An efficient screening method for computer experiments

Alexis Boukouvalas
Aston University, United Kingdom
and
John Paul Gosling
University of Leeds, United Kingdom
and
Hugo Maruri-Aguilar
Queen Mary, University of London, United Kingdom

November 4, 2013

Abstract

Computer simulators of real world processes are often computationally expensive and require many inputs. The problem of the computational expense can be handled using emulation technology; however, highly-multidimensional input spaces may require more simulator runs to train and validate the emulator. We aim to reduce the dimensionality of the problem by screening the simulator's inputs for non-linear effects on the output rather than distinguishing between negligible and active effects. Our proposed method is built upon the elementary effects method for screening (Morris, 1991) and utilises a threshold value to separate the inputs with linear and non-linear effects. The technique is simple to implement and acts in a sequential way in order to keep the number of simulator runs down to a minimum, whilst identifying the inputs that have non-linear effects. The algorithm is applied on a set of simulated examples and a rabies disease simulator where we observe run savings ranging between 28% and 63% compared to the batch elementary effects method. Supplementary materials for this paper are available online.

Keywords: Morris design, Sensitivity analysis, Variable selection.

1 Analysis of complex computer simulators

Formal design and analysis of computer experiments stems from the seminal paper by Sacks et al. (1989). In recent years, computer experiments have been used to make climate projections (Hargreaves et al., 2004), investigate biological processes (Wedge et al., 2009) and estimate national carbon balances (Kennedy et al., 2008). In many applications, emulators are constructed as efficient surrogate alternatives to expensive computer simulators (O'Hagan, 2006). Further analyses are then carried out on the emulator; for example, sensitivity analyses or prediction of response at new points. However, even when using the emulator technology, the analysis of complex simulators can be restricted by the number of input dimensions. In the present paper, we introduce a sequential screening procedure to identify inputs that have a non-linear effect on the output. Our proposal stems from the elementary effects (EE) method introduced in Morris (1991), and consists of sequential estimation coupled with a space-filling criterion. Estimation of model behaviour is a common practice in sensitivity analysis (Morris, 1991) and use of space filling design ensures reasonable coverage of the input space (Campolongo and Braddock, 1999). There exist a wide range of methods in the literature to estimate functional effects. Screening methods allow for efficient estimation of effects using a minimal number of simulator evaluations. Variance-based sensitivity analyses methods (Saltelli et al., 2000) offer more precise results in terms of the percentage of variance explained by each input factor and their interactions. However, these methods require many more model runs and are employed at a later stage of simulator analysis than screening algorithms.

Our aim in this paper is to develop a method that identifies inputs with non-linear effects on the output. Morris (1991) considers screening for factors with linear, non-linear or interacting effects, defined as active. Identification of inputs with active effects is the usual aim of screening, also known as variable selection. Our focus is to distinguish factors with non-linear or interaction effects from factors with linear or negligible effects. It is of interest to identify the latter since they can be dealt with more simply in subsequent stages of simulator analyses as discussed in Section 4. Screening methods are further discussed in Section 2. We review the EE method in Section 3 and introduce our sequential method and its properties in Section 4. We compare the method to other popular screening methods

through a simulated example in Section 5 and finish by demonstrating our method on a computer model of rabies disease spread in Section 6.

2 Screening approaches

In classical experimentation on physical processes, screening usually incorporates some of factorial designs, fractional factorial designs, nonregular orthogonal arrays, or optimal designs. However, without further model assumptions, sequential use of such designs is not straightforward and cannot be utilised to identify a nonlinear effect. When the number of inputs dimensions is very large, but it is expected a priori that only a small subset of them will have an effect on the output, the iterated fractional factorial design (IFFD) of Andres (1997) has been shown to be highly efficient in practice. The IFFD method is most efficient when there are very few highly influential factors (Saltelli et al., 2000). The IFFD is an example of a supersaturated design. These designs use fewer model runs than input dimensions, but typically make assumptions on the number of active inputs or the type of effects on the response (Saltelli et al., 2000, Section 4.8). Kleijnen (2009) proposes sequential bifurcation that can be effective, provided monotonicity of the model output with respect to the inputs is satisfied.

If the simulator is computationally cheap, traditional sensitivity methods can be used such as the Fourier amplitude sensitivity test (Cukier et al., 1973) or the method of Sobol' (Sobol, 1993). These approaches utilise a simulator-based functional ANOVA. The robust application of such methods requires typically many more runs of the simulator than most screening methods.

Response surface methods have been proposed for screening in high-dimensional problems where a surrogate model is utilised to approximate the simulator response. In Linkletter et al. (2006) and Savitsky et al. (2011), the surrogate is a Gaussian-process-based model and a prior encapsulates the assumptions of effect sparsity. Reich et al. (2009) proposes a functional ANOVA decomposition where a spline surrogate model is used to estimate each term. When using any approach that relies on a surrogate model, careful validation is needed to ensure the adequacy of the surrogate model for the purposes of screening.

The elementary effects (EE) method (Morris, 1991) is a popular methodology for sensi-

tivity analysis of computer simulators. The EE method requires no simplifying assumptions to be made on the ratio of active to total number of factors or their effect on the response (Saltelli et al., 2000). Furthermore, the method is easy to implement and computationally efficient. In the next section, we describe the EE method in detail, and, in Section 4, we extend the EE method to a more efficient, sequential approach.

3 The elementary effects method

Consider a deterministic simulator $Y(\cdot)$ with k input variables and design region $[0,1]^k$. The simulator is assumed to be a smooth real valued function with a domain containing the design region. Computation of elementary effects starts from a point $\mathbf{x}$ from which a trajectory is constructed with k random moves of size Δ. One-at-a-time (OAT) moves are performed along each single coordinate axis in turn to end at point $\mathbf{x} + \Delta(\mathbf{e}_1 + \cdots + \mathbf{e}_k)$. The elementary effect for the i-th input variable for the trajectory starting at $\mathbf{x} \in [0,1]^k$ is

$$EE_i(\mathbf{x}) = \frac{\mathbf{Y}\left(\mathbf{x} + \boldsymbol{\Delta}\sum_{\mathbf{j=1}}^{\mathbf{i}} \mathbf{e_j}\right) - \mathbf{Y}\left(\mathbf{x} + \boldsymbol{\Delta}\sum_{\mathbf{j=0}}^{\mathbf{i-1}} \mathbf{e_j}\right)}{\boldsymbol{\Delta}}, \tag{1}$$

where $\Delta > 0$ is fixed. Here $i = 1, \ldots, k$ indexes input factors and $\mathbf{e}_i$ is the unit vector in the direction of the i-th axis where $\mathbf{e_0}$ is defined as $\mathbf{0}$. A total of $k+1$ evaluations of $Y(\cdot)$ are performed, ending with effects $EE_1(\mathbf{x}), \ldots, EE_k(\mathbf{x})$. Each $EE_i(\mathbf{x})$ is a measure of the variation in the output with respect to a change in input i at point $\mathbf{x}$.

Consider R starting points $\mathbf{x}_r$, $r = 1, \ldots, R$. From each point $\mathbf{x}_r$, we perform k OAT moves and compute elementary effects $EE_i(\mathbf{x}_r)$ for every input factor so that the total number of runs used in the EE method is $(k+1) \times R$. The following sample moments are computed for each input factor:

$$\mu_i = \frac{1}{R}\sum_{r=1}^{R} EE_i(\mathbf{x}_r),\ \mu_i^* = \frac{1}{R}\sum_{r=1}^{R} |EE_i(\mathbf{x}_r)| \text{ and } \sigma_i = \sqrt{\sum_{r=1}^{R} \frac{(EE_i(\mathbf{x}_r) - \mu_i)^2}{R-1}}. \tag{2}$$

The moment μ_i is an average effect measure, and high values suggest dominant contribution of the i-th input factor. The moment μ_i^* is a main effect measure, proposed in Campolongo et al. (2004) since μ_i may prove misleading due to cancellation of effects. An input with

a negligible or linear effect will have constant EE_i values and the corresponding σ_i will be zero. Non-linear and interaction effects are identified with relatively large σ_i. An effects plot can be constructed by plotting μ_i or μ_i^* against σ_i. This plot is a visual tool to detect and rank effects. Factor effects close to the origin on this plot are the least influential.

There is interest in doing input screening with as few runs as possible and as the number of input factors k is fixed, the size of the experiment is controlled by R. Usually small values of R are used; for instance, Morris (1991) used $R = 3$ and $R = 4$ in his examples. A value of R between 10 and 50 is mentioned in recent literature (Campolongo et al., 2004, 2007). A larger value of R will improve the quality of the estimations, but at the price of extra simulator runs.

The step size Δ is selected in such a way that all the simulator runs lie in the input space and the elementary effects are computed within reasonable precision. The choice of Δ in the literature is determined by the input space considered for experimentation, which is a k dimensional grid constructed with p uniformly spaced values for each input. The number p is recommended to be even and $\Delta = p/(2(p-1))$, ensuring the elementary effects for each input have an equal probability of selection during the trajectory design generation (Morris, 1991). To ensure that the trajectories remain in the design region, the implementation of the EE method will use $-\Delta$ as necessary in place of Δ. The step Δ is usually kept at the same value for all the inputs, but the method can be generalised to instead use different values of Δ and p for every input. In the original proposal Morris (1991), the points $\mathbf{x}_1, \ldots, \mathbf{x}_R$ were taken at random from the input space grid. Campolongo et al. (2007) proposed spreading runs over the design space by generating a large number of trajectory designs and selecting a subset by maximising the minimum distance between them.

A potential drawback of OAT designs used in the EE method is that design points may fall on top of each other when projected into lower dimensions. This disadvantage becomes more apparent when the design runs are to be used in further modelling after discarding unimportant factors. An alternative is to screen with randomly rotated simplices located at points x_r, see Pujol (2009). The computation of distribution moments μ_i, μ_i^*, σ_i and further analysis is similar to the EE method, with the advantage that projections of the resulting design do not fall on top of existing points, and all observations can be reused in a later stage. A disadvantage of this approach is the loss of efficiency in the computation

of elementary effects; that is, computing effects from a rotated simplex is suboptimal when compared with (1) which is optimal for computing elementary effects, see Pujol (2009).

4 Sequential elementary effects method

We propose a sequential screening method that has the potential to reduce computation time significantly. The methodology aims to distinguish between factors with linear or no effect and factors with non-linear effect. In the EE method, all $(k+1)$ evaluations are performed for each trajectory irrespective of the type of factor effect. Whereas, in our sequential strategy, when effects are identified as non-linear, their corresponding one-step evaluations are removed from subsequent trajectories. This results in fewer model runs. The rationale is that if σ_i is small for a given factor, then we should investigate whether σ_i remains small when adding extra trajectories. At the end of experimentation, those input factors for which σ_i remained small are considered to have linear or no effect, and factors for which σ_i was bigger than a threshold have a non-linear effect on the output. A method of eliciting the choice of threshold is presented in Section 4.1.

The justification of thresholding solely on the variance of the elementary effects σ_i is that independent linear effects of factors may be removed from the simulator output at a preprocessing stage or during the emulation phase. As an example of the latter approach, factors with linear effects may be incorporated in the mean function of a Gaussian Process (GP) emulator while omitted from the covariance specification. If we denote by A the subset of $\{1,\ldots,k\}$ that indexes factors with linear effects, the GP prior may be written as $Y(x) = \beta + \sum_{i=1}^{k} a_i X_i + Z^*$ where Z^* is a stochastic process whose covariance structure depends only on the variables with non-linear effects; that is, the x_i with $i \in \{1,\ldots,k\} \setminus A$. The residual process Z^* is therefore placed in a lower dimensional space simplifying the design and inference tasks.

For our algorithm to run, a space filling design with M points is created. This design provides the sequence of points at which the Morris OAT runs will be tested. Initially, we select a good space filling design, such as a maximin Latin hypercube (LH) (Morris and Mitchell, 1995). The generation of maximin designs is discussed further in (Ravi et al., 1994). The value of M is selected such that $(k+1)M$ is the maximum number of runs that

can be performed during the whole screening process.

A preprocessing stage orders the design points according to the biggest distance between points. The first two points are those whose Euclidean distance is largest; then the third point maximises the minimum distance between itself and the first two points, then a fourth point is ordered in the same way, and so on. This procedure of ordering points mirrors nearest neighbour clustering, but acts in an opposite manner as points are ordered from those farthest apart to those that are closest.

Example 1. For $k = 5$ input factors and $M = 6$ runs, consider a maximin LH design in $[0, 1]^5$ with point coordinates $\mathbf{x}_1 = (0, 4, 1, 4, 4)/5$, $\mathbf{x}_2 = (1, 5, 5, 1, 3)/5$, $\mathbf{x}_3 = (2, 1, 3, 0, 1)/5$, $\mathbf{x}_4 = (3, 0, 4, 5, 2)/5$, $\mathbf{x}_5 = (4, 3, 0, 3, 0)/5$ and lastly $\mathbf{x}_6 = (5, 2, 2, 2, 5)/5$. The preprocessing stage first selects the points $\mathbf{x}_2$ and $\mathbf{x}_5$, which are furthest apart. The next point, $\mathbf{x}_4$, maximises the distance between those remaining points and the first two points chosen. The procedure continues by selecting $\mathbf{x}_6$, then $\mathbf{x}_1$ and finishes with $\mathbf{x}_3$. This preprocessing stage produces the ordered sequence of points $\mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_1, \mathbf{x}_3$, which are relabelled as $\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(6)}$.

The screening algorithm starts with the computation of elementary effects for all input factors at the first two design points of the ordered maximin LH design. OAT runs are created at those two points and elementary effects are computed. With this initial data, a first estimation of the moments μ_i, μ_i^* and σ_i is available. If, for a given input factor, its sample moment σ_i is larger than a specified threshold σ_0, then we say that the output is responding non-linearly to this input. This leads us to declare that input as active and remove it from the list of current input factors. The technique continues by adding OAT runs at the next point, only for factors suspected to be linear. Elementary effects are computed and moments are updated for each added point, removing a factor if the condition for σ_i is met. The methodology ends when all input factors have been removed or after computing elementary effects for all M points. On ending, the input factors are separated into two groups: those having non-linear effect and those with linear or no effect on the output. Algorithm 1 sets out the procedure in pseudo-code form.

Example 2. To show how the proposed sequential algorithm works, consider $Y(x_1, x_2, x_3, x_4, x_5) = cos(x_3/5)(x_2 + 1/2)^4/(x_1 + 1/2)^2 + x_5$ on the design region $[0, 1]^5$. Note that x_4 has negligible

Algorithm 1 The procedure for completing our screening technique.

Screening algorithm

Input: Simulator $Y(\cdot)$ with k inputs; total number of one-at-a-time experiments M; step size Δ; threshold σ_0.
Output: Moments μ_i, σ_i, μ_i^*; lists of factors with linear ($\mathcal{C}$) and with non-linear effect ($\mathcal{A}$).

A. *Preprocessing stage*

1. Set design region to $[0,1]^k$ and create space filling design with M points $\mathbf{x}_1, \ldots, \mathbf{x}_M$.
2. Order the design points using maximum distance between points. Label the ordered points as $\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(M)}$.

B. *Calculating the elementary effects*

1. Set $R := 2$ and the initial design to be $D := \left\{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}\right\}$. Set list of current factors to $\mathcal{C} := \{1, \ldots, k\}$ and list of active effects $\mathcal{A} := \emptyset$.
2. For every point in D, create one-at-a-time runs only for those input factors indexed by $\mathcal{C}$. Run the simulator at those points. This totals $|\mathcal{C}| + 1$ experiments for every point in D.
3. Using simulator runs from step B2 and (1), compute elementary effects $\{EE_i(\mathbf{x}) : \mathbf{x} \in D, i \in \mathcal{C}\}$.
4. If $R = 2$, compute moments μ_i, μ_i^* and σ_i using elementary effects for all factors. If $R > 2$, only update moments for the current list of input factors, indexed by $\mathcal{C}$.
5. For $i \in \mathcal{C}$, if $\sigma_i > \sigma_0$ then update $\mathcal{C} := \mathcal{C} \setminus \{i\}$ and $\mathcal{A} := \mathcal{A} \cup \{i\}$.
6. If $\mathcal{C} = \emptyset$, then all the inputs were identified as non-linear. Algorithm ends.
7. If $R = M$, then all the design points available are exhausted. Algorithm ends.

C. *Producing the next design point*

1. Update R:=R+1; set $D = \{\mathbf{x}_{(R)}\}$.
2. Goto B2.

effect here. The function Y is treated as a simulator, from which the only information we require are its values at design points. We use the same pre-ordered LH design of Example 1; set $p = 10$ for step size $\Delta = 5/9$ and threshold $\sigma_0 = 0.15$. See 4.1 for details on the construction of the threshold σ_0. Random trajectories are constructed with the first two ordered points, giving the following moment estimates $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5) = (-7.87, 4.66, -0.047, 0, 1)$ and $(\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5) = (6.62, 0.27, 0.06, 0, 0)$. Note that the method yields exact zeroes for the fourth variable. The estimates σ_1, σ_2 are greater than the threshold σ_0 and thus x_1 and x_2 are separated as having non-linear effects. As all σ_3, σ_4 and σ_5 are smaller than σ_0, further investigation is required for x_3, x_4 and x_5. At the third design point, OAT experimentation only for those factors produces updated moments $(\mu_3, \mu_4, \mu_5) = (-0.06, 0, 1)$ and $(\sigma_3, \sigma_4, \sigma_5) = (0.05, 0, 0)$, i.e. the factors remain under investigation. The sequential methodology continues for all of them until finishing with all the design points. At this final step, updated moments are $(\mu_3, \mu_4, \mu_5) = (-0.039, 0, 1)$ and $(\sigma_3, \sigma_4, \sigma_5) = (0.04, 0, 0)$, where again for the fourth factor the results are exactly zero. We conclude that linearity of the response in terms of x_3, x_4 and x_5 over the design region could not be rejected. Near zero variances for x_4 and x_5 suggest strong conclusions for those factors.

The total experimental effort was 28 runs, from which the first 12 runs involved trajectories for all factors, while further 16 runs were required for the linear factors under investigation. This is a 22% reduction from the $(5 + 1) * 6 = 36$ runs needed to perform the batch EE method. The moment estimates obtained for the non-linear factors are only rough approximations of the true moment values, and the moment estimates for the linear factor were computed with more information. This asymmetry is apparent when comparing with exact analytic sensitivity results $\mu = (-5.34, 6.62, -0.04, 0, 1)$ and $\sigma = (8.88, 7.42, 0.06, 0, 0)$, where we use $\mu_i = \int_{[0,1]^5} \frac{\partial}{\partial x_i} Y(\mathbf{x}) d\mathbf{x}$ and $\sigma_i^2 = \int_{[0,1]^5} \left(\mu_i - \frac{\partial}{\partial x_i} Y(\mathbf{x}) \right)^2 d\mathbf{x}$.

4.1 Heuristic selection of variance threshold

In Algorithm 1, the elementary effect variance threshold σ_0 is an input. In this section, we present a heuristic to choose σ_0 indirectly by eliciting the expected divergence from linearity of the factor effect as the variance of an auxiliary random variable. A linear (or near-linear)

effect of the variable x_i is represented by an additive noise model:

$$Y(x_i) = ax_i + b + \varepsilon_i\,, \tag{3}$$

where ε_i is an independent and identically distributed normal random variable with zero mean and known variance γ, and a, b are constants. In other words, the marginal effect due to the factor x_i is modelled with a simple regression line. The auxiliary random variable ε_i captures interaction with other factors and non-linearities.

In practice, the variance γ has to be elicited prior to the screening experiment and there are two alternatives that we have considered in the examples in the present paper.

1. We might believe that the factor x_i has a linear effect, but the simulator runs contain a numerical error. In this case, we expect γ to be set to a small value, such as a multiple of machine precision.

2. If we believe that small non-linear effects will not have an appreciable impact on the model output, then γ should be chosen to reflect the level of variation from a straight line to be tolerated.

The elicitation of the variance parameter is in contrast to Kadane et al. (1980) and Garthwaite and Dickey (1988), where full probability distributions are elicited that reflect beliefs about the parameters of the linear model. In the present application, we do not wish to prejudge the behaviour of the model: we want a number specifying how far we can tolerate from being linear.

Given the variance γ, the sampling distribution of the variance of the elementary effects can be calculated according to the following lemma, whose proof is given in Appendix A. For simplicity we omit the subscript i from the quantities σ^2, σ_Φ, γ and Δ, all of which could potentially take different values for each input. In practice we choose a common γ and Δ for all inputs.

Lemma 1. *Let $x_1, \ldots, x_R$ be univariate design points, at each of which trajectories are constructed. Assume that observations taken at design points and trajectories follow the model given in* (3)*. Let elementary effects and moments be defined as in* (1) *and* (2) *and let*

$\sigma_\Phi^2 = \frac{2\gamma}{\Delta^2}$. *Then*

$$\sigma^2 \sim \frac{\sigma_\Phi^2}{R-1}\chi^2_{R-1}. \tag{4}$$

where χ^2_{R-1} *denotes a chi-square random variable with* $R-1$ *degrees of freedom.*

We propose to use the 99% quantile of the cumulative distribution function of the chi-square distribution to derive the EE variance threshold σ_0. We have found this choice of quantile sufficiently conservative for the examples we have investigated. The following equation

$$P\left(\sigma^2 \leq \sigma_0^2\right) = P\left(\frac{\sigma_\Phi^2}{R-1}\chi^2_{R-1} \leq \sigma_0^2\right) = 0.99,$$

when inverted yields the threshold

$$\sigma_0 = \sqrt{\chi^2_{0.99,R-1}\sigma_\Phi^2/(R-1)}, \tag{5}$$

where $\chi^2_{0.99,R-1}$ is the 99% quantile of a chi-squared distribution with $R-1$ degrees of freedom. In other words, σ_0 defines a threshold over which the effect is considered non-linear; that is, if $\sigma_i < \sigma_0$, then the input variable x_i is retained for further testing. Note that Lemma 1 applies directly in a multivariate setting; in which case, the comparison is performed separately for each input variable.

In Example 2, we used a single threshold σ_0 for all variables. In order to obtain $\sigma_0 = 0.385$ using the method described in this section, the values $R = 6$, $\Delta = 5/9$, $\sqrt{\gamma} = 8.7 \times 10^{-2}$ and quantile $\chi^2_{0.99,5} = 15.08$ could be used.

To simplify the algorithm, the threshold σ_0 may be kept fixed for all computations rather than adapting σ_0 to the actual number of trajectories involved. The main difference is in the degrees of freedom for the scaled chi-square distribution in (5). The adaptive approach, which was utilised in the simulation experiments presented, involves recomputing σ_0 with updated degrees of freedom prior to step B5 in Algorithm 1. In the simplified approach using a single value σ_0, the method is more conservative than when varying degrees of freedom; i.e. the rejection rate is higher with fixed threshold than otherwise.

The detection of nonlinearity under the heuristics in this section is further explored in Appendix B. The robustness of the proposed selection of σ_0 in our sequential algorithm is explored under different departures from linear model (3).

5 Simulated high dimensional example

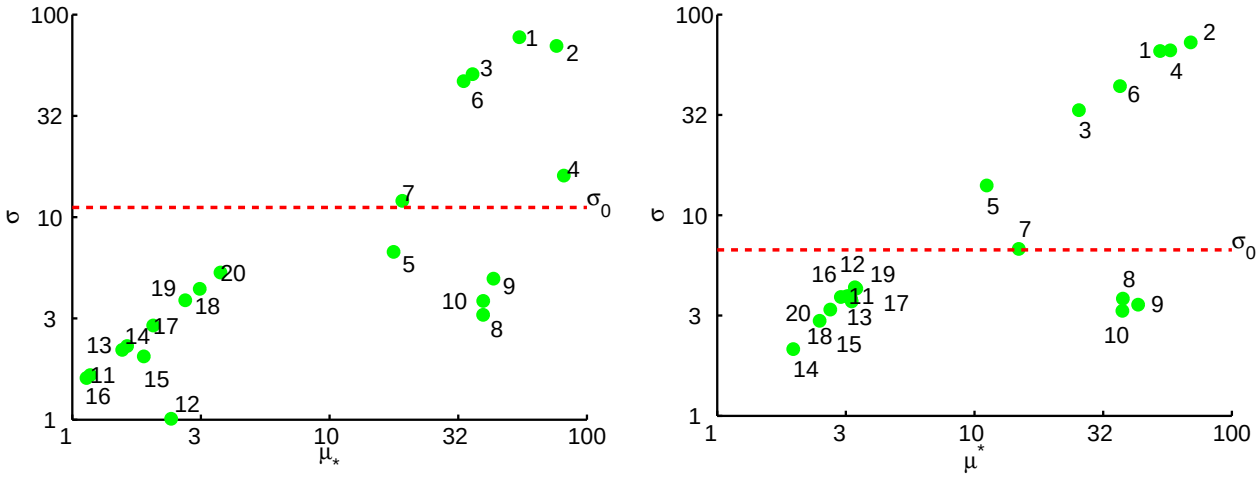
We illustrate the sequential screening method on the synthetic test function introduced in Morris (1991). The function is defined on 20 inputs $\mathbf{x} \in [0,1]^{20}$ as follows:

$$y = \beta_0 + \sum_{i=1}^{20} \beta_i w_i + \sum_{i<j}^{20} \beta_{ij} w_i w_j + \sum_{i<j<l}^{20} \beta_{ijl} w_i w_j w_l + \sum_{i<j<l<s}^{20} \beta_{ijls} w_i w_j w_l w_s, \qquad (6)$$

where $w_i = 2(x_i - \frac{1}{2})$ except for $i = 3, 5, 7$ where $w_i = 2\left(1.1x_i/(x_i + 0.1) - \frac{1}{2}\right)$. The coefficients are set to $\beta_i = 20$ for $i = 1, \ldots, 10$, $\beta_{ij} = -15$ for $i, j = 1, \ldots, 6$, $\beta_{ijl} = -10$ for $i, j, l = 1, \ldots, 5$ and $\beta_{ijls} = 5$ for $i, j, l, s = 1, \ldots, 4$. The remaining first and second order coefficients are generated independently from a zero mean unit variance normal distribution and the remaining third and fourth order coefficients are set to zero. Factors $x_1, \ldots, x_7$ have a non-linear effect on the function output while factors x_8, x_9, x_{10} have a linear effect and factors $x_{11}, \ldots, x_{20}$ have negligible effect (Morris, 1991; Pujol, 2009).

First, we compare the performance of the batch and sequential EE methods in terms of required model runs. The screening experiment was performed under the configuration used in Pujol (2009) for 100 realisations. As in Pujol (2009) the discretisation level has been set to $p = 20$ so $\Delta = 0.5263$ and the number of trajectories to $R = 10$. For the sequential procedure, a threshold value of $\gamma = 2.6$ was used. This corresponds to around 0.005% of the range of the response y in (6), which varies between -225 and 139.

A total of 210 function evaluations are required for the batch EE procedure. For the sequential EE procedure on average 150 function evaluations are required with a standard deviation of 13, that is an average savings of 28% compared to the batch approach. Factors $x_1, \ldots, x_7$ are correctly identified as having non-linear effect in 99% of the realisations because the corresponding σ_i is larger than the threshold σ_0. Factors $x_8, \ldots, x_{10}$ are found to have linear effects in 92% of the realisations because μ_i is large and σ_i is small. Factors $x_{11}, \ldots, x_{20}$ are found to have negligible in all realisations due to small σ_i values. The full batch EE screening results and the first iteration of one realisation of the sequential algorithm are shown in Figure 1. This first iteration is equivalent to running the algorithm up to step B.4 in Algorithm 1. We note that six of the seven factors with non-linear effects are identified at this early stage.



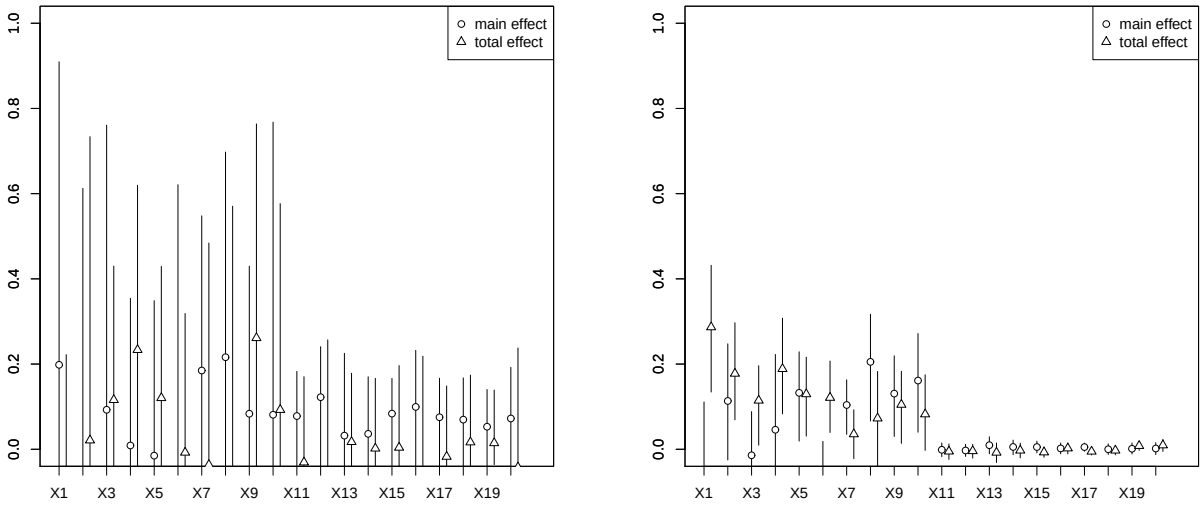
(a) After the first iteration

(b) Average batch results.

Figure 1: Applying the batch and sequential EE screening method on the 20 input factor Morris test function. Horizontal dashed line denotes the σ_0 threshold value for the given iteration.

Finally, we illustrate the efficiency of the EE method by contrasting it to a traditional sensitivity analysis method. Applying the Sobol' sensitivity analysis method (Sobol, 1993) to compute first-order and total indices using a random design of 220 runs, results in large 95% confidence intervals indicating that more model runs are required before any conclusions can be drawn from the examination of the indices (Figure 2). We also show how the uncertainties dramatically reduce when a larger design is used. These results were obtained by using the sensitivity R package (Pujol et al., 2013). This illustration confirms our expectation that screening methods such as the EE method, can be utilised prior to a more detailed sensitivity analysis in order to minimise the number of model runs.

We conclude that the sequential approach results in significant computational savings compared to the batch EE method as factors with clear non-linear effects can be eliminated in the early screening stages with high confidence. Furthermore, we have demonstrated the efficiency of the EE screening methods compared to classical sensitivity analysis as illustrated by the number of runs required to effectively calculate the Sobol's first-order and total indices for this example.



(a) 220 runs

(b) 4378 runs.

Figure 2: Applying Sobol's sensitivity analysis method on data generated using (6). Random designs of 220 and 4378 have been used. The error bars indicate 95% confidence intervals for the first-order and total effect indices.

6 Rabies Model

In this section, we discuss the application of the Morris sequential screening method described in Section 4 on a stochastic model provided by the Food and Environment Research Agency (FERA) (Singer et al., 2008, 2009). An overview of the stochastic simulator is given in Section 6.1, followed by a description of the screening methodology in Section 6.2) and a discussion of the results in Section 6.3.

6.1 Model Description

Although rabies in wild animal populations has been eradicated from large parts of Europe, there is a remaining risk of disease re-introduction. The situation is aggravated by an invasive species, the raccoon dog (*Nyctereutes procyonoides*) that can act as a second rabies vector in addition to the red fox (*Vulpes vulpes*). The purpose of the rabies model is to analyse the risk of rabies spread in this new type of vector community (Singer et al., 2008). The individual-based, non-spatial, discrete-time model incorporates population and disease

dynamical processes such as host reproduction and mortality as well as disease transmission.

The model includes two vector species: raccoon dogs and foxes. The model is non-spatial and disease propagation is calculated solely with respect to population dynamics.

	Description	min	max
1	Number of replications	200	300
2	Fox population winter density (ind./km^2)	0.1	0.5
3	Raccoon dog population winter density (ind./km^2)	0.1	1
4	Shape parameter for the probability distribution of raccoon dog infection	0.39	0.47
5	Dummy variable with no influence	0.9	1.1
6	Fox population mortality	0.9	1.1
7	Raccoon dog population mortality	0.9	1.1
8	Winter hunting proportion	0.9	1.1
9	Fox population birth rate	0.9	1.1
10	Raccoon dog population birth rate	0.9	1.1
11	Fox population infection rate	0.9	1.1
12	Fox population rabies incubation rate	0.95	1.05
13	Raccoon dog rabies incubation rate	0.95	1.05

Table 1: Input parameters for the rabies model and their ranges.

The model has 13 free parameters, shown in Table 1. In addition, three simulator configuration parameters were kept fixed: the maximum number of time steps was set to 400; the cross infection input was 0.002 and the environment area size was kept fixed at 5400 km^2.

The response used is the probability that the rabies disease becomes extinct in both species within 5 years. This output is scaled to lie in the range $[0, 100]$ and is important in deciding on the response to a potential rabies outbreak since it indicates the risk of long term rabies disease persistence (Singer et al., 2008, 2009).

6.2 Screening Methodology

For screening, we utilise a Morris design with the number of trajectories set to $R = 20$ resulting in $(k + 1) \times R = 14 \times 20 = 280$ simulator evaluations, where $k = 13$. The mean μ^* of the absolute values of the Elementary Effects is used to rank the input factors. We set the number of levels to $p = 6$ and $\Delta = p/(2(p - 1)) = 0.6$.

The sequential Morris method discussed in Section 4 allows the specification of a variance γ from which the threshold on the Elementary Effect deviation σ_0 is derived. We set $\gamma = 3.5$,

which reflects a prior belief that individual factor effects on the output are considered near-linear if the effect on the output is within three standard deviations of purely linear, i.e. $\pm 3\sqrt{\gamma} = 5.6$. Since the output is bounded in the range $[0, 100]$ a factor has near-linear effect if the output varies no more than 5.6% from linear. This variability encapsulates both the internal variability of the stochastic model and our prior definition of a near-linear effect.

6.3 Screening Results

Singer and Kennedy (2008) performed sensitivity analysis on this model using the standard Morris method with the same setup as here as well as the Sobol' method. They noted the most important parameters are species winter densities (inputs (2) and (3)) and mortalities (inputs (6) and (7)). They also noted the least influential factors are the dummy variable (5) that has by definition no influence on the model output and parameter 4, a shape parameter for the probability distribution of raccoon dog infection. It is also noted that the Sobol' method is prohibitively expensive and offers low accuracy with a sample size of 300. They suggested increasing the sample size and reducing the dimensionality of the problem by fixing some of the factors to their nominal values. For expensive simulators this motivates the usage of the Morris method.

The standard Morris method variable ranking with $R = 20$ trajectories is presented in Figure 3(a) where the four dominant factors were found to be the winter densities for both species (2,3) and the associated mortality rates (6,7). Singer and Kennedy (2008) have noted that the dominant factors have strong non-linear and interaction effects which is reflected in the high σ value observed in the Morris method.

The sequential Morris method is initialised with $R = 2$ trajectories on all 13 factors requiring $(k + 1)R = 28$ simulator runs. The same step size Δ and the computed threshold σ_0 is the same for all factors. The Morris plot with the associated threshold value is shown in Figure 3(b). The elementary effect variance is over the threshold for four factors, the raccoon dog winter density (3), raccoon dog rabies incubation probability (13), number of replicated runs (1) and raccoon dog population birth rate (10) are eliminated from further consideration since they have strong non-linear effects on the simulator output.

Another trajectory design for the remaining 9 factors is evaluated and requires 10 further simulator evaluations (Figure 3(c)). The parameters fox winter density (2) and mortality

(6) factors are found to have non-linear effects and are removed from further consideration. As evidenced by the Morris plot, the σ value for parameter 2 changed significantly from the previous iteration where the effect was considerably below the threshold and very close to linear.

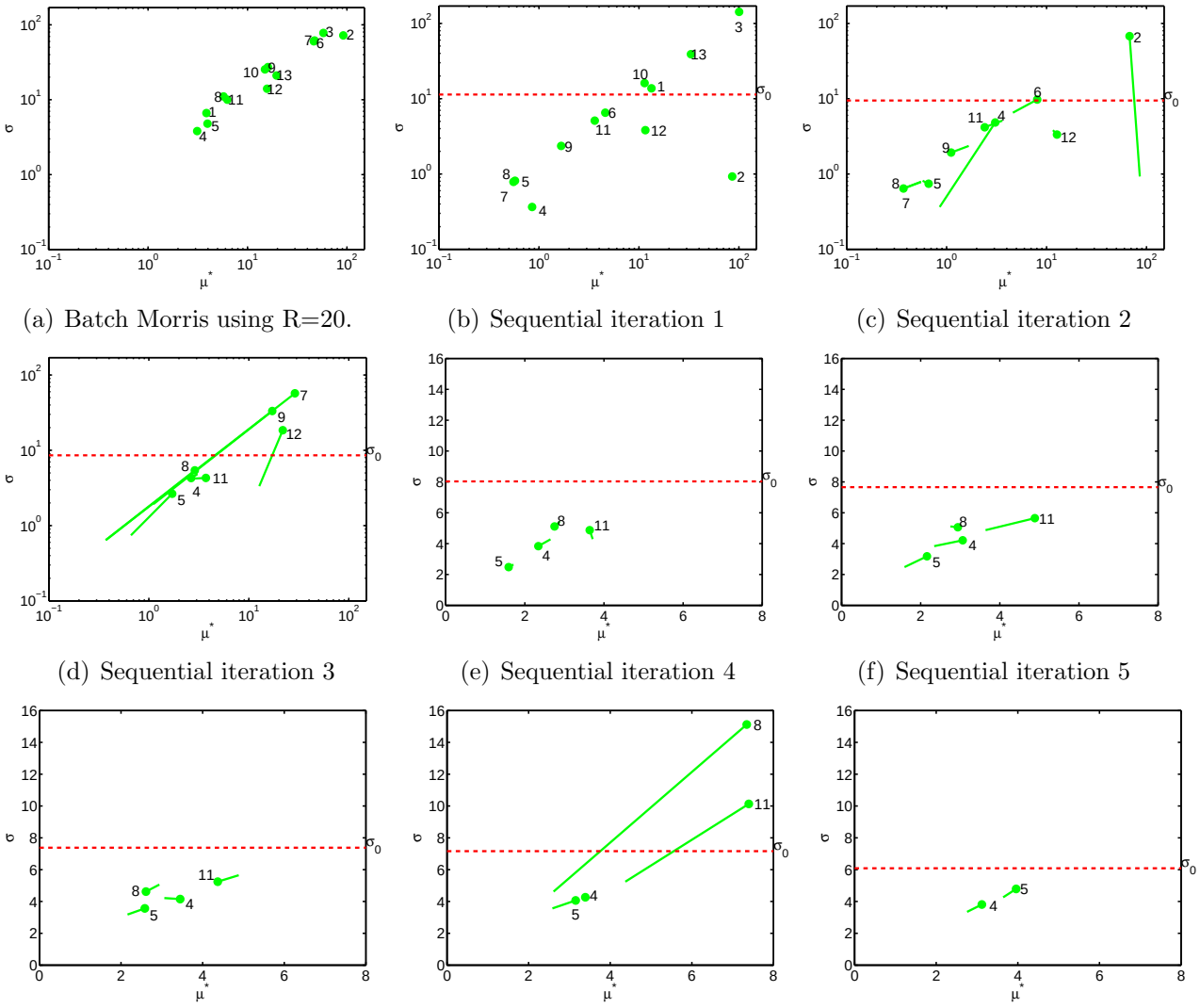
For the third iteration, the seven factor trajectory requires 8 more simulator evaluations (Figure 3(d)). Three further parameters are eliminated, the raccoon dog (7) mortality rate, the fox birth rate (9) and the fox population rabies incubation probability (12) where large changes in the moments of the elementary effects are again observed due to the increased accuracy from the increased Morris design size. No more factors are eliminated until iteration 7 requiring a further $(4 + 1) \times 4 = 20$ simulator evaluations (Figure 3 (e)-(h)). At iteration 7, the winter hunting proportion (8) and fox population infection rate (11) parameters are removed from further consideration.

The remaining two factors, the shape parameter for the probability distribution (4) and the dummy variable (5), are found to be below the σ threshold for all subsequent twelve iterations requiring $(2 + 1)12 = 36$ simulator evaluations. The total number of simulator evaluations for the sequential procedure is 102 compared to the 280 evaluations required by the standard batch Morris method with $R = 20$ trajectories, yielding a 27% reduction in required code runs.

We have also performed the threshold calculation on the full Morris set with $R = 20$ trajectories and the same factors as with the sequential version are identified as near-linear. In summary, the sequential Morris method for the rabies model has been successfully used to identify factors with no or near-linear effects on the simulator response at a significant saving to the standard Morris method.

7 Discussion

We have presented a novel sequential screening method that extends the batch elementary effects method originally proposed by Morris (1991). The method aims to identify inputs with non-linear effects with a minimum number of trajectories. We have empirically demonstrated the computational savings achieved compared to the batch approach on both synthetic data and a real-world simulator. A critical aspect of the method is the specification of the thresh-



(a) Batch Morris using R=20.

(b) Sequential iteration 1

(c) Sequential iteration 2

(d) Sequential iteration 3

(e) Sequential iteration 4

(f) Sequential iteration 5

(g) Sequential iteration 6

(h) Sequential iteration 7

(i) Sequential iteration 19

Figure 3: Batch and Sequential EE screening on the rabies simulator. Solid line denotes path from previous value of (μ^*, σ) for each factor. Horizontal dashed line denotes the σ_0 threshold value for the given iteration.

old σ_0 which is used to determine whether an input has a non-linear effect. The elicitation of this value directly can be challenging and we have presented an indirect approach which utilises an easily interpetable variance value γ specified on the simulator output space.

In order to apply the screening method of the present paper, the analyst must make a number of choices. To create the ordered design of OAT-experiment start-points, M must be specified. We recommend that M is chosen with respect to the computational effort needed to run the simulator; at worst, the simulator will run $(k+1)M$ times. The threshold value σ_0 can be set to zero so that only true linear- and no-effect inputs are investigated. We suggest to use a threshold $\sigma_0 > 0$ as in practice some small non-linearities might be tolerated and replaced instead by a linear function.

In cases where direct elicitation of the EE variance threshold σ_0 is not feasible or straightforward, elicitation of the variance γ may be preferable. The variance γ may be interpreted as the expected divergence of a factor from strict linearity under which the factor effect may still be considered linear for modelling purposes. In the case of stochastic simulators, this parameter includes the internal simulator variability whereas for deterministic simulators near-linear definitions only include errors due to machine precision and degree of departure from truly linear effects on the output. In the future, alternative elicitation methods may be constructed to allow for other non-Gaussian deviations from linear effect. An area of future interest is double classification of factors using the joint distribution of moments μ_i^* and σ_i. This not only would classify a factor as linear or non-linear, but would discriminate between active and non-active factors. A starting point for this would be the t test for μ in Morris (1991).

A variation of our algorithm is to select OAT runs to improve space filling properties of the whole design. The simplest approach is to use random OAT runs. However, OAT runs can be selected to maximise distance between them, as in Campolongo et al. (2007). An adaptation of the OAT runs suggested in Pujol (2009) can also be used where we create a simplex in the current set of input factors. Another criterion is to select OAT runs to improve uniformity of unidimensional projections of the design thus minimising the unidimensional discrepancy of the whole design, conditioned on the initial hypercube. The use of discrepancy to measure uniformity stems from quasi-Monte-Carlo methods (Niederreiter, 1992). We have used Sobol's *low discrepancy* space filling sequence (Niederreiter, 1992). The only

change required in the pseudo-code of Algorithm 1 is to remove step A2. Sampling from low discrepancy sequences has the advantage of sequential generation of points. However, for small sample sizes the spread of points of a low discrepancy sequence may not be as good as that of a maximin space filling design with fixed size.

In this paper, we have not addressed the question of how multiple simulator outputs could be handled in our method. The simplest approach, of generating separate OAT designs for each output, is inefficient. In a sequential setting, the initial design could be shared for all outputs. In our approach, factors are excluded from subsequent screening stages when non-linear effects are detected. Therefore, subsequent stages need only include factors that are under the EE variance threshold across all outputs. As the number of outputs grows, it is more likely that a factor will have a non-linear effect on at least one of the outputs. Therefore fewer simulator runs will be required for screening but it is also more likely fewer factors with only linear or no effects across all outputs are detected. An alternative approach would be to use a functional summary of all simulator outputs as the response for the screening analysis. However the EE variance threshold will need to be elicited for the functional summary rather than for each individual simulator output.

Acknowledgements

This work was carried out within the Managing Uncertainty in Complex Models (EPSRC grant EP/D048893/1). We thank Dan Cornford for many useful discussions. We also thank Graham Smith and Alexander Singer at the Food and Environment Research Agency for providing the rabies model and explaining the simulator structure.

A Proof of Lemma 1.

Proof. All computations in this lemma are defined for a single factor. The elementary effect at point x_i follows a normal distribution $EE(x_i) \sim N(a, \frac{2\gamma}{\Delta^2})$. Independence of elementary effects $EE(x_1), \ldots, EE(x_R)$ follows from independence of observations of model.

The mean of R elementary effects is distributed $\mu \sim N(a, \frac{2\gamma}{R\Delta^2})$. To find the distribution

of $\sigma^2 = \frac{1}{R-1}\sum_{i=1}^{R}(EE(x_i) - \mu)^2$, the following sum of squares is used

$$\sum_{i=1}^{R}\left(\frac{EE(x_i) - a}{\sqrt{\frac{2\gamma}{\Delta^2}}}\right)^2 = (R-1)\frac{\sigma^2}{\frac{2\gamma}{\Delta^2}} + R\frac{(\mu - a)^2}{\frac{2\gamma}{\Delta^2}}.$$

The left hand side above is a sum of squared independent standard normal variables and thus it has a chi-squared distribution with R degrees of freedom. By independence of μ and σ^2, the first summand on the right has a chi-squared distribution with $R-1$ degrees of freedom and the second summand has a chi-squared distribution with one degree of freedom. Therefore $\sigma^2 \sim \frac{2\gamma}{(R-1)\Delta^2}\chi^2_{R-1}$. □

B Power computations

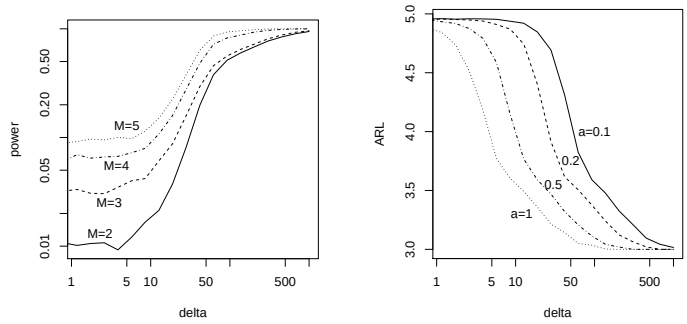
The input x_i is declared non-linear the first time that σ_i satisfies $\sigma_i > \sigma_0$ in step B5 of the algorithm. If the total number of experiments satisfies $M > 2$, then detection of nonlinearity for a given input can occur before the total number of allowed experiments M is reached. Power computations are possible for different departures from the linear model used to determine σ_0, here we explore two.

The first departure we consider is a change in variance, with variance γ being multiplied by a factor δ^2 to become $\gamma' = \delta^2\gamma$. The factor δ quantifies change in variance from the original model. The probability of detecting the change at the step R is $\Pr\left(X^2_{R-1} > \chi^2_{0.99,R-1}/\delta^2\right)$. A geometric walk argument with probabilities changing as more data is included yields the power: $1 - \prod_{R=2}^{M}\left(1 - \Pr\left(X^2_{R-1} > \chi^2_{0.99,R-1}/\delta^2\right)\right)$.

A different departure from linear model is when data follows

$$Y(x_i) = cx_i^2 + ax_i + b + \varepsilon_i\,. \tag{7}$$

Here ε_i still satisfies the same assumptions as for the linear model and a, b and c are constants. The following scenarios were used for power computations through simulations under the (wrong) model (7): $M = 2, 3, 4, 5$, $a = \frac{1}{10}, \frac{1}{5}, \frac{1}{2}, 1$; and the quadratic coefficient depended on the slope through the following relation $c = \delta a$ with δ ranging from 10^{-2} to 10^3. Here the quantity δ is used to quantify departures from the original model through an increasing



(a) (b)

Figure 4: Simulation results: (a) power and (b) ARL against scaling factor δ.

quadratic coefficient. The errors in the models simulated were standard normal, i.e. $\gamma = 1$ and in each scenario used a uniform design of M points over $[0, 1]$. The results of power against different values of M with $a = \frac{1}{10}$ are shown in Figure 4 (a). As would be expected, the test reacts quickly to increasing quadratic trend through δ, but also as M increases and even for very small changes there is already a higher rejection rate. For instance, when $M = 5$ the rate is already larger than 10%.

In each simulation experiment, the run length was recorded. Recall that run length is the number of steps required for a factor to be declared nonlinear and it only takes integer values. A number of $12,000$ simulations were carried out and run lengths obtined were averaged to produce the average run length (ARL). This computation was performed for $M = 5$ and same scenarios for parameters a and c as above. The graph of ARL against δ is shown in Figure 4 (b), where the top curve ($a = \frac{1}{10}$) required more exploration and produced higher ARL than the rest of the cases for a, with the lowest ARL in the simulations achieved for $a = 1$ (bottom curve), where the departure from linear is highest and the method detects it very quickly. Simulations for both Figures 4 (a) and (b) used the same range of δ but in the right figure the results are plotted only for $M = 5$.

Supplementary Materials

Technical Report: Supplementary information on the scalability of the sequential elementary effects method. (pdf)

Matlab code for the sequential screening method: Matlab code that demonstrates the sequential screening method on a synthetic dataset. The example can be started by running seqMorrisExample.m. (GNU zipped tar file)

References

Andres, T. (1997). Sampling methods and sensitivity analysis for large parameter sets. *Journal of Statistical Computation and Simulation*, 57(1-4):77–110.

Campolongo, F. and Braddock, R. (1999). The use of graph theory in the sensitivity analysis of the model output: a second order screening method. *Reliability Engineering & System Safety*, 64(1):1–12.

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling Software*, 22(10):1509–18.

Campolongo, F., Cariboni, J., Saltelli, A., and Schoutens, W. (2004). Enhancing the Morris Method. In Hanson, K. M. and Hemez, F. M., editors, *Proceedings of the 4th International Conference on Sensitivity Analysis of Model Output*, pages 369–79, Santa Fe, New Mexico.

Cukier, R., Fortuin, C., Shuler, K., Petschek, A., and Schaibly, J. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in ratecoefficients. *Journal of Chemical Physics*, 59:38733878.

Garthwaite, P. H. and Dickey, J. M. (1988). Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society. Series B. Methodological*, 50(3):462–474.

Hargreaves, J. C., Annan, J. D., Edwards, N. R., and Marsh, R. (2004). An efficient climate forecasting method using an intermediate complexity Earth System Model and the ensemble Kalman filter. *Climate Dynamics*, 23:745–60.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S., and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, 75(372):845–854.

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, F., Gosling, J., and Heinemeyer, A. (2008). Quantifying uncertainty in the biospheric carbon flux for England and Wales. *Journal of the Royal Statistical Society - Series A*, 171:109–135.

Kleijnen, J. P. C. (2009). Factor screening in simulation experiments: Review of sequential bifurcation. *Advancing the Frontiers of Simulation: A Festschrift in Honor of George S. Fishman*, pages 147–173.

Linkletter, Crystal, Bingham, Derek, Hengartner, Nicholas, Higdon, David, Ye, and Kenny, Q. (2006). Variable selection for gaussian process models in computer experiments. *Technometrics*, 48(4):478–490.

Morris, M. and Mitchell, T. (1995). Exploratory designs for computer experiments. *Journal of Statistical Planning and Inference*, 43:381–402.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–74.

Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

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

Pujol, G. (2009). Simplex-based screening designs for estimating metamodels. *Reliability Engineering & System Safety*, 94:1156–60.

Pujol, G., Iooss, B., and Janon, A. (2013). *sensitivity: Sensitivity Analysis*. R package version 1.7.

Ravi, S. S., Rosenkrantz, D. J., and Tayi, G. K. (1994). Heuristic and special case algorithms for dispersion problems. *Operations Research*, 42(2):pp. 299–310.

Reich, B. J., Storlie, C. B., and Bondell, H. D. (2009). Variable selection in bayesian smoothing spline anova models: Application to deterministic computer codes. *Technometrics*, 51(2):110–120.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409–23.

Saltelli, A., Chan, K., and Scott, E., editors (2000). *Sensitivity Analysis*. New York: Wiley.

Savitsky, T., Vannucci, M., and Sha, N. (2011). Variable Selection for Nonparametric Gaussian Process Priors: Models and Computational Strategies. *ArXiv e-prints*.

Singer, A., Kauhala, F., Holmala, K., and Smith, G. (2008). Towards the elimination of rabies in eurasia. *Developments in Biologicals*, 131:213–222.

Singer, A., Kauhala, F., Holmala, K., and Smith, G. (2009). Rabies in northeastern europe - the threat from invasive raccoon dogs. *Journal of Wildlife Diseases.*, 45(4):1121–1137.

Singer, A. and Kennedy, M. (2008). Sensitivity analysis on the rabies model. Personal Communication.

Sobol, I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Math. Modelling Comput. Exp.*, 1:407–414.

Wedge, D. C., Rowe, W., Kell, D. B., and Knowles, J. (2009). In silico modelling of directed evolution: Implications for experimental design and stepwise evolution. *Journal of Theoretical Biology*, 257:131–41.