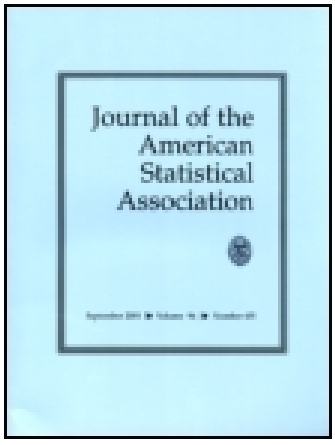

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Bayesian Forecasting for Complex Systems Using Computer Simulators

Peter S Craig[a], Michael Goldstein[a], Jonathan C Rougier[a] & Allan H Seheult[a]
[a] Peter S. Craig is Lecturer, Michael Goldstein is Professor, Jonathan C. Rougier is Senior Research Associate, and Allan H. Seheult is Lecturer, Department of Mathematical Sciences, University of Durham, Durham, DH1 3LE, U.K. . This work has been partially supported by a U.K. Engineering and Physical Sciences Research Council grant GR/L10031. The authors thank Scientific Software- Intercomp for their support, and the referees for helpful comments.
Published online: 31 Dec 2011.

Bayesian Forecasting for Complex Systems Using Computer Simulators

Peter S. CRAIG, Michael GOLDSTEIN, Jonathan C. ROUGIER, and Allan H. SEHEULT

Although computer models are often used for forecasting future outcomes of complex systems, the uncertainties in such forecasts are not usually treated formally. We describe a general Bayesian approach for using a computer model or simulator of a complex system to forecast system outcomes. The approach is based on constructing beliefs derived from a combination of expert judgments and experiments on the computer model. These beliefs, which are systematically updated as we make runs of the computer model, are used for either Bayesian or Bayes linear forecasting for the system. Issues of design and diagnostics are described in the context of forecasting. The methodology is applied to forecasting for an active hydrocarbon reservoir.

KEY WORDS: Bayes linear methods; Calibration; Computer experiments; Design; Diagnostics; History matching; Hydrocarbon reservoir.

1. INTRODUCTION

This article is concerned with forecasting outcomes for a complex system that is simulated using a high-dimensional, computationally expensive, deterministic computer model of the system. The computer model, which we usually call the "simulator," is used to gain understanding of the actual system. Examples of such problems include forecasting future gas production at a particular well in a hydrocarbon reservoir, assessing the future extent of contaminated groundwater, assessing the pattern of radioactive contamination following a nuclear reactor accident, and forecasting weather and climate changes.

Galway and Lucas (1997) summarized the importance and some of the difficulties associated with large-scale simulators of complex systems:

> The computer models …tend to be very large, often with thousands of parameters, and the run times are correspondingly lengthy, so relatively few computer runs are feasible. As a result, relatively sparse "data" exist so that the simulations are almost never validated in the sense of establishing their "closeness" to real-world outcomes. The models are therefore often viewed with some skepticism. However …important decisions must be made, e.g., setting standards and regulations for carbon-dioxide emissions and nuclear waste sites or efficiently managing hydrocarbon production. The first question is how to use such models rigorously and how to account for our uncertainty about the models' relationship to reality.

Substantial difficulties may arise when simulating a complex system:

a. The simulator is not an exact representation of the system.
b. Values of many of the simulator input parameters are unknown.
c. The simulator can be regarded only as a "black box".
d. The simulator is slow and expensive to run.
e. Collections of both past observations and future outcomes may be very large and have complicated spatial/temporal structure.

Although a Bayesian approach seems to us inevitable, in practice prior specification, analysis, and design are often very difficult for complex versions of these problems.

The data available for forecasting usually comprise "past" observations on outcomes of the actual system plus corresponding "past" and "future" output values of the simulator from runs made at a number of different settings for the input parameters. Simulator inputs are usually intended to correspond to determining features of the system; for example, permeability for different rock types in a reservoir and hydraulic conductivity of an aquifer bed in a groundwater model. Calibration, which is concerned with inference about the collection of values for the simulator inputs that best correspond to the system, feeds naturally into forecasting (Beven and Binley 1992), but is discussed only briefly in this account. (For detailed discussion, see, e.g., Craig, Goldstein, Seheult, and Smith 1996, 1997; Kennedy and O'Hagan 2000; O'Hagan, Kennedy, and Oakley 1999.)

Our approach to forecasting is Bayesian. It incorporates a stochastic representation of the deterministic relationship between the inputs and outputs of the simulator, expert assessment of the uncertainty in those input values for which the simulator "best" represents the system, and assessment of the uncertainty about the discrepancy between that best simulator and the actual system. Sacks, Welch, Mitchell, and Wynn (1989) were among the first to study "computer experiments" based on a stochastic representation of deterministic simulators and gave a comprehensive review, and Currin, Mitchell, Morris, and Ylvisaker (1991) gave a Bayesian interpretation. In these articles the emphasis is on predicting simulator output at untried inputs, and design, to minimize some measure of prediction error, was considered by, for example, Bates, Buck, Riccomagno, and Wynn (1996), Morris and Mitchell (1995), and Sacks, Schiller, and Welch (1989). O'Hagan et al. (1999) provided an overview of some of the different problems in the field of computer experiments and develop Bayesian solutions based on a stationary gaussian formulation for simulators with a univariate output and a small number of inputs. (See also O'Hagan 1992 for similar formulations in numerical analysis.) Craig et al. (1996, 1997) described Bayes linear methods for "big" simulators; that is, methods that require only first- and second-order belief specifications for random

Peter S. Craig is Lecturer, Michael Goldstein is Professor, Jonathan C. Rougier is Senior Research Associate, and Allan H. Seheult is Lecturer, Department of Mathematical Sciences, University of Durham, Durham, DH1 3LE, U.K. (E-mail: *P.S.Craig@durham.ac.uk*). This work has been partially supported by a U.K. Engineering and Physical Sciences Research Council grant GR/L10031. The authors thank Scientific Software-Intercomp for their support, and the referees for helpful comments.

quantities. Goldstein (1999) gave an overview of the Bayes linear approach to inference.

In Section 2 we describe our application, forecasting future pressures at a number of wells in an active hydrocarbon reservoir. In Section 3 we formulate a prior model linking system outcomes to corresponding simulator output using a notion of "best" simulator input. We show in Section 4 that, given past system observations and observed outputs from runs on the simulator, the prior modeling augmented with Gaussian assumptions leads to a forecast distribution for future system outcomes that depends on a calibration distribution for the "best" simulator inputs, whose form is a consequence of the prior modeling. The general computational burden required to evaluate the forecast distribution leads us to consider a simpler Bayes linear alternative in Section 5. In Section 6 we illustrate this alternative using our example. A key part of our approach is a statistical model of the the simulator. In Section 7 we describe one such model that allows us to assess the expectations and variances required for forecasting and is also tractable for prior elicitation and design. Design is considered in Section 8. We conclude in Section 9 with a discussion of issues and future directions. Computational details are given in the Appendix.

2. OUR APPLICATION

We illustrate our approach using data from an active hydrocarbon reservoir previously considered in a case study on "history matching," the term for calibration used by reservoir engineers (see Craig et al. 1997 for details). Briefly, the case study concerns a mostly gas-producing reservoir, comprising one mainly onshore field and three offshore fields. The data comprise irregular readings on bottom hole pressure across 109 wells.

For this application, we select six producing wells spanning the onshore part of the reservoir and consider bottom hole pressure at these wells at various times. Our dataset is shown in Figure 1, where the field observations are displayed as ± 2 standard deviation ranges representing the engineer's judgement of reliability. (For these data, a standard deviation of 3% of the actual value was suggested.) The six wells are chosen to have differing numbers of observations available at the notional time of our forecast, which is indicated by the vertical dashed line; in particular, well 56 has no "past" observations. Our forecast time points are shown as circles on the horizontal axis. All wells have a near future and a far future observation available for validation of our forecasts.

In addition to the data, we have available a computer model of the reservoir, constructed by reservoir engineers using commercial software, that includes reservoir structure, geometry, fault patterns, and spatial distributions of permeability and porosity. Among the many simulator inputs, we focus on 7 permeability multipliers (range $[.1, 10.0]$) for the regions into which the reservoir was divided and 33 fault transmissibility multipliers (range $[0, 1]$). Previous experience and the engineer's judgments suggest that the logarithms of the permeability multipliers are more suitable for use in statistical modeling. We also choose to linearly transform so that all input variables vary over the range $[-1/2, 1/2]$. Among the simulator outputs are bottom hole pressures at each well through time.

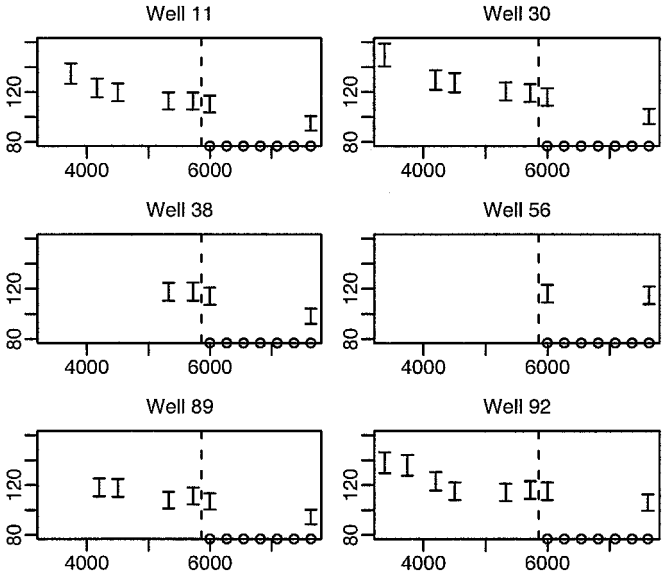

Figure 1. Data Used for Forecasting and Validation, Through Time for Each of the Six Wells. The horizontal axis denotes time in days from the reservoir's inception, and the vertical axis denotes pressure in bars. The error bars represent ±2 s.d. intervals about the ' eld observations, as supplied by the reservoir engineer. The vertical dashed line separates the "past" from the "future." The time points at which we make forecasts are shown as circles on the horizontal axis.

Each run of the computer simulator at a specified set of inputs takes about 10 hours on average. We also have a "coarsened" version of the simulator, with the same inputs and outputs but with larger grid blocks and time steps, which runs in about 3 minutes. We use this to help build the statistical model of the simulator, described in Section 7.

The historical data, information about wells, and all simulator data used in the example are available from http://maths.dur.ac.uk/stats/physpred/software/forecaststudydata.html.

3. PRIOR MODELING

We wish to forecast a vector y_F of future system outcomes, where the elements of y_F will typically be indexed by time and possibly by space for different types of variables. We have observations z_P on a vector y_P of past system outcomes, which we may use to improve our forecasts. We model z_P as $z_P = y_P + e_P$, where the measurement error vector e_P is independent of y_P and has expectation 0 and variance Σ_{e_P}, which is usually diagonal. Also available is a simulator for the system. Corresponding to the combined system outcome vector $y = (y_P, y_F)$, the simulator produces an output vector $s(x) = (s_P(x), s_F(x))$ for any allowable choice of input vector x, where x is a collection of quantities, often high-dimensional, that represent those aspects of the system which affect the outputs.

In our example, y_F is the vector of bottom hole pressures for each of the six wells at the times specified by circles along the horizontal axes of Figure 1, constituting a total of 42 components. The data z_P comprise the 22 observations that lie to the left of the vertical line in Figure 1. The simulator input vector x denotes the linearly transformed fault transmissibilities and logarithms of regional permeability multipliers.

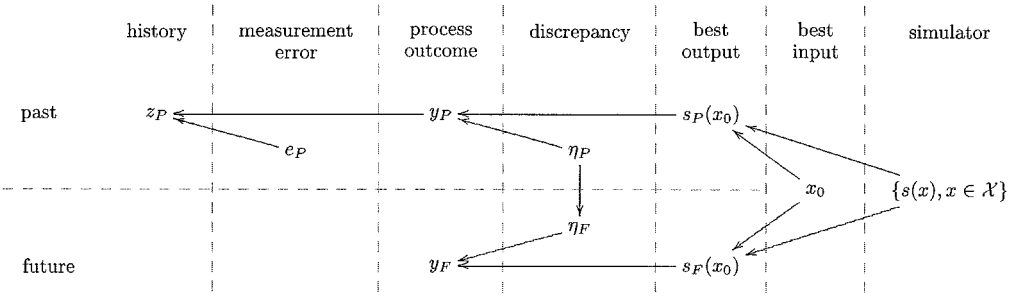

Figure 2. Annotated Directed Conditional Independence Graph for the Components of the Prior Model. All child nodes other than η_F are completely determined by their parents.

We use $\mathcal{X}$ to denote the space of allowable values of x. In this article we restrict attention to computer models for which the outputs are determined solely by the inputs x. Thus we do not consider the additional modeling issues introduced when the outputs are partially influenced by other inputs, such as environmental variables, which vary randomly over time. We also exclude the possibility of design inputs that may be used to partially control the future outputs. In many problems, evaluating the simulator output $s(x)$ for any chosen x is expensive. In such cases, substantial uncertainty will exist about the value of $s(x)$ for large subsets of $\mathcal{X}$. Consequently, we treat $\{s(x),\ x \in \mathcal{X}\}$ as a random function about which we learn as we make simulator runs. In Section 7 we describe a particular statistical representation for $s(x)$ that is tractable and appropriate for our example.

We treat the relationship between the computer model and the physical system analogously to that of a standard parametric statistical model. That is, we make the simplifying assumption that there is a unique but unknown choice $x_0 \in \mathcal{X}$ that corresponds to the true values of the inputs x. The value $s(x_0)$ does not exactly reproduce the physical system, but, given the evaluation $s(x_0)$, we believe that we can learn nothing further about the physical system from the simulator. We express this judgment by the relationship $y = s(x_0) + \eta$, where the discrepancy η between the system and the best representation $s(x_0)$ from the simulator of the system is independent of both x_0 and each $s(x)$, and η has expectation 0 and variance Σ_η. The quantity x_0 is unknown and is independent of each $s(x)$. We separate past and future components of y as $y_P = s_P(x_0) + \eta_P$ and $y_F = s_F(x_0) + \eta_F$, and use Σ_{η_P}, Σ_{η_F}, and $\Sigma_{\eta_{PF}}$ to denote the variances and covariance of η_P and η_F. The variance matrix Σ_η, which must be elicited from a system expert, such as a reservoir engineer, represents the link between the simulator and the system; see Section 6, where we discuss the choice of Σ_η for our example.

Because all prior uncertainties in our description derive from the judgments of system experts, the plausibility of our assumptions in any application derives from the assessment of the expert. For example, for reservoir applications, it is common practice to forecast the future by seeking a value of $\hat{x}$ for the inputs for which $s_P(\hat{x})$ is close to the observed z_P ("history matching"), and then treating the value of $s_F(\hat{x})$ as the forecast for y_F. Some form of sensitivity analysis may then be performed based on the consideration that $\hat{x}$ may not be the best choice. We may criticize this approach for poor treatment of forecast uncertainty, but it does suggest that, to a first approximation, the reservoir engineer is prepared to behave as though there is a best choice for x. In general, the closer the correspondence between the inputs to the model and the features of the physical system, the more our assumptions will correspond to precise statements of belief. In other circumstances, these assumptions would serve as a pragmatic simplification of a more complex belief system, whose additional features would often be difficult to elicit, while only having a minor effect on our analysis.

Our approach to forecasting is thus based on past system observations z_P and any runs of the simulator that we make, in combination with expert prior beliefs about (a) the discrepancy η between the system and the simulator, (b) the value of x_0, and (c) the function $s(x)$. Graphical representation of the elements of our prior modeling is shown in the conditional independence graph in Figure 2.

In Sections 4 and 5 we suggest two Bayesian approaches to forecasting, the first based on augmenting the prior model described so far with specific distributional assumptions, and the second a simpler Bayes linear approach that is particularly useful when, because of the model's dimensionality, the full Bayes formulation becomes difficult to specify meaningfully or leads to intractable calculations.

4. BAYESIAN FORECASTING

We now describe a full Bayes approach to forecasting. The two sources of data are the system observation z_P and outputs $S = (s(x_1), \ldots, s(x_n))$ from n runs of the simulator at inputs $x_1, \ldots, x_n$. We forecast y_F by evaluating the conditional distribution $\mathrm{p}(y_F|z_P, S)$ of y_F given z_P and S, which we compute as

$$\mathrm{p}(y_F|z_P, S) = \int_{\mathcal{X}} \mathrm{p}(y_F|z_P, S, x_0)\, \mathrm{p}(x_0|z_P, S)\, dx_0, \qquad (1)$$

where the posterior density $\mathrm{p}(x_0|z_P, S)$ for x_0 may also be used to calibrate (tune) the simulator to z_P.

For the purposes of the Bayesian approach in this section, we augment the assumptions of Section 3 as follows: (a) the discrepancy between the system and the simulator is such that $\eta = y - s(x_0) \sim \mathrm{N}(0, \Sigma_\eta)$; (b) the simulator output $s(x)$ for $x \in \mathcal{X}$ is a Gaussian process; and (c) the measurement error $e_P = z_P - y_P \sim \mathrm{N}(0, \Sigma_{e_P})$. It follows from (b) that the posterior distribution of $s(x_0)$ given S and x_0 is $\mathrm{N}(\mu(x_0), \Sigma(x_0))$, where

$$\mu(x) = \mathrm{E}[s(x)] + \mathrm{cov}[s(x), S]\mathrm{var}[S]^{-1}(S - \mathrm{E}[S]) \qquad (2)$$

and

$$\Sigma(x) = \mathrm{var}[s(x)] - \mathrm{cov}[s(x), S]\mathrm{var}[S]^{-1}\mathrm{cov}[S, s(x)].$$

When $\mathrm{var}[S]$ is not invertible, it may be replaced by the corresponding Moore–Penrose generalized inverse. We write $V(x) = \Sigma(x) + \Sigma_\eta + \Sigma_e$, where the entries of Σ_e are 0 except for the measurement error variance matrix Σ_{e_P}, and subscripts P and F refer to "past" and "future" partitions of $\mu(x)$ and $V(x)$, so that $V_F(x) = \Sigma_F(x) + \Sigma_{\eta_F}$, $V_{FP}(x) = \Sigma_{FP}(x) + \Sigma_{\eta_{FP}}$, and $V_P(x) = \Sigma_P(x) + \Sigma_{\eta_P} + \Sigma_{e_P}$. Note that $V(x_0)$ is the conditional variance of (z_P, y_F) given S and x_0. It is straightforward to show that the first term in the integrand in (1) is such that

$$y_F | z_P, S, x_0 \sim \mathrm{N}\big(\mu_{F|P}(x_0), \Sigma_{F|P}(x_0)\big),$$

where

$$\mu_{F|P}(x) = \mu_F(x) + V_{FP}(x)V_P(x)^{-1}\big(z_P - \mu_P(x)\big)$$

and

$$\Sigma_{F|P}(x) = V_F(x) - V_{FP}(x)V_P(x)^{-1}V_{PF}(x).$$

Note that $V_P(x)$ must be invertible if all measurement errors are uncorrelated with positive variance.

The second term $\mathrm{p}(x_0|z_P, S)$ in the integrand in (1) can be shown to be proportional to $\mathrm{p}(x_0)\,\mathrm{p}(z_P|S, x_0)$, where $\mathrm{p}(z_P|S, x_0)$ is interpretable as the likelihood for x_0 and is such that $z_P|S, x_0 \sim \mathrm{N}(\mu_P(x_0), V_P(x_0))$. Because the proportionality constant has no closed-form expression, it must be evaluated numerically, which can be a daunting task when x and z_P are high-dimensional.

Because of the complicated form of the posterior distribution for x_0, a simple expression for the forecast distribution (1) is not available. Thus numerical evaluation can be a massive computational burden for all but low-dimensional problems. Alternatives such as direct stochastic simulation using rejection sampling appear to be unacceptably expensive, except again for low-dimensional problems. The computational burden becomes orders of magnitude larger if we consider the design problem that must be solved in selecting at which values of the inputs we should choose to run the simulator to construct the collection S on which our forecast is based. This is of particular importance if $s(\cdot)$ is expensive to run so that we may only make a small number of simulator runs. Therefore, we have developed an alternative Bayes linear approach to forecasting that is simpler to specify and analyze, and which we describe in the following section.

5. BAYES LINEAR FORECASTING

In this section we drop the full distributional assumptions of Section 4 and describe Bayes linear adjustment of future system outcomes y_F given the past observations z_P. All that we require to make such an adjustment is the mean and covariance structure for (z_P, y_F), although in complex problems even this may be hard to obtain. The Bayes linear approach may be viewed either as a pragmatic approximation to a full Bayesian analysis, based on simple aspects of prior beliefs, or, from a foundational viewpoint, as an appropriate method for using second-order prior specifications. An overview of the Bayes linear approach was given by Goldstein (1999). Here the relevant equations are

$$\mathrm{E}_{z_P}[y_F] = \mathrm{E}[y_F] + \mathrm{cov}[y_F, z_P]\mathrm{var}[z_P]^{-1}(z_P - \mathrm{E}[z_P]) \quad (3)$$

and

$$\mathrm{var}_{z_P}[y_F] = \mathrm{var}[y_F] - \mathrm{cov}[y_F, z_P]\mathrm{var}[z_P]^{-1}\mathrm{cov}[z_P, y_F], \quad (4)$$

where $\mathrm{E}_{z_P}[y_F]$ in (3) is the "expectation of the future system output y_F adjusted by past system observations z_P." Thus to forecast future system output y_F, we adjust the prior forecast $\mathrm{E}[y_F]$ linearly in z_P according to the second term on the right side of (3). Forecast uncertainty is assessed by $\mathrm{var}_{z_P}[y_F]$ in (4), which is the "variance of y_F adjusted by z_P." Thus the prior uncertainty $\mathrm{var}[y_F]$ in y_F is reduced by an amount given by the second term on the right of (4) that depends only on the prior covariance structure of (z_P, y_F) and not on the value of z_P itself.

As $y = s(x_0) + \eta$, the correlation between y_F and y_P is in part determined by beliefs about the form of the simulator function $s(x)$. Next, we discuss using the observed simulator runs to update beliefs about the function $s(x)$, using these updated beliefs to determine the covariance structure for (z_P, y_F) used in (3) and (4), and the resulting forecast methodology.

5.1 Adjusting Beliefs About $s(x)$ Using the Simulator Runs

Suppose, as in Section 4, that we observe outputs S from a series of n runs at inputs $x_1, \ldots, x_n$ on the simulator of the system. We update our expectations and variances for the simulator $s(x)$ using the Bayes linear adjusted expectation $\mathrm{E}_S[s(x)]$ for $s(x)$ given S and the adjusted variance $\mathrm{var}_S[s(x)]$, where

$$\mathrm{E}_S[s(x)] = \mathrm{E}[s(x)] + \mathrm{cov}[s(x), S]\mathrm{var}[S]^{-1}(S - \mathrm{E}[S]) \quad (5)$$

and

$$\mathrm{var}_S[s(x)] = \mathrm{var}[s(x)] - \mathrm{cov}[s(x), S]\mathrm{var}[S]^{-1}\mathrm{cov}[S, s(x)]. \quad (6)$$

Thus, given the outputs S from the n simulator runs, we can update our prior expectation $\mathrm{E}[s(x)]$ for the simulator output $s(x)$ at x by the linear adjustment in S given in the second term of (5), where the magnitude of the adjustment depends on the strength of the prior correlations between $s(x)$ and $s(x_1), \ldots, s(x_n)$. The reduction in our prior uncertainty $\mathrm{var}[s(x)]$ about $s(x)$, given by the second term in (6), depends only on the prior correlations between $s(x)$ and $s(x_1), \ldots, s(x_n)$ and not on the values of the simulator outputs S. Equations (5) and (6) are analogous to those used in kriging (see, e.g., Cressie 1991).

Note that $\mathrm{E}_S[s(x)] = \mu(x)$ and $\mathrm{var}_S[s(x)] = \Sigma(x)$, where the expressions for $\mu(x)$ and $\Sigma(x)$ are those given in (2). Bayes linear adjustments make no distributional assumptions; the equivalence here is a consequence of the joint normality of S and $s(x)$ in Section 4. In particular, the joint distribution of the quantities z_P and y_F involved in the Bayes linear adjustment of the future on the past given in (3) and (4) will not be Gaussian even when $\{s(x),\ x \in \mathcal{X}\}$ is a Gaussian process.

5.2 The Mean and Covariance Structure for $s(x_0)$

We now calculate the mean μ and variance Σ for $s(x_0)$ from our prior beliefs for x_0 and our adjusted beliefs about $s(x)$ in (5) and (6) as follows (noting that $\mu(x)$ is also a function of S):

$$\mu = \mathrm{E}[\mu(x_0)|S] \tag{7}$$

and

$$\Sigma = \mathrm{E}[\Sigma(x_0)|S] + \mathrm{var}[\mu(x_0)|S]. \tag{8}$$

Here the expectations and variance on the right side of (7) and (8) are with respect to the prior distribution $p(x_0)$ for x_0, because x_0 and S are independent; for example,

$$\mathrm{E}[\mu(x_0)|S] = \int_x \mu(x_0)p(x_0)dx_0,$$

and similarly for $\mathrm{E}[\Sigma(x_0)|S]$ and $\mathrm{var}[\mu(x_0)|S]$.

Thus the mean and variance of $s(x_0)$ are evaluated as the expectation and variance (with respect to $p(x_0)$) of our current beliefs about the value of $s(x)$ at $x = x_0$. In general, these calculations require extensive numerical integration. However, under certain plausible approaches to prior modeling (discussed in Sec. 7), as we show in the Appendix, explicit closed-form expressions can be derived for μ and Σ when $p(x_0)$ is either uniform or Gaussian. We use Σ_P, Σ_F, and Σ_{PF} to refer to the obvious submatrices of Σ and μ_P and μ_F to refer to the corresponding past and future component vectors of μ. The value μ_F is the forecast for y_F, with variance Σ_F, based on observed simulator runs S, our statistical representation for $s(x)$, and our beliefs about x_0, before observing z_P.

5.3 Adjusting Beliefs About y_F Using the Past z_P

As described at the beginning of Section 5, to compute the forecast for y_F we carry out Bayes linear fitting of the future y_F on the past z_P, assessing the Bayes linear mean and variance for y_F given z_P as

$$\begin{aligned}\mathrm{E}_{z_P}[y_F] &= \mathrm{E}[y_F] + \mathrm{cov}[y_F, z_P]\mathrm{var}[z_P]^{-1}(z_P - \mathrm{E}[z_P]) \\ &= \mu_F + (\Sigma_{FP} + \Sigma_{\eta_{FP}})(\Sigma_P + \Sigma_{\eta_P} + \Sigma_{e_P})^{-1} \\ &\quad \times (z_P - \mu_P)\end{aligned} \tag{9}$$

and

$$\begin{aligned}\mathrm{var}_{z_P}[y_F] &= \mathrm{var}[y_F] - \mathrm{cov}[y_F, z_P]\mathrm{var}[z_P]^{-1}\mathrm{cov}[z_P, y_F] \\ &= (\Sigma_F + \Sigma_{\eta_F}) - (\Sigma_{FP} + \Sigma_{\eta_{FP}})(\Sigma_P + \Sigma_{\eta_P} + \Sigma_{e_P})^{-1} \\ &\quad \times (\Sigma_{PF} + \Sigma_{\eta_{PF}}),\end{aligned} \tag{10}$$

where we take the adjusted beliefs arising from (7) and (8) as prior beliefs for z_P and y_F. The specific forms on the right sides of (9) and (10) follow simply from the relations $y_P = s_P(x_0) + \eta_P$, $y_F = s_F(x_0) + \eta_F$, and $z_P = y_P + e_P$ and from partitioning μ and Σ into past and future components.

A natural interpretation of (9) and (10) is to view μ_F as the forecast from the simulator for y_F that we modify when there are large discrepancies between μ_P, the forecast from the simulator for y_P, and the observed value of z_P. The size of the modification follows from the strength of our prior correlation between past and future discrepancies between the simulator and the system, and the extent to which past and future quantities in the simulator may be correlated through dependence on the common quantity x_0.

5.4 Lower Bounds on Forecast Uncertainty

If we are trying to set targets for forecast uncertainty, then it can be shown that forecast variances cannot be reduced below

$$\Sigma_{\eta_F} - \Sigma_{\eta_{FP}}(\Sigma_{\eta_P} + \Sigma_{e_P})^{-1}\Sigma_{\eta_{PF}} \tag{11}$$

whether we proceed by Bayes linear fitting or by carrying out a full Bayes Gaussian analysis. One argument is as follows. We wish to forecast $y_F = s_F(x_0) + \eta_F$ from $z_P = y_P + e_P = s_P(x_0) + \eta_P + e_P$. If we could run the simulator at all possible inputs and compare the output from each input with the past z_P, then in principle we might be able to determine the values of $s_P(x_0)$ and $s_F(x_0)$, and so also the value of $z_P - s_P(x_0)$. The remaining forecasting uncertainty for y_F is the variance for η_F, given $\eta_P + e_P$ in (11). To put it another way, the simulator is uninformative about the discrepancy η or the measurement error e_P.

In certain circumstances we can find a better, albeit approximate, lower bound on forecast uncertainty. As we increase the number of simulator runs, $\Sigma(x)$ decreases for all x and hence $\mathrm{E}[\Sigma(x_0)|S]$ decreases and has some known lower bound Ξ determined by our choice of statistical model for $s(x)$. When $\mathrm{E}[\Sigma(x_0)|S]$ is small, $\mathrm{var}[\mu(x_0)|S]$ is unlikely to change much. Hence an approximate lower bound for Σ is $\Xi + \mathrm{var}[\mu(x_0)|S]$, which in turn provides an approximate lower bound for forecast uncertainty, obtained by substituting $\Xi + \mathrm{var}[\mu(x_0)|S]$ for Σ in (10). The approximation becomes more reliable as $\mathrm{E}[\Sigma(x_0)|S]$ approaches its limit.

5.5 Forecast Diagnostics

A simple diagnostic assessment for our forecasting approach is based on the comparison of our beliefs about the past values of the physical system and the actual historical values that we have observed. The observed values are the vector z_P, the expectation of z_P given the runs on the simulator is μ_P, and the difference is $D_P = z_P - \mu_P$ with expected value of 0. If D_P is many standard deviations from 0, in some appropriate norm, then we may need to reconsider the basis of our prior modeling. We may monitor the magnitudes of individual scalar components of D_P, and also assess certain standardized multivariate summaries. The simplest such summary measure is

$$S(D_P) = D_P^T \mathrm{var}[D_P]^{-1} D_P, \tag{12}$$

where $\mathrm{var}[D_P]$ is the variance matrix of D_P, namely $\mathrm{var}[D_P] = \Sigma_P + \Sigma_{\eta_P} + \Sigma_{e_P}$. The expected value of $S(D_P)$ is the rank of $\mathrm{var}[D_P]$, which can be used to calibrate the diagnostic.

The foregoing diagnostic assesses the overall discrepancy between the simulator and system, but is not tuned particularly to problems relating to our forecasting methodology. Thus we may partition D_P into uncorrelated subvectors to highlight such problems. Divide the time interval of historical observations into m, not necessarily equal, time

periods $t_1 < t_2 < \cdots < t_m < t_{m+1}$. Let z_i be the vector of observations whose values were recorded in the period $[t_i, t_{i+1})$ for $i = 1, \ldots, m$, and $z_{[i]}$ be the vector of observations whose values were recorded in the period $[t_1, t_i)$, where $z_{[1]}$ is empty. Let D_i denote the vector of forecast errors when Bayes linear forecasting the vector z_i of observations in $[t_i, t_{i+1})$, and let $z_{[i]}$ denote the vector of observations available before time i, namely

$$D_i = z_i - \mathrm{E}_{z_{[i]}}[z_i].$$

The collections $D_1, ..., D_m$ are mutually uncorrelated, with prior mean 0, and may be monitored diagnostically by adapting (12) as

$$S(D_i) = D_i^T \mathrm{var}[D_i]^{-1} D_i.$$

Adjusted expectations for each z_i and variances for D_i are given by the analogous forms to (9) and (10). The collection of values $S(D_i)$ may be monitored to assess whether historically our claimed forecast accuracy has been achieved, whether there are noticeable changes over time in the forecast diagnostics, and so forth. We may also choose to further decompose the diagnostics $D_1, \ldots, D_m$; for example, by assessing separately each scalar component of each discrepancy D_i to assess whether there are any spatial aspects to observations with large forecast diagnostics.

Some elements of the vector D_P may refer to aspects of the past that have no effect on our forecasts. If we want to restrict attention only to that subset of z_P that does affect our forecasts, then we may carry out the diagnostic analysis on changes in forecast for y_F that we make given z_P, by comparing observed to expected values for $D_F = \mathrm{E}_{z_P}[y_F] - \mu_F$. Note that D_F is a linear transform of D_P that selects that subspace of z_P that determines the forecasts. Again, we may monitor individual components of D_F or use summary measures, such as

$$S(D_F) = D_F^T \mathrm{var}[D_F]^{-1} D_F \tag{13}$$

with expected value the rank of $\mathrm{var}[D_F]$.

6. BACK TO OUR EXAMPLE

We now illustrate the Bayes linear approach to forecasting using our example. In Section 7 we give the details of simulator modeling that allows us evaluate the terms in (5) and (6).

6.1 Prior Beliefs for x_0 and η

We use a uniform prior distribution for x_0 over $\mathcal{X} = [-1/2, 1/2]^{40}$, which is consistent with the engineer's beliefs.

In our example we have two sources of information to help us specify Σ_η. First, the reservoir engineer suggested that median absolute error of about 5% would be appropriate for each component of η, although he expressed some doubts about his own judgment. Second, we also have available what was judged to be the "best" run of the simulator in the history matching exercise (from Craig et al. 1996). We performed a simple analysis using the difference between simulator outputs at this best run and the actual field data z_P. This suggested

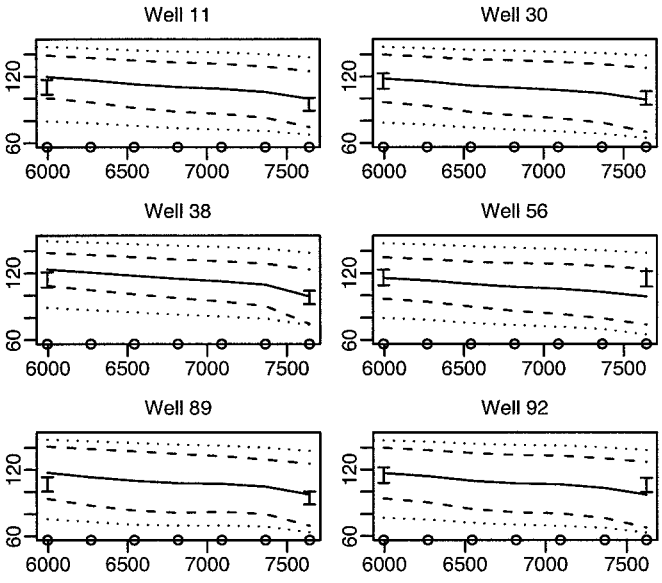

Figure 3. Forecasts for the Six Wells, Without Using z_P. The outer dotted interval represents ±2 s.d. before any simulator runs, and the inner dashed interval the same after six simulator runs, with the mean curve after the six runs shown as a solid line. The forecast time points are shown along the horizontal axis as circles. The data kept back for validation are shown as bars denoting ±2 measurement error s.d. about the ' eld observations.

that there are temporal effects and well effects, a hypothesis supported by the engineer. On this basis, we chose the model

$$\mathrm{cov}[\eta_k, \eta_l] = \sigma_1^2 \exp\{-\theta_1 (T_k - T_l)^2\} + \sigma_2^2 \exp\{-\theta_2 (T_k - T_l)^2\} 1_{W_k = W_l},$$

where the kth output component comes from well W_k at time T_k, and 1_p denotes the indicator function of the proposition p. We assigned values to the four parameters by informal data analysis using a combination of guessing-and-simulating and variogram methods. This resulted in $\sigma_1^2 = 25$, $\theta_1 = (6 \times 10^{-4})^2$, $\sigma_2^2 = 6$, and $\theta_2 = (2 \times 10^{-3})^2$. The resulting variances for the individual components of η are smaller than, but of the same order of magnitude as, those suggested by the engineer. A more sophisticated model for a larger-scale study might replace the well effects by spatial modeling.

6.2 Forecasts

We consider forecasts using four combinations of information: with and without simulator runs S, and with and without past data z_P. We use six simulator runs chosen according to the design procedure described in Section 8. Figure 3 shows our forecasts with and without the simulator runs, without z_P, and Figure 4 shows our forecasts with and without the simulator runs, using z_P. In both figures the mean curve prior to the simulator runs has been omitted for greater clarity.

In Figure 3, where z_P is not used, the effect of simulator runs is seen mainly in the reduction in forecast uncertainty, although there is also some effect on the mean curve. Note that uncertainty does not necessarily increase through time in these panels, as the simulator can be thought of as providing "contemporary" information at all time points. In Figure 4 the

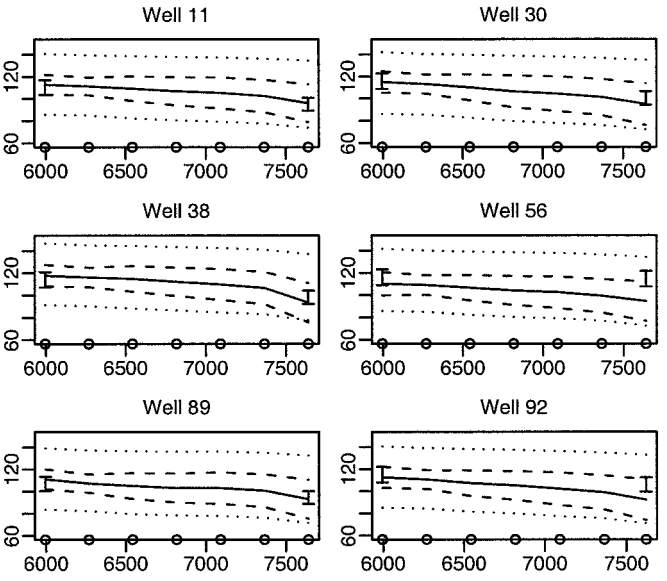

Figure 4. Forecasts for the Six Wells, Using z_P. See the caption for Figure 3 for details.

effect of including z_P is small before running the simulator, but large in conjunction with the simulator runs. This effect extends the full length of the forecast period, although the reduction in uncertainty is most marked at the near end, where there is recently available data in z_P. This is true even for well 56, for which there is no historical data in z_P. The forecasts for well 56 have "borrowed strength" from the data at the other wells through our prior covariance structure on $s(x)$ and on Σ_η. In each panel the error bars representing the data held back for validation are consistent with the forecast intervals.

6.3 Diagnostics

Figure 5 shows the diagnostics given by (12) and (13). All of the diagnostics are somewhat smaller than their expected value, indicating that the prior modeling has been somewhat conservative.

Figure 6 illustrates diagnostics based on partitioning D_P into seven subvectors, $D_1, \ldots, D_7$, by increasing time of observation. We can see that the standardized errors are not unreasonably large, and that there appears to be no systematic overforecast or underforecast at any individual well.

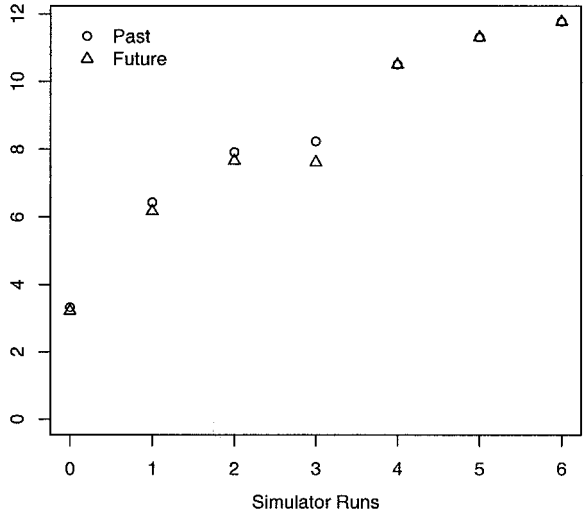

Figure 5. Forecast Diagnostics $S(D_P)$(∘) and $S(D_F)$(Δ) After Each Run of the Simulator. The expected value in each case is 22, the number of past observations.

7. MODELING THE SIMULATOR

To evaluate the expectations and covariances required for forecasting, we must model the form of the simulator function $s(x)$. We now describe a general approach to modeling that we have found useful for reservoir problems. We describe the model first when $s(x)$ is univariate and then when it is multivariate.

7.1 Univariate $s(x)$

We select a subset x^* of components of x that we believe to be most important in explaining the variation in $s(x)$ over x. Often x will be a suitable transform and reparameterization of the original simulator inputs. We let x^* denote the active inputs for $s(x)$, where x^* is usually a small subset of the total available inputs (see also Welch et al. 1992). We model the relationship between $s(x)$ and x as a sum,

$$s(x) = h(x^*) + \epsilon(x^*) + \delta(x), \tag{14}$$

of three independent components. Although this decomposition may raise identifiability issues, these appear not to be a major problem in forecasting.

The quantity $h(x^*)$ in (14) is a regression surface in the active inputs, representing the main global aspects of the variation in $s(x)$ with x. We generally choose the polynomial form $h(x^*) = \beta^T g(x^*)$, where $g(x^*)$ is a vector of monomial functions in the components of x^* and β is a vector of regression

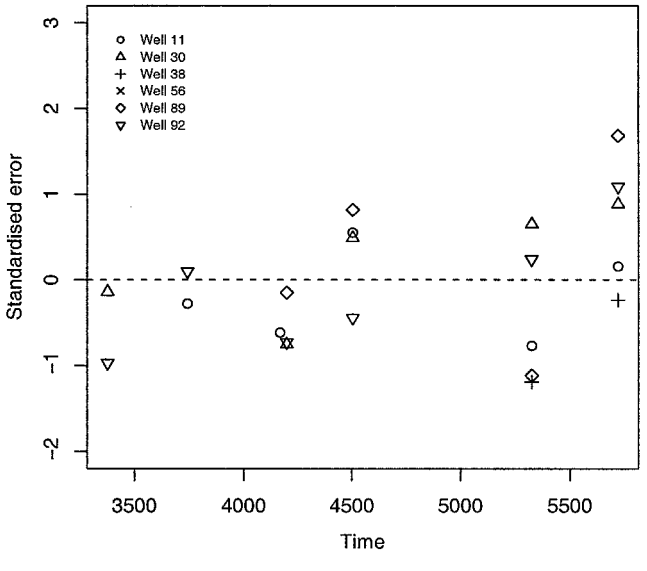

Figure 6. Forecast Diagnostics $D_1, \ldots, D_7$, After Six Runs of the Simulator. (∘, well 11; Δ, well 30; +, well 38; ×, well 56; ◇, well 89; ∇, well 92) Each point represents the standardized forecast error of a component of z_P, using all components of z_P from earlier time points.

coefficients. We specify prior means, variances, and covariances for the components of β.

The quantity $\epsilon(x^*)$ in (14) is a smooth function of x^* that expresses part of the residual between $s(x)$ and $h(x^*)$ as a stationary stochastic process with prior mean 0 and autocovariance function of the form $\text{cov}[\epsilon(x^*), \epsilon(x^{*\prime})] = \sigma_\epsilon^2 r(x^* - x^{*\prime})$ that must be specified. Our intention is for $\epsilon(x^*)$ to describe variation in $s(x)$ that is not captured by $h(x^*)$ but is smoothly dependent on x^*.

Finally, the quantity $\delta(x)$ in (14) is a stationary process in x expressing all of the remaining variation in $s(x)$ that is not explainable by the terms involving x^* alone. In principle, we should specify a covariance function for δ over x. However, we generally choose instead to view $\delta(x)$ as white noise, which is a reasonable approximation provided that σ_δ^2, the variance of $\delta(x)$, is small. This leads to enormous reductions in the complexity and dimensionality of many of the calculations that we require, with no substantial distortion of our beliefs over $s(x)$.

In the computer experiments literature, it is common to use the constant form $h(x^*) = \beta_0$ with a possibly more complicated stationary correlation structure for $\epsilon(x^*) + \delta(x)$ (see, e.g., Sacks et al. 1989). But a nonstationary formulation for $s(x)$ using a nontrivial polynomial form in active variables x^* for $h(x^*)$ has certain advantages:

a. It leads to the tractable approximate design criteria for forecasting described in Section 8.
b. It ensures numerical stability.
c. It allows space/time modeling of the β components.
d. Its inclusion may well reflect expert judgment regarding the behavior of $s(x)$.
e. It can lead to an efficient description of $s(x)$, as evidenced in the example, where we used quadratic functions of the active variables of each component.

Prior descriptions of belief such as (14) may be based on a combination of two sources of information. First, there is elicitation from experts in the system under investigation. This is particularly valuable in identifying the candidate active inputs x^*. Second, we may often construct approximations to the full simulator that can be run much faster, such as coarsening the grids over which the underlying mathematical model of the system is solved. Running these simple simulators many times, carrying out statistical fitting of the models on the outputs, and adjusting the uncertainties to reflect perceived differences between the forms on the different simulators gives us quantified models to form the basis of our analysis. In the simplest case, we suppose that there is a single fast approximation $\tilde{s}(x)$ for $s(x)$ that has the same active variables as $s(x)$ and shares the conditional independence structure of $s(x)$; that is, η, the discrepancy between the system and the full simulator, is also independent of each $\tilde{s}(x)$ given $s(x_0)$. These two approaches for constructing prior beliefs of the form (14) and the multivariate extension in (15) have been discussed by Craig et al. (1998). (see also Kennedy and O'Hagan 2000 and O'Hagan et al. 1999 for discussion of fast approximations.) Using such models for calibration ("history matching") was described by Craig et al. (1996, 1997). In the example here, we have mostly used data from a fast approximate simulator.

7.2 Multivariate $s(x)$

Having specified beliefs when $s(x)$ is univariate, we now complete the prior description when $s(x)$ is multivariate by specifying the covariance structure between each pair of components s_k and s_l of $s(x)$, where now we write

$$s_k(x) = h_k(x_k^*) + \epsilon_k(x_k^*) + \delta_k(x), \tag{15}$$

with active inputs x_k^* and where $\epsilon_k(x_k^*)$ has autocorrelation function $r_k(x_k^* - x_k^{*\prime})$.

There is no unique way to make this completion. However, the three components are independent for an individual output, and we retain that structure in the multivariate model.

Correlations must be specified between all of the regression coefficients. A simple approach, adopted in the example, is to correlate components of the β vectors for two outputs only if they share the same corresponding monomial function. However, using a more complex approach when needed carries no substantial computational penalty.

We restrict ourselves to choosing the same autocorrelation function r_k for any pair of outputs that have the same active variables. We do so because in general it is not easy to choose a coherent nonzero cross-correlation function for ϵ_k and ϵ_l. However, when two outputs share the same active inputs and the same correlation function $r(x^* - x^{*\prime})$, a straightforward approach is to set the cross-correlation to $\nu\, r(x^* - x^{*\prime})$. Moreover, computational savings result when the choice of autocorrelation function depends only on the collection of active variables. Hence, when s_k and s_l have the same active variables, we assign

$$\text{corr}[\epsilon_k(x^*), \epsilon_l(x^{*'})] = \nu_{kl}\, r_k(x_k^* - x_k^{*\prime}),$$

and the correlation is otherwise 0. Further computational savings are realized when $r_k(x^* - x^{*\prime})$ is a product of correlations derived from the individual active variables provided that the same correlation function is used for a given active variable wherever it appears. For Gaussian or uniform prior distributions for x_0, the forms $\exp\{-\theta|x - x'|\}$ and $\exp\{-\theta|x - x'|^2\}$ have computational advantages. Where all active variables have the same range of variation, it is natural to use the same correlation function for each of them.

The white noise correlation structure for each $\delta_k(x)$ limits us to correlating $\delta_k(x)$ and $\delta_l(x')$ only when $x = x'$. We assign a constant correlation τ_{kl} when $x = x'$, and 0 otherwise.

Such choices reduce the joint specification to the specification of several correlation matrices where the structure of each matrix will be determined by the phenomenon being modeled. For example, the subscripts k and l often represent locations in space and time, so that the correlation structures and values can be determined by a combination of spatial/temporal modeling based on output from many runs on a fast version of the simulator and expert elicitation. Diagnostics for simulator modeling have been discussed by Craig et al. (1997).

For this approach to simulator modeling, computation of the adjusted mean and covariance structure in (5) and (6) and the integrals for (7) and (8) are detailed in the Appendix.

7.3 The Example

To build the models that we have described, we used 101 runs of the coarsened version of the simulator, one at the origin and the other 100 on a latin hypercube design. For each output we selected up to three candidate active variables by forward selection, maximizing R^2 on full quadratic fits using ordinary least squares (OLS). We then used backward selection to drop terms, dropping a term if the resulting increase in the residual sum of squares was less than 50%. Models for outputs that shared the same active variables resulting from this process were augmented with extra terms so that all models with the same active variables have the same terms. In two cases we intervened in the model-building process, adding extra active variables to ensure a smooth model time structure.

After completing this process, we were left with only three different sets of active variables, involving only four of the permeability multipliers. Two variables were in all three sets and were the only variables in the largest of the three sets, accounting for 39 of the 64 components. The other two sets each contained one extra variable. This paucity of important input variables is part of the reason why we require only six simulator runs to effect a substantial reduction in forecast variance, as we discuss in Section 8. The choices of active variables were generally in accordance with the engineer's beliefs.

We used our OLS-fitted models on the coarse simulator data to set the prior means and variances for the β coefficients in $h_k(x_k^*) = \beta_k^T g_k(x_k^*)$. We consider each monomial term separately, and do not correlate coefficients across monomial terms, either within or between outputs. For any given monomial term, we have a sample mean $\bar{b}$ and a sample variance s^2 taken from the coarse simulator estimates $b_1, \ldots, b_m$, over the m outputs that contain the specified term. Letting $\tilde{\beta}_1, \ldots, \tilde{\beta}_m$ denote the corresponding fine simulator coefficients of the monomial, we set

$$\mathrm{E}[\tilde{\beta}_i] = \bar{b} + \rho(b_i - \bar{b})$$

and

$$\mathrm{cov}[\tilde{\beta}_i, \tilde{\beta}_{i'}] = s^2(1-\rho^2)/m + s^2(1-\rho^2)1_{i=i'}$$

for $i, i' = 1, \ldots, m$. This choice originates in a simple exchangeable representation for the two sets of coefficients. We set the value of the correlation parameter $\rho = .5$, although we also investigated both $\rho = 0$ and $\rho = .9$ to check the sensitivity of our results.

We chose to use autocorrelation functions r_k that are products of functions depending on the individual active variables. In the example, all of the input variables have the same range. For each active variable, we use the same autocorrelation function $\exp\{-20(x-x')^2\}$, which results in

$$r_k(x_k^* - x_k^{*\prime}) = \exp(-20\|x_k^* - x_k^{*\prime}\|^2),$$

an autocorrelation function which gives infinitely differentiable realisations. The value of $\theta = 20$ was arrived at by simulating random functions to find a value giving realizations that were fairly smooth but had substantial nonquadratic variation.

For each output, we estimated σ_ϵ^2 and σ_δ^2 as follows. Let ϵ and δ denote the vectors of coarse simulator values of $\epsilon(x)$ and $\delta(x)$ for the coarse simulator runs. The vector e of residuals from the OLS fit described earlier is simply $P(\epsilon + \delta)$, where P is the projection operator for the OLS fit. Thus $\mathrm{var}[e] = \sigma_\epsilon^2 PRP^T + \sigma_\delta^2 PIP^T$, where R is the known correlation matrix for the elements of ϵ. Because the two components of $\mathrm{var}[e]$ commute, they can be diagonalized simultaneously in the orthonormal basis $u_1, \ldots, u_n$ with eigenvalues $\sigma_\epsilon^2 \lambda_{1j}$ and $\sigma_\delta^2 \lambda_{2j}$. Thus $v_j = u_j^T e$ are independent with $\mathrm{E}[v_j] = 0$ and $\mathrm{var}[v_j] = \sigma_\epsilon^2 \lambda_{1j} + \sigma_\delta^2 \lambda_{2j}$, where the λ_{ij} are known a priori. This allows us to estimate σ_ϵ^2 and σ_δ^2 by regressing v_j^2 on λ_{1j} and λ_{2j} (without an intercept). Because $\mathrm{var}[v_j^2]$ is likely to be proportional to $\mathrm{E}[v_j^2]^2$, this can be performed using iteratively reweighted least squares; however, for this example we simply used weighted least squares using weights derived from $\sigma_\epsilon^2 = \sigma_\delta^2$. On those few occasions when one or other estimate was negative, we replaced it with 0.

Once σ_ϵ^2 and σ_δ^2 are known, it is easy to calculate the covariances required for Bayes linear adjustment of ϵ and δ by e. We carried out this calculation for each output using our estimates of σ_ϵ^2 and σ_δ^2 to yield estimates of ϵ and δ. We assigned the value of τ_{kl} to be the empirical correlation between the δ vectors for outputs k and l. Similarly, we used empirical correlations between ϵ vectors to assign the value of ν_{kl}, subject to the constraint that the correlation must be 0 when the active variables differ.

8. DESIGN FOR FORECASTING

When the simulator is slow, time constraints may allow only a small number of carefully chosen runs. However, Bayesian design for high-dimensional problems is notoriously intractable (see, e.g., O'Hagan 1978). We now describe a method for selecting runs for Bayes linear forecasting. Suppose that we sequentially choose input vectors $x_1, x_2, \ldots$ at which to run the simulator, where our goal is to optimize a forecast criterion based on some function Q of the adjusted forecast variance matrix in (10). For example, Q might be a weighted sum of its diagonal elements, which is equivalent to minimizing the trace, after standardizing each output variable on a scale that makes reductions in uncertainty comparable over different quantities and reflects the relative importance of good forecasts for the various future quantities. At stage $n+1$, we have n simulator output values $S = (s(x_1), \ldots, s(x_n))$ at n input choices $x_1, \ldots, x_n$ with which we evaluate the adjusted mean and variance of $s(x_0)$, given S, as in (7) and (8).

Let $Q(x, s(x))$ denote the value of the forecast criterion if we now choose input vector x and observe simulator output value $s(x)$. Let $\overline{Q}(x)$ denote the expected value of $Q(x, s(x))$ over $s(x)$, for fixed x. We want to choose x_{n+1} to be the value of x that minimizes $\overline{Q}(x)$. Because this minimization is computationally expensive, we use two approximations to simplify the design calculations.

First, we may approximate the expectation $\overline{Q}(x)$ by the value $Q(x, \mathbf{E}_S[s(x)])$. Second, much of the effort involved in computing $\overline{Q}(x)$ is expended on assessing the contribution to the forecast variance from error components $\epsilon(x^*)$. However, if we may select only a small number of runs, then usually our choices will be well separated, so that the correlation attributed to ϵ will be comparatively small. Thus often we can speed up the design calculations substantially by absorbing the ϵ terms

into the uncorrelated error components $\delta(x)$, with increased error variance contributions $\sigma_\epsilon^2 + \sigma_\delta^2$; Craig et al. (1996) used a similar approximation when designing for simulator calibration. With this approximation, we do not update beliefs about the $\epsilon + \delta$ surface given previous simulator runs, S, when evaluating the design criterion, and Bayes linear updating of beliefs over $s(x)$ given S simplifies to the corresponding Bayes linear update for the β coefficients in the polynomial $h(x^*)$. Details of these design calculations are given in the Appendix. When we have chosen a design point using the approximate criterion and observed the simulator output, then, for this particular choice, we may carry out the full evaluation of the criterion to monitor the quality of our approximations. For example, a possible problem in ignoring the ϵ correlations could be that the chosen design point might be too close to an existing point, in which case we should force the new design choice away from current design choices.

Even with the suggested simplifications, the design search still may require minimization of a complicated nonlinear function over a high-dimensional space. If there is a relatively low-dimensional subspace of x that accounts for most of the variability in $s(x)$, [e.g., if a few components of x are active for most components of $s(x)$], then we may evaluate the design criterion at each point on a lattice or latin hypercube design (Owen 1994) over these active components to select a good choice for x_{n+1}.

We used such a latin hypercube for our example. Figure 7 illustrates our design procedure over six runs of the simulator. Our criterion Q is the trace of the adjusted forecast variance matrix. Figure 7 gives the range of values of the criterion on a 101-point hypercube at each step, plus the actual value obtained after running the simulator at the chosen value. This allows us to verify that the approximation we make in our model of $s(x)$ is reasonable in this case. It also shows the theoretical lower bound of the criterion that could be achieved, as given in (11), and the approximate lower bound that follows from our choice of simulator model. Here we take Ξ in Section 5.4 to be the matrix with entries $\tau_{kl}\sigma_{\delta k}\sigma_{\delta l}$, which is the covariance matrix of the white noise components of the outputs. After about six runs, we see that the design criterion has become very flat across the hypercube, and that further runs will only very slowly decrease the criterion that appears to be close to its lower bound.

In other cases, a practical alternative to select x_{n+1} may be as follows. We search iteratively on the individual components of x. We hold every component of x fixed but one, varying that remaining component over a dynamically chosen grid. For each choice, we evaluate our design criterion and fix that component at the best choice. We now vary another component of x. We continue to iterate this process over components until some search criterion has been met; for example, when the design evaluation has stabilized. If it is expensive to carry out the full iterative calculations for each potential design choice, then when we vary each component of x, we might include in the forecast criterion only those elements of y for which the component is active. This greatly reduces the calculations for almost all components of x. For some components of x, we may choose to eliminate still more elements of $s(x)$, using such criteria as retaining those components of $s(x)$ with the largest current forecast errors and least intergroup correlations. We may search similarly over several coordinates at the same time, if the components occur together as active variables for many components of $s(x)$.

If the design calculations take a long time, then we might run them in parallel with simulator runs, so that the simulator run at the next design point can be made as soon as possible after its current run is finished.

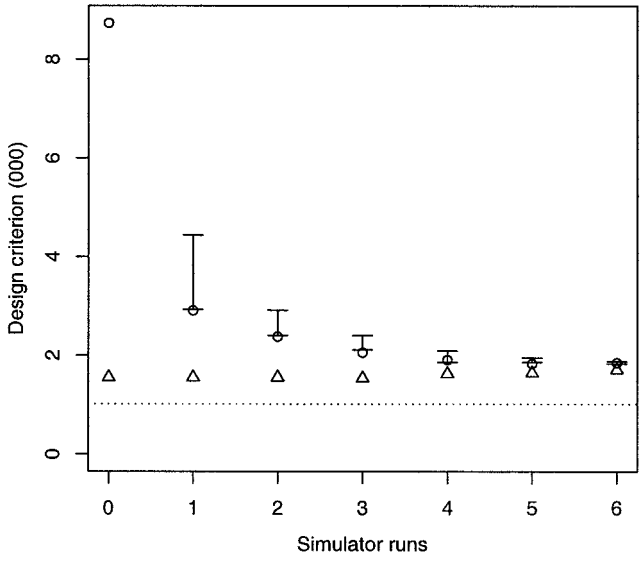

Figure 7. Sequential Design. The criterion is the trace of the adjusted forecast variance matrix. The vertical bar at simulator run i represents the range of criterion values computed using the design approximation at run i − 1, and the circle represents the actual value of the criterion after making run i at the input with the smallest criterion value. The triangle at run i shows our approximate lower bound for the criterion given our choice of simulator model.

9. DISCUSSION

We have described a general approach to forecasting using complex, high-dimensional computer models. We conclude with a brief discussion of further issues arising for our approach.

9.1 Forecasting and Refocusing

We model $y(x)$ over a region $\mathcal{X}$ of potential x_0 values, and evaluate a Bayesian or Bayes linear forecast given several runs of the simulator. If it appears from our runs that much of the region $\mathcal{X}$ can be eliminated, then a natural approach to improving forecast accuracy is to reduce the region to a subset $\mathcal{X}_1$. We may decide to choose quite different models to describe $y(x)$ over $\mathcal{X}_1$. We refer to the operation of reducing the region and refitting the models for $y(x)$ over $\mathcal{X}_1$ as refocusing. The potential improvement in forecast accuracy that may follow from refocusing is based partly on the reduction in uncertainty for x_0 and partly on the improvement in our modeling of $y(x)$ over the restricted range in x, as we may modify the coefficients in the various systematic terms, change the active inputs, and reduce the residual variation in our model fitting.

Therefore, a general procedure is to make several runs of the simulator, refocus on a subregion $\mathcal{X}_1$, refit our models,

sample further, then refocus onto a subregion $\mathcal{X}_2$, refit models again, and continue in this way until eventually we choose not to refocus but instead to forecast.

The refocusing step is based on identifying those elements of $\mathcal{X}$ for which the likelihood of the data is almost 0. The Bayesian analysis is essentially unaffected by removing from $\mathcal{X}$ all values of x with almost 0 likelihood, and rescaling prior beliefs over x_0 to integrate to 1 over $\mathcal{X}_1$ before carrying out the analysis. Therefore, to a close approximation we apply the Bayes analysis by carrying out the forecast calculations with respect to the posterior distribution restricted to $\mathcal{X}_1$.

The effect of refocusing for the Bayes linear forecast is as for the Bayesian version—namely, to remove from $\mathcal{X}$ all of the values that appear implausible (see Craig et al. 1996, 1997), scale up beliefs over $\mathcal{X}_1$ to integrate to 1, and then carry out the Bayes linear forecast calculations with respect to the rescaled prior distribution over the reduced region. Note in particular that choosing an initial uniform prior for x_0 over $\mathcal{X}$ results in a uniform prior over each reduced region, so that tractability over rectangular subregions is straightforward for this specification.

To implement this approach, we need ways to decide at each stage whether to choose runs of the simulator to improve forecasts or to refocus. Thus we need forecast targets and rules of thumb to see how close we are to attaining our targets at each stage. If it appears that we can achieve our forecasting objectives at the current stage, then we carry out the sequential design for forecasting, and, subject to diagnostic checking, make our forecast. Otherwise, we use a design intended for refocusing (for examples, see Craig et al. 1996, 1997). Having refocused, we repeat the foregoing assessments and continue in this way until our forecast aims are attained or until we run out of time.

9.2 The Status of x_0

Our approach is based around the existence of an input value x_0 for which the output $y(x_0)$ is sufficient for the simulator for forecasting the system outcome y. This is both a reasonable assumption when the simulator is a good representation of the system and simply a device that allows us to score each simulator output $y_F(x)$ as a potential predictor for y_F through the proximity of $y_P(x)$ and y_P. This assumption may be weakened in various ways while still preserving our approach; for example, we may not require that precisely the same value of x_0 be appropriate for each component of y. An alternative and more general approach is to construct a covariance structure between $y(x)$ and y that directly expresses the belief that small values of $y_P(x) - y_P$ are likely to correspond to small values of $y_F(x) - y_F$.

9.3 Learning About Variances

We have not described how the residual variances σ_ϵ^2 and σ_δ^2 are modified as we make runs on the simulator. In the Bayesian analysis, we may extend the prior specification over the values of these variances, although this may complicate the resulting analyses significantly. Wilkinson and Goldstein (1996) described how to update the variances directly by Bayes linear fitting using residual sums of squares given relevant fourth-moment specifications. Such modifications might be important if we may make many runs of the simulator. But, we are concerned primarily with simulators for which we do not make many runs, so that there is relatively sparse information on residual variation. In such cases, it is arguably better to make careful prior estimates for variation (e.g., using multiple runs on fast versions of the simulator) and then carefully monitor all of the model diagnostics to identify gross inadequacies in the prior variance specifications.

9.4 Different Approaches to Forecasting

The general problem of forecasting using computer models covers an enormous range of situations. A number of considerations will always be relevant, including how much past data we have for forecasting and for calibrating the simulator, how reliable a representation of the system we believe the simulator to be, how many runs of the simulator we can make in the available time, and the complexity of the input and the output spaces of the simulator. The lower the dimensions, the more feasible it is to carry out a full Bayesian analysis with a realistic prior distribution. The larger the number of runs that we may make, then the more we may be guided by essentially data-analytic approaches to forecasting, particularly if the simulator is a fairly accurate representation of the system and we have little external prior information.

We have described both the Bayesian and the Bayes linear approach to forecasting. The former approach should work well for low-dimensional problems. The latter approach should give a reasonable approximation to the Bayesian analysis for such problems, while effectively becoming the most feasible approach in high-dimensional problems, particularly when we may make only a relatively small number of simulator runs.

APPENDIX: COMPUTATION

Here we set out the calculations explicitly and show how certain simplifications arise for a uniform prior distribution for x_0 and a particular covariance structure for $\epsilon(x)$. Recall that we write the model of the simulator as the sum of three mutually independent terms,

$$s(x) = B\,g(x) + \epsilon(x) + \delta(x), \tag{A.1}$$

where $s(x)$ is a k-vector of outputs, x is a p-vector of inputs, B is a $k \times q$ matrix of unknown coefficients with known mean and covariance, $g(x)$ is a q-vector of known functions of the components of x (possibly including the constant), and $\epsilon(x)$ and $\delta(x)$ are zero mean random k-vectors with known covariance structures. For simplicity, we do not distinguish the different sets of "active" variables in the components of $\epsilon(x)$ at this stage, and we take $g(x)$ to be the union of all of the required terms, so that the mean and covariance of B are usually sparse. We write $H^{\epsilon}_{ii'}(x - x')$ for $\mathrm{cov}[\epsilon_i(x), \epsilon_{i'}(x')]$ and $H^{\delta}_{ii'}(x - x')$ for $\mathrm{cov}[\delta_i(x), \delta_{i'}(x')]$, which, by our modeling assumptions, equals $\tau_{ii'}\sigma_{\delta i}\sigma_{\delta i'}$ when $x = x'$ and is otherwise 0.

If we run the simulator at n points in the design matrix X, then we will have an output matrix S, and (A.1) becomes

$$S = B\,G^T + E + \Delta, \tag{A.2}$$

where S, E, and Δ each are $k \times n$ and $G = g(X)$ is the $n \times q$ transform of the design matrix. At this point we switch from matrix to tensor notation, using Einstein's convention of summing over all repeated

indices in product terms. We use the indices $i \in 1, \ldots, k$ for the output components; $r \in 1, \ldots, n$ for the simulator runs; and $m \in 1, \ldots, q$ for the terms in g, with primes to distinguish indices over the same set. Equation (A.2) then becomes $S_{ir} = B_{im} G_{rm} + E_{ir} + \Delta_{ir}$ with prior mean $\mathrm{E}[S_{ir}] = \mathrm{E}[B_{im}]G_{rm}$ and prior covariance $\mathrm{cov}[S_{ir}, S_{i'r'}] = G_{rm}\mathrm{cov}[B_{im}, B_{i'm'}]G_{r'm'} + H^{\epsilon}_{ii'}(x_r - x_{r'}) + H^{\delta}_{ii'}(x_r - x_{r'})$.

A.1 Prediction for a Known Input

Suppose that we wish to predict the simulator output at some input x having observed the outputs S, where x is not among the runs in X. We write the simulator model for given x as

$$s_i(x) = B_{im}\, g_m(x) + \zeta_i(x), \tag{A.3}$$

where $\zeta_i(x) = \epsilon_i(x) + \delta_i(x)$. We adjust our prior beliefs about B_{im} and $\zeta_i(x)$ by the outputs S using Bayes linear methods, which gives

$$\begin{aligned}
\mathrm{E}_S[B_{im}] &= \mathrm{E}[B_{im}] + \mathrm{cov}[B_{im}, S_{i'r}]R_{i'r}, \\
\mathrm{cov}_S[B_{im}, B_{i'm'}] &= \mathrm{cov}[B_{im}, B_{i'm'}] \\
&\quad - \mathrm{cov}[B_{im}, S_{i''r}]P_{i''ri'''r'}\,\mathrm{cov}[B_{i'm'}, S_{i'''r'}], \\
\mathrm{E}_S[\zeta_i(x)] &= H^{\epsilon}_{ii'}(x - x_r)\, R_{i'r}, \\
\mathrm{cov}_S[\zeta_i(x), \zeta_{i'}(x)] &= H^{\epsilon}_{ii'}(0) + H^{\delta}_{ii'}(0) \\
&\quad - H^{\epsilon}_{ii''}(x - x_r)\, P_{i''ri'''r'}\, H^{\epsilon}_{i'i'''}(x - x_{r'}),
\end{aligned}$$

and

$$\mathrm{cov}_S[B_{im}, \zeta_{i'}(x)] = -\mathrm{cov}[B_{im}, S_{i''r}]P_{i''ri'''r'}\, H^{\epsilon}_{i'i'''}(x - x_{r'}).$$

Note that $\mathrm{cov}[B_{im}, S_{i'r}] = \mathrm{cov}[B_{im}, B_{i'm'}]G_{rm'}$, and we define $R_{ir} = P_{iri'r'}\big(S_{i'r'} - \mathrm{E}[S_{i'r'}]\big)$ and P to be the inverse of $\mathrm{var}[S]$; that is,

$$P_{iri'r'}\,\mathrm{cov}[S_{i'r'}, S_{i''r''}] = \begin{cases} 1 & \text{if } i = i'' \text{ and } r = r'' \\ 0 & \text{otherwise.} \end{cases}$$

Then we substitute these expressions into the mean and covariance of (A.3) to give, for any x not in the design matrix X, the expressions corresponding to (5) and (6) in the text,

$$\mu_i(x) = \big\{\mathrm{E}[B_{im}] + \mathrm{cov}[B_{im}, S_{i'r}]R_{i'r}\big\}g_m(x) + R_{i'r}H^{\epsilon}_{ii'}(x - x_r) \tag{A.4}$$

and

$$\begin{aligned}
\Sigma_{ii'}(x) = {} & \big\{\mathrm{cov}[B_{im}, B_{i'm'}] - \mathrm{cov}[B_{im}, S_{i''r}] \\
& \quad \times P_{i''ri'''r'}\,\mathrm{cov}[B_{i'm'}, S_{i'''r'}]\big\}g_m(x)g_{m'}(x) \\
& + H^{\epsilon}_{ii'}(0) + H^{\delta}_{ii'}(0) \\
& - P_{i''ri'''r'}\, H^{\epsilon}_{ii''}(x - x_r)H^{\epsilon}_{i'i'''}(x - x_{r'}) \\
& - \mathrm{cov}[B_{im}, S_{i''r}]P_{i''ri'''r'}\, H^{\epsilon}_{i'i'''}(x - x_{r'})g_m(x) \\
& - \mathrm{cov}[B_{i'm}, S_{i''r}]P_{i''ri'''r'}\, H^{\epsilon}_{ii'''}(x - x_{r'})g_m(x).
\end{aligned} \tag{A.5}$$

We may generalize these calculations to compute predictions over a collection of x-values, for which we additionally compute covariances between components of $s(x)$ at different x-values.

A.2 Prediction for an Unknown Input

The mean and variance of $s(x)$ when x is the unknown x_0 are computed by integrating out x according to the prior distribution of x_0, to give the general results

$$\mu_i = \mathrm{E}[\mu_i(x_0)|S]$$

and

$$\Sigma_{ii'} = \mathrm{var}[\mu_i(x_0)|S] + \mathrm{E}[\Sigma_{ii'}(x_0)|S].$$

In what follows all expectations and variances are over x with respect to the prior distribution of x_0. Inspecting (A.4) and (A.5), we see that we actually require all components of the mean and the covariance of the joint vector

$$\begin{aligned}
\Upsilon = \big(& g_1(x), \ldots, g_q(x), H^{\epsilon}_{11}(x - x_1), H^{\epsilon}_{12}(x - x_1), \\
& \ldots, H^{\epsilon}_{kk}(x - x_1), H^{\epsilon}_{11}(x - x_2), \ldots, H^{\epsilon}_{kk}(x - x_n)\big),
\end{aligned}$$

which is a vector of length $q + k^2 \times n$.

For the mean μ_i, we require the expectations of $g_m(x)$ and $H^{\epsilon}_{ii'}(x - x_r)$, which we substitute into (A.4) to compute the expression (7) in the text. For the expectation of the covariance, we require the expectation of the product terms $g_m(x) \times g_{m'}(x)$, $g_m(x) \times H^{\epsilon}_{ii'}(x - x_r)$ and $H^{\epsilon}_{ii'}(x - x_r) \times H^{\epsilon}_{i''i'''}(x - x_{r'})$, which we substitute into (A.5). For the covariance of the expectation, we must take the outer product of (A.4), into which we substitute the covariances $\mathrm{cov}[g_m(x), g_{m'}(x)]$, $\mathrm{cov}[H^{\epsilon}_{ii'}(x - x_r), H^{\epsilon}_{i'i'''}(x - x_{r'})]$, and $\mathrm{cov}[g_m(x), H^{\epsilon}_{ii'}(x - x_r)]$. These latter terms are all easily constructed from the previous calculations. After combining the covariance of the expectation and the expectation of the covariance, we can compute expression (8) in the text.

A.3 Computable Cases

The scale of the calculation for finding μ and Σ is effectively proportional to the squared length of Υ. In our example we have $q = 15$, $k = 64$, and $n = 6$, which would require 80,199 separate numerical integrations for arbitrary functional representations of g_m and $H^{\epsilon}_{ii'}$. However, we may substantially reduce the scale of the calculation through two approaches.

First, we can reduce the effective length of Υ. Υ mainly comprises terms of the form $H^{\epsilon}_{ii'}(x - x_r)$. As it stands, we have a full outer product of pairs of output components and simulator runs. But if each $H^{\epsilon}_{ii'}(x - x_r)$ has a product form in which the x component is the same for a collection of ii' components, then we need only n terms per collection rather than n terms for all k^2 individual terms. As detailed in the text, our choice is

$$H^{\epsilon}_{ii'}(x - x_r) = \begin{cases} \sigma^{\epsilon}_{ii'} r(\|x - x_r\|_{\mathcal{A}_i}) & \text{if } \mathcal{A}_i = \mathcal{A}_{i'} \\ 0 & \text{otherwise,} \end{cases}$$

where we write $\sigma^{\epsilon}_{ii'} = \nu_{ii'}\sigma_{\epsilon i}\sigma_{\epsilon i'}$, let $\mathcal{A}_i \subset \{1, \ldots, p\}$ denote the indices of the active variables of output i, and let $\|\cdot\|_{\mathcal{A}_i}$ denote the norm in the active variables of output i. The terms $\sigma^{\epsilon}_{ii'}$ enter multiplicatively once we have accounted for the different sets of active variables, and so the effective number of H components in Υ is reduced to n terms per set of active variables. The three sets of active variables in our example cause a reduction from 384 components to 18 components in Υ. Now, only 594 numerical integrations remain to be performed.

Second, we may reduce the complexity of the individual integrations and, if possible, find analytic or tabulated expressions in place of numerical integrations. Substantial simplifications follow from using monomial expressions in g, using an expression for r that factorizes in x and using a prior in which components of x_0 are independent. In this case the joint integrals over x_0 factor into products

of simple integrals in individual components. In our example we use the Gaussian correlation function

$$r(\|x-x_r\|_{\mathcal{A}_i}) = \exp\{-\theta\textstyle\sum_{j\in\mathcal{A}_i}(x_{0j}-x_{rj})^2\} = \prod_{j\in\mathcal{A}_i}\exp\{-\theta(x_{0j}-x_{rj})^2\}$$

and assume a uniform rectangular prior on x_0 (after transforming the original values). This allows us to find analytic expressions for the means of $g_m(x)$ and $g_m(x)\times g_{m'}(x)$, and to find expressions for all other terms from the values of the Gaussian density. We write

$$\psi(v,v',\theta)=\exp\{-\theta(v-v')^2\}$$

and

$$\Psi(v,v',\theta)=\int_{-\infty}^{v}\psi(v'',v',\theta)\,dv''$$

for scalar v. Then the expectation of the product terms $H^{\epsilon}_{ii'}(x-x_r)\times H^{\epsilon}_{i''i'''}(x-x_{r'})$ becomes the integral of the product of terms such as $\psi(v,v',\theta)\times\psi(v,v'',\theta)$, which can be found from values of the Gaussian density function after completing the square in $(v-v')^2+(v-v'')^2$. All of the expectations of the compound terms $g_m(x)\times H^{\epsilon}_{ii'}(x,x_r)$ can be found in terms of the functions ψ and Ψ using the recurrence relations

$$\int v\psi\,dv = v'\Psi-\frac{1}{2\theta}\psi,$$

$$\int v^2\psi\,dv = v'\int v\psi\,dv-\frac{1}{2\theta}(v'\psi-\Psi),$$

and so on.

It is also straightforward to combine the Gaussian correlation function with monomial g and Gaussian prior beliefs for x_0, where the components of x_0 need not be independent.

A.4 Design Calculations

The design calculations suggest the next run of the simulator, given information about B and about ζ based on output from all of the runs to date, S. These calculations are numerically intensive, so we use a simplified model in which S is used to update B alone. Thus we use the approximation $s_i(x)=\overline{B}_{im}g_m(x)+\bar{\zeta}_i(x)$, where we take the adjusted beliefs $\mathrm{E}_S[B_{im}]$ and $\mathrm{cov}_S[B_{im},B_{i'm'}]$ as means and covariances for $\overline{B}$, and where $\mathrm{E}[\zeta_i(x)]=0$ and $\mathrm{cov}[\zeta_i(x),\zeta_{i'}(x')]=\nu_{ii'}\sigma_{\epsilon i}\sigma_{\epsilon i'}+\tau_{ii'}\sigma_{\delta i}\sigma_{\delta i'}$ when $x=x'$ and otherwise is 0.

We evaluate a potential design point w according to its impact through the adjusted variance of $\overline{B}$ in three stages. First, we adjust the variance of $\overline{B}$ supposing that we observed a run on the simulator at w, giving

$$\mathrm{cov}_{s(w)}[\overline{B}_{im},\overline{B}_{i'm'}]=\mathrm{cov}[\overline{B}_{im},\overline{B}_{i'm'}] - \mathrm{cov}[\overline{B}_{im},s_{i''}(w)]P_{i''i'''}\mathrm{cov}[\overline{B}_{i'm'},s_{i'''}(w)],$$

where $\mathrm{cov}[\overline{B}_{im},s_{i''}(w)]=\mathrm{cov}[\overline{B}_{im},\overline{B}_{i''m'}]g_{m'}(w)$ and P is the inverse of the prior variance matrix of $s(w)$ using our approximation. Second, we compute the variance of $s(x_0)$ using the adjusted variance of $\overline{B}$ as

$$\mathrm{E}_{s(w)}[\overline{B}_{im}]\mathrm{cov}[g_m(x_0),g_{m'}(x_0)]\mathrm{E}_{s(w)}[\overline{B}_{i'm'}] + \mathrm{cov}_{s(w)}[\overline{B}_{im},\overline{B}_{i'm'}]\mathrm{E}[g_m(x_0)g_{m'}(x_0)] + \nu_{ii'}\sigma_{\epsilon i}\sigma_{\epsilon i'}+\tau_{ii'}\sigma_{\delta i}\sigma_{\delta i'},$$

although we substitute $\mathrm{E}[\overline{B}_{im}]$ for $\mathrm{E}_{s(w)}[\overline{B}_{im}]$, as we cannot know the latter without running the simulator. Finally, we compute the adjusted variance for the "future" values of $s(x_0)$ using the past values, via (10). The design point w can then be scored by some function of this posterior predictive variance; in the example we use the trace.

[Received December 1999. Revised September 2000.]

REFERENCES

Bates, R. A., Buck, R. J., Riccomagno, E., and Wynn, H. P. (1996), "Experimental Design and Observation for Large Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 58, 77–94.

Beven, K. J., and Binley, A. M. (1992), "The Future of Distributed Models: Model Calibration and Uncertainty Prediction," *Hydrological Processes*, 6, 279–298.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1996), "Bayes Linear Strategies for History Matching of Hydrocarbon Reservoirs" (with discussion), in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 69–95.

———, (1997), "Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments" (with discussion), in *Case Studies in Bayesian Statistics: Volume III*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, and N. D. Singpurwalla, New York: Springer-Verlag, pp. 36–93.

———, (1998), "Constructing Partial Prior Specifications for Models of Complex Physical Systems" (with discussion), *Journal of the Royal Statistical Society,* Ser. D, 47, 37–53.

Cressie, N. A. C. (1991), *Statistics for Spatial Data*, New York: Wiley.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, 86, 953–963.

Galway, L., and Lucas, T. (1997), Comments on "Pressure Matching for Hydrocarbon Reservoirs" by Craig, Goldstein, Seheult, and Smith, in *Case Studies in Bayesian Statistics: Volume III,* eds. C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, and N. D. Singpurwalla, New York: Springer-Verlag, pp. 87–91.

Goldstein, M. (1999), "Bayes Linear Analysis," in *Encyclopedia of Statistical Sciences: Update Volume 3,* eds. S. Kotz, C. B. Read, and D. L. Banks, New York: Wiley, pp. 29–34.

Kennedy, M. C., and O'Hagan, A. (2000), "Predicting the Output From a Complex Computer Code When Fast Approximations are Available," *Biometrika*, 87, 1–13.

Morris, M. D., and Mitchell, T. J. (1995), "Exploratory Designs for Computational Experiments," *Journal of Statistical Planning and Inference*, 43, 381–402.

O'Hagan, A. (1978), "On Curve Fitting and Optimal Design for Regression" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 40, 1–42.

———, (1992), "Some Bayesian Numerical Analysis" (with discussion), in *Bayesian Statistics 4,* eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 345–363.

O'Hagan, A., Kennedy, M. C., and Oakley, J. E. (1999), "Uncertainty Analysis and Other Inference Tools for Complex Computer Codes" (with discussion), in *Bayesian Statistics 6,* eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 503–524.

Owen, A. B. (1994), "Controlling Correlations in Latin Hypercube Samples," *Journal of the American Statistical Association*, 89, 1517–1522.

Sacks, J., Schiller, S. B., and Welch, W. J. (1992), "Design for Computer Experiments," *Technometrics*, 31, 41–47 .

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments" (with discussion), *Statistical Science*, 4, 409–435.

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.

Wilkinson, D. J., and Goldstein, M. (1996), "Bayes Linear Adjustment for Variance Matrices," in *Bayesian Statistics 5,* eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 791–799.