A Bayesian Calibration Approach to the Thermal Problem

Dave Higdon, Los Alamos National Laboratory
Charlie Nakhleh, Los Alamos National Laboratory
Jim Gattiker, Los Alamos National Laboratory
Brian Williams, Los Alamos National Laboratory
LA-UR-07-0624

Many of the problems we work with at Los Alamos National Laboratory are similar to the thermal problem described in the tasking document. In this paper we describe the tools and methods we have developed that utilize experimental data and detailed physics simulations for uncertainty quantification, and apply them to the thermal challenge problem. We then go on to address the regulatory question posed in the problem description. This statistical framework used here is largely based on the approach of Kennedy and O'Hagan (2001), but has been extended to deal with functional output of the simulation model.

Keywords: computer experiments; predictability; certification; uncertainty quantification; Gaussian process; predictive science; functional data analysis; verification and validation

1 Introduction

Understanding and predicting the behavior of complex physical processes is crucial in a variety of applications we are currently involved with at Los Alamos National Laboratory. Examples range from weapon certification to better understanding of the cosmos. Investigation of these physical systems invariably requires a computer code–a simulator–which simulates the physical process of interest along with field data, collected from experiments or observations of the actual physical system.

Even with these fairly well understood physical processes, uncertainties play an important role in using the code to predict behavior of the physical system. Uncertainties arise from a variety of sources: uncertainty in the specification of initial conditions; uncertainty in the value of important physical constants (e.g., thermal conductivity, equations of state, material strength); inadequate mathematical models in the code to describe physical behavior; and inadequacies in the numerical algorithms used for solving the specified mathematical systems (e.g. unresolved grids).

In this paper we describe an approach for combining observations from field experiments with detailed computer simulations of a physical process to carry out statistical inference. Of particular interest here is determining uncertainty in the resulting predictions. This typically involves estimation of experimental errors, calibration of parameters in the computer simulator, and accounting for inadequate physics in the simulator.

Our methodology, described in gory detail in Section 2, is a multivariate generalization of the Bayesian calibration approach of Kennedy and O'Hagan (2001). The method is then applied to the thermal challenge problem in Section 3, using varying amounts of experimental data. The paper then ends with a brief discussion.

2 Methodology

2.1 Overview

We use a Bayesian statistical approach that combines information from the field experiments (i.e., the ensemble and accreditation experiments) and the simulation model to predict temperature as a function of time at the regulatory compliance conditions. This method is an extension of the basic approach of Kennedy and O'Hagan (2001), which is described in more detail in Higdon et al. (2005).

We take $\eta(x, \theta)$ to denote the time trace of temperature obtained from the mathematical model described by equation (2) in the problem description. We'll refer to $\eta(x, \theta)$ as the simulation model, which depends on input conditions x, and unknown calibration parameters θ. Here we take $x = (q, L)$ to denote the heat flux q and the slab thickness L for a given experiment. The unknown parameters $\theta = (k, \rho C_p)$ are the thermal conductivity k and the volumetric heat capacity ρC_p. We assume there is a true but unknown value for these calibration parameters which is to be estimated using the field experiments. The remaining uncertainty is modeled as experimental variation between replicate samples with a correlated error term.

Although evaluating $\eta(x, \theta)$ is trivial in this challenge problem, we limit ourselves to less than 200 evaluations. This restriction is necessary for any complicated simulation code that takes hours to run. We use a space filling latin hypercube sample to select the input settings $(x_1^*, \theta_1^*), \ldots, (x_m^*, \theta_m^*)$ at which to run the simultor. See Santner et al. (2003) for an overview on the design of computer experiments. For this analysis, we run the simulator at 32 different settings of $(k, \rho C_p)$ for 6 different input conditions: 4 corresponding to the ensemble experimental conditions; 1 corresponding to the accreditation experimental condition; and 1 corresponding to the regulatory compliance condition. The $m = 6 \times 32$ simulations are shown in Figure 1, along with experimental traces for all 18 experiments. It is these simulations, $\eta(x_1^*, \theta_1^*), \ldots, \eta(x_m^*, \theta_m^*)$, along with various subsets of the experimental data, that will be used to address the regulatory compliance question.

We model the 11-point temperature traces $y(x_1), \ldots, y(x_n)$ from experiments at input conditions $x_1, \ldots, x_n$ as noisy versions of the simulator at the true parameter setting θ

$$y(x_i) = \eta(x_i, \theta) + \delta(x_i) + e_i, \; i = 1, \ldots, n. \tag{1}$$

The term $\delta(x_i)$ accounts for the discrepancy between the simulator and the actual system; e_i describes experimental variation between time traces from replicate experiments . We assume each e_i is iid $N(0, \Sigma_e)$, and use experimental replicates to help estimate Σ_e (see Figure 10).

In our analysis, we treat the 192 simulations and the 5, 9, or 18 experimental outcomes as data which are used to reduce uncertainties. We specify priors for the unknowns in (1): θ, $\eta(\cdot, \cdot)$, $\delta(\cdot)$, and Σ_e, and follow the Bayesian paradigm to infer about these quantities. A simple uniform prior whose ranges are determined by the material characterization data is used for θ. Gaussian process priors are used for the unknown functions $\eta(\cdot, \cdot)$ and $\delta(\cdot)$. The structure of Σ_e is estimated up front from the data, and only the scaling is estimated in this analysis.

Looking ahead to results, Figure 7 shows the resulting posterior distribution for the 2-d calibration parameter θ. Figure 8 shows the posterior decomposition of the experimental data into calibrated simulator $\eta(x_i, \theta)$, discrepancy $\delta(x_i)$, and experimental prediction $\hat{y}(x_i) = \eta(x_i, \theta) + \delta(x_i)$ using 5 experiments, corresponding to the "low" condition. Given the posterior distribution for the unknowns, one can readily address the regulation compliance question (Figure 9). Section 2.2 below gives the details of the modeling formulation we use here.

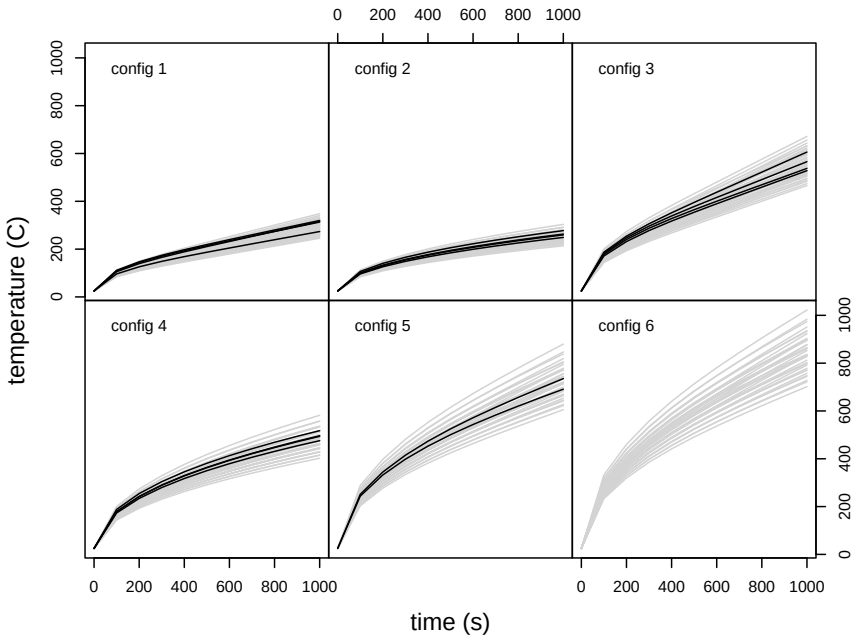

Figure 1: Simulations (gray) and experimental data (black) for each of the ensemble experiments (configs 1–4) and the accreditation experiments (config 5). Also shown are simulations corresponding to the regulatory compliance setting (config 6). The ranges for k and ρC_p are obtained from the ranges of the material characterization experiments. Thirty two simulations are carried out for each configuration over parameter space ($k \times \rho C_p$) according to an orthogonal array-based latin hypercube sample.

2.2 Modeling Details

2.2.1 Modeling Simulator Output

Our analysis requires we develop a probability model to describe the simulator output at untried settings (x, θ). To do this, we use the simulator outputs to construct a GP model that "emulates" the simulator at arbitrary input settings over the (standardized) design space $[0,1]^{p_x+p_\theta}$. To construct this emulator, we model the simulation output using a p_η-dimensional basis representation:

$$\eta(x,\theta) = \sum_{i=1}^{p_\eta} k_i w_i(x,\theta) + \epsilon, \ (x,\theta) \in [0,1]^{p_x+p_\theta}, \tag{2}$$

where $\{k_1, \ldots, k_{p_\eta}\}$ is a collection of orthogonal, n_η-dimensional basis vectors, the $w_i(x,\theta)$'s are GPs over the input space, and ϵ is a n_η-dimensional error term. This type of formulation reduces the problem of building an emulator that maps $[0,1]^{p_x+p_\theta}$ to R^{n_η} to building p_η independent, univariate GP models for each $w_i(x,\theta)$. The details of this model specification are given below.

Output from each of the m simulation runs prescribed by the design results in n_η-dimensional vectors, which we denote by $\eta_1, \ldots, \eta_m$. Since the simulations rarely give incomplete output, the simulation output can often be efficiently represented via principal components (Ramsay and Silverman, 1997). We first standardize the simulations by centering the simulations about the

mean of raw simulation output vectors: $\frac{1}{m}\sum_{j=1}^{m}\eta_j$. We then scale the output by a single value so that its variance is 1. This standardization simplifies some of the prior specifications in our models. We also note that, depending on the application, some alternative standardization may be preferred. Whatever the choice of the standardization, the same standardization is also applied to the experimental data.

We define Ξ to be the $n_\eta \times m$ matrix obtained by column-binding the (standardized) output vectors from the simulations

$$\Xi = [\eta_1; \cdots ; \eta_m].$$

Typically, the size of a given simulation output n_η is much larger than the number of simulations carried out m. We apply the singular value decomposition (SVD) to the simulation output matrix Ξ giving

$$\Xi = UDV^T,$$

where U is a $n_\eta \times m$ orthogonal matrix, D is a diagonal $m \times m$ matrix holding the singular values, and V is a $m \times m$ orthonormal matrix. To construct a p_η-dimensional representation of the simulation output, we define the principal component (PC) basis matrix K_η to be the first p_η columns of $[\frac{1}{\sqrt{m}}UD]$. The resulting principal component loadings or weights is then given by $[\sqrt{m}V]$, whose columns have variance 1.

For the thermal application we take $p_\eta = 3$ so that $K_\eta = [k_1; k_2; k_3]$; the basis functions k_1, k_2 and k_3 are shown in Fig. 2. Note that the k_i's are functions of time.

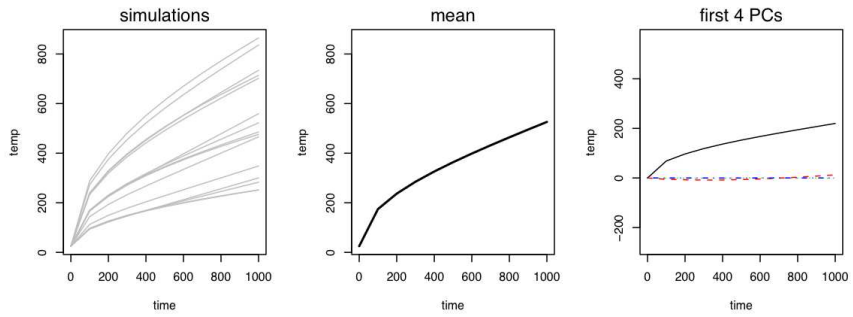

Figure 2: Deriving the basis representation from the simulation output. Here the simulations (left) are represented as the sum of their mean (middle) and a weighted combination of bases (right) obtained via singular value decomposition of the simulations.

We use the basis representation of Eq. (2) to model the n_η-dimensional simulator output over the input space. Each PC weight $w_i(x, \theta)$, $i = 1, \ldots, p_\eta$, is then modeled as a mean 0 GP

$$w_i(x, \theta) \sim \mathrm{GP}(0, \lambda_{wi}^{-1} R((x, \theta), (x', \theta'); \rho_{wi})), \tag{3}$$

where λ_{wi} is the marginal precision of the ith process and the correlation function is given by

$$R((x, \theta), (x', \theta'); \rho_{wi}) = \prod_{k=1}^{p_x} \rho_{wik}^{4(x_k - x'_k)^2} \times \prod_{k=1}^{p_\theta} \rho_{wi(k+p_x)}^{4(\theta_k - \theta'_k)^2}. \tag{4}$$

This is the Gaussian covariance function, which gives very smooth realizations, and has been used previously by Sacks et al. (1989) and Kennedy and O'Hagan (2001) to model computer simulation

output. An advantage of this product form is that only a single additional parameter is required per additional input dimension, while the fitted GP response still allows for rather general interactions between inputs. We use this Gaussian form for the covariance function because the simulators we work with tend to respond very smoothly to changes in the inputs. Depending on the nature of the sensitivity of simulation output to input changes, one may wish to alter this covariance specification to allow for rougher realizations. The parameter ρ_{wik} controls the spatial range for the kth input dimension of the process $w_i(\theta)$. Under this parameterization, ρ_{wik} gives the correlation between $w_i(\theta)$ and $w_i(\theta')$ when the input conditions θ and θ' are identical, except for a difference of 0.5 in the kth component. Note that this interpretation makes use of the standardization of the input space to $[0,1]^{p_\theta}$.

We define the m-vector w_i to be the restriction of the process $w_i(\cdot,\cdot)$ to the input design settings

$$w_i = (w_i(x_1^*, \theta_1^*), \ldots, w_i(x_m^*, \theta_m^*))^T, \quad i = 1, \ldots, p_\eta.$$

In addition we define $R((x^*, \theta^*); \rho_{wi})$ to be the $m \times m$ correlation matrix resulting from applying (4) to each pair of input settings in the design. The $(p_x + p_\theta)$-vector ρ_{wi} gives the correlation distances for each of the input dimensions.

At the m simulation input settings, the mp_η-vector $w = (w_1^T, \ldots, w_{p_\eta}^T)^T$ then has prior disribution

$$w = \begin{pmatrix} w_1 \\ \vdots \\ w_{p_\eta} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda_{w1}^{-1} R((x^*, \theta^*); \rho_{w1}) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_{wp_\eta}^{-1} R((x^*, \theta^*); \rho_{wp_\eta}) \end{pmatrix} \right), \tag{5}$$

which is controlled by p_η precision parameters held in λ_w and $p_\eta \cdot (p_x + p_\theta)$ spatial correlation parameters held in ρ_w. The centering of the simulation output makes the zero mean prior appropriate. The prior above can be written more compactly as

$$w \sim N(0, \Sigma_w),$$

where Σ_w, controlled by parameter vectors λ_w and ρ_w, is given in (5).

We specify independent $\Gamma(a_w, b_w)$ priors for each λ_{wi} and independent beta(a_{ρ_w}, b_{ρ_w}) priors for the ρ_{wik}'s.

$$\begin{aligned} \pi(\lambda_{wi}) &\propto \lambda_{wi}^{a_w - 1} e^{-b_w \lambda_{wi}}, \quad i = 1, \ldots, p_\eta, \\ \pi(\rho_{wik}) &\propto \rho_{wik}^{a_{\rho_w} - 1} (1 - \rho_{wik})^{b_{\rho_w} - 1}, \quad i = 1, \ldots, p_\eta, \; k = 1, \ldots, p_x + p_\theta. \end{aligned}$$

We expect the marginal variance for each $w_i(\cdot,\cdot)$ process to be close to one due to the standardization of the simulator output. For this reason we specify that $a_w = b_w = 5$. In addition, this informative prior helps stabilize the resulting posterior distribution for the correlation parameters which can trade off with the marginal precision parameter (Kern, 2000).

Because we expect only a subset of the inputs to influence the simulator response, our prior for the correlation parameters reflects this expectation of "effect sparsity." Under the parameterization in (4), input k is inactive for PC i if $\rho_{wik} = 1$. Choosing $a_{\rho_w} = 1$ and $0 < b_{\rho_w} < 1$ will give a density with substantial prior mass near 1. We take $b_{\rho_w} = 0.1$, which makes $\Pr(\rho_{wik} < 0.98) \approx \frac{1}{3}$ a priori. In general, the selection of these hyperparameters should depend on how many of the $p_x + p_\theta$ inputs are expected to be active.

If we take the error vector in the basis representation of (2) to be i.i.d. normal, we can then develop the sampling model, or likelihood, for the simulator output. We define the $n_\eta m$-vector η to be the concatenation of all m simulation output vectors

$$\eta = \text{vec}(\Xi) = \text{vec}([\eta(\theta_1^*); \cdots ; \eta(\theta_m^*)]).$$

Given precision λ_η of the errors the likelihood is then

$$L(\eta|w, \lambda_\eta) \propto \lambda_\eta^{\frac{mn_\eta}{2}} \exp\left\{-\tfrac{1}{2}\lambda_\eta(\eta - Kw)^T(\eta - Kw)\right\},$$

where the $n_\eta \times mp_\eta$ matrix K is given by

$$K = [I_m \otimes k_1; \cdots ; I_m \otimes k_{p_\eta}],$$

and the k_i's are the p_η basis vectors previously computed via SVD. A $\Gamma(a_\eta, b_\eta)$ is specified for the error precision λ_η.

Since the likelihood factors as shown below

$$\begin{aligned} L(\eta|w, \lambda_\eta) \quad \propto \quad & \lambda_\eta^{\frac{mp_\eta}{2}} \exp\left\{-\tfrac{1}{2}\lambda_\eta(w - \hat{w})^T(K^TK)(w - \hat{w})\right\} \times \\ & \lambda_\eta^{\frac{m(n_\eta - p_\eta)}{2}} \exp\left\{-\tfrac{1}{2}\lambda_\eta \eta^T(I - K(K^TK)^{-1}K^T)\eta\right\}, \end{aligned}$$

the formulation can be equivalently represented with a dimension reduced likelihood and a modified $\Gamma(a'_\eta, b'_\eta)$ prior for λ_η:

$$L(\hat{w}|w, \lambda_\eta) \quad \propto \quad \lambda_\eta^{\frac{mp_\eta}{2}} \exp\left\{-\tfrac{1}{2}\lambda_\eta(\hat{w} - w)^T(K^TK)(\hat{w} - w)\right\}, \tag{6}$$

where

$$\begin{aligned} a'_\eta &= a_\eta + \frac{m(n_\eta - p_\eta)}{2}, \\ b'_\eta &= b_\eta + \tfrac{1}{2}\eta^T(I - K(K^TK)^{-1}K^T)\eta, \text{ and} \\ \hat{w} &= (K^TK)^{-1}K^T\eta. \end{aligned} \tag{7}$$

Thus the normal-gamma model

$$\eta|w, \lambda_\eta \sim N(Kw, \lambda_\eta^{-1}I_{n_\eta}),\ \lambda_\eta \sim \Gamma(a_\eta, b_\eta)$$

is equivalent to the reduced form

$$\hat{w}|w, \lambda_\eta \sim N(w, (\lambda_\eta K^TK)^{-1}),\ \lambda_\eta \sim \Gamma(a'_\eta, b'_\eta)$$

since

$$L(\eta|w, \lambda_\eta) \times \pi(\lambda_\eta; a_\eta, b_\eta) \propto L(\hat{w}|w, \lambda_\eta) \times \pi(\lambda_\eta; a'_\eta, b'_\eta). \tag{8}$$

The likelihood depends on the simulations only through the computed PC weights $\hat{w}$. After integrating out w, the posterior distribution becomes

$$\begin{aligned} &\pi(\lambda_\eta, \lambda_w, \rho_w|\hat{w}) \propto \\ &\quad \left|(\lambda_\eta K^TK)^{-1} + \Sigma_w\right|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}\hat{w}^T([\lambda_\eta K^TK]^{-1} + \Sigma_w)^{-1}\hat{w}\} \times \\ &\quad \lambda_\eta^{a'_\eta - 1}e^{-b'_\eta\lambda_\eta} \times \prod_{i=1}^{p_\eta} \lambda_{wi}^{a_w - 1}e^{-b_w\lambda_{wi}} \times \prod_{i=1}^{p_\eta}\prod_{j=1}^{p_\theta}(1 - \rho_{wij})^{b_\rho - 1} \end{aligned} \tag{9}$$

This posterior distribution is a milepost on the way to the complete formulation, which also incorporates experimental data. However, it is worth considering this intermediate posterior distribution for the simulator response. It can be explored via MCMC using standard Metropolis updates and we can view a number of posterior quantities to illuminate features of the simulator.

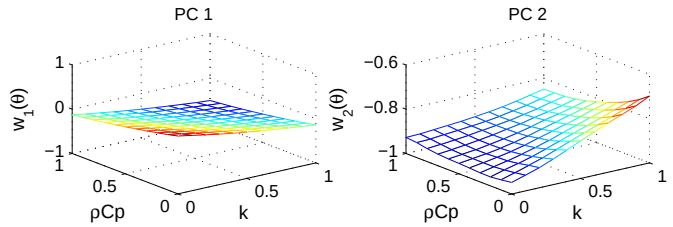

Figure 3: Estimated posterior mean for the response surfaces $w_1(x,\theta)$ and $w_2(x,\theta)$, where the value of x is held fixed corresponding to configuration 4: $x = (q = 2000, L = 2.54)$.

Given the posterior realizations from (9), one can generate realizations from the process $\eta(x,\theta)$ at any input setting $(x^\star, \theta^\star)$. Since

$$\eta(x^\star,\theta^\star) = \sum_{i=1}^{p_\eta} k_i w_i(x^\star,\theta^\star),$$

realizations from the $w_i(x^\star,\theta^\star)$ processes need to be drawn given the MCMC output. For a given draw $(\lambda_\eta, \lambda_w, \rho_w)$ a draw of $w^\star = (w_1(x^\star,\theta^\star),\ldots,w_{p_\eta}(x^\star,\theta^\star))^T$ can be produced by making use of the fact

$$\begin{pmatrix}\hat{w}\\ w^\star\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \left[\begin{pmatrix}(\lambda_\eta K^T K)^{-1} & 0\\ 0 & 0\end{pmatrix} + \Sigma_{w,w^\star}(\lambda_w,\rho_w)\right]\right),$$

where $\Sigma_{w,w^\star}$ is obtained by applying the prior covariance rule to the augmented input settings that include the original design and the new input setting $(\theta^\star)$. Recall $\hat{w}$ is defined in (7). Application of the conditional normal rules then gives

$$w^\star|\hat{w} \sim N(V_{21}V_{11}^{-1}\hat{w}, V_{22} - V_{21}V_{11}^{-1}V_{12}), \tag{10}$$

where

$$V = \begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix} = \left[\begin{pmatrix}(\lambda_\eta K^T K)^{-1} & 0\\ 0 & 0\end{pmatrix} + \Sigma_{w,w^\star}(\lambda_w,\rho_w)\right]$$

is a function of the parameters produced by the MCMC output. Hence, for each posterior realization of $(\lambda_\eta, \lambda_w, \rho_w)$, a realization of $w^\star$ can be produced. The above recipe easily generalizes to give predictions over many input settings at once. Figure 3 shows the posterior mean for the $w_1(x,\theta)$ and $w_2(x,\theta)$ processes, conditional on x being set to $(q = 2000, L = 2.54)$.

Figure 4 shows posterior means for the simulator response η where each of the four inputs were varied over their prior range of $[0,1]$ while the other three inputs were held at their nominal setting of 0.5. The posterior mean response conveys an idea of how the different parameters affect the highly multivariate simulation output. Other marginal functionals of the simulation response can also be calculated such as sensitivity indicies or estimates of the Sobol decomposition (Sacks *et al.*, 1989; Oakley and O'Hagan, 2004).

As a final check before moving on to describe the complete model formulation, we test how well this GP model can predict holdout runs from the simulation model. Figure 5 shows three holdout predictions. These are predictions for the simulation model at input settings that were not part of the original sample used to construct the GP model. Over a collection of 200 holdouts, the GP model was typically within 2^oC of the actual simulations.

2.2.2 Modeling Discrepancy

Here we define the discrepancy model which, like the model for $\eta(x, \theta)$, is constructed using a basis representation, placing GP models on the basis weights. It differs in that the basis weights depend only on input condition x and that the basis specification for $\delta(x)$ is typically nonorthogonal and tailored to the application at hand.

For the thermal application, $\delta(x)$ smoothly adjusts the temperature as a function of time at the 11 time points collected from the experiment. This discrepancy between actual and simulated temperature is constructed as a linear combination of $p_\delta = 5$ basis functions that are defined over time t. Thus

$$\delta(x) = \sum_{k=1}^{p_\delta} d_k(t) v_k(x) = \sum_{k=1}^{p_\delta} d_k v_k(x), \tag{11}$$

where the basis functions d_k, $k = 1, \ldots, p_\delta$, are shown in Fig. 6, and independent GP priors over x are specified for each weight $v_k(x)$.

The basis functions are specified according to what is known about the actual physical process and potential deficiencies in the simulator. Here the width of the kernels indicates that we expect the discrepancy to be smooth, and to persist over time. Higdon (2002) shows that such a representation is nearly equivalent to a stationary GP prior for $\delta(x)$ with a Gaussian covariance function.

We specify independent mean 0 GP priors for each basis weight $v_k(x)$. Thus the p_δ-variate process $v(x) = (v_1(x), \ldots, v_{p_\delta}(x))^T$ is a mean 0 GP with covariance rule given by

$$\mathrm{Cov}(v(x), v(x')) = \lambda_v^{-1} I_{p_\delta} \otimes R(x, x'; \rho_v),$$

where λ_v is the common marginal precision of each $v_k(x)$, ρ_v is a p_x-vector controlling the correlation strength along each component of x, and $R(x, x'; \rho_v)$ is a stationary Gaussian product correlation

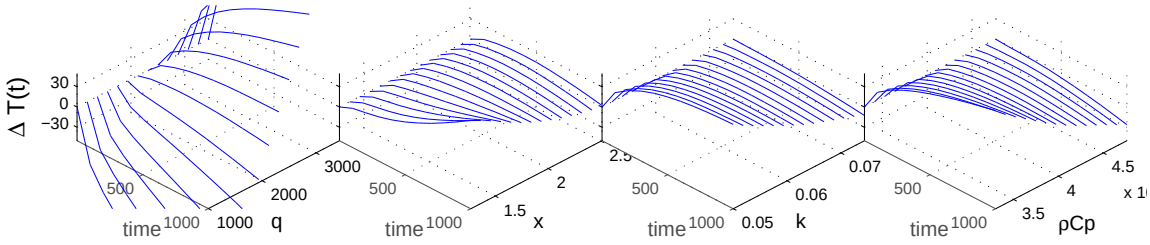

Figure 4: Sensitivity of the simulation model to changes in the input variables. For each input variable $(q, x, k, \rho C_p)$, the figures above show how varying the input setting from its minimum value to its maximum (while leaving the other inputs at their midpoint setting) alters the temperature profile.

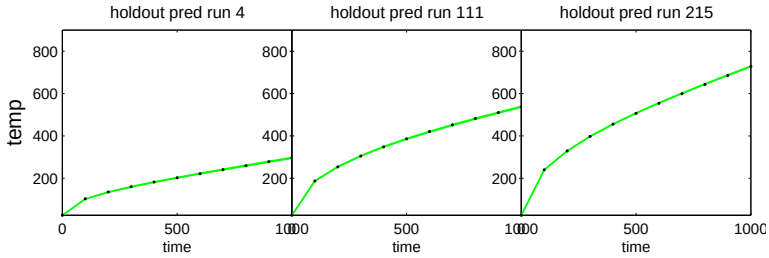

Figure 5: Predicting the simulator response at holdout settings not used to estimate the Gaussian process response surface model. In looking at over 200 holdout predictions, the average absolute difference between the actual simulations and the response surface prediction is about 2^oC.

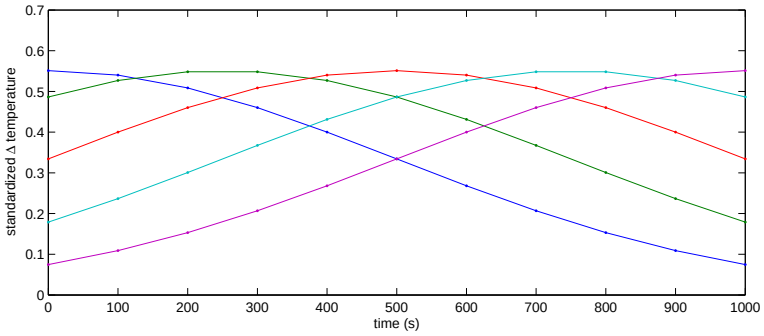

Figure 6: Basis kernels d_k, $k = 1, \ldots, p_\delta$. Each kernel is a normal density with an sd of 500 s. This construction gives realizations of $\delta(x)$ that vary slowly with time.

model given by

$$R(x, x'; \rho_v) = \prod_{k=1}^{p_x} \rho_{vk}^{4(x_k - x'_k)^2}. \tag{12}$$

Note that the Gaussian form of the correlation will enforce a high degree of smoothness for each process $v_k(x)$ as a function of x. We feel this is plausible in this application since we expect any discrepancies to change smoothly with input condition x. Other applications may require an alternate specification.

As with the GP model for the simulator $\eta(x, \theta)$, we complete the discrepancy model formulation by specifying a gamma prior for the precision λ_v and independent beta priors for the components of ρ_v,

$$\begin{aligned} \pi(\lambda_v) &\propto \lambda_v^{a_v - 1} e^{-b_v \lambda_v} \\ \pi(\rho_{vk}) &\propto \rho_{vk}^{a_{\rho_v} - 1} (1 - \rho_{vk})^{b_{\rho_v} - 1}, \quad k = 1, \ldots, p_\delta, \end{aligned}$$

where $a_v = 1$, $b_v = 0.0001$, $a_{\rho_v} = 1$, and $b_{\rho_v} = 0.1$. This results in a rather uninformative prior for the precision λ_v. If the data are uninformative about this parameter, it will tend to stay at large values that are consistent with a very small discrepancy. Like the prior for ρ_w, we take $a_{\rho_v} = 1$ and $b_{\rho_v} = 0.1$ to encourage effect sparsity.

2.2.3 Full model specification

Given the model specifications for the simulator $\eta(x,\theta)$ and the discrepancy $\delta(x)$, we can now consider the sampling model for the experimentally observed data. We assume the data $y(x_1),\ldots,y(x_n)$ are collected for n experiments at input conditions $x_1,\ldots,x_n$. For the thermal problem, there are $n=5,9$, or 18 experiments whose data are shown in Figure 1. Each $y(x_i)$ is a collection of $n_{y_i}=11$ measurements equally spaced over time. The data for experiment i is modeled as the sum of the simulator output at the true parameter setting θ and the discrepancy

$$y(x_i) \;=\; \eta(x_i,\theta)+\delta(x_i)+e_i,$$

where the observation error vector e_i is modeled as $N\left(0,(\lambda_y W_i)^{-1}\right)$. Using the basis representations for the simulator and the discrepancies, this becomes

$$y(x_i) \;=\; K_i w(x_i,\theta)+D_i v(x_i)+e_i.$$

Because the time support of each $y(x_i)$ varies with experiment and isn't necessarily contained in the support of the simulation output, the basis vectors in K_i may have to be interpolated over time and angle from K_η. The discrepancy basis matrix D_i is determined by the functional form given in (11)–the jk element of D_i is given by

$$D_{ijk}=d_k(t_j).$$

where t_j is the j^{th} experimental time point. The sampling model for the observations in experiment i is n_{y_i}-variate normal

$$y(x_i)|w(x_i,\theta),v(x_i),\lambda_y \;\sim\; N\left([D_i;K_i]\begin{pmatrix}v(x_i)\\ w(x_i,\theta)\end{pmatrix},(\lambda_y W_i)^{-1}\right).$$

Note that $(\lambda_y W_i)^{-1}$ gives us Σ_e from (1).

Taking all of the experiments together, the sampling model is n_y variate normal, where $n_y=n_{y_1}+\cdots+n_{y_n}$, is the total number of experimental data points. We define y to be the n_y-vector from concatenation of the $y(x_i)$'s, $v=\text{vec}([v(x_1);\cdots;v(x_n)]^T)$ and $u(\theta)=\text{vec}([w(x_1,\theta);\cdots;w(x_n,\theta)]^T)$. The sampling model for the entire experimental dataset, along with the prior for the observation precision λ_y, can be written as

$$y|v,u(\theta),\lambda_y\sim \text{N}\left(B\begin{pmatrix}v\\ u(\theta)\end{pmatrix},(\lambda_y W_y)^{-1}\right),\ \lambda_y\sim\Gamma(a_y,b_y), \tag{13}$$

where $W_y=\text{diag}(W_1,\ldots,W_n)$,

$$B=[\text{diag}(D_1,\ldots,D_n);\ \text{diag}(K_1,\ldots,K_n)]\begin{pmatrix}P_D^T & 0\\ 0 & P_K^T\end{pmatrix},$$

and P_D and P_K are permutation matricies whose rows are given by

$$P_D(j+n(i-1);\cdot) \;=\; e^T_{(j-1)p_\delta+i},\ i=1,\ldots,p_\delta;\ j=1,\ldots,n$$

$$P_K(j+n(i-1);\cdot) \;=\; e^T_{(j-1)p_\eta+i},\ i=1,\ldots,p_\eta;\ j=1,\ldots,n,$$

where e_j is the vector of 0's, with a one in the i^{th} entry. Note that permutations are required for specifying B since the basis weight components v and $u(\theta)$ are separated in (13). The observation precision W_y is often fairly well-known in practice. Hence we use an informative prior for λ_y that encourages its value to be near one. For the thermal problem we set $a_y = b_y = 3$.

Equivalently (13) can be represented using the normal-gamma form

$$\begin{pmatrix}\hat{v}\\ \hat{u}\end{pmatrix}\left|\begin{pmatrix}v\\ u(\theta)\end{pmatrix}, \lambda_y \sim \mathrm{N}\left(\begin{pmatrix}v\\ u(\theta)\end{pmatrix}, (\lambda_y B^T W_y B)^{-1}\right), \quad \lambda_y \sim \Gamma(a'_y, b'_y),\right.$$

with

$$\begin{aligned}
n_y &= n_{y_1} + \cdots + n_{y_n}, \text{ denoting the total number of experimental data points,}\\
\begin{pmatrix}\hat{v}\\ \hat{u}\end{pmatrix} &= (B^T W_y B)^{-1} B^T W_y y,\\
a'_y &= a_y + \tfrac{1}{2}[n_y - n(p_\delta + p_\eta)], \text{ and}\\
b'_y &= b_y + \frac{1}{2}\left[\left(y - B\begin{pmatrix}\hat{v}\\ \hat{u}\end{pmatrix}\right)^T W_y \left(y - B\begin{pmatrix}\hat{v}\\ \hat{u}\end{pmatrix}\right)\right].
\end{aligned}$$

This equivalency follows from (8) given in Sec. 2.2.1.

The (marginal) distribution for the combined, reduced data obtained from the experiments and simulations given the covariance parameters has the form

$$\begin{pmatrix}\hat{v}\\ \hat{u}\\ \hat{w}\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}\Lambda_y^{-1} & 0\\ 0 & 0 & \Lambda_\eta^{-1}\end{pmatrix} + \begin{pmatrix}\Sigma_v & 0 & 0\\ 0 & & \Sigma_{uw}\\ 0 & & \end{pmatrix}\right), \tag{14}$$

where

$$\begin{aligned}
\Lambda_y &= \lambda_y B^T W_y B,\\
\Lambda_\eta &= \lambda_\eta K^T K,\\
\Sigma_v &= \lambda_v^{-1} I_{p_\delta} \otimes R(x, x; \rho_v),
\end{aligned}$$

and the covariance matrix Σ_{uw}, which links the simulator response $u(\theta)$ at the experimental settings, (x_i, θ), $i = 1, \ldots, n$, to the simulator response w at the design inputs, (x_j^*, θ_j^*), $j = 1, \ldots, m$, is given by

$$\begin{pmatrix}
\lambda_{w1}^{-1} R((x,\theta),(x,\theta);\rho_{w1}) & 0 & 0 & \lambda_{w1}^{-1} R((x,\theta),(x^*,\theta^*);\rho_{w1}) & 0 & 0\\
\vdots & \ddots & 0 & \vdots & \ddots & 0\\
0 & 0 & \lambda_{wp_\eta}^{-1} R((x,\theta),(x,\theta);\rho_{wp_\eta}) & 0 & 0 & \lambda_{wp_\eta}^{-1} R((x,\theta),(x^*,\theta^*);\rho_{wp_\eta})\\
\lambda_{w1}^{-1} R((x^*,\theta^*),(x,\theta);\rho_{w1}) & 0 & 0 & \lambda_{w1}^{-1} R((x^*,\theta^*),(x^*,\theta^*);\rho_{w1}) & 0 & 0\\
\vdots & \ddots & 0 & \vdots & \ddots & 0\\
0 & 0 & \lambda_{wp_\eta}^{-1} R((x^*,\theta^*),(x,\theta);\rho_{wp_\eta}) & 0 & 0 & \lambda_{wp_\eta}^{-1} R((x^*,\theta^*),(x^*,\theta^*);\rho_{wp_\eta})
\end{pmatrix}.$$

Above, $R(x, x; \rho_v)$ denotes the $n \times n$ correlation matrix for the discrepancy process obtained by applying (12) to the input conditions $x_1, \ldots, x_n$ corresponding to the n experiments; $R((x^*, \theta^*), (x, \theta); \rho_{wi})$ denotes the $m \times n$ correlation submatrix for the GP modeling the simulator output obtained by applying (4) to the m simulator input settings $(x_1^*, \theta_1^*), \ldots, (x_m^*, \theta_m^*)$ crossed with the n experimental settings $(x_1, \theta), \ldots, (x_n, \theta)$, with θ denoting the true, but unknown, calibration setting to

be estimated. The remaining components of Σ_{uw} are constructed analogously. Note that only the off-diagonal blocks of Σ_{uw} depend on the unknown calibration parameters contained in θ. The equivalency of (8) reduces the $(n_y + mn_\eta)$-variate normal distribution of $(y^T, \eta^T)^T$ to the $(n(p_\eta + p_\delta) + mp_\eta)$-variate normal distribution of $(\hat{v}^T, \hat{u}^T, \hat{w}^T)^T$ given in (14)–particularly efficient when n_η and n_y are large.

2.2.4 Posterior distribution

If we take $\hat{z}$ to denote the reduced data $(\hat{v}^T, \hat{u}^T, \hat{w}^T)^T$, and $\Sigma_{\hat{z}}$ to be the covariance matrix given in (14), the posterior distribution has the form

$$\begin{aligned}
&\pi(\lambda_\eta, \lambda_w, \rho_w, \lambda_y, \lambda_v, \rho_v, \theta | y, \eta) \propto \qquad (15)\\
&|\Sigma_{\hat{z}}|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\hat{z}^T\Sigma_{\hat{z}}^{-1}\hat{z}\right\} \times \lambda_\eta^{a'_\eta - 1} e^{-b'_\eta \lambda_\eta} \times \prod_{i=1}^{p_\eta} \lambda_{wi}^{a_w - 1} e^{-b_w \lambda_{wi}} \times \\
&\prod_{i=1}^{p_\eta}\prod_{k=1}^{p_x+p_\theta} \rho_{wik}^{a_{\rho_w}-1}(1-\rho_{wik})^{b_{\rho_w}-1} \times \lambda_y^{a'_y - 1} e^{-b'_y \lambda_y} \times \lambda_v^{a_v - 1} e^{-b_v \lambda_v} \times \\
&\prod_{k=1}^{p_x} \rho_{vk}^{a_{\rho_v}-1}(1-\rho_{vk})^{b_{\rho_v}-1} \times I[\theta \in C],
\end{aligned}$$

where C denotes the constraint region for θ, which is typically a p_θ-dimensional rectangle. In other applications C can also incorporate constraints between the components of θ.

Realizations from the posterior distribution are produced using standard, single site MCMC. Metropolis updates are used for the components of ρ_w, ρ_v and θ with a uniform proposal distribution centered at the current value of the parameter. The precision parameters λ_η, λ_w, λ_y and λ_v are sampled using Hastings (1970) updates. Here the proposals are uniform draws, centered at the current parameter values, with a width that is proportional to the current parameter value.

2.2.5 Posterior predictions

As with the pure emulator analysis described in Section 2.2.1, predictions of system behavior can be produced at unobserved input settings $x^\star$. Since

$$\begin{aligned}
\hat{y}(x^\star) &= \eta(x^\star, \theta) + \delta(x^\star) \\
&= Kw(x^\star, \theta) + Dv(x^\star),
\end{aligned}$$

we need only produce draws $w(x^\star, \theta)$ and $v(x^\star)$ given a posterior draw of the parameter vector $(\lambda_\eta, \lambda_w, \rho_w, \lambda_y, \lambda_v, \rho_v, \theta)$. Draws of $w(x^\star, \theta)$ and $v(x^\star)$ can then be used to give posterior realizations for the calibrated simulator $\eta(x^\star, \theta)$, the discrepancy term $\delta(x^\star)$, and predictions $\hat{y}(x^\star)$.

These predictions can be produced from standard GP theory. Conditional on the parameter vector $(\lambda_\eta, \lambda_w, \rho_w, \lambda_y, \lambda_v, \rho_v, \theta)$, the reduced data $\hat{z}$, along with the predictions $w(x^\star, \theta)$ and $v(x^\star)$, have the joint distribution

$$\begin{pmatrix} \hat{z} \\ v(x^\star) \\ w(x^\star, \theta) \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\hat{z}} & \Sigma_{\hat{z}v^\star} & \Sigma_{\hat{z}w^\star} \\ \Sigma_{v^\star\hat{z}} & \lambda_v I_{p_\delta} & 0 \\ \Sigma_{w^\star\hat{z}} & 0 & \mathrm{diag}(\lambda_w) \end{pmatrix} \right),$$

where $\Sigma_{\hat{z}v^\star}$ has nonzero elements due to the correlation between $\hat{v}$ and $v(x^\star)$, and $\Sigma_{\hat{z}w^\star}$ has nonzero elements due to the correlation between $(\hat{u}, \hat{w})$ and $w(x^\star, \theta)$. The exact construction of the matrices $\Sigma_{\hat{z}v^\star}$ and $\Sigma_{\hat{z}w^\star}$ is analogous to the construction of Σ_v and Σ_{uw} in Sec. 2.2.3. Generating simultaneous draws of $v(x^\star)$ and $w(x^\star, \theta)$ is then straightforward using conditional normal rules as is detailed in Sec. 2.2.1.

3 Results for the Thermal Problem

The analysis reveals a number of sources of uncertainty. The posterior distribution for the calibration parameters is shown in Figure 7. Note that there is a fair bit of trade off between these two parameters. This isn't surprising given the sensitivities shown in Figure 4. When the amount of experimental data is medium or high, the correlation in this posterior is much less pronounced. Note also, that the uncertainty regarding θ is more in the low case, but similar for the medium and high cases.

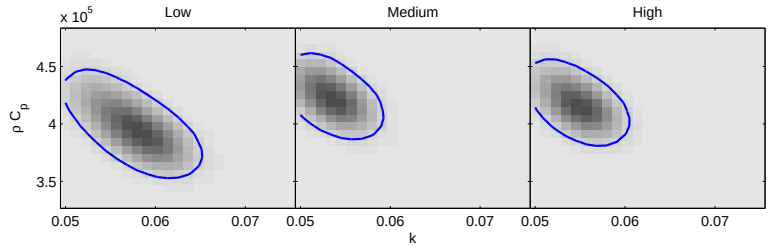

Figure 7: The resulting posterior distribution for the calibration parameters $\theta = (k, \rho C_p)$ under the three analyses using different amounts of experimental data. The analysis with the low amount of experimental data shows greater uncertainty in the posterior distribution for θ.

If the posterior distribution for θ is propagated through the simulation model $\eta(x, \theta)$, the resulting calibrated simulator predictions are shown in the left hand column of Figure 8. Note that the uncertainty due to θ is relatively small.

The posterior uncertainty for the discrepancy $\delta(x)$ is quite small and is shown in the middle column of Figure 8. The scaling required a 10-fold magnification just to be visible. Clearly, the discrepancy term adds little to the overall uncertainty. The fit for a new experiment is shown in the right hand column of the figure. Since the discrepancy is nearly zero, the fitted values look very similar to the calibrated simulator.

If one were to predict the outcome of a new experiment at input condition $x^\star$, the predicted value would be

$$y(x^\star) = \eta(x^\star, \theta) + \delta(x^\star) + e^\star.$$

Figure 9 shows pointwise 90% prediction intervals at six different input conditions using the low amount of experimental data. These predictions are obtained through Monte Carlo draws. For the regulatory condition $(q = 3500, L = 1.90)$, the estimated probability that the temperature exceeds 900^oC at 1000 seconds is 0.07, easily exceeding the requirement of $p_f = 0.01$. Since the uncertainty due to the calibration parameter θ and the discrepancy is quite small, it's clear that this uncertainty is dominated the replicate variability $e^\star$ which has variance Σ_e.

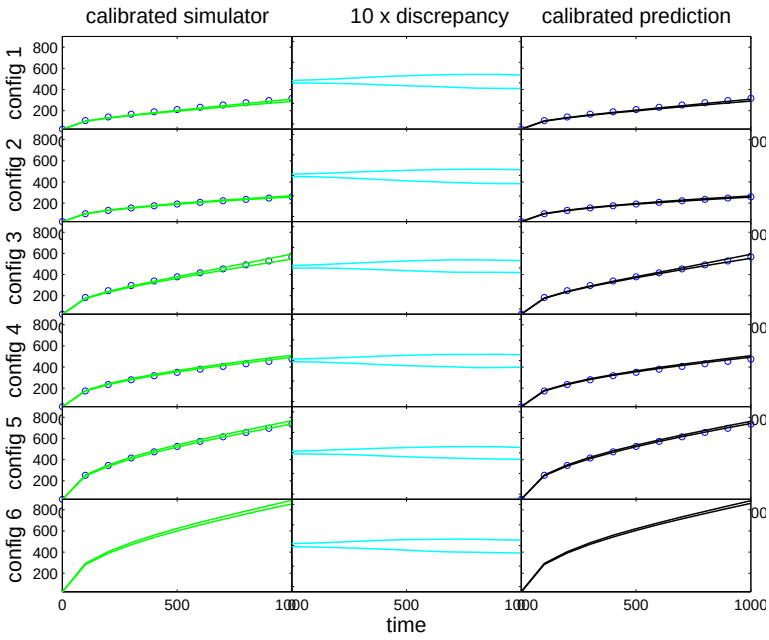

Figure 8: The resulting posterior distribution for the calibrated simulation model (left column), model discrepancy as a function of time (middle column), and the posterior prediction for the field experiment. Experimental configurations 1–4 correspond to the 4 different ensemble validation experimental conditions, configuration 5 corresponds to the accreditation experimental condition, and configuration 6 corresponds to the regulatory compliance experiment. Note the estimated model discrepancy is inflated by a factor of 10. Hence model discrepancy plays a small role in this application.

The replicate variation obtained from taking the first two experiments at each configuration is shown in Figure 10. There is some hint that the variability grows with flux q, but nothing definite. Our model assumes the replicate variation is iid for all experimental conditions. In any case, this variation can be as large as $\pm 30^o$C at 1000 seconds. This is important since it points to one potential mitigation strategy for realizing the certification requirement – reducing the replicate variability in samples.

Finally, the summary of the certification analyses is given in Table 1. The additional information in the medium and high experimental data cases leads to some reduction in the probability of exceeding the 900^o limit, but the additional data can not reduce the replicate variation which is the key source of variation in this problem.

Table 1: Estimated probability of exceeding the failure temperature of 1000^oC at 1000 seconds.

Exp runs	Ensemble Validation N_v	Accreditation N_a	$\mathrm{P}(T(1000) > 900^oC)$
low	4	1	0.07
medium	8	1	0.03
high	16	2	0.03

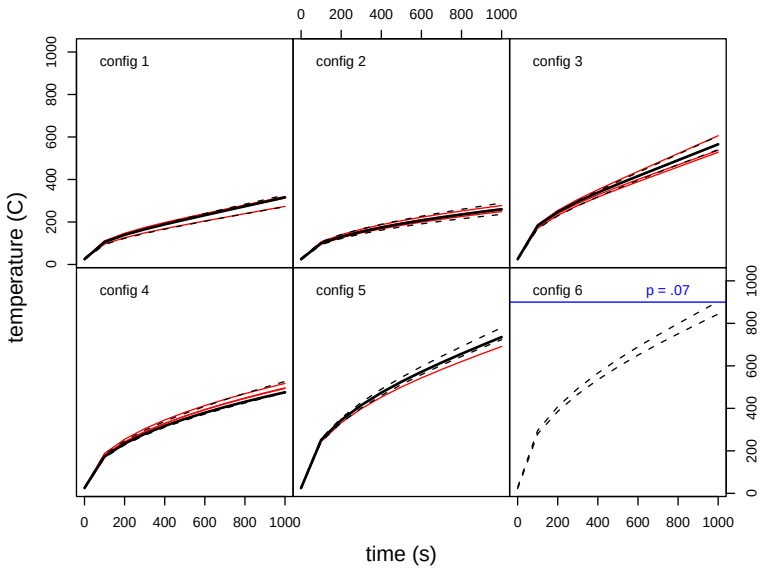

Figure 9: Posterior predictive distributions for temperature as a function of time for each ensemble validation experiment (configs 1–4), the accreditation experiment (config 5) and the regulatory compliance setting (config 6). The black dashed lines give pointwise 90% probability intervals for the temperature of a new experiment as a function of time. The probability that a randomly chosen sample at the regulatory setting ($q = 3500, L = 2$) exceeds 900^oC at 1000 seconds is 0.07. These estimates were constructed using the "low" number of experiments. The actual experimental values used for the analysis are given by the solid black lines. The red lines show experiments that were not used.

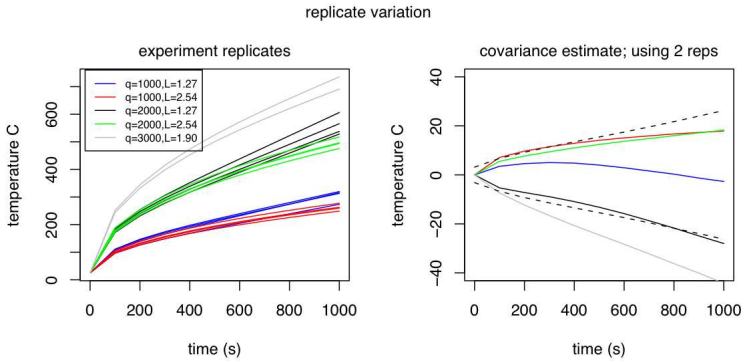

Figure 10: Estimation of experimental replicate variability. By using replicate experiments, the structure of the error covariance matrix Σ_y can be estimated. Left: experimental data by input condition. Right: estimated pointwise 90% uncertainty bounds obtained from the estimate for Σ_y, along with the difference between the two experimental observations used for estimation at each of the 5 configurations.

4 Discussion

We used a statistical approach to answer the certification question posed for this thermal problem. The philosophy of this approach is to treat the experimental outcomes and the simulations both as data that inform about the question of interest. In this setting, there is no formal verification of the model. We are willing to address the problem with whatever imperfect model we have in hand. A better model simply leads to better accuracy in addressing the question.

Clearly the mathematical model given us in this application is imperfect. This is clear from the material characterization data which show the measured k appears to depend on temperature, while the model assumes k is constant. In cases where reality and the simulation model are clearly different, one needs to be careful not to interpret the calibrated parameter settings as physically meaningful.

In this particular application, the lion's share of uncertainty is due to experiment to experiment variability. This may be due to variations in the materials used, or may be due to fluctuations in the experimental apparatus. This clearly has important consequences for the actual certification, as well as for potential mitigation strategies. For example, improving the mathematical model is not likely to reduce prediction uncertainties. However, preselecting materials might. No doubt further study is required to truly solve the certification problem.

References

Bayarri, M., Berger, J., Higdon, D., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J. and Tu, J. (2002). A framework for the validation of computer models, *in* D. Pace and S. Stevenson (eds), *Proceedings of the Workshop on Foundations for V&V in the 21st Century*, Society for Modeling and Simulation International.

Besag, J., Green, P. J., Higdon, D. M. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**: 3–66.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

Higdon, D. (2002). Space and space-time modeling using process convolutions, *in* C. Anderson, V. Barnett, P. C. Chatwin and A. H. El-Shaarawi (eds), *Quantitative Methods for Current Environmental Issues*, Springer Verlag, London, pp. 37–56.

Higdon, D., Gattiker, J. R. and Williams, B. J. (2005). Computer model calibration using high dimensional output, *Technical Report LA-UR-05-6410*, Los Alamos National Laboratory.

Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J. and Ryne, R. D. (2004). Combining field observations and simulations for calibration and prediction, *SIAM Journal of Scientific Computing* **26**: 448–466.

Kennedy, M. and O'Hagan, A. (2001). Bayesian calibration of computer models (with discussion), *Journal of the Royal Statistical Society (Series B)* **68**: 425–464.

Kern, J. (2000). Bayesian process-convolution approaches to specifying spatial dependence structure, Ph.D. Thesis, Institute of Statistics and Decision Sciences, Duke University.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1091.

Oakley, J. and O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs, *Biometrika* **89**(4): 769–784.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*, Springer, New York.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments (with discussion), *Statistical Science* **4**: 409–423.

Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *Design and analysis of computer experiments*, Springer, New York.