Combining discrepancy models with hierarchical Bayesian inference for parameter estimation of very ill posed thermal problems

D Bardot and A F Emery
Department of Mechanical Engineering, University of Washington, Seattle, WA 98195-2600, USA
email: emery@u.washington.edu

Keywords: Discrepancy Model, Hierarchical Bayesian Inference, Parameter Estimation

Abstract - Parameter estimation assumes that the model is an accurate representation of the system being studied and that any deviations are caused by measurement noise. For real experimental data this is often not the case. Clearly, the model will constructed to the highest fidelity by the analyst but when it is deficient, the remedy is not always obvious. One approach is to include a discrepancy function which one hopes will resolve any differences. The paper describes the use of such a function for a very ill posed problem using Bayesian inference effected by Markov Chain Monte Carlo sampling.

1. Introduction
The fundamental assumption when estimating the parameters of a model is that the model accurately reflects the behavior of the system being studied. For typical statistical studies, the model often is simply a device for future predictions, i.e. a regression analysis, and the parameters have little or no physical or phenomenological meaning. However, for engineering and scientific analysis the model is almost always assumed to be an accurate representation of the system’s behavior and the model usually incorporates the conservation equations and other sub-models (i.e., turbulence, combustion) and is said to be a 'true' model. The estimation of the parameters (or properties) of the model is done by comparing the model predictions to the data utilizing the least squares approach (or maximum likelihood which is the equivalent approach if the errors are assumed to be normally distributed) to minimize the differences between the predictions and the measurements.

Sometimes the data are taken to be the deterministic response of the model corrupted by noise and nothing is assumed about the noise other than it is of zero mean and independently and identically distributed. In this case there is no question about the exactness of the model. In contrast, for almost all experimental data, the model is not exact. For example for thermal problems there are always unaccounted for heat losses that cannot be well characterized. For stress problems, whose solution is found using finite element models the surface tractions are treated as nodal point loads and the predicted stresses at the surface rarely agree with the imposed values.

In the least squares approach we define the model as \( M(\Theta) \), where \( \Theta \) represent the parameters. Let the true value of a parameter be denoted by \( \Theta \) and the estimated value by \( \hat{\Theta} \). The measurements are presumed to be corrupted by the noise \( \epsilon \) to give

\[
D = M(\Theta) + \epsilon
\]  

(1)

where \( E[\epsilon] = 0 \) and \( cov[\epsilon] = \Sigma \). For non-linear problems, the parameters are found using a solution that is effected by an iterative procedure based upon linearization. The estimated property, \( \hat{\Theta} \), is
that which minimizes the functional $L(\hat{\Theta})$

$$L(\hat{\Theta}) = r^T(\hat{\theta})\Sigma^{-1}r(\hat{\theta})$$

where the residuals are $r(\hat{\Theta}) \equiv D - M(\hat{\Theta})$ (2)

Upon convergence of the iterations, the estimate $\hat{\Theta}$ satisfies

$$E[\hat{\Theta}] = \Theta \quad \text{with} \quad \text{cov}[\hat{\Theta}] = (A^T\Sigma^{-1}A)^{-1}$$

(3)

For noise that is of zero mean and uncorrelated, $E[r] = 0$ and $\sigma(r) = \sqrt{N-d} \sigma(\epsilon)$ and $\hat{\Theta}$ is an unbiased estimator with minimum variance. $\Sigma$ can be expressed in terms of the correlation matrix $\Omega$ as

$$\Sigma = \sigma_n^2\text{cor}[\epsilon] \equiv \sigma^2_n\Omega$$

(4)

where $\sigma_n$ is the standard deviation of the noise. The adequacy of the model is generally judged by the statistical characteristics of the residuals. These residuals, for a correct model, are presumed to be representative of the noise of the measurements and to be of zero mean and stationary. Residuals that show a bias or non-stationary behavior are taken as evidence that the underlying model is inadequate and needs to be expanded. Unfortunately, most models are thought to be complete when constructed by the analyst and how the model should be augmented is not clear – if it were clear, the model would obviously be modified on sound technical grounds. We will use the statistics of the residuals as a measure of the adequacy of the model.

One way of improving the model is to add a discrepancy function which one hopes will provide a better match with the measured data. Although the form of the function is probably not well defined, otherwise the inadequacy of model would immediately have been perceived, one can add some generic forms, each of course with its own parameters. If fitting the augmented model provides a good match with several experiments with a relatively constant form, then one can argue that it does represent some shortcoming in the original model. An examination of the discrepancy function might then provide insight into what the original model was lacking, e.g., an unexpected heat loss. Unfortunately adding more parameters always reduces the amount of information, as characterized by the Fisher information matrix, increases the imprecision in the estimated parameters, and may lead to an increase in the ill posedness of the inverse problem.

2. The Application to a Calorimeter

One method of determining the specific heat is to observe the change in temperature of a bath when a sample is inserted [1]. Immediately upon insertion of the sample, the liquid bath temperature begins to fall below that of the walls of the calorimeter due to heat transfer between the sample and the fluid. The rate at which this convective heat transfer occurs is a function of the velocity of the fluid circulating around the sample and the mass of the sample. Several seconds after insertion, the rate of bath cooling is significantly reduced for a few seconds and then resumes its fall. This short period of reduced cooling is caused by heat flowing from the calorimeter to the liquid bath. After the material close to the wall has cooled such that a temperature pulse has penetrated into the wall material, the rate at which heat is transferred from the walls of the calorimeter to the liquid is reduced sufficiently such that the liquid resumes its rate of cooling.

The specific heat of the sample is obtained from the conservation of energy, equation (5)

$$c_s = \frac{Q_L + m_wc_w(T_w^i - T_w^f)}{m_s(T_s^f - T_s^i)}$$

(5)

where $c_s, m_s$ and $c_w, m_w$ are the specific heat and mass of the sample and the water respectively and $T_w^i, T_w^f$ and $T_s^i, T_s^f$ are the initial and final temperatures of the water and sample and $Q_L$ is the heat transferred between the bath and the calorimeter. The inaccuracy in estimating the specific heat arises from the uncertain heat loss, $Q_L$. ASTM standards give a lower bound on the imprecision in the specific heat due to $Q_L$ of no less than 10%.
Quantifying this heat transfer is difficult because the convective heat transfer coefficient, $h_{w}$, between the liquid bath and the wall is unknown. In the ASTM procedure, the water temperature is measured over time and extrapolated back to the time of sample insertion, at which time the heat loss is presumed to be zero. Unfortunately it takes a finite time for the sample and the water to equilibrate so the time to extrapolate back to is unclear. One approach is to shorten the time for the sample and bath temperatures to equilibrate by increasing the convective heat transfer between the sample and the bath. However, this increases the heat flow to the wall. Another approach is to maintain the water bath at a constant temperature using some form of proportional-integral-derivative (PID) control. Since the heat lost from the water to the wall when the water is at a fixed temperature can be quantified from calibration experiments, the heat lost during the sample insertion is felt to be better estimated. However using a temperature controlled circulating fluid simply moves the problem from the calorimeter walls to the system used to maintain the temperature of the circulating fluid and little gain in accuracy will be achieved.

2.1 The Calorimeter
In order to achieve better precision in estimating the specific heat we constructed the calorimeter shown in Figure 1. The calorimeter is composed of a plastic beaker, 250 ml, surrounded by a larger polypropylene beaker with foamed in-place insulation. A screen is placed near the bottom to support the sample. To measure $c_s$, the calorimeter is partially filled with hot water. A small immersion heater is used to maintain the water at a fixed temperature, usually near 70°C. The water temperature is measured by a thermistor immersed in the water, the surface of the inside beaker and the exterior surface wall temperatures by RTDs. The calorimeter is placed on a magnetic stirrer and a small cylindrical magnet is located below the screen. The stirrer is operated continuously to ensure that the temperature of the water is well mixed. The sample is composed of small pieces of material which are dropped into the calorimeter and the time history of the temperatures recorded. The sizes of the pieces are chosen to ensure that the sample comes to an isothermal state within a few seconds.

The size of the inner beaker and the mass of the sample were based upon an analysis of the uncertainty [2] of $c_s$ based on preliminary estimates of the imprecision in the temperature measurements and the heat lost to the calorimeter. For most materials, a sample size of 50-100 grams for the 250 ml beaker was determined to be acceptable.

2.2 Analysis of the Heat Lost
One might think that the extent of the problem can be reduced if the calorimeter is better insulated. Unfortunately, the better the insulation the longer the time constant associated with heat transfer
to the calorimeter walls and the less accurately one can estimate $Q_L$. The more rapidly the water is stirred to reduce the time for the sample to equilibrate, the greater the convective heat transfer coefficient and the greater the problem. It is clear that the heat loss will be a function of the convective heat transfer coefficients and the thermal properties of the calorimeter wall. The behavior of the calorimeter was simulated using the finite element method (FEM) with a mesh that includes the thin plastic interior and exterior walls and the foamed in place insulation. The heat loss term $Q_L$ includes the heat loss paths through the walls, and the evaporative losses. The loss is a function of the sample, $\rho_s, c_s, m_s$, the water, $\rho_w, c_w, m_w$, the polypropylene walls, the foamed insulation, $\rho_f, c_f, k_f$, the local ambient air temperature, $T_a$ and the convective heat transfer coefficient, $h_a$ and the geometry

$$Q_L = f(\rho_s, c_s, m_s, \rho_w, c_w, m_w, \rho_f, c_f, k_f, T_a, h_a, T_w, geometry)$$ (6)

Based upon initial measurements and quoted material properties of the insulation, the sensitivity of the heat loss to all of these parameters was determined. Of these sensitivities, only those to the convective heat transfer coefficient between the water and the inner wall, $h_w$, the density times specific heat of the foam, $\rho_f c_f$, the thermal conductivity of the foam, $k_f$, and the convective heat transfer coefficient to the ambient air, $h_a$, were found to be non-negligible and are shown in Figure 2. The figure shows that both $k_f$ and $h_w$ must be known, but that $h_a$ and $c_f$ are unimportant.

3. Determination of $k_f$ and $h_w$

Now that we know that $k_f$ and $h_w$ are the most important parameters needed in evaluating $Q_L$ for the geometry of this specific calorimeter, we need a method to determine them. A series of experiments was performed in which the calorimeter was filled with water, the stirrer turned on and maintained at a constant speed and the temperatures measured as a function of time. In addition, the weight loss due to small amounts of vapor escaping was also measured. Figure 3 shows a typical time history.

![Figure 3 Typical Water Temperature History](image)

![Figure 4 Sensitivity of $T_w$ to $k_f, c_f$ and $h_w$](image)

Figure 4 depicts the sensitivity of the water temperature. The sensitivity to $k_f$ is very much larger than to $c_f$ or to $h_w$. It appears from these sensitivities that we will be able to estimate $k_f$ with reasonable precision, but $c_f$ and $h_w$ are likely to be difficult. Fortunately, $Q_L$ is not sensitive to $c_f$.

3.1 Steady State Experiments

Analyzing the steady state experiments gave the results shown in Table 1. Although the estimate of $k_f$ is essentially the same whether $h_w$ is specified or estimated, the mean of the residuals suggest that $k_f$ and $h_w$ should be estimated simultaneously. Based upon the sensitivities of the heat loss shown in Figure 2, these estimates are sufficiently precise to compute the heat loss with reasonable confidence.
Table 1 Estimated Parameters; Least Squares Minimization; Steady State Tests

|             | Estimating $k_f$ with $h_w = 62$ | Estimating $k_f$ and $h_w$ |
|-------------|---------------------------------|-----------------------------|
|             | $k_f$                            | $k_f$  | $h_w$          |
| average     | 0.0538                           | 0.0539 | 63.6           |
| std         | 0.0031                           | 0.0031 | 3.29           |
| average of residuals | 0.0191                         | 0.0007          |
| std of residuals       | 0.0083                           | 0.0083          |

Table 2 Estimated Parameters; Least Squares Minimization; Transient Tests (Figure 3)

|             | Heating | Cooling |
|-------------|---------|---------|
| $k_f$       | $h_w$   | $k_f$   | $h_w$   |
| average     | 0.0445  | 55.6    | 0.0498  | 38.7    |
| std         | 0.0041  | 9.6     | 0.0120  | 17.5    |
| average of residuals | 0.033   | -0.052  |
| std of residuals       | 0.45    | 0.30    |

3.2 Transient Experiments
The steady state parameters statistics indicated that the $k_f$ and $h_w$ estimates were satisfactory. This created the expectation that transient results would be similarly promising. Unfortunately, this optimism is not borne out when the transient experiments are analyzed as indicated by the values shown in Table 2 where the cooling tests show significant variation in $k_f$ and $h_w$ and with the residuals for both heating and cooling tests being an order of magnitude larger than for the steady state results. In addition, it is clear by comparing the measured and predicted temperature histories that the model is deficient, Figure 5.

![Figure 5 Typical Measured and Predicted Temperatures for Heating; No discrepancy function](image)

4. Discrepancy Function
Unfortunately the nature of the deficiency of the model is unknown. Recognizing that there may be some uncertainty in the energy supplied to the water during heating, to uncertainty in the rate of vapor loss, to uncertainty to heat lost to the cover from the water vapor, and by other heat losses that are likely to be functions of the water temperature, we chose to augment the model by including an unspecified heat loss given by the simple relationship

$$\text{Discrepancy Function} = D_0 + D_1(T_w - T_{ambient})$$

where the coefficients $D_0$ and $D_1$ are to be estimated from the data simultaneously with $k_f$ and $h_w$. Regardless of the algorithm used, because of the ill conditioning we found it unusually difficult
to estimate even the first coefficient, $D_0$ in conjunction with $k_f$ and $h_w$. The problem seemed to be related to the interaction between $D_0$ and $h_w$. Attempts to use different amounts of regularization and under-relaxation factors for each parameter were generally unsuccessful even when using the Levenburg-Marquardt approach. When successful, analyzing a single experiment usually took of the order of several hours and even then we were not confident that a global minimum had been found. Since the interaction between $h_w$ and $k_f$ and $D_0$ appeared to be the problem, we decided to first consider a constant value of $h_w$ and to see if the inclusion of $D_0$ would have an effect. The results for a specific $h_w$ are shown in Figures 6.

![Figure 6](image_url)

**Figure 6** Typical Measured and Predicted Water Temperatures for Heating using the discrepancy function with $D_0 \neq 0$

| Test     | $D_0 = 0$ | $D_0 \neq 0$ |
|----------|-----------|---------------|
|          | mean $k_f$ | std $k_f$ | mean $k_f$ | std $k_f$ | mean $D_0$ | std $D_0$ |
| Heating  | 0.0445    | 0.0041      | 0.0441     | 0.0020     | -0.0379    | 0.2031   |
| mean Residuals | 0.053   | -0.0023     |
| Cooling  | 0.0498    | 0.0120      | 0.0410     | 0.0015     | 0.1101     | 0.1301   |
| mean Residuals | -0.052  | 0.0023      |

The temperatures match very well and the residuals are not only significantly smaller, but show the behavior typical of normally distributed measurement errors. The overall results are given in Table 3 when considering only $D_0$. The heat loss due to the discrepancy function was of the order of 2-5% of the heat supplied in the heating tests. When using only $D_0 \neq 0$, the discrepancy function accounts for a constant heat loss and we do not understand why it improves the results for both heating and cooling. The improvement was always remarkable and the residuals were at least an order of magnitude smaller than when it was not applied.

### 4.1 Hierarchical Bayes

From Figures 5 and 6 and Table 3 we see that including the discrepancy function has substantially reduced the standard deviation of $k_f$ and the magnitude of the residuals. Unfortunately, we were unable to estimate $h_w$ and $D_0$ and $D_1$ simultaneously using least squares even with substantial regularization. Recognizing that the use of Bayesian inference in estimating parameters is equivalent to regularization [3], we examined its use in determining the probability distributions of the parameters.

We assumed that $k_f$ was a constant with some uncertainty, but that $h_w$ and $D_0$ and $D_1$ were random variables drawn from a parent population with a prescribed distribution. Let $\Theta_i$ represent the parameters $k_f$, $h_w$, $D_0$ and $D_1$ appropriate to a given calibration test, $\sigma^2 I$ ($I$ is the identity matrix) represent the temperature measurement error, $\Theta_2$ and $\Sigma_2$ be the mean values and variances...
of the probability distributions from which $\Theta_i$ is to be sampled from. That is $\Theta_2$ are the parent values which represent our best estimate of the true values. Then Bayesian inference can be expressed as

$$p(\Theta_i, \sigma, \Theta_2, \Sigma_2 | D) \propto p(D|\Theta_i, \sigma)p(\Theta_i|\Theta_2, \Sigma_2)p(\Theta_2|\Sigma_2)p(\Sigma_2)p(\sigma)$$

where $D$ represents the data, the measured temperatures, and $p(D|\Theta_i, \sigma)$ is the likelihood of the data based upon a multivariate normal. We chose $p(\Theta_i|\Theta_2, \Sigma_2)$ to be a multivariate normal distribution. The priors necessary to solve equation (8) are $\pi(\Theta_2)$ is multivariate normal, $\pi(\Sigma_2)$ is inverse Wishart and $\pi(\sigma)$ is inverse gamma, the latter two representing minimal prior information.

The marginal posterior distributions of $\Theta_2 = k_f, h_w, D_0, D_1$ are obtained by integrating out $\Theta_i, \Sigma_2, \sigma$ from the left side of equation (8). Unfortunately this hierarchical model results in an expression that cannot be integrated effectively using Gaussian quadrature. Instead, we used a Markov Chain Monte Carlo approach with the Gibbs sampler [4]. Typical results are shown in Figures 7 for the heating and cooling tests. The mean value of the parameters are nearly equal but the distributions are much wider for the heating tests. Of particular interest is the long tail for $k_f$ at the lower values for the heating as compared to the long tail for higher values for the cooling tests and the strange behavior of $h_w$ for the heating tests. There is a clear difference between the two tests and although the pdfs for the cooling test of $h_w$ and $D_0$ are more satisfying, that of $k_f$ is not. Most importantly, the distribution for $k_f$ is far from normal, raising questions about its effect on the heat loss. Figure 8 gives the results for all of the tests combined and it is apparent that the results from the heating tests have the greatest impact on the distributions.

Figure 7a Marginal Probability Density Distribution of $k_f$, $h_w$ and $D_0/2\pi$; heating experiments

Figure 7b Marginal Probability Density Distribution of $k_f$, $h_w$ and $D_0/2\pi$; cooling experiments

Figure 8 Marginal Probability Density Distribution of $k_f$, $h_w$ and $D_0/2\pi$; all experiments
5. Conclusions

We anticipated that the $D_0$ term would exist in the heating experiments because of some variability in the heat input and that there would be a small effect of the $D_1$ term. However, $D_1$ was found to be negligible, circa $10^{-4}$. What we did not expect and could not explain was the presence of a non-zero $D_0$ for the cooling case. Why there should be this constant but small heat loss and how it had such a dramatic effect on reducing the residuals is inexplicable. Although we expected $D_1$ to be important in the cooling case, it was not. This is confusing since any unexplained heat loss due to conduction or to any vapor loss would be expected to be proportional to the temperature of the water.

It is an old adage that given enough free parameters one can fit any data. While this may be true for ad hoc models used in regression analysis, it would not appear to be the case when using a discrepancy model to augment a mathematical model based on physical principles. For these experiments the discrepancy model makes no physical sense and it raises the question of whether it is ever possible to implement such a device. The problem may have been in the assumptions used in forming what is called the ‘mixed model’ [5] and its associated hierarchical model which requires that $h_w$, $D_0$ and $D_1$ be random variables constrained by the nature of the prior probability density. This remains to be tested by other real experiments.

The results obtained in these experiments indicate that the heat transfer model of the calorimeter does not estimate the heat losses with the desired accuracy. Considering the uncertainties in the convective heat transfer between the bath and the calorimeter, inaccuracies in measuring the amount of vapor escaping during the lengthy experiment, the possibility of the thermal conductivity varying with temperature, and variability in the ambient air conditions, it is not yet clear to us how to revise the model. Instead we have constructed a calorimeter that is under PID control. Preliminary experiments suggest that we can maintain a temperature constant to $\pm 0.04^\circ C$. Whether this is sufficient, considering the long time constant, for a dynamic model of the heat transfer to accurately measure $Q_L$ is still problematic. We are also investigating whether using a poorly insulated calorimeter, which thus has a relatively short time constant, under PID control will give improved results.

References

[1] ASTM International, 1999, Standard C351-92b, "Standard Test Method for Mean Specific Heat of Thermal Insulation,"
[2] Coleman, H W and Steele Jr, W G, 1999, Experimentation and Uncertainty Analysis for Engineers, J. Wiley and Sons, N.Y., NY
[3] O’Hagan, A and Forster, J, 2004, Bayesian Inference, Kendall’s Advanced Theory of Statistics, Vol. 2b, Oxford Univ. Press, Oxford, UK
[4] Gelman, A, Carlin, J B, Stern, H S and Rubin, D B, 2004, Bayesian Data Analysis, Chapman & Hall/CRC
[5] McCulloch, C E and Searle, S R, 2001, Generalized, Linear, and Mixed Models J. Wiley and Sons, N.Y., NY