LA-UR-00-2915

Title: Integrated Analysis of Computer and Physical Experiments

Author(s): C. Shane Reese, TSA-1
Alyson G. Wilson, TSA-1
Michael S. Hamada, TSA-1
Harry F. Martz, TSA-1
Kenneth J. Ryan, Iowa State University

Submitted to: Technometrics

Los Alamos
NATIONAL LABORATORY

Form **836 (10/96)**

Integrated Analysis of Computer and Physical Experiments

C. Shane Reese, Alyson G. Wilson, Michael Hamada, Harry F. Martz, and Kenneth J. Ryan *

September 6, 2000

Abstract

In scientific investigations, the situation in which we have data from computer experiment(s) as well as related physical experimental data on the same factors and related response variable(s) is arising more frequently. In addition, there may also exist one or more expert opinions regarding the response of interest. Traditional statistical approaches consider each of these sets of data 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 we simultaneously analyze the combined data 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 the methodology by using it to model the thermodynamic operation point of a top-spray fluidized bed microencapsulation processing unit. Such units are increasingly being 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 were 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 use in predicting Y.

1 Introduction

Computer models are often used to perform experiments before expensive physical experiments are performed. The computer models attempt to reproduce the physical properties of a process by mathematically representing the individual physical sub-processes. For example, in the food industry, fluidized bed processes are increasingly used to coat food particles with preservatives and flavor enhancers. The operation of such fluidized beds is based on physics principles such

*Shane Reese, Alyson Wilson, Michael Hamada, and Harry Martz are Technical Staff Members in the Statistical Sciences group at Los Alamos National Laboratory, Los Alamos, NM, 87545. Kenneth Ryan is a graduate student in the Department of Statistics at Iowa State University, Ames, IA, 50011-1210. The authors wish to thank Val Johnson for his valuable comments.

as heat transfer, fluid flow, and other less well understood principles. As a result, stochastic computer models, based on these principles of physics, are constructed which resemble and simulate the actual physical process. In this paper we analyze data which were collected from three such computer models (each accounting for different effects) as well as data which were 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 expresses the relationship between the factors and the response variable in the combination of the physical experimental data as well as the computer model output data. We assume that, although the response variables of interest in the computer and physical experiments may not be the same, they can be related by a known transfer function. Thus, we effectively consider the same response variable in both types of experiments. However, we do not impose 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. For example, it is sometimes the case that a broad (screening) computer experiment is first performed, that is to be followed later 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 sets of data 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 will illustrate, the simultaneous analysis of such combined data permits the unknown coefficients in an assumed overall regression (or response surface) model to be more precisely estimated, thereby producing a better fitting response surface.

In Section 2 we present the methodology including our implementation of the RBHM. Section

3 contains a description of the mechanics and process variables involved in the fluidized bed example, as well as describe 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. Finally, in Section 4 we discuss the results and methodology.

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 is given in Berry (1996). An RBHM provides a convenient way to sequentially combine the data as follows. A set of initial informative, but diffuse, prior distributions are initially defined for each unknown parameter. If they exist, any available expert opinion data 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 of the available data are recursively used within the context of the model to successively (and more precisely) estimate all 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 et al. (1989) provide 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). Bayesian treatment of design and analysis of computer experiments is presented in Currin et al. (1991). These papers are primarily concerned 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) provides an early treatment of meta analysis. Hedges and Olkin (1987) provide a nice review of meta analytic techniques. Meta analysis has not been viewed without strong criticism (Shapiro 1994 and discussion). Müller et al. (1999) present 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.

The statistical notion of pooling data (sometimes also known as "borrowing strength") underlies the RBHM and analysis to be discussed. Modern methods used to borrow strength have their basis in *hierarchical Bayes modeling*. A nice introduction to both hierarchical Bayes modeling and borrowing strength is given in 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. The practical advantages of borrowing strength for estimating the unknown parameters will be illustrated 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, while computer experiments may be collected at (possibly) different design points. In addition, expert opinions may only be available at a very limited set of design points, such as the center of the statistical design region. Our goal is to combine these sources of information using an appropriately flexible integration methodology which 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:

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

where $\mathbf{X}$ is a design matrix, $\underline{\beta}$ is a vector of unknown coefficients, and $\underline{\varepsilon}$ is a vector of unobserved errors. Note that while this formulation can accomodate a general class of models, $f(\cdot)$, including both linear and nonlinear regression models, in this and the following sections we consider only linear models (that is, $f(\mathbf{X}, \underline{\beta}) = \mathbf{X}\underline{\beta}$).

2.1 Physical Experimental Data

Physical experimental data forms the backbone of the analysis. We assume 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 following familiar model:

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

where the subscript p denotes the "physical experiment." Thus, the physical experimental data are assumed to be normally distributed with mean $\mathbf{X}_p\underline{\beta}$, where $\mathbf{X}_p$ is a model matrix and $\underline{\beta}$ 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 must also be estimated.

If physical experimental data were the only information source considered, this model would typically be fit using either standard least-squares regression methods (Draper and Smith 1981) or standard Bayesian linear model methods (Gelman et al. 1995). However, we want to incorporate information both from experts and computer experimental data to "improve" our estimates of $\underline{\beta}$ and σ^2.

2.2 Expert Opinion

Suppose there are e expert opinions. These opinions do not have to be from distinct experts. The i^{th} expert opinion ($i = 1 \ldots e$) is elicited at design point x_i. 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} (i.e., $Pr(y_i \leq q_{\xi_i}) = \xi_i$)

In addition, we consider the elicited "worth" of the opinion in units of equivalent physical experimental data observations, $m_{o_i}^{(e)}$. In order to use this data, we need to transform these individual pieces of information into probability distributions that provide information about $\underline{\beta}$ and σ^2. Assume for the moment that the three quantities above can be used to create "data" with the following model:

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

As with the physical experimental data, the expert data are assumed to be normally distributed. However, the mean is $\mathbf{X}_o\underline{\beta} + \underline{\delta}_o$, where $\underline{\delta}_o$ is a vector of location biases that are expert specific. The variances are also biased, and the matrix $\mathbf{\Sigma}_o$ 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 over-valuation and are well-documented in the elicitation literature. For example, an expert may be asked to provide what they think is a 0.90 quantile which in reality is actually only a 0.60 quantile (Meyer and Booker 1990). Although responses from experts can be correlated by having non-diagonal elements in $\mathbf{\Sigma}_o$, however, we consider uncorrelated responses; thus,

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

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

$$\begin{aligned} \underline{\beta}|\sigma^2 &\sim N(\underline{\mu}_o, \sigma^2\mathbf{C}_o) \\ \sigma^2 &\sim IG(\alpha_o, \gamma_o), \end{aligned}$$

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

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

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

$$\begin{aligned}
\pi(\underline{\beta}|\sigma^2, \underline{\eta}, \underline{y}_o) \quad &\sim \quad N\Big((\mathbf{X}_o'\mathbf{\Sigma}_o^{-1}\mathbf{X}_o + \mathbf{C}_o^{-1})^{-1}(\mathbf{X}_o'\mathbf{\Sigma}_o^{-1}(\underline{y}_o - \underline{\delta}_o) + \mathbf{C}_o^{-1}\underline{\mu}_o), \\
&\qquad \sigma^2(\mathbf{X}_o'\mathbf{\Sigma}_o^{-1}\mathbf{X}_o + \mathbf{C}_o^{-1})^{-1}\Big) \\
\pi(\sigma^2|\underline{\eta}, \underline{y}_o) \quad &\sim \quad IG\Bigg(\alpha_o + \frac{\sum_{i=1}^{e} m_{o_i}}{2}, \\
&\qquad \gamma_o + .5[(\underline{y}_o - \underline{\delta}_o)'\mathbf{\Sigma}_o^{-1}(\underline{y}_o - \underline{\delta}_o) + \underline{\mu}_o'\mathbf{C}_o^{-1}\underline{\mu}_o - \\
&\qquad (\mathbf{X}_o'\mathbf{\Sigma}_o^{-1}(\underline{y}_o - \underline{\delta}_o) + \mathbf{C}_o^{-1}\underline{\mu}_o)'(\mathbf{X}_o'\mathbf{\Sigma}_o^{-1}\mathbf{X}_o + \mathbf{C}_o^{-1})^{-1}(\mathbf{X}_o'\mathbf{\Sigma}_o^{-1}(\underline{y}_o - \underline{\delta}_o) + \mathbf{C}_o^{-1}\underline{\mu}_o)]\Bigg).
\end{aligned}$$

Given that the full vector of observations $\underline{y}_o$ was not elicited (only sufficient statistics were), we cannot immediately evaluate any term involving this expression. However, consider the following. Suppose m_{o_i} observations were elicited as $\underline{y}_{o_i}$ from the i^{th} expert opinion. Then

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

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

Using a similar argument, we can show that

$$\begin{aligned}(\underline{y}_o - \underline{\delta}_o)'\mathbf{\Sigma}_o^{-1}(\underline{y}_o - \underline{\delta}_o) &= k_{o_1}(\sum_{n=1}^{m_{o_1}}(y_{o_{jn}} - \delta_{o_1})^2) + \ldots + k_{o_e}(\sum_{n=1}^{m_{o_e}}(y_{o_{jn}} - \delta_{o_e})^2) \\ &= \sum_{i=1}^{e} k_{o_i} m_{o_i}(s_i^2 + (y_{o_i} - \delta_{o_i})^2), \qquad (1)\end{aligned}$$

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 $Var(Y) = E[Y^2] - E[Y]^2$.

By a similar argument

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

These representations allow the quantities in the posterior distributions to be calculated based on the elicited values rather than the actual observations.

There are several unknown parameters $\underline{\eta} = (\underline{\delta}_o, \mathbf{\Sigma}_o, \underline{m}_o, \mathbf{C}_o, \underline{\mu}_o, \alpha_o, \gamma_o)$ for which we propose the following prior distributions:

$$\begin{aligned}\underline{\mu}_o &= \underline{k}_{\mu_o} \\ \mathbf{C}_o &= k_{C_o}\mathcal{I} \\ \alpha_o &= k_{\alpha_o} \\ \gamma_o &= k_{\gamma_o} \\ m_{o_i} &\sim Uniform(0.5m_{o_i}^{(e)}, 2.0m_{o_i}^{(e)}) \\ \delta_{o_i} &\overset{iid}{\sim} N(\theta_o, \xi_o^2) \\ \theta_o &\sim N(m_{\theta_o}, s_{\theta_o}^2) \\ \xi_o^2 &\sim IG(a_{\xi_o^2}, b_{\xi_o^2}) \\ k_{o_i} &\overset{iid}{\sim} G(\phi_o, \omega_o) \\ \phi_o &\sim G(a_{\phi_o}, b_{\phi_o}) \\ \omega_o &\sim G(a_{\omega_o}, b_{\omega_o}),\end{aligned}$$

where $G(a, b)$ indicates a gamma distribution with mean ab and variance ab^2.

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 $(\underline{\beta}, \sigma^2)$. However, Zellner (1986) elicits posterior means for $\underline{\beta}$ and σ^2, while we elicit predicted observations $\underline{y}_o$. Agliari and Parisetti (1988) extend Zellner's methods to include a different design matrix 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 $\underline{\beta}$ and σ^2. We continue to update our knowledge about these parameters using data from computer experiments. Consider the following model:

$$\begin{aligned} \underline{Y}_c &\sim N(\mathbf{X}_c\underline{\beta} + \underline{\delta}_c, \sigma^2\mathbf{\Sigma}_c) \\ \underline{\beta}|\sigma^2 &\sim N(\underline{\mu}_c, \sigma^2\mathbf{C}_c) \\ \sigma^2 &\sim IG(\alpha_c, \gamma_c). \end{aligned}$$

The "prior" distributions for $\underline{\beta}|\sigma^2$ and σ^2 are the Stage 1 posterior distributions given the expert opinion data. The only other unspecified prior distributions are:

$$\begin{aligned} \delta_{c_i} &\overset{iid}{\sim} N(\theta_c, \xi_c^2) \\ \theta_c &\sim N(m_{\theta_c}, s_{\theta_c}^2) \\ \xi_c^2 &\sim IG(a_{\xi_c^2}, b_{\xi_c^2}) \\ k_{c_i} &\overset{iid}{\sim} G(\phi_c, \omega_c) \\ \phi_c &\sim G(a_{\phi_c}, b_{\phi_c}) \\ \omega_c &\sim G(a_{\omega_c}, b_{\omega_c}). \end{aligned}$$

In addition to location biases, computer experimental data are likely to have scale biases, as 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, since 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. Although we consider biases only in the intercept term of this model, more general bias structures for the parameters can also be modeled. In Section 3, we motivate these ideas by introducing the operation of fluidized beds and the computer models for that process.

2.4 Incorporating Physical Experimental Data

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

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

After incorporating the computer experimental data into the analysis, we have a Stage 2 posterior that is used as the prior for $(\underline{\beta}, \sigma^2)$ 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). See the Appendix for general information on MCMC and the Metropolis-Hastings algorithm.

3 Application of RBHMs to Fluidized Bed Processes

Fluidized bed processes are used in the food industry to coat certain food products with additives. Dewettinck et al. (1999) describe a physical experiment and several corresponding thermodynamic computer models that were developed for use in 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, and beneath the screen there is an air

pump. Also, there are coating sprayers at the side of the unit.

[Figure 1 about here]

To use the unit, a batch of uncoated food product is placed inside. Then, 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 a 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; namely,

- 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) consider twenty-eight 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 $0\%w/w$ for all 28 runs. Also, T_s was at room temperature for all 28 runs. Table 1 shows 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).

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

There are important differences between the three computational models that are described in detail in 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.

[Table 1 about here]

[Table 2 about here]

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 $T_{2,exp}$ ($r = 0.73$).

[Table 3 about here]

Choice of a model is complicated by the fact that the underlying design is not at all clear. Additionally, the covariance matrix reveals that some of the covariates are highly correlated (as high as 0.82) indicating possible collinearity. Also, we note that the full second-order model is fully saturated.

[Figure 2 about here]

Chipman, Hamada, and Wu (1997) describe a Bayesian variable selection procedure which 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, 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 $\mathbf{X}\underline{\beta}$, where $\mathbf{X}$ is composed of a column of ones (for the intercept) and columns corresponding to T_a, R_f, V_f, and $R_f \times V_f$, whose respective regression parameters are $\underline{\beta} = (\beta_0, \ldots, \beta_4)'$.

[Table 4 about here]

Table 5 contains the OLS fit of the most likely model in Table 4.

[Table 5 about here]

In our example the hyperparameter values are given in Table 6. Note that the same hyperparameters were used for all three computer experiments. As we have no prior knowledge as to the sign of the location bias, we center the distribution of δ_{c_i} at zero (that is, unbiased in location), and allow the mean of that distribution to have a standard deviation of 10. While we believe the computer models are all reasonably good approximations of the physical model, we don't have a good idea about the degree of separation, and thus allow a generous variability for the location biases ($a_{\xi_c^2} = 2000$ and $b_{\xi_c^2} = 3.0$ suggest a mean variability of 2000/(3-1)=1000 and a standard deviation of 2000/((3-1)*(3-2))=1000). 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 the standard deviation of the scale biases should be no greater than 15, thus we let the mean of the scale bias distribution be 1 and the standard deviation 15. This allows a generous range for the scale biases.

[Table 6 about here]

3.3 RBHM Analysis Results

Figures 3-7 show the posterior for $\underline{\beta}$ with only the physical experimental data, the physical with each of the computer experimental data taken separately, and the final posterior distribution

for $\underline{\beta}$ after incorporating all sources of information. Figure 8 shows the corresponding posteriors for σ^2. The figures indicate two important points 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 sources do not necessarily contain the same amount of information (although, in our example, they do have the same number of observations). These are appealing results of the methodology.

In addition to posterior distributions for $\underline{\beta}$ and σ^2, our modeling approach allows us to estimate the bias terms. As an illustration, Figures 9 and 10 present the location and scale bias predictive distributions for each of the computer models. Note that these distributions are integrated over the distribution of individual specific location 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 which have most mass over 0 are less location biased for the physical experimental data. For example, the bias is slightly more concentrated around 0 for the third computer model than for the other two computer models. An additional feature of these plots is that they also reveal the uncertainty associated with the bias terms (a feature which cannot easily be inferred from a casual examination of the data). Note that the third model is the computer model which attempts to account for more phenomena. Figure 10 reveals that all three computer models tend to underestimate the variability in the physical experimental data (that is, the scale bias is less than unity).

[Figures 3-10 about here]

Table 7 contains the maximum likelihood estimate (MLE), 95% confidence intervals (calculated from only the physical experimental data), the posterior mean and 95% highest posterior density (HPD) intervals calculated using the integrated computer and physical experimental data for $\underline{\beta}$ 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.

[Table 7 about here]

4 Conclusions/Discussion

When expert opinion is elicited, an equivalent number of observations $m_{o_i}^{(e)}$ is also stated that gives an idea how many physical observations the expert opinion is "worth." 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 $\underline{\beta}$ and σ^2. If these parameters are used to specify a very diffuse ("non-informative") 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.

We have presented an RBHM that can be used to combine data from both computer and physical experiments. When available, expert opinion data is 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 (this is, averaged out of the analysis using appropriate priors).

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.

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.

Carlin, B. P. and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall.

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.

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.

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

Dewettinck, K., Visscher, A.D., Deroo, L., and Huyghebaert, A. (1999), "Modelling 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), "Chapter 4: Selected Statistical Methodology for Combining Information (CI)" in *Combining Information: Statistical Issues and Opportunities for Research*, eds. Cochran, D. and Farrally, J., National Academy Press.

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 Ap-

plications," *Biometrika*, 57, 97-109.

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

Malakoff, D. (1999), "Bayes Offers a 'New' Way to Make Sense of Numbers," *Science*, 286, 1460-1464.

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 Judgement*, U.S. Nuclear Regulatory Commission, Washington.

M*ü*ller, 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.

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 and Hall.

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," *Bayesian Inference and Decision Techniques*, eds. P. Goel and A. Zellner, New York: Elsevier Science.

A Appendix

A.1 Markov Chain Monte Carlo (MCMC)

Suppose we are interested in making statistical inference about a parameter (possibly vector valued) Θ. We may have some information (or lack of information) about the distribution of Θ which we will call $\pi(\Theta)$ (prior distribution). Data are collected and represented by the likelihood or $f(\mathbf{x}|\Theta)$. In any Bayesian analysis, inference on the parameters is carried out by calculating the posterior distribution

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

In many situations, the denominator of (2) is not a well known integral and must be calculated numerically. In cases where the denominator cannot be calculated explicitly, a technique known as Markov Chain Monte Carlo (MCMC) can be employed. The technique proceeds by letting $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ be an n dimensional vector, and Θ_{-v} be Θ with the v^{th} element removed. A successive substitution implementation of the MCMC algorithm proceeds as follows:

(1) Initialize $\Theta^{(0)}$ and set $t = 1$.

(2) Set $v = 1$.

(3) Generate an observation $\theta_v^{(t)}$ from the distribution of $[\theta_v|\Theta_{-v}^{(t-1)}]$, replacing recently generated elements of $\Theta_{-v}^{(t-1)}$ with elements of $\Theta_{-v}^{(t)}$ if they have been generated.

(4) Increment v by 1 and go to (3) until $v = k$.

(5) If $v = k$ increment t by 1 go to (2).

As $t \to \infty$, and under conditions outlined in Hastings (1970), the distribution of $\{\theta_1^{(t)}, \ldots, \theta_k^{(t)}\}$ tends to the joint posterior distribution of Θ, 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 Θ. 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 $[\theta_v|\Theta_{-v}^{(t-1)}] \propto g(\theta_v)$. Obtaining observations from such distributions is facilitated by implementing a *Metropolis-Hastings* step (Hastings 1970) for step (3) in the algorithm above. This is difficult because the distribution is only known up to a constant.

(1) Initialize $\theta_{v_{old}}^{(0)}$ and set $j = 0$.

(2) Generate an observation $\theta_{v_{new}}^{(j)}$ from a *candidate* distribution $q(\theta_{v_{old}}^{(j)}, \theta_{v_{new}}^{(j)})$, where $q(x, y)$ is a probability density in y with mean x.

(3) Generate a uniform (0,1) observation u.

(4) Let

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

where $\alpha(x, y) = \min\left\{\frac{g(y)q(y,x)}{g(x)q(x,y)}, 1\right\}$.

(5) Increment j and go to (2).

The candidate distribution can be almost any distribution (Gilks et al. 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 in Gilks et al. (1996) suggests that the variance in $q(x, y)$ be one-third of the sample variance of the observed data.

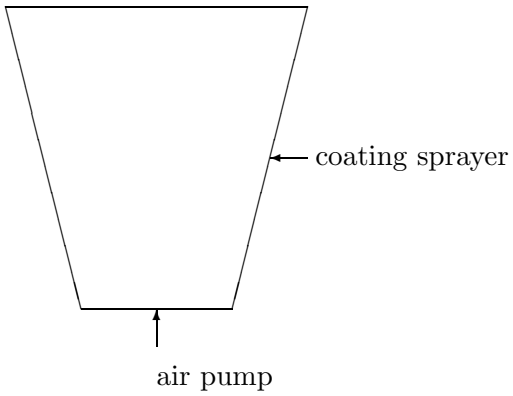

Figure 1: A Glatt GPCG-1 fluidized bed unit.

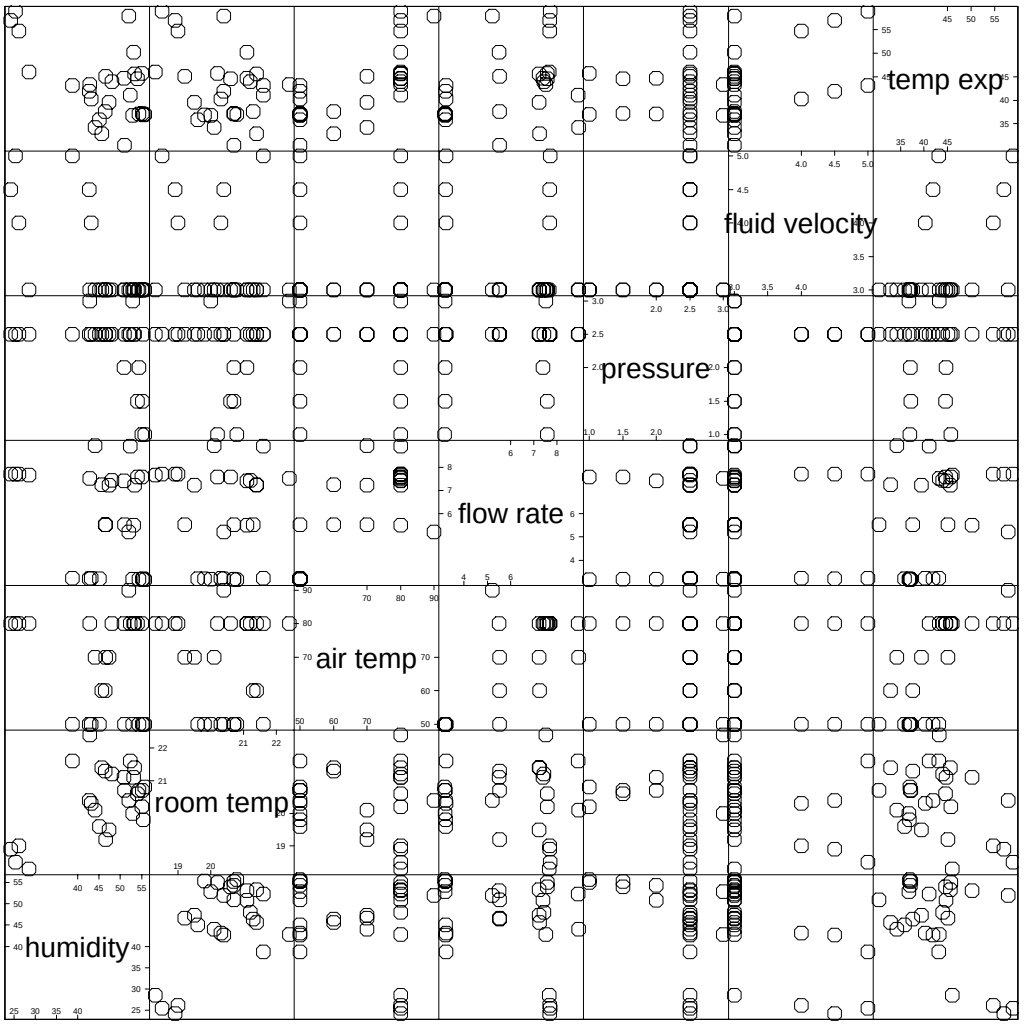

Figure 2: Scatterplot matrix of the experimental response with each of the 6 covariates.

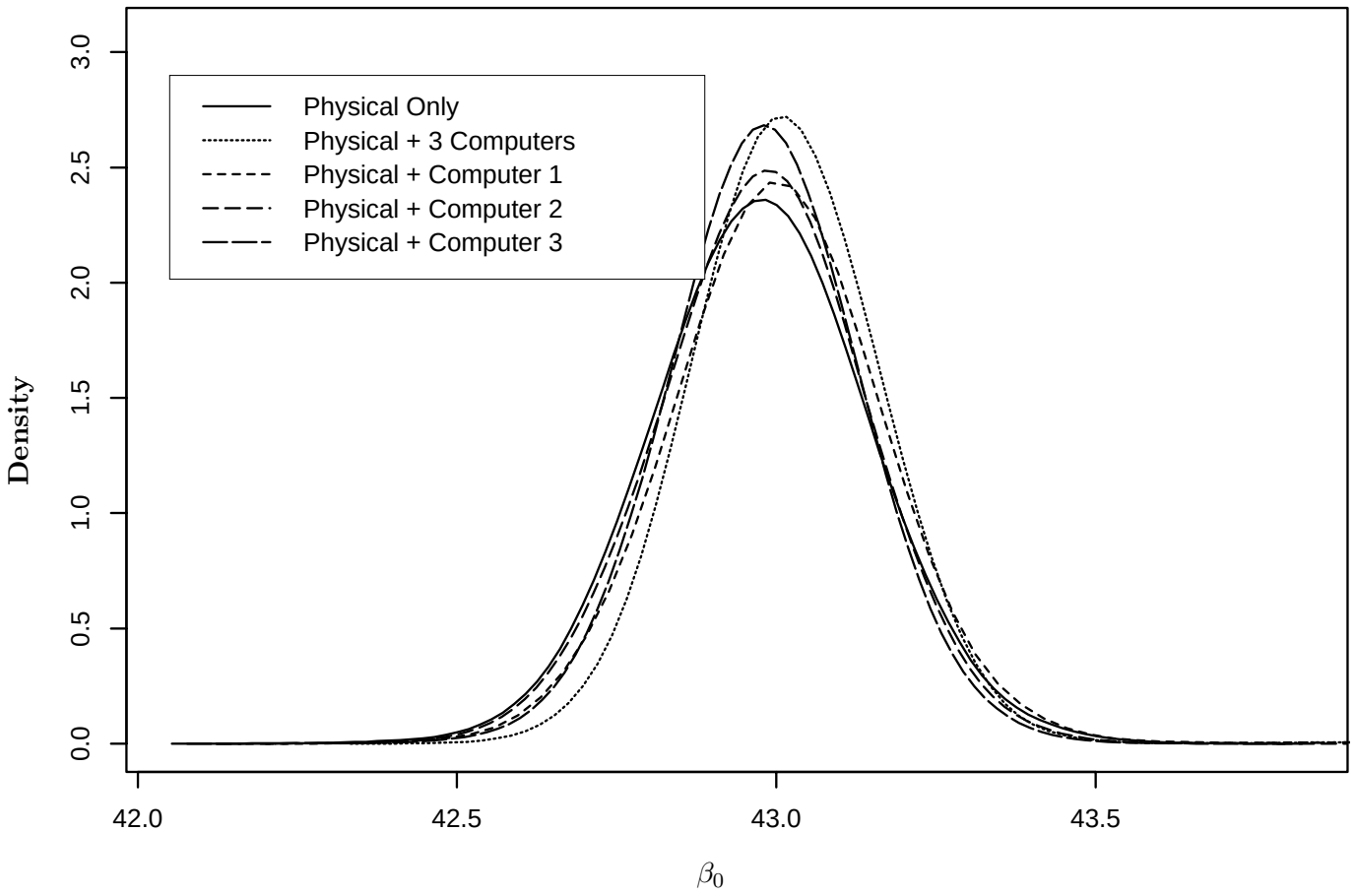

Figure 3: Comparison of posterior distributions for β_0 (intercept) conditional on different sources of information.

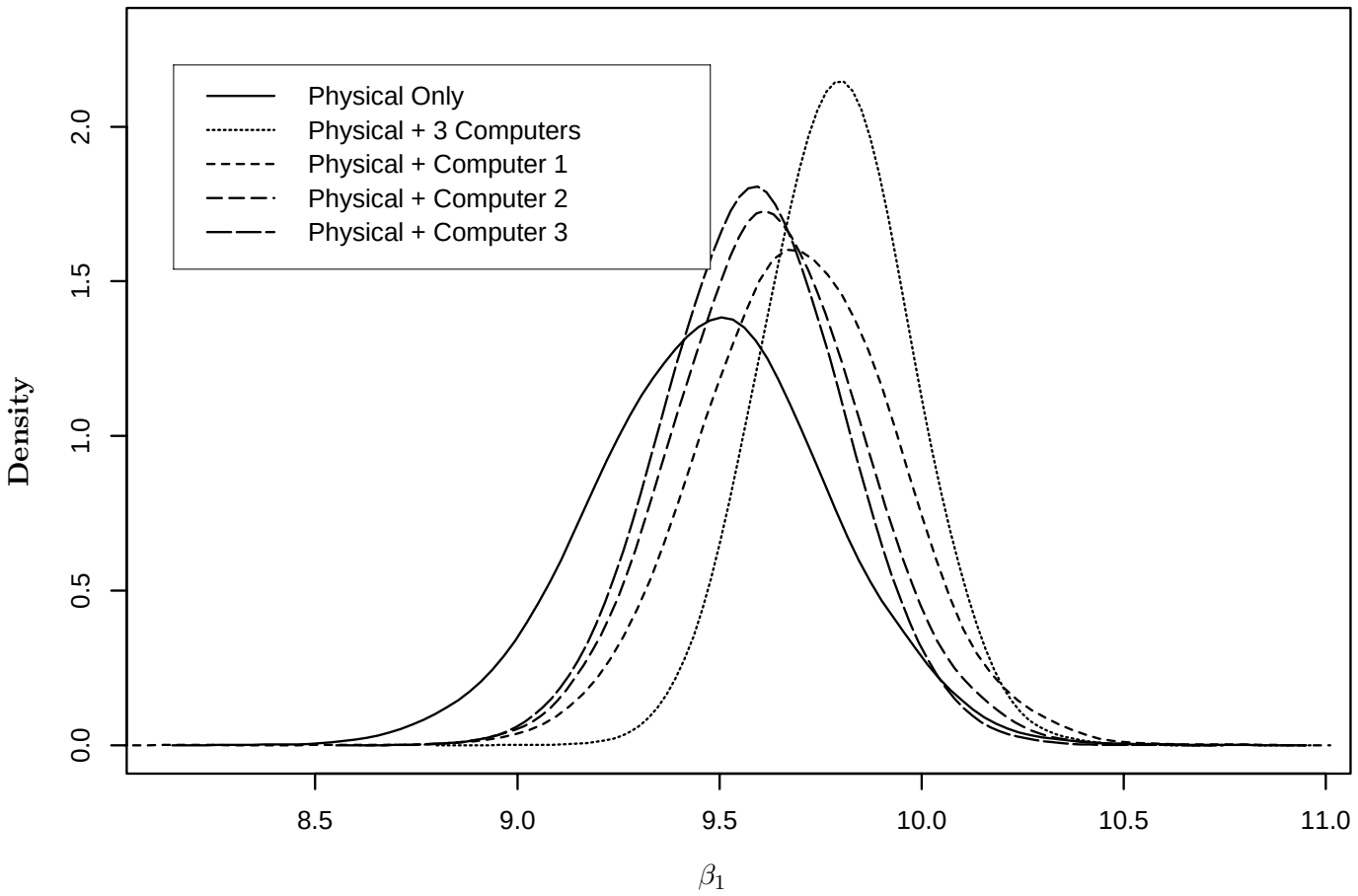

Figure 4: Comparison of posterior distributions for β_1 (regression coefficient for air temperature) conditional on different sources of information.

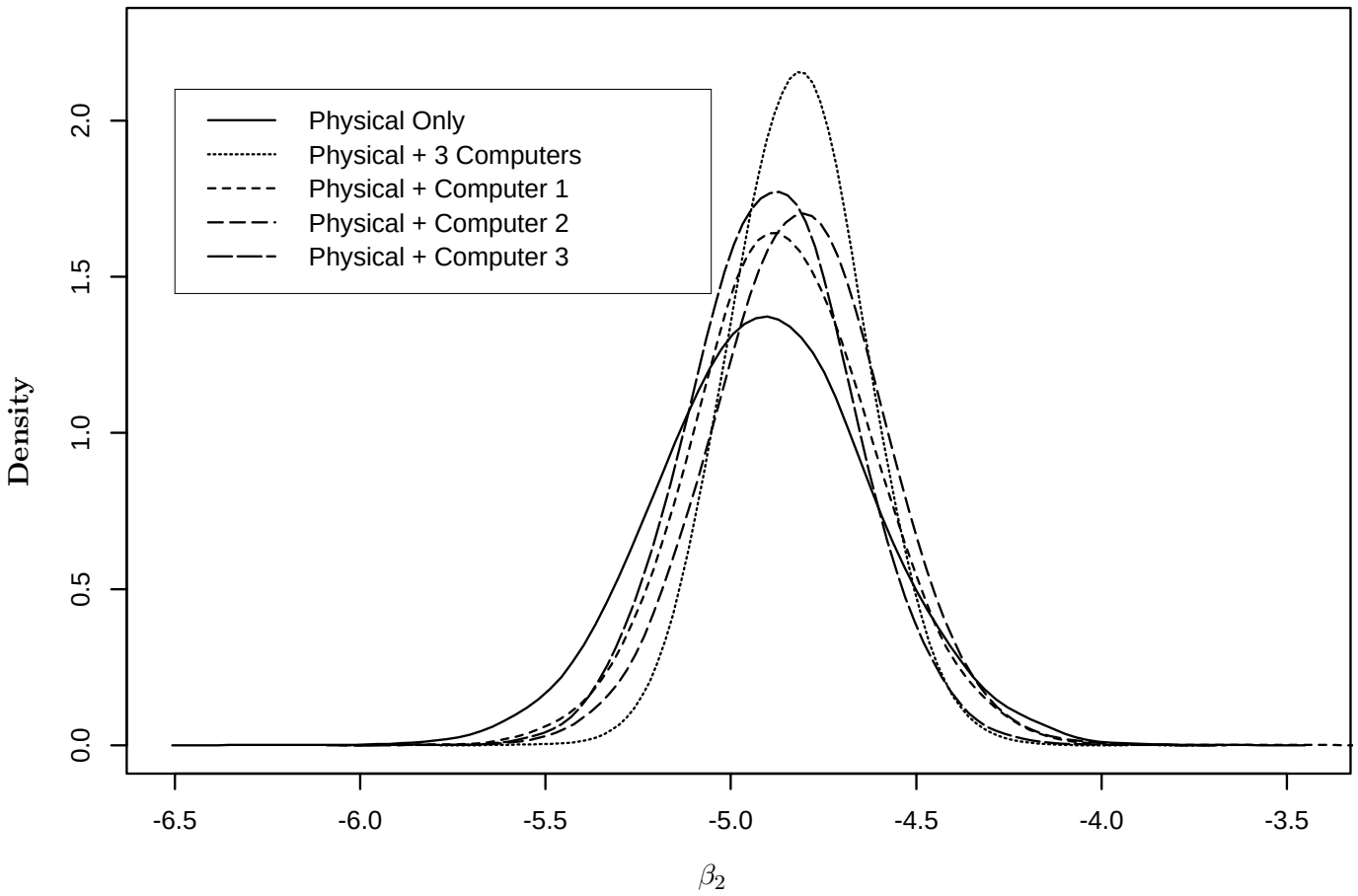

Figure 5: Comparison of posterior distributions for β_2 (regression coefficient for flow rate) conditional on different sources of information.

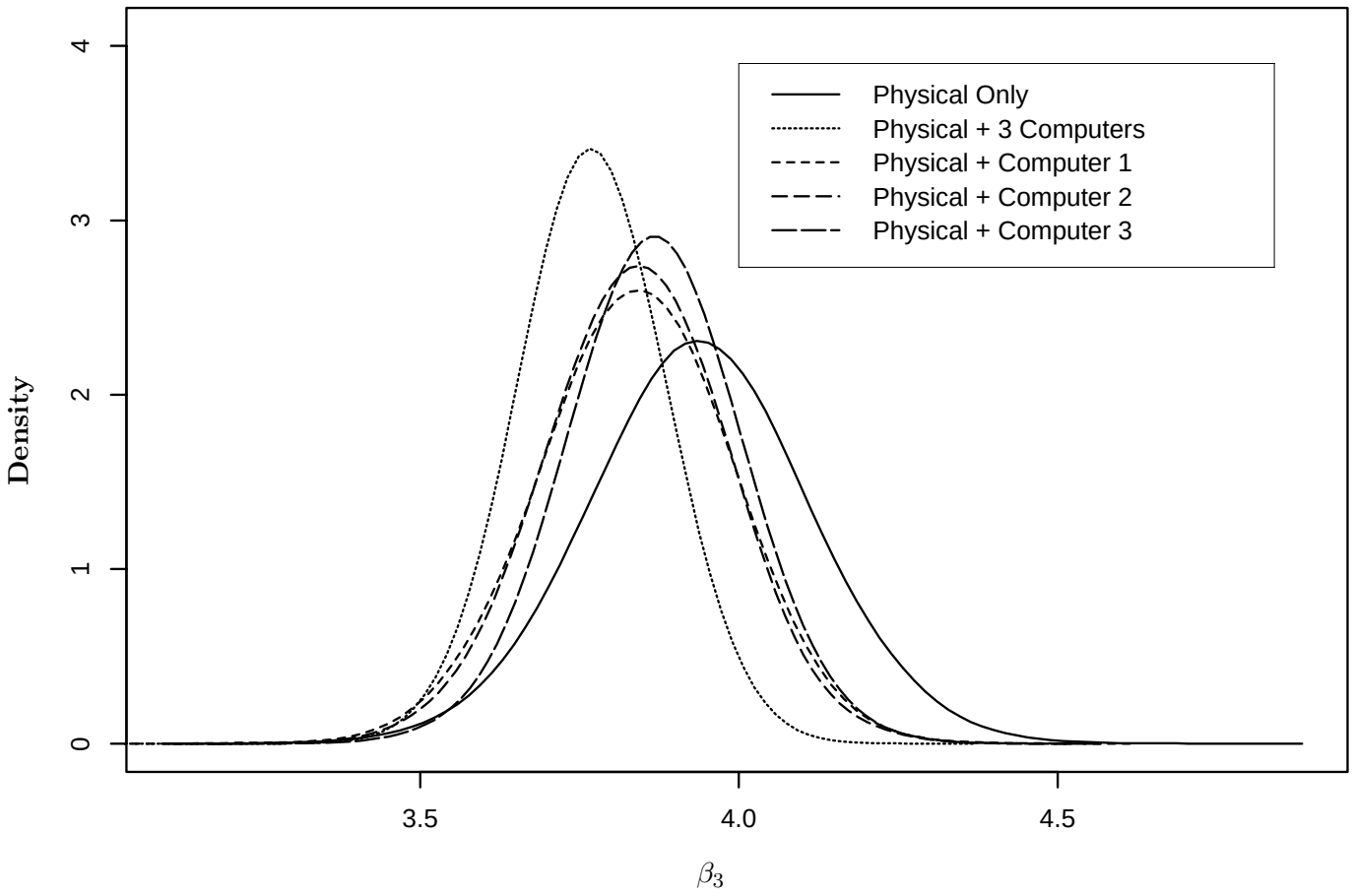

Figure 6: Comparison of posterior distributions for β_3 (regression coefficient for fluid velocity) conditional on different sources of information.

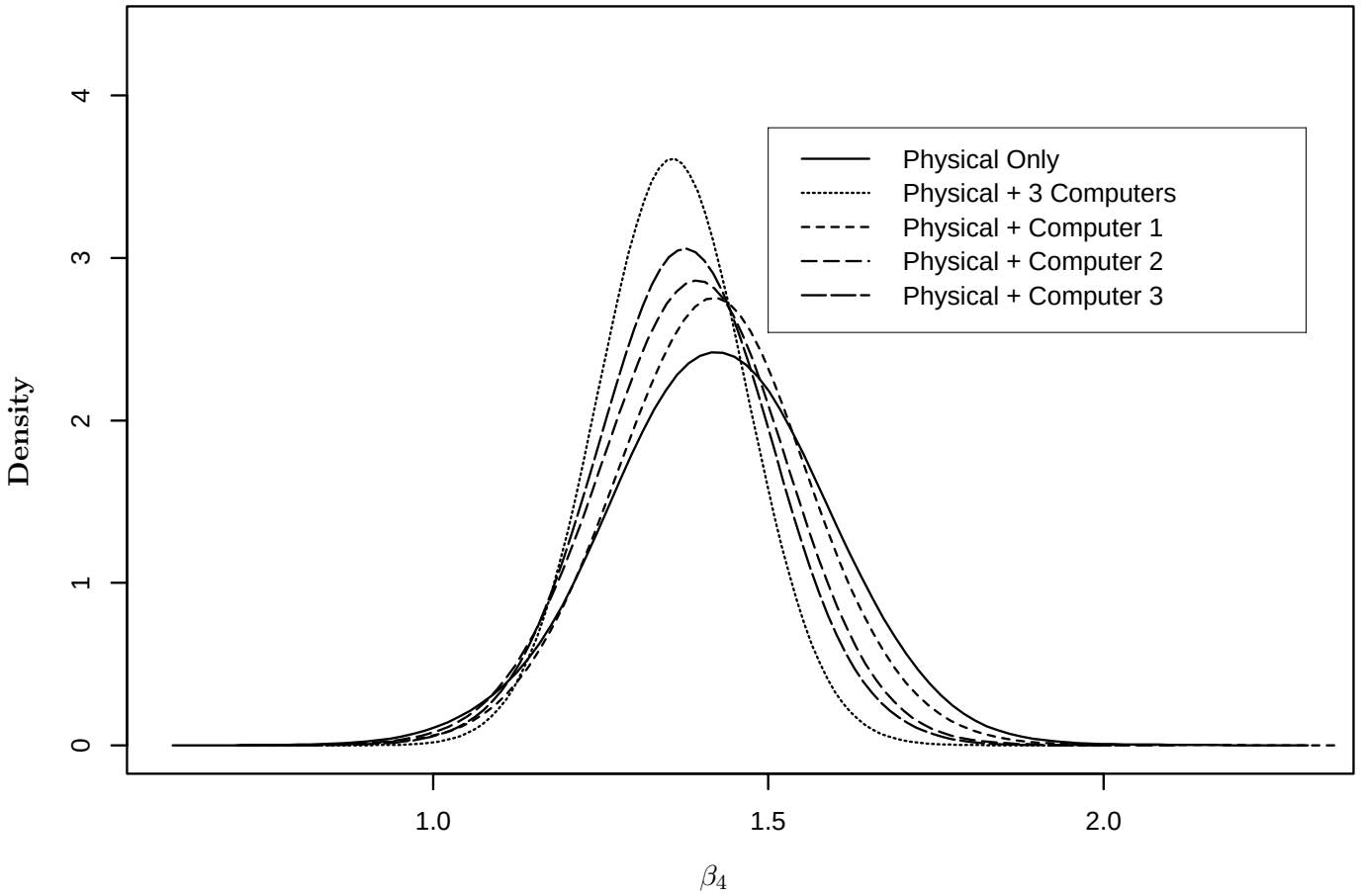

Figure 7: Comparison of posterior distributions for β_4 (regression coefficient for interaction between fluid velocity and flow rate) conditional on different sources of information.

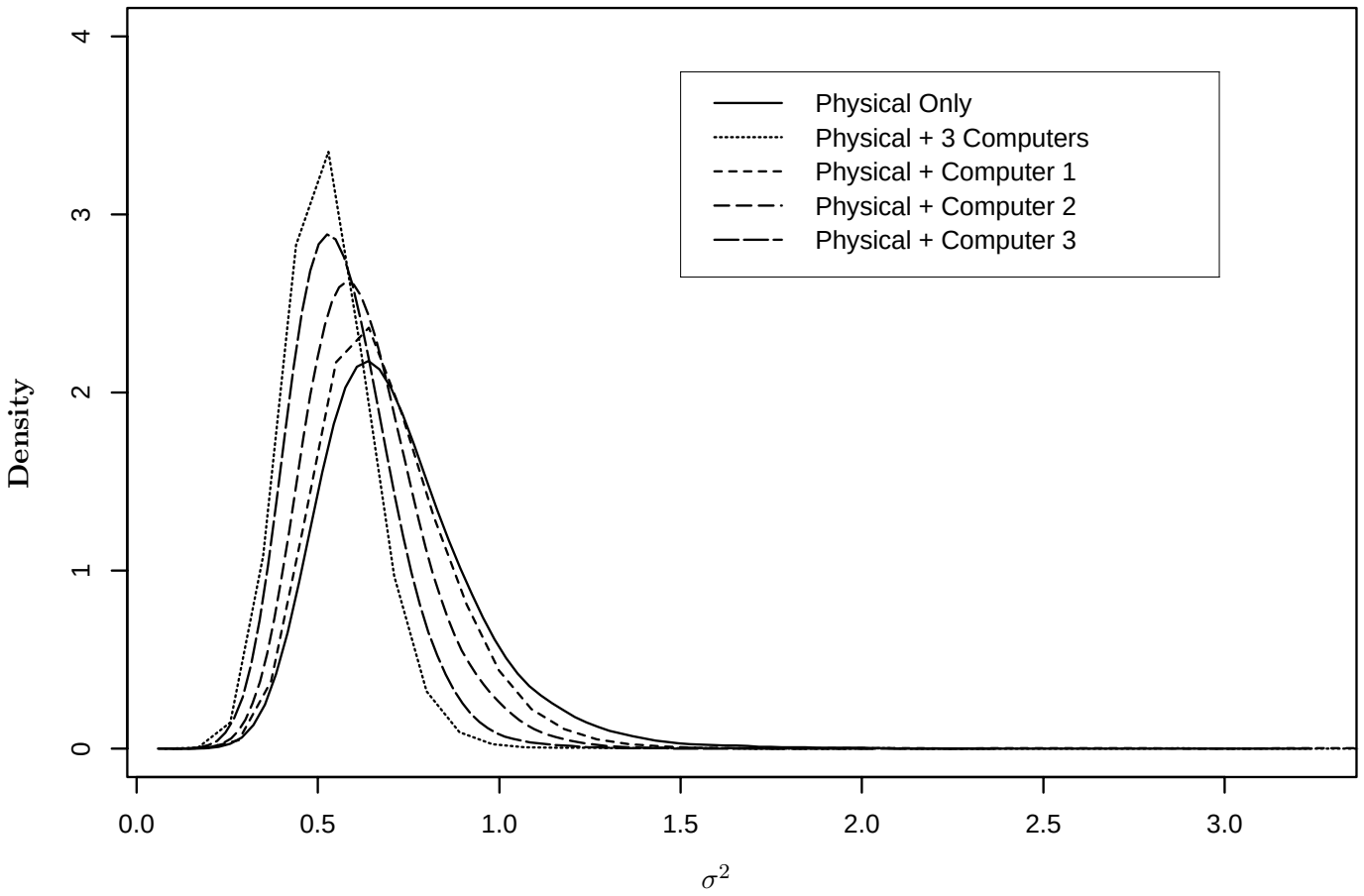

Figure 8: Comparison of posterior distributions for σ^2 (error variance) conditional on different sources of information.

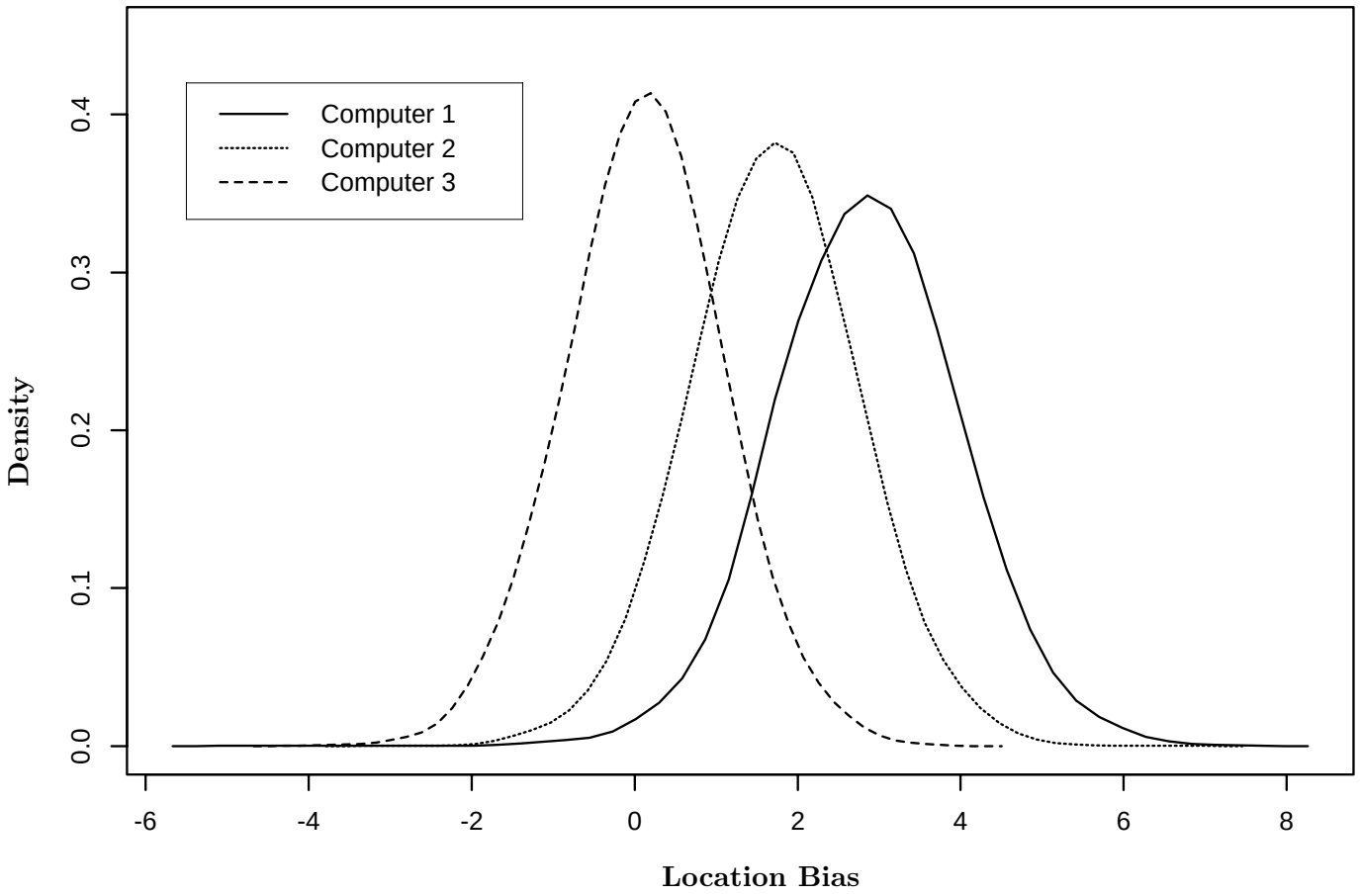

Figure 9: Comparison of location bias predictive distributions for three different computer models of the fluidized bed process.

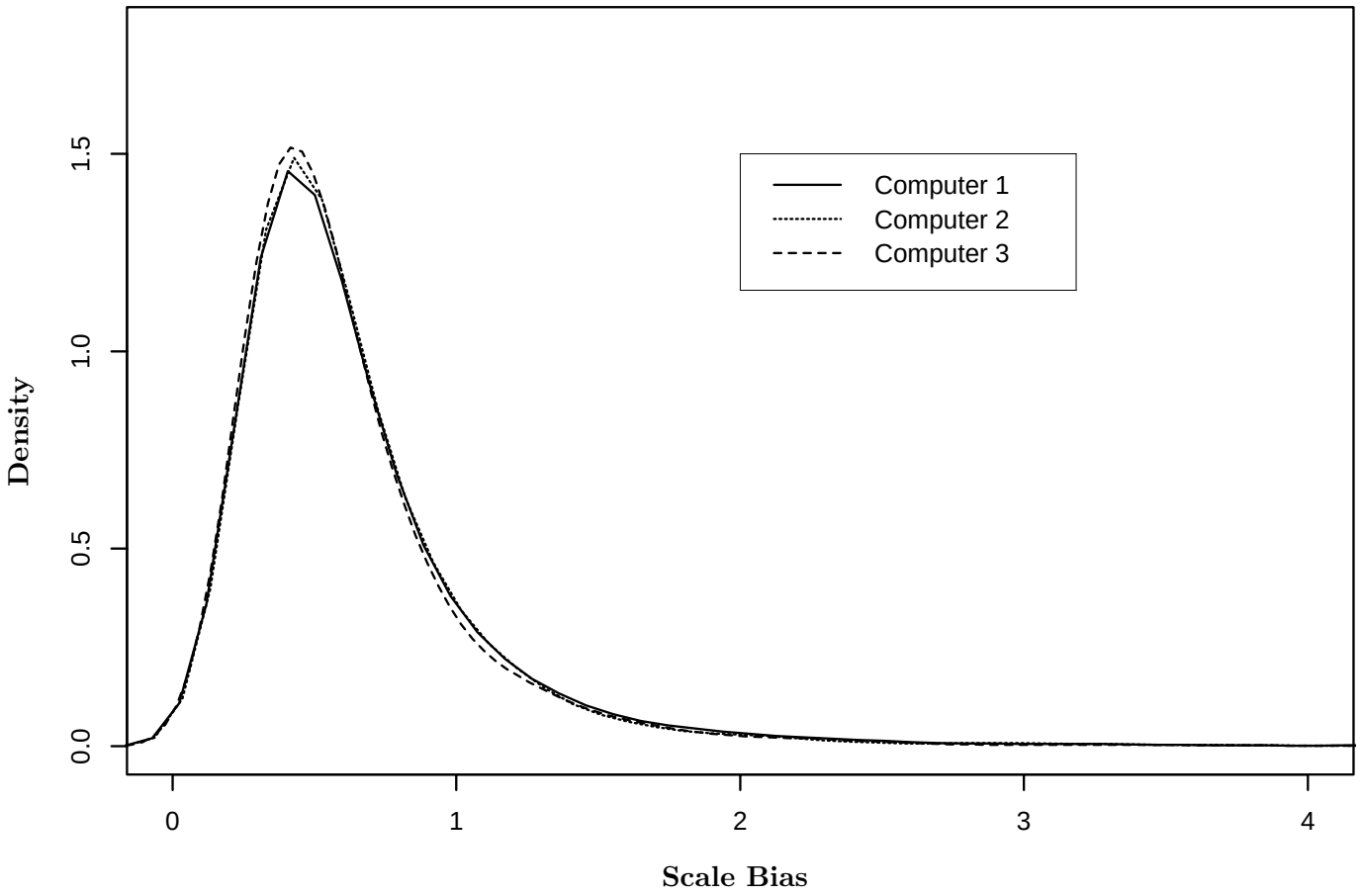

Figure 10: Comparison of scale bias predictive distributions for three different computer models of the fluidized bed process.

Table 1: Process Variables

H_r (%)	T_r ($^\circ C$)	T_a ($^\circ C$)	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

Table 2: Experimental and Computer Model Steady-State Temperatures

$T_{2,exp}$ ($^\circ C$)	$T_{2,1}$ ($^\circ C$)	$T_{2,2}$ ($^\circ C$)	$T_{2,3}$ ($^\circ C$)
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	0.57	−0.26	−0.33	−0.39	−0.69	−0.53
T_r	0.57	1.00	−0.09	−0.07	−0.04	−0.28	−0.37
T_a	−0.26	−0.09	1.00	0.82	0.06	−0.08	0.73
R_f	−0.33	−0.07	0.82	1.00	0.09	−0.10	0.35
P_a	−0.39	−0.04	0.06	0.09	1.00	0.18	0.08
V_f	−0.69	−0.28	−0.08	−0.10	0.18	1.00	0.47
$T_{2,exp}$	−0.53	−0.37	0.73	0.35	0.08	0.47	1.00

Table 4: Bayesian Variable Selection Results

Model	Pr(Model|Data)
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 R_f \times V_f$	0.1169
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 H_r \times T_r + \beta_5 R_f \times V_f$	0.0349
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 H_r^2$	0.0155
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 T_r \times T_a + \beta_5 R_f \times V_f$	0.0141
$\beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 R_f \times V_f + \beta_5 V_f^2$	0.0136
$\beta_0 + \beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 H_r^2$	0.0132
$\beta_1 T_r + \beta_2 T_a + \beta_3 R_f + \beta_4 V_f + \beta_5 R_f \times V_f$	0.0130

Table 5: OLS fit for $T_{2,exp} = \beta_0 + \beta_1 T_a + \beta_2 R_f + \beta_3 V_f + \beta_4 R_f \times V_f + \epsilon$

Variable	DF	Parameter Estimate	Standard Error	T for $H_\circ$: Parameter= 0	Prob> $\|T\|$
Intercept	1	42.9769	0.1714	250.7352	0.0000
T_a	1	9.4756	0.3056	31.0076	0.0000
R_f	1	-4.9048	0.3035	-16.1626	0.0000
V_f	1	3.9345	0.1761	22.3445	0.0000
$R_f \times V_f$	1	1.4263	0.1671	8.5336	0.0000

Table 6: Hyperparameter Values for Parameters in Computer Experiments

Hyperparameter	Value
m_{θ_c}	0.0
$s^2_{\theta_c}$	100.0
$a_{\xi_c^2}$	2000.0
$b_{\xi_c^2}$	3.0
a_{ϕ_c}	5.0×10^{-1}
b_{ϕ_c}	1.0×10^{-5}
a_{ω_c}	5.0×10^{-1}
b_{ω_c}	1.0×10^{-5}

Table 7: Comparison of Confidence and Credible Intervals

		95% Confidence Int.			95% Credible HPD Int.	
	MLE	Lower	Upper	Post. Mean	Lower	Upper
σ^2	0.81	0.49	1.60	0.53	0.36	0.77
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