

How to avoid a perfunctory sensitivity analysis

Andrea Saltelli*, Paola Annoni

Joint Research Center, Institute for the Protection and Security of the Citizen, via E.Fermi, 2749, Ispra VA 21027, Italy

ARTICLE INFO

Article history:

Received 21 October 2009

Received in revised form

14 April 2010

Accepted 15 April 2010

Available online 15 May 2010

Keywords:

Mathematical modeling

Sensitivity analysis

Uncertainty analysis

Robustness

One-at-a-time

ABSTRACT

Mathematical modelers from different disciplines and regulatory agencies worldwide agree on the importance of a careful sensitivity analysis (SA) of model-based inference. The most popular SA practice seen in the literature is that of 'one-factor-at-a-time' (OAT). This consists of analyzing the effect of varying one model input factor at a time while keeping all other fixed. While the shortcomings of OAT are known from the statistical literature, its widespread use among modelers raises concern on the quality of the associated sensitivity analyses. The present paper introduces a novel geometric proof of the inefficiency of OAT, with the purpose of providing the modeling community with a convincing and possibly definitive argument against OAT. Alternatives to OAT are indicated which are based on statistical theory, drawing from experimental design, regression analysis and sensitivity analysis proper.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Existing guidelines and textbooks reviewed here recommend that mathematical modeling of natural or man-made system be accompanied by a 'sensitivity analysis' (SA). More specifically modelers should:

- Characterize the empirical probability density function and the confidence bounds for a model output. This can be seen as the numerical equivalent of the measurement error for physical experiments. The question answered is "How uncertain is this inference?"
- Identify factors or groups of factors mostly responsible for the uncertainty in the prediction. The question answered is "Where is this uncertainty coming from?"

We call task (a) above 'uncertainty analysis' – UA – and task (b) sensitivity analysis – SA (Saltelli et al., 2008). The two tasks, while having different objectives, are often coupled in practice and called "sensitivity analysis". The term 'sensitivity analysis' can also be used to indicate a pure uncertainty analysis (Kennedy, 2007; Leamer, 1990). Whatever the terminology used, SA is not to be intended as an alternative to UA but rather as its complement.

Recent regulatory documents on impact assessment point to the need of performing a quantitative sensitivity analysis on the output

of a mathematical model, especially when this output becomes the substance of regulatory analysis (EC, 2009; EPA, 2009; OMB, 2006). According to the US Environmental Agency (EPA, 2009):

"[SA] methods should preferably be able to deal with a model regardless of assumptions about a model's linearity and additivity, consider interaction effects among input uncertainties, [...], and evaluate the effect of an input while all other inputs are allowed to vary as well."

According to the European Commission (EC, 2009):

"Sensitivity analysis can be used to explore how the impacts of the options you are analysing would change in response to variations in key parameters and how they interact."

The term 'interaction' used in both guidelines naturally links to experimental design and ANOVA. The US office for Management and Budget prescribes that:

"Influential risk assessments should characterize uncertainty with a sensitivity analysis and, where feasible, through use of a numeric distribution. [...] Sensitivity analysis is particularly useful in pinpointing which assumptions are appropriate candidates for additional data collection to narrow the degree of uncertainty in the results. Sensitivity analysis is generally considered a minimum, necessary component of a quality risk assessment report."

Modelers and practitioners from various disciplines (Kennedy, 2007; Leamer, 1990; Pilkey and Pilkey-Jarvis, 2007; Saltelli et al.,

* Corresponding author.

E-mail address: andrea.saltelli@jrc.ec.europa.eu (A. Saltelli).

2008; Santner et al., 2003; Oakley and O'Hagan, 2004; Saisana et al., 2005) share the belief that sensitivity analysis is a key ingredient of the quality of a model-based study.

At the same time most published sensitivity analyses – including a subset reviewed in the present paper by way of illustration – rely on 'one-at-a-time' sensitivity analysis (OAT-SA). OAT use is predicated on assumptions of model linearity which appear unjustified in the cases reviewed. While this is well known to sensitivity analysis practitioners as well as to the statistical community at large, modelers seem reluctant to abandon this practice. The following examples have been retained for the present work: Bailis et al. (2005), Campbell et al. (2008), Coggan et al. (2005), Murphy et al. (2004), Stites et al. (2007).

The purpose of this paper is to present a novel, simple and hopefully convincing argument against OAT, aimed at the modeling community. We also sketch for each of the reviewed cases what alternative technique could be used under which assumption. In most cases the alternative will come at no additional computational cost to the modeler, where computational cost is defined as the number of times the model has to be evaluated. This cost is relevant for models whose execution is computer and analyst time intensive. To put the present criticism of OAT in perspective, it is also important to stress that in recent years more examples of good practices have appeared. Limiting oneself to environmental applications of good practices of sensitivity analysis appearing in the present year, Thogmartin (2010) performs both an OAT and a global sensitivity analysis using Fourier amplitude sensitivity test – FAST, on a model for birds population. This author finds a typical, and for us unsurprising, contradiction between local and global results which is due to the effect of interactions between input parameters. Another good example of global sensitivity analysis concerns a relatively complex crop model described in Varella et al. (2010). In this latter work the Authors apply the Extended FAST method to compute estimates of the first and total order sensitivity indices. Uncertainty analysis using the Latin Hypercube sampling strategy is carried out by Albrecht and Miquel (2010). Note that the motivation for using Latin Hypercube as opposed to random sampling to explore a multidimensional space is precisely to maximize the exploration of the space for a given number of points.

Good practices for sensitivity analysis are also increasingly seen on this journal based on regression analysis (Manache and Melching, 2008), variance based methods (Confalonieri et al., 2010) and meta-modelling (Ziehn and Tomlin, 2009). All these treat the model as a black box. When information is available of the characteristics of the model ad hoc strategies can of course be devised, see e.g. Norton (2008).

Finally, the purpose of the present paper is not to intercompare in earnest methods' performance as this is already accomplished in many publications as, for instance, in Campolongo et al. (1999, 2007), Helton et al. (2006), Cacuci and Ionesco-Bujor (2004).

2. Local, global and OAT sensitivity analysis

Searching the literature in environmental and life sciences with "sensitivity analysis" as a keyword, it is easy to verify that most often SA is performed by changing the value of uncertain factors one-at-a-time (OAT) while keeping the others constant. Some examples of OAT-SA – selected out of a much longer list for the purpose of illustration – are Ahtikoski et al. (2008), Bailis et al. (2005), Campbell et al. (2008), Coggan et al. (2005), de Gee et al. (2008), Hof et al. (2008), Hasnadi and Taylor (2008), Murphy et al. (2004), Nolan et al. (2007), Stites et al. (2007), Van der Fels-Klerx et al. (2008).

The present work is not concerned with the literature of 'local' sensitivity analysis, where factors' importance is investigated by

derivative (of various order) of the output with respect to that factor. The term 'local' refers to the fact that all derivatives are taken at a single point, known as 'baseline' or 'nominal value' point, in the hyperspace of the input factors.¹ At times, e.g. in the treatment of inverse problems (Rabitz, 1989; Turanyi, 1990), or in approximating a model output in the neighborhood of a set of pre-established boundary conditions, it may not be necessary to average information over the entire parameter space and local approaches around the nominal values can still be informative. In other applications 'adjoint sensitivity analysis' can be performed which is capable of handling thousands of input variables, as typical in geophysical and hydrological applications (Castaings et al., 2007; Cacuci, 2003). 'Automated differentiation' techniques and software are also popular for these applications (Grievank, 2000).

In principle, local analyses cannot be used for the robustness of model based inference unless the model is proven to be linear (for the case of first order derivatives) or at least additive (for the case of higher and cross order derivatives). In other words, derivatives are informative at the base point where they are computed, but do not provide for an exploration of the rest of the space of the input factors unless some conditions (such as linearity or additivity) are met in the form of the mathematical model being represented.

When the property of the models is a-priori unknown, a global SA is preferred (Kennedy, 2007; Leamer, 1990; Oakley and O'Hagan, 2004; Saltelli and Tarantola, 2002; Helton et al., 2006). Practitioners call this a model-free setting. Thus we support methods based on exploring the space of the input factors according to the consideration that a handful of data points judiciously thrown in that space is far more effective, in the sense of being informative and robust, than derivative values estimated at a single data point at the centre of the space.

A global approach aims to show that even varying the input assumptions within some plausible ranges some desired inference holds, see Stern (2006). The best illustration of this strategy is due to Leamer (1990):

"I have proposed a form of organized sensitivity analysis that I call "global sensitivity analysis" in which a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighborhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful".

Funtowicz and Ravetz (1990) similarly argue: "GIGO (Garbage In, Garbage Out) Science – [is] where uncertainties in inputs must be suppressed lest outputs become indeterminate."²

Modelers in Ahtikoski et al. (2008), Bailis et al. (2005), Campbell et al. (2008), Coggan et al. (2005), de Gee et al. (2008), Hof et al. (2008), Hasnadi and Taylor (2008), Murphy et al. (2004), Nolan et al. (2007), Stites et al. (2007), Van der Fels-Klerx et al. (2008) cannot assume linearity and additivity as their models come in the form of computer programmes, possibly including differential equations solvers, smoothing or interpolation algorithms, or parameters estimation steps. Using OAT, these authors move one factor at a time from the baseline (or 'nominal') value and looks at the effect that this change has on the output. At the baseline point all k uncertain factors are at their reference or 'best' estimated value. All factors considered in these OAT analyses are taken to be independent from one another. In this setting the space of the input

¹ When derivatives are taken at several points in this space to arrive at some kind of average measure the approach is no-longer local, see Kucherenko et al. (2009).

² For a discussion of the work of Funtowicz and Ravetz in a statistician's perspective (see Zidek, 2006).

factors can be normalized to the unit hyper-cube of side one, and the task of a sensitivity analysis design it to explore this cube with uniformly distributed points. These points can then be mapped onto the real distribution functions of the input factors, using a quantile transformation, see [Shorack \(2000\)](#). Based on these premises, the geometric proof of the inadequacy of the OAT approach is introduced next.

3. OAT can't work. A geometric proof

[Fig. 1](#) illustrates the 'curse of dimensionality'. Hyper-spheres are included in – and tangent to – the unit hypercube (the case of two and three dimensions is shown), and the curve on the plot is the ratio r of the volume of the sphere to that of the cube in k dimensions. Why is this plot relevant here? Because all the points of the OAT design are *by construction* internal to the sphere. The volume of the sphere goes very rapidly to zero with increasing the number of dimensions k .

In a system with just two uncertain factors ($k = 2$) the area of the circle inscribed in the unit square, and hence the ratio of the 'partially' explored to the total area, is $r(2) = \pi(1/2)^2 \sim 0.78$. In three dimensions, as in [Bailis et al. \(2005\)](#), [Campbell et al. \(2008\)](#), this is $r(3) = (4\pi/3)(1/2)^3 \sim 0.52$. The general formula for the volume of the hypersphere of radius $1/2$ in k dimensions is:

$$r(k) = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2} + 1)} \left(\frac{1}{2}\right)^k$$

where $\Gamma(k/2 + 1)$ can be derived for both even and odd k by the following properties: $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$; $\Gamma(x+1) = x\Gamma(x) \forall x \in \mathbb{R} \setminus (-\mathbb{N})$, $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$ ([Abramowitz and Stegun, 1970](#)). It is easy to see that in 12 dimensions, as in [Stites et al. \(2007\)](#), the fraction of the hyperspace explored is equal to $r = 0.000326$, less than one-thousandth of the factors' space ([Fig. 1](#)). This is one of the many possible ways to illustrate the so-called *curse of dimensionality*. See [Hastie et al. \(2001\)](#) for a different interpretation.

An OAT sensitivity analysis thus designed is therefore perfunctory in a model free setting. In a high dimensional space, a derivative based analysis and an OAT based one are practically equivalent, i.e. they are both *local* and thus by definition non-explorative.

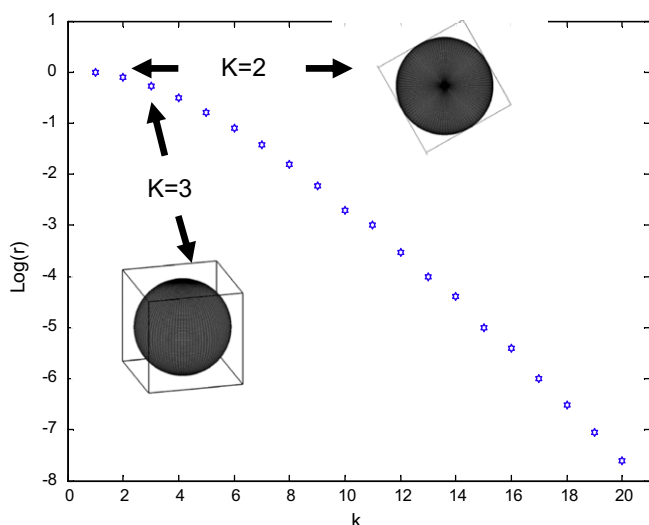


Fig. 1. The curse of dimensionality. In $k = 3$ dimensions the volume of the sphere internal to a cube and tangent to its face is $r \sim 0.5$. r goes rapidly to zero with increasing k .

Further OAT cannot detect interactions among factors because this identification is predicated on the simultaneous movement of more than one factor. If factors are moved OAT the interactions are not activated and hence not detected, i.e. one has no way to know whether the effect of moving X_1 and X_2 is different from the superposition of the individual effects obtained moving first X_1 , coming back to the baseline and then moving X_2 .

The inadequacy of OAT is not limited to sensitivity analysis, e.g. to the quest for the most influential model input factors, but to uncertainty analysis as well. Elementary statistics about the model output (inference), such as its maximum, or mode, can be totally misinterpreted via OAT. We shall return to this point in Section 5, but we discuss OAT's fortune first.

4. Why modelers prefer OAT

For a modeler the baseline is a very important, perhaps the most important, point in the space of the input factors. This is customarily the best estimate point, thus one can understand why in a OAT design one comes back to the baseline after each step. Yet to assume that all what one needs to explore are the neighborhoods of the baseline amounts to imply that all the probability density functions for the uncertain factor has a sharp peak on this multidimensional point. The 'peaked' pdf's assumption does not belong to the examples reviewed here, but were it true why not using local methods? Arguments which might justify the favor enjoyed by OAT are

- the baseline vector is a safe starting point where the model properties are well known;
- all OAT sensitivities are referred to the same starting point;
- moving one factor at a time means that whatever effect is observed on the output (including the case of no effect), this is due *solely* to that factor – no noise is involved unless the model has a stochastic term. No effect does not mean no influence, of course;
- conversely, a non-zero effect implies influence, i.e. OAT does not make type I errors, it never detects uninfluential factors as relevant;
- the chances of the model crashing or otherwise giving unacceptable results is minimized, as these are likely to increase with the distance from the baseline. The model has more chances to crash when all its k factors are changed than when just one is. A global SA is by definition a stress testing practice.

The last point might seem surprising but it is of practical importance. In case of model failure under OAT analysis, the modeler immediately knows which is the input factor responsible of the failure. Further, as one of the reviewer of the present work pointed out, the OAT approach is consistent with the modeler way of thinking one parameter at a time as she/he wants to verify systematically the effect of parameter variation. Taking all these consideration into account, a possible way to correcting OAT is by using the Elementary Effect method, which will be presented in Section 5.3, and is based on a limited number of iterations of the OAT analysis performed by changing the baseline.

5. Suggested practices

There are several alternative to OAT-SA which are based on statistical theory. These are now illustrated, having care to use for each of the OAT studies reviewed a number of model runs close to that of the original OAT analysis. The first two approaches, linear regression and factorial design, belong to the theory of statistics for the analysis of physical experiments or observational data, while the latter two have been developed specifically for the sensitivity

analysis of mathematical models where the data are the output of a numerical computation.

5.1. A two level design

For the analysis by [Bailis et al. \(2005\)](#), [Campbell et al. \(2008\)](#) where $k=3$, a simple Factorial Design - FD - can be suggested ([Box et al., 1978](#)). Let Y be the experiment outcome depending on three input factors X_i . A saturated two-level FD, where each factor X_i can take two possible values (levels), denoted as '1' or '0' can be applied. The design simulates all possible combination of the factor levels with a computational cost in this case of $2^3 = 8$ points.

This simple design can be used to estimate factors' main effects and interactions. As can be seen from [Fig. 2](#), for $Y = Y(X_1, X_2, X_3)$ an eight points, two level design involves all points from $Y(0, 0, 0)$ to $Y(1, 1, 1)$. The two faces of the cube in [Fig. 2](#), respectively, defined by $\{Y(0, 0, 0), Y(0, 1, 0), Y(0, 1, 1), Y(0, 0, 1)\}$ and $\{Y(1, 0, 0), Y(1, 1, 0), Y(1, 1, 1), Y(1, 0, 1)\}$ are used to estimate the main effect of X_1 :

$$\text{Eff}(X_1) = \frac{1}{4}(Y(1, 0, 0) + Y(1, 1, 0) + Y(1, 1, 1) + Y(1, 0, 1)) + \frac{1}{4}(Y(0, 0, 0) + Y(0, 1, 0) + Y(0, 1, 1) + Y(0, 0, 1))$$

i.e. the difference between two faces of the cube along the X_1 direction. The main effect of X_1 is thus defined as the difference between the average experiment outcome for the '1' level of X_1 and the average outcome for the '0' level of X_1 . Analogously for the main effects of X_2 and X_3 . The idea is extended to interactions. Two factors, say X_1 and X_3 , are said to interact if their effect on the outcome Y if the effect of X_1 is different at the two different levels of X_3 . The second order interaction between X_1 and X_3 is defined as half the difference between the main effect of X_1 for $X_3 = '1'$ and the main effect of X_1 for $X_3 = '0'$:

$$\text{Eff}(X_1, X_3) = \frac{1}{4}(Y(0, 0, 0) + Y(0, 1, 0) + Y(1, 0, 1) + Y(1, 1, 1)) + \frac{1}{4}(Y(1, 0, 0) + Y(1, 1, 0) + Y(0, 0, 1) + Y(0, 1, 1))$$

and so on for the other second order terms. The third order term can be computed in a similar way ([Box et al., 1978](#)). In [Campbell et al. \(2008\)](#) five points were used in three dimensions (one dimension was explored with two steps instead of one). In this way two single factor effects are obtained as well as two non-independent estimates for the third factor. With three more points as just described one would have obtained four non-independent estimates for the main effect of each factor plus estimates for the second and third order effects – this approach is not model-free, neither does it explore thoroughly the input factors' space, but at least it is not bound by an assumption of model linearity.

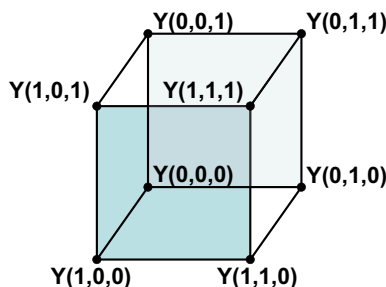


Fig. 2. Two-level, full factorial design for three factors.

5.2. Regression

A simple way to carry out a global SA is using regression techniques, such as standardized regression coefficients. The analysis consists in regressing, usually with an ordinary least squares, one (or more) output variables Y with respect to a set of input factors X 's. The classical form of a regression model is: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$ where the β 's are the unknown parameters and ε is the statistical error. The outcome can be seen as a meta-model where estimated output values are described in terms of linear combination of the input factors. By means of the model coefficient of determination R^2_Y , non-linearity or non-additivity of the model may be detected. In fact, R^2_Y gives by definition the proportion of variability in Y explained by regression on the X 's and is a scale-free normalized number. It can be shown that R^2_Y is the square of the multiple correlation coefficient between Y and the X 's, that is the square of the maximum correlation between Y and any linear function of the X 's ([Weisberg, 1985](#)). The lower R^2_Y the higher the non-linearity of the model. The standardized regression coefficients $\hat{\beta}_i$ are defined as the estimates of regression parameters corresponding to standardized Y and X 's. Parameter $\hat{\beta}_i$ of factor X_i can be used as sensitivity measure for X_i , since $\sum_{i=1}^k \hat{\beta}_i^2 = R^2_Y$ when factors are independent ([Draper and Smith, 1981](#)).

The regression method has three main advantages:

- it explores the entire interval of definition of each factor;
- each 'factor effect' is averaged over that of the other factors;
- standardized regression coefficients give also the sign of the effect of an input factor on the output.

Authors in [Coggan et al. \(2005\)](#) use 40 points for $k=4$ (with some waste given to the non independence of these effects once taken along the same OAT line). These authors could more usefully use the same 40 points in a Monte Carlo plus regression framework. Similar considerations apply to the work of [Murphy et al. \(2004\)](#).

The main pitfall of regression based SA is that, when R^2_Y is low (non-linear and/or non additive model), it is pointless to use the values of $\hat{\beta}_i$ for ranking input factors ([Saltelli et al., 2004](#)). Note that although linear regression is in principle predicated on model linearity, it in fact takes us further, by giving an estimate of the degree of non-linearity R^2_Y , which works at the same time as a measure of the quality of our $\hat{\beta}_i$ -based analysis.

5.3. Elementary effects method

The practice of reverting to the baseline point in order to compute any new effect is what gives to OAT its appealing symmetry as well as its poor efficiency. A good OAT would be one where, after having moved of one step in one direction, say along X_1 , one would straightway move of another step along X_2 and so on till all factors up to X_k have been moved of one step each. This type of trajectory is variously known among practitioners as 'Elementary Effects' (EE), from [Saltelli et al. \(2008\)](#), or 'Morris', from [Morris \(1991\)](#) and [Campolongo et al. \(2007\)](#), or winding stairs, from [Jansen \(1999\)](#). Both when using EE or winding stairs one does not stop to a single trajectory, but tries to have as many as possible compatibly with the cost of running the model. In winding stairs all trajectories are joined, while in EE they are separate. Good albeit non quantitative results can be obtained for screening purposes using between four and ten trajectories but already two trajectories can be quite informative as they give a double estimate for the effect of each factor, and by difference of these an idea of the deviation from linearity. This approach would be an ideal

alternative to the OAT analysis by Stites et al. (2007), where $k = 12$, and $2k + 1 = 25$ points in the space of input factors were used by taking two OAT steps along each direction. Note that using the two-trajectory EE at the cost of $2(k + 1) = 26$ points one would obtain two independent effects, while the two effects taken along the same line in OAT are not independent. EE trajectories are considered as a good practice for factors screening in sensitivity analysis (EPA, 2009; Saltelli et al., 2008).

One may wonder whether there is a way to ‘complete’ the OAT approach – since this is the one naturally preferred by modelers – with additional simulation as to make the result more reliable. In fact this can be achieved by simply iterating the OAT analysis itself. To understand how this may work one has to see OAT as a particular case of EE at one trajectory. Although OAT has a radial symmetry which the trajectory has not, both a one-trajectory EE and an OAT provide a single estimate for each factor’s effect. As a result one can imagine a two-OAT’s (instead of a two trajectories) design whereby, provided that the baseline points are different for the two OAT’s, one still obtains two independent estimate for the effect of each factor. This numerical approach can be termed a ‘radial-EE’ design, and is a legitimate alternative to trajectory-based EE (Saltelli et al., 2010).

5.4. Variance based methods on design points

When modelers can afford more ‘expensive’ travels across the k -dimensional parameter space, variance based sensitivity indices can be computed. The design needed to compute these indices is based on Monte Carlo or some form of stratified sampling, such as for instance the Latin Hypercube Sampling (LHS) (Blower et al., 1991; McKay et al., 1979). Quasi random numbers may be used as well and Sobol’ LP_τ sequences (Sobol’, 1967, 1976) were found to perform better than both crude random sampling and LHS (Homma and Saltelli, 1996).

A first order sensitivity index (or main effect) is defined as:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)} \quad (1)$$

where X_i is the i -th factor and $\mathbf{X}_{\sim i}$ denotes the matrix of all factors but X_i . The meaning of the inner expectation operator is that the mean of Y , the scalar output of interest, is taken over all possible values of $\mathbf{X}_{\sim i}$ while keeping X_i fixed. The outer variance is taken over all possible values of X_i . The variance $V(Y)$ in the denominator is the total (unconditioned) variance. The total sensitivity index is defined as (Homma and Saltelli, 1996; Saltelli and Tarantola, 2002):

$$S_{Ti} = \frac{E_{\mathbf{X}_{\sim i}}(V_{X_i}(Y|\mathbf{X}_{\sim i}))}{V(Y)} \quad (2)$$

Both S_i , S_{Ti} have an intuitive interpretation. Referring to the numerators in Eqs. (1) and (2) above:

- $V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))$ is the expected reduction in variance that would be obtained in X_i could be fixed.
- $E_{\mathbf{X}_{\sim i}}(V_{X_i}(Y|\mathbf{X}_{\sim i}))$ is the expected variance that would be left if all factors but X_i could be fixed.

S_i and S_{Ti} are quite powerful and versatile measures. S_i gives the effect of factor X_i by itself, while S_{Ti} gives the total effect of a factor, inclusive of all its interactions with other factors. For additive models $S_i = S_{Ti}$ for all factors. If the objective of the sensitivity analysis is to fix non-influential factors, then S_{Ti} is the right measure to use (Saltelli et al., 2004). A detailed recipe to compute both S_i and S_{Ti} when the input factors are independent from one another is given for example in Saltelli (2002) and Saltelli et al. (2010) while

the case of non-independent input is discussed in Saltelli and Tarantola (2002) and Da Veiga et al. (2009).

6. One at a time versus elementary effects

As discussed in the previous section, variance based S_i and S_{Ti} are candidate best practices to carry out a sensitivity analysis. Still they are not a good choice for all those model which are computationally expensive. In this section we show that even using a handful of model evaluations, methods better than OAT can be adopted. The method of the Elementary Effects just described (Section 5.3) is used to this purpose. The comparison is ‘fair’ as approximately the same number of points in the input factor space is used for both OAT and EE.

6.1. Toy functions

Performances of the two methods are firstly compared on the basis of the empirical cumulative distributions – CDF – of two toy functions described below. In these terms the comparison is based on an uncertainty analysis and not a sensitivity analysis, as defined in Section 1. We rely here on the fact that the UA is the first step of a SA. If one is wrong about the domain of existence of the target function, there is little chance that a sensitivity analysis run on this space may be reliable.

Two test functions commonly used in SA methods intercomparison are selected for the analysis (Saltelli et al., 2010; Da Veiga et al., 2009; Kucherenko et al., 2009):

- G function of Sobol’, (Archer et al., 1997), defined as:

$$G = G(X_1, X_2, \dots, X_k, a_1, a_2, \dots, a_k) = \prod_{i=1}^k g_i$$

$$g_i = \frac{|4X_i - 2| + a_i}{1 + a_i}$$

where $a_i \in \mathbb{R}^+ \forall i = 1, \dots, k$, k total number of input factors.

- D function, (Da Veiga et al., 2009):

$$D = D(X_1, X_2)$$

$$= 0.2\exp(X_1 - 3) + 2.2|X_2| + 1.3X_2^6 - 2X_2^2 - 0.5X_2^4 - 0.5X_1^4$$

$$+ 2.5X_1^2 + 0.7X_1^3 + \frac{3}{(8X_1 - 2)^2 + (5X_2 - 3)^2 + 1}$$

$$+ \sin(5X_1)\cos(3X_1^2)$$

The first function is more ductile as one can increase ad libitum the number of input factors X_i . Indeed its typology is driven by the dimensionality k as well as by the value of the coefficients a_i . Low values of a_i , such as $a_i = 0$, imply an important first order effect. If more than one factor has low a_i ’s, then high interaction effects will also be present. Function D has been recently used to compare SA methods in a two-dimensional setting (Da Veiga et al., 2009).

In the argument against OAT the support of the two functions is set to $[-1, +1] \times [-1, +1]$, that is we take two input factors varying uniformly in the $[-1, +1]$ interval. The vector of parameters \mathbf{a} for the G function is set to $(0, 0.01)$. Fig. 3 shows the two functions in this case.

In order to show that OAT may fail in catching even the elementary statistics of a multidimensional function, we set up an experiment to estimate the cumulative distribution function CDF of the G and D functions using OAT (Section 2) and EE approach (Section 5.3). All statistics which constitute the subject of a plain uncertainty analysis, such as minimum, maximum, mean, median and variance will of course fail if the percentiles are wrongly estimated.

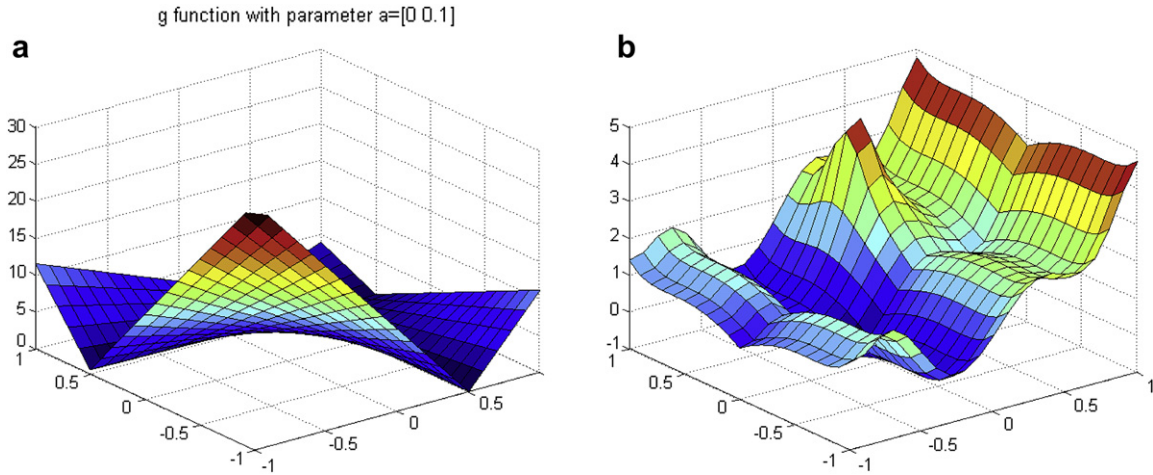


Fig. 3. Plots of the two test functions on the support $[-1, +1] \times [-1, +1]$: (a) G function; and (b) D function.

The OAT experiment is carried out considering the origin 0 as baseline point. For each factor four steps are taken symmetrically in the positive and negative direction of length 0.2 and 0.4. The total number of function evaluations for the OAT experiment is $(1 + 4k)$, with $k = 2$ number of input factors.

For comparability purposes, the EE experiment is carried out using a number γ of trajectories equal to $\text{round}((1 + 4k)/(k + 1))$ with 'round' meaning the nearest integer. In this way the number of function evaluations of the EE experiment is comparable (if not exactly equal) to the number of function evaluations of the OAT experiment, for each input factor number k . For example, for two input factors the total number of function evaluation would be nine.

The 'true' CDF of the two test functions derives from quasi Monte-Carlo experiment using Sobol' sequence of quasi random numbers (Sobol', 1993). The number of function evaluations in this case is set to $1000 \cdot k$. The CDF from the quasi Monte-Carlo experiment is considered as the reference.

Fig. 4 shows OAT and EE empirical CDFs with respect to the true one (solid line) of both test functions. It can be seen that OAT output estimates tend to 'stick' near the baseline in both cases, thus failing in detecting high values.

The decreasing performance of OAT as k increases can be seen from Fig. 5 which shows empirical CDFs by OAT and EE for the G function with a. $k = 5$ and b. $k = 7$ input factors (as noticed already, for the G function only is possible to increase the number of input factors). The vector of input parameters is set equal to $\mathbf{a} = (0, 0.1, 0.2, 0.3, 0.4)$ for $k = 5$ and $\mathbf{a} = (0, 0.1, 0.2, 0.3, 0.4, 0.8, 1)$ for $k = 7$. From these plots it is evident that OAT is not capable to 'move' enough from the reference point while EE trajectories are relatively more accurate on roughly an equal number of points in the input factor space.

As mentioned above, similar results can be obtained by a radial approach to estimate the effects, e.g. trajectories can be replaced by iterated OAT's. Note that the EE method captures interactions, though it is unable to tell them apart from nonlinearities (Morris, 1991), because EE measures the effects at different points in the multidimensional space of input factors.

6.2. An environmental case study

As additional case study of relevance to environmental studies the Bateman equations (Cetnar, 2006) are considered. These describe a chain of species mutating one into another without backward reaction, of the type A transforming into B transforming into C for an arbitrary long chain of species. These could be various type of biota, chemical compounds, or nuclear radioisotopes.

The Bateman equations describe the concentrations N_i of a number k of species in linear chain governed by a rate constants λ_i ($i = 1, \dots, k$):

$$\begin{aligned} \frac{dN_1}{dt} &= -\lambda_1 N_1 \\ \frac{dN_i}{dt} &= \lambda_{i-1} N_{i-1} - \lambda_i N_i; \quad i = 2, \dots, k \end{aligned} \quad (3)$$

If all the concentrations of daughter species at time zero are zero

$$N_1(0) \neq 0 \quad N_i(0) = 0 \quad \forall i > 1$$

the concentration of the last k th species at time t is:

$$\begin{aligned} N_k(t) &= \frac{N_1(0)}{\lambda_k} \sum_{i=1}^k \lambda_i \alpha_i \exp(-\lambda_i t) \\ \alpha_i &= \prod_{j=1, j \neq i}^k \frac{\lambda_j}{(\lambda_j - \lambda_i)} \end{aligned} \quad (4)$$

The rate constants λ_i , ($i = 1, \dots, k$), are considered as uncertain input factors with uniform distributions with support $[a_i, b_i]$. Both a_i and b_i are expressed in s^{-1} . They are randomly sampled in the interval $[1, 100]$ and reordered when $b_i < a_i$. Thus, all the factors (rates in this case) have comparable orders of magnitude.

The case is particularly interesting in our setting because it allows us to run various experiments with a different number of uncertain factors and to show the decreasing level of performance of OAT with respect to EE as the dimensionality of the problem increases. To this aim six scenarios are set-up with different numbers of species: from the simplest (in terms of dimensionality) two species case to the most complex 12 species case.

The system output $N_k(t)$ is a time dependent function. For our experiments a fixed time $t = 0.1$ s is set for all the simulations, while the starting concentration of the first species is set to $N_1(0) = 100$ (in arbitrary units). As for the toy function case (Section 6.1), the comparison between OAT and EE methods is based upon empirical CDFs. OAT points are taken starting with 0.5 as central point and taking four additional points for each factor corresponding to the sequence $\{0.1; 0.3; 0.7; 0.9\}$. Values of input factors λ_i are computed by inverting their cumulative distribution functions. The total number of function evaluations for the OAT experiment is $1 + 4k$, with k number of species.

As for the previous case, the EE experiment is carried out with a number of trajectories $\gamma = \text{round}((1 + 4k)/(k + 1))$. This allows for the OAT and EE experiments to be comparable in terms of number of function evaluations.

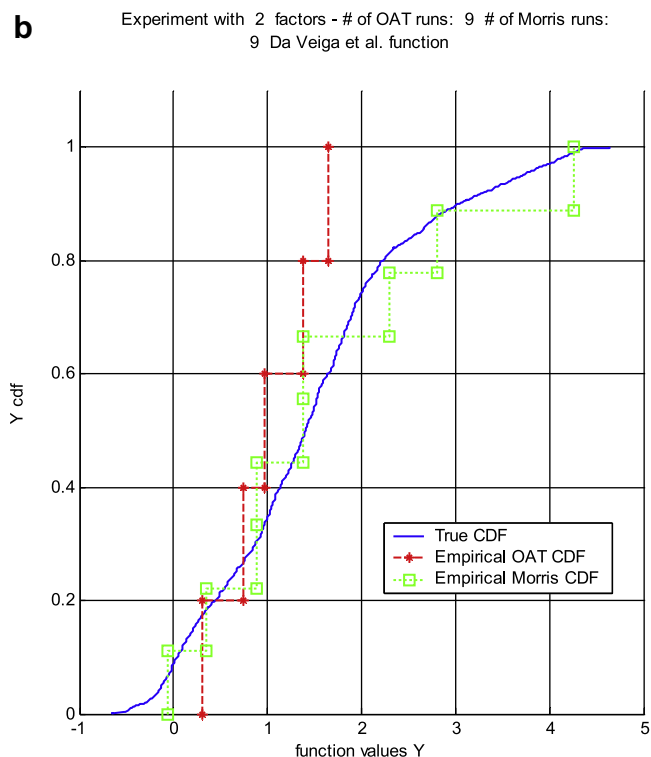
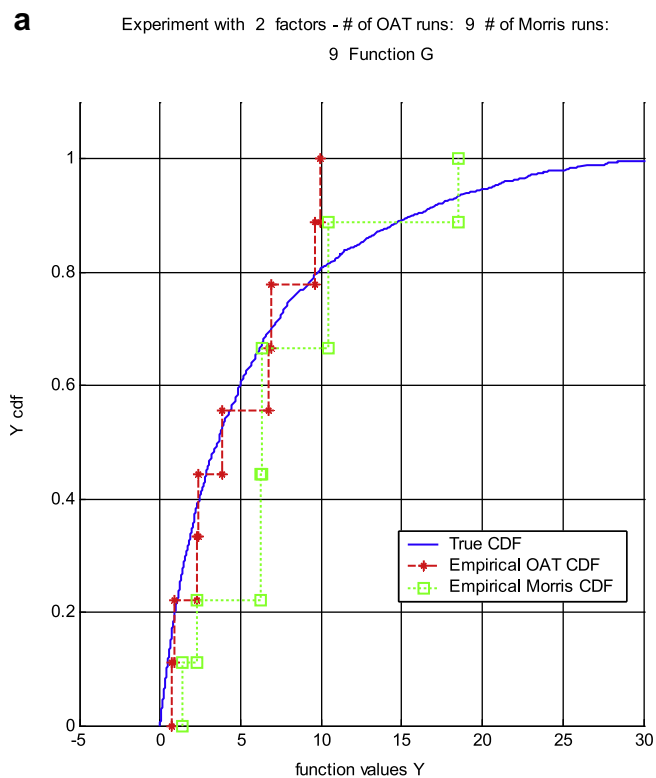


Fig. 4. OAT and EE comparison for the (a) G and (b) D function, $k = 2$.

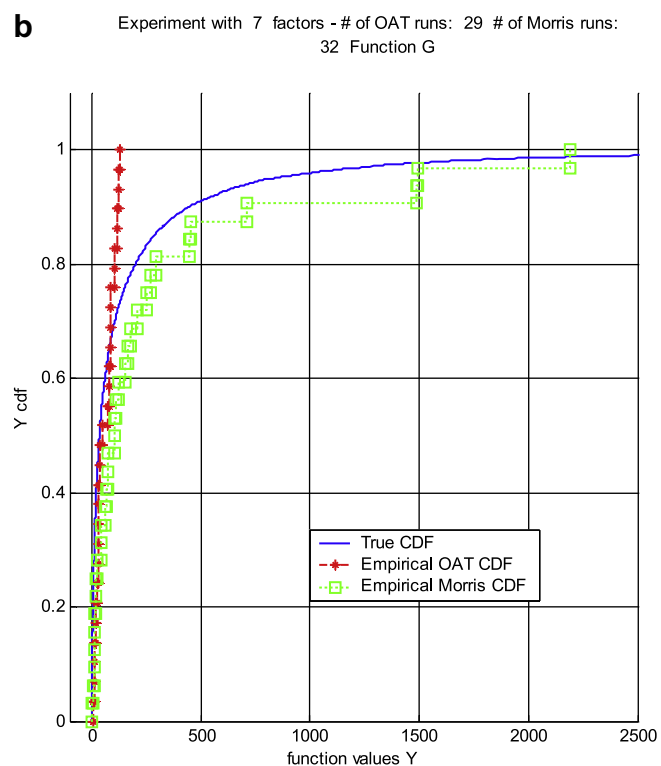
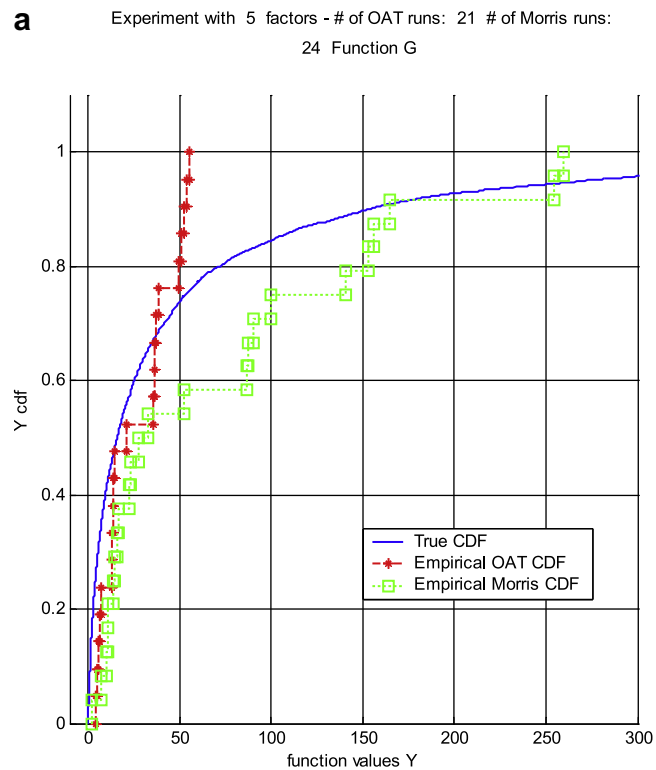


Fig. 5. OAT and EE comparison for the G function with (a) five and (b) seven input factors.

The 'true' CDF of the two test functions derives from quasi Monte-Carlo experiment using Sobol' sequence of quasi random numbers (Sobol', 1993) with $1000 \cdot k$ function evaluations. The CDF from the quasi Monte-Carlo experiment is again considered as reference.

Fig. 6 shows the comparison of the two methods with an increasing number of uncertain input factors (k goes from 2 to 12 going from the upper-left side to the bottom-right side of Fig. 6). We tested the two methods with a minimum of 9 runs (for the two-

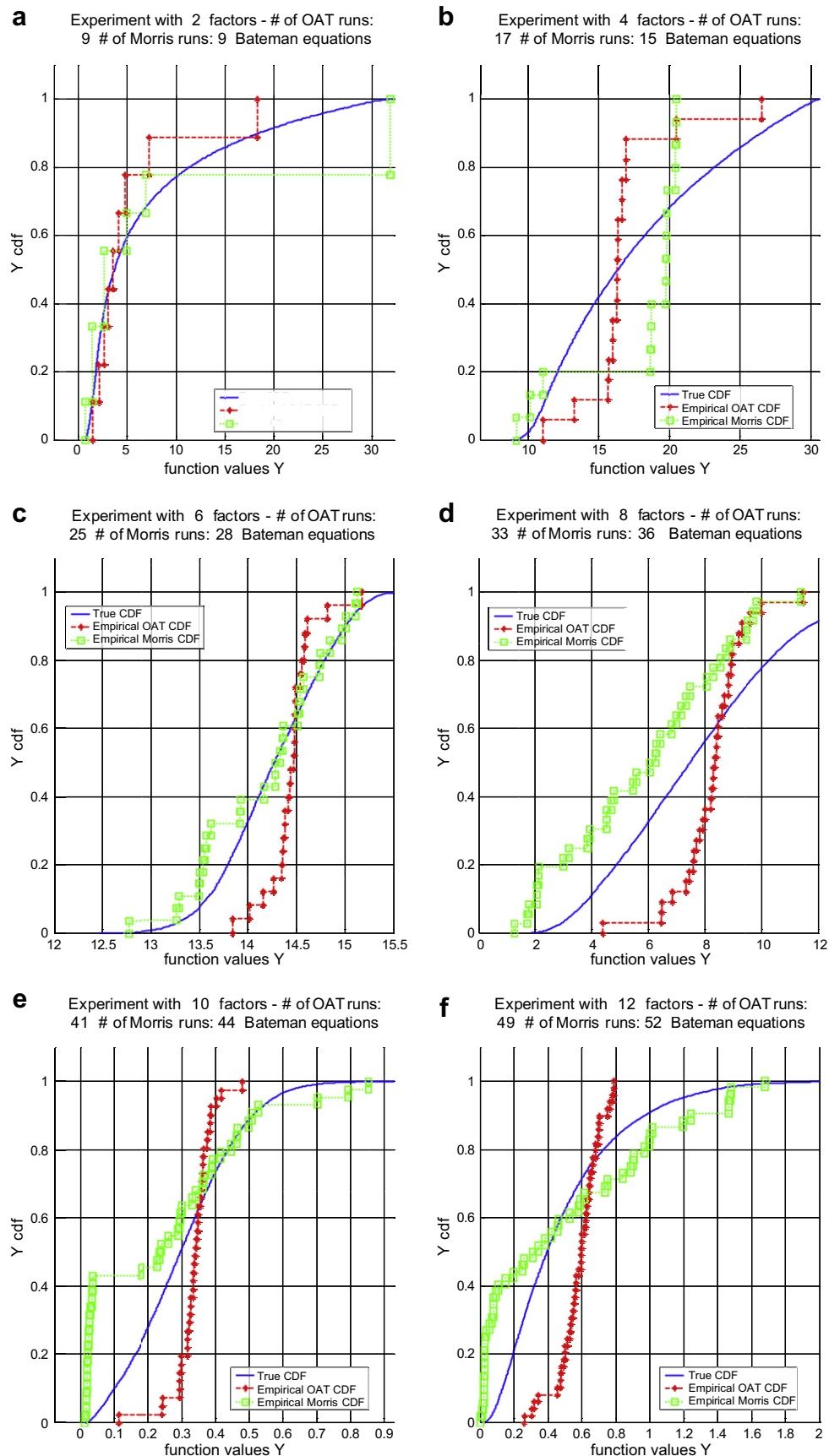


Fig. 6. OAT and EE comparison for the Bateman equations. From up-left side to down-right side: (a) 2 species series; (b) 4 species series; (c) 6 species series; (d) 8 species series; (e) 10 species series; and (f) 12 species series.

dimensional case) and a maximum of about 50 runs (for the 12-dimensional case). At first sight, Fig. 6 suggests that the range of variation of estimated values by OAT method is always lower than that by EE. As the number of dimensions increases (from Fig. 6.c onward) it is easy to note that OAT points are stuck in the neighborhood of a central point and are less capable of following the shape of the real CDF than the EE points.

7. Conclusions

A novel geometric argument was introduced to demonstrate the inadequacy of an OAT-SA. A few practices have been illustrated as possible alternatives. Of these, the full factorial design, the regression and the EE design could have been applied at the same sample size adopted by the authors reviewed, that is at no extra cost.

The variance based methods require a larger investment in computer time which might or might not be affordable depending on the computational cost of the model. A wealth of approaches has been omitted in this short note. Among these, an active line of research focus on emulators, e.g. on trying to estimate the model at untried points, thus facilitating the computation of the sensitivity measures, especially variance based. An early example of this approach is due to Sacks et al. (1989), while recent ones are by Oakley and O'Hagan (2004), Ratto et al. (2007).

When modelers are constrained by computational costs, a recommended practice is to perform a screening step by means of EE trajectories, followed by application of more computationally intensive methods to a smaller set of input factors, as exemplified in Mokhtari et al. (2006).

One aspect worth noticing in sensitivity analysis is that similar recommendations and recipes can be found in Economics (Kennedy, 2007; Leamer, 1990) as well as in Ecology (EPA, 2009; Pilkey and Pilkey-Jarvis, 2007). The importance of SA in impact assessment (EC, 2009; OMB, 2006) would suggest that more attention needs to be paid to the theory of sensitivity analysis to increase the transparency of – and trust in – model-based inference. According to Zidek (2006) “[Statisticians] should be able to take their legitimate place at the head table of science and to offer clear and intelligent criticism as well as comment.” A anonymous reviewer of *Science*, in reply to our critique of the papers just reviewed, noted: “Although [the above] analysis is interesting [...] it [is] more appropriate for a more specialized journal.” Although this is likely a standard reply form, the discussion presented in the present paper is plain, rather than specialized, and we consider hence urgent that some concern for a well designed sensitivity analysis seeps all the way to the less-specialized journals.

References

- Abramowitz, M., Stegun, I., 1970. Handbook of Mathematical Functions. Dover Publications, Inc., New York.
- Ahtikoski, A., Heikkilä, J., Alenius, V., Siren, M., 2008. Economic viability of utilizing biomass energy from young stands – the case of finland. *Biomass and Bio-energy* 32, 988–996.
- Albrecht, A., Miquel, S., 2010. Extension of sensitivity and uncertainty analysis for long term dose assessment of high nuclear waste disposal sites to uncertainties in the human behaviour. *Journal of Environmental Radioactivity* 101 (1), 55–67.
- Archer, G., Saltelli, A., Sobol, I., 1997. Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation* 58, 99–120.
- Bailis, R., Ezzati, M., Kammen, D., 2005. Mortality and greenhouse gas impacts of biomass and petroleum energy futures in Africa. *Science* 308, 98–103.
- Blower, S., Hartel, D., Dowlatabadi, H., Anderson, R.M., May, R., 1991. Drugs, sex and hiv: a mathematical model for New York city. *Philosophical Transactions: Biological Sciences* 331 (1260), 171–187.
- Box, G., Hunter, W., Hunter, J., 1978. Statistics for Experimenters. An Introduction to Design, Data Analysis and Model Building. Wiley, New York.
- Cacuci, D.G., 2003. Theory. In: Sensitivity and Uncertainty Analysis, vol. 1. Chapman and Hall Publisher.
- Cacuci, D.G., Ionesco-Bujor, M., 2004. A comparative review of sensitivity and uncertainty analysis of large-scale systems-II: statistical methods. *Nuclear Science and Engineering* 147, 204–217.
- Campbell, J.E., Carmichael, G.R., Chai, T., Mena-Carrasco, M., Tang, Y., Blake, D.R., Blake, N.J., Vay, S.A., Collatz, G.J., Baker, I., Berry, J.A., Montzka, S.A., Sweeney, C., Schnoor, J., Stanier, C.O., 2008. Photosynthetic control of atmospheric carbonyl sulfide during the growing season. *Science* 322, 1085–1088.
- Campolongo, F., Tarantola, S., Saltelli, A., 1999. Tackling quantitatively large dimensionality problems. *Computer Physics Communications* 117, 75–85.
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software* 22, 1509–1518.
- Castangs, W., Dartus, D., Le Dimet, F.X., Saulnier, G.M., 2007. Sensitivity analysis and parameter estimation for the distributed modeling of infiltration excess overland flow. *Hydrology and Earth System Sciences Discussions* 4, 363–405.
- Cetnar, J., 2006. General solution of Bateman equations for nuclear transmutations. *Annals of Nuclear Energy* 33, 640–645.
- Coggan, J., Bartol, T., Esquenazi, E., Stiles, J., Lamont, S., Martone, M., Berg, D., Ellisman, M., Sejnowski, T., 2005. Evidence for ectopic neurotransmission at a neuronal synapse. *Science* 309, 446–451.
- Confalonieri, R., Bellocchi, G., Tarantola, S., Acutis, M., Donatelli, M., Genovese, G., 2010. Sensitivity analysis of the rice model WARM in Europe: exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environmental Modelling and Software* 25, 479–488.
- Da Veiga, S.D., Wahl, F., Gamboa, F., 2009. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics* 51, 439–451.
- de Gee, M., Lof, M., Hemerik, L., 2008. The effect of chemical information on the spatial distribution of fruit flies: II parametrization, calibration and sensitivity. *Bulletin of Mathematical Biology* 70, 1850–1868.
- Draper, N.R., Smith, H., 1981. Applied Regression Analysis. Wiley, New York.
- EC, 2009. Impact Assessment Guidelines, 15 January 2009. Technical Report 92, SEC. http://ec.europa.eu/governance/impact/docs/key_docs/iag_2009_en.pdf, 24 pp. (accessed 10.04.09).
- EPA, 2009. March. Guidance on the Development, Evaluation, and Application of Environmental Models. Technical Report EPA/100/K-09/003. Office of the Science Advisor, Council for Regulatory Environmental Modeling. http://www.epa.gov/crem/library/cred_guidance_0309.pdf, 26 pp. (accessed 10.04.09).
- Funtowicz, S., Ravetz, J., 1990. Uncertainty and Quality in Science for Policy. Springer, Dordrecht.
- Grievank, A., 2000. Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation. SIAM Publisher.
- Hasmadi, M., Taylor, J., 2008. Sensitivity analysis of an optimal access road location in hilly forest area: a GIS approach. *American Journal of Applied Sciences* 5 (12), 1686–1692.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Data Mining, Inference and Prediction. In: Springer Series in Statistics. Springer.
- Helton, J., Johnson, J., Sallaberry, C., Storlie, C., 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering and System Safety* 91 (10–11), 1175–1209.
- Hof, A., den Elzen, M., van Vuuren, D., 2008. Analysing the costs and benefits of climate policy: value judgements and scientific uncertainties. *Global Environmental Change* 18 (3), 412–424.
- Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety* 52 (1), 1–17.
- Jansen, M., 1999. Analysis of variance designs for model output. *Computer Physics Communications* 117, 35–43.
- Kennedy, P., 2007. A Guide to Econometrics, fifth ed. Blackwell Publishing.
- Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., Shah, N., 2009. Monte carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering and System Safety* 94, 1135–1148.
- Leamer, E., 1990. Let's take the con out of econometrics, and Sensitivity analysis would help. In: Modelling Economic Series. Clarendon Press, Oxford.
- Manache, G., Melching, C.S., 2008. Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters. *Environmental Modelling and Software* 23, 549–562.
- McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245.
- Mokhtari, A., Frey, H., Zheng, J., 2006. Evaluation and recommendation of sensitivity analysis methods for application to stochastic human exposure and dose simulation models. *Journal of Exposure Science and Environmental Epidemiology* 16, 491–506.
- Morris, M., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33 (2), 161–174.
- Murphy, J., Sexton, D., Barnett, D., Jones, G., Webb, M., Collins, M., Stainforth, D., 2004. Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature* 430, 768–772.
- Nolan, B., Healy, R., Taber, P., Perkins, K., Hitt, K., Wolock, D., 2007. Factors influencing ground-water recharge in the eastern united states. *Journal of Hydrology* 332, 187–205.
- Norton, J.P., 2008. Algebraic sensitivity analysis of environmental models. *Environmental Modelling and Software* 23, 963–972.
- Oakley, J., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society B* 66, 751–769.

- OMB, 2006, January. Proposed Risk Assessment Bulletin. Technical report. The Office of Management and Budget's/Office of Information and Regulatory Affairs (OIRA). http://www.whitehouse.gov/omb/inforeg/proposed_risk_assessment_bulletin_010906.pdf, 16–17 pp. (accessed 10.04.09).
- Pilkey, O., Pilkey-Jarvis, L., 2007. *Useless Arithmetic. Why Environmental Scientists Can't Predict the Future*. Columbia University Press, New York.
- Rabitz, H., 1989. System analysis at molecular scale. *Science* 246, 221–226.
- Ratto, M., Pagano, A., Young, P., 2007. State dependent parameter metamodeling and sensitivity analysis. *Computer Physics Communications* 177, 863–876.
- Sacks, J., Welch, W., Mitchell, T., Wynn, H., 1989. Design and analysis of computer experiments. *Statistical Science* 4, 409–435.
- Saisana, M., Saltelli, A., Tarantola, S., 2005. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society - A* 168 (2), 307–323.
- Saltelli, A., 2002. Making best use of model valuations to compute sensitivity indices. *Computer Physics Communications* 145, 280–297.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* 181, 259–270.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global Sensitivity Analysis. The Primer*. John Wiley and Sons.
- Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models. John Wiley and Sons Publishers.
- Saltelli, A., Tarantola, S., 2002. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of American Statistical Association* 97, 702–709.
- Santner, T., Williams, B., Notz, W., 2003. *Design and Analysis of Computer Experiments*. Springer-Verlag.
- Shorack, G.R., 2000. *Probability for Statisticians*. Springer-Verlag.
- Sobol', I.M., 1993. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment* 1, 407–414. Translated from Russian. Sobol', I.M., 1990. Sensitivity estimates for nonlinear mathematical models. *Matematicheskoe Modelirovani* 2, 112–118.
- Sobol', I.M., 1976. Uniformly distributed sequences with an addition uniform property. *USSR Computational Mathematics and Mathematical Physics* 16, 236–242.
- Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics* 7, 86–112.
- Stern, N., 2006. *Stern Review on the Economics of Climate Change*. Technical Report. UK Government Economic Service, London. www.sternreview.org.uk Technical Annex to the Postscript. http://www.hm-treasury.gov.uk/media/1/8/Technical_annex_to_the_postscript_P1-6.pdf.
- Stites, E., Trampont, P., Ma, Z., Ravichandran, K., 2007. Network analysis of oncogenic ras activation in cancer. *Science* 318, 463–467.
- Thogmartin, W.E., 2010. Sensitivity analysis of North American bird population estimates. *Ecological Modelling* 221 (2), 173–177.
- Turanyi, T., 1990. Sensitivity analysis of complex kinetic systems. Tools and applications. *Journal of Mathematical Chemistry* 5, 203–248.
- Van der Fels-Klerx, H., Tromp, S., Rijgersberg, H., Van Asselt, E., 2008. Application of a transmission model to estimate performance objectives for salmonella in the broiler supply chain. *International Journal of Food Microbiology* 128, 22–27.
- Varela, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. *Environmental Modelling and Software* 25 (3), 310–319.
- Weisberg, S., 1985. *Applied Linear Regression*. John Wiley & Sons.
- Zidek, J., 2006. Editorial: (post-normal) statistical science. *Journal of the Royal Statistical Society A* 169 (Part 1), 1–4.
- Ziehn, T., Tomlin, A.S., 2009. GUI-HDMR – a software tool for global sensitivity analysis of complex models. *Environmental Modelling and Software* 24, 775–785.