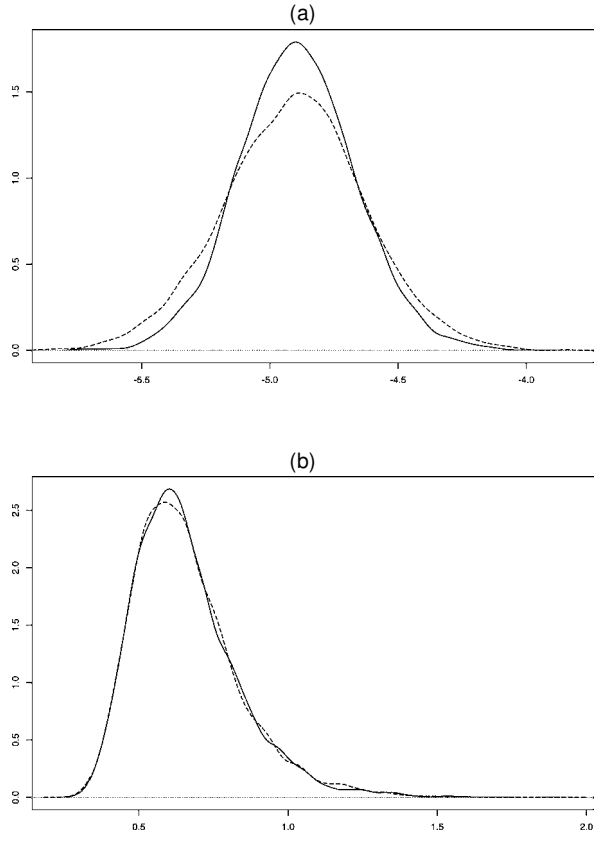

Figure 5. Comparison of Posterior Distributions for (a) β_2, Flow Rate and (b) σ^2. The solid line represents the posterior distribution conditional on the artificial expert opinion with one computer model and the physical experimental data; the dotted line, the posterior distribution with only the physical experimental data and one computer model. (—— EO included; ----- EO excluded.)

Table 9. Hyperparameter Values for Sensitivity Analysis

Hyperparameter	*Low*	*High*
C_C	1.0×10^{-7}	.1
α_C	.1	5.0
β_C	.1	5.0
$s^2_{\theta_C}$	50	500
$a_{\xi_C^2}$	100	1,000
$b_{\xi_C^2}$	1	10
a_{ϕ_C}	1.0×10^{-2}	2
b_{ϕ_C}	1.0×10^{-2}	2
a_{ω_C}	1.0×10^{-2}	2
b_{ω_C}	1.0×10^{-2}	2

4. PRIOR SENSITIVITY

Bayesian analyses that contain many parameters have the potential to rely heavily on prior distributions and prior parameter choice. To assess the impact of our choices of prior parameters (and hyperparameters), we conducted a small sensitivity analysis. To address hyperparameter sensitivity, we designed a 2^{10-5} fractional factorial design in which we chose "high" and "low" values that we deemed feasible. The values that we chose are given in Table 9.

Marginal posterior distributions for the regression coefficients $(\beta_0, \ldots, \beta_4)$ and the error variance (σ^2) are shown in Figure 6. Because the posterior distributions do not lie ex-

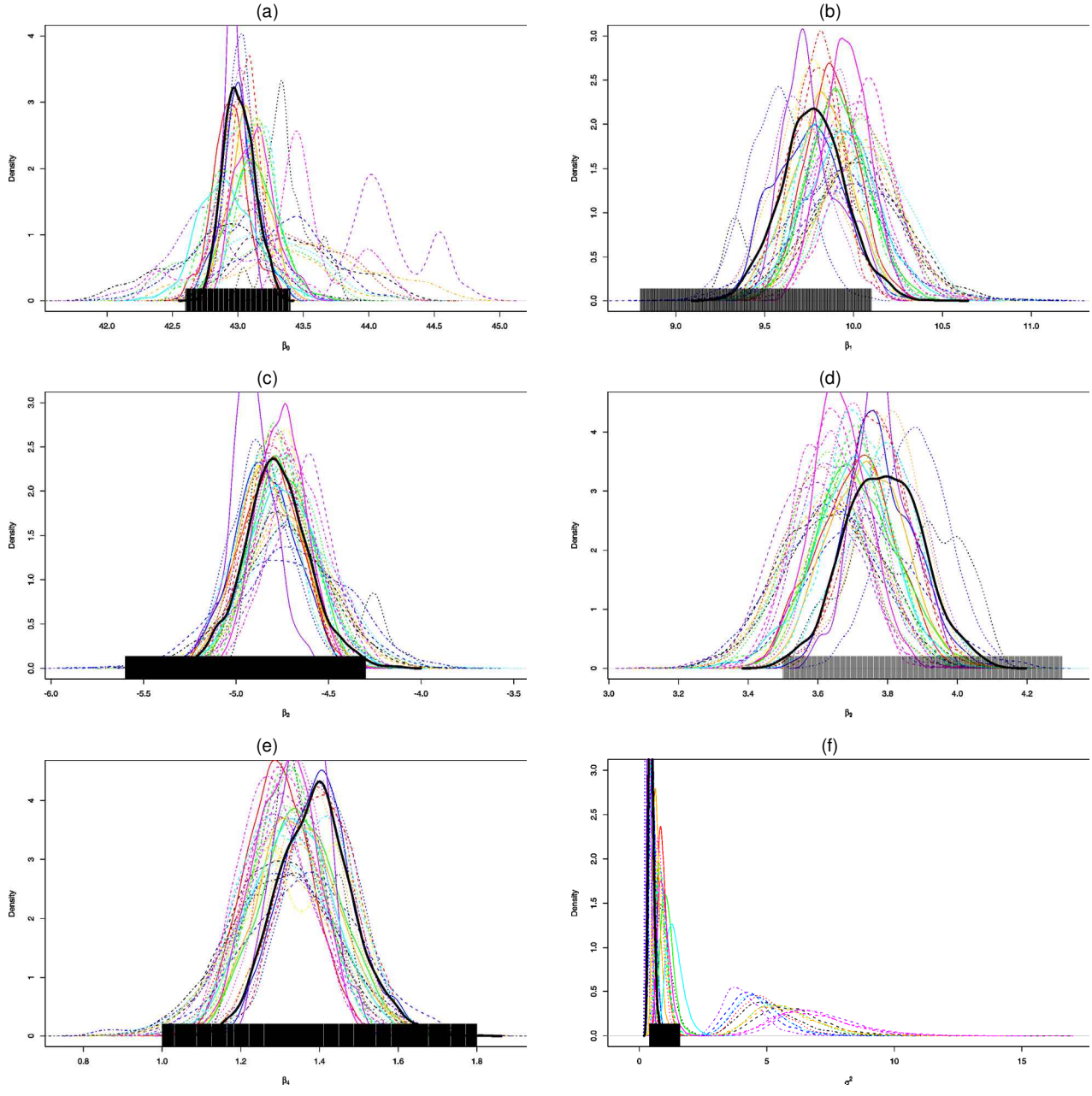

Figure 6. Sensitivity Analysis for Selected Hyperparameters in the RBHM Formulation: (a) β_0, Intercept; (b) β_1, Air Temperature; (c) β_2, Flow Rate; (d) β_3, Fluid Velocity; (e) β_4, Interaction Between Flow Rate and Fluid Velocity; and (f) σ^2. The different lines indicate a different factorial combination in the sensitivity analysis, and the thick solid line indicates the posterior at the original settings. The "rug" at the bottom of each picture is the frequentist confidence interval based on the physical data only fit to the five-parameter linear model.

actly on one another, there is clearly some sensitivity to prior specification. The differences in the posterior distributions are not significant, however. The only clear deviations are 6 of the 32 fractional factorial combinations for σ^2 and 2 of the 32 fractional factorial combinations for β_0. These produce significant departure from the posterior distributions presented in Section 3.3. We note that each of these stems from a prior distribution that includes nearly no mass around the posterior distribution; that is, they represent infeasible prior distributions. This indicates that care should be taken when specifying prior parameters on the variability. Sensitivity is observed only when priors are completely misspecified.

5. DISCUSSION AND CONCLUSIONS

When expert opinion is elicited, an equivalent number of observations, $m_{o_i}^{(e)}$, is also stated that reflects its worth in terms of a number of equivalent physical observations. This parameter is not required for the computer experimental data, because this information is captured in the prior parameters $\theta_c, \xi_c^2, \phi_c$, and ω_c. These parameters control the prior information about the location and scale biases for the computer experimental data. If the biases are known exactly (a point mass prior), then each computer observation counts as exactly one physical observation—no information must be used to estimate the biases, and it can all be used to estimate $\boldsymbol{\beta}$ and σ^2. If these parameters are used to specify a very diffuse ("noninformative") prior with close to infinite variances, then each computer observation counts for only a tiny fraction of a physical observation. If the parameters specify an informative prior, then the computer observations account for some intermediate fraction of a physical observation.

The model that we have used in our example treats each computer model independently. In the extreme, this implies that if the three models were identical, then we would count each observation three times the fraction of a physical observation implied by the prior distributions. We can change this by modeling correlations between the computer models. There are two obvious ways to do this. The simplest way is to add a hierarchical structure on the hyperparameters (θ_c, ξ_c, ϕ_c, and ω_c) of the various computer models. As discussed in Section 2.3, this induces correlations in the unconditional distributions of the computer observations. A second is to model the entire vector of observations from the three computer models directly as a multivariate normal and to specify an appropriate covariance structure. This choice would be especially appropriate in the case in which we had precise information about the differences in the physics modeled by the individual computer models. For this example, we have insufficient knowledge about the precise similarities/differences between the three computer models to permit the use of any of them.

We have not imposed the requirement that the computed (or measured) value of the response variable be considered at the same factor values in both experiments. We only require that there exist some common set of factors (either all or at least some) for both experiments. Although the example does not fully illustrate this, it is an important feature in the general model. As the analysis proceeds by using information from one type of experiment to update the distribution of the parameters, if there are no data at a particular design point for a particular experiment, then the distribution for the parameter remains unchanged, except for correlations that may exist in the parameters.

As with any Bayesian analysis, there is sensitivity to the specification of the prior distributions for the hyperparameters. Fortunately, however, the sensitivity is particularly acute only when the priors are completely misspecified. Although some of the hyperparameter selections in Section 3.2 are somewhat arbitrary, they illustrate the kinds of discussions that the analyst would engage in with the data owner to arrive at "reasonable" hyperparameter distributions. If at all possible, we prefer diffuse but informative prior distributions using expert input.

In this example we included all three sets of computer data, even though we believed that the models were successively improved. We made this choice for two reasons: first, we believe that by appropriate modeling of biases, there is information in all of the codes that should not be discarded; and second, it is often of interest to characterize the biases of each code relative to the physical data.

We have presented an RBHM that can be used to combine data from both computer and physical experiments. When available, expert opinion data are also used to "sharpen" the initial informative, but rather diffuse prior distributions. Appropriate biases are introduced as a way to account for differences in these data sources. Sample results indicate that significantly more precise estimates of the regression coefficients and error variance are obtained by means of this method. In addition, the methodology can be used to recursively estimate those unknown biases of particular interest. Biases that are not particularly interesting can be marginalized (i.e., averaged out of the analysis using appropriate priors). Obviously, not all problems involving combination of computer models and physical experiments are well suited to combination through statistical (response surface) models. In our example, however, the approach is well suited to the data collected, and the biases seem to reflect the actual differences between the computer models and the physical data.

The methodology can also be used to combine various other kinds of experimental information. Similarly, information from more than two physical and/or computer experiments can also be combined using the RBHM simply by considering an appropriate bias structure for each data source and by increasing the number of stages in the analysis accordingly.

ACKNOWLEDGMENTS

The authors thank Valen Johnson for his valuable comments. They also thank the editor, an associate editor, and the referees for their suggestions, which improved the manuscript.

APPENDIX: COMPUTATIONAL DETAILS

A.1 Markov Chain Monte Carlo

Suppose that we are interested in making statistical inference about a parameter (possibly vector valued) $\boldsymbol{\Theta}$. We characterize our information (or lack of information) about the distribution of $\boldsymbol{\Theta} = \{\theta_1, \theta_2, \ldots, \theta_n\}$ as $\pi(\boldsymbol{\Theta})$ (prior distribution). Data are

collected and represented by the likelihood or by $f(\mathbf{x}|\mathbf{\Theta})$. In any Bayesian analysis, inference on the parameters depends on the calculated posterior distribution

$$\pi(\mathbf{\Theta}|\mathbf{x}) = \frac{\pi(\mathbf{\Theta})f(\mathbf{x}|\mathbf{\Theta})}{\int_{\mathbf{\Theta}} \pi(\mathbf{\Theta})f(\mathbf{x}|\mathbf{\Theta})\, d\mathbf{\Theta}}. \quad (2)$$

In many situations, the denominator of (2) is not a well-known integral and must be calculated numerically by, for example, MCMC. Let $\mathbf{\Theta}_{-v}$ be $\mathbf{\Theta}$ with the vth element removed. A successive substitution implementation of the MCMC algorithm proceeds as follows:

1. Initialize $\mathbf{\Theta}^{(0)}$ and set $t = 1$.
2. Set $v = 1$.
3. Generate an observation $\boldsymbol{\theta}_v^{(t)}$ from the distribution of $[\theta_v|\mathbf{\Theta}_{-v}^{(t-1)}]$, replacing recently generated elements of $\mathbf{\Theta}_{-v}^{(t-1)}$ with elements of $\mathbf{\Theta}_{-v}^{(t)}$ if they have been generated.
4. Increment v by 1 and go to step 3 until $v = n$.
5. If $v = n$ increment t by 1 go to step 2.

Under conditions outlined by Hastings (1970), as $t \to \infty$ the distribution of $\{\theta_1^{(t)}, \ldots, \theta_n^{(t)}\}$ tends to the joint posterior distribution of $\mathbf{\Theta}$, as desired.

Typical implementation of the algorithm generates an initial "large" number of iterations (called the "burn-in") until the observations have converged. The burn-in samples are discarded, and the observations generated thereafter are used as observations from the posterior distribution of $\mathbf{\Theta}$. Nonparametric density estimators (Silverman 1986) can then be used to approximate the posterior distribution.

A.2 Metropolis–Hastings

Some complete conditional distributions may not be available in closed form; that is, it may be difficult to sample from $[\theta_v|\mathbf{\Theta}_{-v}^{(t-1)}] \propto g(\theta_v)$. Obtaining observations from such distributions is facilitated by implementing a *Metropolis–Hastings* step (Hastings 1970) for step 3 in the algorithm given in Section A.1. This is difficult, because the distribution is known only up to a constant. The procedure is as follows:

1. Initialize $\theta_{v_{\text{old}}}^{(0)}$ and set $j = 0$.
2. Generate an observation $\theta_{v_{\text{new}}}^{(j)}$ from a *candidate* distribution $q(\theta_{v_{\text{old}}}^{(j)}, \theta_{v_{\text{new}}}^{(j)})$, where $q(x, y)$ is a probability density in y with mean x.
3. Generate a uniform(0, 1) observation u.
4. Let

$$\theta_{v_{\text{new}}}^{(j+1)} = \begin{cases} \theta_{v_{\text{new}}}^{(j)} & \text{if } u \le \alpha\left(\theta_{v_{\text{old}}}^{(j)}, \theta_{v_{\text{new}}}^{(j)}\right) \\ \theta_{v_{\text{old}}}^{(j)} & \text{otherwise,} \end{cases}$$

   where $\alpha(x, y) = \min\{\frac{g(y)q(y,x)}{g(x)q(x,y)}, 1\}$.
5. Increment j and go to step 2.

The candidate distribution can be almost any distribution (Gilks, Richardson, and Spiegelhalter 1996), although a symmetric distribution such as the normal results in a simplification of the algorithm, and is called a *Metropolis step* (as opposed to a Metropolis–Hastings step). A common choice for $q(x, y)$ is a normal distribution with mean x and some variance which allows the random deviates to be a representative sample from the entire complete conditional distribution. A rule of thumb given by Gilks et al. (1996) suggests that the variance in $q(x, y)$ be one-third of the sample variance of the observed data.

[Received June 2000. Revised October 2003.]

REFERENCES

Agliari, A., and Parisetti, C. C. (1988), "A g Reference Informative Prior: A Note on Zellner's g-Prior," *The Statistician*, 37, 271–275.

Berry, D. A. (1996), *Statistics: A Bayesian Perspective*, New York: Duxbury Press.

Broemeling, L. (1985), *Bayesian Analysis of Linear Models*, New York: Marcel Dekker.

Chipman, H., Hamada, M., and Wu, C. F. J. (1997), "A Bayesian Variable-Selection Approach for Analyzing Designed Experiments With Complex Aliasing," *Technometrics*, 39, 372–381.

Craig, P. S., Goldstein, M., Rougier, J. C., and Seheult, A. H. (2001), "Bayesian Forecasting for Complex Systems Using Computer Simulators," *Journal of the American Statistical Association*, 96, 717–729.

Currin, C., Mitchell, T. J., Morris, M. D., and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, 86, 953–963.

Dewettinck, K., Visscher, A. D., Deroo, L., and Huyghebaert, A. (1999), "Modeling the Steady-State Thermodynamic Operation Point of Top-Spray Fluidized Bed Processing," *Journal of Food Engineering*, 39, 131–143.

Draper, D., Gaver, D. P., Goel, P. K., Greenhouse, J. B., Hedges, L. V., Morris, C. N., Tucker, J. R., and Waternaux, C. M. (1992), "Selected Statistical Methodology for Combining Information (CI)," in *Combining Information: Statistical Issues and Opportunities for Research*, eds. D. Cochran and J. Farrally, Washington, DC: National Academy Press, Chap. 4.

Draper, N. R., and Smith, H. (1998), *Applied Regression Analysis*, New York: Wiley.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, London: Chapman & Hall.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Hedges, L. V., and Olkin, I. (1987), *Statistical Methods for Meta Analysis*, New York: Wiley.

Kennedy, M. C., and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 63, 425–464.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code," *Technometrics*, 21, 239–245.

Meyer, M. A., and Booker, J. M. (1990), *Eliciting and Analyzing Expert Judgment*, Washington, DC: U.S. Nuclear Regulatory Commission.

Müller, P., Parmigiani, G., Schildkraut, J., and Tardella, L. (1999), "A Bayesian Hierarchical Approach for Combining Case-Control and Prospective Studies," *Biometrics*, 55, 858–866.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409–423.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer-Verlag.

Shapiro, S. (1994), "Meta-Analysis/Shmeta-Analysis," *American Journal of Epidemiology*, 140, 771–791.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

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.

Zeckhauser, R. (1971), "Combining Overlapping Information," *Journal of the American Statistical Association*, 66, 91–92.

Zellner, A. (1986), "On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions," in *Bayesian Inference and Decision Techniques*, eds. P. Goel and A. Zellner, New York: Elsevier Science, pp. 233–243.