J. R. Statist. Soc. B (2004)
66, *Part* 3, *pp.* 751–769

Probabilistic sensitivity analysis of complex models: a Bayesian approach

Jeremy E. Oakley and Anthony O'Hagan

University of Sheffield, UK

[Received May 2002. Revised December 2003]

Summary. In many areas of science and technology, mathematical models are built to simulate complex real world phenomena. Such models are typically implemented in large computer programs and are also very complex, such that the way that the model responds to changes in its inputs is not transparent. Sensitivity analysis is concerned with understanding how changes in the model inputs influence the outputs. This may be motivated simply by a wish to understand the implications of a complex model but often arises because there is uncertainty about the true values of the inputs that should be used for a particular application. A broad range of measures have been advocated in the literature to quantify and describe the sensitivity of a model's output to variation in its inputs. In practice the most commonly used measures are those that are based on formulating uncertainty in the model inputs by a joint probability distribution and then analysing the induced uncertainty in outputs, an approach which is known as probabilistic sensitivity analysis. We present a Bayesian framework which unifies the various tools of probabilistic sensitivity analysis. The Bayesian approach is computationally highly efficient. It allows effective sensitivity analysis to be achieved by using far smaller numbers of model runs than standard Monte Carlo methods. Furthermore, all measures of interest may be computed from a single set of runs.

Keywords: Bayesian inference; Computer model; Gaussian process; Sensitivity analysis; Uncertainty analysis

1. Introduction

Consider a deterministic model that is represented by $y = \eta(\mathbf{x})$, where $\mathbf{x}$ is a vector of input variables and y is the model output. We suppose that $\eta(\cdot)$ is a complex model, such that the way that the model responds to changes in its inputs is not transparent. Sensitivity analysis is concerned with understanding how changes in the inputs $\mathbf{x}$ influence the output y. This may be motivated simply by a wish to understand the implications of a complex model but often arises because there is uncertainty about the true values of the inputs that should be used for a particular application. We have a 'base-line' or central estimate $\mathbf{x}_0$ for $\mathbf{x}$ but are then interested in how the true output $y = \eta(\mathbf{x})$ might differ from the base-line output $y_0 = \eta(\mathbf{x}_0)$.

Local sensitivity analysis is based on derivatives of $\eta(\cdot)$ evaluated at $\mathbf{x} = \mathbf{x}_0$ and indicates how y will change if the base-line input values are perturbed slightly. This is clearly of limited value in understanding the consequences of real uncertainty about $\mathbf{x}$, which would in practice entail more than infinitesimal changes in the inputs. Global sensitivity analysis considers these more substantial changes in $\mathbf{x}$. However, there is then the question of how far to perturb individual inputs. Perturbing each input to the limits that might be considered plausible gives some kind of limits of plausibility for y, but the resulting range is usually unrealistically wide if many inputs

Address for correspondence: Jeremy E. Oakley, Department of Probability and Statistics, University of Sheffield, Sheffield, S3 7RH, UK.
E-mail: j.oakley@sheffield.ac.uk

 1369–7412/04/66751

are perturbed together, or unrealistically narrow if they are only perturbed individually. Such difficulties are overcome by acknowledging the uncertainty in $\mathbf{x}$ and formally treating it as a random variable with a specified distribution.

Thus, denoting the unknown true inputs by $\mathbf{X}$, it follows that the corresponding output $Y = \eta(\mathbf{X})$ is also unknown. We suppose that our uncertainty about the elements of $\mathbf{X}$ is described by some probability distribution G. The approach to sensitivity analysis which exploits this probabilistic setting is known as *probabilistic sensitivity analysis*. The most immediate question in this case is to characterize the distribution of Y that is induced by giving $\mathbf{X}$ the distribution G. Although this may be regarded as an aspect of probabilistic sensitivity analysis, it is known by the separate name of *uncertainty analysis*. Sensitivity analysis proper is generally seen as going beyond uncertainty analysis by exploring how individual inputs or groups of inputs contribute to uncertainty in Y. In particular, an important problem is to identify which elements in $\mathbf{X}$ are the most influential, in some sense, in inducing the uncertainty in Y.

We shall be particularly interested in the case where the model is so complex that simply computing the output y for any given set of input values is a non-trivial task. For instance, large process models in engineering, environmental science, chemistry, etc. are often implemented in complex computer codes that require many minutes, hours or even days for a single run. We call such models *expensive*, whereas a model that can be run many thousands of times in a reasonable time is called *cheap*. The distinction is important when it comes to calculating any desired measures of sensitivity, since 'brute force' computation that may be the simplest solution for cheap models is usually impractical for expensive models.

Sensitivity and uncertainty analysis are important techniques for exploring complex models. Saltelli, Tarantola and Campolongo (2000) and Kleijnen (1997) clearly show the key role of these tools within the wider context of the building, validation and use of process models. We present here

(a) a discussion of sensitivity analysis that unifies various other approaches that are considered in the literature and
(b) a Bayesian method building on the approach of O'Hagan *et al.* (1999) that is both robust and highly efficient, allowing sensitivity analysis to be applied to expensive models.

2. Probabilistic sensitivity analysis

We write $\mathbf{x} = \{x_1, \ldots, x_d\}$, and we refer to x_i as the ith element of $\mathbf{x}$ or the ith uncertain model input. We shall denote the subvector (x_i, x_j) by $\mathbf{x}_{i,j}$, and in general if p is a set of indices then $\mathbf{x}_p$ is the subvector of $\mathbf{x}$ whose elements have those indices. Finally, $\mathbf{x}_{-i}$ is the subvector of $\mathbf{x}$ containing all elements except x_i.

Note, however, that much of the analysis that is discussed here holds if the x_is are not simple scalars, so the notation $\mathbf{x} = \{x_1, \ldots, x_d\}$ could denote only a partial decomposition of $\mathbf{x}$ into d subvectors.

2.1. Main effects and interactions

Some widely used methods of sensitivity analysis can be seen in terms of a decomposition of the function $\eta(\cdot)$ into main effects and interactions:

$$y = \eta(\mathbf{x}) = E(Y) + \sum_{i=1}^{d} z_i(x_i) + \sum_{i<j} z_{i,j}(\mathbf{x}_{i,j}) + \sum_{i<j<k} z_{i,j,k}(\mathbf{x}_{i,j,k}) + \ldots + z_{1,2,\ldots,d}(\mathbf{x}), \qquad (1)$$

where

$$z_i(x_i) = E(Y|x_i) - E(Y),$$

$$z_{i,j}(\mathbf{x}_{i,j}) = E(Y|\mathbf{x}_{i,j}) - z_i(x_i) - z_j(x_j) - E(Y),$$

$$z_{i,j,k}(\mathbf{x}_{i,j,k}) = E(Y|\mathbf{x}_{i,j,k}) - z_{i,j}(\mathbf{x}_{i,j}) - z_{i,k}(\mathbf{x}_{i,k}) - z_{j,k}(\mathbf{x}_{j,k}) - z_i(x_i) - z_j(x_j) - z_k(x_k) - E(Y),$$

and so on. We refer to $z_i(x_i)$ as the *main effect* of x_i, to $z_{i,j}(\mathbf{x}_{i,j})$ as the first-order *interaction* between x_i and x_j, and so on.

Note that the definitions of these terms depend on the distribution G of the uncertain inputs. Consider, for instance, the very simple model $\eta(x_1, x_2) = x_1$. We have $E(Y) = E(X_1)$ and $z_1(x_1) = x_1 - E(X_1)$. If G is such that X_1 and X_2 are independent then $z_2(x_2) = 0$ and $z_{12}(x_1, x_2) = 0$. In this case, the representation reflects the structure of the model itself, comprising a linear effect of x_1 with no x_2-effect and no interaction. If, however, X_1 and X_2 are not independent, we have $z_2(x_2) = E(X_1|x_2) - E(X_1) = -z_{12}(x_1, x_2)$, which will not in general be 0.

Computing and plotting the main effects and first-order interactions is a powerful visual tool for examining how the model output responds to each individual input, and how those inputs interact in their influence on y.

2.2. Variance-based methods

Variance-based methods of probabilistic sensitivity analysis quantify the sensitivity of the output Y to the model inputs in terms of a reduction in the variance of Y.

This approach is reviewed by Saltelli, Chan and Scott (2000). Two principal measures of the sensitivity of Y to an individual x_i are proposed. The first is

$$V_i = \mathrm{var}\{E(Y|X_i)\}.$$

The motivation for this measure is that it is the expected amount by which the uncertainty in Y will be reduced if we learn the true value of x_i. Thus, if we were to learn x_i, then the uncertainty about Y would become $\mathrm{var}(Y|x_i)$, a difference of $\mathrm{var}(Y) - \mathrm{var}(Y|x_i)$. Since we do not know the true value of x_i, the expected difference is $\mathrm{var}(Y) - E\{\mathrm{var}(Y|x_i)\} = V_i$, by a well-known identity. Although $\mathrm{var}(Y) - \mathrm{var}(Y|x_i)$ can be negative for some x_i, its expectation V_i is always positive, so this is the expected reduction in uncertainty due to observing x_i. Note also that $V_i = \mathrm{var}\{z_i(X_i)\}$ and so is based on the main effect of x_i.

The second measure, first proposed by Homma and Saltelli (1996), is

$$V_{Ti} = \mathrm{var}(Y) - \mathrm{var}\{E(Y|\mathbf{X}_{-i})\},$$

which is the remaining uncertainty in Y that is unexplained after we have learnt everything except x_i.

Both measures are converted into scale invariant measures by dividing by $\mathrm{var}(Y)$:

$$S_i = V_i/\mathrm{var}(Y), \tag{2}$$

$$S_{Ti} = V_{Ti}/\mathrm{var}(Y) = 1 - S_{-i}. \tag{3}$$

Thus, S_i may be referred to as the main effect index of x_i, and S_{Ti} is known as the total effect index of x_i. The relative importance of each input in driving the uncertainty in Y is then gauged by comparing their indices.

As well as indicating the relative importance of an individual x_i in driving the uncertainty in Y, equation (2) can be seen as indicating where to direct effort in future to reduce that uncertainty. If it were possible to observe one of the x_is, to learn its true value exactly, and the cost of that observation would be the same for each i, then we should choose that with the largest S_i. In practice, of course, it is rarely possible to learn the true value of any of the uncertain inputs exactly; nor is the cost of gaining more information likely to be the same for each input. Nevertheless, the analysis does suggest where there is the greatest potential for reducing uncertainty through new research. We do not believe that there is any comparable interpretation of S_{Ti} in terms of guiding research effort.

It does not follow that the two inputs with the largest main effect variances will be the best two inputs to observe. We would need to calculate

$$V_{i,j} = \mathrm{var}\{E(Y|\mathbf{X}_{i,j})\} = \mathrm{var}\{z_i(X_i) + z_j(X_j) + z_{ij}(\mathbf{X}_{i,j})\} \tag{4}$$

for all i and j, since this is the part of $\mathrm{var}(Y)$ that is removed on average when we learn both x_i and x_j. The search for the most informative combinations of inputs is considered further by Saltelli and Tarantola (2002). In general, $V_p = \mathrm{var}\{E(Y|\mathbf{X}_p)\}$ is the expected reduction in variance that is achieved when we learn $\mathbf{x}_p$.

2.3. Variance decomposition

When G is such that the elements of $\mathbf{X}$ are mutually independent, we have already remarked that the definitions of main effects and interactions will directly reflect the model structure. In this case, we can also decompose the variance of Y into terms relating to the main effects and various interactions between the input variables. A decomposition like an analysis of variance is given in Cox (1982):

$$\mathrm{var}(Y) = \sum_{i=1}^{d} W_i + \sum_{i<j} W_{i,j} + \sum_{i<j<k} W_{i,j,k} + \ldots + W_{1,2,\ldots,d}, \tag{5}$$

where $W_p = \mathrm{var}\{z_p(\mathbf{X}_p)\}$. This result holds because when the X_is are independent it is straightforward to show that all the terms in equation (1) are uncorrelated. Equation (5) gives us a partition of the variance into terms that are the variances of the main effects and interaction terms in equation (1).

We have $W_i = V_i$, i.e. the variance of the main effect is the reduction in $\mathrm{var}(Y)$ that is obtained by learning the true value of x_i. $W_{i,j}$ is the component of $\mathrm{var}(Y)$ due solely to uncertainty about the interaction between inputs x_i and x_j. Note that equation (4) becomes $V_{i,j} = W_i + W_j + W_{i,j} = V_i + V_j + W_{i,j}$, so $W_{i,j}$ is an extra amount of variance removed when we learn both x_i and x_j, over the main effect variances V_i and V_j.

It is clear that when equation (5) holds we can identify $V_{-i} = \mathrm{var}\{E(Y|\mathbf{X}_{-i})\}$ with the sum of all the W_p-terms not including the subscript i. Therefore the total effect index (3) is the proportion of $\mathrm{var}(Y)$ that is accounted for by all the terms in equation (5) with a subscript i, and so $S_{Ti} \geqslant S_i$. It is also clear that $\Sigma_{i=1}^{d} S_i \leqslant 1 \leqslant \Sigma_{i=1}^{d} S_{Ti}$, with equalities only when all interactions are 0.

Independence between the input variables, therefore, allows a tidy decomposition of the total variance into component variances that are directly related to the quantities that were discussed in Section 2.2. An analogy is that in regression analysis we have a nice partition of the total sum of squares when the regressors, or groups of regressors, are orthogonal. Without orthogonality, we can still define the sum of squares attributable to any set of variables, but sums of squares for different sets of regressors no longer partition the total sum of squares.

2.4. Regression components

The analogy with regression analysis becomes clearer if we consider the variance of Y as an expected squared error of prediction. Thus, if we wish to predict Y without gaining any further information about $\mathbf{x}$, then the best prediction (in terms of minimizing the expected squared error) is $E(Y)$. Then $\mathrm{var}(Y)$ is the expected squared error of this prediction. Similarly, if we learn the true value of the subvector $\mathbf{x}_p$, then the best predictor of Y becomes $E(Y|\mathbf{x}_p)$ and results in an expected squared error of $E\{\mathrm{var}(Y|\mathbf{X}_p)\}$.

Consider predicting $Y=\eta(\mathbf{x})$ by a linear model of the form

$$\hat{\eta}(\mathbf{x})=\alpha+\mathbf{g}(\mathbf{x})^{\mathrm{T}}\gamma. \tag{6}$$

The components of the vector function $\mathbf{g}(\mathbf{x})$ are supposed given. We wish to choose α and γ to obtain an approximation of the form (6) to minimize the expected squared prediction error $E\{Y-\hat{\eta}(\mathbf{X})\}^2$. We then find that the optimal approximation is given by

$$\gamma=\mathrm{var}\{\mathbf{g}(\mathbf{X})\}^{-1}\,\mathrm{cov}\{\mathbf{g}(\mathbf{X}),Y\}$$

and $\alpha=E(Y)-E\{\mathbf{g}(\mathbf{X})\}^{\mathrm{T}}\gamma$. The expected squared error is $\mathrm{var}(Y)-V_{\mathbf{g}(\mathbf{x})}$, where

$$V_{\mathbf{g}(\mathbf{x})}=\mathrm{cov}\{\mathbf{g}(\mathbf{X}),Y\}^{\mathrm{T}}\,\mathrm{var}\{\mathbf{g}(\mathbf{X})\}^{-1}\,\mathrm{cov}\{\mathbf{g}(\mathbf{X}),Y\}. \tag{7}$$

The inclusion of the constant term α leads to $E\{Y-\hat{\eta}(\mathbf{X})\}=0$. Furthermore, it then holds that $\hat{\eta}(\mathbf{X})$ is uncorrelated with $\eta(\mathbf{X})-\hat{\eta}(\mathbf{X})$, and so we have the variance decomposition

$$\mathrm{var}(Y)=V_{\mathbf{g}(\mathbf{x})}+\mathrm{var}\{\eta(\mathbf{X})-\hat{\eta}(\mathbf{X})\}. \tag{8}$$

The interpretation of equation (8) is that $V_{\mathbf{g}(\mathbf{x})}$ is the component of $\mathrm{var}(Y)$ that is explained by this fitted approximation, and that the second term measures its lack of fit. Setting $\mathbf{g}(\mathbf{x})=x_i$ gives a variance component

$$V_{x_i}=\mathrm{cov}(X_i,Y)^2/\mathrm{var}(X_i)$$

for a best prediction of Y by a linear function of x_i alone. Now, since $\mathrm{cov}(X_i,Y)=\mathrm{cov}\{X_i,E(Y|X_i)\}$, this is also the best linear predictor of $E(Y|x_i)$, and hence of the main effect of x_i. Therefore the difference $V_i-V_{x_i}$ measures the lack of linearity of this main effect.

Remembering that one objective of sensitivity analysis is to understand the way that the output responds to changes to the inputs, these linear variance components and their complementary lack-of-fit components give further insight into the behaviour of the model. By introducing quadratic and higher order polynomial terms in $\mathbf{g}(\mathbf{x})$ we can refine this understanding further. We could equally well look at other regressor variables if the nature of the phenomenon that is being modelled suggested them, such as harmonic terms for a cyclic input.

It should be noted that regression coefficients, correlation coefficients and related sums of squares have been widely used in sensitivity analysis. Various sensitivity diagnostics based on a Monte Carlo sample of runs are presented by Kleijnen and Helton (1999) and the use of regression coefficients in particular is discussed in Helton and Davis (2000). However, our approach is different in some important respects.

Their approach is based on a Monte Carlo sample $\{(\mathbf{x}_s,y_s),s=1,2,\ldots,N\}$ that is obtained by sampling the input vectors $\mathbf{x}_s$ from the distribution G and then running the model at each sampled $\mathbf{x}_s$ to compute output $y_s=\eta(\mathbf{x}_s)$. They regarded the regression or correlation coefficient between y and each input variable x_i as measures of sensitivity of the output to that x_i. Implicitly, they are fitting the statistical model $y_s=\alpha+\Sigma_{i=1}^{d}\beta_i x_{is}+\varepsilon_s$, with random-error term ε_s (although Kleijnen and Helton (1999) also considered rank regression and other analyses of the Monte Carlo data).

In contrast, we define the regression coefficients γ in terms of a best-fitting regression approximation (6) to $Y = \eta(\mathbf{x})$, judged by the expected squared error. In practice, it is easy to see that the regression coefficients of Helton and Davis (2000) are estimates of our optimal coefficients γ in the corresponding regression fit. However, the interpretation is different, we allow for non-linear fits and we add the very important step of interpreting the difference between the regression variance component and the corresponding main effect variance as a lack-of-fit variance component.

Probably the most important difference is that we define the regression fits and corresponding sums of squares as properties of the model $\eta(\cdot)$, without involving any particular sample of model runs. We see the standard regression approach in sensitivity analysis as a way to compute estimates of these measures, but our definition opens the way for non-sampling-based ways to estimate them, as in Section 3.

2.5. Discussion

The preceding subsections have presented a very broad perspective on probabilistic sensitivity analysis. There are many reasons for conducting a sensitivity analysis of a model, as set out for instance in French (2003). The techniques of sensitivity analysis are correspondingly diverse, but it is very useful to be able to see a range of techniques in a common framework. Our formulation unifies a variety of current approaches and offers new measures, to provide a deeper understanding of a model and its dependence on the uncertain model inputs.

The idea of estimating and plotting main effects and interactions has been used before by Welch *et al.* (1992), the variance-based sensitivity measures of main effect and total effect indices are widely used, as described in Saltelli, Chan and Scott (2000), and sample-based regression measures of sensitivity have been widely used by Helton and Davis (2000) and others, but these are usually seen as distinct and unrelated approaches to sensitivity analysis. We define new population-based regression measures that provide a link between the sample measures and variance-based sensitivity analysis. Our proposal to use the difference between V_i and V_{x_i} to measure non-linearity in $z_i(x_i)$ is novel and, we believe, powerful. A similar idea, comparing V_{x_i} with the same measure based on ranked data, was proposed by Saltelli and Sobol' (1995).

We contend that a fuller understanding of how individual inputs (and groups of inputs) influence the model output can be obtained through studying all the various measures—main effects and interactions, variance-based sensitivity indices and regression components—rather than relying on just one of these approaches.

We end this section by briefly addressing some other issues.

2.5.1. Local sensitivity

Local sensitivity analysis is based on partial derivatives of the function $\eta(\cdot)$, evaluated at the base-line inputs $\mathbf{x}_0$. Defining $\eta^i(\mathbf{x}) = \partial\eta(\mathbf{x})/\partial x_i$, a measure of local sensitivity of the output to input i that is invariant to scale changes in both inputs and output is

$$D_i = \eta^i(\mathbf{x}_0)\sqrt{\{\mathrm{var}(X_i)/\mathrm{var}(Y)\}}.$$

It is then straightforward to show that D_i^2 is the proportional reduction in $\mathrm{var}(Y)$ if we predict Y by using a linear predictor with slope $\eta^i(\mathbf{x}_0)$. So in general $D_i^2\,\mathrm{var}(Y) \leqslant V_{x_i}$. Baker (2001) suggested approximating $\eta(\cdot)$ by a first-order Taylor series and derived the D_i^2 as measures of sensitivity.

2.5.2. Value of information

Our use of squared prediction error as a criterion can be justified formally in decision theoretic terms by using the squared error loss. It may then be shown that V_p is the expected value of gaining perfect information about $\mathbf{x}_p$. More generally, wherever the computer model is to be used for decision-making, we could again measure sensitivity by the expected value of information, but now defined with respect to the relevant utility or loss function and the available decisions. An example of this approach is Oakley (2002a), where the model was used as the basis of a decision on relative cost-effectiveness of competing medical technologies.

2.5.3. Computation

In principle, if $\eta(\cdot)$ is sufficiently tractable it would be possible to derive *analytically* any of the sensitivity measures that were discussed in the preceding sections. For models of sufficient complexity for it not to be obvious how the output would respond to the model inputs, we cannot hope for such tractability and must instead seek to obtain the desired measures computationally.

If $\eta(\cdot)$ is sufficiently cheap to evaluate for many different inputs, simple Monte Carlo methods can be used to estimate $\mathrm{var}(Y)$ or the component of variance $V_{\mathbf{g}(\mathbf{x})}$ for any regression fit with negligible error. It is not so straightforward to obtain V_i (and hence S_i), S_{Ti} or the main effects or interactions as suggested in Section 2.1, since they depend on evaluating conditional expectations. The method of Sobol' (1993) and the Fourier amplitude sensitivity test, devised by Cukier *et al.* (1973) and extended by Saltelli *et al.* (1999), are techniques that have been developed specifically to compute some of these measures. Nevertheless, sensitivity analysis by these techniques demands many thousands of function evaluations. For an expensive function, where the evaluation of $\eta(\mathbf{x})$ at a single $\mathbf{x}$ might take minutes or even hours, such methods are impractical.

3. Bayesian sensitivity analysis

In this section, we shall develop Bayesian inference tools for estimating all the quantities of interest in sensitivity analysis, for the case of expensive functions. In addition to making it feasible to carry out sensitivity analysis with a much smaller number of model runs, a key benefit of our approach is that it can estimate all the many sensitivity measures that were discussed in Section 2, from a single set of runs.

The essence of the Bayesian approach is that the model $\eta(\cdot)$ is treated as an unknown function. In an absolute sense, of course, $\eta(\cdot)$ is certainly not unknown, since it implements a model that has been specified in precise mathematical form by someone (or some group of people). Nevertheless, in a pragmatic sense $\eta(\mathbf{x})$ is unknown for any particular input configuration $\mathbf{x}$ until we actually run the model for those inputs.

We therefore formulate a prior distribution for the function $\eta(\cdot)$. This is updated according to the usual Bayesian paradigm, using as data the outputs $y_i = \eta(\mathbf{x}_i)$, $i = 1, 2, \ldots, N$, from a set of runs of the model. The result is a posterior distribution for $\eta(\cdot)$, which we then use to make formal Bayesian inferences about the various sensitivity measures that were introduced in Section 2. Similar methods to estimate some of these sensitivity measures have been described in Welch *et al.* (1992) and applied by Mrawira *et al.* (1999) to a model for highway management.

3.1. Inference about functions using Gaussian processes

We first develop the prior model for $\eta(\cdot)$ in the form of a Gaussian process prior distribution and derive the posterior distribution.

Gaussian processes have been used before for modelling computer codes; examples are Currin *et al.* (1991) and Haylock and O'Hagan (1996). The key requirement is that $\eta(\cdot)$ is believed to be a smooth function, so if we know the value of $\eta(\mathbf{x})$ we should have some idea about the value of $\eta(\mathbf{x}')$ for $\mathbf{x}$ close to $\mathbf{x}'$. It is this property of $\eta(\cdot)$ that will give us the opportunity to improve on Monte Carlo sampling, since the extra information that is available after each code run is ignored in the Monte Carlo approach.

For any set of points $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, we represent our uncertainty about $\{\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)\}$ through a multivariate normal distribution. The mean of $\eta(\mathbf{x})$ is given by

$$E\{\eta(\mathbf{x})|\beta\} = \mathbf{h}(\mathbf{x})^{\mathrm{T}}\beta, \tag{9}$$

conditional on β. The vector $\mathbf{h}(\cdot)$ consists of q known regression functions of $\mathbf{x}$, and β is a vector of coefficients. The choice of $\mathbf{h}(\cdot)$ is arbitrary, though it should be chosen to incorporate any beliefs that we might have about the form of $\eta(\cdot)$. It is quite distinct from the $\mathbf{g}(\cdot)$ that was introduced in Section 2.4. The covariance between $\eta(\mathbf{x})$ and $\eta(\mathbf{x}')$ is given by

$$\mathrm{cov}\{\eta(\mathbf{x}), \eta(\mathbf{x}')|\sigma^2\} = \sigma^2\, c(\mathbf{x}, \mathbf{x}'), \tag{10}$$

conditional on σ^2, where $c(\mathbf{x}, \mathbf{x}')$ is a function which decreases as $|\mathbf{x} - \mathbf{x}'|$ increases and also satisfies $c(\mathbf{x}, \mathbf{x}) = 1$ for all $\mathbf{x}$. The function $c(\cdot, \cdot)$ must ensure that the covariance matrix of any set of outputs $\{y_1 = \eta(\mathbf{x}_1), \ldots, y_n = \eta(\mathbf{x}_n)\}$ is positive semidefinite.

We regard β and σ^2 as unknown hyperparameters, and we discuss their prior distribution below. The other components of equations (9) and (10) are the vector $\mathbf{h}(\cdot)$ of regressor functions to model beliefs about the general way that the output will respond to the inputs, and a correlation function $c(\cdot, \cdot)$ to model beliefs about the smoothness of the model output. Considerable flexibility is available in the choice of these functions. In particular, $c(\cdot, \cdot)$ is typically represented hierarchically in terms of further unknown hyperparameters. An extensive discussion of these modelling issues is given in Kennedy and O'Hagan (2001).

For mathematical tractability, the conjugate prior form for β and σ^2, the normal inverse gamma distribution, is assumed:

$$p(\beta, \sigma^2) \propto \sigma^{-(d+q+2)/2} \exp[-\{(\beta - \mathbf{z})^{\mathrm{T}} V^{-1} (\beta - \mathbf{z}) + a\}/2\sigma^2]. \tag{11}$$

In the examples that are given in this paper, we use the weak form of this prior, $p(\beta, \sigma^2) \propto \sigma^2$. This implies an infinite prior variance of $\eta(\mathbf{x})$, whereas in practice we expect there to be cases when the model developer can provide some proper prior knowledge about the function $\eta(\cdot)$. Elicitation of such prior information is described in Oakley (2002b).

The output of $\eta(\cdot)$ is observed at n design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to obtain data $\mathbf{y}$. In contrast with Monte Carlo methods, these points are not chosen randomly but are selected to give good information about $\eta(\cdot)$. In practice, the design points will be spread to cover $\mathcal{X}$, the sample space of $\mathbf{X}$, although their choice will also depend on G. The choice of design points is discussed in Sacks *et al.* (1989). Given these data it can be shown that

$$\frac{\eta(\mathbf{x}) - m^*(\mathbf{x})}{\hat{\sigma}\sqrt{c^*(\mathbf{x}, \mathbf{x})}} | \mathbf{y} \sim t_{d+n}, \tag{12}$$

where

$$m^*(\mathbf{x}) = \mathbf{h}(\mathbf{x})^{\mathrm{T}}\hat{\beta} + \mathbf{t}(\mathbf{x})^{\mathrm{T}} A^{-1} (\mathbf{y} - H\hat{\beta}), \tag{13}$$

$$\begin{aligned} c^*(\mathbf{x}, \mathbf{x}') = {} & c(\mathbf{x}, \mathbf{x}') - \mathbf{t}(\mathbf{x})^{\mathrm{T}} A^{-1} \mathbf{t}(\mathbf{x}') \\ & + (\mathbf{h}(\mathbf{x})^{\mathrm{T}} - \mathbf{t}(\mathbf{x})^{\mathrm{T}} A^{-1} H)(H^{\mathrm{T}} A^{-1} H)^{-1} (\mathbf{h}(\mathbf{x}')^{\mathrm{T}} - \mathbf{t}(\mathbf{x}')^{\mathrm{T}} A^{-1} H)^{\mathrm{T}}. \end{aligned} \tag{14}$$

$$\mathbf{t}(\mathbf{x})^{\mathrm{T}} = (c(\mathbf{x}, \mathbf{x}_1), \ldots, c(\mathbf{x}, \mathbf{x}_n)),$$

$$H^{\mathrm{T}} = (\mathbf{h}(\mathbf{x}_1)^{\mathrm{T}}, \ldots, \mathbf{h}(\mathbf{x}_n)^{\mathrm{T}}),$$

$$A = \begin{pmatrix} 1 & c(\mathbf{x}_1, \mathbf{x}_2) & \ldots & c(\mathbf{x}_1, \mathbf{x}_n) \\ c(\mathbf{x}_2, \mathbf{x}_1) & 1 & & \vdots \\ \vdots & & \ddots & \\ c(\mathbf{x}_n, \mathbf{x}_1) & \ldots & & 1 \end{pmatrix},$$

$$\hat{\beta} = V^*(V^{-1}\mathbf{z} + H^{\mathrm{T}}A^{-1}\mathbf{y}),$$

$$\hat{\sigma}^2 = \{a + \mathbf{z}^{\mathrm{T}}V^{-1}\mathbf{z} + \mathbf{y}^{\mathrm{T}}A^{-1}\mathbf{y} - \hat{\beta}^{\mathrm{T}}(V^*)^{-1}\hat{\beta}\}/(n + d - 2),$$

$$V^* = (V^{-1} + H^{\mathrm{T}}A^{-1}H)^{-1},$$

$$\mathbf{y}^{\mathrm{T}} = (\eta(\mathbf{x}_1), \ldots, \eta(\mathbf{x}_n)).$$

The outputs corresponding to any set of inputs will now have a multivariate t-distribution, with covariance between any two outputs given by equation (14). Full details of the prior to posterior analysis can be found in O'Hagan (1994).

Note that the t-distribution arises as a marginal distribution for $\eta(\mathbf{x})$ after integrating out the hyperparameters β and σ^2. In practice, as mentioned earlier, typically further hyperparameters will be associated with the modelling of $c(\cdot, \cdot)$, and it is generally impossible to integrate the posterior analytically with respect to these further parameters. Although it is possible to integrate numerically or to use Markov chain Monte Carlo (MCMC) sampling (Bayarri *et al.*, 2002; Neal, 1999) this is a highly intensive computation, and experience suggests that it is adequate simply to estimate the hyperparameters of $c(\cdot, \cdot)$ from the posterior distribution and then to substitute these estimates into $c(\cdot, \cdot)$ wherever it appears in the above formulae; see Kennedy and O'Hagan (2001).

The following sections set out how inferences about the various sensitivity measures can be estimated from this posterior distribution. It may be helpful first, however, to relate this approach to that outlined for cheap functions in Section 2.5. Any method for computing a given measure can be viewed as estimating it. Like Monte Carlo methods, our Bayesian approach estimates such measures by formal statistical methods, and our estimates can be accompanied by standard errors or standard deviations to indicate their accuracy. Monte Carlo methods applied to very cheap functions typically employ many thousands of model runs, so that the estimation error is very small. When we consider expensive functions, it will rarely be feasible to do enough runs for the estimation error to be negligible. A key benefit of our methods, however, is that the standard deviations that are associated with estimates are generally very much smaller, often by orders of magnitude, than those which are obtained from a Monte Carlo method with the same number of model runs. It is this that allows us to achieve useful sensitivity analyses of complex expensive models without having to make prohibitively many runs.

3.2. Inference for main effects and interactions

First consider inference about

$$E(Y|\mathbf{x}_p) = \int_{\mathcal{X}_{-p}} \eta(\mathbf{x}) \mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p),$$

using obvious notation for the space of possible values for $\mathbf{x}_{-p}$ and for its conditional distribution given $\mathbf{x}_p$. Since this is a linear functional of the Gaussian process $\eta(\cdot)$, its posterior distribution will be t_{d+n}, after standardizing as in distribution (12) by subtracting its posterior mean and dividing by its posterior standard deviation; see O'Hagan (1991). We can derive the posterior mean as follows. Note that we denote expectations, variances and covariances defined with respect to the posterior distribution of $\eta(\cdot)$ by E^*, var* and cov* respectively.

$$E^*\{E(Y|\mathbf{x}_p)\} = R_p(\mathbf{x}_p)\hat{\beta} + T_p(\mathbf{x}_p)\mathbf{e}, \tag{15}$$

where

$$R_p(\mathbf{x}_p) = \int_{\mathcal{X}_{-p}} \mathbf{h}(\mathbf{x})^{\mathrm{T}} \mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p), \tag{16}$$

$$T_p(\mathbf{x}_p) = \int_{\mathcal{X}_{-p}} \mathbf{t}(\mathbf{x})^{\mathrm{T}} \mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p), \tag{17}$$

$$\mathbf{e} = A^{-1}(\mathbf{y} - H\hat{\beta}).$$

Similarly, $E^*\{E(Y)\} = R\hat{\beta} + T\mathbf{e}$, where R and T are the special cases of equations (16) and (17) when p is the null set.

We can now readily find the posterior mean of any main effect or interaction. For instance,

$$E^*\{z_i(x_i)\} = \{R_i(x_i) - R\}\hat{\beta} + \{T_i(x_i) - T\}\mathbf{e},$$

$$E^*\{z_{i,j}(\mathbf{x}_{i,j})\} = \{R_{i,j}(\mathbf{x}_{i,j}) - R_i(x_i) - R_j(x_j) - R\}\hat{\beta} + \{T_{i,j}(\mathbf{x}_{i,j}) - T_i(x_i) - T_j(x_j) - T\}\mathbf{e}.$$

All the main effects and interactions are linear functionals of $\eta(\cdot)$, and so their posterior distributions are t_{d+n} after appropriate standardization. We have derived their means and now require to obtain standard deviations. All these can be found from the following general result:

$$\begin{aligned}\mathrm{cov}^*\{E(Y|\mathbf{x}_p), E(Y|\mathbf{x}'_q)\} &= \hat{\sigma}^2 \int_{\mathcal{X}_{-p}} \int_{\mathcal{X}_{-q}} c^*(\mathbf{x}, \mathbf{x}')\, \mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\, \mathrm{d}G_{-q|q}(\mathbf{x}'_{-q}|\mathbf{x}'_q) \\ &= \hat{\sigma}^2[U_{p;q}(\mathbf{x}_p, \mathbf{x}'_q) - T_p(\mathbf{x}_p)A^{-1}T_q(\mathbf{x}'_q)^{\mathrm{T}} \\ &\quad + \{R_p(\mathbf{x}_p) - T_p(\mathbf{x}_p)A^{-1}H\}W\{R_q(\mathbf{x}'_q) - T_q(\mathbf{x}'_q)A^{-1}H\}^{\mathrm{T}}],\end{aligned} \tag{18}$$

where

$$U_{p;q}(\mathbf{x}_p, \mathbf{x}'_q) = \int_{\mathcal{X}_{-p}} \int_{\mathcal{X}_{-q}} c(\mathbf{x}, \mathbf{x}')\, \mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\, \mathrm{d}G_{-q|q}(\mathbf{x}'_{-q}|\mathbf{x}'_q), \tag{19}$$

$$W = (H^{\mathrm{T}}A^{-1}H)^{-1}.$$

The fact that all these results are expressed in terms of integrals, such as equation (16) or (19), is not a problem in practice. For some common formulations of $\mathbf{h}(\cdot)$, $c(\cdot, \cdot)$ and G, as used in the examples of Section 4, it is possible to evaluate these integrals analytically. Even if this is not possible, they may readily be computed numerically, since the integrands are very cheap functions. The same will hold for all the inferences in the following subsections.

For a single input x_i, we can plot the posterior mean of the main effect $E^*\{z_i(x_i)\}$ against x_i, with bounds of, for instance, plus and minus two posterior standard deviations. If we standardize each input variable, we can give $E^*\{z_i(x_i)\}$ for $i = 1, \ldots, d$ on a single plot, and this will provide a good graphical summary of the influence of each variable.

However, unless n is large this combined plot may give a false impression of the importance of each input, relative to that obtained from considering the main effect variance terms V_i. From the plot, it is tempting to think of the inputs showing the greatest variation as the most important, but $\mathrm{var}[E^*\{z_i(X_i)\}]$ is not the same as the posterior mean of V_i, i.e. $E^*[\mathrm{var}\{z_i(X_i)\}]$. This is why it is important to consider the posterior variance of $z_i(x_i)$ as well as its posterior mean. From a graph showing an individual $E^*\{z_i(x_i)\}$ with standard deviation bounds, we could visually judge (admittedly rather crudely) both $\mathrm{var}[E^*\{z_i(X_i)\}]$ and $E[\mathrm{var}^*\{z_i(X_i)\}]$. Since

$$E^*[\mathrm{var}\{z_i(X_i)\}] = \mathrm{var}[E^*\{z_i(X_i)\}] + E[\mathrm{var}^*\{z_i(X_i)\}] - \mathrm{var}^*[E\{z_i(X_i)\}],$$

and, since $E\{z_i(X_i)\} = 0$ for all i, we have

$$E^*(V_i) = \mathrm{var}[E^*\{z_i(X_i)\}] + E[\mathrm{var}^*\{z_i(X_i)\}].$$

If the design set is sufficiently large and well chosen to make $\mathrm{var}^*\{z_i(x_i)\}$ small for almost all x_i (with respect to G) then the second term can be ignored, but this will not usually be so in practice.

3.3. Inference for variances

We now consider posterior inference for V_i and V_{Ti}. Note that these are quadratic functionals of $\eta(\cdot)$, and their posterior distributions will no longer be obtainable analytically. We can, however, derive posterior means and variances. For instance, Haylock and O'Hagan (1996) derived the posterior mean and variance of $\mathrm{var}(Y)$. We generalize their approach to derive the posterior mean of $V_p = \mathrm{var}\{E(Y|\mathbf{X}_p)\}$ for any subvector $\mathbf{x}_p$. First note that

$$\begin{aligned}\mathrm{var}\{E(Y|\mathbf{X}_p)\} &= E\{E(Y|\mathbf{X}_p)^2\} - E\{E(Y|\mathbf{X}_p)\}^2 \\ &= E\{E(Y|\mathbf{X}_p)^2\} - E(Y)^2.\end{aligned}$$

Since $E^*\{E(Y)^2\}$ can be derived from results that we already have for $\mathrm{var}^*\{E(Y)\}$ and $E^*\{E(Y)\}$, we just need $E^*[E\{E(Y|\mathbf{X}_p)^2\}]$. (Note that the expression for $\mathrm{var}^*\{E(Y)\}$ in Haylock and O'Hagan (1996) was incorrect, and we have given the correct result here via equation (18).)

$$\begin{aligned}&E^*[E\{E(Y|\mathbf{X}_p)^2\}] \\ &\quad= \int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} E^*\{\eta(\mathbf{x})\,\eta(\mathbf{x}^*)\}\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p) \\ &\quad= \int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} \{\hat{\sigma}^2\,c^*(\mathbf{x},\mathbf{x}^*) + m^*(\mathbf{x})\,m^*(\mathbf{x}^*)\}\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p),\end{aligned}$$

where $G_p(\cdot)$ denotes the marginal distribution of $\mathbf{X}_p$ and $\mathbf{x}^*$ denotes the vector with elements made up of $\mathbf{x}_p$ and $\mathbf{x}'_{-p}$ in the same way as $\mathbf{x}$ is composed of $\mathbf{x}_p$ and $\mathbf{x}_{-p}$. We have

$$\begin{aligned}&\int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} \hat{\sigma}^2\,c^*(\mathbf{x},\mathbf{x}^*)\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p) \\ &\qquad= \hat{\sigma}^2[U_p - \mathrm{tr}(A^{-1}P_p) + \mathrm{tr}\{W(Q_p - S_pA^{-1}H - H^{\mathrm{T}}A^{-1}S_p^{\mathrm{T}} + H^{\mathrm{T}}A^{-1}P_pA^{-1}H)\}]\end{aligned}$$

and

$$\begin{aligned}&\int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} m^*(\mathbf{x})\,m^*(\mathbf{x}^*)\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-1}(\mathbf{x}'_{-1}) \\ &\qquad= \mathrm{tr}(\mathbf{e}^{\mathrm{T}}P_p\mathbf{e}) + 2\,\mathrm{tr}(\hat{\beta}S_p\mathbf{e}) + \mathrm{tr}(\hat{\beta}Q_p\hat{\beta}),\end{aligned}$$

where

$$U_p = \int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} c(\mathbf{x},\mathbf{x}^*)\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p),$$

$$P_p = \int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} t(\mathbf{x})t(\mathbf{x}^*)^{\mathrm{T}}\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p),$$

$$Q_p = \int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} h(\mathbf{x})h(\mathbf{x}^*)^{\mathrm{T}}\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p),$$

$$S_p = \int_{\mathcal{X}_p}\int_{\mathcal{X}_{-p}}\int_{\mathcal{X}_{-p}} h(\mathbf{x})t(\mathbf{x}^*)^{\mathrm{T}}\,\mathrm{d}G_{-p|p}(\mathbf{x}_{-p}|\mathbf{x}_p)\,\mathrm{d}G_{-p|p}(\mathbf{x}'_{-p}|\mathbf{x}_p)\,\mathrm{d}G_p(\mathbf{x}_p).$$

The formula for $\mathrm{var}^*[\mathrm{var}\{E(Y|\mathbf{X}_p)\}]$ is complex and will not be presented here. As before, all the required integrals can be done numerically if necessary but are available analytically for certain common modelling choices.

From this derivation of $E^*(V_p)$ the posterior means of the main effect variance V_i and the complementary effect variance V_{Ti} follow immediately. For inference about the indices S_i and S_{Ti}, it is natural to divide these estimates by $E^*\{\mathrm{var}(Y)\}$. This does not of course give $E^*(S_i)$ or $E^*(S_{Ti})$, but the posterior expectations of ratios cannot be derived analytically.

In the case of independent X_is, the variance decomposition (5) is of interest. To compute posterior means of interaction variances W_p, it is straightforward to show that (for independent inputs)

$$\mathrm{cov}\{E(Y|\mathbf{X}_p), E(Y|\mathbf{X}_q)\} = \mathrm{var}\{E(Y|\mathbf{X}_{p\cap q})\}.$$

3.4. Inference for regression fits

To derive inferences for the regression fits and corresponding variance components in Section 2.4, it is helpful first to note that

$$\mathrm{cov}\{\mathbf{g}(\mathbf{X}), Y\} = E\{\mathbf{g}^*(\mathbf{X})Y\} = \int_{\mathcal{X}} \mathbf{g}^*(\mathbf{x})\,\eta(\mathbf{x})\,\mathrm{d}G(\mathbf{x}),$$

where $\mathbf{g}^*(\mathbf{x}) = \mathbf{g}(\mathbf{x}) - E\{\mathbf{g}(\mathbf{X})\}$. The posterior mean of the regression fit may now be obtained from

$$E^*(\gamma) = \mathrm{var}\{\mathbf{g}(\mathbf{X})\}^{-1}\int_{\mathcal{X}} \mathbf{g}^*(\mathbf{x})\,m^*(\mathbf{x})\,\mathrm{d}G(\mathbf{x}),$$

then expanding $m^*(\mathbf{x})$ from equation (13) and integrating term by term, as in equation (15).

Similarly, the posterior mean of the regression variance component $V_{\mathbf{g}(\mathbf{x})}$ is obtained from

$$E^*\{V_{\mathbf{g}(\mathbf{x})}\} = \mathrm{tr}[\mathrm{var}\{\mathbf{g}(\mathbf{X})\}^{-1}]\int_{\mathcal{X}}\int_{\mathcal{X}} \mathbf{g}^*(\mathbf{x})\,\mathbf{g}^*(\mathbf{x}')^{\mathrm{T}}\{c^*(\mathbf{x},\mathbf{x}') + m^*(\mathbf{x})\,m^*(\mathbf{x}')\}\,\mathrm{d}G(\mathbf{x})\,\mathrm{d}G(\mathbf{x}'),$$

then expanding $c^*(\mathbf{x},\mathbf{x}')$ and $m^*(\mathbf{x})\,m^*(\mathbf{x}')$ from equations (14) and (13), and integrating term by term. All the resulting integrals may be computed numerically and may be obtained analytically for common modelling choices.

Posterior variances may be obtained, and since γ is a linear functional of $\eta(\cdot)$ its posterior distribution is t_{d+n} after appropriate standardization.

We also note that it is possible to derive posterior inferences about any desired derivatives. Relevant theory is given for one dimension in O'Hagan (1992) and is easily generalized to higher dimensions. We find, for instance, that the posterior distribution of $\eta^i(\mathbf{x})$ has mean $\partial m^*(\mathbf{x})/\partial x_i$, and is t_{d+n} after appropriate standardization. Inference about D_i^2 can then also be derived.

4. Examples

We present two illustrative examples, which are typical of a variety of models that we have considered. To apply the techniques of Section 3 in practice, it is necessary to identify the functions $\mathbf{h}(\cdot)$ and $c(\cdot,\cdot)$ that represent prior beliefs about the function $\eta(\cdot)$, and the distribution $G(\cdot)$ that defines the uncertainty about the model inputs.

The general principles for specifying $\mathbf{h}(\cdot)$ and $c(\cdot,\cdot)$ are discussed in Kennedy and O'Hagan (2001). In both examples our choice for $c(\cdot,\cdot)$ is the Gaussian form

$$c(\mathbf{x},\mathbf{x}') = \exp\{-(\mathbf{x}-\mathbf{x}')^{\mathrm{T}} B(\mathbf{x}-\mathbf{x}')\}, \tag{20}$$

where B is a diagonal matrix of (positive) roughness parameters. This implies a belief that the output is an analytic differentiable function of its inputs.

The distribution $G(\cdot)$ in these examples is multivariate normal. Together with the choices of $\mathbf{h}(\cdot)$ and $c(\cdot,\cdot)$ above, a normal $G(\cdot)$ allows all the integrals in Section 3 to be done analytically, conditional on B.

4.1. Synthetic example

We illustrate our methodology first with a synthetic example. The following test function is used:

$$\eta(\mathbf{x}) = \mathbf{a}_1^{\mathrm{T}}\mathbf{x} + \mathbf{a}_2^{\mathrm{T}}\sin(\mathbf{x}) + \mathbf{a}_3^{\mathrm{T}}\cos(\mathbf{x}) + \mathbf{x}^{\mathrm{T}} M \mathbf{x}. \tag{21}$$

A 15-dimensional input vector $\mathbf{x}$ is considered. The elements of the unknown true input $\mathbf{X}$ all have independent $N(0,1)$ distributions. The weights $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are chosen so that one group of five input variables accounts for the majority of the variance of $Y = \eta(\mathbf{X})$, another group of five variables makes a relatively small contribution to the variance, and the remaining five have very little effect. The values of these and of the matrix M can be downloaded from `www.sheffield.ac.uk/st1jeo`.

In the prior mean function we set $h(\mathbf{x}) = 1$, to represent no prior knowledge about how the model output relates to its inputs. A more usual choice would be $\mathbf{h}(\mathbf{x})^{\mathrm{T}} = (1, \mathbf{x}^{\mathrm{T}})$, representing a belief that the output will be approximately linear in all the inputs. However, the function (21) includes a linear component, and by not including this in our prior information we shall pose a stronger test for our methods.

We evaluated $\eta(\mathbf{x})$ at 250 design points, chosen to make the expected posterior variance

$$\int_{\mathcal{X}} c^*(\mathbf{x},\mathbf{x})\,\mathrm{d}G(\mathbf{x}) \tag{22}$$

small, conditional on guessed values of the roughness parameters in B. This is done by using a suboptimal, but computationally simple, scheme. We start with a Latin hypercube sample of 250 points and evaluate integral (22). We then consider exchanging each design point in turn with a randomly drawn candidate design point. If the new design decreases the value of the integral, the candidate design point is exchanged for the current point. This process is repeated until reductions in integral (22) become small.

Given the 250 runs, we found the posterior mode of B, and conditional on this estimate $\mathrm{var}(Y)$ and V_i by their posterior expectations for each of the 15 input variables. We then estimated the

Table 1. True and estimated main effects and linear components: synthetic example

	$100S_i$	$100\hat{S}_i$	$100V_{x_i}/var(Y)$	$100\hat{V}_{x_i}/\widehat{var}(Y)$
X_1	0.1560	0.1738	0.1258	0.1369
X_2	0.0186	0.0944	0.0164	0.0355
X_3	0.1307	0.1331	0.1130	0.1136
X_4	0.3045	0.3025	0.0921	0.1731
X_5	0.2905	0.3382	0.0719	0.0559
X_6	2.3035	2.2980	1.7123	1.8968
X_7	2.4151	1.9681	1.6283	1.3838
X_8	2.6517	2.8173	2.3651	2.5833
X_9	4.6036	4.4774	2.2383	2.1671
X_{10}	1.4945	1.3914	1.3189	1.3283
X_{11}	10.1823	9.4742	7.9852	7.7383
X_{12}	13.5708	13.2545	11.6408	11.8991
X_{13}	10.1989	10.7672	8.3444	8.5742
X_{14}	10.5169	10.2401	9.1020	8.4137
X_{15}	12.2818	12.8899	10.9820	11.9776

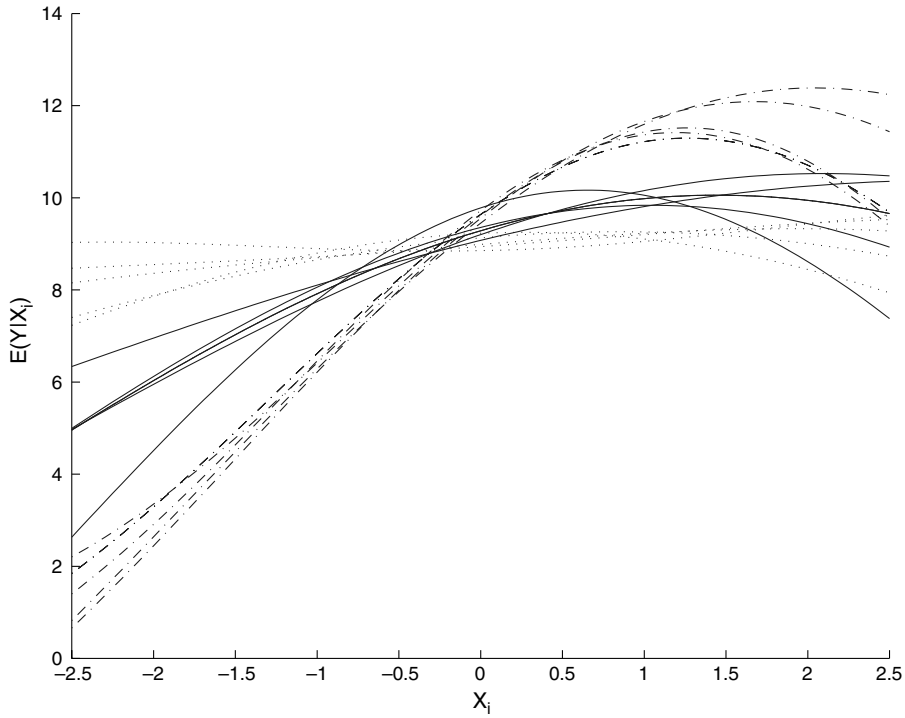

Fig. 1. Posterior expectation of $E(Y|x_i)$ against x_i for each input variable: synthetic example (·····, X_1, X_2, X_3, X_4, X_5; ———, $X_6, X_7, X_8, X_9, X_{10}$; · – · – ·, $X_{11}, X_{12}, X_{13}, X_{14}, X_{15}$)

main effect indices S_i by the ratio $\hat{S}_i$ of these posterior means. The true values of these terms can also be determined analytically. We also estimated the contributions to the variance from the best linear fit, defined as V_{x_i}. These results are summarized in Table 1.

From Table 1 we can see that, although there is some error in the estimates of the main effects and linear components, we have successfully identified the three distinct groups of variables,

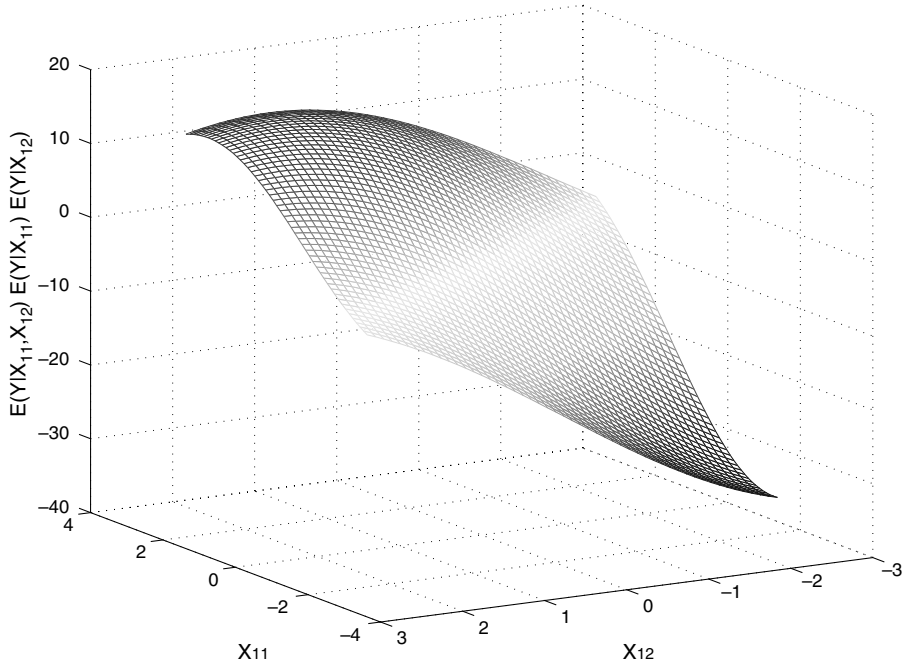

Fig. 2. Posterior expectation of $E(Y|x_{11,12}) - E(Y|x_{11}) - E(Y|x_{12})$ against x_{11} and x_{12}: synthetic example

and we have obtained the correct order of magnitude for the effect of each variable. To estimate the posterior uncertainty about each S_i, a simulation method presented in Oakley and O'Hagan (2002) is used. (It is possible to derive formulae for var*[var$_X$ $\{E(Y|X)\}$ and var* $\{\text{var}(Y)\}$ for certain modelling choices, though the number of terms is very large.) The simulation method involves generating many additional runs of the code $\eta(\cdot)$ from its posterior distribution and re-estimating S_i each time. For the three groups $X_1, \ldots, X_5$, $X_6, \ldots, X_{10}$ and $X_{11}, \ldots, X_{15}$ approximate standard errors of the corresponding estimates $100\hat{S}_i$ were of the order of 0.2, 0.5 and 1 respectively. The Bayesian method is clearly validated in this case by the agreement between these figures and the errors that are seen in Table 1 between $100\hat{S}_i$ and the true values $100S_i$. In comparison, Saltelli and colleagues (personal communication) report that, for this example, 1024 runs are needed per input factor (so 15 360 runs in total) to achieve a standard error of 1. Clearly, our approach based on 250 runs has demonstrated a substantial gain in efficiency.

Note that the main effect proportions do not sum to 100% of the variance. The remaining variance after the main effects is estimated as 29% of the total variance (true value 28%). Since in this example we have independent inputs, this represents the sum of the interaction components.

Plotting the posterior expectation (with respect to the unknown function $\eta(\cdot)$) of $E(Y|x_i)$ against x_i for each variable also allows us to identify the three groups of variables. This is illustrated in Fig. 1. To illustrate the effect of interactions between x_i and x_j, we can plot the posterior expectation of $E(Y|\mathbf{x}_{i,j}) - E(Y|x_i) - E(Y|x_j)$ against x_i and x_j, and an example is given in Fig. 2.

4.2. Oil-field simulator

We now give a second example involving a computer model of a hydrocarbon reservoir. This model was used in Craig *et al.* (1997, 2001) to demonstrate their methodology for calibration

and forecasting. The model predicts pressures at various wells throughout the reservoir, at various time points. As in Craig *et al.* (2001), we consider 40 uncertain inputs. These include seven inputs ranging between 0.1 and 10 related to permeability in different regions of the reservoir, and 33 inputs ranging between 0 and 1 related to fault transmissibility. We choose notional distributions for these inputs; we first take log-transformations of the permeability inputs as in Craig *et al.* (2001). A further linear transformation is taken so that each input is on the same scale. We then suppose that each input has a normal distribution, with the ranges of each input representing six standard deviations.

We perform a sensitivity analysis on the output at a single well at a single time point. We have 101 runs of the code, with the design points chosen to form a Latin hypercube. For the prior mean function we set $\mathbf{h}(\mathbf{x})^{\mathrm{T}} = (1, \mathbf{x}^{\mathrm{T}})$. We then estimate the roughness parameters in B by their posterior mode and compute the main effect and best linear fit components. Out of the 40 inputs, it was found that three inputs accounted for almost all the variance. In Table 2 we

Table 2. Main effect indices and linear components: reservoir example

	$100\hat{S}_i$	$100\hat{V}_{x_i}/\widehat{var}(Y)$
X_2	6.3921	6.3921
X_3	70.4647	66.7173
X_4	17.2029	11.0936
Others	4.2252	4.0175

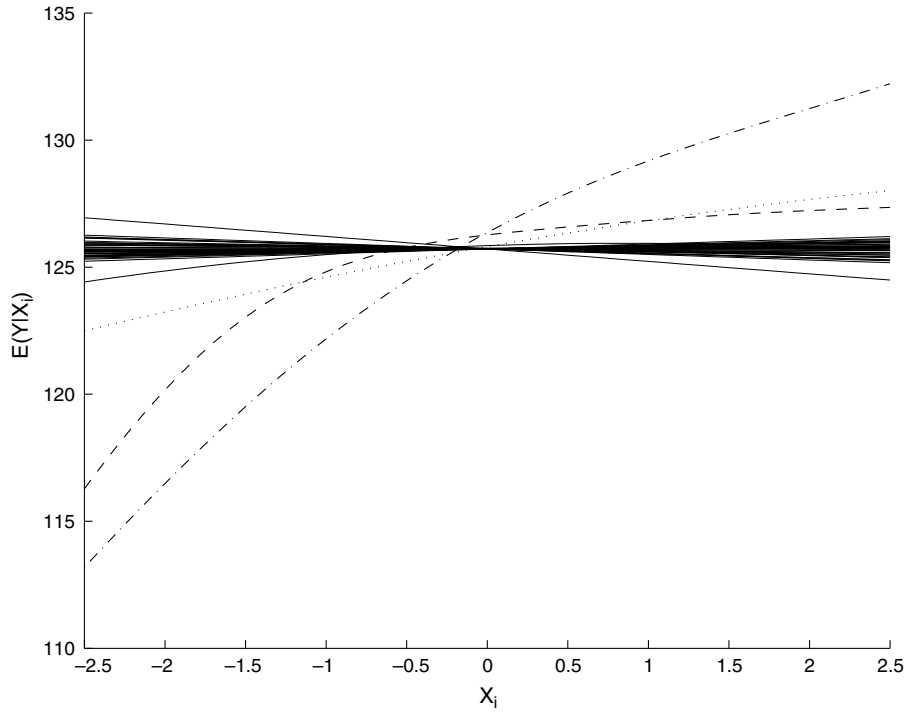

Fig. 3. Posterior expectation of $E(Y|x_i)$ against x_i for each input variable: reservoir example (·····, X_2; ·–·–·, X_3; – – –, X_4; ———, others)

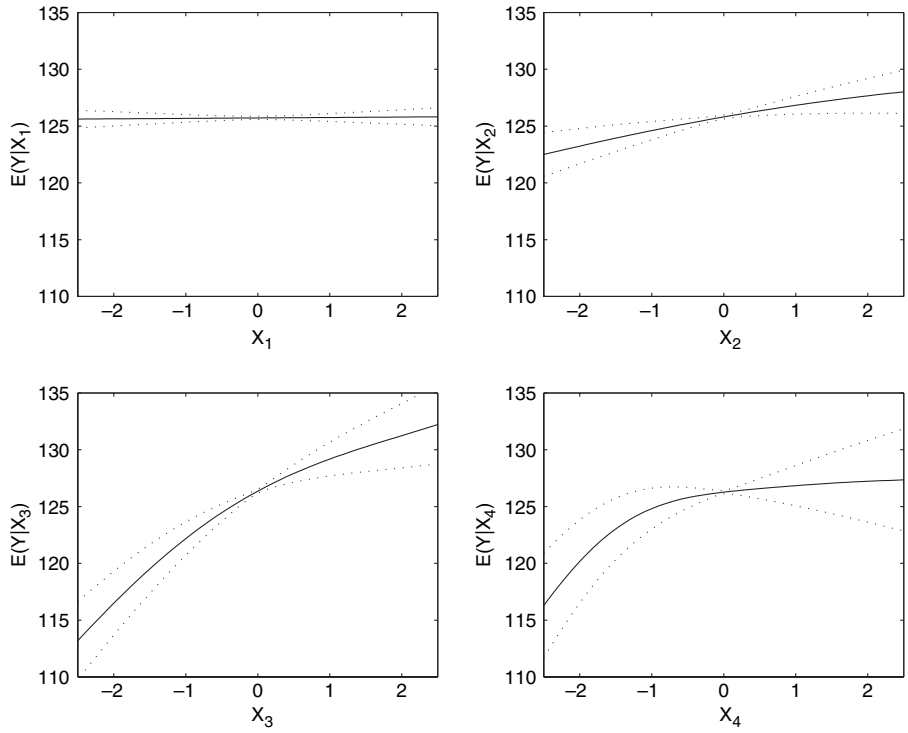

Fig. 4. Posterior expectation of $E(Y|x_i)$ against x_i for four input variables (·····, ±2 standard deviations): reservoir example

show the main effect indices and best linear fit components for these three inputs, and the sums of the main effects and for the other 37 inputs.

The sum of the main effect indices is close to 100%, so there is little evidence of interactions between the inputs. From Table 2 we can see evidence of non-linearity, particularly in input x_4. This can also be seen in Fig. 3, where we plot the conditional expectations of the output for each input.

Finally, to illustrate the uncertainty that we still have about the model, we show conditional means with 95% intervals of the output for four inputs, in Fig. 4.

Though we do not show the results here, we have also performed the same sensitivity analysis for different wells in the reservoir, and at different time points. In each case inputs x_3 and x_4 were by far the most influential, though in some cases in reverse order.

5. Conclusions

We have presented a Bayesian approach to probabilistic sensitivity analysis that builds on a range of existing measures and tools. Our method facilitates a deep and thorough analysis of the sensitivity of a model output to variation in its inputs—through decomposition of the output variance into components representing main effects and interactions, through further decomposition of individual terms into components for linear or other regression-based fits, and for non-linearity, and through graphical presentation of main effects and first-order interactions.

The method is highly efficient computationally. It will typically require far fewer model runs than are needed for conventional Monte-Carlo-based methods (even those employing variance

reduction techniques such as Latin hypercube sampling). This is particularly important in the case of expensive models, since Monte Carlo methods become infeasible if each model run takes an appreciable amount of computer time. The Bayesian approach also allows the complete range of sensitivity measures to be computed from a single set of model runs. It may therefore be valuable even for cheap models if a full decomposition of the output variance is desired, since other methods typically demand large repeat runs to compute each variance component.

The greater efficiency is achieved through a prior belief in smoothness of the function $\eta(\cdot)$. The methods that are presented here will not be appropriate if it is known that the model can respond erratically or discontinuously to changes in its inputs.

We have presented two examples to illustrate the power of this method. It is worth noting that the kinds of model for which a sensitivity analysis is required in many areas of science and technology will usually have large numbers of uncertain inputs. Our examples involve 15 and 40 uncertain model inputs and are therefore of realistic, albeit moderate, dimensionality. In contrast, most other available methods experience severely increasing computational demands as the number of inputs increases, so they are generally only applied in low dimensional problems.

Acknowledgements

We are grateful for the perceptive comments of two referees and an Associate Editor on an earlier draft of this paper.

References

Baker, R. D. (2001) Sensitivity analysis for health care models fitted to data by statistical methods. *IMA J. Mangmnt Sci.*, **12**, 1–17.

Bayarri, M., Berger, J., Higdon, D., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C. and Tu, J. (2002) A framework for validation of computer models. In *Proc. Wrkshp Foundations for Verification and Validation in the 21st Century* (eds D. Pace and S. Stevenson). San Diego: Society for Modeling and Simulation International.

Cox, D. C. (1982) An analytical method for uncertainty analysis of nonlinear output functions, with applications to fault-tree analysis. *IEEE Trans. Reliab.*, **31**, 265–268.

Craig, P. S., Goldstein, M., Rougier, J. C. and Seheult, A. H. (2001) Bayesian forecasting for complex systems using computer simulators. *J. Am. Statist. Ass.*, **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*, vol. III (eds C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi and N. D. Singpurwalla), pp. 36–93. New York: Springer.

Cukier, R. I., Fortuin, C. M., Schuler, K. E., Petschek, A. G. and Schaibly, J. H. (1973) Study of the sensitivity of coupled systems to uncertainties in rate coefficients. *J. Chem. Phys.*, **59**, 3873–3878.

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

French, S. (2003) Modelling, making inferences and making decisions: the roles of sensitivity analysis. *Top*, **11**, 229–252.

Haylock, R. G. and O'Hagan, A. (1996) On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. In *Bayesian Statistics 5* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 629–637. Oxford: Oxford University Press.

Helton, J. C. and Davis, F. J. (2000) Sampling-based methods. In *Sensitivity Analysis* (eds A. Saltelli, K. Chan and E. M. Scott), ch. 6, pp. 101–153. New York: Wiley.

Homma, T. and Saltelli, A. (1996) Importance measures in global sensitivity analysis of model output. *Reliab. Engng Syst. Safty*, **52**, 1–17.

Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc.* B, **63**, 425–464.

Kleijnen, J. P. C. (1997) Sensitivity analysis and related analyses: a review of some statistical techniques. *J. Statist. Computn Simuln*, **57**, 111–142.

Kleijnen, J. P. C. and Helton, J. C. (1999) Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. *Reliab. Engng Syst. Safty*, **65**, 147–185.

Mrawira, D., Welch, W. J., Schonlau, M. and Haas, R. (1999) Sensitivity analysis of computer models: World Bank HDM-III model. *J. Transprtn Engng*, **125**, 421–428.
Neal, R. (1999) Regression and classification using gaussian process priors. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 69–95. Oxford: Oxford University Press.
Oakley, J. (2002a) Value of information for complex cost-effectiveness models. *Technical Report 533/02*. Department of Probability and Statistics, University of Sheffield, Sheffield.
Oakley, J. (2002b) Eliciting Gaussian process priors for complex computer codes. *Statistician*, **51**, 81–97.
Oakley, J. E. and O'Hagan, A. (2002) Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, **89**, 769–784.
O'Hagan, A. (1991) Bayes-hermite quadrature. *J. Statist. Planng Inf.*, **91**, 245–260.
O'Hagan, A. (1992) Some Bayesian numerical analysis. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 345–363. Oxford: Oxford University Press.
O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics*, vol. 2B, *Bayesian Inference*. London: Arnold.
O'Hagan, A., Kennedy, M. and Oakley, J. E. (1999) Uncertainty analysis and other inference tools for complex computer codes (with discussion). In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 503–524. Oxford: Oxford University Press.
Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989) Design and analysis of computer experiments. *Statist. Sci.*, **4**, 409–435.
Saltelli, A., Chan, K. and Scott, M. (eds) (2000) *Sensitivity Analysis*. New York: Wiley.
Saltelli, A. and Sobol', I. M. (1995) About the use of rank transformation in sensitivity analysis of model output. *Reliab. Engng Syst. Safty*, **50**, 225–239.
Saltelli, A. and Tarantola, S. (2002) On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *J. Am. Statist. Ass.*, **97**, 702–709.
Saltelli, A., Tarantola, S. and Campolongo, F. (2000) Sensitivity analysis as an ingredient of modeling. *Statist. Sci.*, **15**, 377–395.
Saltelli, A., Tarantola, S. and Chan, K. (1999) A quantitative model independent method for global sensitivity analysis of model output. *Technometrics*, **41**, 39–56.
Sobol', I. M. (1993) Sensitivity analysis for nonlinear mathematical models. *Math. Modlng Comput. Expt*, **1**, 407–414.
Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992) Screening, predicting, and computer experiments. *Technometrics*, **34**, 15–25.