Risk Analysis, Vol. 22, No. 3, 2002

Sensitivity Analysis for Importance Assessment

Andrea Saltelli*

We review briefly some examples that would support an extended role for quantitative sensitivity analysis in the context of model-based analysis (Section 1). We then review what features a quantitative sensitivity analysis needs to have to play such a role (Section 2). The methods that meet these requirements are described in Section 3; an example is provided in Section 4. Some pointers to further research are set out in Section 5.

KEY WORDS: Uncertainty analysis; quantitative sensitivity analysis; computational models; assessment of importance; risk analysis

1. INTRODUCTION

Sensitivity analysis (SA) is the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input. The SA is hence considered by some as a prerequisite for model building in any setting, be it diagnostic or prognostic, and in any field where models are used. Kolb (quoted in Rabitz, 1989) noted that theoretical methods are sufficiently advanced, so that it is intellectually dishonest to perform modeling without SA. Fürbinger (1996) muses:

> Sensitivity analysis for modellers?
> Would you go to an orthopaedist who didn't use X-ray?

In our opinion, among the reasons for an increased role of SA in the scientific discourse is last decade's change in the role of science in society. Quantitative sensitivity analysis (QSA) is increasingly invoked for corroboration, quality assurance, and the defensibility of model-based analysis. Issues such as relevance and transparency become critical in this context, as we shall try to illustrate in this section.

According to Hornberger and Spear (1981), "most simulation models will be complex, with many parameters, state-variables and non linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be made to produce virtually any desired behavior, often with both plausible structure and parameter values." Examples of instrumental use of models can be found in the literature, especially when models are used for making decisions having a large social and economic impact. Thus, it is not surprising to meet cynical opinions about models. An example is in the journal *The Economist*, where one reads that "based largely on an economic model ... completing K2R4 [a nuclear reactor] in 2002 has a 50% chance of being 'least cost'" (*The Economist*, 1998). Given that the model was used to contradict a panel of experts on the opportunity to build the aforementioned reactor, *The Economist* comments:

> Cynics say that models can be made to conclude anything provided that suitable assumptions are fed into them.

The problem highlighted by Hornberger and Spear and illustrated by the example above is acutely felt in the modeling community. Economist Edward

* Address correspondence to Dr. Andrea Saltelli, Institute for the Protection and Security of the Citizen (IPSC), The European Commission, Joint Research Centre TP 361, 21020 ISPRA(VA) Italy; tel: +39 0 332 78 9686; fax: +39 0 332-78 5733; andrea.saltelli@jrc.it.

0272-4332/02/0600-0579$22.00/1

E. Leamer suggests the following: "I have proposed a form of organised sensitivity analysis that I call 'global sensitivity analysis' in which a neighbourhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighbourhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful" (Leamer, 1990). This awareness of the dangers implicit in selecting a model structure as true and working happily thereafter leads naturally to the attempt to map rigorously alternative model structures or working hypotheses into the space of the model predictions. The natural extension of this is the analysis of how much each source of uncertainty weights on the model prediction. One possible way to apportion the importance of the input factor with respect to the model output is to apply global QSA methods. Here the expression "global sensitivity analysis" takes on an additional meaning, with respect to that proposed by Leamer, in that a decomposition of the total uncertainty is sought. A combination of uncertainty and SA is a line of action that we have recommended in a number of review works, and in a multiauthor book on the subject published recently (Saltelli *et al.*, 2000).

Hornberger's concern is better known in the scientific community as the problem of the GIGO models (garbage in-garbage out).[1] There is apparently even an operative definition of a GIGO principle: "Precision of outputs goes up as accuracy of inputs goes down" (Stirling, 2000a). In other words, one way of GIGOing is to obtain precise outputs by arbitrarily restricting the input space.

Andrew Stirling studies "precautionary" and "science-based" approaches to risk assessment and environmental appraisal. In a recent work, which is a compilation of four different studies on the subject, he studies what the precautionary principle implies and how it can be operationalized (Stirling, 2000b). One of the recommendations he arrives at is "Express Analytical Results Using Sensitivity Analysis":

> It has been shown in this interim report that—in a variety of areas—risk assessment results are often presented with a very fine degree of numerical precision. Such a style conveys the impression of great accuracy, and distracts attention from the crucial question of the sensitivity of final results to changes in starting assumptions. This problem is particularly acute, where the values obtained—and even the ordering of different options—are quite volatile under the perspectives in appraisal associated with different social constituencies and economic interests. A practical and well-established way of dealing with such a problem lies in "sensitivity analysis"—a technique involving the explicit linking of alternative framing assumptions with the results which they yield. Rather than being expressed as discrete scalar numbers, then, risk assessment results might be expressed as ranges of values, with the ends of the ranges reflecting extremities in the framing assumptions associated with different stakeholders in the appraisal process.

Stirling introduces in this text the value-laden nature of different framing assumptions, which is a very crucial topic in present-day discourse on governance (see also Lemons *et al.*, 1997). We want to illustrate now how SA (or the lack of it) might impinge on the defensibility of a model-based analysis.

The Economist (2001) reported recently on the work of a team led by Daniel Esty of Yale University, with support from Columbia University, about a new Environmental Sustainability Index (ESI, 2001) produced on behalf of the World Economic Forum and presented to the 2001 annual Davos summit. This study contains a detailed assessment of dozens of variables that influence the environmental health of economies, producing an overall index that allows countries to be ranked. Mathis Wackernagel, mental father of the "Ecological Footprint" and thus an authoritative source in the sustainable development expert community, concludes a critique of the study done by Daniel Esty *et al.* by noting: "Overall, the report would gain from a more extensive peer review and a sensitivity analysis. The lacking sensitivity analysis undermines the confidence in the results since small changes in the index architecture or the weighting could dramatically alter the ranking of the nations" (Wackernagel, 2001). It is clear from this example that index numbers, such as ESI, can be considered as models. Tarantola *et al.* (2000) have shown how SA can be used to put an environmental debate on track by showing that the uncertainty in the decision on whether to burn or dispose solid urban waste depends on the choice of the index and not on the quality of the available data (e.g., emission factors).

Oreskes *et al.* (1994), in an article in *Science* entitled "Verification, Validation and Confirmation of Numerical Models in the Earth Sciences," puts SA in an apparently different context. The SA is not treated as a tool to build or improve a model, but

[1] Assuming one has already got rid of garbage in between, i.e., numerical or conceptual code errors.

instead represents one of the possible licit uses to which the model can be put. According to Oreskes, who takes a Popperian stance on the issue, natural systems are never closed and models put forward as description of these are never unique. Hence, models can never be "verified" or "validated" but only "confirmed" or "corroborated" by the demonstration of agreement (noncontradiction) between observation and prediction. Since confirmation is inherently partial, models are qualified by a heuristic value: models are representations, useful for guiding further study, but not susceptible to proof. Under Oreskes *et al.*'s point of view: "Models can corroborate a hypothesis. ... Models can elucidate discrepancies with other models. Models can be used for sensitivity analysis—for exploring 'what if' questions—thereby illuminating which aspects of the system are most in need of further study, and where more empirical data are most needed."

Modeling as craftsmanship is also the subject of Rosen (1991). We would like to end this introductory section with P. Høeg, a Danish novelist, who notes in his excellent *Borderliners*:

> That is what we meant by science. That both question and answer are tied up with uncertainty, and that they are painful. But that there is no way around them. And that you hide nothing; instead, everything is brought out into the open. (Høeg, 1995)

Høeg, like Oreskes, seems to think that uncertainty is not an accident of the scientific method, but its substance.

2. DESIRED PROPERTIES AND SETTINGS

Increasingly, the role of scientists in society is not so much of that of revealing truth as that of providing evidence, based on incomplete knowledge, sometimes in the form of probability, before systems of conflicting stakes and beliefs (Funtowicz *et al.*, 1996). This consideration, and the resulting need for the scientist to provide evidence that is defensible, poses some demand on the SA, which is one of the constituent elements of all model-based analyses. This is what has driven us in our choice of a methodology for SA.

2.1. Properties

We shall ignore here local methods, screening methods, regression-based methods, and others that can be found in the reference book mentioned above (Saltelli *et al.*, 2000). Our focus will be on methods that are global, quantitative, and model free,[2] capable of testing the robustness and relevance of a model-based analysis in the presence of uncertainties. Our choice is the variance-based methods, also known as importance measures or sensitivity indices. These methods provide a factor-based decomposition of the output variance and implicitly assume that this moment is sufficient to describe output variability. This may not be the case if one is interested, for example, in the tails of the output distribution. We list below what are the desirable properties of such methods.

- *Cope with the influence of scale and shape.* The influence of the input should incorporate the effect of the range of input variation and the form of its probability density function (pdf). It matters whether the pdf of an input factor is uniform or normal, and what are the distribution parameters.
- *Include multidimensional averaging.* In a perturbative approach to SA, one computes partial derivatives: the effect of the variation of a factor when all others are kept constant at the central (nominal) value. A global method should instead evaluate the effect of a factor while all others are varying as well.
- *Be model independent.* The method should work regardless of the additivity or linearity of the test model. A global sensitivity measure must be able to appreciate the so-called interaction effects, especially important for nonlinear, nonadditive models. These arise when the effect of changing two factors is different from the sum of their individual effects.
- *Be able to treat grouped factors as if they were a single factor.* This property of synthesis is essential for the agility of the interpretation of the results. One would not want to be confronted with a SA made of dense tables of input-output correlations.

At the same time we would like that the setting for the SA itself be as stringent as possible. It may well happen that using different measures of sensitivity, different experts obtain different relative ranking of the influence of the various input factors (see OECD/NEA PSAG User Group, 1993 for an

[2] In the sense of independent from assumptions about the model, such as linearity, additivity, and so on.

example). This happens if the objective of the analysis is left unspecified. In the same way as there are several definitions of risk (*Risk Newsletter*, 1987), there may be several definitions of importance. In the following, we shall offer two alternative rigorous settings for SA that will help us in our analysis (Saltelli and Tarantola, 2002).

Our point of departure is a mathematical or computational model $Y = f(X_1, X_2, \ldots, X_k)$, where some of the input factors are uncertain. We know something about their range of uncertainty. This knowledge might come from a variety of sources: measurements, expert opinion, physical bounds, analogy with factors for similar species, compounds, etc. This latter may be seen as a particular case of expert opinion. We may further have information (e.g., via observation) on the joint probability distribution of the factors.

The model may be used in a prognostic (forecast) or diagnostic (estimation) mode. In the former, all our knowledge about model input is already coded in the joint probability distribution of the input factors. In the latter, the input information constitutes *a priori* and the analysis might be aimed at updating either the distribution of the input factors or the model formulation based on the evidence.

A "forecast" mode for the model is assumed in the following unless otherwise specified. We select one among the many outputs produced by the given model and call this our output of interest. This might also be in the form of an averaged mean over more model outputs. The output of interest should be in the form a single quantity, possibly a scalar Y, whose value is taken as the top-most information that the model is supposed to provide. This could be, for instance, the ratio of an environmental pressure over the selected target value; it could be the maximum or averaged number of health effects in a given area and timespan; it could be the estimated failure probability for a system; and so on. We express this by saying that a sensitivity analysis should not focus on the model output as such, but rather on the answer the model is supposed to provide, on the thesis that it is supposed to prove or disprove. In $Y = f(X_1, X_2, \ldots, X_k)$, one does not need to assume f as constant, as it is customary to propagate uncertainty through different model structures or formulations. In this case some of the input factors are triggers that drive the selection of a structure versus another, and f stands for the computational code where all this takes place. Let us assume that we are able to compute the model output as much as we like, possibly sampling from the best joint probability distribution of input that we can come up with. This procedure is called by some a parametric bootstrap, in the sense that we sample with replacement the factors that enter into a model and reevaluate the model each time. Let us further assume for simplicity that each factor indeed has a true, albeit unknown, value. We know that often factors are themselves lumped entities called in as a surrogate for some more complex underlying process, but we now assume that they are simply data imprecisely known because of lack of sufficient observations. This clearly does not apply to stochastic uncertainties, such as the time of occurrence of an earthquake in a given area, although one might have frequency information for the area based on geological or historical records. Even in this case it is useful to think of the stochastic factor as possessing a true value for the sake of assessing its importance relative to all other factors. We can at this point introduce our first setting for SA.

2.2. Settings

2.2.1. Setting 1

The objective of SA is to identify the most important factor. This is defined to be the one that, if determined (i.e., fixed to its true, albeit unknown, value), would lead to the greatest reduction in the variance of the top statement Y. Likewise, one can define the second most important factor and so on till all factors are ranked in order of importance.

One might notice that we have made the concept of importance more precise, linking it to a reduction of the variance of the target function. It should also be noted that, in general, one would not be able to meet the objective of Setting 1, as this would imply knowing what the true value of a factor is. The purpose of Setting 1 is to allow a rational choice under uncertainty.

Another thing worth noting about Setting 1, which will be elaborated below, is that it assumes that factors are fixed one at a time. This will prevent the detection of interactions, i.e., in adopting Setting 1, we accept the risk of remaining ignorant about important features of the model that is the object of the SA. In this setting, the presence of interactions that would normally be something worth knowing about a model may escape the analysis altogether.

The ideal use for Setting 1 is for the prioritization of research, which is one of the most common uses of SA. Under the hypothesis that all uncertain factors are susceptible of determination, at the same cost per factor, Setting 1 allows the identification of the factor most deserving better experimental measurement.

A second setting that we have found useful when SA is part of a risk assessment study is the following. The objective of the analysis is the reduction of the variance of the target function Y from its unconditional value $V(Y)$ to a lower preestablished threshold value.

2.2.2. *Setting 2*

One must obtain a variance of Y equal or smaller than a given target variance $V_r < V(Y)$ by fixing simultaneously the smallest number of factors. Even in this case we have to make an informed choice without knowing where the true values of the factors lie.

Also, here we are only allowed to make an informed choice, rather than finding the optimum that would need the true factors' value to be known. Setting 2 allows factors to be fixed in groups, and the solution in this case can be influenced by the interactions among factors, if these are present.

We do not claim to have exhausted here all possible settings for SA. Our point is a different one. One setting must be defined for the analysis to be unambiguously implemented. Settings 1 and 2 will be sufficient for the purpose of the present review.

At the same time one should not forget that, as mentioned at the beginning of this section, we assume that one is interested in describing the output uncertainty in terms of its variance. In some decision contexts, there may be other measures that are more important, depending on the preferences of the decisionmaker (e.g., 95th percentile). Moreover, in some cases we may be concerned about shifts in central tendency of a model output attributable to an input factor, regardless of its contribution to the variance in the model. In OECD (1993), an analysis was performed by shifting the entire distribution of each factor a given fraction (5%), and the resulting shift in the model output was used to rank the factors. This approach has some drawbacks, as discussed in Saltelli and Tarantola (2002). It is insensitive to model nonmonotonicity and dependent on the fraction shift (5%) in the input distributions. Probably more thought should be given on how to shape the analysis on settings where the emphasis is not on the variance. Krykacz-Hausmann (2001) has criticized the use of variance as a measure of output uncertainty, and suggested using entropy H instead, defined as either $H(Y) = -\int f(y)\ln(f(y))$ or $H(Y) = -\sum p_i \ln(p_i)$ depending on whether the distribution of Y is continuous (f) or discrete (p). Krykacz-Hausmann's argument is that the largest uncertainty for Y should be that associated with a uniform distribution for Y in its range. By some intuitive examples, he shows that H is better than V in capturing this aspect of the problem.

3. METHODS

The previous discussion on settings should now help us introduce the recommended methods in a fairly natural way.

Let us put ourselves in Setting 1. We want to rank factors according to how much the unconditional variance $V(Y)$ of Y is reduced by fixing the various factors to their true value. The factors could then be ranked according to $V(Y|X_i = X_i^*)$, the variance being taken over all factors but X_i, or equivalently to $V(Y|X_i = X_i^*)/V(Y)$, where $V(Y|X_i = X_i^*)$ is the variance obtained by fixing X_i to its true value X_i^*. Note that $V(Y|X_i = X_i^*)$ could even be larger than $V(Y)$ for particular values of X_i^*. The problem is that we do not know what X_i^* is for each X_i. It will hence seem sensible to look at the factor with the smallest weighted average of the above measure over all possible values X_i^* of X_i, i.e., over $E(V(Y|X_i))$. We have dropped the dependence from X_i^* in the inner variance as this is eliminated by the outer mean.

Statistical science tells us that $V = V(E(Y|X_i)) + E(V(Y|X_i))$ so that betting on the highest $V(E(Y|X_i))$ is equivalent to betting on the lowest $E(V(Y|X_i))$.

Unsurprisingly, many practitioners of SA have come up with different estimates of $V_i = V(E(Y|X_i))$ as a measure of sensitivity (some have called V_i or $S_i = V_i/V(Y)$ importance measures, sensitivity indices, etc.). See Chan *et al.* (2000), and Saltelli *et al.* (1999, 2000) for reviews.

In conclusion, under Setting 1, $S_i = V_i/V(Y)$ is a proper measure of sensitivity to use to rank the input factors in order of importance even if the model is nonadditive and the input factors are correlated (Saltelli and Tarantola, 2002). The coefficients S_i are nicely scaled in $[0, 1]$.

Before we proceed with the method, we would like to illustrate another path to the same measure of importance. Imagine that we have just completed an estimation step in a model and that we have obtained for the factors under analysis a joint posterior distribution. We can now perform a parametric bootstrap, sampling with replacement from the joint pdf of the input and recomputing at each trial the sum of residuals between computed and experimental values of the output. The resulting scatter plots can look something like Fig. 1 from Planas and Depoutot (2000).

Clearly, we have more faith in the estimation of the factor THETA(1) than for factor LENGTH OF MONTHS. This is because the conditional variance of the target function Y, where Y is now sum of residuals between model boot-prediction and the best fit prediction, is small on average for all values of THETA(1), i.e., $E(V(Y|X_i))$ is much smaller for factor THETA(1) than for factor LENGTH OF MONTHS. We have found again the same importance measure in a context of parameter estimation. This shows that SA can tell us something about the quality of the estimation process. Scatter plots such as those in Fig. 1 have been used by Young (1999).

We can now close our digression on estimation and go back to our description of the methods.

When the input factors are independent, the conditional variances $V(E(Y|X_i))$ can be seen in the context of a general variance decomposition scheme proposed by Sobol (1990), whereby the total unconditional variance for a model with k factors can be decomposed as:

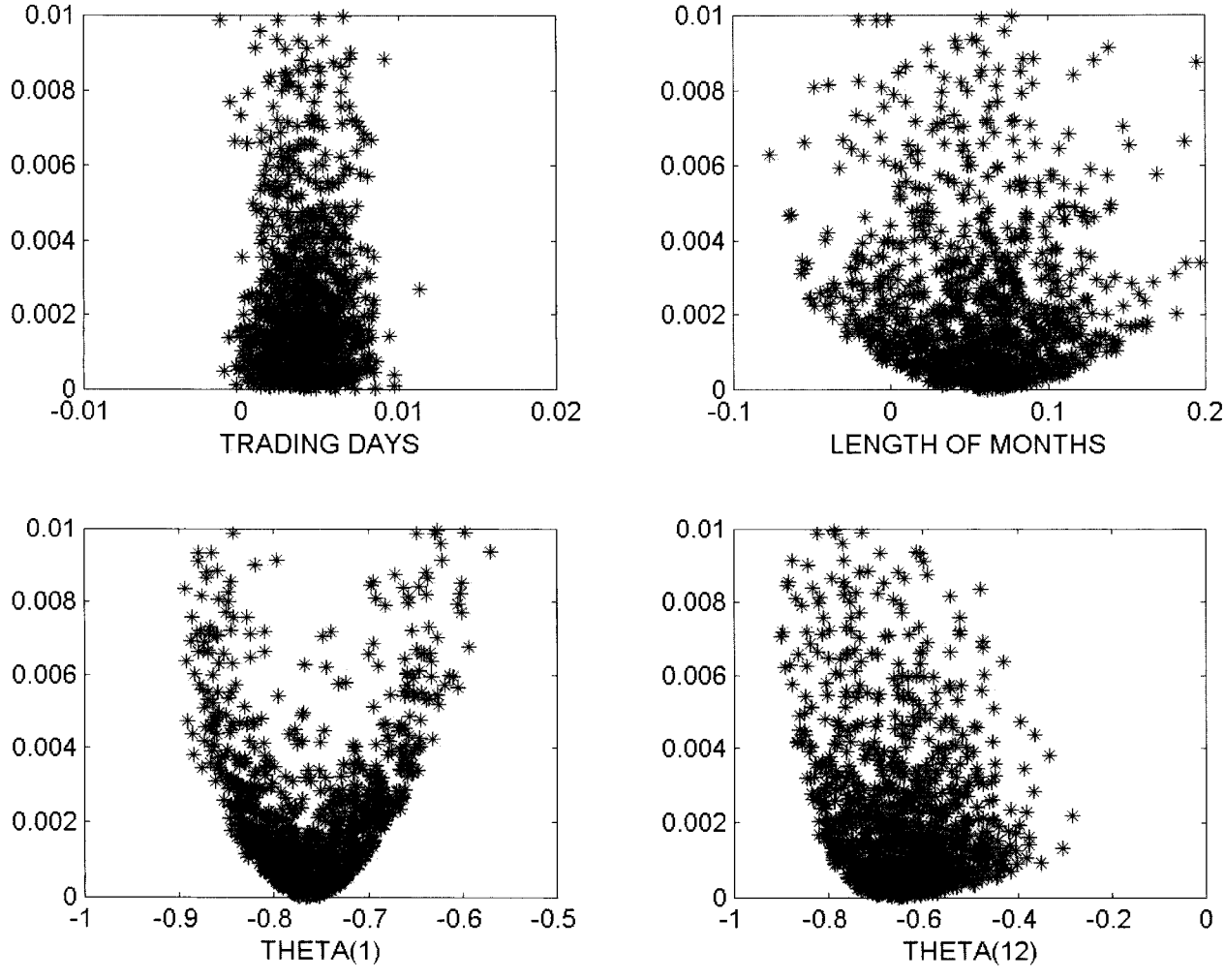

Fig. 1. Importance of input factors on trend estimates.

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12\ldots k}, \quad (1)$$

where

$$V_i = V(E(Y|X_i)), \quad (2)$$

$$V_{ij} = V\big(E(Y|X_i, X_j)\big) - V_i - V_j, \quad (3)$$

and so on. The development in Equation (1) contains k terms of the first order V_i, $k(k-1)/2$ terms of the second order V_{ij} and so on, till the last term of order k, for a total of $2^k - 1$ terms. The V_{ij} terms are the second-order (or two-way) terms, analogous to the second-order effects described in experimental design textbooks (see, e.g., Box *et al.*, 1978). The V_{ij} terms capture that part of the effect of X_i and X_j that is not described by the first order terms. Equation (1) has a long history, and various authors have proposed different versions of it. A discussion can be found in Archer *et al.* (1997), as well as in Rabitz *et al.* (1999, 2000). Sobol's decomposition is based on a decomposition of the function f itself into terms of increasing dimensionality, i.e.,

$$f(Y) = f_0 + \sum_i f_i + \sum_i \sum_{j>i} f_{ij} + \cdots + f_{12\ldots k}, \quad (4)$$

where each term is a function only of the factors in its index, i.e., $f_i = f_i(X_i), f_{ij} = f_{ij}(X_i, X_j)$ and so on. The decompositions in Equations (1) and (4) are unique provided that the input factors are independent and that the individual terms in $f_{i_1 i_2 \ldots i_s}$ are square integrable over the domain of existence.

One important aspect of Sobol's development is that similar decompositions can be written by taking the factors into subsets. Imagine that the factors have been partitioned into a trial set $\mathbf{u} = (X_{i_1}, X_{i_2}, \ldots X_{i_m})$, and the remaining set $\mathbf{v} = (X_{l_1}, X_{l_2}, \ldots X_{l_{k-m}})$. Then, according to Sobol, the total variance associated with $\mathbf{u}$ can be computed as

$$V(E(Y|\mathbf{u})) + V(E(Y|\mathbf{u}, \mathbf{v})) = V(Y) - V(E(Y|\mathbf{v})), \quad (5)$$

In Equation (4), $V(E(Y|\mathbf{u}))$ is the first-order effect of the set $\mathbf{u}$, while $V(E(Y|\mathbf{u}, \mathbf{v}))$ is the interaction term between the sets $\mathbf{u}$ and $\mathbf{v}$.

Before we proceed we need to introduce a new notation for a sensitivity measure closed within a subset of factors. Let us call it $V^c_{i_1 i_2 \ldots i_s}$, where the superscript c stands for closed, i.e., $V^c_{i_1 i_2 \ldots i_s}$ is the sum of all $V_{i_1 i_2 \ldots i_s}$ terms in Equation (1) that is closed in the indices $i_1, i_2, \ldots i_s$: $V^c_1 = V_1$, $V^c_{ij} = V_i + V_j + V_{ij}$, and so on. Likewise $V^c_{-i_1 i_2 \ldots i_s}$ will indicate the sum of all $V_{l_1 l_2 \ldots l_{k-s}}$ that are closed within the complementary set of $i_1, i_2, \ldots i_s$. Note that these also can be written in the usual Bayesian notation as, e.g., $V^c_{ij} = V\big(E\big(Y|X_i X_j\big)\big)$, see Equation (3).

We introduce now one last conditional variance (Homma and Saltelli, 1996), $V\big(E\big(Y|\mathbf{X}_{-j}\big)\big)$. This is the total contribution to the variance of Y due to non-X_j. This implies that the difference $V(Y) - V\big(E\big(Y|\mathbf{X}_{-j}\big)\big)$ is equal to the sum of all terms in the variance decomposition (Equation (1)) that include X_j. We illustrate this for the case $k = 3$:

$$S_1^T = \frac{V(Y) - V(E(Y|\mathbf{X}_{-1}))}{V(Y)} = \frac{E(V(Y|\mathbf{X}_{-1}))}{V(Y)} = S_1 + S_{12} + S_{13} + S_{123}, \quad (6)$$

where, e.g., $S_1 = V(E(Y|X_1))/V(Y)$, and analogous expressions can be written for S_2^T, S_3^T. We have called the S_j^Ts "total effect" terms. The total effects are useful for the purpose of SA, as discussed in Saltelli *et al.* (1999), as they give information on the nonadditive part of the model. It may be useful to observe here that for a purely additive model, $\sum_{i=1}^k S_i = 1$, whereas for a given factor X_j an important difference between S_j^T and S_j flags an important role of interactions for that factor in Y. Clearly, the same information could be obtained by computing all terms in Equation (1), but these are as many as $2^k - 1$. This problem has been referred to by Rabitz *et al.* (2000) as "the curse of dimensionality." For this reason we customarily tend to compute the set of all S_i plus the set of all S_i^T, which gives a fairly good description of the model sensitivities at the more reasonable cost. Many applications of this strategy to different models can be found in various chapters of Saltelli *et al.* (2000).

We are now posed to suggest a strategy to tackle problem Setting 2. This is particularly complex, especially for the general case where the input factors are not independent. The problem with correlated input, in brief, is that the reduction in variance that can be achieved fixing one factor depends on whether other factors have been fixed, and the incremental reduction in variance for each factor depends on the order in which factors are fixed. Equation (1) loses its uniqueness in this case. One can still compute closed variances such as

$V_{ij}^c = V(E(Y|X_iX_j))$, but this can no longer be decomposed as first order and interaction effects in a unique way.

In Saltelli and Tarantola (2002), we suggested the following empirical procedures for Setting 2.

3.1. Procedure 1: Case of Uncorrelated Input

We compute the full set of V_j s and V_{Tj} s and use the latter to rank the factors. A sequence V_{TR_1}, $V_{TR_2}, \ldots, V_{TR_k}$ is thus generated where $V_{TR_1} > V_{TR_2} > \cdots > V_{TR_k}$. If V_{TR_1} is greater than $V(Y) - V_r$, then the problem is solved. Otherwise, we take the factor with the second highest total index, i.e., V_{TR_2}. If $V_{R_1R_2}^c = V_{R_1} + V_{R_2} + V_{R_1R_2} > V(Y) - V_r$, then end the procedure, and so on.

This procedure has an alternative in a brute force search of all combination of factors yielding the desired reduction of variance, but this would again confront us with the curse of dimensionality. The procedure for the correlated case is more complex.

3.2. Procedure 2: Case of Correlated Input

This time we cannot compute the total effect indices, and thus we rank the single factors in order of importance using the values of the first-order terms V_j, obtaining a sequence V_{R_1}, $V_{R_2}, \ldots, V_{R_k}$, where $V_{R_1} > V_{R_2} > \cdots > V_{R_k}$. If $V_{R_1} > V(Y) - V_r$, then the problem is solved. Otherwise, we compute the second-order term $V(E(Y|X_{R_1}, X_j))$, where X_j, the second factor to be "fixed," has been selected on the basis of a "figure of merit" M_j that shall be defined below. $V(E(Y|X_{R_1}, X_j))$ gives us the reduction of the variance of $V(Y)$ that can be achieved by fixing the pair X_{R_1}, X_j. If $V(E(Y|X_{R_1}, X_j)) > V(Y) - V_r$ then the pair X_{R_1}, X_j solves our problem. Otherwise, a third factor X_m is selected using the same measure (we select the factor with the highest M_m), and the third-order index $V(E(Y|X_{R_1}, X_j, X_m))$ is computed. If $V(E(Y|X_{R_1}, X_j, X_m)) > V(Y) - V_r$ then the triplet X_{R_1}, X_j, X_m solves the problem, otherwise we continue in the same fashion.

The formula proposed for M_j in Saltelli and Tarantola (2002) is:

$$M_j = S(R_j)\left(1 - \max_{i \in u} |c_{ij}|\right)\left(1 + \frac{V_{Tj}^{NC} - V_j^{NC}}{V^{NC}}\right)^2, \tag{7}$$

where $S(R_j)$ is a Savage score (Savage, 1956)

$$S(R_j) = \sum_{r=R_j}^{k} \frac{1}{r}, \tag{8}$$

and u is the subset of the input factors that have already been fixed. The first two terms in the product of Equation (7) mean that the rank R_j of the candidate factor (obtained via the first-order terms) is first converted into a Savage score and then penalized by an amount related to the highest correlation figure involving that factor and those in the set u already fixed. c_{ij} is the correlation coefficient between factors X_j, X_i and is known *a priori*.

The last term contains the sensitivity coefficient (total and first order) for the noncorrelated case and the corresponding unconditional variance, hence the *NC* in the subscript. The rationale for this term is that a candidate factor for inclusion that interacts with one or more of the factors already fixed should be prized, as it could lead to a higher variance reduction than it is implied by its V_j.

The procedure and Equation (7) suggested above imply that besides computing terms such as V_j, $V(E(Y|X_{R_1}, X_j))$, and $V(E(Y|X_{R_1}, X_j, X_m))$, for the correlated case, we also compute the full set of first-order and total-order coefficients for the associated uncorrelated problem.

In conclusion, we see that for the uncorrelated case, a rational selection strategy for the subset of interest is based on the computation of the full sets of S_j and S_j^T. This strategy is meant to fight the curse of dimensionality, as attempting all combination of factors in a brute force search for the smallest subset of $\mathbf{X}$ that gives the desired reduction in $V(Y)$ would be computationally prohibitive; to start with, one would have to compute all $2^k - 1$ terms in Equation (1). The iterative procedure described above for the uncorrelated case includes as a step the computation of the full set of S_j and S_j^T.

For the correlated case, one might still engage in a brute force search computing all possible closed terms $V_{i_1 i_2 \ldots i_s}^c$. Note that for the correlated case the $V_{i_1 i_2 \ldots i_s}^c$ can no longer be decomposed meaningfully into a sum of lower dimensionality terms, but would still allow a perfectly informed choice, as would the $V_{i_1 i_2 \ldots i_s}$ in the uncorrelated case. Also for the correlated case, our suggested alternative involves the computation of the S_j and S_j^T for the noncorrelated problem.

We do not detail here how to estimate terms such as V_j, $V(E(Y|X_i, X_j))$, $V(E(Y|X_i, X_j, X_m))$, and hence S_j and S_j^T, etc. It will suffice to say that in

the case of noncorrelated input, accelerated computation procedures are available. For these, the reader is referred to the work of Sobol (1990) and Homma and Saltelli (1996), where Monte-Carlo-based strategies are offered. An alternative is the FAST method (Fourier Amplitude Sensitivity Test) in its extended version (Saltelli *et al.*, 1999). A review can be found in Chan *et al.* (2000). When the input is correlated, a different estimation procedure is available for the first-order terms $V(E(Y|X_j))$. This procedure (see McKay, 1995) uses the replicated Latin hypercube sampling design (r-LHS). For higher-order terms and correlated input a more laborious procedure is necessary that involves the explicit estimation of the multidimensional integral associated to terms such as $V(E(Y|X_i, X_j))$.

4. WORKED EXAMPLE

The Level E test case describes a problem of mass transfer governed by advection, dispersion, chemical retention, and radioactive decay in a multilayered medium (OECD/NEA PSAC User Group, 1989). This test case was used as a benchmark for computer codes used in the assessment of the performance of a geological disposal for radioactive waste. The Level E model is of medium complexity. It involves the numerical solution of a system of partial differential equations. The output variable considered in this study is the total annual dose to humans due to all the migrating radionuclides (^{129}I and the chain $^{237}Np \rightarrow {}^{233}U \rightarrow {}^{229}Th$).

The factors for the Level E exercise and their distribution were decided by a panel of experts who designed this exercise for the benchmark (OECD/NEA PSAC User Group, 1989). The factors are assumed independent, but a correlated version was generated for the purpose of Saltelli and Tarantola (2002). The Level E, with its strong nonmonotonic and nonadditivite nature, was used as a test model by several practitioners. It was thus instrumental in the development of the new methods for sensitivity analysis.

The radiological dose due to the four nuclides at $t = 10^5 y$ in the future is the quantity of interest in the study. The overall predictive uncertainty about that dose is due to uncertainties in model parameters (both intrinsic, such as the time of occurrence of a geological event, or due to our poor knowledge of the system). Twelve uncertain parameters are indeed taken into account in the simulation model, including relevant physical constants of the process, retention coefficients for radioactive species, length of pathways, and so on (Table I).

4.1. Cases

Four cases are sequentially tackled in Saltelli and Tarantola (2002):

1. Setting 1 and parameters assumed not correlated.
2. Setting 2 and parameters assumed not correlated.
3. Setting 1 and some parameters assumed correlated.
4. Setting 2 and some parameters assumed correlated.

4.1.1. Case 1

The first-order terms V_i are computed, which rank the factors in the order $v^{(1)}$, W, and $l^{(1)}$, with $S_i = V_i/V$ equal to 0.18, 0.04, and 0.02, respectively.

Table I. Input Factors for Level E

Notation	Definition	Distribution	Range	Units
T	containment time	uniform	/100, 1000/	*yr*
k_I	leach rate for Iodine	log-uniform	$/10^{-3}, 10^{-2}/$	yr^{-1}
k_C	leach rate for Np chain nuclides	log-uniform	$/10^{-6}, 10^{-5}/$	yr^{-1}
$v^{(1)}$	water vel. in geosphere's 1st layer	log-uniform	$/10^{-3}, 10^{-1}/$	*m/yr*
$l^{(1)}$	length of geosphere's 1st layer	uniform	/100, 500/	*m*
$R_I^{(1)}$	retention factor for I (1st layer)	uniform	/1, 5/	—
$R_C^{(1)}$	factor to compute ret. coeff. for Np (1st layer)	uniform	/3, 30/	—
$v^{(2)}$	water vel. in geosphere's 2nd layer	log-uniform	$/10^{-2}, 10^{-1}/$	*m/yr*
$l^{(2)}$	length of geosphere's 2nd layer	uniform	/50, 200/	*m*
$R_I^{(2)}$	retention factor for I (2nd layer)	uniform	/1, 5/	—
$R_C^{(2)}$	factor to compute ret. coeff. for Np (2nd layer)	uniform	/3, 30/	—
W	stream flow rate	log-uniform	$/10^5, 10^7/$	m^3/yr

The V_i for the other factors are negligible. The total cost of this analysis was of 8,192 times 12 (the number of factors) simulations using Sobol's method (1990).

4.1.2. Case 2

A 40% variance reduction has been used as a target. The total indices V_{Ti} are computed for the 12 factors and the three most important factors are $v^{(1)}$, W, and $l^{(1)}$. Fixing $v^{(1)}$ alone does not meet our target, so that the second-order term $V_{v^{(1)}W}$ must be estimated. The sum $S_{v^{(1)}} + S_W + S_{v^{(1)}W}$ is 0.32. That still does not give us the required 40% reduction in the variance. We then compute the two second-order terms $V_{v^{(1)}l^{(1)}}$, $V_{l^{(1)}W}$, and the third-order term $V_{v^{(1)}l^{(1)}W}$. The third-order term $V_{v^{(1)}l^{(1)}W}$ is found to be high, i.e., $V_{v^{(1)}l^{(1)}W}/V = 0.23$, and this allows our target to be met by simultaneously fixing the three factors $v^{(1)}$, W, and $l^{(1)}$. The cost of this analysis was 8,192 × 12 to compute the 12 pairs V_i, V_{Ti}, plus four times 8,192 to compute the four partial variances $V_{v^{(1)}W}, V_{v^{(1)}l^{(1)}}, V_{l^{(1)}W}$, and $V_{v^{(1)}l^{(1)}W}$.

Let us now consider Cases 3 and 4, where some of the input factors are correlated (Table II).

4.1.3. Case 3

Computing all the V_i shows that $v^{(1)}$ is no longer the most influential factor, and the two top-ranked factors are W and k_C (Table III). This may be due to the interplay between interaction and correlation. The cost of the analysis was of 1,000 runs using an efficient scheme suggested by McKay (1995).

4.1.4. Case 4

Here, the target variance reduction is again fixed at 40%. The ranking based on V_i yields W as the most important factor, but the average output variance reduction that we would obtain by fixing W is only 5% (Table III). The largest figure of merit M_i is obtained for $v^{(1)}$, and based on our Procedure 2 we compute $V\big(E\big(Y|W, v^{(1)}\big)\big)$, which corresponds to a variance reduction of about 28%. The factor k_c is hence included in the analysis and the term $V\big(E\big(Y|W, v^{(1)}, k_c\big)\big)$ is estimated. This time the target is reached, with a variance reduction of about 43%. Note that the sum of the first-order terms for the three selected factors is not high, while their cooperative effect is important. One-thousand runs were needed to compute the first-order terms, while two samples of size 10,000 were needed to compute $V\big(E\big(Y|W, v^{(1)}\big)\big)$, $V\big(E\big(Y|W, v^{(1)}, k_c\big)\big)$.

Table II. Correlations

Pairs of Correlated Factors	Correlation
k_I, k_C	0.5
$R_I^{(1)}, R_C^{(1)}$	0.3
$R_I^{(2)}, R_C^{(2)}$	0.3
$T, v^{(1)}$	0.7
$v^{(1)}, v^{(2)}$	0.5
$R_I^{(1)}, R_I^{(2)}$	0.5
$R_C^{(1)}, R_C^{(2)}$	0.5

Table III. Correlated Case: The V_i, the Ratios V_i/V, the Ranking on V_i, and the Merit Figures M_i are Given

Factor	$V_i * 10^{17}$	V_i/V	Ranking on V_i	M_i	V_i^{NC}/V	V_{Ti}^{NC}/V
T	7.17	0.01	4	1.27	0	0
k_I	0	0	6	0.08	0	0
k_C	22.1	0.031	2	2.10	0	0
$v^{(1)}$	18	0.026	3	4.51	0.18	0.86
$l^{(1)}$	0	0	6	0.39	0.02	0.51
$R_I^{(1)}$	0	0	6	0.61	0	0.49
$R_C^{(1)}$	0	0	6	0.41	0	0.03
$v^{(2)}$	0	0	6	0.68	0	0.15
$l^{(2)}$	5.1	~0	5	0.95	0	0.08
$R_I^{(2)}$	1.2	~0	5	0.79	0	0.10
$R_C^{(2)}$	6.8	~0	5	1.09	0	0.03
W	34.9	0.05	1	—	0.04	0.64

Note: In the last two columns, the ratios V_i^{NC}/V and V_{Ti}^{NC}/V for the uncorrelated set are also given, as they serve to compute the M_i.

5. FINAL REMARKS

We have mostly focused on a prognostic use of models. We would like to conclude this review work by coming back to a diagnostic use of models. There is in fact some resonance between the quantitative sensitivity approach defended here and Monte-Carlo-based approaches to model calibration presented in recent studies:

- *Monte Carlo (MC) filtering* (Rose *et al.*, 1991) is the process of rejecting sets of model simulations that fail to meet some prespecified criteria of model performance (acceptable behavior). This process can be presented as an objective method for model calibration: the subsets of model parameters that generate acceptable model simulations can be

regarded as equally satisfactory model calibrations. Contrary to the approach defended in the present article, Rose *et al.* (1991) selected the subset of factors for the analysis using a local SA method.

More radically, Fedra *et al.* (1981) present *MC filtering* as an alternative to the concept of local calibration. The analyst should refrain from searching for an optimal solution and instead be satisfied with the plausible ones.

- "*Regionalized (or Generalized) Sensitivity Analysis*" is a term used by Hornberger and Spear (1981), Spear *et al.*, (1994), and Young *et al.* (1996) to indicate the use of the outcome from a MC filtering experiment for SA purposes. The model output from the MC run is categorized into either *acceptable* (A) or *unacceptable* (B). The sets of model inputs that lead to acceptable behavior are then statistically compared with those that do not. Specifically, for a given input variable X, two subsets of possible values are identified: those that lead to acceptable behavior, subset X_A, and those that lead to unacceptable behavior, subset X_B. A statistical test of the hypothesis is then applied to check if the two subsets are samples from the same statistical distribution. An input variable is regarded as important when the generated sample distribution functions are statistically different. In our view, this approach has a limitation in that it takes into account only the output variation along the acceptable-unacceptable direction, while ignoring the variations of the output within the class of the acceptable values. In other words, an influent parameter could escape such an analysis only because it drives variation within the acceptable range.
- *Generalized Likelihood Uncertainty Estimation technique (GLUE)* is based on the concept of Bayesian inference for uncertainty estimation and has been developed from an acceptance of the possible equifinality of models, i.e., different sets of model parameters may be equally likely as simulators of the real system. It works with multiple sets of factors, typically via MC sampling. Model realizations are weighted and ranked on a likelihood scale via conditioning on observations. In practice, the likelihood for each realization in GLUE is inversely proportional to the sum of squared differences between data and model predictions in correspondence to a known set of model forcing functions (e.g., meteorological data). Based on the likelihood, weights are defined and used to formulate a cumulative distribution of the model output when the model is used in conjunction with a new set of forcing functions for which observations are not available. Thus, no true calibration is done in the GLUE approach; all parameter sets of the MC sample are used to compute the model output by simply applying weights.

We recently (Ratto *et al.*, 2001) performed quantitative sensitivity analysis (QSA) on the likelihood measure itself, thus coupling QSA and GLUE. The QSA allows a quantitative assessment of model factors mainly driving model behavioral runs. The use of GLUE, through the definition of a likelihood measure for each model run, allows the performance of a QSA conditioned to observations. The likelihood measure provides an estimate of the posterior joint pdf of the input factors and its analysis allows a description of the interaction structure between factors, connected to model overparameterization. The QSA allows a quantitative decomposition of the likelihood variance with respect to the input factors, including high-order terms. Factors providing negligible contributions to the likelihood variation can be clearly identified, allowing the modeler to exclude them from the calibration procedure and to fix them at a nominal value. On the other hand, factors having a significant impact on the likelihood measure (either as a main effect or as a total effect in interaction with all the other factors) have to be accounted for in calibration because they are able to drive behavioral runs of the model.

ACKNOWLEDGMENTS

Stefano Tarantola, who gave invaluable assistance in the revision of this manuscript, is gratefully acknowledged.

REFERENCES

Archer, G., A. Saltelli, and I. M. Sobol (1997), "Sensitivity measures, ANOVA like techniques and the use of bootstrap," *Journal of Statistical Computation and Simulation*, 58, 99–120.

Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978), *Statistics for experimenters*, New York, John Wiley and Sons.

Chan, K., S. Tarantola, A. Saltelli, and I. M. Sobol (2000), "Variance based methods," in *Sensitivity analysis*, A. Saltelli, K. Chan, and M. Scott, eds., 167–197, New York, John Wiley & Sons.

The Economist (1998), "More fall-out from Chernobyl," June 27, 98.

The Economist (2001), "Green and growing," Jan. 25, available at www.economist.com.

ESI (2001), available from http://www.ciesin.org/indicators/ESI/index.html.

Fedra, K., G. Van Straten, and M. B. Beck (1981), "Uncertainty and arbitrariness in ecosystems modeling: A lake modeling example," *Ecological Modelling*, 13, 87–110.

Funtowicz, S., M. O'Connor, S. Faucheux, G. Froger, and G. Munda (1996), "Emergent complexity and procedural rationality: Post-normal science for sustainability," in *Getting down to Earth: Practical applications of ecological economics*, R. Costanza, S. Olman, J. Martinez-Alier, O. Seguna, eds., Island Press.

Fürbinger, J. M. (1996), "Sensitivity analysis for modellers," *Air Infiltration Review* 17(4). Available at http://www.aivc.org/Air/17_4/sam.html.

Homma T., and A. Saltelli (1996), "Importance measures in global sensitivity analysis of model output," *Reliability Engineering and System Safety*, 52(1), 1–17.

Hornberger G. M., and R. C. Spear (1981), "An approach to the preliminary analysis of environmental systems," *Journal of Environmental Management*, 12, 7–18.

Høeg, P. (1995), *Borderliners*, 19, Toronto, McClelland-Bantam, Inc.

Krykacz-Hausmann, B. (2001), "Epistemic sensitivity analysis based on the concept of entropy," in *Proceedings of SAMO2001*, P. Prado and R. Bolado, eds., 31–35, Madrid, CIEMAT.

Leamer, E. E. (1990), "Let's take the con out of econometrics" and "Sensitivity Analysis would help," in *Modelling economic series*, C. W. J. Granger, ed., Oxford, Clarendon Press.

Lemons, J., K. Shrader-Frechette, and C. Cranor (1997), "The precautionary principle: Scientific uncertainty and type I and II errors," *Foundations of Science*, 2, 207–236.

McKay, M. D. (1995), *Evaluating prediction uncertainty*, NUREG/CR-6311, U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory.

OECD/NEA PSAC User Group (1989), *PSACOIN Level E intercomparison. An international code intercomparison exercise on a hypothetical safety assessment case study for radioactive waste disposal systems*, prepared by B. W. Goodwin, J. M. Laurens, J. E. Sinclair, D. A. Galson, and E. Sartori, Paris, OECD–NEA.

OECD/NEA PSAG User Group (1993), *PSACOIN Level S intercomparison. An international code intercomparison exercise on a hypothetical safety assessment case study for radioactive waste disposal systems*, Paris, OECD–NEA.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), "Verification, validation, and confirmation of numerical models in the earth sciences," *Science*, 263, 641–646.

Planas, C., and R. Depoutot (2000), "Sensitivity analysis for signal extraction in economic time series," in *Sensitivity analysis*, A. Saltelli, K. Chan, and M. Scott, eds., 293–307, New York, John Wiley & Sons.

Rabitz, H. (1989), "System analysis at molecular scale," *Science*, 246, 221–226.

Rabitz, H., and Ö. F. Ali (2000), "Managing the tyranny of parameters in mathematical modelling of physical systems," in *Sensitivity analysis*, A. Saltelli, K. Chan, and M. Scott, eds., 199–223, New York, John Wiley & Sons.

Rabitz, H., Ö. F. Ali, J. Shorter, and K. Shim (1999), "Efficient input-output model representations," *Computer Physics Communications*, 117, 11–20.

Ratto, M., S. Tarantola, and A. Saltelli (2001), "Sensitivity analysis in model calibration: GSA-GLUE approach," *Computer Physics Communications*, 136, 212–224.

Risk Newsletter (1987), "Defining Risk," *Risk Newsletter*, 7(3), 5.

Rose, K. A., E. P. Smith, R. H. Gardner, A. L. Brenkert, and S. M. Bartell (1991), "Parameter sensitivities, monte carlo filtering, and model forecasting under uncertainty," *Journal of Forecasting*, 10, 117–133.

Rosen, R. (1991), *A comprehensive inquiry into the nature, origin, and fabrication of life*, New York, Columbia University Press.

Saltelli A., and S. Tarantola (2002), "On the relative importance of input factor in mathematical models," to appear in *Journal of American Statistical Association*.

Saltelli, A., S. Tarantola, and F. Campolongo (2000), "Sensitivity analysis as an ingredient of modelling," *Statistical Science*, 15(4), 377–395.

Saltelli, A., S. Tarantola, and K. Chan (1999), "A quantitative, model independent method for global sensitivity analysis of model output," *Technometrics*, 41(1), 39–56.

Saltelli A., K. Chan, and M. Scott, eds. (2000), *Sensitivity analysis*, New York, John Wiley & Sons.

Savage, I. R. (1956), "Contributions to the theory of rank order statistics—The two-sample case," *Annals of Mathematical Statistics*, 27, 590–615.

Sobol, I. M. (1990), "Sensitivity estimates for nonlinear mathematical models, *Matematicheskoe Modelirovanie*, 2, 112–118 (translated as I. M. Sobol (1993), "Sensitivity analysis for non-linear mathematical models," *Mathematical Modelling and Computational Experiment*, 1, 407–414).

Spear R. C., T. M. Grieb, and N. Shang (1994), "Factor uncertainty and interaction in complex environmental models," *Water Resources Research*, 30, 3159–3169.

Stirling, A. (2000a), as quoted by Jerry Ravetz, personal communication.

Stirling, A. (2000b), "On science and precaution in the management of technological risk," a synthesis report of studies conducted by O. Renn, A. Rip, A. Salo, and A. Stirling, Final Report of a project for the EC unit of forward studies (FSU), SPRU, University of Sussex.

Tarantola, S., J. Jesinghaus, M. Puolamaa (2000), "Global sensitivity analysis: A quality assurance tool in environmental policy modelling," in *Sensitivity Analysis*, A. Saltelli, K. Chan, and M. Scott, eds., 385–397, New York, John Wiley & Sons Publishers.

Young, P. (1999), "Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis," *Computer Physics Communication*, 117, 113–129.

Young, P. C., S. D. Parkinson, and M. Lees (1996), "Simplicity out of complexity: Occam's razor revisited," *Journal of Applied Statistics*, 23, 165–210.

Wackernagel, M. (2001), "Shortcomings of the Environmental Sustainability Index," *Redefining Progress*, March 6, 2001, available at http://www.rprogress.org/about/.