Bayes Linear Uncertainty Analysis for Oil Reservoirs Based on Multiscale Computer Experiments

Jonathan Cumming and Michael Goldstein

Department of Mathematical Sciences, Durham University, U.K.

June 26, 2008

1 Introduction

Reservoir simulators are important and widely-used tools for oil reservoir management. These simulators are computer implementations of high-dimensional mathematical models for reservoirs, where the model inputs are physical parameters, such as the permeability and porosity of various regions of the reservoir, the extent of potential faults, aquifer strengths and so forth. The outputs of the model, for a given choice of inputs, are observable characteristics such as pressure readings, oil and gas production levels and so forth, for the various wells in the reservoir.

Usually, we are largely uncertain as to the physical state of the reservoir, and thus we are unsure as to appropriate choices of the input parameters for a reservoir model. Therefore, an uncertainty analysis for the model often proceeds by first calibrating the simulator against observed production history at the wells and then using the calibrated model to forecast future well production, and therefore to act as an information tool for the efficient management of the reservoir.

In a Bayesian analysis, all of our uncertainties are incorporated into our system forecasts. In addition to the uncertainty about the input values, there are three other basic sources of uncertainty. First, although the simulator is deterministic, an evaluation of the simulator for a single choice of parameter values can take hours or days, so that the function is unknown to us, except at the subset of values which we have chosen to evaluate. Secondly, the reservoir simulator, even at the best choice of input values, is only a model for the reservoir, and we must therefore take into account the discrepancy between the model and the reservoir. Finally, the historical data which we are calibrating against is observed with error.

This problem is typical of a very wide and important class of problems each of which may be broadly described as an uncertainty analysis for a complex physical system based on a model for the system. Such problems arise in almost all areas of scientific enquiry; typical examples are the use of climate models to study climate change, or universe simulation models to explore the origins and generating principles of the universe. In all such applications, we must deal with the same four basic types of uncertainty, namely input uncertainty, function uncertainty, model discrepancy and observational error. A general methodology has been developed to deal with this class of problems. Our aim, in this article, is to provide an introduction to this methodology and to show

1

how it may be applied for a reservoir model of realistic size and complexity. We shall therefore analyse a particular problem in reservoir description, based upon a description of our general approach to uncertainty analysis for complex models. In particular, we will highlight the value of fast approximate versions of the computer simulator for making informed prior judgements relating to the form of the full simulator. Our account is based on the use of Bayes linear methodology for simplifying the specification and analysis for complex high dimensional problems, and so this article also serves as an introduction to the general principles of this approach.

2   Uncertainty analysis for complex physical models

2.1   General formulation

We now describe a general formulation for our approach. There is a collection, x^+, of system properties. These properties influence system behaviour, as represented by a vector of system attributes, $y = (y_H, y_P)$, where y_H is a collection of historical values and y_P is a collection of values that we may wish to predict. We have an observation vector, z_H, on y_H. We write

$$z_H = y_H + e \tag{1}$$

and suppose that the observational error, e, is independent of y with $E(e) = 0$. For the reservoir example, x^+ is a detailed description of the reservoir geology (initially unknown), y_H, y_P are vectors of past and future production values at the various wells, and z_H is the observed well log record for y_H.

Ideally, we would like to construct a deterministic computer model, $f(x) = (f_H(x), f_P(x))$, embodying the laws of nature, which satisfies $y = f(x^+)$. In practice, however, our actual model f usually simplifies the physics and approximates the solution of the resulting equations. Therefore, our uncertainty description must allow both for the possible differences between the physical value of x^+ and the best choice of inputs to the simulator, and also for the discrepancy between the model outputs, evaluated at this best choice, and the true values of the system attributes, y.

Therefore, we must make explicit our assumptions relating the computer model $f(x)$ and the physical system, y. In general, this will be problem dependent. The simplest and most common way to relate the simulator and the system is the so-called "Best Input Approach". We proceed as though there exists a value x^*, independent of the function f, such that the value of $f^* = f(x^*)$ summarises all of the information that the simulator conveys about the system, in the following sense. If we define the *model discrepancy* as the difference between y and f^*, so that

$$y = f^* + \epsilon \tag{2}$$

then our assumption is that ϵ is independent of both f and x^*. (Here, and onwards, all probabilistic statements relate to the uncertainty judgements of the analyst.) For some models, this assumption will be justified as we can identify x^* with the true system values x^+. In other cases, this should be viewed more as a convenient simplifying assumption which we consider to be approximately true because of an approximate identification of this type. For many problems, whether this formulation is appropriate is, itself, the question of interest; for example, in large scale universe simulations, the fundamental question is whether the underlying physics has been correctly captured, and one way of testing this is to see whether there are any choices of inputs which result in model outputs which are in approximate correspondence with our observed universe.

For a careful discussion of the status of the best input approach, and a more general formulation of the nature of the relationship between simulators and physical systems, see Goldstein and Rougier (2007).

2.2 Bayes linear uncertainty analysis for complex models

Our overall aim is to tackle previously intractable problems arising from the uncertainties inherent in imperfect computer models of highly complex physical systems, using a Bayesian formulation. This involves a specification for (i) the prior probability distribution for best input x^*, (ii) a probability distribution for the computer function f, (iii) a probabilistic discrepancy measure relating $f(x^*)$ to the system y, (iv) a likelihood function relating historical data z to y.

This full probabilistic description provides a formal framework to synthesise expert elicitation, historical data and a careful choice of simulator runs. From this synthesis, we aim to learn about appropriate choices for the simulator inputs and to assess, and possibly to control, the future behaviour of the system. For problems of moderate size, this approach is appropriate, practical and highly effective (Kennedy and O'Hagan, 2001; Santner, Williams, and Notz, 2003). As the scale of the problem increases, however, the full Bayes analysis becomes increasingly difficult because (i) it is difficult to give a meaningful full prior probability specification over high dimensional spaces; (ii) the computations, for learning from data (observations and computer runs), particularly in the context of choosing informative sets of input values at which to evaluate the simulator, become technically difficult and extremely computer intensive; (iii) the likelihood surface tends to be very complicated, so that full Bayes calculations may become highly non-robust.

However, the idea of the Bayesian approach, namely capturing our expert prior judgements in stochastic form and modifying them by appropriate rules given observations, is conceptually appropriate, and indeed there is no obvious alternative. In this article, we therefore describe the Bayes linear approach to uncertainty analysis for complex models.

The Bayes linear approach is (relatively) simple in terms of belief specification and analysis, as it is based only on the mean, variance and covariance specifications for the system. These specifications are made directly as, following de Finetti (De Finetti, 1974, 1975), we take expectation as our primitive quantification of uncertainty. The adjusted expectation and variance for a random vector y, given random vector z, are as follows.

$$E_z(y) = E(y) + C(y,z)V(z)^{-1}(z - E(z)), \tag{3}$$
$$V_z(y) = V(y) - C(y,z)V(z)^{-1}C(z,y) \tag{4}$$

(If $V(z)$ is not invertible, then an appropriate generalised inverse is used in the above forms.)

For the purpose of this account, we may either view Bayes linear analysis as a simple approximation to a full Bayes analysis, or as the appropriate analysis given a partial specification based on expectation. We give more details of the rationale and practicalities of the Bayes linear approach in the appendix to this chapter. For a detailed treatment, see Goldstein and Wooff (2007).

2.3 Representing beliefs about f using emulators

In many problems involving high dimensional computer simulators of complex processes, the simulator is sufficiently slow to evaluate that we may only make a relatively small number of evaluations of the simulator at different choices of input values, x. Therefore, we must express our uncertainty about the value of the function for all input values at which we have not evaluated the simulator.

A probabilistic belief specification for a deterministic function is often referred to as an *emulator*. A common form for such an emulator, for component i of f, takes the following form:

$$f_i(x) = \sum_j \beta_{ij} \, g_{ij}(x) + u_i(x) \tag{5}$$

In this formulation, $B = \{\beta_{ij}\}$ are unknown scalars, and g_{ij} are known deterministic functions of x. The component $Bg(x)$ expresses the global variation in f, namely that portion of the variation in $f(x)$ which we can resolve without having to make evaluations for f at input choices which are near to x. $u(x)$ expresses local variation, and is often taken to be a weakly stationary stochastic process with constant variance for which $\text{Corr}(u_i(x), u_i(x'))$ is a function of $\|x - x'\|$. Within the Bayes linear formulation, we need only specify the mean, variance and covariance across the elements of B and the variance and covariance function for each $u_i(x)$. For a full Bayesian analysis, we make a full prior specification for these quantities, often choosing a Gaussian form. There is an extensive literature on functional emulation (Sacks, Welch, Mitchell, and Wynn, 1989; Currin, Mitchell, Morris, and Ylvisaker, 1991; Craig, Goldstein, Rougier, and Seheult, 2001; Santner et al., 2003; O'Hagan, 2006).

In order to fit an emulator, we must choose the functions $g_{ij}(x)$, specify prior uncertainties for the coefficient matrix B and update these by carefully chosen evaluations of the simulator, and choose an appropriate form for the local variation $u(x)$. Often, we discover that global variation, for output component f_i is determined by a relatively small subset, $x_{[i]}$ say, of the input quantities. In such cases, we may often be able to further simplify our emulation, as

$$f_i(x) = \sum_j \beta_{ij} \, g_{ij}(x_{[i]}) + u_i(x_{[i]}) + v_i(x) \tag{6}$$

where $u_i(x_{[i]})$ is a stationary process in $x_{[i]}$ and $v_i(x)$ is a "nugget" expressing all of the residual variation which is attributable to the remaining inputs, which we treat as a function with zero correlation length. When variation in this residual term is small, and the number of remaining inputs is large, this simplification enormously reduces the dimension of the computations that we must make, while usually having only a small impact on the accuracy of our results.

If the simulator is quick to evaluate, emulator choice may be made on the basis of a very large collection of simulator evaluations. If simulator evaluations are more costly, then we must rely on prior information to direct the choice of evaluations and the form of the collection $G = \cup_{i,j}\{g_{ij}(\cdot)\}$. A common source of prior information in such problems is access to a fast approximate version of the simulator. We may make many runs of the fast simulator, which allows us to develop a preliminary view of the form of the function, and therefore to make a preliminary choice of the function collection G and to suggest an informed prior specification for the random quantities that determine the emulator for f. This specification is then updated by careful choice of evaluations of the full simulator, supported by a diagnostic analysis, for example based on looking for systematic structure in the emulator residuals. In our example, there are certain natural coarsenings of the grid over which the reservoir simulator is evaluated which provide a useful starting point for such an analysis.

The emulator expresses prior uncertainty judgements about the function. From the emulator, we may extract those aspects of our judgements about the function that we require. In the Bayes linear approach, we are particularly concerned with the mean, variance and covariance for the function, at each input value x, and these are determined by the first and second order specifications for each of the terms of the emulator.

2.4   History matching and calibration

Most models go through a series of iterations before they are judged to give an adequate representation of the physical system. This is the case for reservoir simulators, where a key stage in assessing simulator quality is termed *history matching*, namely identifying the set X^* of possible choices of x^* for the reservoir model (i.e. those choices of input geology which give a sufficiently good fit to historical observations, relative to model discrepancy and observational error). If our search reveals no possible choices for x^*, this is usually taken to indicate structural problems with the underlying model, provided that we can be reasonably confident that the set X^* is indeed empty. This can be difficult to determine, as the input space over which we must search may be very high dimensional, the collection of outputs over which we may need to match may be very large, and each single function evaluation may take a very long time. We now describe the method followed in Craig, Goldstein, Seheult, and Smith (1997). For comparison, a direct Bayesian approach to model calibration is described in Kennedy and O'Hagan (2001).

History matching is based on the comparison of simulator output with historical observations. If we evaluate the simulator at a value, x, then we can judge whether x is a member of X^* by comparing $f(x)$ with data z. We do not expect an exact match, due to observational error and model discrepancy, and so we only require a match at some specified tolerance, often expressed in terms of the number of standard deviations between the function evaluation and the data. In practice, we cannot usually make a sufficient number of function evaluations to determine X^* in this way. Therefore, using the emulator, we obtain, for each x, the values $E(f(x))$ and $V(f(x))$. We seek to rule out regions of x space for which we expect that the evaluation $f(x)$ is likely to be a very poor match to observed z.

For a particular choice of x, we may assess the potential match quality, for a single output f_i, by evaluating

$$I_{(i)}(x) = \frac{|E(f_i(x)) - z_i|^2}{V(f_i(x) - z_i)}, \tag{7}$$

which we term the *implausibility* that $f_i(x)$ would give an acceptable match to z_i. For given x, implausibility may be evaluated over the vector of outputs, or over selected sub-vectors or over a collection of individual components. In the latter case, the individual component implausibilities must be combined, for example by using

$$I_M(x) = \max_i I_{(i)}(x). \tag{8}$$

We may identify regions of x with large $I_M(x)$ as implausible, i.e. unlikely to be good choice for x^*. These values are eliminated from our set of potential history matches, X^*.

If we wish to assess the potential match of a collection of q outputs f, we use a multivariate implausibility measure analogous to (7) given by

$$I(x) = \frac{(E(f(x)) - z)^T V(f(x) - z)^{-1} (E(f(x)) - z)}{q}, \tag{9}$$

where $I(x)$ is scaled to have expectation 1 if we set $x = x^*$. Unlike (7), the calculation of $I(x)$ from (9) requires the specification of the full covariance structure between all components of z and f, for any pair of x values.

We refocus our analysis, by re-sampling and refitting our emulator over the reduced set X^* and repeat the process until either no further refocusing is possible or we have exhausted our budget. This process is a form of iterative global search aimed at finding all choices of x^* which would give good fits to historical data. In a typical sequence, the early stages of the process are based on matching those outputs which have simple and informative global terms. As the space X^* shrinks, it often becomes progressively more straightforward to emulate functional outputs which were difficult to represent over the original parameter space. Therefore, we are often able to introduce additional constraints at each refocusing step. If the set X^* does turn out to be empty, then this is indicative of a problem, either with the data or the discrepancy judgements or the computer code or the chosen ranges and choices of parametrisation for the model inputs. It is common for several of these features to undergo various changes during the history matching process, and we should therefore be cautious in considering model calibration until the history matching phase has been completed successfully.

Bayesian calibration, namely placing a posterior probability distribution on the "true value" of x^*, is meaningful to the extent that the notion of a true value for x^* is meaningful. In such cases, we may make a direct Bayesian evaluation over the reduced region X^*, based on careful sampling and emulation within this region. If our history matching has been successful, the space over which we must calibrate will have been sufficiently reduced that calibration should be tractable and effective, provided our prior specification is sufficiently careful. As a simple approximation to this calculation, we may re-weight the values in this region by some function of our implausibility measure. The Bayes linear approach to prediction that we shall now describe does not need a calibration stage and so may be used directly following a successful history match.

2.5 Bayes linear prediction

We wish to predict the collection y_P of future well production using the observed well history z_H. This is achieved by making an appropriate specification for the joint mean and variance of the collection (y_P, z_H), and so our prediction for y_P using the history z_H is the adjusted expectation and variance of y_P given z_H. This Bayes linear approach to forecasting is discussed extensively in Craig et al. (2001).

The Bayes linear forecast equations for y_P given z_H are given by

$$E_{z_H}(y_P) = E(y_P) + C(y_P, z_H)V(z_H)^{-1}(z_H - E(z_H)), \tag{10}$$

$$V_{z_H}(y_P) = V(y_P) - C(y_P, z_H)V(z_H)^{-1}C(z_H, y_P). \tag{11}$$

Given the relations (1) and (2), we can express this forecast in terms of the "best" simulator run f^*, the model discrepancy ϵ, the observed history z_H, and the observational error e. From (2) we write the expectation and variance of y as $E(y) = E_F(f^*) + E(\epsilon)$ and $V(y) = V_F(f^*) + V(\epsilon)$, namely the adjusted expectation and variance of the best simulator run f^*, given the collection of available simulator evaluations F, plus the model discrepancy. For simplicity of presentation, we introduce the shorthand notation $\mu^* = E_F(f^*)$, $\Sigma^* = V_F(f^*)$, $\Sigma^\epsilon = V(\epsilon)$, and $\Sigma^e = V(e)$, and we again use the subscripts H, P to indicate the relevant sub-vectors and sub-matrices of these quantities corresponding to the historical and predictive components. We also assume $E(\epsilon) = 0$ to reflect the

belief that there are no systematic biases in the model known a priori. The Bayes linear forecast equations are now fully expressed as follows

$$E_{z_H}(y_P) = \mu_P^* + (\Sigma_{PH}^* + \Sigma_{PH}^\epsilon)(\Sigma_H^* + \Sigma_H^\epsilon + \Sigma_H^e)^{-1}(z_H - \mu_H^*), \tag{12}$$

$$V_{z_H}(y_P) = (\Sigma_P^* + \Sigma_P^\epsilon) - (\Sigma_{PH}^* + \Sigma_{PH}^\epsilon)(\Sigma_H^* + \Sigma_H^\epsilon + \Sigma_H^e)^{-1}(\Sigma_{HP}^* + \Sigma_{HP}^\epsilon). \tag{13}$$

Given a specification for Σ^ϵ and Σ^e, we can assess the first and second order specifications $E(f_i(x))$, $C(f_i(x), f_i(x'))$ from our emulator of f for every $x, x' \in X^*$. We may therefore obtain the mean and variance of $f^* = f(x^*)$ by first conditioning on x^* and then integrating out with respect to an appropriate prior specification over X^* for x^*. Hence $E_F(f^*)$ and $V_F(f^*)$ are calculated to be the expectation and variance (with respect to our prior belief specification about x^*) of our adjusted beliefs about $f(x)$ at $x = x^*$ given the model evaluations F. Specifically, this calculation requires the computation of the expectations, variances and covariances of all $g_{ij}(x_{[i]}^*)$, $u_i(x_{[i]}^*)$, $v_i(x^*)$, which, in general, may require substantial numerical integration.

This analysis makes predictions without a preliminary calibration stage. Therefore, the approach is tractable even for large systems, as are search strategies to identify collections of simulator evaluations chosen to minimise adjusted forecast variance. The approach is likely to be effective when global variation outweighs local variation and the overall collection of global functional forms $g(x)$ for f_H and f_P are similar. It does not exploit the local information relevant to the predictive quantities, as represented by the residual terms $u_i(x)$ in the component emulators. If some quantities that we wish to predict have substantial local variation, then we may introduce a Bayes linear calibration stage before forecasting, whilst retaining tractability, (Goldstein and Rougier, 2006).

3 Application

3.1 Model Description

The focus of our application is a simulation of a hydrocarbon reservoir provided to us by Energy SciTech Ltd. The model is a representation of the Gullfaks oil and gas reservoir located in the North Sea, and the model is based around a 3-dimensional grid of size $38 \times 87 \times 25$ where each grid cell represents a cuboid region of subterranean rock within the reservoir. Each grid cell has different specified geological properties and contains varying proportions of oil, water and gas. The reservoir also features a number of wells which, during the course of the simulation, either extract fluids from or inject fluids into the reservoir. The overall purpose of the simulation is thus to model changes in pressure, and the flows and changes in distribution of the different fluids throughout the reservoir, thereby giving information on the pressures and production levels at each of the wells. A simple map of the reservoir is shown in Figure 1.

The inputs to the computer model are a collection of scalar multipliers which adjust the magnitudes of the geological properties of each grid cell uniformly across the entire reservoir. This results in four field multipliers – one each for porosity (ϕ), x-permeability (k_x), z-permeability (k_z), and critical saturation (crw). There is no multiplier for y-permeability as the (x, y) permeabilities are treated as isotropic. In addition to these four inputs, we have multipliers for aquifer permeability (A_p) and aquifer height (A_h) giving a total of 6 input parameters. The input parameters and their ranges are summarised in Table 1.
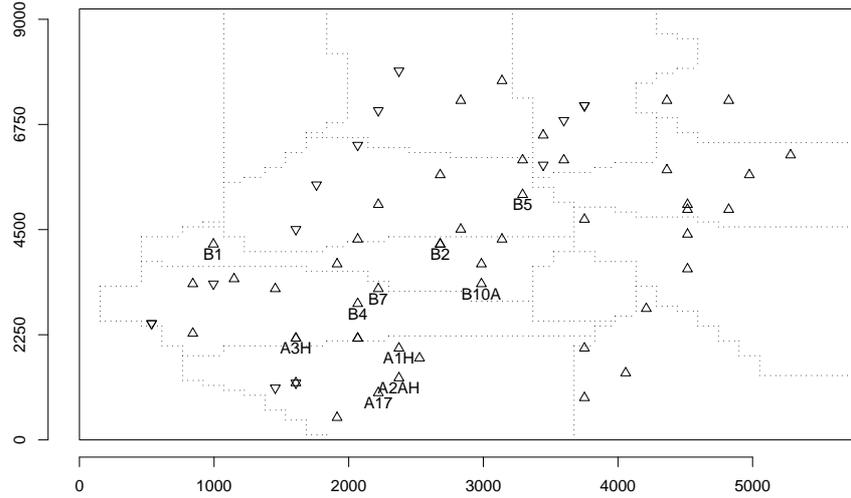
Figure 1: Map of the Gullfaks reservoir grid. Production wells are marked △, injection wells are marked ▽, wells considered in our analysis are labelled, the boundaries of different strucural regions of the reservoir are indicated by dotted lines.

Description	Symbol	Initial range
Porosity	ϕ	$[0.5, 1.5]$
x-permeability	k_x	$[0.25, 6]$
z-permeability	k_z	$[0.1, 0.75]$
Critical saturation	crw	$[0.4, 1.6]$
Aquifer height	A_h	$[50, 500]$
Aquifer permeability	A_p	$[300, 3000]$

Table 1: The six input parameters to the hydrocarbon reservoir model

The outputs of the model are collections of time series of monthly values of various production quantities obtained for each well in the reservoir. The output quantities comprise monthly values of oil, water and gas production rates, oil, water and gas cumulative production totals, water-cut, gas-oil ratio, bottom-hole pressure and tubing-head pressure. For the purposes of our analysis, we shall focus exclusively on oil production rate since this is the quantity of greatest practical interest and has corresponding historical observations. In terms of the time series aspect of the output, we shall focus on a 3-year window in the operation of the reservoir beginning at the start of the third year of production. We smooth these 36 monthly observations by taking four-month averages. By making these restrictions, we focus our attention on the 10 production wells which are operational throughout this period and so our outputs now consist of a collection of 10 time series with each 12 time points.

For the purposes of our analysis, it will be necessary to have access to a fast approximate version of the simulator. To obtain such an approximation, we coarsened the vertical gridding of the model by a factor of 10. The evaluation time for this coarse model is between 1 to 2 minutes, compared to 1.5 to 3 hours for the full reservoir model.

3.2 Overview of the analysis

The evaluation of complex computer models, such as the hydrocarbon reservoir simulation, at a given choice of input parameters is often a highly expensive undertaking both in terms of the time and the computation required. This expense typically precludes a large-scale investigation of the behaviour of the simulation with respect to its input parameters on the basis of model evaluations alone. Therefore, since the number of available model evaluations is limited by available resources there remains a substantial amount of uncertainty about the function and its behaviour, which we must represent by means of an emulator.

For some problems, an approximate version of the original simulation may also be available. This *coarse* simulator, denoted f^c, can be evaluated in substantially less time and for substantially less cost, albeit with a consequent lower degree of accuracy. Since, both this coarse simulator and the original *accurate* simulator, f^a, are models of the same physical system, it is reasonable to expect that there will be strong qualitative similarities between the two models. Therefore, with an appropriate belief framework to link the two simulators, we can use a large batch of evaluations of f^c to construct a detailed emulator of the coarse simulator, which we can then use to inform our beliefs about f^a and supplement the sparse collection of available full model evaluations. This is the essence of the multiscale emulation approach.

Our multiscale analysis of the hydrocarbon reservoir model has been chosen to illustrate each of the stages described in Section 2. We proceed in the following stages

1. Initial model runs and screening — we perform a large batch of evaluations of $f^c(x)$ and then identify which wells are most informative and therefore most important to emulate.

2. Emulation of the coarse simulator — given the large batch of evaluations of $f^c(x)$, we now emulate each of the remaining outputs after the screening process.

3. Linking the coarse and accurate emulators — we use our emulators for $f^c(x)$ to construct an informed prior specification for the emulators of $f^a(x)$, which we then update by a small number of evaluations of $f^a(x)$.

4. History matching — using our updated emulators of $f^a(x)$, we apply the history matching techniques of Section 2.4 to identify a set X^* of possible values for the best model input x^*.

5. Re-focusing — we now focus on the reduced space identified by our history matching process. In addition to the previous outputs we now consider an additional time point 12 months after the end of our original time series, which is to be the object of our forecast. We then build our emulators in the reduced space for $f^c(x)$ and $f^a(x)$ over the original time series and the additional forecast point.

6. Forecasting — using our emulators of $f^a(x)$ within the reduced region, we forecast the 'future' time point using the methods from Section 2.5.

3.3 Initial model runs and screening

We begin by evaluating the coarse model $f^c(x)$ over a 1000-point Latin hypercube design (McKay, Beckman, and Conover, 1979; Santner et al., 2003) in the input parameters. Since emulation,

Well name	Well h_i	Cumulative % of variation
B2	4526.0	77.7
A3H	26.5	81.8
B1	19.5	84.6
B5	14.1	87.1
B10A	10.5	94.7
B7	7.0	95.2
A2AH	7.3	99.2
A17	1.1	99.7
B4	1.0	99.7
A1H	1.1	100.0

Table 2: Table of summary statistics for the selection of principal wells

history matching and forecasting are computationally demanding processes, we choose to screen the collection of 120 outputs and determine an appropriate subset which will serve as the focus of our analysis. In order to identify this reduced collection, we will apply the principal variable selection methods of Cumming and Wooff (2007) to the correlation matrix of the output vectors $\{f^c(x_i)\}$, $i = 1, \ldots, 1000$.

Given a collection of outputs, $y_{1:q}$, with correlation matrix R, the principal variable selection procedure operates by assigning a score $h_i = \sum_{j=1}^{q} \text{Corr}(y_i, y_j)^2$ to each output y_i. The first principal variable is then identified as the output which maximises this score. Subsequent outputs are then selected using the partial correlation given the set of identified principal variables. This allows for the choice of additional principal variables to be made having removed any effects of those variables already selected. To calculate this partial correlation, we first partition the correlation matrix into block form

$$R = \left(\begin{array}{cc} R_{11} & R_{12} \\ R_{21} & R_{22} \end{array} \right),$$

where R_{11} corresponds to the correlation matrix of the identified principal variables, R_{22} is the correlation matrix of the remaining variables, and R_{12} and R_{21} are the matrices of correlations between the two groups. We then determine the partial correlation matrix $R_{22\cdot1}$ as

$$R_{22\cdot1} = R_{22} - R_{21} R_{11}^{-1} R_{12}.$$

The process continues until sufficient variables are chosen that the partial variance of each remaining output is small, or a sufficient proportion of the overall variability of the collection has been explained. In general, outputs with large values of h_i have, on average, large loadings on important principal components of the correlation matrix and thus correspond to structurally important variables.

The outputs from the hydrocarbon reservoir model have a group structure, with groups formed by the different wells, and different time points. We intend to retain all time points at a given well to allow for a multivariate temporal treatment of the emulation. Therefore we make our reduction in the number of wells in the model output by applying a modified version of the above procedure where, rather than selecting a single output, y_i, at each stage, we instead select all 12 outputs corresponding to the well with highest total h_i score. We then continue as before, though selecting a block of 12 outputs at each stage. The results from applying this procedure to the 10 wells in the model are given in Table 2.

We can see from the results that there is a substantial amount of correlation among the outputs at each of the wells, as the first identified principal well accounts for 77.7% of the variation of the collection. Introducing additional wells into the collection of principal outputs only increases the amount of variation of all the outputs explained by the principal variables by a small amount. On the basis of this information one could choose to retain the first four or five principal wells and capture between 87% and 95% of the variation in the collection. For simplicity, we choose to retain four of the ten wells, namely B2, A3H, B1 and B5.

3.4 Constructing the coarse emulators

We have accumulated a large collection of simulator runs and identified which production wells in the reservoir are of particular interest. Prior to emulation, the design was scaled so that all inputs took the range $[-1, 1]$, and all outputs from f^c were scaled by the model runs to have mean 0 and variance 1. We now describe the emulator of component $i = (w, t)$ of the coarse simulator f^c, where w denotes the well, and t denotes the time associated with the ith output component of the computer model. Using the form (6), we write our emulator as

$$f_i^c(x) = g_i(x_{[i]})^T \beta_i^c + u_i^c(x_{[i]}) + v_i^c(x), \tag{14}$$

where $g_i(x_{[i]}) = (g_{i1}(x_{[i]}), \ldots, g_{ip_i}(x_{[i]}))$, and $\beta_i^c = (\beta_{i1}^c, \ldots, \beta_{ip_i}^c)$. From this prior form, the prior mean and variance of the coarse emulator are given by

$$E(f_i^c(x)) = g_i(x_{[i]})^T E(\beta_i^c) + E(u_i^c(x_{[i]})) + E(v_i^c(x)), \tag{15}$$

$$V(f_i^c(x)) = g_i(x_{[i]})^T V(\beta_i^c) g_i(x_{[i]}) + V(u_i^c(x_{[i]})) + V(v_i^c(x)), \tag{16}$$

where a priori we consider $\{\beta_i^c, u_i^c(x_{[i]}), v_i^c(x)\}$ as independent.

The residual process $u_i^c(x_{[i]})$ is a weakly stationary process in $x_{[i]}$ which represents the residual variation in the emulator that is not captured by our linear trend in the active variables. As such, we believe that residual values will be strongly correlated for neighbouring values of $x_{[i]}$. We therefore specify a prior covariance structure over values of $u_i^c(x_{[i]})$ which is a function of the separation of the active variables. The prior form we use is the Gaussian covariance function

$$C(u_i^c(x_{[i]}), u_i^c(x_{[i]}')) = \sigma_{u_i}^2 \exp\left(-\theta_i^c ||x_{[i]} - x_{[i]}'||^2\right), \tag{17}$$

where $\sigma_{u_i}^2$ is the point variance at any given x, θ_i is a correlation length parameter which controls the strength of correlation between two separated points in the input space, and $||\cdot||$ is the Euclidean norm.

The nugget process $v_i^c(x)$ expresses all the remaining variation in the emulator attributable to the inactive inputs. The magnitude of the nugget process is often small and so is treated as uncorrelated random noise with $V(v_i^c(x)) = \sigma_{v_i}^2$. We consider the point variances of these two processes to be proportions of the overall residual variance of the computer model given the emulator trend, σ_i^2, so that $\sigma_{u_i}^2 = (1 - \delta_i)\sigma_i^2$, and $\sigma_{v_i}^2 = \delta_i \sigma_i^2$, for some σ_i^2 and some typically small value of δ_i.

The first step in constructing the emulator is to identify, for each output component f_i^c, the subset of active inputs $x_{[i]}$ which drive the majority of global variation in f_i^c. Using the large batch of coarse model runs, we make this determination via a stepwise model search using simple linear regression. We begin by fitting each f_i^c on all linear terms in x using ordinary least squares. We

Well	Time	$x_{[i]}$	No. Model Terms	Coarse Simulator R^2	Accurate Simulator $\tilde{R}^2$
B2	4	ϕ, crw, A_p	9	0.886	0.951
B2	8	ϕ, crw, A_p	7	0.959	0.958
B2	12	ϕ, crw, A_p	10	0.978	0.995
B2	16	ϕ, crw, k_z	7	0.970	0.995
B2	20	ϕ, crw, k_x	11	0.967	0.986
B2	24	ϕ, crw, k_x	10	0.970	0.970
B2	28	ϕ, crw, k_x	10	0.975	0.981
B2	32	ϕ, crw, k_x	11	0.980	0.951
B2	36	ϕ, crw, k_x	11	0.983	0.967
A3H	4	ϕ, crw, A_p	9	0.962	0.824
A3H	8	ϕ, crw, k_x	7	0.937	0.960
A3H	12	ϕ, crw, k_z	10	0.952	0.939
A3H	16	ϕ, crw, k_z	7	0.976	0.828
A3H	20	ϕ, crw, k_x	11	0.971	0.993
A3H	24	ϕ, crw, k_x	10	0.964	0.899
A3H	28	ϕ, k_z, A_p	10	0.961	0.450
A3H	32	ϕ, crw, k_z	11	0.968	0.217
A3H	36	ϕ, crw, k_x	11	0.979	0.278

Table 3: Emulation summary for wells B2 and A3H

then perform a stepwise delete on each regression, progressively pruning away inactive inputs until we are left with a reduced collection $x_{[i]}$ of between 3 and 5 of the original inputs. The active variables for a subset of the wells of f_i^c are presented in the third column of Table 3. We can see from these results that the inputs ϕ and crw are active on almost all emulators for those two wells, a pattern that continues on the remaining two wells. Clearly ϕ and crw are important in explaining the global variation of f^c across the input space. Conversely, the input variable A_h appears to have no notable effect on model output.

The next stage in emulator construction is to choose the functions $g_{ij}(x_{[i]})$ for each $f_i^c(x)$ which form the basis of the emulator trend. Again, since we have an ample supply of computer evaluations we determine this collection by stepwise fitting. For each $f_i^c(x)$, we define the maximal set of basis functions to include an intercept with linear, quadratic, cubic and pairwise interaction terms in $x_{[i]}$. The saturated linear regression over these terms is then fitted using the coarse model runs and we again prune away any unnecessary terms via stepwise selection.

For each component f_i^c, we have now identified a subset of active inputs $x_{[i]}$ and a collection of p_i basis functions $g_i(x_{[i]})$ which adequately capture the majority of the global behaviour of f_i^c. The next stage is to quantify beliefs about the emulator coefficients β_{ij}^c. We fit our linear description in the selected active variables using ordinary least squares assuming uncorrelated errors to obtain appropriate estimates for these coefficients. The value of $E(\beta_{ij}^c)$ is then taken to be the estimate $\hat{\beta}_{ij}^c$ from the linear regression and $V(\beta_{ij}^c)$ is taken to be the variance of the corresponding estimates. As we have 1000 evaluations in an approximately orthogonal design, the estimation error is negligible.

The results of the stepwise selection and model fitting are given in the first five columns of Table 3. We can see from the R^2 values that the emulator trends are accounting for a very high proportion of the variation in the model output. We observe similar performance on the emulators of the remaining wells, with the exceptions of the emulators of well B5 at times $t = 4$ and $t = 8$

which could not be well-represented using any number or combination of basis functions. These two model outputs were therefore omitted from our subsequent analysis leaving us with a total of 34 model outputs.

The final stage is to make assessments for the values of the hyperparameters in our covariance specifications for $u_i(x_{[i]})$ and $v_i(x)$. We estimate σ_i^2 by the estimate of variance of the emulator trend residuals, and then obtain estimates for θ_i and δ_i by applying the robust variogram methods of Cressie (1991). We then use these estimates as plug-in values for θ_i, δ_i^2, and σ_i^2.

This approach to emulator construction relies on having a large number of model runs to perform the model search and fitting processes adequately, which is made possible by exploiting the speed of the coarse simulator. Having such an ample data set allows for model selection and parameter estimation to be carried out independently for each component of f^c and to be driven solely by the information from the model runs. If we do not have such a large amount of runs, then we must use expert judgement to specify appropriate choices for $x_{[i]}$ and $g_i(x_{[i]})$, and quantify prior expectations and variances for β_i^c and the residual processes, which we update by the available model evaluations via the Bayes linear adjustment formulae (3) and (4) (see Craig, Goldstein, Seheult, and Smith, 1998).

3.5 Linking the coarse and accurate emulators

We now develop an emulator for the accurate version of the computer model $f^a(x)$. We consider that $f^c(x)$ is sufficiently informative for $f^a(x)$ that it serves as the basis for an appropriate prior specification for this emulator. We initially restrict our emulator for component i of $f^a(x)$ to share the same set of active variables and the same basis functions as its coarse counterpart $f_i^c(x)$. Since the coarse model $f^c(x)$ is well-understood due to the considerable number of model evaluations, we consider the coarse emulator structure as known and fixed and use this as a structural basis for building the emulator of $f^a(x)$. Thus we specify a prior accurate emulator of the form

$$f_i^a(x) = g_i(x_{[i]})\beta_i^a + \beta_{w_i}^a w_i^c(x) + w_i^a(x), \tag{18}$$

where $w_i^c(x) = u_i^c(x_{[i]}) + v_i^c(x_i)$, $w_i^a(x) = u_i^a(x_{[i]}) + v_i^a(x_i)$, and we have an identical global trend structure over the inputs albeit with different coefficients. On this accurate emulator, we also introduce some unknown multiple of the coarse emulator residuals $\beta_{w_i}^a w_i^c(x)$, and include a new residual process $w_i^a(x)$ which will absorb any structure of the accurate computer model that cannot be explained by our existing set of active variables and basis functions. Alternative methods for constructing such a multiscale emulator can be found in Kennedy and O'Hagan (2000) and Qian and Wu (2008).

As we have performed a large number of evaluations of $f^c(x)$, over a roughly orthogonal design, our estimation error from the model fitting is negligible and so we consider the β^c as known, and further for any x at which we have evaluated $f^c(x)$, the values of $w^c(x)$ are also known. Thus we incorporate the $w^c(x)$ into our collection of basis functions with associated coefficient $\beta_{w_i}^a$. Absorbing $w^c(x)$ into the basis functions and $\beta_{w,i}^a$ into the coefficient vector β_i^a, we write the prior expectation and variance for the accurate simulator as

$$E(f^a(x)) = g_i(x_{[i]})^T E(\beta_i^a) + E(w_i^a(x_{[i]})), \tag{19}$$

$$V(f^a(x)) = g_i(x_{[i]})^T V(\beta_i^a) g_i(x_{[i]}) + V(w_i^a(x_{[i]})), \tag{20}$$

where now $g_i(x_{[i]}) = (g_{i1}(x_{[i]}), \ldots, g_{ip_i}(x_{[i]}), w^c(x))$, and $\beta_i^c = \left(\beta_{i1}^c, \ldots, \beta_{ip_i}^c, \beta_{w_i}^a\right)$. We also specify the expectation and variance of the residual process $w^a(x)$ to be

$$E(w_i^a(x)) = 0, \tag{21}$$

$$C(w_i^a(x), w_i^a(x'),) = \sigma_{u_i}^2 \exp\left(-\theta_i^c||x_{[i]} - x_{[i]}'||^2\right) + \sigma_{v_i}^2 \mathbb{I}, \tag{22}$$

where the covariance function between any pair of residuals on the accurate emulator has the same prior form and hyperparameter values as that used for $u^c(x_{[i]})$ in (17).

We now consider the prior form of β^a in more detail. If we believe that each of the terms in the emulator trend corresponds to a particular qualitative physical effect, then we may expect that the magnitude of these effects will change differentially as we move from the coarse to the accurate simulator. This would suggest allowing the contribution of each $g_{ij}(x_{[i]})$ to the trend of $f^a(x)$ to be changed individually. One prior form which allows for such changes is

$$\beta_{ij}^a = \rho_{ij}\beta_{ij}^c + \gamma_{ij} \tag{23}$$

where ρ_{ij} is an unknown multiplier which scales the contribution of β_{ij}^c to β_{ij}^a, and γ_{ij} is a shift that can accommodate potential changes in location. We consider ρ_{ij} to be independent of γ_{ij}.

Next we assign prior expectations of ρ and γ to be $E(\rho_{ij}) = 1$, and $E(\gamma_{ij}) = 0$ to correspond to a judgement that the expected values of the fine emulator coefficients are, a priori, the same as those of the coarse emulator. We now describe the covariance structure for the ρ_{ij} and γ_{ij} parameters. Every ρ_{ij} (and similarly γ_{ij}) is associated with a unique well w and time point t via the simulator output component $f_i^a(x)$. Additionally, every ρ_{ij} is associated with a unique regression basis function $g_{ij}(\cdot)$. Given these associations, we consider there to be two sources of correlation between the ρ_{ij} at a given well. First, for a given well w we consider there to be temporal effects correlating all $(\rho_{ij}, \rho_{i'j'})$ pairs to a degree governed by their separation in time. Secondly, we consider that there are model term effects which introduce additional correlation when both ρ_{ij} and $\rho_{i'j'}$ are multipliers for coefficients of the same basis functions, i.e. $g_{ij}(\cdot) \equiv g_{i'j'}(\cdot)$.

To express this covariance structure concisely, we extend the previous notation and write ρ_{ij} as $\rho_{(w,t,k)}$ where w and t correspond to the well and time associated with $f_i^a(x)$, and where k indexes the unique regression basis function associated with ρ_{ij}, namely the single element of the set of all basis functions $G = \cup_{i,j}\{g_{ij}(\cdot)\}$. Under this notation, for a pair of multipliers $(\rho_{(w,t,k)}, \rho_{(w,t',k')})$ then $k = k'$ if and only if both are multipliers for coefficients of the same basis function, say ϕ^2, albeit on different emulators. On this basis, we write the covariance function for $\rho_{(w,t,k)}$ as

$$C(\rho_{(w,t,k)}, \rho_{(w,t',k')}) = \left(\sigma_{\rho 1}^2 + \sigma_{\rho 2}^2 \mathbb{I}_{k=k'}\right) R_T(t,t'), \tag{24}$$

where $\sigma_{\rho 1}^2$ governs the contribution of the overall temporal effect to the covariance, and $\sigma_{\rho 2}^2$ controls the magnitude of the additional model term effect, $R_T(t,t') = \exp\{-\theta_T(t-t')^2\}$ is a Gaussian correlation function over time, and $\mathbb{I}_p$ is the indicator function of the proposition p. Our covariance specification for the $\gamma_{(w,t,k)}$ takes the same form as (24) albeit with variances $\sigma_{\gamma 1}^2$ and $\sigma_{\gamma 2}^2$.

To complete our prior specification over $f^a(x)$, we assign $\sigma_{\rho 1}^2 = \sigma_{\gamma 1}^2 = 0.1$, and $\sigma_{\rho 2}^2 = \sigma_{\gamma 2}^2 = 0.1$ for all output components, which correspond to the belief that coefficients are weakly correlated with other coefficients on the same emulator, and that the model term effect has a similar contribution to the covariance as the temporal effect. We also assigned $\theta_T = 1/12^2$ to allow for a moderate amount of correlation across time.

We now evaluate a small batch of 20 additional runs of the accurate simulator. The runs were chosen by generating a large number of Latin hypercube designs and selecting that which would be most effective at reducing our uncertainty about β^a by minimising $\mathbf{tr}\{V(\hat{\beta}^a)\}$ by least squares. As information on $f_w^a(x) = \left(f_{(w,4)}^a(x), f_{(w,8)}^a(x), \ldots, f_{(w,36)}^a(x)\right)$ for each well w is now available in the form of the model runs F_w^a over the design X^a, we can make a diagnostic assessment of the choices made in specifying prior beliefs about f_w^a. As our prior statements about $f_w^a(x)$ describe our beliefs about the uncertain value of the simulator output, we can use the observational data to compare our prior expectations to what actually occurs. A simple such comparison is achieved by the *discrepancy ratio* for F_w^a

$$\mathrm{Dr}(F_w^a) = \frac{\{F_w^a - E(f_w^a(X^a))\}^T V(f_w^a(X^a))^{-1}\{F_w^a - E(f_w^a(X^a))\}}{\mathbf{rk}\{V(f_w^a(X^a))\}}, \tag{25}$$

which has prior expectation 1, and where $\mathbf{rk}\{V(f_w^a(X^a))\}$ corresponds to the rank of the matrix $V(f_w^a(X^a))$. Very large values of $\mathrm{Dr}(F_w^a)$ may suggest a mis-specification of the prior expectation or a substantial underestimation of the prior variance. Conversely, very small values of $\mathrm{Dr}(F_w^a)$ may suggest an overestimation of the variability of $f_w^a(x)$. In the case of our prior specifications for the multivariate emulators for each well, we obtain discrepancy ratio values of 0.86, 1.14, 0.67, and 1.07 suggesting our prior beliefs are broadly consistent with the behaviour observed in the data. For more detailed diagnostic methods for evaluating our prior and adjusted beliefs see Goldstein and Wooff (2007).

Given the prior emulator for $f_w^a(x)$ and the simulator evaluations F_w^a, we now update our prior beliefs about $f_w^a(x)$ by the model runs via the Bayes linear adjustment formulae (3) and (4). Thus we obtain an adjusted expectation and variance for $f_w^a(x)$ given F_w^a

$$E_{F_w^a}(f_w^a(x)) = E(f_w^a(x)) + C(f_w^a(x), F_w^a)V(F_w^a)^{-1}[F_w^a - E(F_w^a)], \tag{26}$$

$$V_{F_w^a}(f_w^a(x)) = V(f_w^a(x)) - C(f_w^a(x), F_w^a)V(F_w^a)^{-1}C(F_w^a, f_w^a(x)). \tag{27}$$

The constituent elements of this update can be derived from our prior specifications $f^a(x)$ from (19) and (20), and our subsequent belief statements made above.

To assess the adequacy of fit for the updated fine emulators of the outputs updated using the 20 runs of $f^a(x)$, we calculate a version of the R^2 statistic by using the residuals obtained from the adjusted emulator trend, denoted $\tilde{R}^2$. These are given in the final column of Table 3. It is clear that the majority of the fine emulators perform well and accurately represent the fine simulator, except for the emulators at well A3H and times $t = 28, 32, 36$, which display poor fits to the fine model due to the behaviour of f^c at those locations being uninformative for the corresponding accurate model. For additional illustration, the coefficients for the coarse emulator and the adjusted expected coefficients of the accurate emulator of a particular output are given in the first two rows of Table 4. We can see from these values in this case that both emulators have good fits to the simulator despite the different coefficients. The results from fitting the same regression trend using a large sample of 183 accurate model runs by least squares are also presented for comparison. There are strong similarities between the coefficients of the accurate emulator obtained by adjustment of the coarse emulator and the large-sample OLS fit suggesting that the multiscale emulation approach is effective in this case.

Model	Int	ϕ	ϕ^2	ϕ^3	crw	crw^2	crw^3
Coarse	0.663	-0.326	-1.858	2.313	-0.219	0.064	0.079
Accurate	0.612	-0.349	-0.599	0.811	0.313	-0.331	-0.822
Large-sample Accurate	0.547	-0.219	-0.420	0.709	-0.013	0.006	-0.0330
Refocused Coarse	-0.149	2.204	-0.614	-0.858	-0.586	0.386	0.119
Refocused Accurate	0.678	0.402	-0.456	-0.098	-0.057	-0.055	0.053

	$\phi \times crw$	A_p	kz	$w^c(x)$	R^2
Coarse	0.407	-0.008			0.904
Accurate	0.072	0.111		0.112	0.952
Large-sample Accurate	0.062	0.014		0.095	0.851
Refocused Coarse	0.206	0.044	-0.037		0.905
Refocused Accurate	-0.045	-0.034	-0.045	-0.109	0.945

Table 4: Table of coefficients from the emulator of oil production rate at well B1 and time 28

3.6 History Matching

The next stage in our analysis is to use our updated emulator of $f^a(x)$ to history match the reservoir simulator and thereby reduce the space of possible input parameter combinations. As described in Section 2.4, the Bayes linear approach to history matching is based on a comparison between the expected simulator output at a given choice of inputs, x, and the corresponding observed historical observations, z, made via the implausibility measure.

At a given well, we allow outputs corresponding to different times to be temporally correlated. Thus we apply the multivariate implausibility measure (9) to obtain an assessment of the potential match quality of a given input x at each well. Incorporating the definitions of z and y from (1) and (2) into the implausibility formulation, we can write the implausibility function as

$$I(x) = \{E_{F^a}(f^a(x)) - z\}^T \{V_{F^a}(f^a(x)) + V(e) + V(\epsilon)\}^{-1} \{E_{F^a}(f^a(x)) - z\}, \qquad (28)$$

which is a function of the adjusted expectations and variances of our emulator for $f^a(x)$ given the model evaluations, combined with the corresponding observational data, z, and the covariances for the observational error, e, and the model discrepancy, ϵ.

We now specify our prior expectation and variance for the observational error e and the model discrepancy ϵ. We do not have any prior knowledge of biases of the simulator or the data therefore we assign $E(e) = 0$ and $E(\epsilon) = 0$. It is believed that our available well production history has an associated error of approximately $\pm 10\%$, therefore we assign $2 \times \text{sd}(e_i) = 0.1 \times z_i$ for each emulator component $f_i^a(x)$ and we assume that there is no prior correlation between observational errors. Assessing model discrepancy is a more conceptually challenging task requiring assessment of the difference between the model evaluated at the best, but unknown, input, x^*, and the true, also unknown, value of the system. For simplicity, we assign the variance of the discrepancy to be twice that of the observational error to reflect a belief that the discrepancy has a potentially important and proportional effect. In contrast to observational errors, we introduce a relatively strong temporal correlation over the $\epsilon_i = \epsilon_{(w,t)}$ such that $\text{Corr}(\epsilon_{(w,t)}, \epsilon_{(w,t')}) = \exp\{-\theta_T(t-t')^2\}$, where we assign $\theta_T = 1/36^2$ to allow the correlation to persist across all 12 time points spanning the 36-month period. We specify such a correlation over the model discrepancy since we believe that if the simulator is, for example, substantially under-estimating the system at time t, then it will be highly likely that it will also under-predict at time $t + 1$.

To assess how the implausibility of input parameter choices changes with x, we construct a grid

over the collection of active inputs spanning their feasible ranges and we evaluate (28) for each of the four selected wells, at every x point in that grid. We then have a vector of four implausibilities for every input parameter combination in the grid. To collapse these vectors into a scalar for each x, we use the maximum projection (8) where we maximise over the different wells to obtain a single measure $I_M(x)$. This gives a conservative measure for the implausibility of a parameter choice, x, since if x is judged implausible on any one of the wells then it is deemed implausible for the collection. Thus the implausibility scores are combined in such a way that a particular input point x is only ever considered a potential match to the simulator if it is an acceptable match across all wells.

Thus we obtain a quantification for the match quality for a representative number of points throughout the possible input space. The domain of the implausibility measure is a 5-dimensional cube and, as such, it is hard to visualise the implausibility structure within that space. To address this problem, we project this hypercube of implausibility values down to 2-D spaces in every pair of active inputs using the method described in Craig et al. (1997). If we partition x such that $x = (x', x'')$, then we obtain a projection of $\hat{I}(x)$ onto the subspace x' of x by calculating

$$\min_{x''} I_M(x),$$

which is a function only of x'.

$I_M(x)$ is a Mahalanobis distance over the four time points for each well. We produce the projections of the implausibility surface in Figure 2, colouring by appropriate χ_4^2 quantiles for comparison. The first plot in Figure 2(a) shows the approximate proportion of the implausibility space that would be excluded if we were to eliminate all points x with $I_M(x)$ greater than a number of the standard deviations from the re-scaled χ^2 distribution. For example, thresholding at 3 standard deviations, corresponding to $I_M(x) \simeq 4$, would excludes approximately 90% of of input space. The subsequent plots in Figure 2(b) to Figure 2(f) are a subset of the 2-D projections of the implausibility surface onto pairs of active variables. It is clear from these plots that there are regions of low implausibility corresponding to values of ϕ less than approximately 0.8 which indicates a clear region of potential matches to our reservoir history. Higher values of ϕ are much more implausible and so would be unlikely history matches. Aside from ϕ, there appears to be little obvious structure on the remaining active variables. This is reinforced by Figure 2(f), which is representative of all the implausibility projections in the remaining active inputs. This plot clearly shows that there is no particular set of choices for k_x or k_z that could reasonably be excluded from consideration without making very severe restrictions of the input space. Therefore, we decide to define our region of potential matches, X^*, by the set $\{x : I_M(x) \leq 4\}$. Closer investigation revealed that this set can be well-approximated by the restriction that ϕ should be constrained to the sub-interval $[0.5, 0.79]$.

3.7   Re-focusing

Given the reduced space of input parameters, we now re-focus our analysis on this subregion with a view towards our next major task – forecasting. Our intention for the forecasting stage is, for each well, to use the last four time points in our existing series to forecast an additional time point located 12-months beyond the end of our original time series at $t = 48$ months. Therefore, we no longer continue investigating the behaviour of well B2 since it ceases production shortly after our original 3-year emulation period.
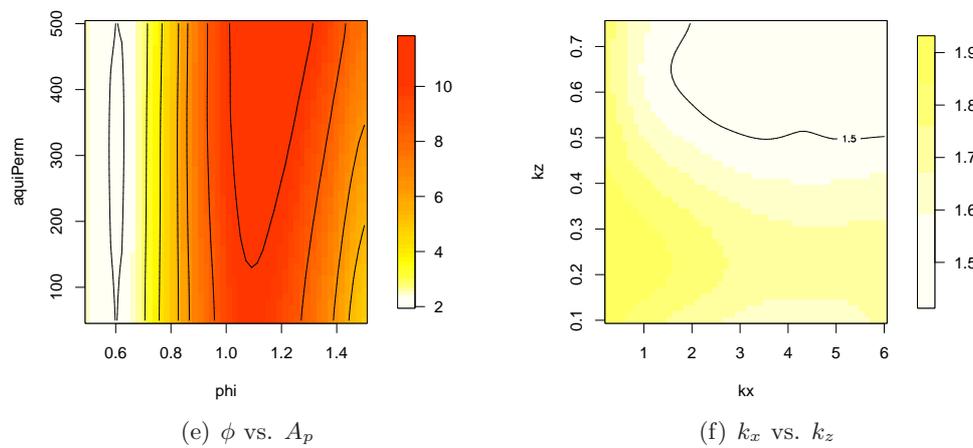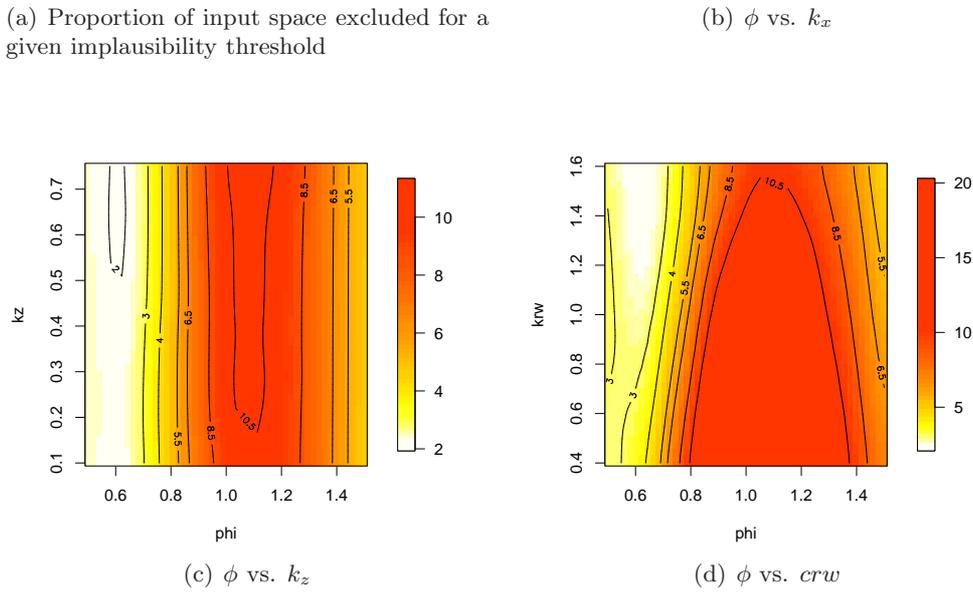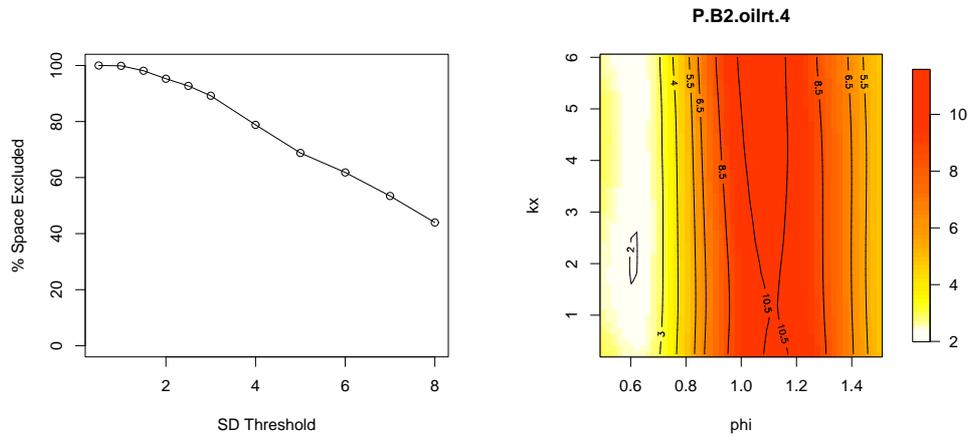
(a) Proportion of input space excluded for a given implausibility threshold

(b) ϕ vs. k_x

(c) ϕ vs. k_z

(d) ϕ vs. crw

(e) ϕ vs. A_p

(f) k_x vs. k_z

Figure 2: Implausibility summary and projections for the hydrocarbon reservoir simulator

Well	Time	$x_{[i]}$	No. Model Terms	Coarse Trend R^2	Accurate Trend $\tilde{R}^2$
A3H	24	ϕ, crw, k_x, k_z	8	0.981	0.974
A3H	28	ϕ, crw, k_x, k_z, A_p	11	0.971	0.989
A3H	32	ϕ, crw, k_x, k_z	11	0.973	0.958
A3H	36	ϕ, crw, k_x	10	0.958	0.917
A3H	48	ϕ, crw, k_x	10	0.981	0.888
B1	24	ϕ, crw, k_x, A_p	11	0.894	0.982
B1	28	ϕ, crw, k_z, A_p	9	0.905	0.945
B1	32	ϕ, crw, k_x	11	0.946	0.953
B1	36	ϕ, crw, k_x	11	0.953	0.927
B1	48	ϕ, crw, k_x, A_p	11	0.941	0.880

Table 5: Re-focused emulation summary for wells A3H and B1

To forecast, we require emulators for each the four historical time points as well as the additional predictive point, which we now construct over the reduced space X^*. To build the emulators, we follow the same process as described in Section 3.4 and Section 3.5, requiring a batch of additional model runs. Of our previous batch of 1000 evaluations of $f^c(x)$, 262 were evaluated within X^* and so can be used at this stage. Similarly, of the 20 runs of $f^a(x)$ a total of 6 remain valid. These runs will be supplemented by an additional 100 evaluations of $f^c(x)$ and then by an additional 20 evaluations of $f^a(x)$.

Adopting the same strategy as Section 3.4, we construct our coarse emulators from the information contained within the large pool of model evaluations, albeit with two changes to the process. Since we have already emulated these output components in the original input space (with the exception of our predictive output), we already have some structural information in the form of the $x_{[i]}$ and the $g_i(x_{[i]})$ for each $f_i^c(x)$ that we obtained from the original emulation. Rather than completely re-executing the search for active variables and basis functions, we shall begin our searches using the original $x_{[i]}$ and $g_i(x_{[i]})$ as the starting point. We allow for the emulators to pick up any additional active variables, but not to exclude previously active inputs; and we allow for basis functions in the new $x_{[i]}$ to be both added and deleted to refine the structure of the emulator.

An emulation summary for $f^c(x)$ within the reduced region X^* is given in Table 5 for wells A3H and B1. We can see that the emulator performance is still good with high R^2 indicating that the emulators are still explaining a large proportion of the variation in model output. Observe that many of the emulators have picked up an additional active input variable when we re-focus in the reduced input space.

Considering the emulator for $f^a(x)$, we make a similar belief specification as before to link our emulator of $f^c(x)$ to that of $f^a(x)$. We choose to make the same choices of parameters $(\sigma_{\rho 1}, \sigma_{\rho 2}, \sigma_{\gamma 1}, \sigma_{\gamma 2}, \theta_T)$ to reflect a prior belief that the relationship between the two emulators in the reduced space is the similar to that over the original space. Comparing this prior with the data via the discrepancy ratio (25) showed that it was again reasonably consistent with $\mathrm{Dr}(F^a)$ taking values of 2.14, 0.98, and 2.02, although perhaps we may be slightly understating our prior variance. The prior emulator was then updated by the runs of the accurate model. Looking at the final column at Table 5 we see that the emulator trend fits the data well, although the $\tilde{R}^2$ values appear to be decreasing over time. Example coarse and accurate emulator coefficients for the re-focused emulator are also given in the bottom two rows of Table 4, which shows more variation in the coefficients as we move from coarse to fine in the reduced space and also shows the presence of

an additional active input.

3.8 Forecasting

We now apply the forecasting methodology, as described in Section 2.5, to the three wells under consideration from the hydrocarbon reservoir model. The goal of this forecasting stage is to forecast the collection of future system output, y_P, using the available historical observations z_H. For a given well in the hydrocarbon reservoir, we will the vector of four average oil production rates at times $t = 24$, 28, 32, and 36 as historical values, and the quantity to be predicted is the corresponding production rate observed 12 months later at time $t = 48$. As we actually have observations z_P on y_P, these may act as a check on the quality of our assessments.

By history matching the hydrocarbon reservoir, we have determined a region X^* in which we believe it is feasible that an acceptable match x^* should lie. For our forecast, we shall consider for simplicity that x^* is equally likely to be any input point contained within the region X^*. This means that we take our prior for x^* to be uniform over the hypercube X^*. As we take the $g_{ij}(x)$ to be polynomials in x, then the expectations, variances and covariances of the $g_{ij}(x^*)$ can be found analytically from the moments of a multivariate uniform random variable which greatly simplifies the calculations of μ^* and Σ^*. We consider that $f^* = f^a(x^*)$ and we now refine our previously generous specification for $V(e)$. Since each output quantity is an average of four monthly values, we now take $V(e)$ to be 1/2 its previous value to reflect the reduced uncertainty associated with the mean value.

Before we can obtain a prediction for y_P given z_H we require an appropriate belief specifcation for the model discrepancy ϵ, both at the past time points, and also at the future time point to be predicted. The role of the model discrepancy, ϵ, is important in forecasting as it quantifies the amount by which we allow the prediction to differ from our mean simulator prediction, $\mu^* = E(f^*)$, in order to move closer to the true system value y_P. If the specified discrepancy variance is too small, then we will obtain over-confident and potentially inaccurate forecasts located in the neighbourhood of μ^*. If the discrepancy variance is too large then the forecast variances could be unfavourably large or we could over-compensate by the discrepancy and move too far from y_P. Therefore, we briefly consider the specification of $V(\epsilon)$.

Consider the plots in Figure 3 depicting the outputs from the reservoir model over the time period of our predictions. Observe that for time points 24 to 36 at wells B1 and B5, the mean values μ^* (indicated by the solid black circles) underestimate the observational data (indicated by the thick solid line). However at well A3H, the simulator "overestimates" observed production. Furthermore, we observe that for well A3H the size of $|\mu^* - z|$ is a decreasing function of time, for well B1 this distance increases over time, and for well B5 $|\mu^* - z|$ appears to be roughly constant.

Given the specification for the observational error used in Section 3.6 and the observed history, we can compare $V(e_H)$ to the observed values of $(\mu^* - z_H)^2$ at each well and historical time point to obtain a simple order of magnitude data assessment for the discrepancy variance at the historical time points. To obtain our specification for $V(\epsilon)$, we took a weighted combination of prior information in the form of our belief specification for $V(\epsilon)$ from Section 3.6 and these order of magnitude numerical assessments. As the value of z_P is unknown at the prediction time $t = 48$, we must make a specification for $V(\epsilon_P)$ in the absence of any sample information. To make this assessment, we performed simple curve fitting to extrapolate the historical discrepancy variances to the forecast point $t = 48$. The resulting specification for $V(\epsilon)$ is given in Table 6; the correlation structure over
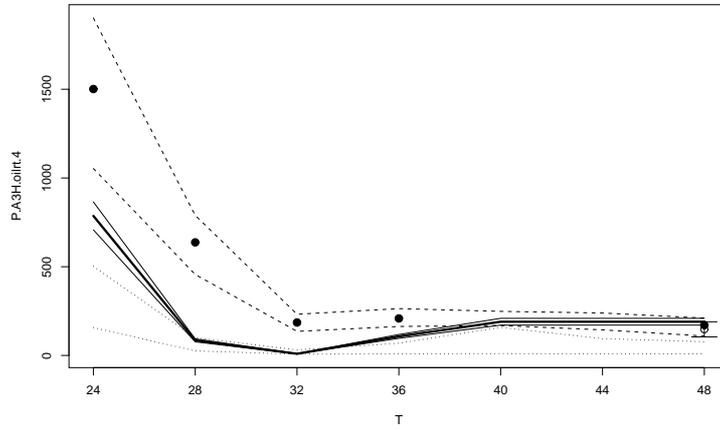
(a) Well A3H
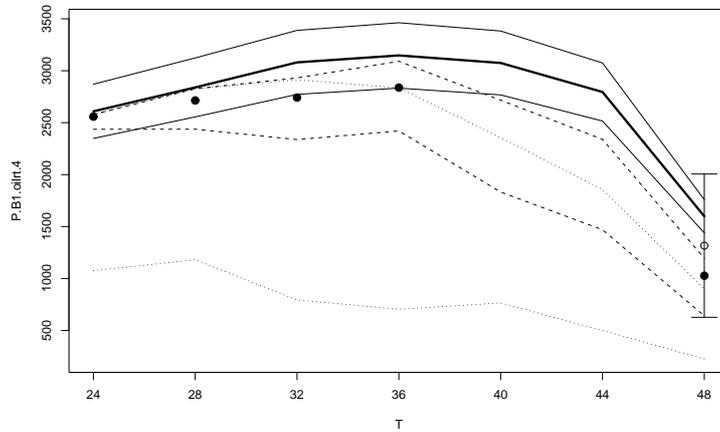
(b) Well B1
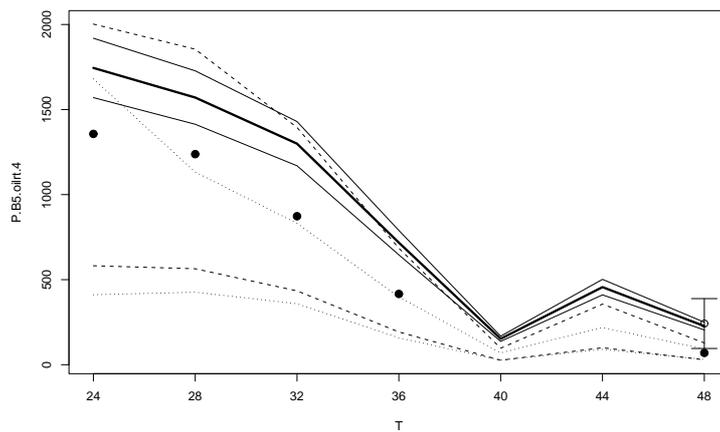
(c) Well B5

Figure 3: Simulator outputs, observational data and forecasts for each well. The solid lines indicate z with error bounds of $2 \times \mathrm{sd}(e)$. The dotted and dashed lines represent the maximum and minimum values of the runs of $f^c(x)$ and $f^a(x)$ respectively in X^*. The solid black dots correspond to μ^*. The forecast is indicated by a hollow circle with attached error bars.

$2 \times \mathrm{sd}(\epsilon_{(w,t)})$	$t = 24$	$t = 28$	$t = 32$	$t = 36$	$t = 48$
A3H	504.9	390.4	124.5	71.5	14.7
B1	130.5	142.0	260.4	245.1	408.0
B5	284.3	239.3	305.8	214.7	260.8

Table 6: Table of specified values for $2 \times \mathrm{sd}(\epsilon_{(w,t)})$

		A3H	B1	B5
Observation	z_P	190.0	1598.6	227.0
	$2 \times \mathrm{sd}(e)$	19.0	159.9	22.7
Prior	$E(y_P)$	170.2	1027.1	69.4
	$2 \times \mathrm{sd}(y_P)$	62.9	595.1	270.3
Forecast	$E_{z_H}(y_P)$	170.3	1207.1	299.8
	$2 \times \mathrm{sd}_{z_H}(y_P)$	39.1	348.9	167.4

Table 7: Forecasts of y_P at $t = 48$ using z_H for three wells in the hydrocarbon reservoir

the discrepancies is the same as in Section 3.6

Given all these specifications, we evaluate the forecast equations (12) and (13). The results of the forecasts are presented in Table 7 alongside the corresponding prior values, and the actual observed production z_P at $t = 48$. The forecasts and their errors are also shown on Figure 3 by a hollow circle with attached error bars.

We can interpret the prediction $E_{z_H}(y_P)$ as the forecast from the simulator for y_P which is modified in the light of the discrepancy between $E_{z_H}(y_H)$ and the observed z_H. In the case of the wells tabulated, the simulator for well A3H over-estimates the value of z_H during the period $t = 24, \ldots, 36$ resulting in a negative discrepancy and a consequent downward correction to our forecast. However, interestingly in the intervening period the simulator changes from under-estimating to over-estimating observed well production, and so on the basis of our observed history alone we under-predict the oil production rate for this well. The wells B1 and B5 behaved in a more consistent manner with a constant under-prediction of the observed data compared to the observed history being reflected by an increase to our forecast. Note that whilst some of the best runs of our computer model differ substantially from z_P, the corrections made by the model discrepancy result in all of our forecast intervals being within the measurement error of the data.

In practice, the uncertainty analysis for a reservoir model is an iterative process based on monitoring the sizes of our final forecast intervals and investigating the sensitivity of our predictions to the magnitude of the discrepancy variance and correlationm, for example by repeating the forecasts using the discrepancy variance values from Table 6 scaled by some constant α, for various values of α and varying the degree of temporal correlation across discrepancy terms. If we obtain forecast intervals which are sufficiently narrow to be useful to reservoir engineers then the we can end our analysis at this stage. If, however, our prediction intervals are unhelpfully large then we return to and repeat earlier stages of the analysis. For example, introducing additional wells and other aspects of historical data into our analysis could be helpful in further reducing the size of X^*, and allow us to refocus again and therefore reduce the uncertainties attached to our emulator, and so narrow the forecast interval.

4   Acknowledgements

APPENDIX

A   Foundations - Bayes Linear Analysis

The Bayes linear approach is similar in spirit to conventional Bayes analysis, but derives from a simpler system for prior specification and analysis, and so offers a practical methodology for analysing partially specified beliefs for large problems. The approach uses expectation rather than probability as the primitive for quantifying uncertainty; see De Finetti (1974, 1975). In the Bayes linear approach, we make direct prior specifications for that collection of means, variances and covariances which we are both willing and able to assess. Given two random vectors, $\mathbf{B}$, $\mathbf{D}$, the *adjusted* expectation for element B_i, given $\mathbf{D}$, is the linear combination $a_0 + \mathbf{a}^T\mathbf{D}$ minimising $E((B_i - a_0 - \mathbf{a}^T\mathbf{D})^2)$ over choices of $a_0, \mathbf{a}$. The adjusted expectation vector, $E_{\mathbf{D}}(\mathbf{B})$ is evaluated as

$$E_{\mathbf{D}}(\mathbf{B}) = E(\mathbf{B}) + C(\mathbf{B}, \mathbf{D})(V(\mathbf{D}))^{-1}(\mathbf{D} - E(\mathbf{D}))$$

(If $V(\mathbf{D})$ is not invertible, then we use an appropriate generalised inverse). The *adjusted variance matrix* for $\mathbf{B}$ given $\mathbf{D}$, is

$$V_{\mathbf{D}}(\mathbf{B}) = V(\mathbf{B} - E_{\mathbf{D}}(\mathbf{B})) = V(\mathbf{B}) - C(\mathbf{B}, \mathbf{D})(V(\mathbf{D}))^{-1}C(\mathbf{D}, \mathbf{B})$$

Stone (1963), and Hartigan (1969) are among the first to discuss the role of such assessments in partial Bayes analysis. A detailed account of Bayes linear methodology is given in Goldstein and Wooff (2007), emphasising the interpretive and diagnostic cycle of subjectivist belief analysis.

The basic approach to statistical modelling within this formalism is through second order exchangeability. An infinite sequence of vectors is second-order exchangeable if the mean, variance and covariance specification is invariant under permutation. Such sequences satisfy the second order representation theorem which states that each element of such a sequence may be decomposed as the uncorrelated sum of an underlying 'population mean' quantity and an individual residual, where the residual quantities are themselves uncorrelated with zero mean and equal variances. This is similar in spirit to de Finetti's representation theorem for fully exchangeable sequences but is sufficiently weak, in the requirements for prior specification, that it allows us to construct statistical models directly from simple collections of judgements over observable quantities.

Within the usual Bayesian view, adjusted expectation offers a simple, tractable approximation to conditional expectation, and adjusted variance is a strict upper bound for expected posterior

variance, over all prior specifications consistent with the given moment structure. The approximations are exact in certain special cases, and in particular if the joint probability distribution of $\mathbf{B}$, $\mathbf{D}$ is multivariate normal. Adjusted expectation is numerically equivalent to conditional expectation when $\mathbf{D}$ comprises the indicator functions for the elements of a partition, i.e. each D_i takes value one or zero and precisely one element D_i will equal one. We may therefore view adjusted expectation as a generalisation of de Finetti's approach to conditional expectation based on 'called-off' quadratic penalties, where we remove the restriction that we may only condition on the indicator functions for a partition. Geometrically, we may view each individual random quantity as a vector, and construct the natural inner product space based on covariance. In this construction, the adjusted expectation of a random quantity Y, by a further collection of random quantities $\mathbf{D}$, is the orthogonal projection of Y into the linear subspace spanned by the elements of $\mathbf{D}$ and the adjusted variance is the squared distance between Y and that subspace. This formalism extends naturally to handle infinite collections of expectation statements, for example those associated with a standard Bayesian analysis.

A more fundamental interpretation of the Bayes linear approach derives from the temporal sure preference principle, which says, informally, that if it is necessary that you will prefer a certain small random penalty A to C at some given future time, then you should not now have a strict preference for penalty C over A. A consequence of this principle is that you must judge now that your actual posterior expectation, $E_T(\mathbf{B})$, at time T when you have observed $\mathbf{D}$, satisfies the relation $E_T(\mathbf{B}) = E_{\mathbf{D}}(\mathbf{B}) + \mathbf{R}$, where $\mathbf{R}$ has, a priori, zero expectation and is uncorrelated with $\mathbf{D}$. If $\mathbf{D}$ represents a partition, then $E_{\mathbf{D}}(\mathbf{B})$ is equal to the conditional expectation given $\mathbf{D}$, and $\mathbf{R}$ has conditional expectation zero for each member of the partition. In this view, the correspondence between actual belief revisions and formal analysis based on partial prior specifications is entirely derived through stochastic relationships of this type.

References

Craig, P. S., Goldstein, M., Rougier, J. C., and Seheult, A. H. (2001), "Bayesian forecasting for complex systems using computer simulators," *Journal of the American Statistical Association*, 96, 717–729.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1997), "Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments," in *Case Studies in Bayesian Statistics*, eds. Gatsonis, C., Hodges, J. S., Kass, R. E., McCulloch, R., Rossi, P., and Singpurwalla, N. D., New York: Springer-Verlag, vol. 3, pp. 36–93.

— (1998), "Constructing partial prior specifications for models of complex physical systems," *Applied Statistics*, 47, 37–53.

Cressie, N. (1991), *Statistics for Spatial Data*, Wiley.

Cumming, J. A. and Wooff, D. A. (2007), "Dimension reduction via principal variables," *Computational Statistics & Data Analsysis*, 52, 550–565.

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.

De Finetti, B. (1974), *Theory of Probability*, vol. 1, Wiley.

— (1975), *Theory of Probability*, vol. 2, Wiley.

Goldstein, M. and Rougier, J. C. (2006), "Bayes Linear Calibrated Prediction for Complex Systems," *Journal of the American Statistical Association*, 101, 1132–1143.

— (2007), "Reified Bayesian Modelling and Inference for Physical Systems," *Journal of Statistical Planning and Inference*.

Goldstein, M. and Wooff, D. A. (2007), *Bayes Linear Statistics: Theory and Methods*, Wiley.

Hartigan, J. A. (1969), "Linear Bayes methods," *Journal of the Royal Statistical Society, Series B*, 31, 446–454.

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.

— (2001), "Bayesian Calibration of computer models," *Journal of the Royal Statistical Society, Series B*, 63, 425–464.

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

O'Hagan, A. (2006), "Bayesian analysis of computer code outputs: A tutorial," *Reliability Engineering and System Safety*, 91, 1290–1300.

Qian, Z. and Wu, C. F. J. (2008), "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments," *Technometrics*, to appear.

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

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer-Verlag.

Stone, M. (1963), "Robustness of non-ideal decision procedures," *Journal of the American Statistical Association*, 58, 480–486.