

# Combining Experimental Data and Computer Simulations, With an Application to Flyer Plate Experiments

LA-UR-06-

Brian Williams, Los Alamos National Laboratory  
Dave Higdon, Los Alamos National Laboratory  
Jim Gattiker, Los Alamos National Laboratory  
Leslie Moore, Los Alamos National Laboratory  
Michael McKay, Los Alamos National Laboratory  
Sallie Keller-McNulty, Rice University

A flyer plate experiment involves forcing a plane shock wave through stationary test samples of material and measuring the free surface velocity of the target as a function of time. These experiments are conducted to learn about the behavior of materials subjected to high strain rate environments. Computer simulations of flyer plate experiments are conducted with a two-dimensional hydrodynamic code developed under the Advanced Strategic Computing (ASC) program at Los Alamos National Laboratory. This code incorporates physical models that contain parameters having uncertain values. The objectives of the analyses presented in this paper are to assess the sensitivity of free surface velocity to variations in the uncertain inputs, to constrain the values of these inputs to be consistent with experiment, and to predict free surface velocity based on the constrained inputs. We implement a Bayesian approach that combines detailed physics simulations with experimental data for the desired statistical inference (Kennedy and O'Hagan, 2001; Higdon *et al.*, 2004).

The approach given here allows for:

- uncertainty regarding model inputs (i.e. calibration);
- accounting for uncertainty due to limitations on the number of simulations that can be carried out;
- discrepancy between the simulation code and the actual physical system;
- and uncertainty in the observation process that yields the actual field data on the true physical system.

The resulting analysis accomplishes the objectives within a unified framework.

Keywords: flyer plate experiments; calibration; computer experiments; predictability; uncertainty quantification; Gaussian process; model validation; sensitivity analysis; predictive science

## 1 Introduction

A flyer plate experiment involves forcing a plane shock wave through stationary test samples of material and measuring the free surface velocity of the target as a function of time. These experiments are conducted to learn about the behavior of materials subjected to high strain-rate environments relevant to integrated hydrodynamic calculations of implosions driven by high explosives. Figure 1 is an example of an experimental facility capable of producing impactor plate velocities of 3.5 to 6.5 km/s (Trunin, 1998). In this setup, velocity is measured using pin detectors coated with insulating enamel. These detectors are comprised of oppositely charged electrodes that come in contact as

the shock front passes through, sending a signal that is recorded by a measuring device. Modern experiments measure free surface velocity using Velocity Interferometer System for Any Reflector (VISAR) technology.

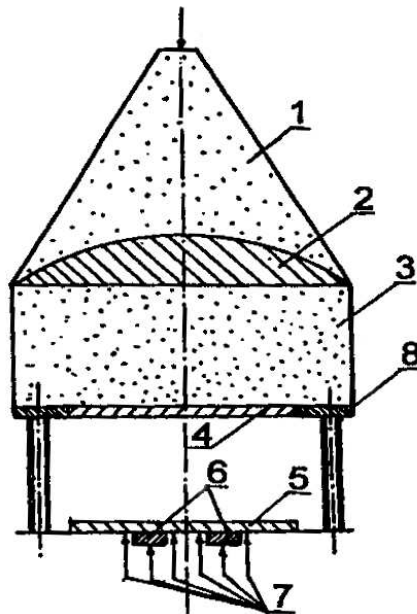


Figure 1: Diagram of flyer plate experiment with an accelerated aluminum impactor: (1) lens-shaped high-explosive charge; (2) correcting lens; (3) main charge; (4) impactor plate; (5) shield from a standard material; (6) tested sample; (7) shorting-pin detectors; (8) peripheral steel ring. (Adapted from Trunin, 1998.)

Figure 2 shows a notional VISAR velocity profile of the shocked material as a function of time. During loading, this material passes through elastic and plastic states prior to reaching peak velocity, at which it remains for the duration of the pulse. A material experiencing stress behaves elastically if it returns to its initial shape upon cessation of loading, and plastically if the loading is strong enough to cause permanent deformation of the material. Phase transitions in the material during the plastic phase are indicated by boundaries in the velocity profile where the slope of the free surface velocity changes abruptly. Unloading occurs elastically and then plastically, and wave interactions in the material can lead to fracturing (spalling).

The analysis of this paper makes use of a single experimental dataset containing the VISAR free surface velocity profile from a tantalum flyer plate experiment. Tantalum is used as a surrogate for hazardous fissile materials. An absolute time scale is not available for these data, so they were uniformly shifted in time by an amount that forced the time at which half the maximum free surface velocity is observed prior to reaching peak velocity to match the corresponding time from a simulation run conducted with nominal values for all input parameters. The analysis objectives are to explore the sensitivity of the simulated velocity profiles to the unknown model parameters, and to calibrate these model parameters to observed flyer plate data, utilizing the general statistical approach for computer model calibration of Kennedy and O'Hagan (2001). The statistical methods of relevance to analysis of the flyer plate data are outlined in Section 2.2. The simulations are carried out using code developed for the Shavano project at the Los Alamos National Laboratory.

Currently substantial statistical research is focusing on the development of methodology for utilizing detailed simulator codes to carry out inference. Issues such as sensitivity analysis, calibration

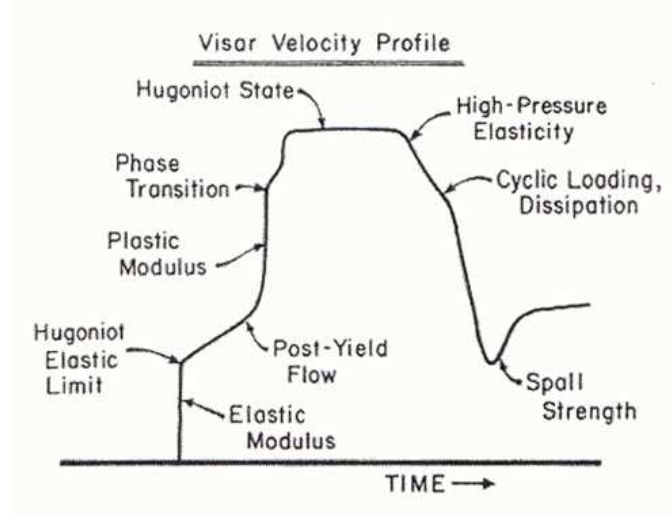


Figure 2: Generic velocity profile of condensed material subjected to shock loading. (Adapted from Meyers, 1994.)

of simulator input parameters, generation of predictions, and characterization of prediction uncertainty are of particular interest. Simulation of well-understood physical processes is typically based on fundamental physical principles. In such problems, the actual amount of observed field data from this process is usually very limited. It is the simulator code that contains the structure of the actual process it is modeling. Because of this, useful inference is possible even with only minimal amounts of observed data on the actual physical system. In contrast, data mining is the opposite extreme – massive amounts of data are used to inform about unknown, and often unconsidered, structure.

To date, much of the statistical methodology relevant to the analysis of computer simulations can be roughly broken into the following nonexclusive categories: experimental design; interpolation or emulation; uncertainty and sensitivity analysis; calibration; and prediction.

## 2 Statistical Methods

In this section we give an overview of the methods used for statistical analysis of the flyer plate data. These include the experimental design used to determine an initial collection of input settings at which to run the simulator as well as statistical model formulation, which allows for estimation of unknown calibration parameters and includes a Gaussian process-based *emulator* which models the simulator output at untried settings. As a byproduct of this formulation, the emulator model can be queried to give a variety of sensitivity measures to describe how changing input settings affect the simulation output.

### 2.1 Computer Experiment Design

To begin the analysis, a simulation campaign consisting of a sequence of computer model runs is required. We require a design that leads to accurate emulator-based predictions over a pre-specified range of input settings. A large number of empirical studies have lead us to favor variants of orthogonal array-based latin hypercube (OA-based LH) designs. Such designs are similar to the space-filling latin hypercubes calculated by the software package Algorithms for the Construction

of Experimental Designs (Welch, 1985) and are a slight extension of the OA-based LH designs given in Tang (1993) and Owen (1994).

We take  $\eta(x, t)$  to denote simulator output given input vector  $(x, t)$ , where the  $p$ -vector  $x$  holds observable, and often controllable, inputs and the  $q$ -vector  $t$  holds additional unobservable calibration and tuning parameters which are required to run the code. When the simulator output is multivariate, we use a component of  $x$  to index the output so that  $\eta(x, t)$  may be treated as univariate. An example of this partitioning of the simulator inputs  $(x, t)$  is given in the flyer plate experiment where  $x$  is time,  $t$  is a collection of uncertain parameters in models describing fundamental properties of the target material, and  $\eta(x, t)$  is the free surface velocity of the target material at time  $x$  for material properties  $t$ .

Given an initial range for each input setting, which we standardize to  $[0, 1]$  for each input, the experimental design effort determines a set of  $m$  input settings  $(x_1^*, t_1^*), \dots, (x_m^*, t_m^*)$  over the  $p + q$ -dimensional hypercube at which to run the simulator. The resulting output  $\eta(x_j^*, t_j^*)$ ,  $j = 1, \dots, m$ , is then used for carrying out the statistical analysis, which includes a Gaussian process (GP) model for the simulator  $\eta(x, t)$  at untried input settings. This GP representation of the simulator response can also be used to explore sensitivities to input variations as is shown in the flyer plate analysis of Section 3. Note that additional constraints on the input settings may lead to additional complications to consider in design specification.

As an example, 2-d projections of a 16 run orthogonal array (OA) design are shown in the lower triangle of Figure 3. This design is over 3 factors, with each factor having 4 levels. The design in Figure 3 has strength 2. This means that for any two inputs every factor level combination occurs an equal number of times. For the 16 run OA, there are precisely 16 values associated with any 2 of the 3 inputs. In terms of an analysis of variance model, effects associated with 4 levels for all 3 inputs would be estimable but interaction effects would not be resolvable. However, if the response is dominated by a subset of the inputs, say only 2, then this strength 2 array would allow estimation of an up to third degree polynomial response surface model. Generally, a strength 3 or higher OA is required to obtain interaction information if many inputs are active. Higher strength of an OA design also assures that the  $m$  design points inherit better space filling properties in higher dimensional projections which is important for fitting a GP model to the simulation output.

Although the OA designs give good coverage in higher dimensions, the one-dimensional projections are quite discrete with only 4 values occurring for each input. On the other hand, LH designs give one-dimensional projections which are a sample of  $m$  values, in this example  $m = 16$  values on each input, but a LH design is effectively a strength 1 OA. Strict OA designs with very discretized input levels can lead to inefficiency when fitting a response surface to a simulator whose response is dominated by a small number of input factors. Linkletter *et al.* (2003) show that by spreading out an OA design so that it becomes a Latin hypercube sample (LHS), the prediction error in response surface models can go down substantially. This LH modification of 2-level OA designs is introduced in Tang (1993) and Owen (1994) and is straightforward to carry out for designs with factors having 2 or more levels. There are catalogues of, and many algorithms for generating, 2-level orthogonal arrays with good properties. Tang (1993) and Owen (1994) argue that an OA-based construction of a LH design results in a LHS with good space-filling properties and smaller correlation between inputs. LH designs constructed in this way have better maximin distance than a random LHS and guaranteed good *binning* properties with the underlying OA *skeleton*. The binning induced by the orthogonal array skeleton is useful for sensitivity studies. Figure 3 shows a LHS construction from the previously mentioned 16 run OA design. Note that this OA-based LH design maintains the same 2-d binning properties of the original OA design. The simulation campaign for the flyer plate

analysis was based on a LH design constructed from an OA of strength 3 having 128 runs and two levels per factor.

From an OA-based LH design, sensitivity analysis can be conducted essentially as suggested in Moore and McKay (2002), comparing  $R^2$  values evaluated from the OA skeleton structure underlying the OA-based LH design or, equivalently, by binning values of the inputs. However, there is added variability in calculations and, thereby, inefficiency in comparing  $R^2$  values based on binned values of inputs specified in a LH design. Modification of OA designs so that they are LH designs is preferred for fitting the simulator output with GP models (see Higdon *et al.* (2004), for example). In cases where a nonparametric sensitivity analysis is the main goal, such as the  $R^2$  analysis of McKay (1995), methods will perform better utilizing an OA design. OA-based LH designs are used to obtain an experiment design with some potential to conduct both analyses, sensitivity and prediction, desired in the simulation studies undertaken here. Another option for conducting sensitivity analysis in the context of a calibration and prediction study is to generate samples of the response surface on a standard OA design and calculate  $R^2$  values from each sample—this is done for the flyer plate analysis of this paper. The result is predicted  $R^2$  indices with associated uncertainty estimates.

Orthogonal arrays for many ( $> 10$ ) factors with more than 2- or 3-levels can dictate more runs than are acceptable, and, as the move to OA-based LH designs indicates, often more than

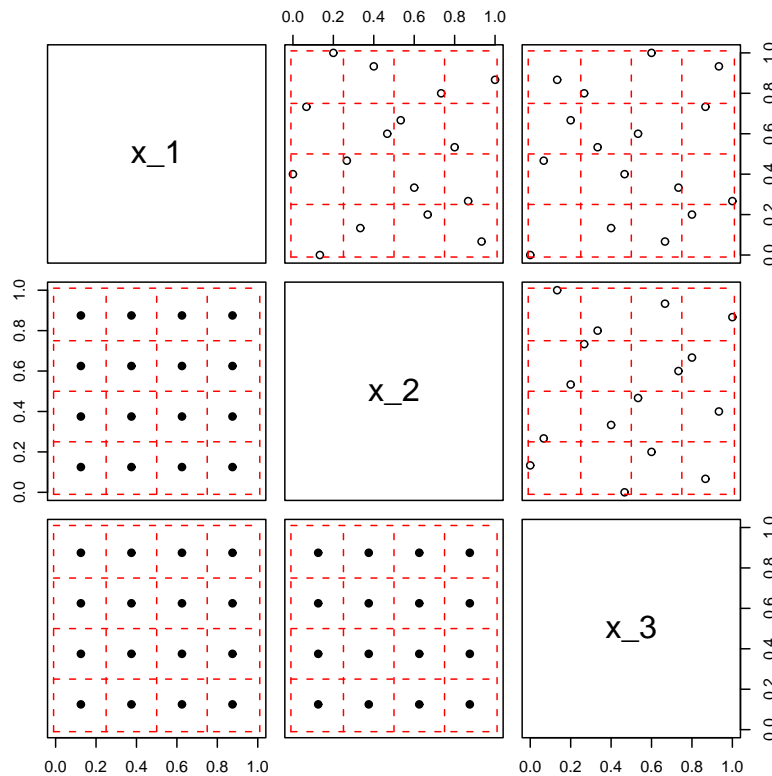


Figure 3: Lower triangle of plots: 2-dimensional projections of a  $m = 16$  point orthogonal array design. Upper triangle of plots: An OA-based LH design obtained by spreading out the OA design so that each 1-dimensional projection gives an equally spaced set of points along  $[0,1]$ .

2- or 3-level factors are desired, even for sensitivity analysis. Strategies for dealing with run size limitations compromise adherence to strict orthogonality. The two main strategies we consider are near orthogonal arrays (Xu (2002)) and combined orthogonal arrays, an idea presented in Moore *et al.* (2004). With an alternative design relaxing strict orthogonality, again LHS construction from these designs as *skeletons*, as with OA-based LH designs, is possible. Additional design references related to this section are included in Santner *et al.* (2003).

## 2.2 Statistical model formulation for calibration and prediction

At various settings for  $x_i$ ,  $n$  observations  $y_i$  are made of the physical system

$$y(x_i) = \zeta(x_i) + \epsilon(x_i), \quad i = 1, \dots, n,$$

where  $\zeta(x_i)$  denotes the response of the actual physical system and the  $\epsilon(x_i)$ 's denote observation error. In a flyer plate experiment,  $\zeta(\cdot)$  represents the actual, unobserved free surface velocity of the target material as a function of time, and  $y(x_i)$  is the observed velocity at time  $x_i$ . Often the size and nature of the  $\epsilon(x_i)$ 's are sufficiently well characterized that their distribution can be treated as known. We take  $y = (y(x_1), \dots, y(x_n))^T$  to denote the physical observations. Often highly multivariate observations are taken from the system; in this case certain components of each  $x_i$  can index the multivariate, observed data so that each  $y(x_i)$  is still univariate. These observed data are then modeled statistically using the simulator  $\eta(x, \theta)$  at the best calibration value  $\theta$  according to

$$y(x_i) = \eta(x_i, \theta) + \delta(x_i) + \epsilon(x_i), \quad i = 1, \dots, n,$$

where the stochastic term  $\delta(x_i)$  accounts for discrepancy between the simulator  $\eta(x_i, \theta)$  and reality  $\zeta(x_i)$ , and  $\theta$  denotes the best, but unknown, setting for the calibration inputs  $t$ . In the context of flyer plate simulations,  $\eta(x_i, \theta)$  represents the calculated free surface velocity of the target material at time  $x_i$ , when the uncertain target material properties have values  $\theta$ . In some cases, the discrepancy term can be ignored; in other cases it plays a crucial role in the analysis.

We treat the fixed set of  $m$  simulation runs

$$\eta(x_j^*, t_j^*), \quad j = 1, \dots, m,$$

as data to be used in the analysis. We are in the situation where the computational demands of the simulation code are so large that only a fairly limited number of runs can be carried out. In this case, a GP model for  $\eta(x, t)$  is required for input combinations  $(x, t)$  for which the simulator has not been run. Note that we use  $t$  to denote an input setting for the calibration parameters here. We reserve  $\theta$  to denote the “best” value of the calibration parameters, which is a quantity about which we wish to infer.

If  $x$  is a vector in  $R^p$  and  $t$  a vector in  $R^q$ , then the function  $\eta(\cdot, \cdot)$  maps  $R^{p+q}$  to  $R$ . We utilize a Gaussian process to model this unknown function (O’Hagan, 1978; Sacks *et al.*, 1989; Santner *et al.*, 2003). A mean function  $\mu(x, t)$  and covariance function  $\text{Cov}((x, t), (x', t'))$  are required to fully specify a GP prior model for  $\eta(x, t)$ . Following Sacks *et al.* (1989) and Kennedy and O’Hagan (2001) we scale all inputs to the unit hypercube, take  $\mu(x, t)$  to be a constant and specify a product covariance having power exponential form

$$\text{Cov}((x, t), (x', t')) = \frac{1}{\lambda_\eta} \prod_{k=1}^p (\rho_k^\eta)^{|2(x_k - x'_k)|^{\alpha_\eta}} \times \prod_{k=1}^q (\rho_{p+k}^\eta)^{|2(t_k - t'_k)|^{\alpha_\eta}}, \quad \rho_k^\eta = \exp(-\beta_k^\eta/4), \quad (1)$$

where the parameter  $\lambda_\eta$  controls the reciprocal of the marginal variance of  $\eta(\cdot, \cdot)$ , the  $(p+q)$ -vector  $\rho^\eta$  controls the dependence strength in each of the component directions of  $x$  and  $t$ , and  $\alpha_\eta$  controls

the smoothness of  $\eta(\cdot, \cdot)$ . A value of  $\alpha_\eta = 2$  leads to a smooth, infinitely differentiable representation for  $\eta(\cdot, \cdot)$ , while taking smaller values of  $\alpha_\eta$  give rougher representations. Our experience has been that the simulator output  $\eta(x_j^*, t_j^*)$ ,  $j = 1, \dots, m$ , does not typically give much information about  $\alpha_\eta$ . Hence we usually fix  $\alpha_\eta$  based on prior information regarding  $\eta(\cdot, \cdot)$  or based on computational considerations. For  $\alpha_\eta = 2$ ,  $\rho_k^\eta$  is the correlation between outputs evaluated at inputs that vary in only the  $k$ -th dimension by half their domain. We note that it is often useful to add a small white noise component to the covariance model (1) to account for small numerical fluctuations in the simulation. Such fluctuations can be caused by slight changes in adaptive meshing or tolerances caused by changes in the input settings. For models with random outputs, such as epidemiological or agent based models, an additional independent error term will be required in (1) above.

The prior model specification is completed by specifying independent priors for the parameters controlling  $\eta(\cdot, \cdot)$ :

$$\begin{aligned}\pi(\mu) &\propto \exp\left\{-\frac{1}{2v}\mu^2\right\} \\ \pi(\lambda_\eta) &\propto \lambda_\eta^{a_\eta-1} \exp\{-b_\eta\lambda_\eta\}, \lambda_\eta > 0 \\ \pi(\rho^\eta) &\propto \prod_{k=1}^{p+q} (1 - \rho_k^\eta)^{-.9}, \quad 0 \leq \rho_k^\eta \leq 1.\end{aligned}$$

Because of the standardization, we can simplify the parameterization and MCMC by fixing  $\mu$  at 0 (i.e.,  $v = 0$ ) and encouraging  $\lambda_\eta$  to be close to 1 by taking  $a_\eta = b_\eta = 5$ . The prior for  $\rho^\eta$  encourages strong dependence in each of the component directions so that prior realizations for  $\eta(\cdot, \cdot)$  are generally quite flat. Hence it will be the data that move the  $\rho_k^\eta$ 's away from 1 in the posterior.

We specify a GP model for the discrepancy term  $\delta(x)$  with mean function of 0, and a covariance function of the form

$$\text{Cov}(x, x') = \frac{1}{\lambda_\delta} \prod_{k=1}^p (\rho_k^\delta)^{|2(x_k - x'_k)|^{\alpha_\delta}}, \quad \rho_k^\delta = \exp(-\beta_k^\delta/4). \quad (2)$$

The prior specification for the parameters governing the GP model for  $\delta(\cdot)$  mirrors the specification for  $\eta(\cdot, \cdot)$ ,

$$\begin{aligned}\pi(\lambda_\delta) &\propto \lambda_\delta^{a_\delta-1} \exp\{-b_\delta\lambda_\delta\}, \lambda_\delta > 0 \\ \pi(\rho^\delta) &\propto \prod_{k=1}^p (1 - \rho_k^\delta)^{-.7}, \quad 0 \leq \rho_k^\delta \leq 1.\end{aligned}$$

Here .7 in the prior for  $\rho^\delta$  gives  $\delta(\cdot)$  a slightly weaker tendency towards flatness than  $\eta(\cdot, \cdot)$ .

We define  $y = (y(x_1), \dots, y(x_n))^T$  to be the vector of field observations and  $\eta = (\eta(x_1^*, t_1^*), \dots, \eta(x_m^*, t_m^*))^T$  to be the simulation outcomes from the experimental design. Now we define the joint  $(n+m)$ -vector  $\mathcal{D} = (y^T, \eta^T)^T$  which has associated simulation input values  $(x_1, \theta), \dots, (x_n, \theta)$  for its first  $n$  components and  $(x_1^*, t_1^*), \dots, (x_m^*, t_m^*)$  for its final  $m$  components. The sampling model, or likelihood, for the observed data  $\mathcal{D}$  is then

$$L(\mathcal{D}|\theta, \mu, \lambda_\eta, \rho^\eta, \lambda_\delta, \rho^\delta, \Sigma_y) \propto |\Sigma_{\mathcal{D}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathcal{D} - \mu\mathbf{1}_{n+m})^T \Sigma_{\mathcal{D}}^{-1} (\mathcal{D} - \mu\mathbf{1}_{n+m})\right\}, \quad (3)$$

where  $\mathbf{1}_{n+m}$  is the  $(n+m)$ -vector of ones and

$$\Sigma_{\mathcal{D}} = \Sigma_\eta + \begin{pmatrix} \Sigma_y + \Sigma_\delta & 0 \\ 0 & 0 \end{pmatrix},$$

where  $\Sigma_y$  is the  $n \times n$  observation covariance matrix, elements of  $\Sigma_\eta$  are obtained by applying (1) to each pair of the  $n + m$  simulation input points corresponding to  $\mathcal{D}$ , and  $\Sigma_\delta$  is a  $n \times n$  matrix obtained by applying (2) to each pair of the  $n$  input settings  $x_i$ ,  $i = 1, \dots, n$ , that correspond to the observed field data  $y$ . Note that  $\Sigma_\eta$  depends on the experimental input conditions  $x_i$ , the simulator input conditions  $(x_j^*, t_j^*)$ , and the parameter value  $\theta$ . Hence updating  $\theta$  affects  $\Sigma_\eta$  which means its determinant and a linear solve need to be recomputed to evaluate (3).

The resulting posterior density has the form

$$\pi(\theta, \mu, \lambda_\eta, \rho^\eta, \lambda_\delta, \rho^\delta | \mathcal{D}) \propto L(\mathcal{D} | \theta, \mu, \lambda_\eta, \rho^\eta, \lambda_\delta, \rho^\delta, \Sigma_y) \times \pi(\mu) \times \pi(\lambda_\eta) \times \pi(\rho^\eta) \times \pi(\lambda_\delta) \times \pi(\rho^\delta) \quad (4)$$

which can be explored via Markov chain Monte Carlo (MCMC) or some other numerical integration scheme. We use tuned, univariate random walk metropolis updates in our MCMC implementation (Metropolis *et al.*, 1953). Figure 4 shows the results from a simple example consisting of a one-dimensional  $x$  and  $t$ , using  $m = 20$  simulations and  $n = 5$  experimental observations.

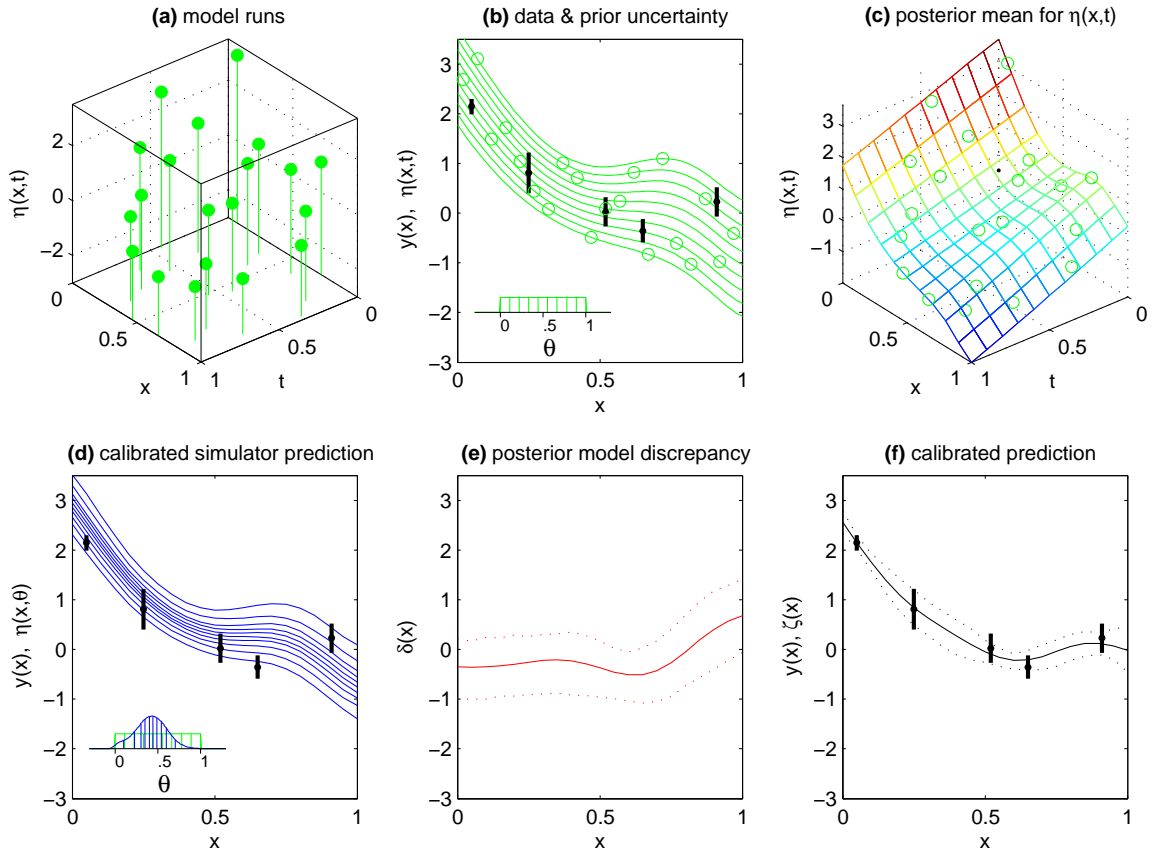


Figure 4: Basic model formulation. (a) An initial set of simulation runs are carried out over the input settings  $(x_j^*, t_j^*)$ ,  $j = 1, \dots, m$ . (b) Experimental data are collected at  $n$  input settings; data are given by the black dots; 90% uncertainties are given by the black lines. The green circles correspond to the  $m$  simulation runs. (c) Posterior mean estimate for the simulator output  $\eta(x, t)$ . (d) Posterior distribution for the calibration parameter  $\theta$  and the resulting simulator-based predictions (blue lines). (e) Posterior mean estimate and pointwise 90% prediction intervals for the model discrepancy term  $\delta(x)$ . (f) Posterior mean estimate and pointwise 90% prediction intervals for the physical system  $\zeta(x)$ .



### 2.3 Computer Model Emulation and Sensitivity Analysis

Given the MCMC output of parameters governing the various GP models, posterior realizations of  $\eta(x, t)$  can be generated using standard theory. As an example, we consider modeling the output of the 3-d function

$$\eta(x) = (x_1 + 1) \cos(\pi x_2) + 0x_3.$$

Although  $\eta(x)$  takes in a 3-dimensional input  $x$ , the third factor does not affect the output. The estimated emulator can account for this by estimating a value for  $\rho_3^\eta$  which is close to 1. We evaluate  $\eta(x)$  according to the  $n = 16$  run OA-based LH design from Figure 3. The computed values  $\eta(x_i^*)$  along with the posterior mean for  $\eta(x)$  are shown in Figure 5 as a function of  $x_1$  and  $x_2$ . The resulting posterior medians for  $(\rho_1^\eta, \rho_2^\eta, \rho_3^\eta)$  are (.95, .68, .999). Hence, the resulting

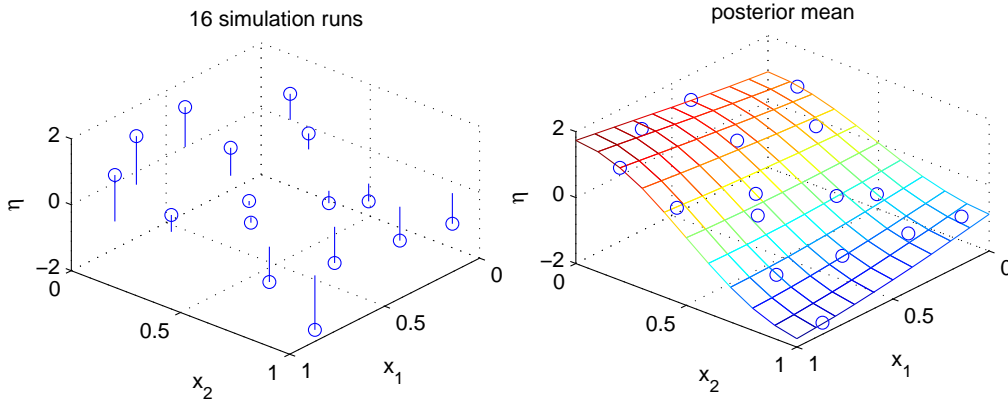


Figure 5: Left: simulator output for the  $\eta(x) = (x_1 + 1) \cos(\pi x_2) + 0x_3$  as a function of  $x_1$  and  $x_2$ . Right: posterior mean for  $\eta(x)$  as a function of  $x_1$  and  $x_2$ . The estimated emulator correctly does not depend on  $x_3$ .

emulator is effectively independent of  $x_3$ . The posterior mean estimate for  $\eta(x)$  is shown as a function of  $(x_1, x_2)$  in the right frame of Figure 5. This analysis also gives uncertainties regarding this emulator.

The posterior distribution of  $\rho^\eta$  is one measure of simulator output sensitivity to inputs. More generally, sensitivity analysis studies how variation in simulator response can be apportioned to the various code inputs. Formal sensitivity analyses can be carried out in a variety of ways. Empirical or sampling based approaches are discussed in Saltelli *et al.* (2000); model based approaches which first build an emulator and carry out sensitivity analyses on this emulator are discussed in Welch *et al.* (1992) and, from a Bayesian perspective, in Oakley and O'Hagan (2004).

One approach we have been using is the empirical  $R^2$  measure of McKay (1995). This measure is one of a host of measures based on sensitivity indices which make use of an ANOVA-type decomposition (Sobol', 1993) of the simulator  $\eta(x)$ . For simplicity we assume  $\eta(x)$  is defined over  $x \in [0, 1]^p$ .

Sobol' (1993) shows that there is a unique decomposition

$$\eta(x_1, \dots, x_p) = \eta_0 + \sum_{k=1}^p \eta_k(x_k) + \sum_{1 \leq k < \ell \leq p} \eta_{k\ell}(x_k, x_\ell) + \dots + \eta_{1,2,\dots,p}(x_1, x_2, \dots, x_p), \quad (5)$$

for which

$$\eta_0 = \int_{[0,1]^p} \eta(x_1, \dots, x_p) dx_1 \cdots dx_p \quad \text{and} \quad \int_0^1 \eta_{k_1, \dots, k_s}(x_{k_1}, \dots, x_{k_s}) dx_{k_i} = 0$$

for any  $1 \leq i \leq s$  and has orthogonal components.

So, for example, the main effect for factor  $k$  of the simulator is given by

$$\eta_k(x_k) = \int_{[0,1]^{p-1}} \eta(x) dx_{-k} - \eta_0$$

where  $x_{-k}$  denotes all input factors except  $k$ . The three main effects estimated from the emulator in our simple 3-d example are given by the lines in Figure 6. The two-factor interaction effect for factors  $j$  and  $k$  of the simulator is given by

$$\eta_{jk}(x_j, x_k) = \int_{[0,1]^{p-2}} \eta(x) dx_{(-j,-k)} - \eta_j(x_j) - \eta_k(x_k) - \eta_0$$

where  $x_{(-j,-k)}$  denotes all input factors except  $j$  and  $k$ .

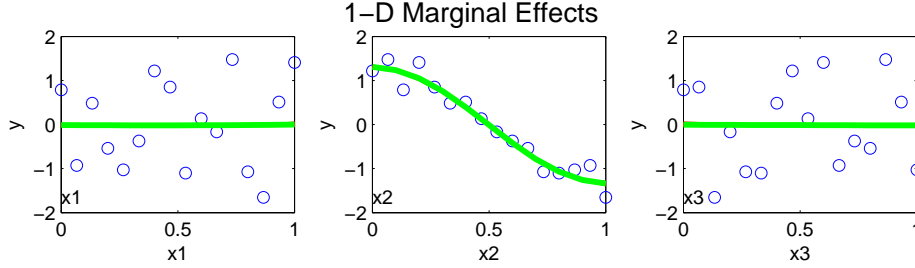


Figure 6: 1-d marginalizations of the posterior mean fit shown in Figure 5. The main effect functions  $\eta_i(x_i)$  are given by the green lines and the 16 calculations are given by the blue circles. Here factor 2 is clearly important giving an  $S_2 = .98$ . The  $S_k$  for factors 1 and 3 is 0.

The  $R^2$  measure of McKay (1995) is a variance-based measure of sensitivity which complements these plot diagnostics. Before getting to  $R^2$ , we first describe a family of variance-based measures of  $\eta(x)$ . The total variance  $V$  of  $\eta(x)$  is defined to be

$$V = \int_{[0,1]^p} \eta^2(x) dx - \eta_0^2.$$

Partial variances are computed from each of the terms of Sobol's decomposition (5) as

$$V_{k_1, \dots, k_s} = \int_{[0,1]^s} \eta_{k_1, \dots, k_s}^2(x_{k_1}, \dots, x_{k_s}) dx_{k_1} \cdots dx_{k_s}$$

for  $s = 1, \dots, p$  and  $1 \leq k_1 < \dots < k_s \leq p$ . If both sides of (5) are squared and integrated over  $[0, 1]^p$ , one obtains

$$V = \sum_{k=1}^p V_k + \sum_{1 \leq k < \ell \leq p} V_{k\ell} + \cdots + V_{1,2, \dots, p}$$

due to the orthogonality of the terms in (5). Sensitivity measures  $S_{k_1, \dots, k_s}$  are given by

$$S_{k_1, \dots, k_s} = \frac{V_{k_1, \dots, k_s}}{V}.$$

The  $R^2$  sensitivity index applied to the factor group  $\{k_1, \dots, k_s\}$  measures its combined effect on the output,

$$R_{k_1, \dots, k_s}^2 = \frac{\text{Var}[\mathbb{E}(\eta(x) | x_{k_1}, \dots, x_{k_s})]}{V} = \sum_{i=1}^s \sum_{\substack{\omega \subset \{k_1, \dots, k_s\}; \\ |\omega|=i}} S_\omega.$$

For the simple 3-d example, the theoretical values of the sensitivity indices are  $S_2 = 27/28$  and  $S_{12} = 1/28$ , with all others equal to zero. The quantities  $S_k$ ,  $k = 1, \dots, p$ , are the first order sensitivities (main effects). The theoretical values of  $R^2$  are  $R_2^2 = 27/28$ ,  $R_{12}^2 = 1$ ,  $R_{23}^2 = 27/28$ , and  $R_{123}^2 = 1$ .  $(S_1, S_2, S_3)$  and  $(R_1^2, R_2^2, R_3^2)$  are estimated from the emulator to be  $(0, .98, 0)$ . Note here that if one were only to look at the first order sensitivities, factor 1 would not be chosen as important. From Figure 5 it is clear that input  $x_1$  does affect  $\eta(x)$ . Higher order sensitivities do show input 1 to be active as a second order interaction with input 2 (estimated  $S_{12} = .02$ ). In the special case that  $[0, 1]^p$  is discretized to a grid with discrete uniform measure, the partial variances become the sums of squares and the sensitivity measures become the  $R^2$  values of a standard ANOVA decomposition. For sufficiently high-dimensional inputs, calculations of sensitivity indices on the entire grid become infeasible due to the complexity of the calculations involved, in terms of the number of simulations required or the order of the linear solve required to generate realizations from the emulator. A subset of grid points is used in place of the entire grid. However, it is now impossible to obtain uncorrelated estimates of all sensitivity indices, but a properly chosen design can allow the lower order effects to be estimated cleanly if it is assumed that higher order effects are negligible. Orthogonal array designs are often used for this purpose.

Designs having at least three levels per variable are desirable, allowing quadratic main effects to be estimated. Main effects will be confounded with three-factor or higher interactions in strength 3 OA designs, and with four-factor or higher interactions in strength 4 OA designs, while two-factor interactions will be confounded with other two-factor or higher interactions in strength 3 OA designs and with three-factor or higher interactions in strength 4 OA designs. Strength 2 OA designs are not desirable as some main effects will be confounded with two-factor interactions. In short, if three-factor or higher interactions are assumed negligible, a strength 3 OA design will allow main effects to be estimated cleanly and a strength 4 OA design will allow main effects and all two-factor interaction effects to be estimated cleanly.

The main effect and two-factor interaction effect sensitivity indices and functions can be estimated using an emulator of the response surface constructed from posterior realizations generated on a sequence of runs specified by an appropriate  $\text{OA}(N, s_1 s_2 \dots s_p, t)$  design, where  $N$  denotes the number of runs in the design,  $p$  the number of input parameters,  $s_k$  the number of levels for input  $k$ , and  $t$  the strength of the array:

1. Generate a draw  $\eta^*$  given the data and a posterior realization of the parameters  $(\theta, \mu, \lambda_\eta, \rho^\eta, \lambda_\delta, \rho^\delta)$  from (4) at untried input settings  $x_i^*$  taken from the OA design.
2. Let  $\overline{\eta_{-k}^*}(x_{k,i})$  denote the average of the  $\eta^*$  having the  $k$ -th input set to  $x_{k,i}$  ( $i = 1, \dots, s_k$  and  $k = 1, \dots, p$ ), and  $\overline{\eta^*}$  denote the overall mean. Estimate the main effect of input  $k$  at  $x_{k,i}$  by

$$\widehat{\eta}_k(x_{k,i}) = \overline{\eta_{-k}^*}(x_{k,i}) - \overline{\eta^*}.$$

3. Estimate  $S_k$  for each input factor  $k$  by

$$\widehat{S}_k = \frac{\frac{N}{s_k} \sum_{i=1}^{s_k} \widehat{\eta}_k(x_{k,i})^2}{\sum_{i=1}^N (\eta^*(x_i^*) - \overline{\eta^*})^2}.$$

4. Let  $\overline{\eta_{-j,-k}^*}(x_{j,h}, x_{k,i})$  denote the average of the  $\eta^*$  having the  $j$ -th input set to  $x_{j,h}$  and the  $k$ -th input set to  $x_{k,i}$  ( $h = 1, \dots, s_j$ ,  $i = 1, \dots, s_k$ ,  $k = j + 1, \dots, p$  and  $j = 1, \dots, p - 1$ ). Estimate the two-factor interaction effect of inputs  $j$  and  $k$  at  $(x_{j,h}, x_{k,i})$  by

$$\widehat{\eta}_{j,k}(x_{j,h}, x_{k,i}) = \overline{\eta_{-j,-k}^*}(x_{j,h}, x_{k,i}) - \widehat{\eta}_j(x_{j,h}) - \widehat{\eta}_k(x_{k,i}) - \overline{\eta^*}.$$

5. Estimate  $S_{jk}$  for each pair of inputs  $j$  and  $k$  by

$$\widehat{S}_{jk} = \frac{\frac{N}{s_j s_k} \sum_{h=1}^{s_j} \sum_{i=1}^{s_k} \widehat{\eta}_{j,k}(x_{j,h}, x_{k,i})^2}{\sum_{i=1}^N (\eta^*(x_i^*) - \overline{\eta^*})^2}.$$

6. Repeat the previous steps for multiple posterior realizations. The result is a sample from the posterior predictive distributions of the sensitivity indices, and of the main effect and two-factor interaction functions on a grid of input values defined by the OA.

Posterior means and quantiles are easily derived from these samples and constitute point estimates of and probability bounds on these indices, and pointwise estimates/bounds on the functions restricted to the grid of inputs from the OA.

### 3 Analysis of a Flyer Plate Experiment

In this analysis of a flyer plate experiment, ten parameters are calibrated to experimental data. The results of this calibration can be used to restrict variation in these parameters for downstream integrated calculations that involve potentially many other uncertain parameters that may be independently calibrated using a similar approach. The free surface velocity is a function of both equation of state (EOS) and material strength. One calibration parameter specifies deviations from the nominal tantalum EOS table. Material strength is incorporated through the Preston-Tonks-Wallace (PTW) model (Preston, Tonks and Wallace, 2003). Seven PTW model parameters that describe the plastic stress-strain relationship are calibrated. A parameter specifying the spall strength of tantalum in the damage model is calibrated. Finally, the flyer plate impact velocity is calibrated due to deficiencies in the mechanism used to measure this velocity in the experiment.

For this simulation study, the allowed domain of variability for the seven free PTW parameters is described by the boundaries of a minimum volume, hyper-rectangular probability region (covering as much probability as possible up to the target of 95%) based on the joint posterior distribution of these parameters, determined by fitting the PTW model simultaneously to several Hopkinson bar and quasi-static stress-strain datasets under a variety of strain rate and temperature conditions (Fugate *et al.*, 2005; Hanson, 2004). Table 1 summarizes the ten calibration parameters and their domains for this analysis.

The simulator was run at 128 unique combinations of the ten calibration parameters as specified by an OA-based LH design generated from a two-level, strength 3 OA design (Tang, 1993). The design values for  $\epsilon$  were binned to the eleven integer values in the range  $[-5, 5]$ . The left panel of Figure 7 shows the resulting free surface velocity profiles in gray, with the experimental velocity profile in blue and the mean of the calculations in red. The right panel of Figure 7 shows more detail, with the mean of the calculations subtracted out of each simulated profile and the experimental data. The dashed vertical segments at the bottom of the right panel indicate one hundred locations along the time axis at which calculations and data were extracted for this statistical analysis.

The single physical parameter  $x$  in this analysis is time, and the ten calibration parameters  $\theta$  are described in Table 1. The calculations  $\eta(x, t)$  are taken at the locations  $(x_1, t_1), \dots, (x_{100}, t_1)$ ,

Input	Description	Domain	
		Min	Max
$\epsilon$	Perturbation of EOS table from nominal	-5%	5%
$\theta_0$	Initial strain hardening rate	$2.78 \times 10^{-5}$	0.0336
$\kappa$	Material constant in thermal activation energy term — relates to the temperature dependence	0.438	1.11
$\gamma$	Material constant in thermal activation energy term — relates to the strain rate dependence	$6.96 \times 10^{-8}$	$6.76 \times 10^{-4}$
$y_0$	Maximum yield stress (at 0 K)	0.00686	0.0126
$y_\infty$	Minimum yield stress ( $\sim$ melting)	$7.17 \times 10^{-4}$	0.00192
$s_0$	Maximum saturation stress (at 0 K)	0.0126	0.0564
$s_\infty$	Minimum saturation stress ( $\sim$ melting)	0.00192	0.00616
pspall	Spall strength	-0.055	-0.045
$v_s$	Flyer plate impact velocity	329.5	338.5

Table 1: Calibration parameters with input domains.

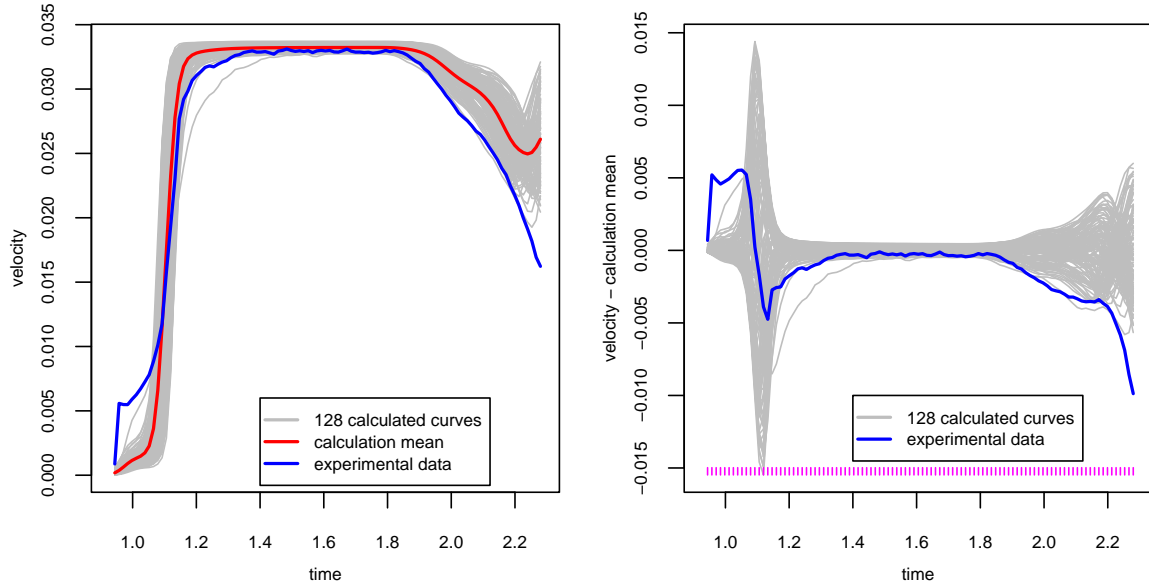


Figure 7: Free surface velocity profiles in gray, with the experimental velocity profile in blue and the mean of the calculations in red (left panel). Mean of the calculations subtracted out of each simulated profile and the experimental data (right panel).

$\dots, (x_1, t_{128}), \dots, (x_{100}, t_{128})$ , and the experimental data  $y(x)$  are taken at the locations  $x_1, \dots, x_{100}$ . The calculations and experimental data are standardized by subtracting out the mean of the calculations at each  $x$  location, and dividing these residuals by the overall standard deviation of the calculations. The statistical analysis takes advantage of Kronecker product structure in the covariance matrices to make likelihood calculations tractable; without this structure, the analysis would be limited to a more restrictive extraction of calculations and experimental data. Table 2 gives the prior distributions on all of the parameters involved in the statistical model.

The Gaussian process model for  $\eta(x, t)$  includes a nugget effect with precision having prior

Parameters	Description	Prior Distribution
$\theta$	Calibration parameters defined in Table 1	Uniform on hyper-rectangle defined by domains in Table 1
$\rho^\eta$	Coefficient in simulator covariance model (1)	Beta(1, 0.1)
$\rho^\delta$	Coefficient in discrepancy covariance model (2)	Beta(1, 0.3)
$\lambda_\eta$	Simulator precision	Gamma(5, 5)
$\lambda_\delta$	Discrepancy precision	Gamma(1, 0.0001)
$\lambda_y$	Measurement precision	fixed at 1

Table 2: Prior distributions on statistical model parameters.

distribution Gamma(1, 0.0001). The measurement error variance at location  $x_i$  is assumed to be  $C_i^2/\lambda_y$ , where  $C_i = 0.05371$  is the standard deviation of the empirical free surface velocities in the relatively constant peak velocity portion of the profile prior to unloading. The prior distribution for  $\lambda_y$  is fixed at 1, so that the measurement error standard deviation takes the value  $C_i$  at location  $x_i$ . The Markov chain was run for 10,000 iterations, after discarding the first 2,100 samples. Figure 8 shows histograms of the retained posterior samples for the calibration parameters  $\theta_0$  and  $y_0$ , with the assumed uniform prior indicated by the straight line segment. The posterior distributions of these parameters shows a distinct gain in information on the plausible range and concentration of values useful for calibrating this simulator to the experimental data. In fact, the parameters  $\epsilon$ ,  $\kappa$ ,  $\gamma$ ,  $\text{psall}$  and  $v_s$  calibrate strongly to the lower end of their ranges, while  $y_0$ ,  $y_\infty$  and  $s_0$  calibrate to the higher end of their ranges. The parameters  $\theta_0$  and  $s_\infty$  calibrate to the extremes of their ranges, with this effect most pronounced for  $\theta_0$ .

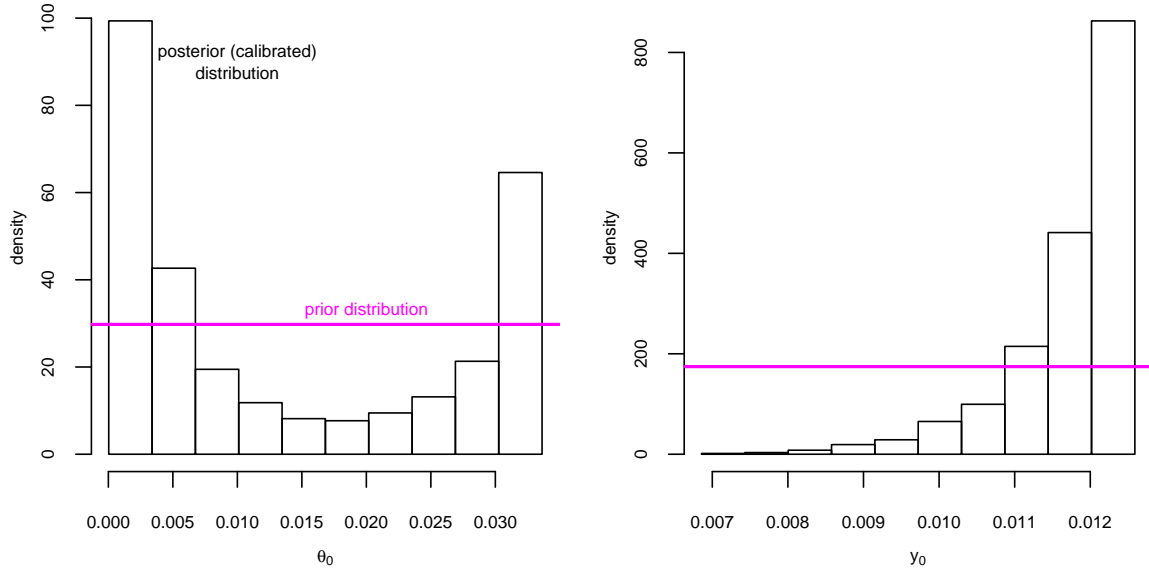


Figure 8: Histograms of the retained posterior samples for the calibration parameters  $\theta_0$  and  $y_0$ . The assumed uniform prior is indicated by the straight line segment.

Figure 9 shows 90% pointwise probability intervals for the stress-strain relationship calculated

from the PTW model based on samples from the prior and posterior distributions of the 7 PTW parameters of Table 1. The main implication of this figure is the tantalum sample used in this flyer plate experiment is slightly less compressible than the tantalum samples used in the Hopkinson bar and quasi-static experiments. Although the posterior distribution excludes the higher compressibility regions admitted by the prior, there is considerable overlap in the stress-strain relationship allowed by these distributions. Calibration to multiple flyer plate experiments conducted under a variety of environments is required to reach conclusions robust to non-physical sources of bias or variation.

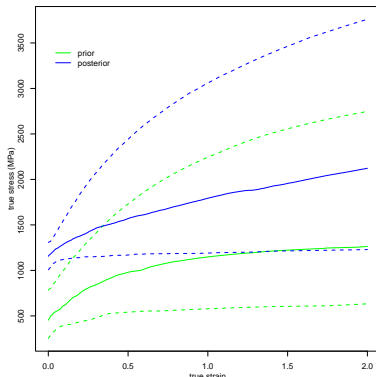


Figure 9: Pointwise 90% stress-strain curves calculated from the PTW model based on prior (green) and posterior (blue) samples of the 7 PTW parameters. The pointwise median curves are given by the solid lines.

Figure 10 shows kernel density estimates of the bivariate posterior distributions of the ten calibration parameters. The importance of essentially every parameter (with the exception of  $s_\infty$ ) to the calibration is indicated by a concentration of probability on a subset of their domains. Bivariate relationships between calibration parameters can be discerned from this figure. In this example, there is strong negative correlation between  $\epsilon$  and  $\theta_0$  ( $-0.89$ ) and moderate correlations between several other parameter pairs:  $y_0$  and  $\theta_0$  ( $-0.43$ ),  $\epsilon$  and  $y_0$  ( $0.38$ ),  $\epsilon$  and  $v_s$  ( $-0.3$ ), and  $y_0$  and  $s_\infty$  ( $0.28$ ). The bimodal nature of the posterior distribution for  $\theta_0$  (and to a lesser extent,  $s_\infty$ ) is due in part to these correlation relationships between parameters. For example, increasing  $\epsilon$  or  $\theta_0$  increases the stiffness of the target material. These parameters are able to trade-off against each other to achieve similar velocity profiles. Additional sources of information beyond the single flyer plate experiment available for this analysis are necessary to restrict the feasible trade-off spaces among these parameters.

Figure 11 shows a posterior prediction of the experimental data based solely on the calibrated simulator  $\eta_c(x)$ , i.e. without the discrepancy adjustment  $\delta(x)$ . Specifically  $\eta(x, \theta)$ , given all of the observed data  $\mathcal{D}$  and  $\theta$ , is a Gaussian process. Let  $\xi = (\theta, \lambda_\eta, \rho^\eta, \lambda_\delta, \rho^\delta, \lambda_y)$ . Given posterior samples  $\xi_1, \dots, \xi_N$ , the distribution of the calibrated simulator at  $x$  is estimated to be

$$\pi(\eta_c(x) | \mathcal{D}) = \int \pi(\eta(x, \theta) | \mathcal{D}, \xi) \pi(\xi | \mathcal{D}) d\xi \approx \frac{1}{N} \sum_{i=1}^N \pi(\eta(x, \theta_i) | \mathcal{D}, \xi_i).$$

The solid red line is the median of  $\pi(\eta_c(x) | \mathcal{D})$ , and the lower and upper dashed red lines are the 5% and 95% quantiles. Note the reduction in the uncertainty of this posterior prediction relative to prior uncertainty represented by variability in the gray traces corresponding to calculations on the uniform space-filling experimental design.

In the right panel of Figure 11, note the presence of substantial discrepancy between the calibrated simulator and the data prior to, and following, the peak velocity portion of the profile prior

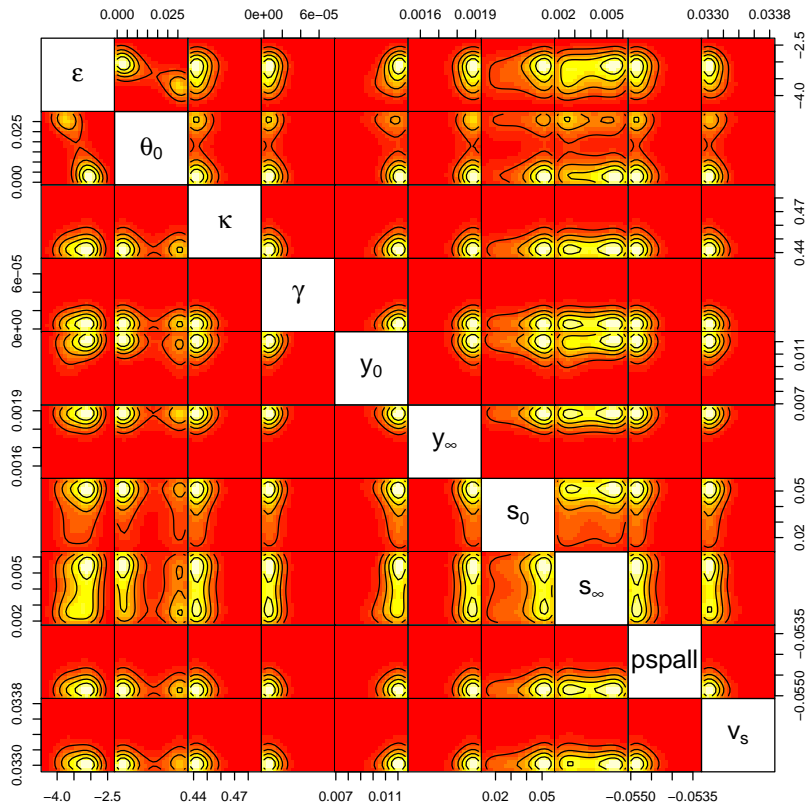


Figure 10: Kernel density estimates of the bivariate posterior distributions of the ten calibration parameters.

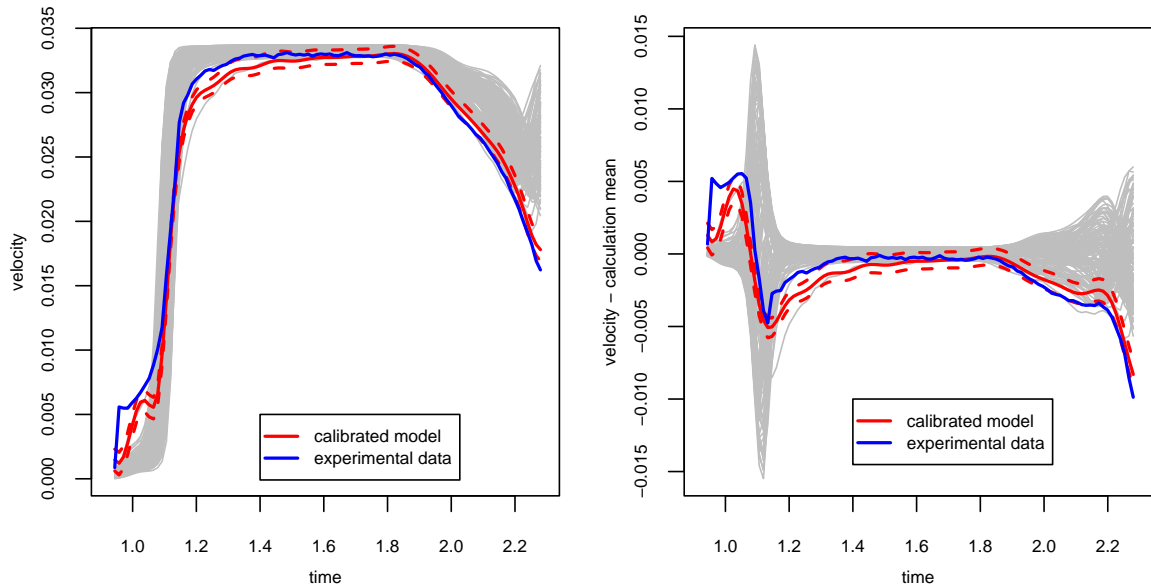


Figure 11: Posterior prediction of the experimental data based solely on the calibrated simulator  $\eta_c(x)$ .



to unloading, particularly in the vicinity of the Hugoniot elastic limit (HEL) at  $x = 0.95$ . Figure 12 shows a posterior prediction of the experimental data based on the calibrated simulator  $\eta_c(x)$  adjusted for discrepancy  $\delta(x)$ . Specifically, this prediction is based on the posterior distribution of  $\zeta(x)$ ,

$$\pi(\zeta(x) | \mathcal{D}) = \int \pi([\eta(x, \theta) + \delta(x)] | \mathcal{D}, \xi) \pi(\xi | \mathcal{D}) d\xi \approx \frac{1}{N} \sum_{i=1}^N \pi([\eta(x, \theta_i) + \delta(x)] | \mathcal{D}, \xi_i).$$

This correction improves the prediction quality along most of the velocity curve, but lack of fit remains in the vicinity of the HEL, where discrepancy between the data and simulation model is particularly abrupt and pronounced. The inability to adjust for deviations of this nature is a consequence of the fact that discrepancy is modeled as a stationary Gaussian process, which can only correct *smooth* deviations of the calibrated simulator from the experimental data.

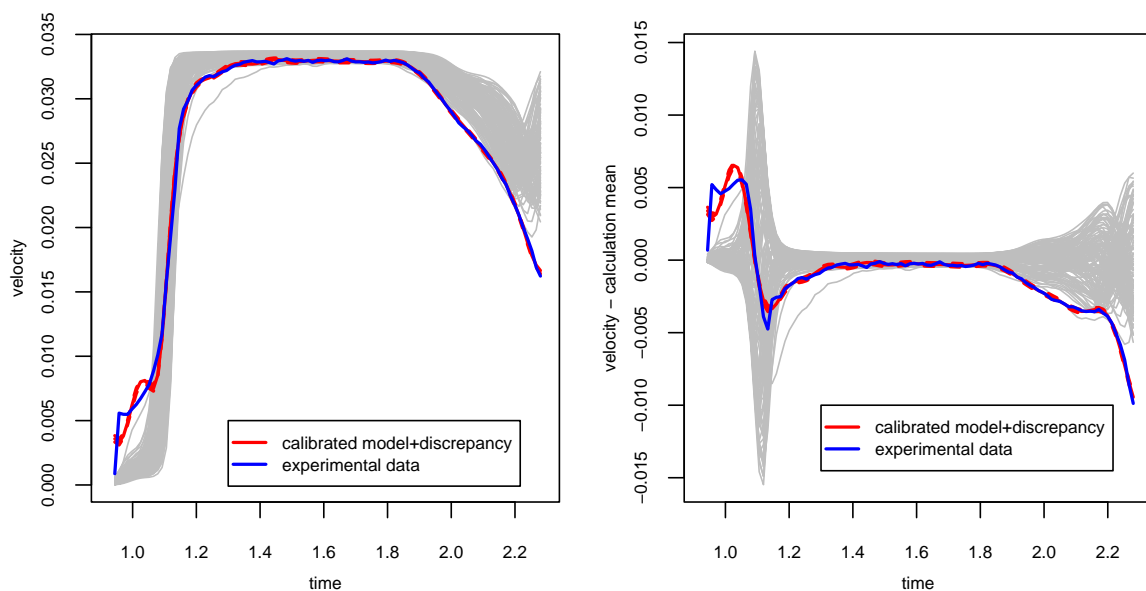


Figure 12: Posterior prediction of the experimental data based on the calibrated simulator  $\eta_c(x)$  adjusted for discrepancy  $\delta(x)$ .

Figure 13 shows results from two sensitivity analyses conducted on the flyer plate simulation output. The goal of these analyses is to provide some information about which input parameters are most influential in explaining output variability. The first order sensitivity indices of Section 2.3 define the metric applied to assess input importance. Each index corresponds to an input parameter, estimating the proportion of total variation in the output explained by variation in that parameter. The first analysis used outputs extracted from the simulation runs at times corresponding to two important features in the experimental data: the HEL and the damage point. The second analysis was based on a principal components decomposition of the mean-centered simulation runs as described in McKay *et al.* (2005). The outputs were the three sets of coefficients corresponding to the first three empirical orthogonal basis functions. This type of analysis is useful when each basis function has a physical interpretation. In this event, the partitioning of total variability in the coefficients of a particular basis function into components explained by the input main effects provides information about the degree to which each input influences the physical characteristics described by that basis function.

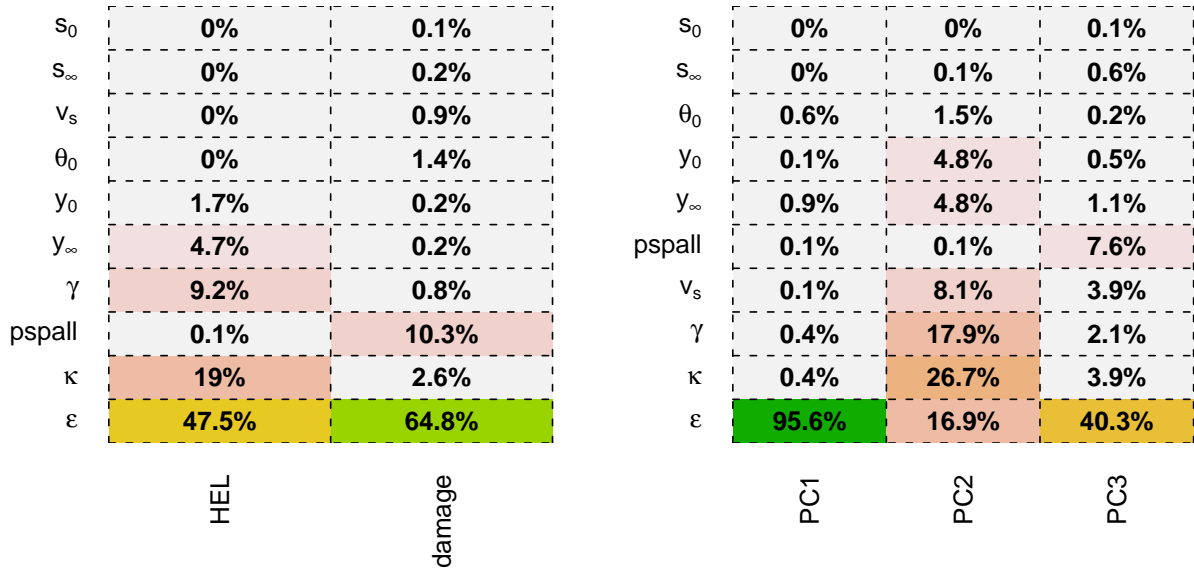


Figure 13: First order sensitivity indices for each of the ten calibration parameters. Outputs are simulated free surface velocities extracted at the HEL and damage points in the experimental data (left panel), and coefficients of the first three principal components calculated from the mean-centered simulation runs (right panel).

In Figure 13, the ten calibration inputs are listed on the  $y$ -axis and the outputs are listed on the  $x$ -axis of each plot. The ordering of the inputs is based on their degree of importance as defined by the mean of their main effect indices across the outputs, with more influential inputs situated lower on the  $y$ -axis. The ordering of the outputs is based on the degree to which they can be explained by main effects, with outputs having higher proportion of total variation due to main effects situated to the left on the  $x$ -axis. The parameter  $\epsilon$  specifying deviations from the nominal EOS table is the most influential input in both analyses, followed by the PTW parameter  $\kappa$ . The PTW parameter  $\gamma$  and the damage model parameter pspall are also influential in both analyses. The EOS perturbation  $\epsilon$  is the only important input that contributes significantly to explaining total variability in every output analyzed. Figure 14 shows 90% probability intervals for the predictive distributions of the main effect sensitivity indices corresponding to the three sets of principal component coefficients. The point estimates of these indices are given in the right panel of Figure 13.

The first principal component explains 80.56% of the total dispersion in the mean-centered simulation runs. The left panel of Figure 15 shows the first component primarily affects the times at which initial loading (rise to peak velocity), initial unloading (release) and pull-back occur. This timing is determined by the speed of the shock wave traveling through the material, which is a function of the tantalum EOS and material strength. The right panel of Figure 15 shows the EOS perturbation  $\epsilon$  drives first order variation in the sign of the coefficient of this component, while the other parameters have negligible first order effects. As  $\epsilon$  traverses its range from negative values to positive, this coefficient linearly decreases from positive to negative values, changing sign at  $\epsilon = 0$ . This behavior translates to initial loading, unloading and damage occurring earlier as  $\epsilon$  increases, which is consistent with the fact that the sound speed is an increasing function of  $\epsilon$ . The main effect functions for all inputs are estimated using the algorithm of Section 2.3, based on a 625-run, five-level, strength 3 orthogonal array.

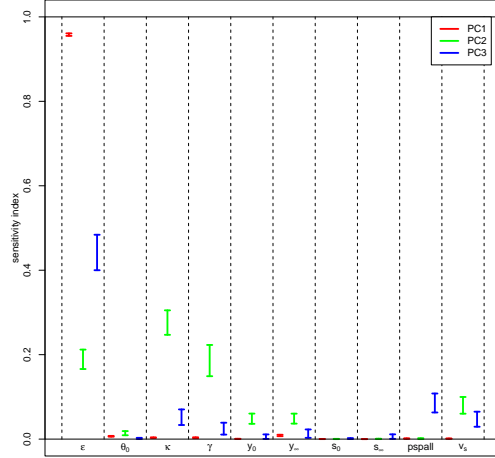


Figure 14: Uncertainties in the first order sensitivity indices corresponding to coefficients of the first (red), second (green) and third (blue) principal components for each of the ten calibration parameters. The 5th and 95th quantiles of the predictive distributions are indicated by horizontal bars.

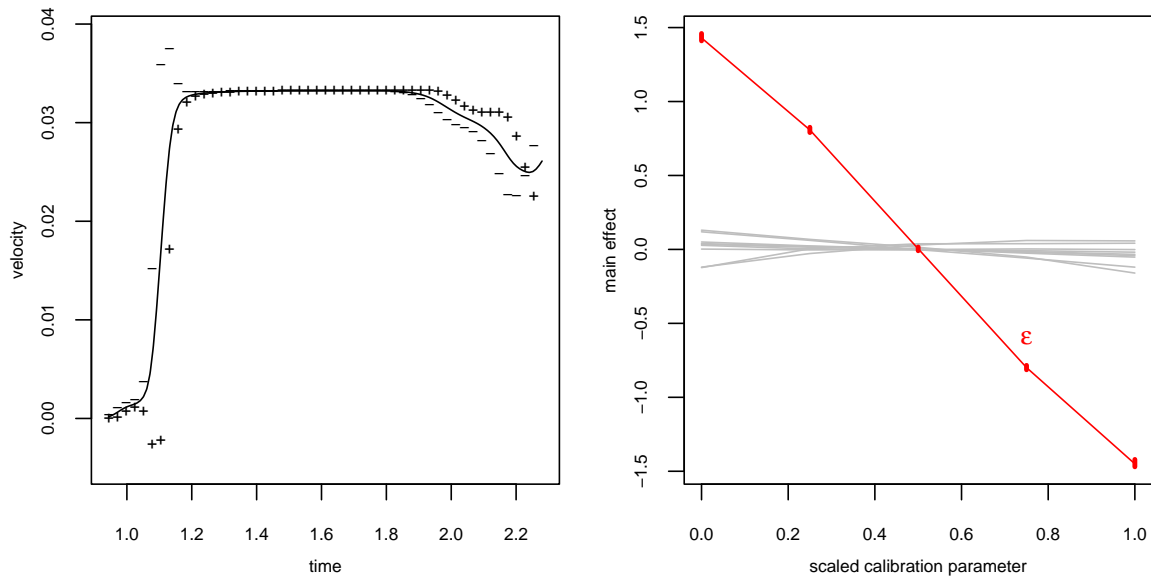


Figure 15: Mean of simulation runs and first principal component added with positive (+) and negative (-) coefficients (left panel), and 1-d marginalizations of the posterior mean fit to the observed coefficients of the first component with the most active input  $\epsilon$  in red and all others in grey (right panel). Vertical bars provide 95% probability regions for the main effect of  $\epsilon$  at the five input values it was estimated.

The second principal component picks up an additional 9.58% of the total dispersion in the mean-centered simulation runs. The left panel of Figure 16 shows the second component makes adjustments to the free surface velocities across the entire velocity profile. This behavior is determined primarily by material strength. Elastic waves travel faster in stiffer materials, resulting in higher free surface velocities at times in the vicinity of the HEL. The elastic waves reflect off the leading edge of the target material and reduce the amplitude of the approaching plastic wave. Greater amplitude reduction is achieved by faster elastic waves, resulting in lower free surface ve-

locities at times subsequent to the HEL. A positive coefficient on this component is consistent with greater material strength relative to a negative coefficient. The right panel of Figure 16 shows the main effect functions for  $\kappa$ ,  $\gamma$ ,  $v_s$  and  $\epsilon$ . Smaller values of  $\kappa$  and  $\gamma$  are associated with increasing coefficient values, consistent with the fact that material strength increases as  $\kappa$  and  $\gamma$  decrease in the PTW model.

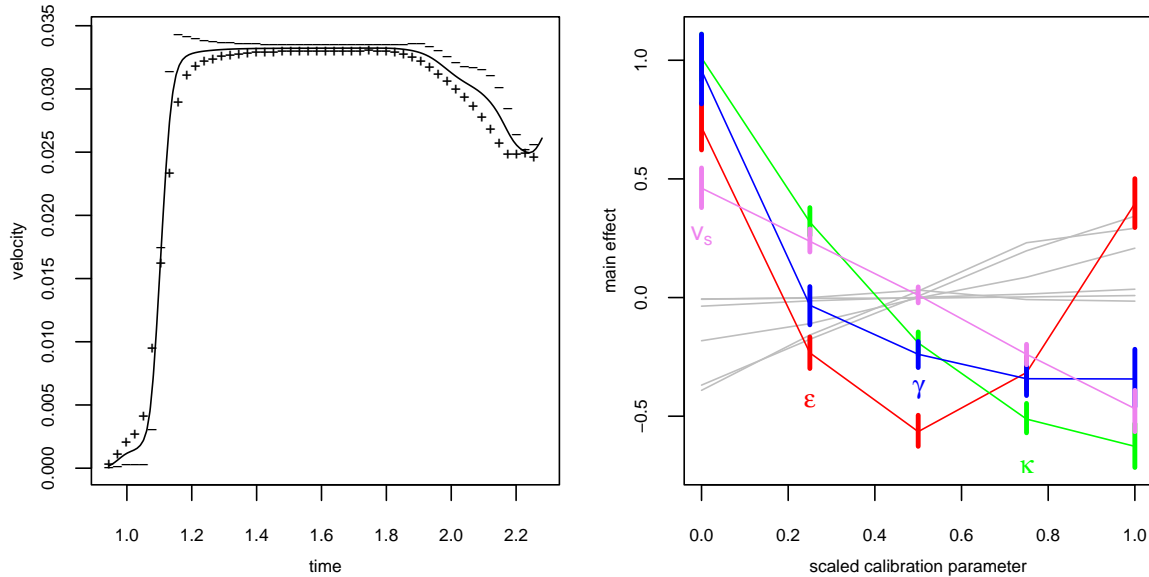


Figure 16: Mean of simulation runs and second principal component added with positive (+) and negative (−) coefficients (left panel), and 1-d marginalizations of the posterior mean fit to the observed coefficients of the second component with the most active inputs ( $\epsilon$ ,  $\kappa$ ,  $\gamma$  and  $v_s$ ) in color and all others in grey (right panel). Vertical bars provide 95% probability regions for the main effects of the active inputs at the five input values they were estimated.

The third principal component picks up an additional 5.71% of the total dispersion in the mean-centered simulation runs. This component makes adjustments to the free surface velocities primarily in the region of the release phase near the damage point. The EOS parameter  $\epsilon$  influences the shape of the velocity profile in the release phase. The pspall parameter specifies an amount of tension the tantalum can withstand before spalling; it is expressed as a negative pressure so smaller pspall values correspond to a higher tension threshold. The difference between peak velocity and velocity at the damage point is an increasing function of this tension threshold. The left panel of Figure 17 shows a positive coefficient corresponds to damage occurring later at lower free surface velocity relative to a negative coefficient. The right panel of Figure 17 shows the main effect functions for  $\epsilon$  and pspall. The main effect for  $\epsilon$  first increases then decreases. The main effect function for pspall decreases monotonically. Therefore, smaller values of pspall are associated with damage occurring later at lower free surface velocity, consistent with a higher tension threshold.

## 4 Discussion

We have demonstrated a Bayesian approach for fusing model simulations and observed field data to carry out model sensitivity, calibration and prediction on a flyer plate experiment. Uncertainties arising from unknown calibration parameters, limited simulation runs, and discrepancy between

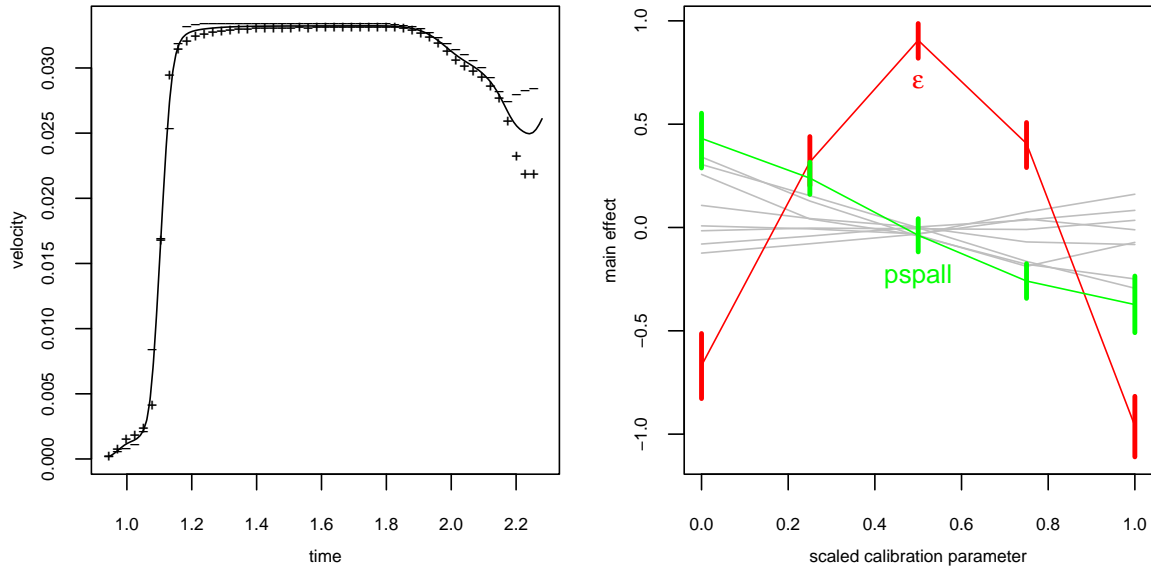


Figure 17: Mean of simulation runs and third principal component added with positive (+) and negative (−) coefficients (left panel), and 1-d marginalizations of the posterior mean fit to the observed coefficients of the third component with the most active inputs ( $\epsilon$  and  $\text{pspall}$ ) in color and all others in grey (right panel). Vertical bars provide 95% probability regions for the main effects of the active inputs at the five input values they were estimated.

simulator and reality are incorporated here. Sensitivity analysis of the functional flyer plate simulation output highlighted the insights gained by identifying physical interpretations of the principal component representation. The availability of such interpretations provides the necessary context for understanding the relative importance and behavior of the uncertain model parameters. Calibration of simulation model parameters to a flyer plate experiment highlighted the extent to which uncertain material properties can be constrained, including the identification of trade-offs among parameters that cannot be resolved by available experimental data. The flyer plate simulation model provides a compelling example of the need to account for uncertainty due to simulation model inadequacies. Including a discrepancy component in the statistical model focuses the calibration effort on physics in the simulation model that provides a sufficiently complete picture of the physical processes generating the experimental data.

This work is a piece of the overall verification and validation (V&V) process for simulation models (see Oberkamp *et al.* (2003), for example), however we focus on the question of whether or not the simulation model is useful for prediction. Given our focus on prediction, the usefulness of the simulator is apparent from the reduction in prediction uncertainty it affords when compared to a purely field data driven analysis. Another important issue is whether or not including simulator runs in the analysis improves our ability to extrapolate.

Reality checks are very important in analyses such as these that build much structure into models – simulators as well as statistical models – and use comparatively small amounts of field data. Comparing predictions to hold out data which have not been used to estimate model parameters can be very useful. In cases for which field data is too scant for such an approach, one needs to be more careful (see Bayarri and Berger (2000), Robbins *et al.* (2000) and accompanying discussion). Also, the investigation of the sensitivity of predictions to changes in modeling assumptions can be

revealing.

This basic framework will likely have to be adjusted to deal with highly multivariate output. We were able to describe the multivariate output of the flyer plate traces under this framework, but to incorporate a very dense trace will require reformulating the statistical model to make the computations tractable. Approaches based on wavelets or functional data analysis techniques seem promising here (Higdon *et al.* (2005), for example).

The statistical modeling of the simulator function  $\eta(\cdot, \cdot)$  becomes increasingly difficult as the dimensionality of the input space increases since the limited number of runs must now cover a high dimensional space. Adaptive design as well as dimension reduction strategies have been proposed (Craig *et al.*, 2001) for dealing with this issue. As an alternative, it may be wise to develop a faster, more approximate simulator and use a strategy that avoids modeling  $\eta(\cdot, \cdot)$ . Another promising approach is to replace the simulator altogether with a statistical model that incorporates key features of the physical process as in Wikle *et al.* (2001).

## Acknowledgements

The authors thank Jonathan Boettger, Kevin Honnell and Charles Nakhleh for helpful discussions on the physics of flyer plate experiments, and Margaret Short for suggesting a computational technique that allowed for large quantities of calculations and data in the flyer plate analysis. The editor and two referees suggested revisions that improved the presentation of this paper.

## References

- Bayarri, M. J. and Berger, J. O. (2000). *P* values for composite null models, *Journal of the American Statistical Association* **95**(452): 1127–1142.
- Craig, P. S., Goldstein, M., Rougier, J. C. and Seheult, A. H. (2001). Bayesian forecasting using large computer models, *Journal of the American Statistical Association* **96**: 717–729.
- Fugate, M., Williams, B., Higdon, D., Hanson, K. M., Gattiker, J., Chen, S.-R. and Unal, C. (2005). Hierarchical Bayesian analysis and the Preston-Tonks-Wallace model, *Technical Report LA-UR-05-3935*, Los Alamos National Laboratory.
- Hanson, K. M. (2004). Inference about the plastic behavior of materials from experimental data, *Technical Report LA-UR-04-3529*, Los Alamos National Laboratory.
- Higdon, D., Gattiker, J. and Williams, B. (2005). Computer model calibration using high dimensional output, *Technical Report LA-UR-05-6410*, Los Alamos National Laboratory.
- Higdon, D., Kennedy, M., Cavendish, J., Cafo, 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.
- Linkletter, C., Bingham, D., Sitter, R., Ye, K., Hengartner, N. and Higdon, D. (2003). Comparing designs of experiments for screening and prediction in computer experiments, *Technical Report LA-UR-03-8524*, Los Alamos National Laboratory.
- McKay, M. D. (1995). Evaluating prediction uncertainty, *Technical Report LA-12915-MS*, Los Alamos National Laboratory.
- McKay, M. D., Campbell, K. S. and Williams, B. J. (2005). Sensitivity analysis when model outputs are functions, in K. M. Hanson and F. M. Hemez (eds), *Sensitivity Analysis of Model Output*, Los Alamos National Laboratory, Los Alamos, 2005 (<http://library.lanl.gov/ccw/samo2004/>), pp. 81–89.
- 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.
- Meyers, M. A. (1994). *Dynamic Behavior of Materials*, J. Wiley & Sons, New York.
- Moore, L. M. and McKay, M. D. (2002). Orthogonal arrays for computer experiments to assess important inputs, in D. W. Scott (ed.), *Proceedings of PSAM6, 6th International Conference on Probabilistic Safety Assessment and Management*, Vol. LA-UR-02-1434, pp. 546–551.
- Moore, L., McKay, M. D. and Campbell, K. (2004). Combined array experiment design, *Technical Report LA-UR-04-0374*, Los Alamos National Laboratory.
- Oakley, J. E. and O’Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach, *Journal of the Royal Statistical Society (Series B)* **66**: 751–769.

- Oberkampf, W. L., Trucano, T. G. and Hirsch, C. (2003). Verification, validation, and predictive capability in engineering and physics, *SAND2003-3796*, Sandia National Laboratories.
- O'Hagan, A. (1978). Curve fitting and optimal design for prediction, *Journal of the Royal Statistical Society (Series B)* **40**: 1–24.
- Owen, A. B. (1994). Controlling correlations in Latin hypercube samples, *Journal of the American Statistical Association* **89**: 1517–1522.
- Preston, D. L., Tonks, D. L. and Wallace, D. C. (2003). Model of plastic deformation for extreme loading conditions, *Journal of Applied Physics* **93**: 211–220.
- Robins, J. M., van der Vaart, A. and Ventura, V. (2000). Asymptotic distribution of  $P$  values in composite null models, *Journal of the American Statistical Association* **95**(452): 1143–1156.
- 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.
- Saltelli, A., Chan, K. and Scott, E. M. (2000). *Sensitivity Analysis*, John Wiley & Sons.
- Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *Design and Analysis of Computer Experiments*, Springer, New York.
- Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models, *Mathematical Model. Comput. Exp.* **1**: 407–414.
- Tang, B. (1993). Orthogonal array-based Latin hypercubes, *Journal of the American Statistical Association* **88**: 1392–1397.
- Trunin, R. F. (1998). *Shock Compression of Condensed Materials*, Cambridge University Press, Cambridge, UK.
- Welch, W. J. (1985). ACED: Algorithms for the construction of experimental designs, *The American Statistician* **39**: 146.
- Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992). Screening, predicting, and computer experiments, *Technometrics* **34**: 15–25.
- Wikle, C. K., Milliff, R. F., Nychka, D. and Berliner, L. M. (2001). Spatio-temporal hierarchical Bayesian modeling: Tropical ocean surface winds, *Journal of the American Statistical Association* **96**: 382–397.
- Xu, H. (2002). An algorithm for constructing orthogonal and near orthogonal arrays with mixed levels and small runs, *Technometrics* **44**: 356–368.