Constructing partial prior specifications for models of complex physical systems

Peter S. Craig, Michael Goldstein, Allan H. Seheult & James A. Smith
Department of Mathematical Sciences
University of Durham
Durham DH1 3LE

Abstract

Many large scale problems, particularly in the physical sciences, are solved using complex high dimensional models whose outputs, for a given set of inputs, are expensive and time-consuming to evaluate. The complexity of such problems forces us to focus attention on those limited aspects of uncertainty which are directly relevant to the tasks for which the model will be used. We discuss methods for constructing the relevant partial prior specifications for these uncertainties, based on the prior covariance structure. Our approach combines two sources of prior knowledge. First, we elicit both qualitative and quantitative prior information based on expert prior judgements, using computer based elicitation tools for organising the complex collection of assessments in a systematic way. Secondly, we test and refine these judgements using detailed experiments based on versions of the model which are cheaper to evaluate. While the approach is quite general, we illustrate it in the context of matching hydrocarbon reservoir history.

KEYWORDS Bayes linear methods; computer experiments; elicitation tools; history matching; variable selection; covariance specification

1 Introduction

Many practical problems, particularly in the physical sciences, involve complex, high dimensional models, usually in the form of partial differential equations, which must be approximated by computer simulators. In many tasks for which such models are used, there are several major sources of uncertainty, so that it is important to combine analysis of the model with appropriate expert judgements. In principle, such uncertainties may be analysed using Bayesian methodology. In practice, the size and complexity of such problems may make prior specification very difficult.

An overview of general approaches to prior elicitation is given in [8], and practical developments of such elicitations are given in [11]. In contrast, this paper is concerned with ways to make detailed and informative prior specification for problems which are sufficiently complex that it is not feasible to quantify directly all of the uncertainties. Much of the difficulty comes from the complexity of the inter-relationships within the model, and the crucial challenge for Bayesian elicitation is to develop practical methods for identifying and quantifying the most important features of such joint variation.

In this paper, we discuss systematic approaches for quantifying those partial aspects of expert prior knowledge which are directly relevant for specific tasks. While elicitation is strongly context dependent, we suggest certain unifying ideas which may be helpful across a wide variety of applications, and which therefore throw light on the fundamental problems of structuring and eliciting prior beliefs for complex, high dimensional problems. We outline a general strategy for making such partial prior specifications, with provisional suggestions for its implementation. Our approach is based on assessing the prior covariance structure over the inputs and outputs of the model, by combining qualitative and quantitative expert judgements with information gained by careful experimentation on simple, approximate versions of the model, which are cheaper to evaluate. Such experimentation allows us to bring a wide variety of statistical ideas to the problems of creating informative prior specifications, while the linkage with expert judgement, through appropriate elicitation tools, ensures that the specifications are consistent with prior experience. We illustrate the approach with a particular practical application of importance in the oil industry.

2 Models and computer experiments

Computer models are widely used for analysing complex physical phenomena, in problems as diverse as circuit design, reservoir management and weather forecasting. Among the types of difficulty that arise are choosing inputs to tune the model using possibly noisy data, optimising certain criteria for outputs, and gaining information useful for general inferential and predictive purposes; for a discussion of statistical issues arising in the treatment of such models, see [12].

In this paper, we describe a general approach for specifying prior beliefs which is applicable to a wide class of such computer models. Our intention is to produce a methodology for simple, semi-automatic belief construction for routine analyses which are not overly sensitive to prior assessments, while allowing careful and detailed prior specifications for analyses where careful use of prior knowledge is important to achieve the objectives of the study.

In this account, we simplify the elicitation task by restricting attention to partial prior specifications, and specifically to means, variances and covariances. Thus, we shall describe methods for eliciting certain high dimensional covariance structures. Within the usual Bayes formalism, this would give a complete prior specification under the assumption of multivariate normality. Alternatively, the approach may be viewed as the first stage in a more detailed belief specification, where we begin by eliciting first and second order structure and then add more detailed probabilistic information, either by further elicitation or by some semi-automatic method, such as choosing a reference prior, which respects the specified second order structure plus any further constraints, such as positivity, that we may wish to impose, but otherwise attempts to be as neutral as possible with respect to the quantities of interest; see, for example, [1]. However, even without the assumption of normality, often the second order specification alone will be adequate for the task at hand, using Bayes linear methodology which is based on linear fitting and updating of beliefs. An overview of Bayes linear methodology is given in [5], while the foundational logic for using this methodology to make inferences based on second order prior specifications is discussed in [7].

The particular example which we use for illustrative purposes concerns history matching for hydrocarbon reservoirs. Here, the model for a reservoir is represented by a computer code, called a reservoir simulator, where inputs represent the physical description of reservoir geology, outputs correspond to the production at the wells, and we seek those input geologies which produce outputs consistent with the observed historical production record for the reservoir; for an overview of history matching, see [9].

History matching has many of the difficulties typical of large scale physical models: (i) very high dimensional input and output spaces; (ii) possible mismatches between the model and the physical system; (iii) measurement errors in the historical record; (iv) possibly many or no solutions which match the record. Further, each run of the simulator is very expensive in CPU time, so that we may only make a very restricted number of runs to achieve a match. For an introduction to the Bayes linear approach to history matching, see [3]. The illustrative case study to which we will refer is described in [4], and concerns the choice of aspects of reservoir geology to match observed pressure readings in a particular reservoir.

In history matching, we make several runs of the simulator, each with different choices of inputs, to seek an acceptable match. Therefore, to guide our search, we develop a prior specification which gives a simple form for the prior mean vector and variance matrix for the outputs, as a function of the inputs. This type of prior specification is useful for any high-dimensional problem in which we would like to evaluate the model at a variety of possible input settings, but where each such assessment is very time consuming.

3 Structuring Uncertainties

For high dimensional models, it is often difficult to find any coherent prior specification which is in rough agreement with expert prior knowledge. Our general approach is to build a coherent formal structure which is intended to be sufficiently flexible to incorporate the most important features of prior uncertainty, and then to quantify this structure. In this account, we are only concerned with the mean and covariance structure. We describe the relationships between model inputs and outputs by means of simple forms of description, such as quadratic surfaces in a small number of input variables, with appropriately correlated residual structures. Such prior specification is considered to be provisional; our only requirement is that it be adequate for the current problem.

We now describe the formal elements of the prior description. The model takes as input a vector $x = (x_1, ..., x_d)^T$ and returns an output vector $y = (y_1, ..., y_p)^T$. We write $y = y(x) = (y_1(x), ..., y_p(x))^T$ when we need to emphasise the dependence of y on x. We outline a way to structure our prior beliefs which is of particular use for any problem in which we will want to search over a high-dimensional input space; for example, in a decision analysis where we seek inputs which yield outputs optimising some utility criterion.

[**1**] Beliefs linking inputs x to outputs y are formulated as

$$y(x) = Mg(x) + \gamma(x) \tag{1}$$

where M is a highly structured matrix of unknown coefficients, $g(x)$ is a deterministic vector function of x, and $\gamma(x)$ is some random process in x, usually stationary, with prior expectation zero, which is uncorrelated with entries in M.

[**2**] For many problems, the matrix M is very sparse, as we judge that uncertainty about the value of each output quantity, $y_i(x)$, is principally influenced by a comparatively small number, r_i, of combinations of the input quantities, $g_{i1}(x), ..., g_{ir_i}(x)$ say, which we

term the active combinations of inputs for output y_i. In such cases, we further decompose $\gamma(x)$ into the sum of two uncorrelated components $\epsilon(x)$ and $\delta(x)$. The auto-correlation function for component i of $\epsilon(\cdot)$ depends only on appropriate combinations of the active inputs for y_i, while the auto-correlation function for $\delta(\cdot)$ is a very simple function of all the inputs, often zero, expressing the (hopefully small) residual correlation which has not been described by the active inputs.

[3] To specify (1), we must identify active combinations of inputs, structure the matrix M, quantify beliefs for the non-zero elements of M and specify the variance and auto-correlation functions for $\epsilon(\cdot)$ and $\delta(\cdot)$. Partly this comes from expert elicitation, using appropriate computer based elicitation tools. We may also construct simple approximations to the physical model, which are cheaper and faster to evaluate, but which we expect to preserve the qualitative relationships between the inputs and outputs. Here, we suppose that we have a single such approximating model, $y^*(x)$, with the same input and output space as the full model. More general combinations of approximating models may be treated similarly.

Corresponding to (1), we have, for the simple model,

$$y^*(x) = M^* g(x) + \gamma^*(x) \tag{2}$$

where we usually suppose that y and y^* have the same active combinations of inputs for each output, so that M^* and M have common non-zero entries, and that the process $\gamma^*(x)$ may be decomposed as the uncorrelated sum of $\epsilon^*(x)$ and $\delta^*(x)$, where appropriate. As we may make many evaluations of the simple model, we use these evaluations to form reasonable order of magnitude sample estimates $\hat{M}^*$, for M^*, and to assess the auto-correlation function for $\gamma^*(\cdot)$.

[4] These assessments are transformed to prior beliefs over (1) by expert elicitation of the uncertainties relating the two models. For this purpose, we may use the relations

$$\hat{M}^* = M^* + \hat{M}_0 \tag{3}$$

to link our judgements for the estimates $\hat{M}^*$ and the values M^*, where specifications for the random matrix $\hat{M}_0$ express beliefs as to the adequacy of the estimates for M^*. We link our judgements for the values M and M^* in the two models, using the relations

$$M = \Lambda \star M^* + M_0 \tag{4}$$

Here $\Lambda \star M^*$ is the matrix defined as $\text{vec}(\Lambda \star M^*) = \Lambda \text{vec}(M^*)$, where $\text{vec}(A)$ unravels the matrix A into a vector, Λ is a matrix of constants to be specified expressing our judgements about the systematic relationships between the two models, and specifications for the random matrix M_0 express our beliefs as to the degree of similarity of the two models. The elements of M_0 are uncorrelated with the elements of $\hat{M}_0$, and the elements of each are uncorrelated with the elements of M^*. As y and y^* have the same active combinations, all entries of M_0 and $\hat{M}_0$ corresponding to zero entries of M^* are zero.

If a full prior elicitation of the mean and variance structure of the terms in (3) and (4) is available from the expert, then we may determine the full Bayes linear adjustment for M given the assessments $\hat{M}^*$. Otherwise, we may choose simple approximations, based on, for example, identifying collections of elements of M which we judge to be qualitatively similar, and pooling the corresponding elements of $\hat{M}^*$ to give plausible order of magnitude assessments for M. We usually use similar parameter values for the auto-correlation structures for the residual processes $\gamma(x)$ and $\gamma^*(x)$, though we may rescale the variances of the processes and we may choose to introduce a correlation between corresponding components of γ and γ^*.

[**5**] The forms that we have described above tend to be parameterised by relatively few quantities which are hopefully sufficient to capture the most important aspects of prior relationships. However, as we have very high dimensional descriptions, even these specifications may impose a very heavy burden of elicitation. An important step in simplifying the elicitation is to make exchangeability assessments over groups of outputs for which, roughly, we consider uncertainties to be similar, so that we may make careful elicitations for an individual member of the group, and use the same assessments over the remaining elements of the group. Different aspects of the specification may lead to partially exchangeable groupings; for example, for some outputs influenced by a particular input, the elements of a corresponding column of M might be judged exchangeable, even though the output variables would not be exchangeable as they might be influenced by differing further input variables.

[**6**] In most models, some of the inputs represent control parameters, while others correspond to fixed, but possibly unknown, physical counterparts. In the latter case, we may further elicit a prior mean and variance structure or, at the least, a region of values within which we have high prior probability, for the values of the physical counterparts. We may also have an observation vector y_H corresponding to the output y_C of the model evaluated at input x_T, which corresponds to the true physical state. As there may be substantial differences between the physical system and the model for the system, and physical data used to calibrate the model usually has associated measurement errors, we link the observations y_H and the model outputs by the relations

$$y_H = y_T + y_E \qquad\qquad y_T = y_C + y_D \tag{5}$$

where y_T is the vector of true values of the physical quantities on which the observations were taken, y_E is the vector of observational errors, and y_D is the vector of differences between the model at x_T and the physical system. As observational errors are features of the measurement process, while the discrepancies between the model and the physical system are features of the modelling process, we usually make the simplifying assessment that y_C, y_D and y_E are mutually uncorrelated vectors a priori. For a typical problem, we might anticipate (i) detailed and relatively easily elicited beliefs about the components of y_E, which we may treat as uncorrelated, with zero mean; (ii) very large prior uncertainty over the value of y_C; (iii) strong correlations between certain of the elements of y_D, because closely related aspects of the model may be expected to differ from the physical system in similar ways. To produce a sensible joint specification over y_D may therefore require careful graphical modelling; see [6] for a discussion of the general approach to Bayes linear graphical modelling based on the identification of common underlying 'components of variation' which are unobservable, but meaningful, quantities over which we can partition uncertainties for observables.

In the illustrative study, the structure was as follows. The input x was a 40 dimensional vector of features of the geology of a particular reservoir, relating to permeabilities of various regions and the states of various potential faults, and the output y was a 77 dimensional vector of pressure readings for wells in the reservoir, recorded at various times. We had a vector y_H of historical pressure values. The aim of the study was to identify those collections of inputs x for which the output $y(x)$ was close to y_H on some appropriate scale. A particular region of inputs was identified by a geologist as likely to contain the vector x_T representing the true geology.

For each output variable, y_i, we selected (at most) three influential inputs, $x_{[i]} = (x_{i1}, x_{i2}, x_{i3})$, termed the active variables for y_i, and, with $x_{i0} = 1$, wrote (1) as

$$y_i(x) = \sum_{j=0}^{3} \sum_{k \leq j} \beta_{ijk} x_{ij} x_{ik} + \epsilon_i(x_{[i]}) + \delta_i(x) \tag{6}$$

where $\delta_i(.)$ and $\epsilon_i(.)$ were uncorrelated processes with

$$\mathrm{Cov}(\epsilon_i(x_{[i]}), \epsilon_i(x'_{[i]})) = \sigma^2_{\epsilon_i} \exp(-\theta_i |x_{[i]} - x'_{[i]}|^2) \tag{7}$$

where each θ_i was a scaling to be chosen, $\mathrm{Var}(\delta_i(x)) = \sigma^2_{\delta_i}$ for all x and we set $\mathrm{Cov}(\delta_i(x), \delta_i(x')) = 0$ for $x \neq x'$. A fast approximation to the reservoir simulator was constructed by reducing the number of layers in the model and combining neighbouring cells in the solver grid, resulting in a reduction in time for a single run from several days to a few minutes.

In general applications, the prior specification may be updated each time the full model is evaluated at a particular choice of input settings, for example by Bayes linear adjustment of each non-zero element of the matrix M. Formal adjustment of the variance forms in the prior specification requires additional, and more complex, prior elicitation. If we want to avoid this extra level of complexity, then we may develop simple approximate methods for such adjustments, but these are outside the scope of this paper.

In the following sections we discuss and illustrate firstly the use of elicitation tools, and secondly the use of simple versions of the physical model, in order to structure and quantify prior beliefs over the inputs and outputs of the model.

4 Elicitation tools and expert judgements

Within the structure described in the previous section, beliefs about the elements of (4) and (5) can only be obtained by elicitation from an expert. In this section, we discuss methods for eliciting qualitative and quantitative expert opinion. In the following section, we consider the role of experiments with the simple model in forming the prior description. In practice, these processes form an iterative cycle, in which the expert checks the plausibility of the prior descriptions suggested by experiments on the simple model, which in turn check aspects of the prior uncertainties suggested by the expert.

We now discuss the use of computer-based elicitation tools for the kinds of prior structure described in section 3. While elicitation can proceed without such tools, the structuring, support and feedback that they provide can greatly simplify the elicitation and lead to more reliable overall specifications. Thus, the challenge lies in designing tools which are sufficiently focussed to be useful for individual elicitations, while being sufficiently general that they may be reconfigured for a wide variety of applications. An example of an interactive graphical tool for a relatively simple elicitation task can be found in [2], which describes an elicitation method for the parameters of a bivariate distribution. For Bayes linear analyses, a general purpose elicitation tool called Foresight is described in [6].

For any physical model which has many inputs and outputs, there are very many quantities for which we need to elicit beliefs. An elicitation tool will have several purposes: (i) to provide an environment within which the elicitation can be managed and which ensures that all necessary beliefs are elicited and are recorded for later use; (ii) to simplify the elicitation by assisting the expert to form collections of related quantities, using, for example, judgements of partial exchangeability; (iii) to support the expert in the individual quantitative elicitation tasks, which are often complex; (iv) to aid reconciliation of prior information from diverse sources; and (v) to provide feedback on the joint implications of beliefs which have been individually specified.

We want to quantify mean and variance structure. To save space, we restrict attention to aspects of the uncertainties arising in the relations (1), which are are well suited to show some of the potential of elicitation tools for supporting the task of prior specification.

4.1 Eliciting overall structure

We have so many relations to elicit from the expert that, to avoid becoming overwhelmed with details, we make detailed elicitations for certain 'strategic pairs' y_i and x_j of output and input variables which are judged typical of the relationships that we wish to elicit, and make all other elicitations by comparison with these strategic elicitations, exploiting partial exchangeability judgements over related quantities. An elicitation tool should assist the expert to make such judgements.

There are many different types of belief to be specified, which may include magnitudes of measurement errors, collections of active inputs for each output and the details of the influence on outputs of each of those inputs. The expert must organise into groups, outputs which are deemed to be effectively exchangeable in relation to each type of belief. Only a single quantitative elicitation, to be described later, is then required for each group. For example, the influence of a particular input may be deemed to be largely the same for a group of outputs. By eliciting beliefs about its influence for a single output, we are effectively eliciting beliefs for every output in the group. We may also elicit beliefs about the difference in its influence on a general pair of members of the group, and so construct the full second order exchangeability representation for the group (see [5]).

The likely origin of most such exchangeability statements is that some aspect of the structure of the physical system is similar for a number of outputs. For fluid flow models, such as those used for hydrocarbon reservoirs, spatial and temporal proximity between measurements is an obvious similarity. A less obvious similarity is that between outputs for which the same measurement method was used. A useful elicitation tool must expose the physical structure in a way which makes it easy to group similar outputs together. In the tool developed for the illustrative study, we show spatial proximity by a map and temporal proximity by a dot-plot of the times of measurements of outputs (see figure 1). In other contexts, other metaphors may be used for the physical structure. For example, where the model describes a mechanical system, a schematic diagram of the system might replace the map and time plot.

Whatever metaphors are being used, the tool should make it easy for the expert to indicate which outputs are to be formed into a group and to provide designations for groups which facilitate recognition of individual groups a later stage. An important concept is that of the *selection*, a notion used in much software that has a graphical user interface. The selection is a collection of outputs (or sometimes of inputs), which are highlighted wherever they appear in the tool. The selection makes a visual connection between windows which display outputs and inputs organised in various different ways. It seems natural that the interface for creating and modifying the selection should be mouse-based, using the usual techniques of pointing, clicking and dragging.

As part of our work on history matching, we have developed an elicitation tool which we describe below. Obviously, the windows which provide the physical metaphors are tuned to that context, although it would be straightforward to adapt them to many other physical contexts. The rest of the interface and code, for eliciting quantitative beliefs and manipulating groups of outputs, are designed for generality. For simplicity, we describe them in the context of the history matching tool.

Figure 1 shows a schematic diagram of the tool displaying a simple fictitious reservoir with three regions, thirteen wells and six faults. Four windows are shown:

(i) The *effects elicitation window* is used to elicit active inputs and their effects for each output. The window has three panels. The top panel shows the groups into which the outputs are organised. Outputs in the same group share the same active variables and beliefs about their effects. Each group has a menu of possible actions which include adding and removing outputs or active inputs, and elicitation of beliefs about effects of changing active inputs. New groups may be created or existing ones eliminated at any stage. The name of a group will be partially or completely highlighted if some or all of its members are part of the selection. In the figure, there are three groups shown, namely the right hand side (RHS) of the reservoir, the middle of the reservoir and well 2, the last of which is highlighted. The middle panel shows a list of all outputs. Outputs which have already been placed in a group appear in italics. Outputs which are part of the selection, in this case the two outputs for well 2, are highlighted. The bottom panel lists inputs, using secondary highlighting for those active for any of the outputs in the selection. In this figure, permeability 1 (the permeability for the region containing well 2) and fault 2 (controlling transmissibility of the fault adjacent to well 2) are highlighted. We may also reverse the roles of inputs and outputs, allowing the selection to be a collection of input variables and we then secondarily highlight all outputs for which any of the selected inputs are active.

(ii) The *map window* shows the map of the reservoir, and is used for display and selection of inputs and outputs. Well 2, at the bottom of the left region is highlighted because the group of associated outputs is currently selected. The fault just above and the region surrounding well 2 are secondarily highlighted as their associated inputs have been selected as active for the outputs at the selected well. The groups shown in the effects window may naturally be created by making selections in the map window using the mouse.

(iii) The *time plot window* shows the times of the historical measurements. Two are highlighted, namely the two readings for well 2.

(iv) The *window manager* controls which windows are visible at any time. The windows which are not shown in the figure are windows for elicitation of other quantitative details such as measurement error magnitudes.

When we are satisfied with the qualitative structure, we then carry out the quantitative elicitation for the group, by eliciting beliefs about the effects of changing active inputs on an individual member of the group.

4.2 Quantifying individual prior judgements

We elicit the expert's beliefs about changes in y_i as we vary x_j, when all of the remaining input variables are fixed at some initial prior values. Specifically, we elicit beliefs about $y_i^-(x_j) = y_i(x_j) - y_i(\hat{x}_j)$, where $\hat{x}_j$ is some fixed, possibly central, value for x_j. It is possible to elicit beliefs directly about values of y_i, but we found that our experts had quite strong beliefs about the effects of changing input values while being fairly uncertain about the overall level of individual outputs. We believe that this may be typical for many physical models. We suppose that we have already chosen the vector $\hat{x}$. For those inputs which correspond to physical parameters, a natural choice for $\hat{x}$ might be the expert's 'best guess' for x_T, for example as a prior mean or mode. We may also choose to represent the scales so that we are eliciting the percentage change in y_i for a given percentage change in x_j. The rationale is that some outputs may be partially exchangeable on percentage scale when they are not in the original units.

To simplify the exposition, we will suppose that there are two terms, $g_1(x_j)$ and $g_2(x_j)$ involving x_j in relation (1) which determine changes in y_i for these fixed values of the remaining terms in x, and that we may transform the scale so that each $g_r(\hat{x}_j) = 0$, so that we may write

$$y_i^-(x_j) = y_i(x_j) - y_i(\hat{x}_j) = m_1 g_1(x_j) + m_2 g_2(x_j) + \gamma(x_j) - \gamma(\hat{x}_j) \qquad (8)$$

where m_1 and m_2 are the elements of M corresponding to terms g_1 and g_2. In the illustrative study, we restricted attention to quadratic surfaces, in appropriately transformed versions of the input variables, so that under this scaling, $g_1(x_j) = x_j$ and $g_2(x_j) = x_j^2$.

Our elicitation tool shows two axes, y_i^- and x_j. The width of the x_j scale corresponds to the a priori range of possible values of interest. We first describe, hypothetically, how we might make a prior elicitation for (8) by prior specification strictly over observable quantities. Subsequently, we describe a less direct, but possibly simpler, method which introduces unobservables.

To elicit for observables, we would begin by asking: given a certain change in x_j alone, what is the expert's expectation for the change in y_i? We would also ask the expert to provide prior confidence bands around the mean curve, choosing whatever percentage is most natural (usually 50%, 66% or 95%). The expert would either draw these curves on the computer screen, or simply specify beliefs about $y_i^-(x_j)$ for certain selected values of x_j, from which we would extrapolate smooth curves; both methods are difficult to implement. The expectation curve would give the prior means for m_1 and m_2 provided that it roughly followed the form $\hat{m}_1 g_1(x_j) + \hat{m}_2 g(x_j)$ for some values $\hat{m}_1$ and $\hat{m}_2$, as $\gamma(x_j)$ has zero prior mean. If no such curve roughly corresponded to the assessed values then we would have to consider either transforming the scales of the variables or introducing new active input combinations. The confidence bands might be used to deduce the covariance structure of m_1 and m_2 and γ. Beliefs are required for at least four values of x_j to provide sufficient information to deduce the covariance structure of m_1 and m_2 and a single variance parameter for γ. If γ has more than one auto-covariance parameter, correspondingly more beliefs are required.

As far as possible, we prefer to elicit uncertainties about observable quantities, because these may be specified on the basis of meaningful and unambiguous questions. However, the procedure outlined above is potentially highly sensitive to small changes in elicited values, as determining the mean and variance structure from the position of the confidence bands for a number of values of x_j is an exercise in curve fitting. Further, the expert will often be uncertain about the bands, but will still have beliefs about the overall shape of the function. Then, particularly when direct elicitations implicitly demand sophisticated calculations, it is helpful to elicit beliefs about convenient intermediary quantities. Therefore, we offer a complementary approach to quantifying these variances, where the expert specifies uncertainties for m_1, m_2 directly, using feedback about the resulting assessments for uncertainty about y_i^- to assist the process. This is the approach that was used for the illustrative study.

Figure 2 gives a schematic view of an elicitation tool which allows us to build up uncertainties directly about the elements of the curve of best (for example, minimum mean squared error) fit in $g_1(x_j)$ and $g_2(x_j)$. The top window shows the expectation curve for $y_i^-(x_j)$, which in this case is $\mathrm{E}(m_1)x_j + \mathrm{E}(m_2)x_j^2$, and bounds around the curve which represent either a certain percentage of the prior probability for the value or a certain number of prior standard deviations. These bounds are assessed by direct elicitation of the variance and covariance terms of m_1 and m_2 and of σ_ϵ^2, each of which is elicited in its own section of the screen. This part of the elicitation does not incorporate variation due to $\delta(x)$ in (1), as we are thinking about changes only to active inputs. Thus, two screens show the contribution to overall uncertainty arising from uncertainties about the values of the best fit coefficients m_1 and m_2. In each case, the heavy solid curve shows the mean around which bounds are being chosen. The other solid curves show the chosen bounds while the dashed and dotted curves suggest possible choices for the bounds. The expert is able to change the mean and bounds curves by pointing using a mouse and clicking an appropriate button. On the computer screen, we are able to use colour to assist the user in identifying these different components. In the version shown, we use a rationale of progressive introduction of complexity. The linear screen ignores curvature and concerns only the line of best fit. The quadratic screen incorporates the mean for the linear term in the mean curve but the bounds only involve changes to the quadratic term. Finally, we use sliders to control both the correlation between m_1 and m_2 and also the value of σ_ϵ^2. Often, we may keep $\mathrm{Corr}(m_1, m_2)$ at zero, and usually we look for prior descriptions in which variation in ϵ is a small part of the total uncertainty.

Other elicitations may proceed similarly. For example, we may elicit 'interactions' by asking whether there is any further input, x_k, for which changing the fixed value of this input to some other value, would change any of the above quantifications. In the case study, we introduced a new bi-linear term for each interaction, so, for example, if the expectation line does not change but the variance bounds increase when x_k changes from zero, then this suggests an interaction, $g_3(x) = x_j x_k$, with coefficient m_3 having zero prior expectation and variance derived from the change in variance bands.

Our experience in the illustrative study with our elicitation tool was positive. The reservoir engineer from whom we elicited beliefs found the tool very useful because it enabled him both to visualise the reservoir and to interact with a map of the reservoir to indicate the outputs to which his statements referred. He found it natural to specify which inputs were likely to influence individual outputs and natural also to organise the outputs into various groups. He seemed happy to express beliefs about the effects on outputs of changing the active inputs, provided that such beliefs were about percentage changes in outputs.

5 Using a simple model to generate empirical prior descriptions

The elicitation outlined in section 4 may be difficult and time-consuming, and we seek a systematic and partially automated approach to the acquisition of additional prior information. Where this is feasible, such information may be obtained from a large number of evaluations of an approximate simpler version of the model, which we believe to be qualitatively similar to the full model. Judgements as to the degree of similarity between the two models must be elicited from the expert, and may, for example, be refined by observing discrepancies between the outputs of each model at common input runs. The data from the evaluations of the simple model are used to develop descriptions relating each model output to a small collection of inputs and where appropriate to supplement and verify prior information elicited from the expert. In particular, $\hat{M}^*$ and $\text{Var}(\hat{M}_0)$ in (3) are obtained from the simple model evaluations. In this section, we discuss the issues associated with generating such empirical prior information and combining it with that elicited from the expert.

[1] The relative cost of an evaluation of the simple model, based on, for example, typical computer run times is usually small. In general, the number of evaluations of the simple model will be guided by the number of inputs, the cost of the runs and the discrepancies between the two models. In the illustrative study, we could afford 200 runs of the fast simulator for the cost of two full simulator runs. As there were 40 input variables, 200 runs was considered sufficient to obtain an adequate description of each pressure output in terms of a few of the inputs.

Having decided on the number of simple model evaluations, we need to choose a design for the settings of the inputs x. Ideally, this design should be able to identify any subset of inputs as important for any output. One such design, suggested by the high dimensionality of x, is a latin hypercube, which varies all the components of x simultaneously; see, for example, [10]. Essentially, their construction for an n–point design is to (i) partition the range of each component of x into n intervals; (ii) select an interval at random for each component; (iii) choose the first design point to be either a point at random in the selected n-dimensional hypercube or a fixed point such as its centre of gravity; and repeat the selection procedure without replacement for each component of the remaining $n-1$ design points. The resulting randomly chosen design has high probability of having approximate orthogonality properties, which are important for model selection.

When the number of inputs is large, the expert's beliefs as to which subsets of the inputs affect each output are an important consideration for design choice. In the illustrative study, the expert held the general belief that well pressures depended strongly on local permeabilities. We might therefore choose a design which is efficient for certain important components of y which we expect to depend on a small and similar collection of combinations of inputs, such as a factorial design over these inputs, possibly in combination with a latin hypercube design over the other inputs. When the overall collection of common inputs is large, a highly fractionated factorial design could be used.

We may further choose to combine two designs sequentially. For example, we might use a latin hypercube design for constructing descriptions of the relationships between inputs and outputs, followed by a second design to evaluate the residual covariance structure. The descriptions built from the output to the first design may identify a small number of active combinations of inputs which are informative for variation in a large number of outputs. In the second design we can vary those combinations, while holding all the other inputs fixed at nominal settings; for more detail, see [**4**] of this section.

[**2**] We now discuss how to quantify beliefs about the elements of (1) and (2) assuming that we have already chosen the active combinations of inputs for each output. We will discuss how the choice of active combinations should be made in [**3**].

Let us suppose we know the active combinations of inputs. We now consider how to estimate the matrix of coefficients M^*, as given in (2), and then to adjust beliefs about M, as given in (1). If we have sufficient evaluations of the simple model, then we can use the sample-based estimates for the component models to assess the coefficients in M^* and the auto-correlation function of γ^*. Otherwise, we may use exchangeability judgements over component models to pool these data to improve accuracy of the coefficient estimates. In the illustrative study, all the components of y were pressure readings, and, as part of the preliminary elicitation with the expert, well location and time of measurement suggested natural groupings of similar outputs, which were subsequently borne out by the analysis of the simple model evaluations.

Having produced a sample estimate $\hat{M}^*$ for M^*, we must now assess prior means and variances for the elements of M. As M and M^* are linked by relation (4) through Λ and M_0, Bayes linear adjustment of M by $\hat{M}^*$ is straightforward when both the coefficient matrix Λ is specified and the second order prior structure for the matrices in (3) and (4) has been fully assessed. If this requires too detailed a level of prior specification to be practical for a given application, then we may make pragmatic simplifications. For example, when n is large, taking M^* to have mean $\hat{M}^*$ and variance matrix equal to the corresponding sample variance matrix of the estimates $\hat{M}^*$ are reasonable general pragmatic choices. Additionally, in the illustrative study, we chose Λ to be the identity matrix and M_0 to have prior mean zero. The variances of the diagonal elements of M_0 were assessed as a percentage of the sample estimates, where the percentage difference was chosen to reflect the expert's judgement as to the degree of difference between the two models.

Correlations between the elements of M_0, and thus between those of M, are assessed by combining two sources of information — initial discussion with the expert and data analysis of the descriptions fitted to the evaluations on the simple model. As an example of the former, in the illustrative study we tried a covariance function which was a product of exponential forms, one for time and one for distance. The distance component of the covariance function was suggested from discussions with the expert. Analysis of his beliefs suggested a spatial correlation of the form $\exp(-6d^2)$ where d is well separation in kilometres, adjusted for intervening faults. A similar calibration was used for pressures taken at different times. As an illustration of the second source of information, many of the collections of active variables $x_{[i]}$ in the illustrative study contained a particular important input x_3. We therefore looked for groupings among the sample coefficients $\{\hat{\beta}_3\}$ for the linear terms in x_3 in the component descriptions. Figure 3 shows a plot of the collection of sample coefficients $\{\hat{\beta}_3\}$ against the times of the pressure readings in which this input x_3 occurs. The plot indicates a strong clustering of the coefficients corresponding to groups of wells in the same region at any given time: all negative coefficients correspond to offshore wells and decrease with time, whereas all the positive ones correspond to onshore wells and increase with time, suggesting natural groupings of strongly related $\hat{\beta}_3$ coefficients. Further analysis indicated a strong spatial pattern within each cluster. Such discoveries suggest correlation structure between the corresponding elements of M_0 in (4). These suggestions may then be fed back to the expert, emphasising the iterative nature of the elicitation process whereby the expert provides initial beliefs, which are used to direct the data analysis on the simple model evaluations which check and refine those prior assessments and may suggest further structuring of beliefs by, for example, suggesting simplifying exchangeability assessments or inducing correlation structure.

[3] We now consider how to select, for each output y_i, a small active subcollection $x_{[i]}$ of the inputs x. Thus, we start with a "sparsity of effects" hypothesis that variation in each y_i can be mostly explained by combinations of just a few components of x. The validity of the hypothesis can be judged to some extent by analysis of regression residuals. In the illustrative study, quadratic models with at most three inputs, as in (6), explained a large percentage (90% - 99%) of variation in most of the 77 outputs, as measured by the R^2 values for the ordinary least squares fits.

We can identify $x_{[i]}$ in two stages. First, for each y_i we use a straightforward, easy to compute, fast screening process on the simple model evaluations to eliminate those inputs (usually a pre-specified number) which account for just a small percentage of the total variation in y_i. Secondly, we use a tractable and computable task-specific criterion to choose $x_{[i]}$ from among those inputs not eliminated in the first stage, possibly also discounting combinations of inputs considered infeasible by the expert. The expert's beliefs about $x_{[i]}$, where available, may be used as a starting point for the first stage. Sometimes we may want to include terms in the prior descriptions that the expert believes should be present in the full model, even though they may not have been selected using the simple model evaluations alone; for example, part of a study might be to examine the effects of certain inputs on particular outputs. Alternatively, the expert might judge that a collection of outputs should all have the same input terms present in their description, allowing us to combine the selection criteria for the subset, and thus reduce the number of evaluations in the first stage search.

Further simplifications arise when certain combinations of inputs account for substantial variation in output. These might be identified from one of, or a combination of (a) expert judgements; (b) data transformations; and (c) data analysis: for example, a regression, in which many of the inputs have similar coefficients, might suggest that the sum of the corresponding inputs is an important combination.

In the illustrative study, we used the stepwise routine of S-Plus ([13]) in the first stage to eliminate all but six inputs for consideration. As the objective was to find inputs x to match each model output $y_i(x)$ to a corresponding reservoir pressure y_{T_i}, we used a model-choice criterion that measured the ability of the model to identify input domains that were unlikely to lead to a pressure match. Specifically, for each i we chose that group of three input variables $x_{[i]}$ from among the six from the first stage which maximised

$$E\left(\int \frac{E_D^2(y_i(x)-y_{T_i})}{\mathrm{Var}_D(y_i(x)-y_{T_i})}\,dx\right) \approx \int \frac{\mathrm{Var}(y_i(x))-\mathrm{Var}_D(y_i(x))+(\mathrm{E}(y_i(x)-y_{T_i}))^2}{\mathrm{Var}_D(y_i(x))+\mathrm{Var}(y_{E_i})}\,dx \quad (9)$$

where D refers to the collection of potential outputs of y_i at the points in a pre-chosen simple design in the six input components under consideration, and where E_D and Var_D are the Bayes linear adjusted mean and variance as described in [5] based on prior model forms (6) and (7) using the chosen $x_{[i]}$. The expression on the right-hand-side of (9) shows that the criterion can be straightforwardly evaluated using only our prior second order belief specification (including the variance of the measurement error), as the value of $\mathrm{Var}_D(y_i(x))$ only depends on the choice of design and not on the values of the potential outputs. The criterion in (9) is an integrated measure of the 'implausibility' that each x is a 'match point' (measured by the coefficient of variation of $y_i(x) - y_{T_i}$ adjusted by D)

and therefore selects models which we expect to be well-suited for identifying large input domains which are unlikely to produce good pressure matches.

Note that, using a criterion such as (9), we need prior assessments for $\mathrm{Var}(M_0)$ and $\mathrm{Var}(M)$ for all possible combinations of each choice of three inputs from the six selected at the first stage. To elicit coefficients for all possible terms in the model, allowing for the possibility that all of the entries for M and therefore M_0 can be non-zero, might impose a large burden of elicitation. To reduce this burden, we would try to group coefficients into a small number of exchangeable groups. An alternative is to make the necessary prior elicitation between the first and second stages of the screening process, so that we only need to elicit for those combinations which are suggested by the preliminary screening.

[**4**] We now consider the γ^* term in (2), begining with the choice of covariance structure for a component y_i of y. We have already discussed in [**1**] the possibility of using special designs for the simple model to facilitate estimation of the covariance structure of γ^*. In general, given appropriate choice of the second stage design, this becomes a problem in spatial estimation. For example, we might estimate the parameters of the covariance process in (7) for a given component $\gamma_i^*(x) = \epsilon_i^*(x_{[i]}) + \delta_i^*(x)$ by (i) holding $x_{[i]}$ fixed at nominal values and varying the components in its complement $x \sim x_{[i]}$ to estimate the variance of δ_i^*; (ii) holding the complement fixed at nominal values and vary $x_{[i]}$ to estimate the variance of ϵ_i^*; and (iii) using the resulting residuals to estimate the autocorrelation parameter θ_i. Exchangeability judgements allow us to share assessments of parameters of the γ^* process across large subclasses of outputs.

The sample correlation matrix of the residuals for the component models identifies those sub-collections of γ_i^* with large correlations. It is comparatively straightforward to create prior correlation structures between residual components that have common active combinations of inputs and similar correlation forms. However, obtaining a coherent second order structure in general is far from easy, and requires further study.

Having assessed the parameters in γ_i^*, we next assess parameters for γ_i. Subject to scrutiny of the expert, we might use a form for γ_i in (1) similar to that we assess for γ_i^*. For the special form in (7), we would typically assess θ_i to be same for both, but the variances for γ_i may be different and most likely larger than those for γ_i^*, reflecting our doubts about the qualitative similarity between the two models.

In the illustrative study, the prior specification obtained for the full model using the above ideas was sufficiently informative to lead to a satisfactory pressure match in an acceptable number of runs of the full reservoir simulator.

6 Reconciliation and refinement of beliefs

A natural elicitation sequence is first to establish a provisional prior structure using expert judgements and then to modify these judgements in the light of experiments with a simple model as we have described. We then use the elicitation tool again to give feedback to the expert as to the nature and consequences of the new prior descriptions, by qualitative assessment of the new structure and by quantitative comparison of the new uncertainties for various observables with direct assessments of variation.

A prior specification, once created, may be updated, for example using the Bayes linear approach, based on evaluations of the model. Eventually, we may decide that such routine updating of the specification is no longer adequate. This may be because of poor performance of the specification; for example we may observe substantial discrepancies between observations and predictions. Alternatively, we may have sufficiently reduced uncertainty over the elements of M that it is now practical to identify additional active inputs, or we may change the task for which the prior specification is intended. At such time, we may carry out the specification process again, based on further analysis of the simplified model, expert elicitation and whatever evaluations we have made of the full model. For example, in history matching, we found it useful to re-specify our beliefs using a further collection of evaluations of the simple model within a sub-region of the input space.

7 Concluding comments

It is often considered to be difficult to elicit prior means and variances directly. However, for very high dimensional problems the mean and covariance structure may be the most useful collection of quantities for which to elicit. We have described, in outline, an approach to such elicitation based on the synthesis of two types of prior assessment.

First, although the task is to construct a subjective prior specification, we may find it very helpful to construct this specification, at least in part, from careful statistical analysis of data from related contexts. The approach that we have outlined allows fast, routine, semi-automatic assessment of prior specifications by using large samples on the simple model to generate prior forms which may be turned into prior descriptions on the full model by exploiting partial exchangeability judgements.

Secondly, for complex problems we use elicitation tools which are designed to integrate the subjective assessments of the expert with any such auxiliary statistical analysis. Therefore, the approach also encourages, by use of such tools, a more careful and interactive approach, which is particularly appropriate if the expert is involved in a sequence of such elicitations through which experience and expertise may progressively be gained. In such a detailed prior specification, the expert tries out various alternative prior descriptions which are tested and refined on the simple model, and then further refined by the expert, with respect to both the individual quantifications of uncertainty and the qualitative structuring of the prior description.

This approach raises a number of questions. Are our prior forms adequate for the tasks they address? How reliable is our approach to using the simple model? What are the most useful forms for the elicitation tools? These questions are problem-specific, but should also be judged in the context of the wider issues that arise when considering prior elicitation for complex systems. How much detail should we expect to capture in our prior description? How much trust should we place in the resulting elicitations? How much effort is it worth expending to improve our elicitations? How should we weigh the benefits of routine elicitation methods against those which are specifically tuned to particular tasks? The general answers to such questions may only be gathered by experience in large structured elicitations. The approach that we have described is intended to offer a framework within which such questions may be addressed by describing how we may begin to construct informative prior specifications for such large and challenging problems.

Acknowledgements

We are especially grateful to our industrial collaborators at Scientific Software-Intercomp (UK) for their support, especially Bob Parish, whose support made the case study possible, and Adam Little, whose beliefs we elicited for the study. This work has been supported by a United Kingdom Engineering and Physical Sciences Research Council grant under their Complex Stochastic Systems Initiative.

References

[1] Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory.* Chichester: Wiley.

[2] Chaloner, K., Church, T., Louis, T.A. and Matts, J.P. (1989) Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*, **42**, 341-353.

[3] Craig, P.S., Goldstein, M., Seheult, A.H. and Smith, J.A. (1996) Bayes linear strategies for matching hydrocarbon reservoir history (with discussion). In *Bayesian Statistics 5* (eds J.M.Bernardo, J.O.Berger, A.P.Dawid and A.F.M.Smith), pp.69–95. Oxford: Oxford University Press.

[4] Craig, P.S., Goldstein, M., Seheult, A.H. and Smith, J.A. (1995) Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments. To appear in *Case studies in Bayesian Statistics III*. Springer.

[5] Farrow, M. and Goldstein, M. (1993) Bayes linear methods for grouped multivariate repeated measurement studies with application to crossover trials. *Biometrika*, **80**, 39-59.

[6] Farrow, M., Goldstein, M., and Spiropoulos, T. (1997) Developing a Bayes Linear Decision Support System for a Brewery. To appear in *The Practice of Bayesian Analysis* (eds S. French and J.Q. Smith). Edward Arnold.

[7] Goldstein, M. (1996) Prior inferences for posterior judgements. To appear in the proceedings of the 10th International Congress of Logic, Methodology and Philosophy of Science (eds M.L.D. Chiara et al). Kluwer.

[8] Kadane, J.B. and Wolfson, L.J. (1996) Experiences in elicitation. In submission as a discussion paper for JRSSB.

[9] Mattax, C.C. and Dalton, R.L. (1990) *Reservoir Simulation*. Society of Petroleum Engineers.

[10] McKay, M.D., Conover, W.J. and Beckham, R.J. (1979) A comparison of three methods for selecting values of input variables in the analysis of output of computer code. *Technometrics*, **21**, 239-245.

[11] O'Hagan, A. (1996) Eliciting expert beliefs in substantial practical applications. In submission as a discussion paper for JRSSB.

[12] Sacks, J. Welch, W., Mitchell, T. and Wynn, H. (1989) Design and analysis of computer experiments. *Statistical Science*, **4**, 409-435.

[13] Statistical Sciences Inc. (1993) *S-PLUS Reference Manual, Version 3.2.* Seattle: StatSci, a division of MathSoft Inc.

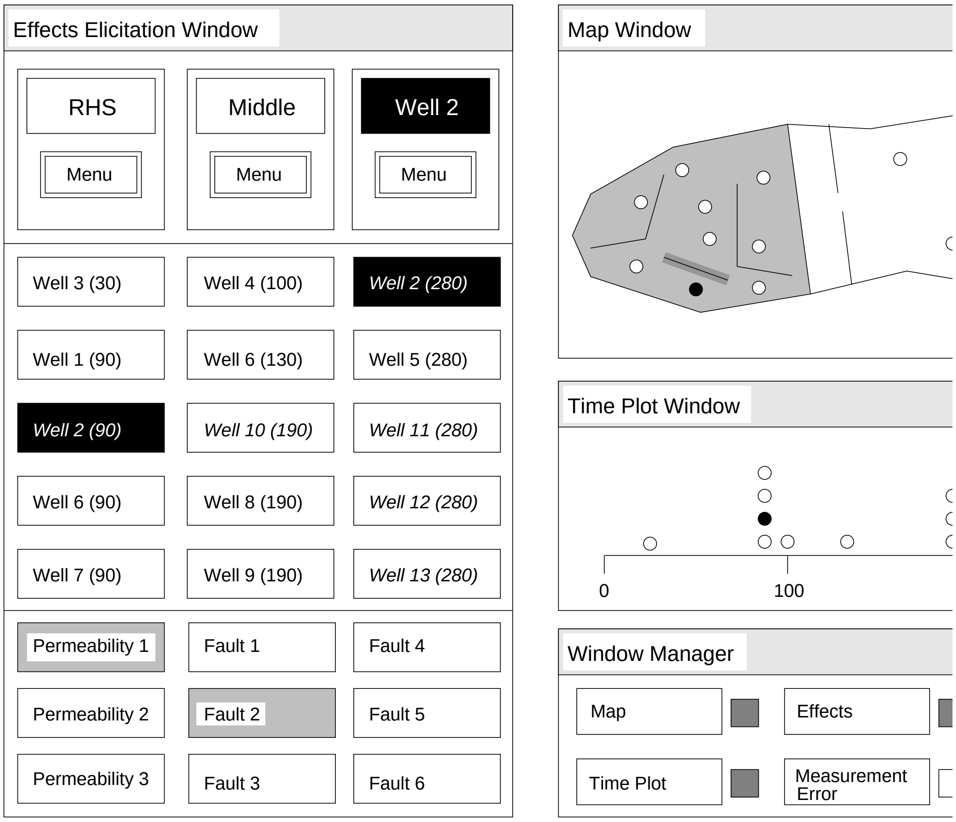

Figure 1: Schematic diagram of an elicitation tool for history matching.

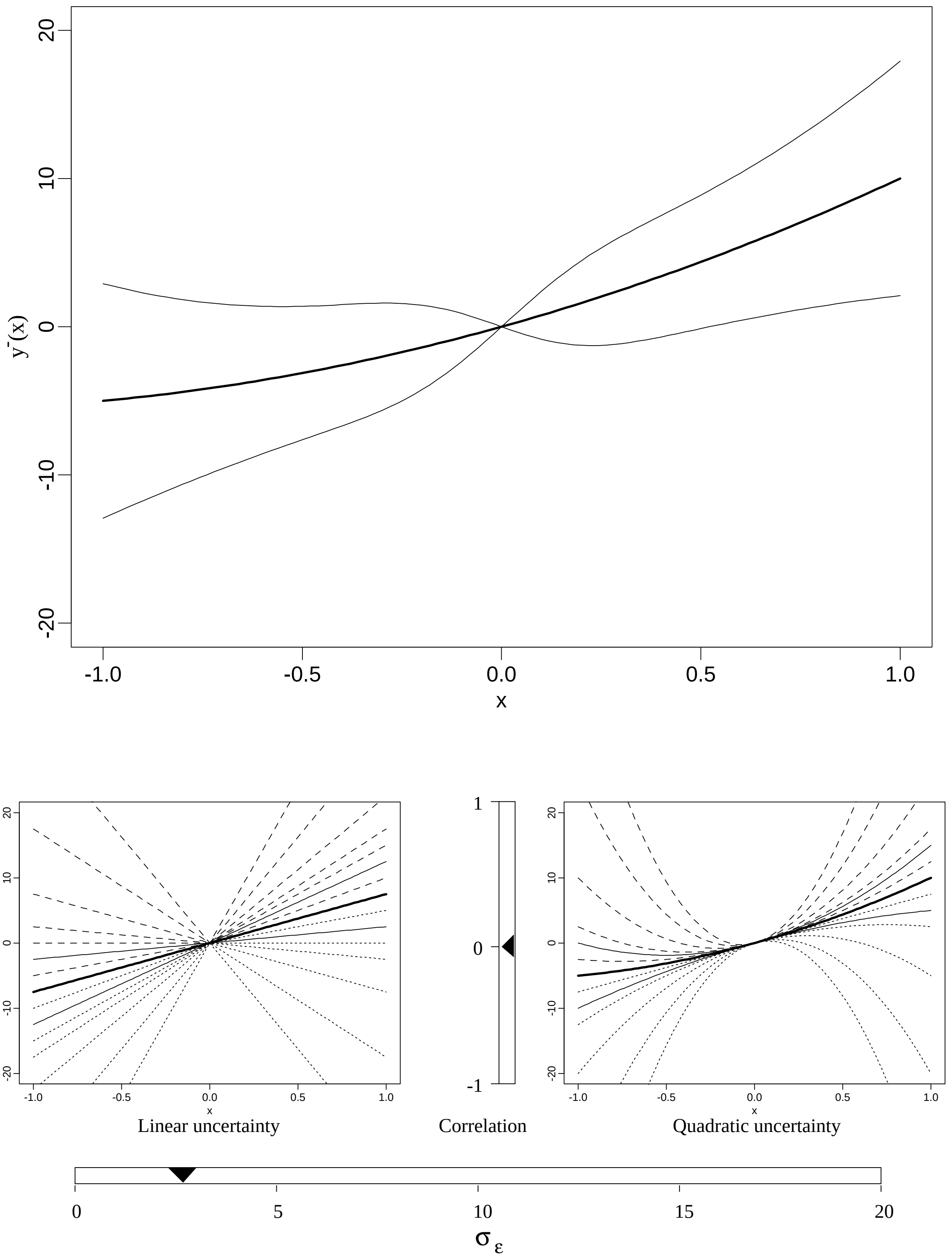

Figure 2: Diagram of a tool for elicitating beliefs about the effects on an output variable of changing an input variable.

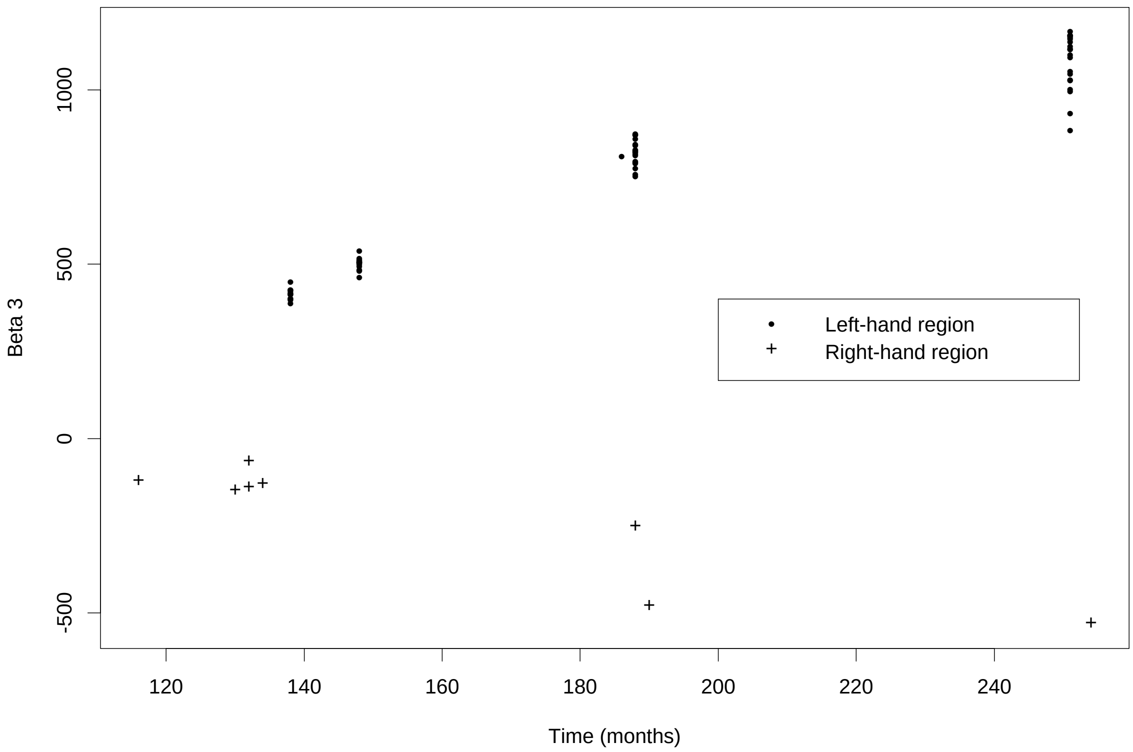

Figure 3: Temporal distribution of estimates from the approximate model of the coefficients of one input variable.