SIAM J. SCI. COMPUT.
Vol. 26, No. 2, pp. 448–466

COMBINING FIELD DATA AND COMPUTER SIMULATIONS FOR CALIBRATION AND PREDICTION*

DAVE HIGDON†, MARC KENNEDY‡, JAMES C. CAVENDISH§, JOHN A. CAFEO¶, AND ROBERT D. RYNE‖

Abstract. We develop a statistical approach for characterizing uncertainty in predictions that are made with the aid of a computer simulation model. Typically, the computer simulation code models a physical system and requires a set of inputs—some known and specified, others unknown. A limited amount of field data from the true physical system is available to inform us about the unknown inputs and also to inform us about the uncertainty that is associated with a simulation-based prediction. The approach given here allows for the following:

- 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;
- uncertainty in the observation process that yields the actual field data on the true physical system.

The resulting analysis yields predictions and their associated uncertainties while accounting for multiple sources of uncertainty. We use a Bayesian formulation and rely on Gaussian process models to model unknown functions of the model inputs. The estimation is carried out using a Markov chain Monte Carlo method. This methodology is applied to two examples: a charged particle accelerator and a spot welding process.

Key words. calibration, computer experiments, predictability, uncertainty quantification, Gaussian process, model validation, simulator science

AMS subject classifications. 60G15, 62F15, 62M30, 62P30, 62P35

DOI. 10.1137/S1064827503426693

1. Introduction.

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

*Received by the editors April 30, 2003; accepted for publication (in revised form) October 10, 2003; published electronically December 22, 2004. This research was supported in part by the National Science Foundation, National Institute of Statistical Sciences, General Motors, Office of Science, and Los Alamos National Laboratory LDRD. A portion of this research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC03-76SF00098. This work was performed in part under the auspices of the Los Alamos National Laboratory, an affirmative action/equal opportunity employer, operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.
http://www.siam.org/journals/sisc/26-2/42669.html
†Los Alamos National Laboratory, Statistical Sciences Group, MS F600, Los Alamos, NM 87545 (dhigdon@lanl.gov).

‡Department of Probability and Statistics, University of Sheffield, Hicks Building, Sheffield, UK (m.kennedy@sheffield.ac.uk).

§Manufacturing Systems Research Lab, General Motors R&D Center, Warren, MI 48090 (james.c.cavendish@gm.com).

¶Vehicle Development Research Lab, General Motors R&D Center, Warren, MI 48090 (john.cafeo@gm.com).

‖Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mail Stop 71J100A, Berkeley, CA 94720-0103 (rdryne@lbl.gov).

principles. In such problems, the actual amount of observed field data from this process is typically very limited. It is the simulator code that contains the structure of the actual process it's 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 provide information on unknown, and often unconsidered, structure.

The first computer experiment ever conducted appears to have been carried out by Enrico Fermi and colleagues (Strogatz (2003)) in Los Alamos back in 1953. The experiment simulated vibrations over 64 lattice points on the MANIAC computer, the top computer of its day. The application of statistical methods to analyses involving computer simulations began decades later.

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: determining input settings at which to carry out a sequence of simulation designs (McKay, Beckman, and Conover (1979); Johnson, Moore, and Ylvisaker (1990); Mitchell, Moore, and Ylvisaker (1995)).
- Interpolation/emulation: given simulation output at a set of input settings, estimating the simulation output at a new, untried input setting (Sacks et al. (1989); Currin et al. (1991); Kennedy and O'Hagan (2001)).
- Uncertainty and sensitivity analysis: determining the variation in simulation output due to uncertainty or changes in the input settings (Helton (1997); Saltelli, Chan, and Scott (2000); Oakley and O'Hagan (2002)).
- Calibration: using field observations and simulation runs to estimate simulation model parameters or to update the uncertainty regarding these parameters (Kennedy and O'Hagan (2001); Higdon, Lee, and Holloman (2003)).
- Prediction: using the calibrated simulator to give predictions (with uncertainty bounds) of the actual physical system (Poole and Raftery (2000); Hegstad and Omre (2001); Kennedy and O'Hagan (2001); Craig et al. (2001); Bayarri et al. (2002)).

In this paper we outline a general statistical approach for combining scant field observations with simulator runs to calibrate parameters in the simulator and to characterize uncertainty in simulator-based predictions. We take a Bayesian approach that closely follows that of Kennedy and O'Hagan (2001). This approach explicitly models uncertainty in model inputs, uncertainty due to limited numbers of simulation runs, and discrepancy between the simulator and the actual physical system. In the following section, we outline our approach in an incremental fashion, making use of a simple application. We then consider two applications: a charged particle accelerator (Qiang et al. (2000)) and a spot welding experiment (Wang and Hayden (1999)).

Our formulation utilizes standard statistical models, including Gaussian process models (Sacks et al. (1989); Stein (1999)). We use these models here in this paper because of their convenience, flexibility, and fairly broad generality. However, we realize that other models may prove more appropriate depending on the application at hand. It is the general modeling strategy for carrying out inference in applications involving simulator code that we wish to emphasize here. The resulting posterior distributions are then explored using a Markov chain Monte Carlo (MCMC) method (Besag et al. (1995); Gilks, Richardson, and Spiegelhalter (1998); Robert and Casella (1999)).

2. Statistical formulation. We take $\eta(x,t)$ to denote simulator output given input vector (x,t), where x holds observable, and often controllable, inputs and t holds

additional unobservable calibration and tuning parameters which are required to run the code. Though there is clearly a distinction between calibration parameters which have physically interpretable meaning and tuning parameters which may be notional and of little or no meaning in the physical system, we will take the term calibration parameters to include both types of parameters throughout this paper. We certainly admit there will be applications where this distinction will make it necessary to treat calibration and tuning parameters differently. Examples of this partitioning of the simulator inputs (x, t) are given in the applications of sections 3 and 4. Ideally, for an appropriate choice of $t = \theta$, $\eta(x, \theta)$ simulates a physical system $\zeta(x)$; note the actual physical system $\zeta(x)$ does not depend on θ.

At various settings for x, observations y are made of the physical system

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

where the $\epsilon(x_i)$'s denote observation error. 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 $y = (y(x_1), \ldots, y(x_n))^T$ to denote the physical observations. Often very 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 true calibration value θ according to

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

where the stochastic term $\delta(x_i)$ accounts for discrepancy between the simulator $\eta(x_i, \theta)$ and reality $\zeta(x_i)$, and θ denotes the "true," but unknown, setting for the calibration inputs t. In some cases, the discrepancy term can be ignored; in other cases, it plays a crucial role in the analysis.

Depending on the application, one may also wish to treat a fixed set of m simulation runs

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

as data to be used in the analysis. This is typically the case when 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 statistical model for $\eta(x, t)$ will be required for input combinations (x, t) for which the simulator has not been run. This will be discussed in more detail in section 2.2 below. Note that we use t to denote an input setting for the calibration parameters here. We reserve θ to denote the "best" or "true" value of the calibration parameters, which is a quantity about which we wish to infer. The following subsections will step through increasingly more complicated statistical analyses based on simulation runs and observed field data for which the details of the statistical modeling and estimation will be described.

2.1. Unlimited simulation runs. We now consider a notional example for which both x and t are one-dimensional and $\eta(x, t)$ simulates a physical system $\zeta(x)$ when t is set to the true, but unknown, value of $t = \theta$. We obtain $n = 3$ field observations at three different values of x as shown in the first plot of Figure 1. We assume field observations are noisy measurements of the true system $\zeta(x)$, and this observation noise is known to be normal, with a standard deviation of 0.25.

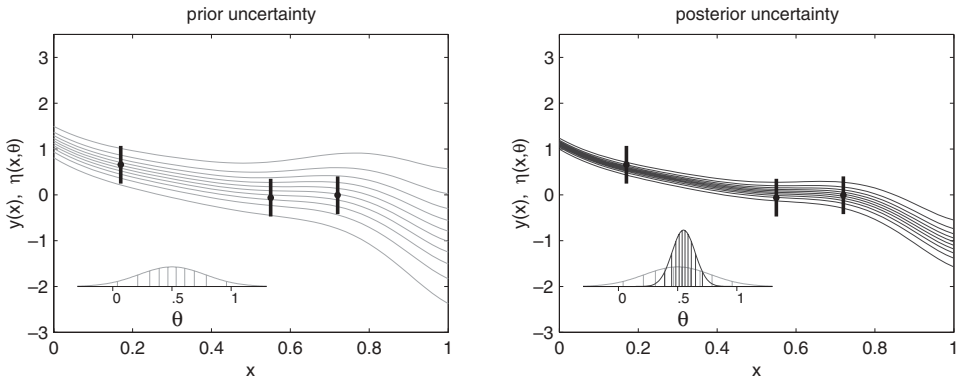

FIG. 1. *Prior and posterior uncertainty in model predictions. Left: Prior density for θ (light lines) and the implied simulations $\eta(x,\theta)$ using quantiles of the prior distribution. The black dots show field observations and the corresponding black lines give 95% uncertainty bounds on the observations. Right: The darker lines show the updated posterior density for θ and the resulting predictions are shown for quantiles of the posterior distribution for θ. Lines correspond to $5th, 15th, \ldots, 95th$ percentiles.*

For now we also assume that the simulator sufficiently represents the physical system so that the model

$$y(x_i) = \eta(x_i, \theta) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

is appropriate. Because the distribution of the ϵ_i's is assumed to be independently and identically distributed $N(0, .25^2)$, the sampling model for y is then

$$L(y|\eta(\theta)) \propto \exp\left\{-\frac{1}{2}(y-\eta(\theta))^T \Sigma_y^{-1}(y-\eta(\theta))\right\},$$

where $y = (y(x_1), \ldots, y(x_n))^T$, $\eta(\theta) = (\eta(x_1,\theta), \ldots, \eta(x_n,\theta))^T$, and the observation covariance matrix $\Sigma_y = I_n.25^2$.

The Bayesian formulation is completed by specifying a prior distribution $\pi(\theta)$ for the unknown calibration parameter θ. For this example we specify that θ is a priori $N(.5, .25^2)$, as shown in the first plot of Figure 1. This prior uncertainty about θ induces prior uncertainty about the simulator output $\eta(x,\theta)$. The lines in the left-hand plot of Figure 1 show simulations of $\eta(x,\theta)$ for $x \in [0,1]$ and the prior quantiles for θ.

The resulting posterior distribution for θ is then given by

$$\begin{aligned}\pi(\theta|y) &\propto L(y|\eta(\theta)) \times \pi(\theta) \\ &\propto \exp\left\{-\frac{1}{2}(y-\eta(\theta))^T \Sigma_y^{-1}(y-\eta(\theta)) - \frac{1}{2(.25^2)}(\theta-.5)^2\right\}.\end{aligned} \tag{2}$$

If $\eta(\theta)$ were a linear map, then $\pi(\theta|y)$ could be obtained analytically. However, since the simulator is typically nonlinear, (2) usually results in an intractable expression for the posterior. In this case, a fairly general purpose approach is to generate a sequence of realizations $\theta^1, \ldots, \theta^N$ from $\pi(\theta|y)$. This could be carried out using importance sampling (Evans and Swartz (1995)) or an MCMC method (Besag et al. (1995); Gilks, Richardson, and Spiegelhalter (1998)), which constructs a Markov chain whose stationary distribution is the posterior distribution.

A very simple MCMC implementation uses the Metropolis algorithm (Metropolis et al., 1953) and goes as follows:

1. Initialize θ^1 at some value.
2. Given the current realization is θ^t, generate θ^* from a symmetric distribution (i.e., the chance of generating θ^* given θ^t is the same as generating θ^t given θ^*).
3. Compute the Metropolis acceptance probability
$$\alpha = \min\left\{1, \frac{\pi(\theta^*|y)}{\pi(\theta^t|y)}\right\}.$$
4. Set
$$\theta^{t+1} = \begin{cases} \theta^* & \text{with probability } \alpha, \\ \theta^t & \text{with probability } 1-\alpha. \end{cases}$$
5. Iterate over steps 2–4.

This simple but general recipe has a number of features that are worth mentioning. This approach is applicable even when θ is a high dimensional vector. In this case components of θ can be updated individually or in larger groupings. The above MCMC references give details regarding such implementations. In step 3 above, one only needs to compute ratios of the posterior density. Hence $\pi(\theta|y)$ need only be specified up to a constant of integration. Often, thousands, or tens of thousands, of MCMC steps are required to sufficiently sample the posterior distribution. Again, the previously mentioned references give guidance on selecting the number of MCMC steps which will be application specific. Since the simulator must be run to compute the acceptance probability in step 3 above, this recipe above is only practicable if $\eta(x, \theta)$ can be computed very quickly.

This direct MCMC-based approach has a number of advantages that make it worth using if it is at all possible. It can readily handle a very large dimensional θ—see Hegstad and Omre (2001) or Lee et al. (2002), for example. It can also deal with large numbers of nuisance parameters as well as highly multivariate output. If a prediction problem has these features, it may well be worth reducing the fidelity in the simulator in return for gains is simulation speed, the point being that a thorough exploration of an approximate posterior (which uses a low-fidelity simulator) may give more information than a very limited exploration of a more accurate posterior (which uses a high-fidelity simulator). In fact, the linear accelerator application of section 3 makes use of the former approach.

The second plot in Figure 1 shows posterior distribution for θ which was computed from 5000 MCMC samples using the Metropolis scheme outlined above. Conditioning on the three noisy field observations leads to reduced uncertainty about θ. The induced posterior distribution for $\eta(x, \theta)$ also shown in the figure is computed from the MCMC realizations from $\pi(\theta|y)$.

2.2. Limited simulation runs. Quite often the computational demands of the simulator make it impossible to use an estimation approach—such as the one outlined above—that requires vast numbers of simulation runs. In this case, one is limited to a fixed number of simulation runs,

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

The actual choice of which inputs $(x_j^*, t_j^*),\ j = 1, \ldots, m$, at which to carry out these

simulation is an important question, but is not the focus of this paper. We point the interested reader to the design references at the beginning of this paper.

A statistical analysis requires that we treat $\eta(x,t)$ as unknown for pairs (x,t) that are not included in the original set of m simulator runs. If x is a vector in R^p and t a vector in R^ℓ, then the function $\eta(\cdot,\cdot)$ maps $R^{p+\ell}$ to R. A standard prior model for an unknown function is a Gaussian process (GP) (O'Hagan (1978); Sacks et al. (1989)). A mean function $\mu(x,t)$ and covariance function $\mathrm{Cov}((x,t),(x',t'))$ 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 typically take $\mu(x,t)$ to be a constant and specify a product correlation of the form

$$\mathrm{Cov}((x,t),(x',t')) = \frac{1}{\lambda_\eta}\exp\left\{-\sum_{k=1}^{p}\beta_k^\eta|x_{ik}-x'_{ik}|^\alpha - \sum_{k'=1}^{\ell}\beta_{p+k'}^\eta|t_{ik'}-t'_{ik'}|^\alpha\right\}, \tag{3}$$

where the parameter λ_η controls the reciprocal of the marginal variance of $\eta(\cdot,\cdot)$, the $p+\ell$-vector β^η controls the dependence strength in each of the component directions of x and t, and α controls the smoothness of $\eta(\cdot,\cdot)$. A value of $\alpha = 2$ leads to a smooth, infinitely differentiable representation for $\eta(\cdot,\cdot)$, while taking smaller values of α gives rougher representations. The simulator output $\eta(x_j^*,t_j^*)$, $j = 1,\ldots,m$, does not typically give any information about α. Hence we typically fix α based on prior information regarding $\eta(\cdot,\cdot)$ or based on computational considerations. We note that it is often useful to add a white noise component to the covariance model (3) for numerical stability and/or to account for small numerical fluctuations in the simulation. For models with random outputs, such as epidemiological or agent-based models, an additional independent error term will be required in (3) above.

As before, we assume model (1) is appropriate and define the field observation vector $y = (y(x_1),\ldots,y(x_n))^T$. To hold the simulation outcomes we define $\eta = (\eta(x_1^*,t_1^*),\ldots,\eta(x_m^*,t_m^*))^T$. Now we define the joint $n+m$-vector $z = (y^T,\eta^T)^T$ which has associated simulation input values $(x_1,\theta),\ldots,(x_n,\theta)$ for its first n components and $(x_1^*,t_1^*),\ldots,(x_m^*,t_m^*)$ for its final m components. The sampling model, or likelihood, for the observed data z is then

$$L(z|\theta,\mu,\lambda_\eta,\beta^\eta,\Sigma_y) \propto |\Sigma_z|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(z-\mu\mathbf{1}_{n+m})^T\Sigma_z^{-1}(z-\mu\mathbf{1}_{n+m})\right\}, \tag{4}$$

where $\mathbf{1}_{n+m}$ is the $n+m$-vector of 1s,

$$\Sigma_z = \Sigma_\eta + \begin{pmatrix}\Sigma_y & 0\\ 0 & 0\end{pmatrix},$$

Σ_y is the $n\times n$ observation covariance matrix, and elements of Σ_η are obtained by applying (3) to each pair of the $n+m$ simulation input points above.

The first plot of Figure 2 shows the same $n=3$ field observations as before. But now we have only $m=20$ simulation runs carried out at the (x,t) pairs, marked by the circle plotting symbols in Figure 2. In this example we standardize the input points so that they are contained in $[0,1]^{p+\ell}$. Also, d is transformed so that the vector η has mean 0 and variance 1. This facilitates the prior specification for the parameters governing the GP model for $\eta(\cdot,\cdot)$. Independent prior distributions are then specified

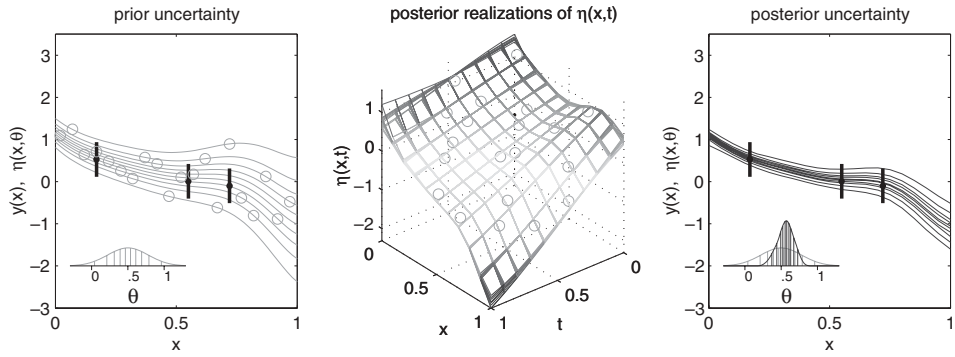

FIG. 2. *Prior and posterior uncertainty in model predictions when the number of model runs is limited. Left: Model runs have been carried out only at the* $m = 20$ (x^*, t^*) *pairs shown by the circle plotting symbols. As with Figure* 1*, the prior density for* θ *and the implied simulations* $\eta(x, \theta)$ *using quantiles of the prior distribution are given by the light lines. The black dots show field observations, and the corresponding black lines give* 95% *uncertainty bounds on the observations. Inference is carried out by conditioning on the* $n = 3$ *data points as well as the* $m = 20$ *model runs. Center: Selected realizations from the posterior distribution of* $\eta(\cdot, \cdot)$*; the actual model runs are marked by the circle plotting symbols. Right: The darker lines show the updated posterior density for* θ*, and the resulting predictions are shown for quantiles of the posterior distribution for* θ*. Uncertainty in the model prediction is due to the spread in the posterior distribution for* θ *as well as uncertainty in* $\eta(x, \theta)$*.*

for each of the parameters for $\eta(\cdot, \cdot)$:

$$\pi(\mu) \propto \exp\left\{-\frac{1}{2v}\mu^2\right\},$$
$$\pi(\lambda_\eta) \propto \lambda_\eta^{a_\eta - 1} \exp\{-b_\eta \lambda_\eta\},\ \lambda_\eta > 0,$$
$$\pi(\beta^\eta) \propto \prod_{k=1}^{p+\ell} (1 - e^{-\beta_k^\eta})^{-.5} e^{-\beta_k^\eta},\ \beta_k^\eta > 0.$$

Because of the standardization, we can simplify the parameterization and MCMC by fixing μ at 0 (i.e., $v = 0$) and encouraging λ_η to be close to 1 by taking $a_\eta = b_\eta = 5$. The prior for β^η 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 β_k^η's away from 0 in the posterior.

Conditioning on the augmented observation vector $z = (y^T, \eta^T)^T$ results in the posterior distribution

$$\pi(\theta, \mu, \lambda_\eta, \beta^\eta | z) \propto L(z|\theta, \mu, \lambda_\eta, \beta^\eta, \Sigma_y)\pi(\theta)\pi(\mu)\pi(\lambda_\eta)\pi(\beta^\eta),$$

which we explore using MCMC. Given a realization of the parameters $(\theta, \mu, \lambda_\eta, \beta^\eta)$, a posterior realization of $\eta(x, t)$ can be drawn using standard multivariate normal theory (Rencher (2002)). The second plot of Figure 2 shows posterior realizations of $\eta(\cdot, \cdot)$ over a grid on $[0, 1]^2$. Uncertainty about the function $\eta(\cdot, \cdot)$ is larger in regions where no simulation points (x^*, t^*) are nearby.

The resulting posterior inference for θ and $\eta(x, \theta)$ is summarized in the final plot of Figure 2. For this simple example, there is very little additional uncertainty that can be ascribed to the restriction to $m = 20$ simulation runs; compare Figures 1 and 2. This uncertainty due to limited evaluations of $\eta(x, t)$ plays a larger role as the

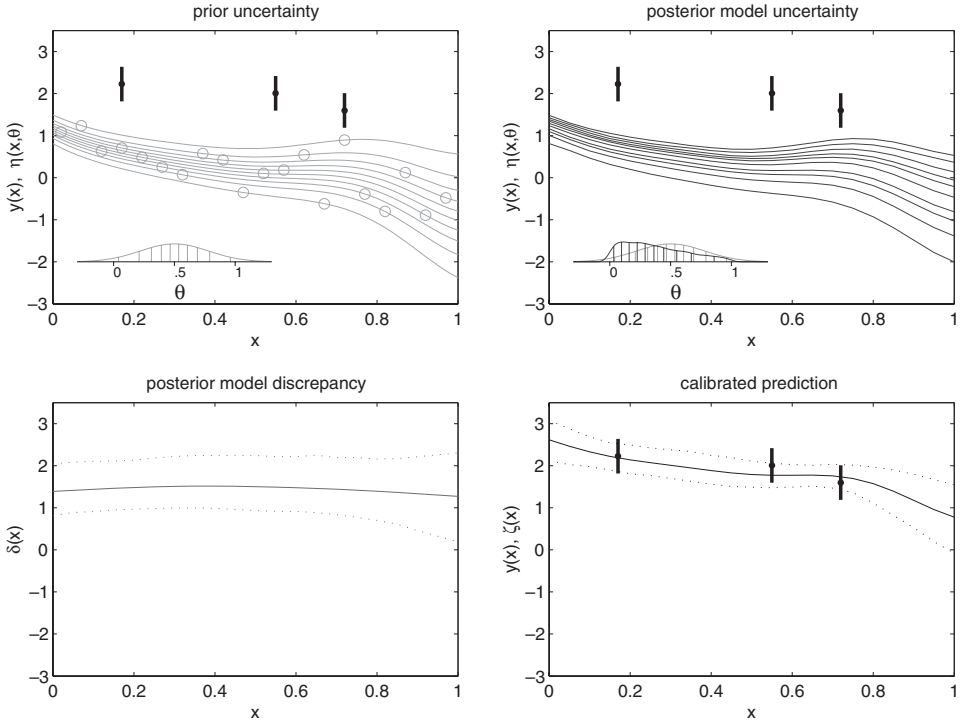

FIG. 3. *Prior uncertainty and posterior decomposition of model terms. Top left: Model runs have been carried out only at the* $m = 20$ (x, θ) *pairs shown by the circle plotting symbols. As with Figures* 1 *and* 2*, the prior density for* θ *and the implied simulations* $\eta(x, \theta)$ *using quantiles of the prior distribution are given by the light lines. The black dots show field observations, and the corresponding black lines give* 95% *uncertainty bounds on the observations. Inference is carried out by conditioning on the* $n = 3$ *data points as well as the* $m = 20$ *model runs. Top right: The darker lines show pointwise* $5th$*,* $15th, \ldots, 95th$ *percentiles for posterior density for* $\eta(x, \theta)$*. The same percentiles are also shown for the posterior distribution for* θ*. Bottom left: Pointwise* 90% *posterior credible intervals for the model discrepancy term* $\delta(x)$*. Bottom right: Pointwise* 90% *posterior credible intervals for the predictions* $\zeta(x) = \eta(x, \theta) + \delta(x)$*. Note that the inclusion of the model discrepancy term makes the posterior distribution for* θ *less precise and less interpretable.*

dimensions of x and t increase. An application one of the current authors is currently involved with has $m = 400$ and $\ell = 10$. Here, uncertainty due to the limitation on simulation runs plays a big role. For such problems we are hopeful that sensitivity analysis, combined with sequential approaches to refine the choice of simulation inputs to run, will prove effective. In any case, restrictions on the number of simulations that can be carried out will necessarily limit the number of input dimensions that can be considered.

2.3. Accounting for model discrepancy. The final alteration to our statistical formulation is motivated by the top left plot of Figure 3. Here the field observations are inconsistent with the simulations no matter what value of θ is used. This discrepancy between field observations and simulations is not uncommon in such applications—the spot weld application in section 4 is one such example. In this case, the simulator may still be of use in prediction if this discrepancy is fairly systematic so that changes in the simulation response track changes in the physical system. This is apparent in the example in Figure 3, where the simulator discrepancy varies smoothly

with input condition x.

To account for this discrepancy, we augment the model formulation with a discrepancy term $\delta(x)$ which may vary with controllable input x:

$$y(x_i) = \eta(x_i, \theta) + \delta(x_i) + \epsilon(x_i), \quad i = 1, \ldots, n. \tag{5}$$

Here $\delta(x_i)$ models the difference between the simulator and the physical system $\zeta(x_i) - \eta(x_i, \theta)$, and as before, θ denotes the true calibration value for t and $\epsilon(x_i)$ gives the observation error for the ith data point.

The modeling for the previously introduced terms remains unchanged. We specify a GP model for the discrepancy term $\delta(x)$ with mean function of 0, and a covariance function of the form

$$\mathrm{Cov}(x, x') = \frac{1}{\lambda_\delta} \exp\left\{ -\sum_{k=1}^{p} \beta_k^\delta |x_{ik} - x'_{ik}|^{\alpha_\delta} \right\}. \tag{6}$$

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

$$\pi(\lambda_\delta) \propto \lambda_\delta^{a_\delta - 1} \exp\{-b_\delta \lambda_\delta\}, \ \lambda_\delta > 0, \tag{7}$$

$$\pi(\beta^\delta) \propto \prod_{k=1}^{p} (1 - e^{-\beta_k^\delta})^{-.6} e^{-\beta_k^\delta}, \ \beta_k^\delta > 0. \tag{8}$$

Here .6 in the prior for β^δ gives $\delta(\cdot)$ a slightly stronger tendency towards flatness than $\eta(\cdot, \cdot)$.

This gives a likelihood for the joint data vector z of exactly the same form as (4), but now

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

where Σ_y is the $n \times n$ observation covariance matrix, elements of Σ_η are obtained by applying (3) to each pair of the $n+m$ simulation input points above, and Σ_δ is an $n \times n$ matrix obtained by applying (6) to each pair of the n input settings x_i, $i = 1, \ldots, n$, that correspond to the observed field data y.

The resulting posterior density now depends on $(\theta, \mu, \lambda_\eta, \beta^\eta, \lambda_\delta, \beta^\delta)$. As with $\eta(x, t)$, posterior realizations can also be produced for $\eta(x, \theta)$, $\delta(x)$, and $\zeta(x) = \eta(x, \theta) + \delta(x)$. Figure 3 shows posterior means and pointwise credible intervals for each of these quantities. In this example, posterior realizations of the discrepancy term $\delta(x)$ are quite smooth. Because of this, the field observations predominantly inform about the level of $\zeta(x)$, and the simulations predominantly inform about the shape of $\zeta(x)$. This results in fairly narrow prediction bands for the physical process, even though the magnitude of the discrepancy is relatively large. Note that this analysis gives very little information about the calibration parameter θ. Here inclusion of the discrepancy term makes it very difficult to learn anything about θ.

As regards extrapolation past the range of the data, the amount of trust one puts in the extrapolated predictions will depend on both the confidence one has in the simulator and the extendibility of the bias model in this new, untested region. Judgments regarding the appropriateness of extrapolated predictions will need to be made on a case-by-case basis. It may be that an alternative model for $\delta(x)$ is preferred when extrapolation is necessary, perhaps one that incorporates some physical principles in the system under study. See Bayarri et al. (2002) for more details regarding extrapolation.

Finally we note that our choice of an additive decomposition of the data in (5) is fairly general, but in other circumstances, some other form may be more appropriate. For example, Kennedy and O'Hagan (2001) use a slightly different form.

3. Charged particle accelerator application. Here we make use of a code (Dragt et al. (1988); Qiang et al. (2000)) that simulates the beam in a charged particle accelerator by moving anywhere from thousands to millions of individual particles along the beamline. This charged particle beam evolves with time as it is influenced by specifically generated magnetic fields along the beamline and by interactions between the particles themselves.

We consider a specific application involving a proton beam moving along a linear path that is influenced by a series of eight pairs of quadrupole magnets as shown in Figure 4. Each quadrupole magnet takes up .36 meters of length along the beamline and is separated by .688 meters of drift space. A quadrupole magnet acts on the particle beam as a lens acts on a beam of light. A focusing quadrupole causes the beam to converge in the x dimension and to diverge in the y dimension; a defocusing quadrupole causes the beam to diverge in the x dimension and to converge in the y dimension. The beamline is 3.38 meters long and is influenced by eight pairs of equally spaced quadrupole magnets. The first of each pair is a defocusing quadrupole, represented by the dark shaded region in the beamline plots of Figure 4; the second is a focusing quadrupole, represented by the light shaded region in the beamline plots of Figure 4.

Given an initial description of the particles at the beginning of the beamline, the code then steps the particles along the beamline, taking account of the effects of the quadrupole magnets and electric forces from the interacting particles. Each particle in the simulation is initially described by a six-parameter phase space $(\mathsf{x}, \mathsf{p}_\mathsf{x}, \mathsf{y}, \mathsf{p}_\mathsf{y}, \tau, \mathsf{p}_\tau)$: x position; x momentum; y position; y momentum; differential time of flight; and its conjugate momentum. The x and y beamline plots of Figure 4 show the 5th, 15th, ..., and 95th percentiles of the x and y positions of the particles as they're stepped along the beamline. This beam is the result of starting with a particular initial beam configuration. Various plots that describe this initial beam configuration are shown in the first row of Figure 4. The code also can output the full six-dimensional phase space description of each particle at various positions along the beamline. Similar plots are also shown of the phase space at various steps along the beamline.

To investigate the behavior of a beam as it moves along the beamline, wire scanners can be placed at specific locations to collect physical observations pertaining to the beam. Typically a wire scan collects counts of particle hits which are binned according to position in either the x or y directions. In the synthetic experiment described here, wire scans are carried out at the nine beamline locations given by the dashed lines in the beamline plots of Figure 4. The resulting wire scans are shown in the rows labeled "x wire scan" and "y wire scan." Each scan consists of a count at each of 256 bins that are equally spaced along a length of .01 meters, extending $\pm$.005 meters from the beam center. The observed counts corresponding to the 1st, 3rd, 5th, 7th, and 9th wire scans in each direction are given in the wire scan plots of Figure 4.

The goal of the statistical analysis is to combine these 10 wire scans—5 x wire scans and 5 y wire scans—with computer simulation runs to estimate the initial phase space configuration of the charged particle beam. In addition, we can compare the predicted wire scans at the 2nd, 4th, 6th, and 8th wire scan locations to the observed data that we held back.

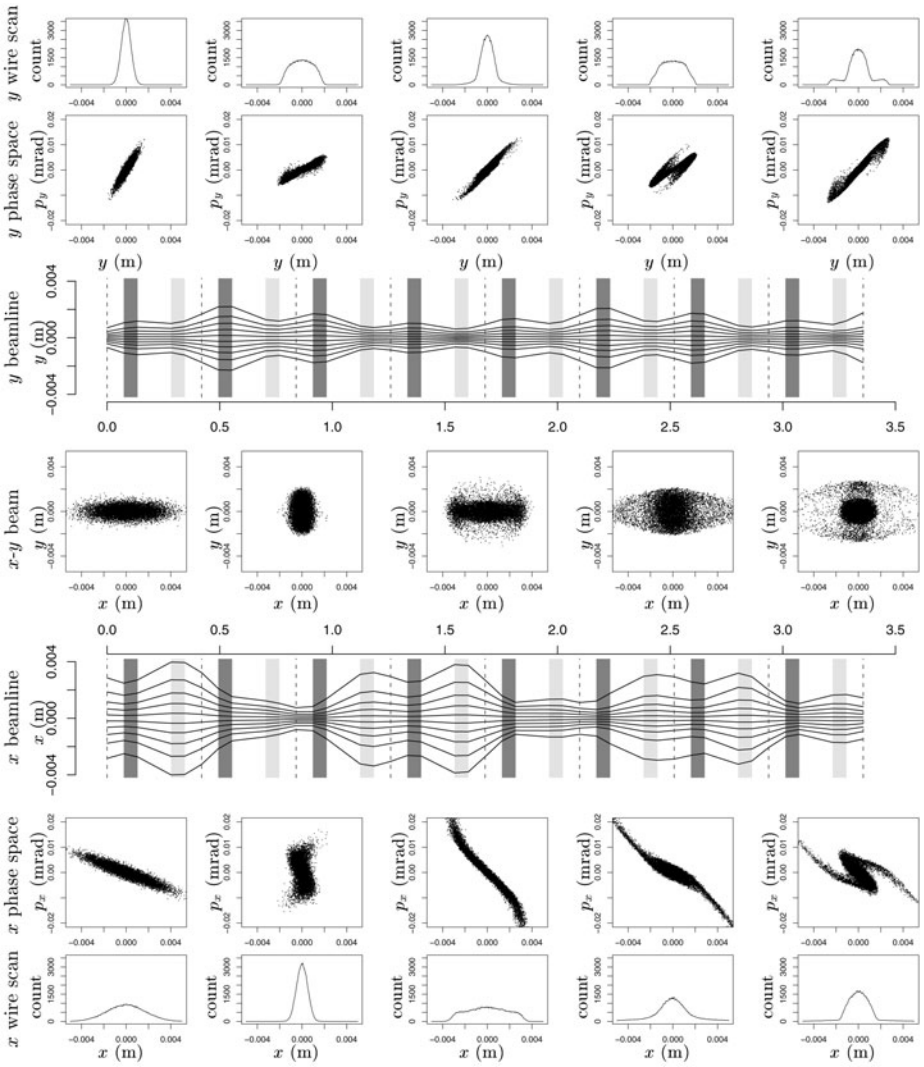

FIG. 4. *Simulation of a proton beam through a series of quadrupole magnets. The simulation involves* 100,000 *particles, each of which is described by its location and momentum in* x *and* y. *The* x *and* y *beamline figures show the progression of particles as they move along the beamline, which consists of eight pairs of quadrupole magnets denoted by the shaded regions in the beamline figures. The lines denote the* $5th, 15th, \ldots, 95th$ *percentiles of position in* x *and* y. *The dashed lines in the beamline figures denote the location of wire scanners which detect the position of the particles separately for the two dimensions* x *and* y. *The wire scan data obtained from the* 1*st,* 3*rd,* 5*th,* 7*th, and* 9*th wire scans are shown in the top and bottom rows labeled* x *and* y *wire scans. In addition, plots of phase space—*x *vs. the* x *momentum* p_x, y *vs.* p_y*—and beam width—*x *vs.* y*—are also shown at the beamline positions corresponding to the odd wire scans.*

3.1. Statistical formulation. The above description uses x and y to denote the horizontal and vertical beam dimensions. In this description of the statistical modeling, x and y in their standard font carry their original meaning: controllable input setting and observed field data.

We take x to be three-dimensional with components corresponding to the wire scan bin (1–256); the beam component dimension (horizontal x or vertical y); and the

wire scan (1–9). The calibration parameter vector θ holds the phase space description of the initial beam cloud. In this example, the beam energy does not change with time, so the τ and p_τ components of the phase space are known; it is the remaining four dimensions of the beam cloud that are parameterized by θ. In this application, it is expected that the initial x phase space components $(\mathsf{x}, \mathsf{p}_\mathsf{x})$ are essentially independent of the initial y phase space components $(\mathsf{y}, \mathsf{p}_\mathsf{y})$ so that the initial four-dimensional cloud can be represented as two independent two-dimensional clouds. We assume that the initial phase space for the two beam dimensions can be modeled as bivariate normal densities with unknown location, standard deviations, and correlation. Hence θ is a ten-dimensional vector—5 dimensions per phase space cloud.

We construct the field data y by generating x and y wire scans from a single high-fidelity, 100,000-particle simulation shown in Figure 4. The particles were simulated from independent bivariate normal distributions for the x and y phase spaces using a true θ of $(\mu_\mathsf{x}, \mu_{\mathsf{p}_\mathsf{x}}, \sigma_\mathsf{x}, \sigma_{\mathsf{p}_\mathsf{x}}, \rho_\mathsf{x}) = (0, 0, .0015, .002, -.93)$ and $(\mu_\mathsf{y}, \mu_{\mathsf{p}_\mathsf{y}}, \sigma_\mathsf{y}, \sigma_{\mathsf{p}_\mathsf{y}}, \rho_\mathsf{y}) = (0, 0, .00086, .0031, .93)$. Here μ denotes mean, σ denotes standard deviation, and ρ denotes correlation. The plots in the first column of the figure show features of this initial beam configuration.

In this analysis we do not model the simulator $\eta(x, t)$ statistically, but instead evaluate the simulator as required, as in the example of section 2.1. In order to make the simulator sufficiently fast we reduce the number of particles to 8000 and minimize the number of "slices" along the beamline where the simulator accounts for nonlinear force interactions. The reduced number of particles leads to much noisier simulated wire scans as compared to those produced by the high-fidelity simulation. To remove the effect of randomness due to different 8000 particle realizations of the initial beam configuration as prescribed by θ, an initial 8000×4 draw of standard normals is established. Now for a given θ, this established set of draws is transformed to form the initial beam. In contrast, the large number of observed counts in the field data leads to a rather small observation error. This leaves the discrepancy term $\delta(x)$ to account for the differences between the calibrated 8000 particle simulation and the observed y. We take $y(x)$ and $\eta(x, \theta)$ to be the square root of the wire scan counts to stabilize the variance and use the GP model for $\delta(x)$ with a covariance specified in (6), but modified to only allow dependence within a wire scan and to include a white noise component to account for the noise in the 8000 particle simulations. Also, since the first x and y wire scans are applied directly to the initial beam, only the white noise component of the discrepancy term is used for the first scan ($x_3 = 1$). The resulting covariance function for $\delta(x)$ is then given by

$$\mathrm{Cov}(x, x') = \frac{1}{\lambda_{\delta 1}} \exp\left\{-\beta_1^\delta |x_1 - x_1'|^{\alpha_\delta}\right\} 1_{(x_2 = x_2', x_3 = x_3' > 1)} + \frac{1}{\lambda_{\delta 2}} 1_{(x = x')}.$$

Here $1_{(\cdot)}$ is the indicator function and we set $\alpha_\delta = 2$. The resulting posterior has the form

$$\pi(\theta, \beta^\delta, \lambda_{\delta 1}\lambda_{\delta 2} | y) \propto |\Sigma_\delta|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - \eta(\theta))^T \Sigma_\delta^{-1} (y - \eta(\theta))\right\}$$
$$\times \pi(\theta) \times \pi(\beta^\delta) \times \pi(\lambda_{\delta 1}) \times \pi(\lambda_{\delta 2}),$$

where Σ_δ is constructed according to the covariance function above, θ has a vague uniform prior (i.e., $\pi(\theta) \propto 1$), $\pi(\lambda_{\delta 1})$ and $\pi(\lambda_{\delta 2})$ both follow (7), and $\pi(\beta^\delta)$ is given in (8).

The posterior is explored using MCMC, evaluating $\eta(x, \theta)$ whenever called for in the Metropolis updating scheme. Figure 5 shows a representation of the x and y initial phase space distribution—posterior realizations bivariate, two standard deviation

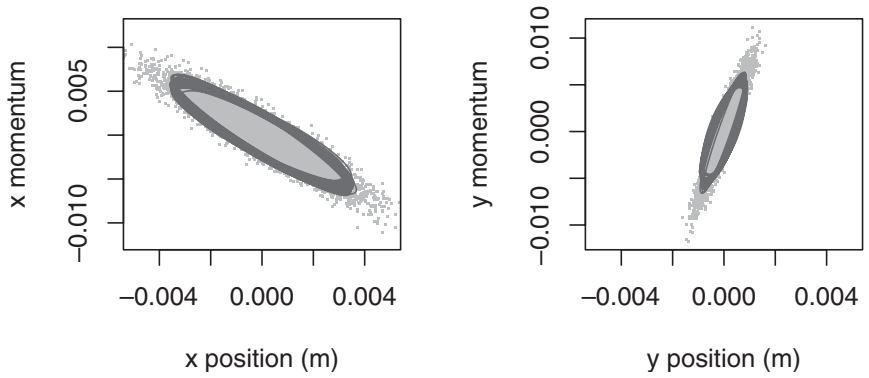

FIG. 5. *Posterior uncertainty of the initial* x *and* y *phase space configurations. The ellipses correspond to realizations of the two standard deviation contours of the bivariate normal kernels that represent the initial phase space particle density. The points are a subset of the "true"* 100,000*-particle* x *and* y *phase space.*

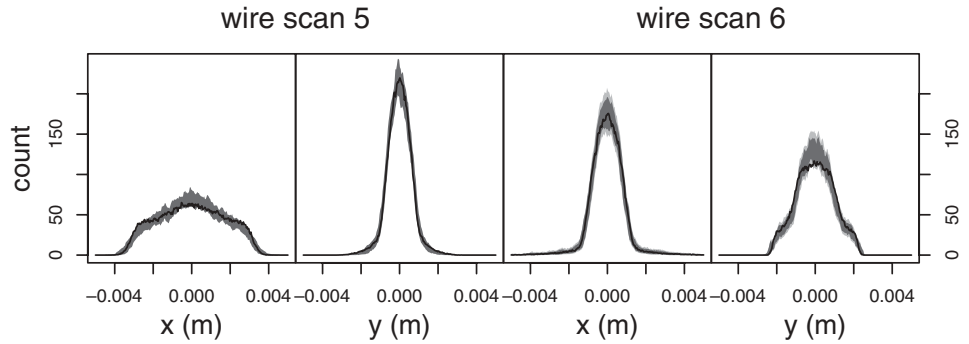

FIG. 6. *Posterior predictions for wire scans* 5 *and* 6*. The dark shaded regions show pointwise* 90% *credible intervals for the simulator* $\eta(x, \theta)$*. For wire scan* 6*, the light shaded regions show* 90% *intervals for the prediction* $\zeta(x)$*. The observed wire scans are given by the black line. The data from scan* 6*, along with the other even scans, were not used to fit the model.*

elliptical contours of the initial particle distribution in x and y phase space. The highly parametric representation of this initial phase space distribution results in a rather narrow range of uncertainty. The posterior distribution for the simulated wire scans $\eta(x, \theta)$ and predictions $\zeta(x) = \eta(x, \theta) + \delta(x)$ are shown in Figure 6 for scans 5 and 6, along with the observed scans. Scan 5 is treated as field data in the analysis. Scan 6 was held out and therefore was not used to calibrate θ or estimate the discrepancy term $\delta(x)$.

4. Spot welding application. In resistance spot welding, two metal sheets are compressed by water-cooled copper electrodes, under an applied load L. Figure 7 is a simplified representation of the spot weld process, illustrating some of the essential features for producing a weld. A direct current of magnitude C is supplied to the sheets via the two electrodes to create concentrated and localized heating at the interface where the two sheets have been pressed together by the applied load (this interface is called the faying surface). The heat produced by the current flow across the faying surface leads to melting that produces a weld nugget after cooling.

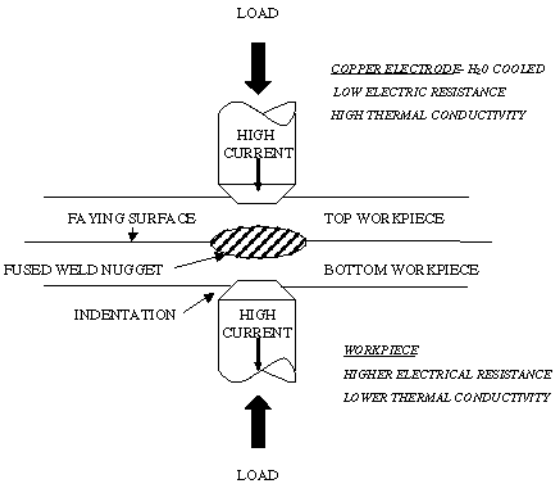

FIG. 7. *Resistance spot welding process.*

The resistance offered at the faying surface is particularly critical in determining the amount of heat generated. Because contact resistance at the faying surface, as a function of temperature, is poorly understood, a nominal function relating resistivity to temperature is specified. This function is then typically tuned to the field data. Here we will focus on a single parameter θ which governs this nominal function.

The physics of the spot weld process are modeled by a coupling of partial differential equations that govern the heat and electrical conduction with those that govern temperature dependent, elastic/plastic mechanical deformation (Wang and Hayden (1999)). Additional details regarding uncertainties in this model can be found in Bayarri et al. (2002).

We let x hold the three-dimensional inputs: load L; current C; and gauge G, the thickness of the aluminum sheets being welded together. We take the calibration parameter θ to be a univariate value that serves as a multiplicative constant in the notional function that relates contact resistance at the faying surface to temperature.

Field data consist of weld nugget diameters measured for 10 replicates taken at a 2×3 configuration of load by current settings for each of the two gauge settings. In all, this makes $n = 120$ field observations $y(x)$. The actual data measurements are apparent in the figures from this application. Simulation runs were carried out at $m = 47$ different (x, t) combinations. The input settings and simulator outcomes are apparent in Figure 8. This application fits into the specification of section 2.3. The only slight modification is that we assume $\Sigma_y = I_n/\lambda_y$ and use the data to provide information on λ_y. We use a gamma prior for λ_y, which is identical to that of λ_δ in (7).

The resulting posterior mean estimate for $\eta(\cdot, \cdot)$ is shown in Figure 8 for a lattice of load, current, and gauge values for various values of the calibration parameter t. The resulting posterior distribution for θ is fairly wide and shows bimodality, with modes at about 2.7 and 6.4. The posterior mean decomposition for model fit is given in Figure 9. It gives the posterior mean for $\eta(x, \theta)$, $\delta(x)$, and the resulting fit $\zeta(x)$. Finally, Figure 10 shows the same decomposition with 90% posterior credible intervals as a function of current for load and gauge values where the replicate field data were observed. As with the example in section 2.3, the magnitude of the discrepancy term

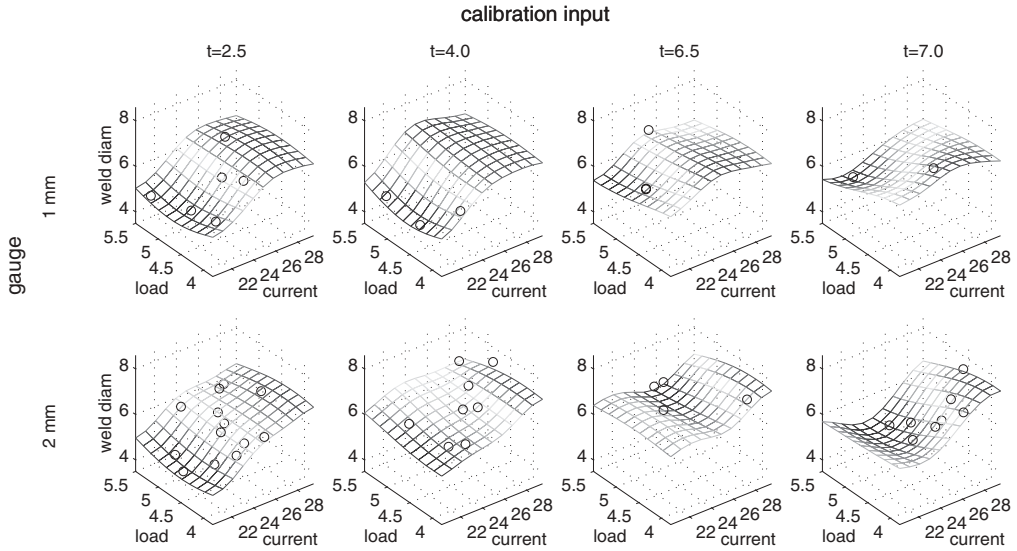

FIG. 8. *Posterior mean estimate for the simulator $\eta(x,t)$. The surfaces show weld diameter (in mm) for the two gauge values as a function of load (in kN) and current (in kA) for various values of the calibration input t. The plotting symbols denote the outcomes of the $m = 47$ simulator runs and are shown in the surface plots for which their value of t was closest.*

$\delta(x)$ is fairly large. Here, however, posterior realizations of the bias term do not stay mostly constant over x. Hence the predictions $\zeta(x)$ become uncertain for x's away from the data. From Figures 9 and 10 it is clear that the sign of the discrepancy flips for predictions of 1mm and 2mm gauge. This suggests investigating aspects of the model that deal with thickness of the workpieces being welded. This large discrepancy, combined with a wide posterior for $\eta(x,\theta)$, leads to a high degree of uncertainty regarding θ.

5. Discussion. We've demonstrated a Bayesian approach for fusing model simulations and observed field data to carry out model calibration and prediction. Uncertainties arising from unknown calibration parameters, limited simulation runs, and discrepancy between simulator and reality are incorporated here. This work is a piece of the overall verification and validation (V&V) process for simulation models (see Oberkampf, Trucano, and Hirsch (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, which 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 the amount of field data is too scant for such an approach, one needs to be more careful (see Bayarri and Berger (2000), Robins, van der Vaart, and Ventura (2000), and accompanying discussion). Also, the investigation of the sensitivity of predictions to changes in modeling assumptions can be revealing.

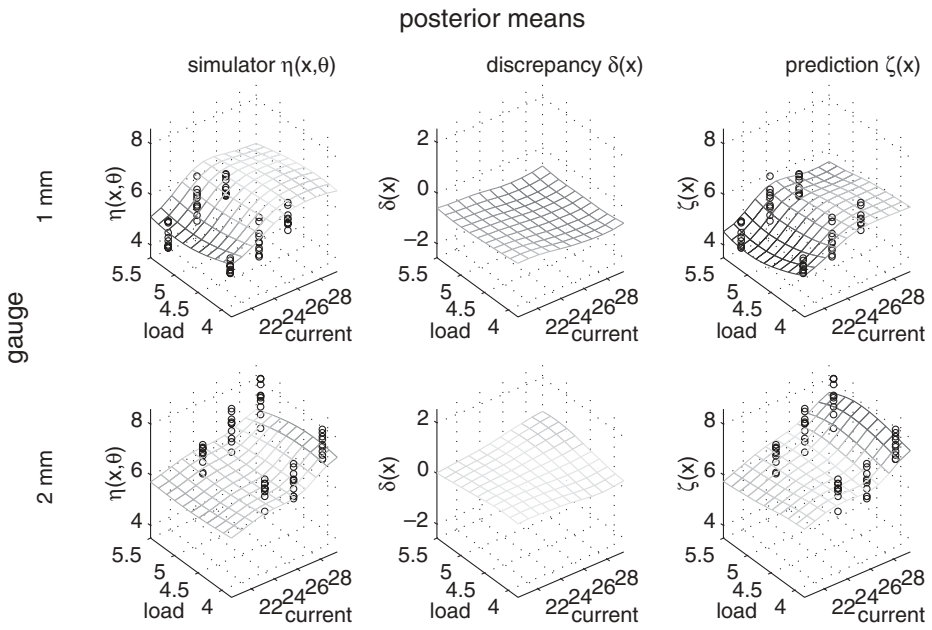

FIG. 9. *Posterior mean estimates for the calibrated simulator prediction $\eta(x, \theta)$, the discrepancy $\delta(x)$, and the calibrated prediction $\zeta(x) = \eta(x, \theta) + \delta(x)$. The plotting symbols show the $n = 120$ field observations. Here the response is weld diameter in mm, load is in kN, and current is in kA.*

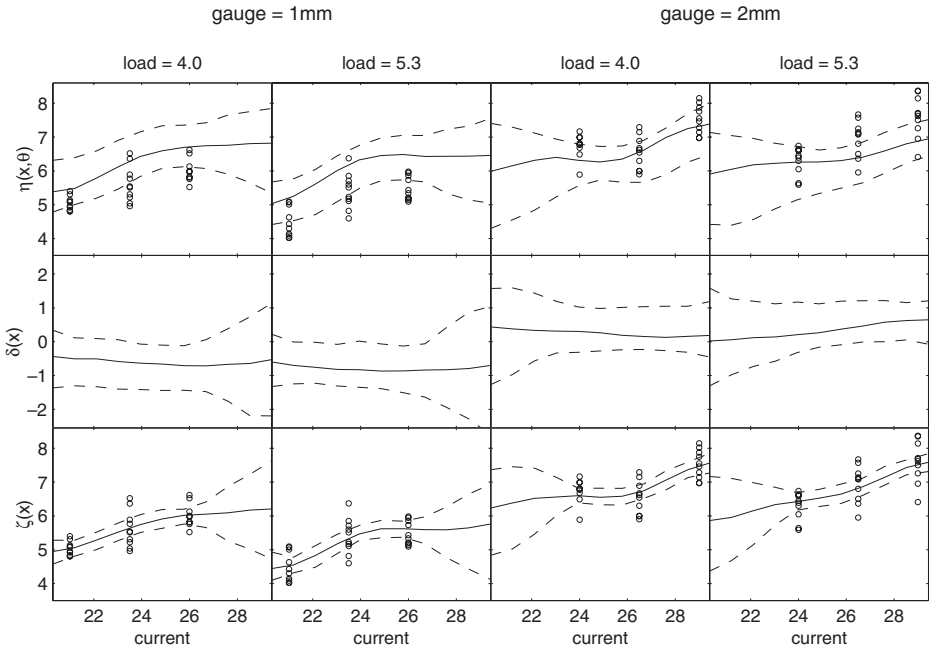

FIG. 10. *Posterior mean and pointwise 90% posterior credible intervals for the calibrated simulator prediction $\eta(x, \theta)$, the discrepancy $\delta(x)$, and the calibrated prediction $\zeta(x) = \eta(x, \theta) + \delta(x)$. The posterior quantities are shown as a function of current for values of load and gauge for which field data y were observed. The plotting symbols show the $n = 120$ field observations. Here the response is weld diameter in mm, load is in kN, and current is in kA.*

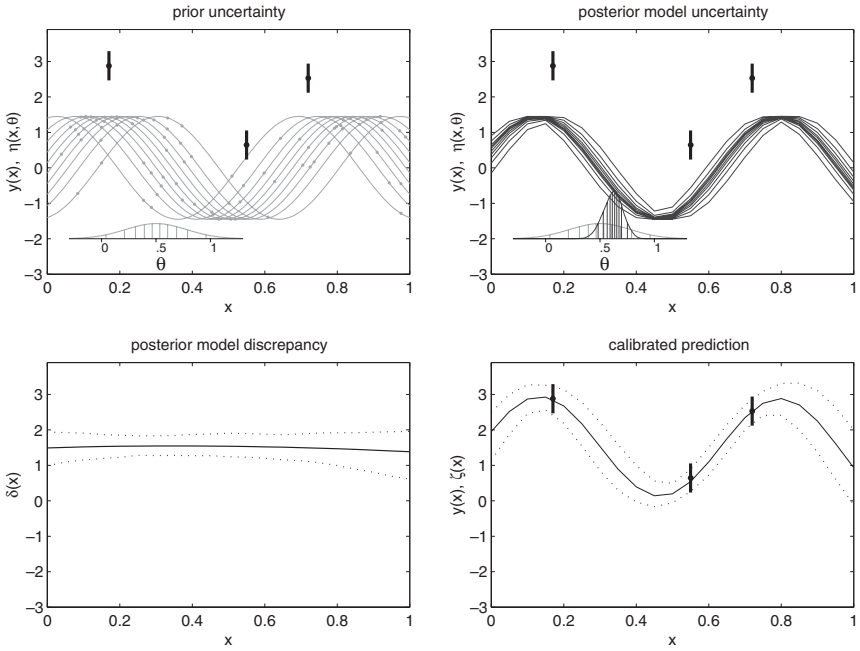

FIG. 11. *Prior uncertainty and posterior decomposition of model terms. Top left: Model runs have been carried out only at the* $m = 60$ (x, θ) *pairs shown by the plotting symbols. As with Figure* 3 *the prior density for* θ *and the implied simulations* $\eta(x, \theta)$ *using quantiles of the prior distribution are given by the light lines. The black dots show field observations, and the corresponding black lines give* 95% *uncertainty bounds on the observations. Inference is carried out by conditioning on the* $n = 3$ *data points as well as the* $m = 60$ *model runs. Top right: The darker lines show pointwise* $5th, 15th, \ldots, 95th$ *percentiles for posterior density for* $\eta(x, \theta)$*. The same percentiles are also shown for the posterior distribution for* θ*. Bottom left: Pointwise* 90% *posterior credible intervals for the model discrepancy term* $\delta(x)$*. Bottom right: Pointwise* 90% *posterior credible intervals for the predictions* $\zeta(x) = \eta(x, \theta) + \delta(x)$*. For this analysis, the inclusion of the model discrepancy term does not make the posterior distribution for* θ *less precise. Here the data were constructed using a* θ *of* .6.

As mentioned previously, the posterior distribution of the discrepancy term $\delta(x)$ is an indicator of how well the simulator is matching reality. Our experience has been that the presence of a large discrepancy term makes interpretation of the posterior distribution for the calibration parameter θ difficult. How this discrepancy term affects the posterior for θ depends on the prior specification as well as the system being modeled. For example, Figure 11 shows a contrived example that mirrors the example in Figure 3, only with a different simulator and field data. Here the discrepancy term is large in magnitude, but the posterior distribution for θ is nearly the same as that obtained when the discrepancy is 0.

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 particle accelerator simulator under this framework, but other multivariate applications will likely require reformulating the statistical model. Examples dealing with such data can be found in Bayarri et al. (2002) and Higdon, Lee, and Holloman (2003).

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 dimen-

sion 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 similar to that of section 2.1 to avoid 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).

Acknowledgments. Thanks to Susie Bayarri, Jim Berger, Jerry Sacks, German Molina, Rui Paulo, and Anasthasios Kottas for their input and insight.

REFERENCES

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

M. J. Bayarri and J. O. Berger (2000), *P values for composite null models*, J. Amer. Statist. Assoc., 95, pp. 1127–1142.

J. Besag, P. J. Green, D. Higdon, and K. Mengersen (1995), *Bayesian computation and stochastic systems (with discussion)*, Statist. Sci., 10, pp. 3–66.

P. S. Craig, M. Goldstein, J. C. Rougier, and A. H. Seheult (2001), *Bayesian forecasting using large computer models*, J. Amer. Statist. Assoc., 96, pp. 717–729.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker (1991), *Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments*, J. Amer. Statist. Assoc., 86, pp. 953–963.

A. Dragt, F. Neri, G. Rangarjan, D. Douglas, L. Healy, and R. Ryne (1988), *Lie algebraic treatment of linear and nonlinear beam dynamics*, Ann. Rev. Nuclear Particle Sci., 38, pp. 455–496.

M. Evans and T. Swartz (1995), *Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems (with discussion)*, Statist. Sci., 10, pp. 254–272.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (1998), *Markov chain Monte Carlo in practice*, Chapman & Hall, London.

B. K. Hegstad and H. Omre (2001), *Uncertainty in production forecasts based on well observations, seismic data and production history*, Soc. Petrol. Engineers J., 6, pp. 409–424.

J. C. Helton (1997), *Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty*, J. Statist. Comput. Simul., 57, pp. 3–76.

D. Higdon, H. Lee, and C. Holloman (2003), *Markov chain Monte Carlo–based approaches for inference in computationally intensive inverse problems*, in Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, eds., Oxford University Press, London, pp. 181–197.

M. E. Johnson, L. M. Moore, and D. Ylvisaker (1990), *Minimax and maximin distance designs*, J. Statist. Planning and Inference, 26, pp. 131–148.

M. Kennedy and A. O'Hagan (2000), *Predicting the output from a complex computer code when fast approximations are available*, Biometrika, 87, pp. 1–13.

M. Kennedy and A. O'Hagan (2001), *Bayesian calibration of computer models (with discussion)*, J. R. Stat. Soc. Ser. B Stat. Methodol., 68, pp. 425–464.

H. Lee, D. Higdon, Z. Bi, M. Ferreira, and M. West (2002), *Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media*, Technometrics, 44, pp. 230–241.

M. D. McKay, R. J. Beckman, and W. J. Conover (1979), *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21, pp. 239–245.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), *Equations of state calculations by fast computing machines*, J. Chem. Phys., 21, pp. 1087–1091.

T. J. Mitchell, M. D. Morris, and D. Ylvisaker (1995), *Two-level fractional factorials and Bayesian prediction*, Statist. Sinica, 5, pp. 559–573.

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

W. L. Oberkampf, T. G. Trucano, and C. Hirsch (2003), *Verification, Validation, and Predictive Capability in Engineering and Physics*, Tech. Report SAND2003-3796, Sandia National Laboratories.

A. O'Hagan (1978), *Curve fitting and optimal design for prediction*, J. R. Stat. Soc. Ser. B Stat. Methodol., 40, pp. 1–24.

A. O'Hagan, M. C. Kennedy, and J. E. Oakley (1999), *Uncertainty analysis and other inference tools for complex computer codes*, in Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., Oxford University Press, London, pp. 503–524.

D. Poole and A. Raftery (2000), *Inference for deterministic simulation models: The Bayesian modeling approach*, J. Amer. Statist. Assoc., 95, pp. 1244–1255.

J. Qiang, R. D. Ryne, S. Habib, and V. Decyk (2000), *An object-oriented parallel particle-in-cell code for beam dynamics simulation in linear accelerators*, J. Comput. Phys., 163, pp. 434–451.

A. C. Rencher (2002), *Methods of Multivariate Analysis*, John Wiley & Sons, New York.

C. P. Robert and G. Casella (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

J. M. Robins, A. van der Vaart, and V. Ventura (2000), *Asymptotic distribution of P values in composite null models*, J. Amer. Statist. Assoc., 95, pp. 1143–1156.

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

A. Saltelli, K. Chan, and E. M. Scott (2000), *Sensitivity Analysis*, John Wiley & Sons, Chichester, UK.

M. Stein (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York.

S. Strogatz (2003), *The real scientific hero of* 1953, Editorial/Op-Ed., New York Times, March 4.

P. C. Wang and D. B. Hayden (1999), *Computational Modeling of Resistance Spot Welding of Aluminum*, GM Research Report R&D-9152, General Motors Research & Development Center, Warren, MI.

C. K. Wikle, R. F. Milliff, D. Nychka, and L. M. Berliner (2001), *Spatio-temporal hierarchical Bayesian modeling: Tropical ocean surface winds*, J. Amer. Statist. Assoc., 96, pp. 382–397.