Screening the Input Variables to a Computer Model Via Analysis of Variance and Visualization

Matthias Schonlau *
William J. Welch **

An experiment involving a complex computer model or code may have tens or even hundreds of input variables and, hence, the identification of the more important variables (screening) is often crucial. Methods are described for decomposing a complex input-output relationship into effects. Effects are more easily understood because each is due to only one or a small number of input variables. They can be assessed for importance either visually or via a functional analysis of variance. Effects are estimated from flexible approximations to the input-output relationships model of the computer model. This allows complex nonlinear and interaction relationships to be identified. The methodology is demonstrated on a computer model of the relationship between environmental policy and the world economy.

1 Introduction

Computer models, also known as "math models" or "codes", are now frequently used in engineering, science and many other disciplines. To run the computer model software, the experimenter provides quantitative values for various input (explanatory) variables. The code then computes values for one or more output (response) variables. For instance, in a model of Arctic sea ice (Chapman et al., 1994), the input variables included rate of snowfall and ice albedo and the code produced values of ice mass, and so on. In circuit-design models (see, for example, Aslett et al., 1998), the input variables are transistor widths and other engineering parameters, and the output variables are measures of circuit performance such as time delays.

* RAND Corporation, 201 N Craig Street, Suite 202, Pittsburgh, PA 15213 matt@rand.org
** Department of Statistics, University of British Columbia, 333-6356 Agricultural Road, Vancouver, BC, Canada V6T 1Z2 will@stat.ubc.ca

Often, the computer model will be expensive to run, for example, if it solves a large number of differential equations which may require several hours or more of computer time. Thus, in a *computer experiment*, that is, an experiment with several runs of the computer model, there is need for careful design or choice of the values of the input variables and careful analysis of the data produced.

One major difference from traditional design and analysis of experiments with physical measurements is that computer models are often deterministic. Two runs of the code with the same set of values for the input variables would give identical results across the two runs for each output variable. Nonetheless, there will often be considerable uncertainty in drawing conclusions about the behavior of the input-output relationships from a limited number of runs, and statistical methods are required to characterize the uncertainty. The management of uncertainty is especially critical when the computer model has a high-dimensional set of input variables.

In applications such as the Arctic sea ice model (Chapman et al., 1994) mentioned above, a strategic objective of a preliminary computer experiment is screening: finding the important input variables. Screening is not a trivial task because the computer model is typically complex, and the relationships between input variables and output variables are not obvious. A common approach is to approximate the relationship by a statistical surrogate model, which is easier to explore. This is particularly useful when there are many input variables.

An example, which is discussed in this chapter, is the "Wonderland" model of Lempert et al. (2003), adapted from Herbert and Leeves (1998). In this case study, 41 input variables are manipulated, relating to population growth, economic activity, changes in environmental conditions, and other economic and demographic variables. The output is a quasi global human development index (HDI) which is a weighted index of net output per capita, death rates, annual flow of pollution, and the carrying capacity of the environment, spanning both "northern" and "southern" countries. The model has many output variables under various policy assumptions; we consider only one, corresponding to a "limits to growth" policy. Under this scenario, economic growth is intentionally limited by a constraint on global emissions. After 2010, both hemispheres must set carbon taxes high enough to achieve zero growth in emissions levels. Larger values of HDI correspond to greater human development and are better; see Lempert et al. (2003) for a full description of this measure.

Figure 1 shows scatter plots of the raw data from a "Latin hypercube" experimental design (see McKay, Conover, and Beckman, 1979) with 500 runs of the Wonderland code. The output variable HDI is plotted against two of the input variables shown in Section 6 to be important: economic innovation in the north (e.inov.n) and sustainable pollution in the south (v.spoll.s).

The first plot suggests a slight upward trend in HDI with e.inov.n. It will be shown in Section 6 that the trend is actually very strong; it looks weak here because there is considerable masking from other variables. The second plot

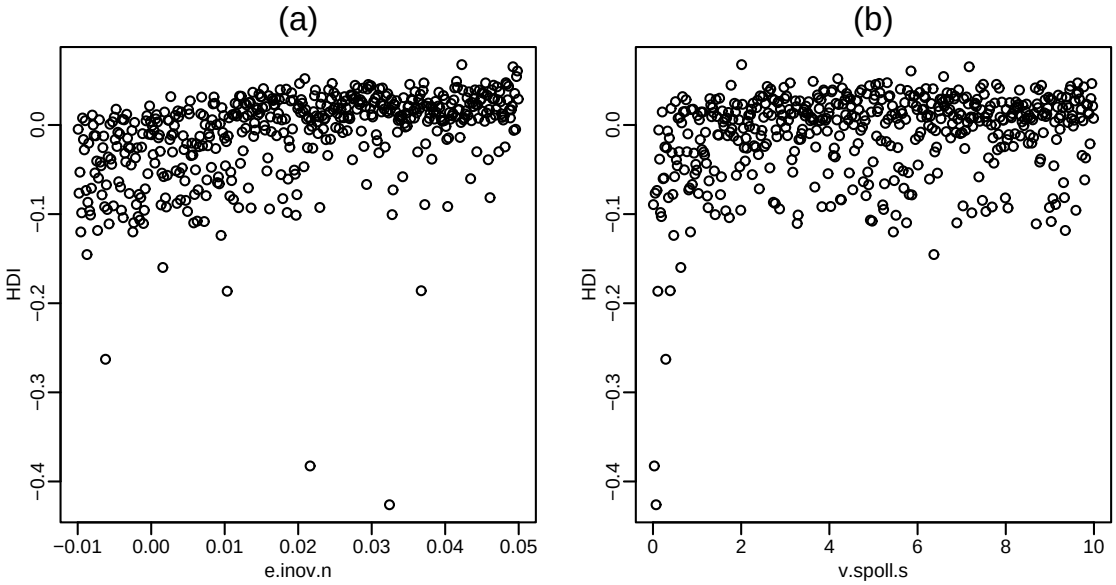

Fig. 1. Human development index (HDI) from the Wonderland model plotted against (a) economic innovation in the north (e.inov.n) and against (b) sustainable pollution in the south (v.spoll.s).

shows a very rapid drop off in HDI for low values of v.spoll.s. This nonlinearity in the computer model would have gone unnoticed without the three points on the left with the lowest HDI values. Thus, a design with fewer runs, or with fewer levels of each input variable, may well have missed the region containing these three points. (Note also that very small values of v.spoll.s do not always give such extreme values of HDI.) In our experience, nonlinear effects are common in computer experiments because the input variables often cover wide ranges. We shall explore the Wonderland application further in Section 6, but hopefully we have already illustrated some of the potential difficulties in screening the input variables of a computer model: large dimensionality, complex nonlinearities and masking.

There is a spectrum of methods proposed for screening variables in a computer experiment. They differ mainly in the assumptions they make about the form of an input-output relationship: with stronger assumptions, fewer runs are typically required.

Iman and Conover (1980) built a rank-regression approximation of a computer model of the discharge of a nuclear isotope from radioactive waste. With seven input variables, they used 200 runs in a Latin hypercube design. A sensitivity analysis followed from the least-squares estimates of the coefficients in the first-order rank-regression model. (A sensitivity analysis, which explores

how sensitive the output variable is to changes in the input variables, is similar to screening.) Morris (1991) described a screening method where the number of runs is a small multiple of the number of input variables. The method does not attempt to model the input-output relationship(s) of a computer code, rather it attempts to divide the variables qualitatively into three categories: unimportant; those with linear and additive effects; and those with more complex, nonlinear or interaction, effects. Twenty variables in a heat-transfer model were investigated using 108 runs.

With even fewer runs relative to the number of input variables, Bettonvil and Kleijnen (1996) used a sequential bifurcation algorithm (see Chapter **??**) to analyze a large deterministic global-climate model. The output is the world-wide CO_2 circulation in the year 2100. The model has 281 input variables, 15 of which were identified as important after 154 runs. The sequential bifurcation algorithm makes several strong assumptions to enable an experiment with fewer runs than input variables (a supersaturated design—see Chapter **??**). Each variable is considered at only two levels, and effects are assumed to be linear and additive. Moreover, the direction (sign) of each effect must be known a priori. The sequential bifurcation method was followed up with a traditional resolution IV design (Chapter **??**) for the most important factors in order to estimate a response surface model.

Gu and Wahba (1993) used a smoothing-spline approach with some similarities to the method described in this chapter, albeit in a context where random error is present. They approximated main effects and some specified two-variable interaction effects by spline functions. Their example had only three explanatory variables, so screening was not an issue. Nonetheless, their approach parallels the methodology we describe in this chapter, with a decomposition of a function into effects due to small numbers of variables, visualization of the effects, and an analysis of variance (ANOVA) decomposition of the total function variability.

The approach to screening the input variables in a computer model described in this chapter is based on a Gaussian random-function approximator to an input-output relationship. This is a flexible, data-adaptive paradigm with a long history in the analysis of computer experiments. Similarly, decomposing the random-function approximator into low-order effects for the purposes of identifying and examining the important effects has been in use for some time. The estimated effects are visualized or quantified via a functional ANOVA decomposition. In an experiment with six input factors and 32 runs, Sacks et al. (1989) identified the important (nonlinear) main effects and two-variable interaction effects. Welch et al. (1992) described a stepwise method for adding important input variables to the statistical approximator and visualized the important effects. They were able to find the important nonlinear and interaction effects among 20 input variables with 50 runs. Chapman et al. (1994) and Gough and Welch (1994) performed sensitivity analyses of climate models with 13 and 7 input variables, respectively, and Mrawira et al. (1999) were able to deal with 35 input variables in a civil-engineering appli-

cation. With up to 36 variables, Aslett et al. (1998) and Bernardo et al. (1992) used visualization of important effects to guide the sequential optimization of electronic circuit designs. Santner, Williams and Notz (2003, Chapter 7) also summarized this approach.

Thus, decomposition of a random-function approximator of a computer model into low-dimensional effects, in order to identify the important effects and examine them visually and quantitatively, has been widely applied and reported by many authors. However, the implementation of these methods has not been described, with the partial exception of Schonlau (1997), a shortcoming which we address in this chapter.

The chapter is organized as follows. Section 2 reviews the key underlying random-function approximator. Effects are defined in Section 3, leading to a functional ANOVA, and Section 4 describes their estimation. Section 5 summarizes the steps in the work flow for identifying and visualizing the important estimated effects. This approach has proved to be a powerful screening tool, as evidenced by the above examples. In Section 6, we return to the Wonderland model and demonstrate how the methodology is used. Some concluding remarks are given in Section 7. Some details of the derivation of the best linear unbiased predictor of an effect are provided in Appendix A, and Appendix B shows how the high-dimensional integrals required for the estimated effects, for their pointwise standard errors, and for the ANOVA decomposition boil down to a series of low-dimensional integrals under certain, fairly common, conditions.

2 The random-function approximator

Here, we give a brief review of methods for the analysis of computer experiments, concentrating on statistical approximation of the computer model. Strategies for the design and analysis of computer experiments have been described by many authors, including Currin et al. (1991), Koehler and Owen (1996), Sacks et al. (1989), Santner, Williams, and Notz (2003), and Welch et al. (1992). All these authors take into account the deterministic nature of a code, such as the Wonderland model, and also provide uncertainty measures via a statistical approximation model.

In general, suppose that a code is run n times in a computer experiment. Each run has a different set of values for the d-dimensional vector of input variables, $\boldsymbol{x} = (x_1, \ldots, x_d)^T$. A particular output variable is denoted by $y(\boldsymbol{x})$. With several output variables, each is treated separately. The data consist of n input vectors, $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}$, chosen from an input region of interest, $\mathcal{X}$, and the vector of n corresponding output values, denoted by $\boldsymbol{y}$.

Following the approach of the above authors, the output variable $y(\boldsymbol{x})$ is treated as a realization of a random function:

$$Y(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\beta} + Z(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), \ldots, f_h(\boldsymbol{x})]'$ is a vector of h known regression functions, $'$ denotes transpose, $\boldsymbol{\beta}$ is a $h \times 1$ vector of parameters to be estimated, and Z is a Gaussian stochastic process indexed by $\boldsymbol{x}$. It is assumed that $Z(\boldsymbol{x})$ has mean zero and constant variance, σ^2, for all $\boldsymbol{x}$. The covariance between $Z(\boldsymbol{x})$ and $Z(\tilde{\boldsymbol{x}})$ at two input vectors, $\boldsymbol{x} = (x_1, \ldots, x_d)'$ and $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_d)'$ is

$$\mathrm{Cov}[Z(\boldsymbol{x}), Z(\tilde{\boldsymbol{x}})] = \sigma^2 R(\boldsymbol{x}, \tilde{\boldsymbol{x}}), \tag{2}$$

where $R(\cdot, \cdot)$ is a "correlation function" and $\tilde{\boldsymbol{x}}$ denotes a different set of input values from $\boldsymbol{x}$.

The correlation function $R(\cdot, \cdot)$ in (2) is central to this statistical model. The *power-exponential* class of correlation functions is a popular choice, for its computational simplicity and because it has been successful in many applications. The power-exponential correlation function is

$$R(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \prod_{j=1}^{d} \exp(-\theta_j |x_j - \tilde{x}_j|^{p_j}), \tag{3}$$

where $\theta_j \geq 0$ and $0 < p_j \leq 2$ $(j = 1, \ldots, d)$ are parameters which can be estimated from the data, often via maximum likelihood. The p_j can be interpreted as smoothness parameters—the output surface is smoother with respect to x_j as p_j increases. For $p_j = 2$, the surface is infinitely differentiable. For $0 < p_j < 2$, the surface is continuous, but not differentiable. As p_j increases between 0 and 2, however, the surface appears to fluctuate less and less, and in this sense could be said to be smoother. The θ_j indicate the extent to which the variation in the output function is local with respect to x_j. If θ_j is large, the correlation (3) between observations or outputs at $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ falls rapidly with the distance between x_j and $\tilde{x}_j$, and the function is difficult to predict in the x_j direction.

We next describe the first steps in the derivation of the *best linear unbiased predictor* (BLUP) of $Y(\boldsymbol{x})$ at an untried input vector $\boldsymbol{x}$ (see, for example, Sacks et al., 1989). Similar steps will be used in Section 4 to estimate the effects of one, two or more input variables. It will then be apparent how to adapt results and computational methods for *predicting* $Y(\boldsymbol{x})$ to the problem of *estimating* such effects.

Following the random-function model (1), consider the prediction of $Y(\boldsymbol{x})$ by $\hat{Y}(\boldsymbol{x}) = \boldsymbol{a}'(\boldsymbol{x})\boldsymbol{y}$, that is, a linear combination of the n values of the output variable observed in the experiment. The best linear unbiased predictor is obtained by minimizing the mean squared error of the *linear predictor* or *approximator*, $\hat{Y}(\boldsymbol{x})$. The mean squared error, $\mathrm{MSE}[\hat{Y}(\boldsymbol{x})]$, is

$$\begin{aligned}
\mathrm{E}[Y(\boldsymbol{x}) - \hat{Y}(\boldsymbol{x})]^2 &= \mathrm{E}[\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\beta} + Z(\boldsymbol{x}) - \boldsymbol{a}'(\boldsymbol{x})(\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{z})]^2 \\
&= \{[\boldsymbol{f}'(\boldsymbol{x}) - \boldsymbol{a}'(\boldsymbol{x})\boldsymbol{F}]\boldsymbol{\beta}\}^2 \\
&\quad + \mathrm{Var}[Z(\boldsymbol{x})] + \boldsymbol{a}'(\boldsymbol{x})\mathrm{Cov}(\boldsymbol{z})\boldsymbol{a}(\boldsymbol{x}) - 2\boldsymbol{a}'(\boldsymbol{x})\mathrm{Cov}[Z(\boldsymbol{x}), \boldsymbol{z}],
\end{aligned}$$

where $\boldsymbol{F}$ is the $n \times k$ matrix with row i containing the regression functions $\boldsymbol{f}'(\boldsymbol{x}^{(i)})$ for run i in the experimental plan, and $\boldsymbol{z} = [Z(\boldsymbol{x}^{(1)}), \ldots, Z(\boldsymbol{x}^{(n)})]'$ is the $n \times 1$ vector of random Z values, with element i corresponding to run i. From the covariance function (2) we can write $\mathrm{Cov}(\boldsymbol{z})$ as $\sigma^2\boldsymbol{R}$, where $\boldsymbol{R}$ is an $n \times n$ matrix with element (i, j) given by $R(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$, and $\mathrm{Cov}[Z(\boldsymbol{x}), \boldsymbol{z})]$ as $\sigma^2\boldsymbol{r}(\boldsymbol{x})$, where $\boldsymbol{r}(\boldsymbol{x})$ is an $n \times 1$ vector with element i given by $R(\boldsymbol{x}, \boldsymbol{x}^{(i)})$. With this notation, the mean squared error of $\hat{Y}(\boldsymbol{x})$ is

$$\mathrm{MSE}[\hat{Y}(\boldsymbol{x})] = \{[\boldsymbol{f}'(\boldsymbol{x}) - \boldsymbol{a}'(\boldsymbol{x})\boldsymbol{F}]\boldsymbol{\beta}\}^2 + \mathrm{Var}[Z(\boldsymbol{x})] + \sigma^2\boldsymbol{a}'(\boldsymbol{x})\boldsymbol{R}\boldsymbol{a}(\boldsymbol{x}) - 2\sigma^2\boldsymbol{a}'(\boldsymbol{x})\boldsymbol{r}(\boldsymbol{x}). \quad (4)$$

Some further simplification of this expression is possible, for example, by using the fact that $\mathrm{Var}[Z(\boldsymbol{x})] = \sigma^2$, by assumption. We leave the mean squared error in this form, however, to facilitate comparison with its counterpart in Section 4 for the estimated effect of a group of variables.

We now choose $\boldsymbol{a}(\boldsymbol{x})$ to minimize (4). To avoid an unbounded contribution from the first term on the right-hand side of (4) from large elements in $\boldsymbol{\beta}$, the contribution is eliminated by imposing the constraint

$$\boldsymbol{F}\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}).$$

This constraint is also sometimes motivated by unbiasedness, that is, from $\mathrm{E}[\hat{Y}(\boldsymbol{x})] = \mathrm{E}[Y(\boldsymbol{x})]$ for all $\boldsymbol{\beta}$. Thus, the best linear unbiased predictor (BLUP), or optimal value of $\boldsymbol{a}(\boldsymbol{x})$, results from the following optimization problem:

$$\min_{\boldsymbol{a}(\boldsymbol{x})} \mathrm{Var}[Z(\boldsymbol{x})] + \sigma^2\boldsymbol{a}'(\boldsymbol{x})\boldsymbol{R}\boldsymbol{a}(\boldsymbol{x}) - 2\sigma^2\boldsymbol{a}'(\boldsymbol{x})\boldsymbol{r}(\boldsymbol{x}) \quad (5)$$

subject to

$$\boldsymbol{F}\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})\,.$$

The optimal $\boldsymbol{a}(\boldsymbol{x})$ turns out to give the following form for the BLUP (or approximator) (see, for example, Sacks et al., 1989):

$$\hat{Y}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})\hat{\boldsymbol{\beta}} + \boldsymbol{r}'(\boldsymbol{x})\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}), \quad (6)$$

where $\hat{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}\boldsymbol{R}^{-1}\boldsymbol{y}$ is the generalized least squares estimator of $\boldsymbol{\beta}$. If we put the optimal $\boldsymbol{a}(\boldsymbol{x})$ into the expression for the mean squared error (4), we obtain the following standard error, $\mathrm{se}[\hat{Y}(\boldsymbol{x})]$, for $\hat{Y}(\boldsymbol{x})$:

$$\mathrm{se}^2[\hat{Y}(\boldsymbol{x})] = \mathrm{Var}[Z(\boldsymbol{x})] - \sigma^2\boldsymbol{r}(\boldsymbol{x})'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x}) + \sigma^2[\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x})]'(\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}[\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{r}(\boldsymbol{x})] \quad (7)$$

This formula ignores the uncertainty from estimating the correlation parameters, for example, the θ_j and p_j in (3). Some comments on this issue are made in Section 7.

3 Effects

The important input variables are those that have large effects on the output variable. As with traditional analysis of variance, we can look at the main effects of single variables, or the joint or interaction effects of several variables at a time.

Suppose that we are interested in the effect of a subset of input variables, held in a vector $\boldsymbol{x}_e$, where e denotes the set of subscripts of the variables of interest. The vector of remaining variables is denoted by $\boldsymbol{x}_{-e}$. For example, when interest is in the effects of x_1 and x_2 among $d > 2$ variables, we have $e = \{1, 2\}$ and $\boldsymbol{x}_e = (x_1, x_2)$, whereupon $\boldsymbol{x}_{-e} = (x_3, \ldots, x_d)$. Without loss of generality we rearrange the order of the input variables so that we can write $\boldsymbol{x} = (\boldsymbol{x}_e, \boldsymbol{x}_{-e})$. To obtain a unique and workable definition of the effect of $\boldsymbol{x}_e$ is essentially the problem of how to deal with the variables in $\boldsymbol{x}_{-e}$. We next discuss several ways of approaching this problem.

Keeping the variables in $\boldsymbol{x}_{-e}$ fixed requires little new methodology. We consider $y(\boldsymbol{x}_e, \boldsymbol{x}_{-e})$ as a function of $\boldsymbol{x}_e$, with $\boldsymbol{x}_{-e}$ fixed, for example at the variable midranges. Estimates and pointwise standard errors of such effects follow immediately from (6) and (7), with $\hat{Y}(\boldsymbol{x}_e, \boldsymbol{x}_{-e})$ and $\mathrm{se}[\hat{Y}(\boldsymbol{x}_e, \boldsymbol{x}_{-e})]$ considered as functions of $\boldsymbol{x}_e$. There are two serious disadvantages of this method, however. First, in the presence of interaction effects involving one or more variables in $\boldsymbol{x}_e$ and one or more variables in $\boldsymbol{x}_{-e}$, the magnitude of the effect of $\boldsymbol{x}_e$ may change depending on the levels chosen for $\boldsymbol{x}_{-e}$, and thus the effect of $\boldsymbol{x}_e$ is not isolated. Consequently, there is no straightforward decomposition of the total variation in $y(\boldsymbol{x})$, or its predictor $\hat{Y}(\boldsymbol{x})$, into contributions from various effects.

Alternatively, we may define an effect by "integrating out" the other variables. Under certain conditions, this leads to a simple decomposition of $y(\boldsymbol{x})$ into contributions from various effects, with a corresponding decomposition of the total variance of $y(\boldsymbol{x})$ over $\mathcal{X}$. Moreover, as we shall see in Section 4, these effects and their variance contributions can be easily estimated. Hence, defining an effect by integrating out the other variables is the method pursued for the remainder of this chapter.

For a convenient decomposition of $y(\boldsymbol{x})$, we need two conditions on the region of interest of the input variables. First, $\mathcal{X}$ is assumed to be a direct product of one-dimensional regions, which we write as

$$\mathcal{X} = \otimes_{j=1}^{d} \mathcal{X}_j, \tag{8}$$

where $\mathcal{X}_j$ denotes the values of interest for variable x_j, for instance a continuous interval or a discrete set of points (for which integration is interpreted as summation). Secondly, we assume that integration is with respect to a weight function, $w(\boldsymbol{x})$, which is a product of functions of one input variable at a time:

$$w(\boldsymbol{x}) = \prod_{j=1}^{d} w_j(x_j) \qquad \text{for } x_j \in \mathcal{X}_j,\ j = 1, \ldots, d. \tag{9}$$

Often, the $w_j(x_j)$ are chosen to be equal, representing uniform interest across the range of values for x_j. In other applications, x_j might be a variable in the computer code because its value in nature is uncertain. If this uncertainty is represented by a given statistical distribution, for example, a normal distribution, then the distribution would be used as the weight function, $w_j(x_j)$. The conditions (8) and (9) occur frequently in applications; a minor relaxation of them is discussed in Section 4.

Under the assumptions (8) and (9), the *marginal effect*, $\bar{y}_e(\boldsymbol{x}_e)$, of $\boldsymbol{x}_e$ is defined by integrating out the other variables,

$$\bar{y}_e(\boldsymbol{x}_e) = \int\limits_{\otimes_{j\notin e}\mathcal{X}_j} y(\boldsymbol{x}_e, \boldsymbol{x}_{-e}) \prod_{j\notin e} w_j(x_j) d\, x_j \qquad \text{for } \boldsymbol{x}_e \in \otimes_{j\in e}\mathcal{X}_j. \tag{10}$$

Note that a marginal effect is the overall effect of all the variables in $\boldsymbol{x}_e$. With just one variable in $\boldsymbol{x}_e$, we call this a *main* effect; with two or more variables, we call this a *joint* effect.

We use the marginal effects (10) to decompose $y(\boldsymbol{x})$ as follows into *corrected* or *adjusted* effects involving no variables, one variable at a time, two variables at a time, and so on, up to the contribution from all the variables:

$$y(\boldsymbol{x}) = \mu_0 + \sum_{j=1}^{d} \mu_j(x_j) + \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} \mu_{jj'}(x_j, x_{j'}) + \cdots + \mu_{1\ldots d}(x_1, \ldots, x_d) \tag{11}$$

for $\boldsymbol{x} \in \mathcal{X}$, where

$$\mu_0 = \int\limits_{\mathcal{X}} y(\boldsymbol{x}) w(\boldsymbol{x}) d\, \boldsymbol{x}$$

is an overall average,

$$\mu_j(x_j) = \bar{y}_j(x_j) - \mu_0 \qquad \text{for } x_j \in \mathcal{X}_j \tag{12}$$

is the corrected main effect of x_j,

$$\mu_{jj'}(x_j, x_{j'}) = \bar{y}_{jj'}(x_j, x_{j'}) - \mu_j(x_j) - \mu_{j'}(x_{j'}) - \mu_0 \quad \text{for } x_j, x_{j'} \in \mathcal{X}_j \otimes \mathcal{X}_{j'} \tag{13}$$

is the corrected joint effect or *interaction* effect of x_j and $x_{j'}$, and so on. Thus, each corrected effect is the corresponding marginal effect corrected for all lower-order terms.

For example, suppose interest centers on the variables x_1 and x_2. If their interaction effect, $\mu_{12}(x_1, x_2)$, has an important magnitude, it is not meaningful to consider the effects of x_1 or x_2 in isolation. We would look at their overall joint effect,

$$\bar{y}_{12}(x_1, x_2) = \mu_0 + \mu_1(x_1) + \mu_2(x_2) + \mu_{12}(x_1, x_2) \qquad \text{for } x_1, x_2 \in \mathcal{X}_1 \otimes \mathcal{X}_2.$$

Similar comments apply to higher-order effects. In practice, we will have to estimate the marginal effects, and hence the corrected effects, to decide which are important.

The effects (11) are orthogonal with respect to the weight function $w(\boldsymbol{x})$, leading to a decomposition of the total variance of $y(\boldsymbol{x})$, called the *ANOVA decomposition* or *functional analysis of variance* as follows:

$$\begin{aligned}\int_{\mathcal{X}} [y(\boldsymbol{x}) - \mu_0]^2 w(\boldsymbol{x}) d\,\boldsymbol{x} \quad &= \quad \sum_{j=1}^{d} \int_{\mathcal{X}_j} \mu_j^2(x_j) w_j(x_j) d\,x_j \\ &+ \quad \sum_{j=1}^{d-1} \sum_{j'=j+1}^{d} \int_{\mathcal{X}_j \otimes \mathcal{X}_{j'}} \mu_{jj'}^2(x_j, x_{j'}) w_j(x_j) w_{j'}(x_{j'}) d\,x_j d\,x_{j'} \\ &+ \cdots + \int_{\mathcal{X}} \mu_{1\ldots d}^2(x_1, \ldots, x_d) \prod_{j=1}^{d} w_j(x_j) d\,x_j. \end{aligned} \tag{14}$$

A quantitative measure of the importance of any effect, and hence the associated variables, follows from the percentage contribution of each term on the right-hand side to the total variance on the left. The functional analysis of variance (ANOVA) in (14) goes back at least as far as Hoeffding (1948).

4 Estimating the effects

Estimating the marginal (main or joint) effects $\bar{y}_e(\boldsymbol{x}_e)$ in (10) is key to our approach for assessing the importance of variables. From the estimated marginal effects, we can also estimate the corrected effects in (11) and the ANOVA decomposition (14). Furthermore, when visualizing the large estimated effects it is easier to interpret main or joint effects than their corrected counterparts.

If $y(\boldsymbol{x})$ is treated as if it is a realization of the random function $Y(\boldsymbol{x})$ in (1), it follows that $\bar{y}_e(\boldsymbol{x}_e)$ is a realization of the analogously integrated random function,

$$\bar{Y}_e(\boldsymbol{x}_e) = \bar{\boldsymbol{f}}_e'(\boldsymbol{x}_e)\boldsymbol{\beta} + \bar{Z}_e(\boldsymbol{x}_e) \qquad \text{for } \boldsymbol{x}_e \in \otimes_{j \in e} \mathcal{X}_j. \tag{15}$$

Here, $\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e)$ and $\bar{Z}_e(\boldsymbol{x}_e)$ have the input variables not in $\boldsymbol{x}_e$ integrated out as in (10):

$$\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e) = \int_{\otimes_{j \notin e} \mathcal{X}_j} f(\boldsymbol{x}_e, \boldsymbol{x}_{-e}) \prod_{j \notin e} w_j(x_j) d\,x_j \qquad \text{for } \boldsymbol{x}_e \in \otimes_{j \in e} \mathcal{X}_j \tag{16}$$

and

$$\bar{Z}_e(\boldsymbol{x}_e) = \int_{\otimes_{j \notin e} \mathcal{X}_j} Z(\boldsymbol{x}_e, \boldsymbol{x}_{-e}) \prod_{j \notin e} w_j(x_j) d\,x_j \qquad \text{for } \boldsymbol{x}_e \in \otimes_{j \in e} \mathcal{X}_j. \tag{17}$$

The statistical properties of the stochastic process $\bar{Z}_e(\boldsymbol{x}_e)$ and the derivation of the BLUP of $\bar{Y}_e(\boldsymbol{x}_e)$ are derived in Appendix A. It is shown that the BLUP of $\bar{Y}_e(\boldsymbol{x}_e)$ is

$$\hat{\bar{Y}}_e(\boldsymbol{x}_e) = \bar{\boldsymbol{f}}_e(\boldsymbol{x}_e)\hat{\boldsymbol{\beta}} + \bar{\boldsymbol{r}}_e'(\boldsymbol{x}_e)\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{\beta}}), \tag{18}$$

and its standard error is given by

$$\begin{aligned} \mathrm{se}^2[\hat{\bar{Y}}_e(\boldsymbol{x}_e)] \;&=\; \mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)] - \sigma^2\bar{\boldsymbol{r}}_e(\boldsymbol{x})'\boldsymbol{R}^{-1}\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e) \\ + \sigma^2[\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e) - \boldsymbol{F}'\boldsymbol{R}^{-1}\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e)]'&(\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}[\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e) - \boldsymbol{F}'\boldsymbol{R}^{-1}\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e)], \end{aligned} \tag{19}$$

where $\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e)$ is defined following (22) in Appendix A.

In other words, software for computing the BLUP of $Y(\boldsymbol{x})$ and its standard error is easily modified for estimating effects and, hence, the ANOVA decomposition, provided that we can compute $\mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)]$ from (21) in Appendix A, $\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e)$ following (22), and $\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e)$ in (16), quantities which will involve high-dimensional integrals in high-dimensional problems. These computations are described in AppendixB.

It is possible to relax the product-region condition (8) in some experiments. For example, Mrawira et al. (1999) dealt with several groups of variables where there were constraints like $x_1 \leq x_2$. The triangular input space for such a group had a product arrangement with all other variables or groups of variables. Thus, in all the above formulas for estimated effects or their standard errors, we merely treat the variables in a group together as if they were a single variable. This means, however, that the estimated effect for a group cannot be decomposed further into contributions from its constituent variables.

5 Steps for Identifying and visualizing the important estimated effects

To screen the input variables, we carry out the following steps.

1. Estimate by maximum likelihood the unknown parameters, $\boldsymbol{\beta}$ in (1), σ^2 in (2), and the correlation parameters, for example, the θ_j and p_j in (3).
2. Before continuing with a screening analysis, it is prudent to check the overall accuracy of the approximator in (6) and the validity of its standard error (7) by cross validation (see, Jones, Schonlau and Welch, 1998).
3. Compute the estimated marginal effects defined in (18) by carrying out the required integrations as described in Appendix B. This will usually be done for all main effects and all two-variable joint effects.
4. For each estimated marginal effect, compute the corresponding estimated corrected effect by subtracting all estimated lower-order corrected effects. This is best done recursively, correcting the main effects first, then correcting the two-variable effects, and so on.
5. Using the estimated corrected effects, compute the estimated contributions in the functional analysis of variance (14).

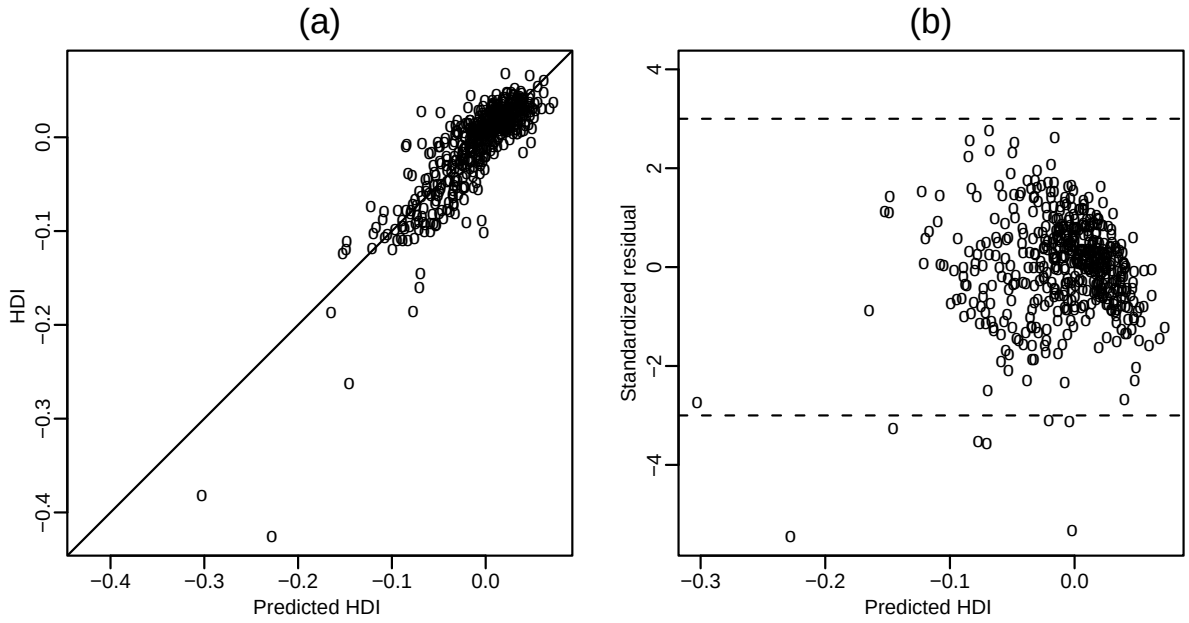

Fig. 2. Diagnostics for the Wonderland approximating model: (a) actual human development index (HDI) values versus their cross-validation predictions; (b)standardized cross-validation residuals versus cross-validation predictions.

6. If an estimated interaction effect makes a substantial contribution to the ANOVA decomposition, the corresponding joint effect (18) is plotted against the relevant input variables as a contour or perspective plot. The standard error (19) can be also be plotted against the same input variables in a separate plot.
7. Any input variable that has a large ANOVA contribution from an estimated (corrected) main effect but does not appear in any large ANOVA contributions from interaction effects has its estimated (uncorrected) main effect plotted. Approximate pointwise confidence intervals based on the standard error can also be shown.

6 Application: the Wonderland model

We illustrate these methods using the Wonderland computer model outlined in Section 1. This model exemplifies the type of screening problem we have in mind, as we shall find that it has highly nonlinear, interactive effects that demand a flexible, data-adaptive statistical modeling strategy. The computer model has 41 input variables and we focus on one particular quasi global

human development index (HDI) output variable resulting from a policy which might be called "limits to growth". The data consist of 500 model runs from a "space-filling" Latin hypercube design (see McKay, Conover and Beckman, 1979).

The first step in the analysis is to fit the random-function model (1) and to check the accuracy of the resulting approximator. We use a simple random-function model:

$$Y(\boldsymbol{x}) = \beta_0 + Z(\boldsymbol{x}),$$

where the regression component is just a constant, β_0. The unknown parameters, β_0, σ^2 in (2) and the correlation parameters θ_j and p_j for $j = 1, \ldots, 41$ in (3), are estimated by maximum likelihood. Figure 2(a) shows the actual HDI value $y(\boldsymbol{x}^{(i)})$ from run i of the Wonderland model versus its leave-one-out cross-validated prediction, $\hat{Y}_{-i}(\boldsymbol{x}^{(i)})$ for $i = 1, \ldots, 500$.

The subscript $-i$ indicates that the approximator (6) is built from all the data except run i. (The random-function correlation parameters are not re-estimated.) Figure 2(a) shows fairly good accuracy of approximation, though with some over-prediction of the extremely low HDI values. Figure 2(b) plots the standardized cross-validated residual,

$$[y(\boldsymbol{x}^{(i)}) - \hat{Y}_{-i}(\boldsymbol{x}^{(i)})]/\mathrm{se}_{-i}[\hat{Y}(\boldsymbol{x})], \tag{20}$$

versus $\hat{Y}_{-i}(\boldsymbol{x}^{(i)})$ for $i = 1, \ldots, 500$, where the standard error $\mathrm{se}_{-i}[\hat{Y}(\boldsymbol{x})]$ is computed from (7), again without the data from run i. The plot shows some standardized residuals falling outside the bands at ± 3, indicating that the error of approximation is sometimes a little larger in magnitude than is suggested by the standard error.

The next step is to compute the estimated marginal effects (18). Following our usual practice, this is done for all main effects and all two-variable joint effects. The required integrations over the remaining 40 or 39 variables, respectively, are computed as described in Appendix B.

Table 1. Estimated main effects and two-variable interaction effects accounting for more than 1% of the total variance of the predictor; the variable names are defined in Table 2

Effect	% of Total Variance	Effect	% of Total Variance
e.inov.n	24.3	v.spoll.s × v.drop.s	2.7
v.spoll.s	13.5	e.grth.n × e.inov.n	1.9
e.inov.s	12.1	v.drop.s	1.9
e.cinov.s	5.3	e.finit.s	1.5
v.spoll.s × v.cfsus.s	4.6	e.inov.n × e.inov.s	1.4
v.drop.s × v.cfsus.s	3.7	v.cfsus.s	1.2

Each estimated marginal effect leads to the corresponding estimated corrected effect in (12) or (13). This is done recursively: the estimated main effects are corrected first, followed by the two-variable interaction effects.

The functional analysis of variance in (14) is then computed from the estimated corrected effects. Here, the 41 main effects and 820 two-factor-interaction effects together account for about 89% of the total variance of the predictor. Hence, about 11% of the predictor's total variability is due to higher-order effects. Table 1 shows the estimated main effects and interaction effects that contribute at least 1% to the functional ANOVA. These 12 effects together account for about 74% of the total variation.

Only six variables appear in these 12 effects; they are described in Table 2.

Table 2. Wonderland input variables that appear in the important estimated effects of Table 1. Prefix "e." or "v." indicates an economic or environmental variable, respectively; suffix ".n" or ".s" indicates the northern region or the southern region, respectively

Variable	Description
e.finit	Flatness of initial decline in economic growth
e.grth	Base economic growth rate
e.inov	Innovation rate
e.cinov	Effect of innovation policies (pollution taxes) on growth
v.spoll	Sustainable pollution
v.cfsus	Change in level of sustainable pollution when natural capital is cut in half
v.drop	Rate of drop in natural capital when pollution flows are above the sustainable level

The ANOVA suggests that e.inov.n (economic innovation in the north) is an important input variable. Its estimated main effect in Figure 3(a) shows a strong, approximately linear trend. The estimated increase in HDI is fairly substantial: about 0.06 over the e.inov.n range. This was not obvious from the scatter plot in Figure 1(a); certainly any guess as to the magnitude of the increase would have been much smaller. The relationship was masked by other variables.

The estimated main effect of v.spoll.s (sustainable pollution in the south) in Figure 3(b) confirms the same nonlinearity that we could see in the scatter plot in Figure 1(b). The drop in HDI over the first twentieth of the range of v.spoll.s is substantial. Given that we sampled 500 points we would suspect roughly one twentieth or 25 of the HDI values to be low. However, the scatter plot in Figure 1(b) shows only three low points. This hints at a highly local interaction.

The analysis of variance in Table 1 does indeed identify several estimated interaction effects involving v.spoll.s; the largest is that with v.cfsus.s (change

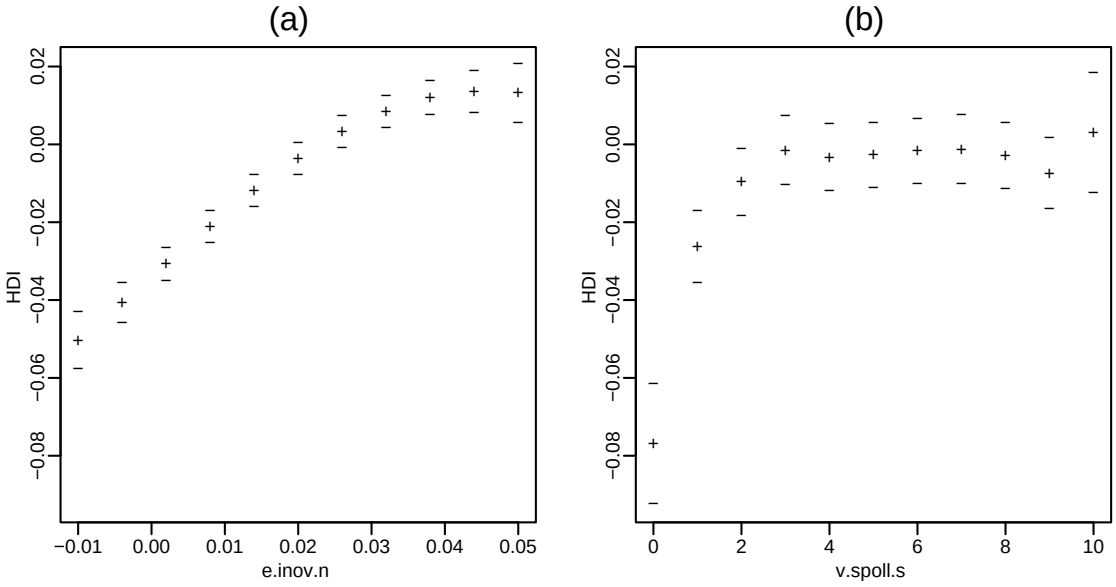

Fig. 3. Estimated main effects on HDI in the Wonderland model: (a) estimated main effect of e.inov.n (economic innovation in the north); (b) estimated main effect of v.spoll.s (sustainable pollution in the south). The estimated effects are denoted by "+" and approximate 95% pointwise confidence limits are denoted by "−".

in sustainable pollution in the south). Figure 4(a) shows the estimated joint effect of these two input variables on HDI. The surface is fairly flat for most of the area. As previously seen in the main effect plot, HDI increases rapidly with v.spoll.s when v.spoll.s is close to its lower limit. Now we see that this increase is larger for high values of v.cfsus.s (an increase from -0.12 to 0) than for low values of v.cfsus.s (an increase from -0.06 to 0). This difference appears to be substantial relative to the standard errors shown in Figure 4(b), which are roughly of order 0.01.

For comparison, we also use stepwise regression to select variables, specifically the R function `step` (R Development Core Team, 2005), which uses Akaike's information criterion (see Akaike, 1973). The selection from all 41 input variables results in a first-order model (main effects model) with 15 variables, but e.finit.s and v.cfsus.s in Table 1 are not included. Extending the model search space to allow all second-order terms also, that is, the 41 squares and 820 bilinear interaction effects of the input variables, yields a final model with 62 terms. Again e.finit.s and v.cfsus.s do not appear. Thus, the bilinear v.spoll.s × v.cfsus.s interaction effect is not included, contrary to Table 1. (Note that a two-factor interaction effect is defined to be a more general, non-additive effect in the random-function model.)

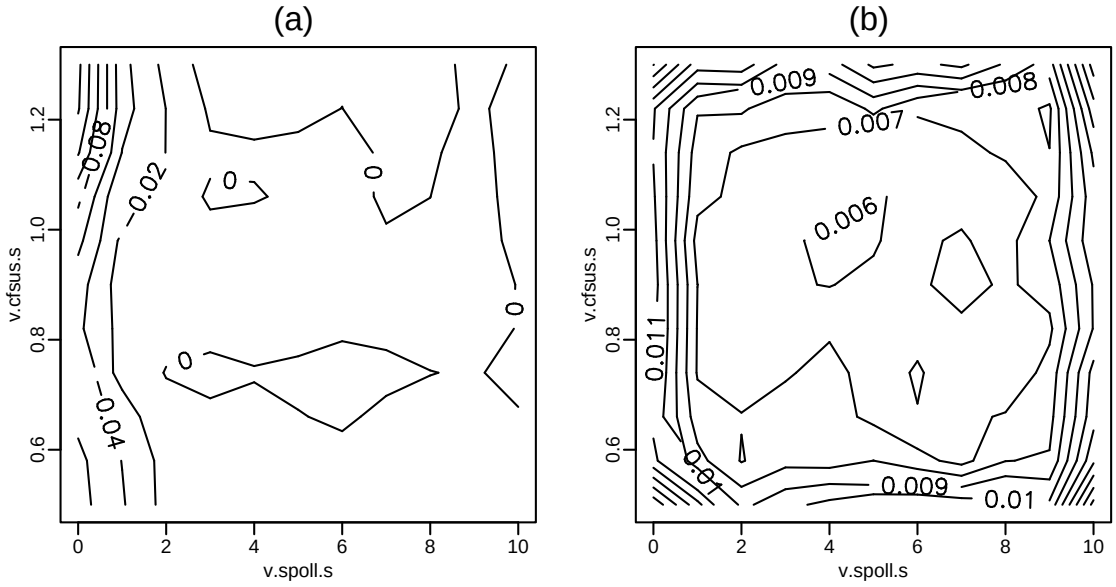

Fig. 4. Joint effect of sustainable pollution in the south (v.spoll.s) and change in sustainable pollution in the south (v.cfsus.s) on HDI in the Wonderland model: (a) estimated effect; (b) pointwise standard error of the estimated effect.

The two regression models have lower prediction accuracy than the random-function model when assessed using "cross-validated root mean squared error of prediction". This quantity is simply the root mean of the squared cross-validated residuals $y(\boldsymbol{x}^{(i)}) - \hat{Y}_{-i}(\boldsymbol{x}^{(i)})$, for $i = 1, \ldots, n$ in the numerator of (20). The cross-validated root mean squared error values are 0.040 and 0.035 for the first-order and second-order regression models, respectively, compared with 0.026 for the random-function model. Figure 5 shows that both regression models are particularly poor at predicting extremely low values of HDI. The true relative importances of the effects of the input variables are not known for the Wonderland model, but it is arguable that the screening results from the random-function model are more credible because of the better prediction performance of this model.

7 Discussion

The Wonderland model illustrates that, at least in some applications, very complex effects involving highly nonlinear, interactive relationships, can exist. Naturally, these are difficult to model and identify. The approach that we have described starts with a random-function model that is data-adaptive

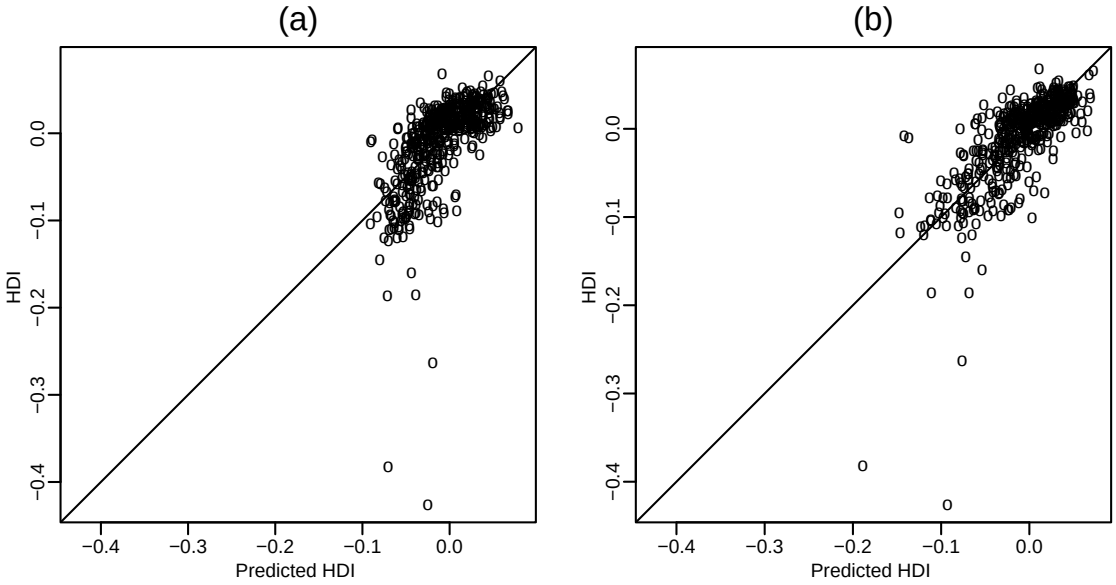

Fig. 5. Actual human development index (HDI) values versus their cross-validation predictions from regressions models: (a) selected from all first-order terms; (b) selected from all second-order terms.

to such complexities, given enough computer-model runs. Similarly, the estimated effects derived from the random-function approximator can be fairly complex if demanded by the data. To detect such subtle effects, the experimental design has to allow exploration of the input region densely, at least for a few variables at a time. "Space-filling" designs such as Latin hypercubes, used in this chapter, and orthogonal arrays (see Chapter **??**) have good projective properties and are desirable in this respect (see Koehler and Owen, 1996, for a review of designs for computer experiments). The design does not have to be balanced in any sense. The ANOVA decomposition is of the approximator (that is, the predictor from the surrogate model); it is not a traditional analysis of variance computed directly from the data.

In the Wonderland model, a pathological scenario was identified of very low values of the human development index, dependent on extreme values of two of the 41 variables. In the experiment with 500 runs, only three runs exhibited this behavior; fewer runs in the design or a less flexible approximation strategy may well have missed this feature.

In practice, one is often faced with choosing a model that is easily interpretable but may not approximate a response very well, such as a low-order polynomial regression, or with choosing a black box model, such as the random-function model in equations (1)–(3). Our approach makes this black box model interpretable in two ways: (a) the ANOVA decomposition provides

a quantitative screening of the low-order effects, and (b) the important effects can be visualized. By comparison, in a low-order polynomial regression model, the relationship between input variables and an output variable is more direct. Unfortunately, as we have seen, the complexities of a computer code may be too subtle for such simple approximating models.

Throughout, we have used "plug-in" estimates of the correlation parameters in (3). These parameters are estimated by maximum likelihood but the estimates are treated as the true values thereafter. The uncertainty from estimating the correlation parameters is not propagated through to the standard errors of estimated effects. In principle, though, this is easily overcome with a Bayesian prior distribution on the correlation parameters. We could: (1) sample say 10–100 sets of values of the correlation parameters from their Bayesian posterior distribution (see, for example, Robert and Casella, 2004); (2) estimate effects using the methods in this chapter, conditional on each set of values of the correlation parameters; and (3) combine the analyses using standard probability results to compute a standard error taking account of parameter uncertainty. As the analysis in this chapter is relatively straightforward computationally, repeating it 10-100 times would not be onerous; rather, the difficulty would be with sampling from the Bayesian posterior for applications with many input variables.

Appendix A. Derivation of the best linear unbiased predictor of an effect

The the best linear unbiased predictor (BLUP) of $\bar{Y}_e(\boldsymbol{x}_e)$ in (18) follows from the properties of $\bar{Z}_e(\boldsymbol{x}_e)$ in (17). Clearly, $\bar{Z}_e(\boldsymbol{x}_e)$, like $Z(\boldsymbol{x})$, has expectation zero. Its variance, however, is not constant:

$$\begin{aligned}\mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)] &= \int_{\otimes_{j\notin e}\mathcal{X}_j} \int_{\otimes_{j\notin e}\mathcal{X}_j} \mathrm{Cov}[Z(\boldsymbol{x}_e,\boldsymbol{x}_{-e}), Z(\boldsymbol{x}_e,\tilde{\boldsymbol{x}}_{-e})] \prod_{j\notin e} w_j(x_j)w_j(\tilde{x}_j)d\,x_j d\,\tilde{x}_j \\ &= \sigma^2 \int_{\otimes_{j\notin e}\mathcal{X}_j} \int_{\otimes_{j\notin e}\mathcal{X}_j} R[(\boldsymbol{x}_e,\boldsymbol{x}_{-e}),(\boldsymbol{x}_e,\tilde{\boldsymbol{x}}_{-e})] \prod_{j\notin e} w_j(x_j)w_j(\tilde{x}_j)d\,x_j d\,\tilde{x}_j. \end{aligned} \tag{21}$$

The steps for deriving the BLUP of $\bar{Y}_e(\boldsymbol{x}_e)$ with a standard error closely follow those in Section 2 for the BLUP of $Y(\boldsymbol{x})$. Again, consider predictors that are linear in the n observed output values, $\hat{\bar{Y}}_e = \boldsymbol{a}_e'(\boldsymbol{x}_e)\boldsymbol{y}$. From the random-function model (15), the mean squared error of $\hat{\bar{Y}}_e(\boldsymbol{x}_e)$ is

$$\mathrm{E}[\bar{Y}_e(\boldsymbol{x}_e) - \hat{\bar{Y}}_e(\boldsymbol{x}_e)]^2 = \mathrm{E}[\bar{\boldsymbol{f}}_e'(\boldsymbol{x}_e)\boldsymbol{\beta} + \bar{Z}_e(\boldsymbol{x}_e) - \boldsymbol{a}_e'(\boldsymbol{x}_e)(\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{z})]^2$$

$$= \{[\bar{\boldsymbol{f}}'_e(\boldsymbol{x}_e) - \boldsymbol{a}'_e(\boldsymbol{x}_e)\boldsymbol{F}]\boldsymbol{\beta}\}^2 + \mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)] + \boldsymbol{a}'_e(\boldsymbol{x}_e)\mathrm{Cov}(\boldsymbol{z})\boldsymbol{a}_e(\boldsymbol{x}_e) - 2\boldsymbol{a}'_e(\boldsymbol{x}_e)\mathrm{Cov}[\bar{Z}_e(\boldsymbol{x}_e), \boldsymbol{z}].$$

Element i of the $n \times 1$ vector $\mathrm{Cov}[\bar{Z}_e(\boldsymbol{x}_e), \boldsymbol{z}]$ is computed from

$$\mathrm{Cov}[\bar{Z}_e(\boldsymbol{x}_e), Z(\boldsymbol{x}^{(i)})] = \int_{\otimes_{j\notin e}\mathcal{X}_j} \mathrm{Cov}[Z(\boldsymbol{x}_e, \boldsymbol{x}_{-e}), Z(\boldsymbol{x}_e^{(i)}, \boldsymbol{x}_{-e}^{(i)})] \prod_{j\notin e} w_j(x_j)d\,x_j$$
$$= \sigma^2 \int_{\otimes_{j\notin e}\mathcal{X}_j} R[(\boldsymbol{x}_e, \boldsymbol{x}_{-e}), (\boldsymbol{x}_e^{(i)}, \boldsymbol{x}_{-e}^{(i)})] \prod_{j\notin e} w_j(x_j)d\,x_j . \quad (22)$$

Thus, we have to integrate out the variables not in $\boldsymbol{x}_e$ from the correlation function. We will write $\sigma^2\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e)$ for $\mathrm{Cov}[\bar{Z}_e(\boldsymbol{x}_e), \boldsymbol{z}]$.

Again imposing a constraint to eliminate the contribution to the mean squared error from the term involving $\boldsymbol{\beta}$, the optimal choice of $\boldsymbol{a}'_e(\boldsymbol{x}_e)$ is formulated as:

$$\min_{\boldsymbol{a}_e(\boldsymbol{x}_e)} \mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)] + \sigma^2\boldsymbol{a}'_e(\boldsymbol{x}_e)\boldsymbol{R}\boldsymbol{a}_e(\boldsymbol{x}_e) - 2\sigma^2\boldsymbol{a}'_e(\boldsymbol{x}_e)\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e) \quad (23)$$

subject to

$$\boldsymbol{F}\boldsymbol{a}_e(\boldsymbol{x}_e) = \bar{\boldsymbol{f}}_e(\boldsymbol{x}_e).$$

The constrained optimization problems leading to the BLUPs of $Y(\boldsymbol{x})$ and $\bar{Y}_e(\boldsymbol{x}_e)$ are very similar: $\mathrm{Var}[Z(\boldsymbol{x})]$, $\boldsymbol{r}(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x})$ in (5) have simply been replaced by $\mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)]$, $\bar{\boldsymbol{r}}_e(\boldsymbol{x}_e)$ and $\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e)$, respectively, in (23).

Appendix B. Computation of the integrals required for the estimated effects and the ANOVA decomposition

To compute the BLUP (18) of a marginal effect we need the vectors $\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e)$ in (16) and $\bar{\boldsymbol{r}}_e(\boldsymbol{x})$ following (22), both of which involve integration over all the variables not in e. For computational convenience in performing these integrations, we will need two further "product-structure" conditions, in addition to (8) and (9). They relate to the properties of the random-function model, specifically the regression functions, $\boldsymbol{f}(\boldsymbol{x})$, in (1) and the correlation function, $R(\boldsymbol{x}, \tilde{\boldsymbol{x}})$, in (2).

First, we assume each regression function is a product of functions in just one input variable, that is, element k of $\boldsymbol{f}(\boldsymbol{x})$ can be written

$$f_k(\boldsymbol{x}) = \prod_{j=1}^{d} f_{kj}(x_j) \qquad (k = 1, \ldots, h). \quad (24)$$

Fortunately, the polynomial regression models commonly used are made up of functions $f_k(\boldsymbol{x})$ which are products of powers of single variables. With (24), element k of $\bar{\boldsymbol{f}}_e(\boldsymbol{x}_e)$ in (16) is

$$\int\limits_{\otimes_{j\notin e}\mathcal{X}_j} f_k(\boldsymbol{x}_e, \boldsymbol{x}_{-e}) \prod_{j\notin e} w_j(x_j)d\,x_j = \prod_{j\in e} f_{kj}(x_j) \int\limits_{\otimes_{j\notin e}\mathcal{X}_j} \prod_{j\notin e} f_{kj}(x_j)w_j(x_j)d\,x_j.$$

The integral on the right-hand side of this equation is clearly a product of one-dimensional integrals,

$$\int\limits_{\mathcal{X}_j} f_{kj}(x_j)w_j(x_j)d\,x_j,$$

which can be evaluated using simple techniques such as Simpson's rule.

Secondly, we assume similarly that the correlation function is a product of one-dimensional correlation functions, that is,

$$R(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \prod_{j=1}^{d} R_j(x_j, \tilde{x}_j). \tag{25}$$

The power-exponential correlation function (3), for example, is of this product form. To compute $\bar{\boldsymbol{r}}_e(\boldsymbol{x})$, the integral on the right-hand side of (22) is evaluated as

$$\prod_{j\in e} R_j(x_j, x_j^{(i)}) \int\limits_{\otimes_{j\notin e}\mathcal{X}_j} \prod_{j\notin e} R_j(x_j, x_j^{(i)})w_j(x_j)d\,x_j,$$

and the integral involved is a product of one-dimensional integrals,

$$\int\limits_{\mathcal{X}_j} R_j(x_j, x_j^{(i)})w_j(x_j)d\,x_j.$$

For the standard error (19), we also need $\mathrm{Var}[\bar{Z}_e(\boldsymbol{x}_e)]$ in (21). With condition (25), the double integral on the right-hand side of (21) is computed as:

$$\prod_{j\in e} R_j(x_j, x_j) \prod_{j\notin e} \int\limits_{\mathcal{X}_j}\int\limits_{\mathcal{X}_j} R_j(x_j, \tilde{x}_j)w_j(x_j)w_j(\tilde{x}_j)d\,x_j d\,\tilde{x}_j.$$

Thus, two-dimensional numerical quadrature is sufficient. Further simplification follows by noting that the correlation function should satisfy $R_j(x_j, x_j) = 1$ when modeling a continuous function.

To visualize the estimated effect $\hat{\bar{Y}}_e(\boldsymbol{x}_e)$ and its standard error, $\mathrm{se}[\hat{\bar{Y}}_e(\boldsymbol{x}_e)]$, these quantities are computed for a grid of values of $\boldsymbol{x}_e$. The required one- and two-dimensional integrals depend only on the variables *not* in $\boldsymbol{x}_e$ and need be computed only once for all grid points.

From the estimated marginal effects, it is straightforward to compute estimates of the corrected main effects (12), the two-variable interaction effects (13), and so on. The ANOVA contributions on the right-hand side of (14) for these low-order effects involve correspondingly low-dimension integrals.

Acknowledgments

We are indebted to Joe Hendrickson for supplying the data for the Wonderland model and answering our numerous questions about the application. Schonlau's research was supported by core funding from the RAND corporation. Welch's research was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, pages 267–281. Akademia Kiadó, Budapest.

Aslett, R., Buck, R.J., Duvall, S.G., Sacks, J. and Welch, W.J. (1998). Circuit optimization via sequential computer experiments: design of an output buffer. *Journal of the Royal Statistical Society, C*, **47**, 31–48.

Bernardo, M.C., Buck, R., Liu, L., Nazaret, W.A., Sacks, J. and Welch, W.J. (1992). Integrated circuit design optimization using a sequential strategy. *IEEE Transactions on Computer-Aided Design*, **11**, 361–372.

Bettonvil, B. and Kleijnen, J.P.C. (1996). Searching for important factors in simulation models with many factors: sequential bifurcation. *European Journal of Operational Research*, **96**, 180–194.

Chapman, W.L., Welch, W.J., Bowman, K.P., Sacks, J. and Walsh, J.E. (1994). Arctic sea ice variability: model sensitivities and a multidecadal simulation. *Journal of Geophysical Research C* , **99**, 919–935.

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

Gough, W.A. and Welch, W.J. (1994). Parameter space exploration of an ocean general circulation model using an isopycnal mixing parameterization. *Journal of Marine Research*, **52**, 773–796.

Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA with componentwise Bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, **2**, 97–117.

Herbert, R.D. and Leeves, G.D. (1998). Troubles in Wonderland. *Complexity International*, **6**.
www.complexity.org.au/ci/vol06/herbert/herbert.html

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, **19**, 293–325.

Iman, R.L. and Conover, W.J. (1980). Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics A - Theory and Methods*, **9**, 1749–1842.

Jones, D.R., Schonlau, M. and Welch, W.J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, **13**, 455–492.

Koehler, J.R. and Owen, A.B. (1996). Computer experiments. In *Handbook of Statistics*, Volume 13. Editors: S. Ghosh and C.R. Rao. Elsevier, Amsterdam.

Lempert, R.J., Popper, S.W. and Bankes, S.C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-term Policy Analysis*. RAND, Santa Monica, California.
www.rand.org/publications/MR/MR1626

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

Morris, M.D. (1991). Factorial sampling plans for preliminary computer experiments. *Technometrics*, **33**, 161–174.

Mrawira, D., Welch, W.J., Schonlau, M. and Haas, R. (1999). Sensitivity analysis of computer models: World Bank HDM-III model. *Journal of Transportation Engineering*, **125**, 421–428.

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
www.R-project.org

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, second edition. Springer, New York.

Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989). Design and analysis of computer experiments (with discussion). *Statistical Science*, **4**, 409–435.

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

Schonlau, M. (1997). *Computer Experiments and Global Optimization.* PhD thesis, University of Waterloo, Waterloo, Ontario.

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.