Statistical Science
2000, Vol. 15, No. 4, 377–395

Sensitivity Analysis as an Ingredient of Modeling

A. Saltelli, S. Tarantola and F. Campolongo

Abstract. We explore the tasks where sensitivity analysis (SA) can be useful and try to assess the relevance of SA within the modeling process.

We suggest that SA could considerably assist in the use of models, by providing objective criteria of judgement for different phases of the model-building process: model identification and discrimination; model calibration; model corroboration.

We review some new global quantitative SA methods and suggest that these might enlarge the scope for sensitivity analysis in computational and statistical modeling practice. Among the advantages of the new methods are their robustness, model independence and computational convenience.

The discussion is based on worked examples.

Key words and phrases: Global sensitivity analysis, quantitative sensitivity measure, screening, numerical experiments, predictive uncertainty, reliability and dependability of models, model transparency.

1. INTRODUCTION

We illustrate here some uses of uncertainty and sensitivity analysis (SA), by anticipating some of our worked examples.

Our first example (Figure 1) is a study of model quality assessment in environmental policy, based on data of the European Environment Agency (EEA) for Austria in 1994. The model assists a hypothetical policy maker in selecting, on the basis of their environmental impact, between incineration and landfill for the disposal of solid waste. The model reads emission data from the EEA-CORINAIR database and supplies a pressure index (PI) for each policy option. The PI is proportional to the overall hazard to the environment of the corresponding option. The model output Y, called the pressure-to-decision index, is defined as a suitable combination of these PI's and is aimed at quantifying the convenience of disposing Austrian solid waste by landfill versus disposal by incineration. The model for Y is described in detail in Section 3.

A Monte Carlo input generation process is taken to estimate the distribution function of Y (Figure 1). Several input factors are sampled from distributions derived from the literature or expert opinion. Also sampled are "structural" modeling elements, such as the choice of the system of environmental indicators used to build the pressure indices.

Figure 1 represents an example of Monte Carlo–based uncertainty analysis, whereby we can appreciate the following:

- the bimodal nature of the output;
- the fact that apparently no clear choice is dictated by the model Y;
- the fact that, according to the captions, incineration is preferred by the EUROSTAT set of indicators while landfill is supported by the Finnish set.

Figure 2 represents a decomposition of the variance of the output Y of interest according to source. The decomposition procedure will be detailed in Section 2. It suffices here to note that various groups of simulated input (sampled from their respective pdf's) have different impacts on the variance of Y. In particular, as mentioned before, we

A. Saltelli is Doctor and Head of Applied Statistics (APPST) sector, Institute for Systems, Informatics and Safety, European Commission, Joint Research Centre, TP 361, 21020 *Ispra (VA), Italy. S. Tarantola is Doctor, Institute for Systems, Informatics and Safety, European Commission, Joint Research Centre, TP* 361, 21020 *Ispra (VA), Italy. F. Campolongo is Doctor, Institute for Systems, Informatics and Safety, European Commission, Joint Research Centre, TP* 361, 21020 *Ispra (VA), Italy.*

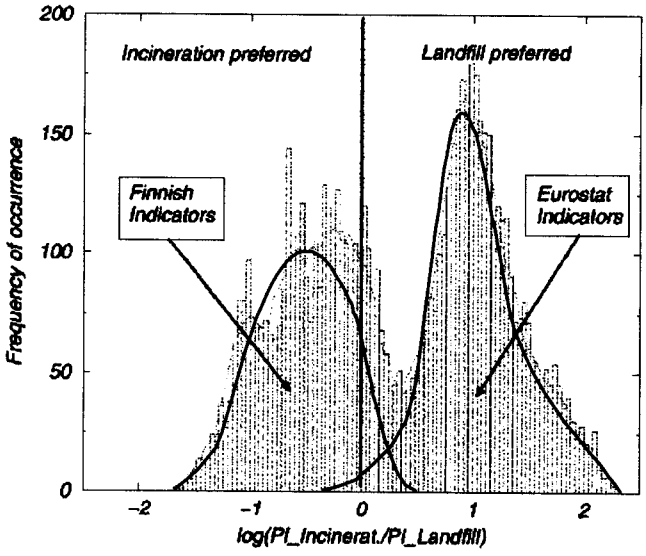

FIG. 1. *Uncertainty analysis for the environmental indicators test case.*

now see clearly how much of the variation in Y is driven by the choice of the set of indicators (the factor [E/F]).

Figure 2 is a sensitivity analysis. It complements Figure 1 (uncertainty analysis) by coloring the grey uncertainty area around the sample mean for Y with colors that identify the contributors to the uncertainty (the variance) of Y. Roughly speaking, the slices of the pie in Figure 2 give an idea of how much the variation of Y could be reduced if we could fix some of the uncertain inputs.

An analyst would read Figure 2 as: "At the present stage, money should not be spent to improve the quality of the data (e.g., emission factors), but to reach a consensus among experts about standard indicator systems."

Our next example is a case of model building under uncertainty (Saltelli and Hjorth, 1995). Figure 3 shows a chemical reaction scheme for the oxidation of a climatically active trace gas (dimethyl sulphide). The scheme in the figure is to a large extent speculative; that is, it reflects the present understanding of the system favored by a particular team of investigators. Uncertain are the value of the kinetic coefficients involved and their temperature dependence. Also uncertain is the existence of some of the reaction branches. The model KIM computes time histories for all trace gases in the system given the input thermodynamic and kinetic data (the k's) and the initial conditions. Because the uncertainty in the k's is large and can span orders of magnitude, a Monte Carlo analysis is also performed here. For each chemical species and time point, the analyst can build a histogram (uncertainty analysis) as well as a pie chart (sensitivity analysis).

One of the questions asked of the KIM model in Saltelli and Hjorth (1995) was whether one of the branches of reaction (k_{23}) was making any significant contribution to the direct formation of sulphate, by direct decomposition of an intermediate ($CH_3SO_3\cdot$) to form SO_3. Because all k's are uncertain, it is not sufficient to explore this system via what-if simulations, changing only the value of k_{23}. The relevance of k_{23} as a pathway also depends upon

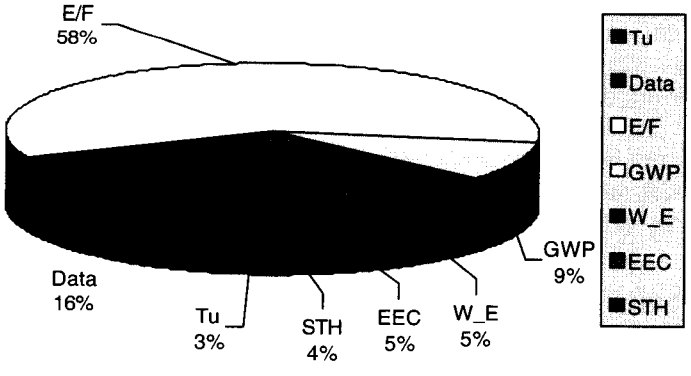

FIG. 2. *Results of SA for the first exercise: both the Finnish and the EUROSTAT sets of indicators have been used, driven by the factor [E/F] acting as a switch. The other (groups of) factors, and their dimensions, are as follows: TU, territorial unit, two levels of spatial aggregation for collecting data are possible (1 factor to select which one); DATA (in total 176 factors), made up of activity rates (120 factors), emission factors (37 factors) and national emissions (19 factors); GWP, weight for greenhouse indicator (in the Finnish set), 3 time-horizons are possible (1 factor to select which one); W_E, weights for EUROSTAT indicators (11 factors); EEC, a single factor that selects the approach for evaluating environmental concerns [Target values (Adriaanse, 1993) or expert judgement (Puolamaa, Kaplas and Reinikainen, 1996)]; STH, a single factor that selects one class of Finnish Stakeholders from a set of 8 possible classes.*

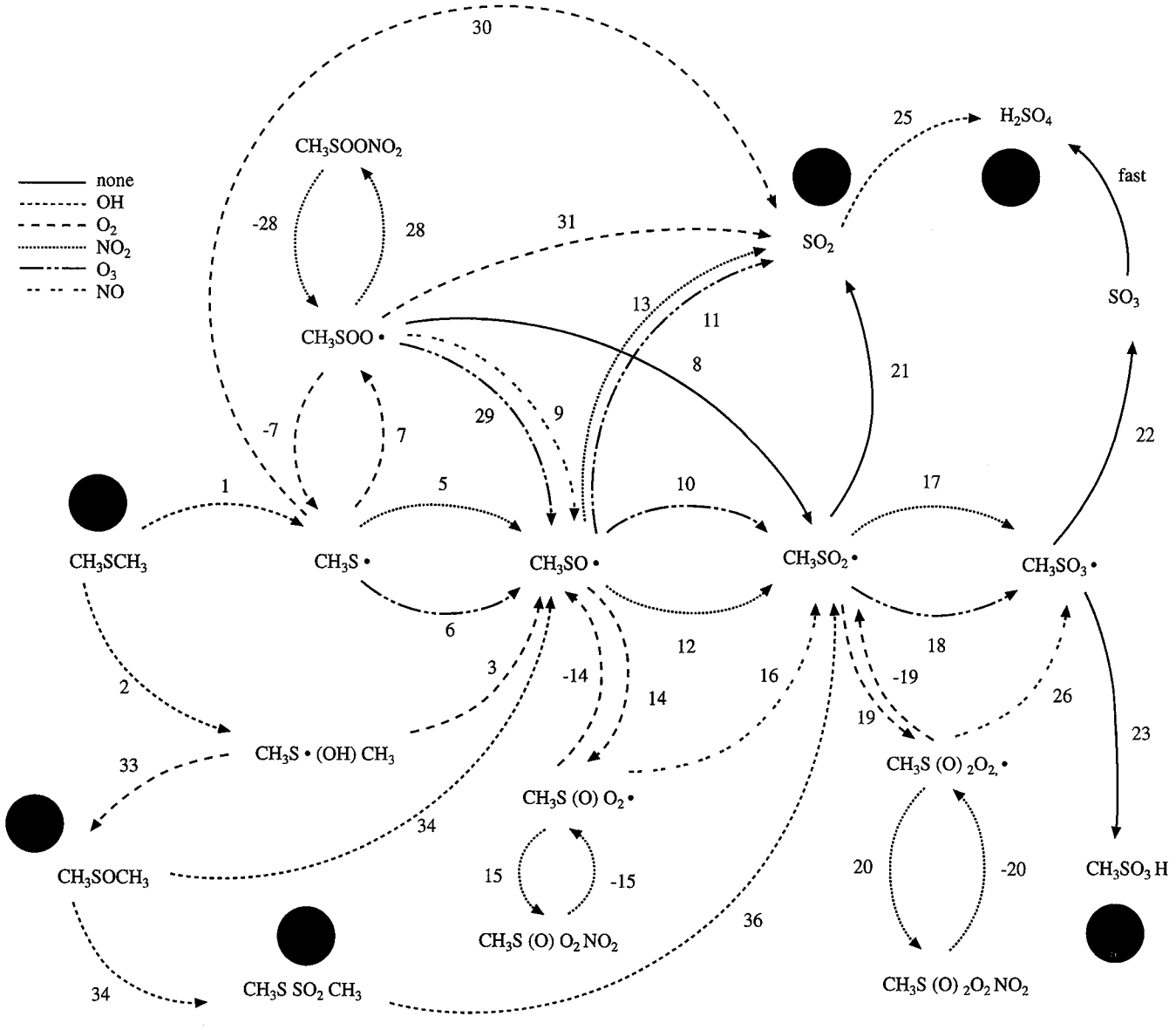

FIG. 3. *Scheme of KIM: the spheres indicate where the chemical species can diffuse to water droplets, where liquid phase chemistry is at play. The scheme for the liquid phase chemistry is not given here.*

coefficients that are upstream (such as k_{12}, k_{17}) or are competing (such as k_{21}). A Monte Carlo analysis where all k's were varied simultaneously allowed the relevance of k_{23} to be assessed globally. Further, sensitivity analysis allowed the identification of the most influential k's, so that further simulations could be targeted to the objective to ascertain under what conditions (what values of the influential k's) could k_{23} become an important pathway for the formation of sulphate. In the end it was found that this pathway is not an efficient one for sulphate production.

Our last example is the level E model. This model is slightly more complex than KIM, being made up of partial differential equations (PDE) rather than ordinary differential equations (ODE) and describes mass transfer with chemical reaction in a porous medium. The migrating species are radioactive, and the variable of interest Y is the total dose to man resulting from all of the radionuclides, once these reach the biosphere (Section 3). Figure 4e shows how the total normalized variation of Y can be decomposed into two sets of uncertain input factors:

- those pertaining to the engineered barriers (such as the lifetime of the steel canister containing the radionuclides);
- the natural barriers (such as the strength of the chemical interaction between nuclides and the porous medium).

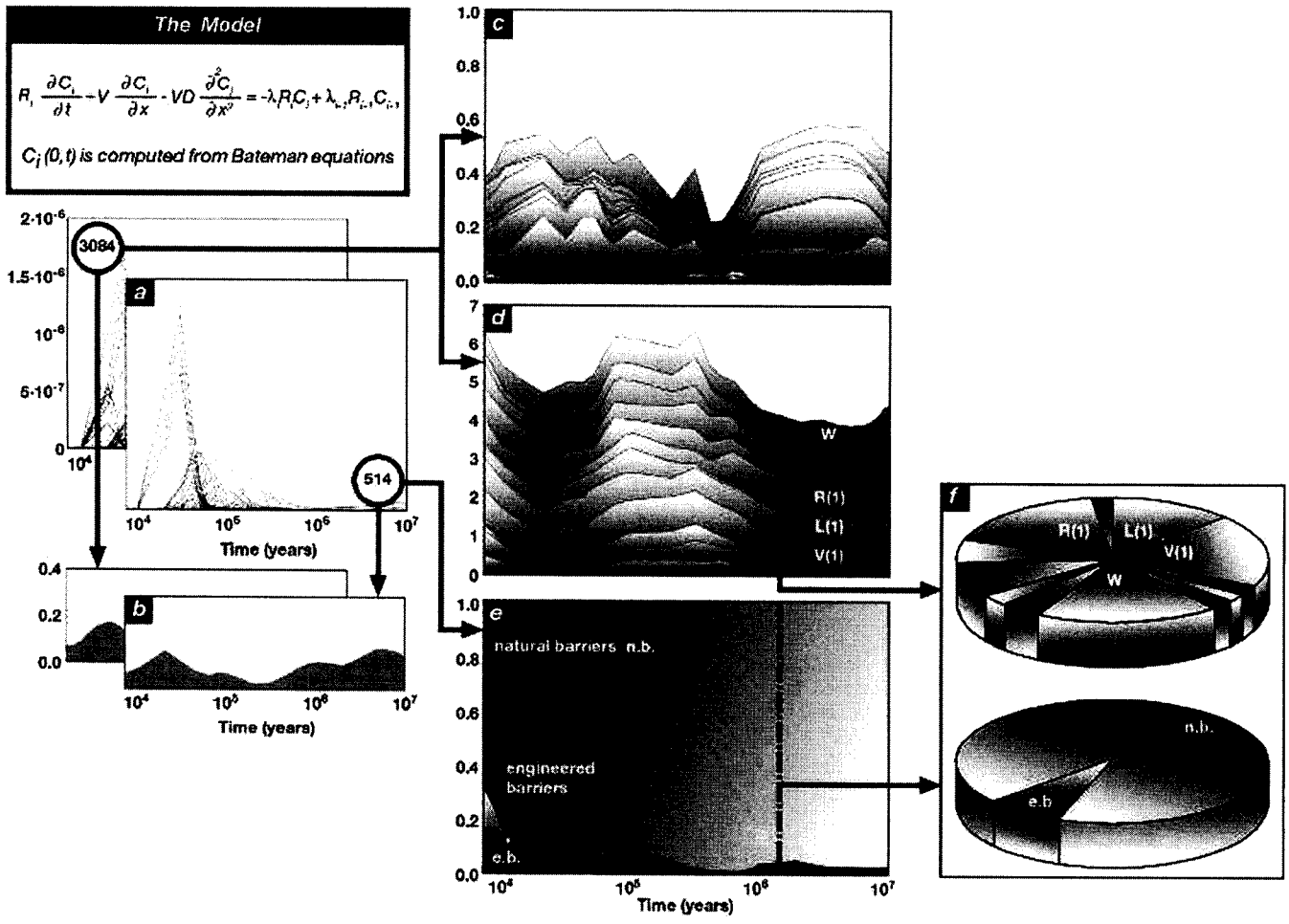

FIG. 4. *Level E test case: the panels display the results as functions of time (years).* (a) *Each of the two curves represents the uncertainty analysis on the total dose, performed on a set of* 514 *and* 3,084 *model evaluations, respectively.* (b) *The* R_y^2 *estimate as obtained from the set of model runs.* (c) *and* (d) *Time-dependent pie charts of the* S_i *and the* S_{Ti} *for the* 12 *factors of the model, as obtained from the set of* 3,084 *model runs.* (d) *The four most influential factors.* (e) *The normalized* S_{Ti} *for an analysis conducted on two groups of factors, "natural barriers" and "engineered barriers."* (f) *Pie chart of the results of the sensitivity analysis at a given time.*

Also in this case the variance of Y has been decomposed using quantitative methods. The analysis suggests (as practitioners know) that the safe containment of the nuclides over a long time span depends more on the geology and geochemistry of the host formation, and less on the characteristics of the disposal design. An upgraded version of the level E model is treated in Draper et al. (1999), where different scenarios for the release of radionuclides are visited. In this latter exercise, sensitivity analysis allowed the importance of scenario uncertainty to be contrasted with factor uncertainty (Section 3).

The three examples above, and another presented in Section 3, will be invoked to illustrate and defend the following theses:

- Uncertainty and sensitivity analysis are an integral part of the modeling process.
- Quantitative sensitivity analysis methods, able to decompose the variance of Y, are useful and easy to interpret. They contribute to the transparency of the analysis.
- Sensitivity analysis can be employed also when dealing with uncertain or competing model structures or scenarios, treating the choice of the model as one of the sources of uncertainty. This process contributes to the evaluation of model-based inference.

In Section 2, we review some newly developed global quantitative methods for sensitivity analysis. In Section 3, our examples are presented and the methods are applied to the examples. In Section 4 we discuss the examples and we also try to contrast the variance-based methods introduced in Section 2 with other available approaches.

2. METHODS

We introduce first the properties of the global methods for sensitivity analysis that are the primary focus of this paper (Section 2.1). In Section 2.2, we give computational details on two methods: Sobol' and extended FAST. In Section 2.3 we contrast these methods against other available techniques for sensitivity analysis.

2.1 Global Quantitative Methods

Partitioning the variance V of the objective function Y is one possible way of performing sensitivity analysis. To the reader of *Statistical Science* this might recall classic regression analysis. If we have generated via Monte Carlo a set of model evaluations $y_i, i = 1, \ldots, N$, corresponding to N different sampled values $\mathbf{x_i}$ of the vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ of input factors, then we can build a regression model for Y using standard least squares analysis. The standardized regression coefficients β_j are a way to measure the sensitivity of Y to the factors X_j. If the factors are independent, $\sum_j \beta_j^2 = V$; that is, the β_j's give the fractional contribution of each factor to the variance of Y. However, the effectiveness of the β_j's as sensitivity measures is judged by the model coefficient of determination

$$R_y^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^N (y_i - \bar{Y})^2}. \tag{1}$$

If R_y^2 is close to 1, then the regression model is accounting for most of the varability in the y_i, and the β_j can be used to gain insight into the relative importance of the input factors. Conversely, low values for R_y^2 suggest that the model has nonlinear behavior. Regression coefficients are described in Draper and Smith (1981) and their application to sensitivity analysis is reviewed by Helton (1993). In the rank-based version of the standardized regression coefficients, both the input and the output values are replaced by their ranks (Iman and Helton, 1988). Rank-based β_j's can be used for the purpose of model sensitivity analysis for nonlinear, albeit monotonic, models. A drawback of rank-based sensitivity measures is that the properties of the model Y are grossly distorted by the rank transformation (Saltelli and Sobol', 1995).

Another class of sensitivity measures that also express fractional contributions to the total (or unconditional) variance of Y are those obtained from extending classic experimental design (DOE, Box, Hunter and Hunter, 1978) to simulation experiments (Kleijnen, 1998). By varying the factors among a set of preestablished values (levels), and running the model with selected multivariate samples of the factors (a design), one obtains output values that permit the estimation of individual effects of the factors (main-order terms) as well as of the interaction effects (second- and higher-order terms).

Experimental designs can provide an estimate of the first-order fractional variance $V[E(Y|X_i = x_i^*)]$, where $E(Y|X_i = x_i^*)$ denotes the expectation of Y conditional on X_i having a fixed value x_i^*, and the variance is taken over all possible values of x_i^*. The conditional expectation is taken over all X_j, $j \neq i$, with X_i fixed to a given value x_i^* (it involves hence an integration over $k-1$ dimensions), and the fractional variance V is only computed over all the possible values of x_i^* (integration over one dimension).

The usefulness of $V[E(Y|X_i = x_i^*)]$ as a measure of the sensitivity of Y to X_i is easy to grasp. An influential factor X_i will be associated with a small value of $V(Y|X_i = x_i^*)$ (i.e., fixing X_i to x_i^* reduces appreciably the variance of Y), as well as to a small value of $E[V(Y|X_i = x_i^*)]$, the expected reduced variance that would be achieved if X_i could be fixed. Because

$$\begin{aligned} V(Y) &= V\big[E(Y|X_i = x_i^*)\big] \\ &\quad + E\big[V(Y|X_i = x_i^*)\big], \end{aligned} \tag{2}$$

a small $E[V(Y|X_i = x_i^*)]$ is equivalent to a large $V[E(Y|X_i = x_i^*)]$ and the latter can be used as a sensitivity measure. Estimation procedures for $V[E(Y|X_i = x_i^*)]$ are discussed in Section 2.2.

The $V[E(Y|X_i = x_i^*)]$ estimate is a more general sensitivity measure than a regression coefficient, as it also works for nonlinear models. Yet this measure does not account for interactions among factors. $V[E(Y|X_i = x_i^*)]$ is in fact the main effect term of X_i on Y. In a full ANOVA-like decomposition, the total variance V of the output is apportioned to all the effects; that is, it is decomposed as a sum of terms of increasing dimensionality:

$$\begin{aligned} V(Y) &= \sum_i V_i + \sum_{i<j} V_{ij} \\ &\quad + \sum_{i<j<m} V_{ijm} + \cdots + V_{12\cdots k}, \end{aligned} \tag{3}$$

where $V_i = V[E(Y|X_i = x_i^*)]$, $V_{ij} = V[E(Y|X_i = x_i^*, X_j = x_j^*)] - V[E(Y|X_i = x_i^*)] - V[E(Y|X_j = x_j^*)]$ and so on. Decomposition (3) has a long history; see Archer, Saltelli and Sobol' (1997) and Rabitz, Alĭ, Shorter and Shim (1999) for a discussion. Formula (3) only holds if the factors X_i are independent of each other. As mentioned before, straightforward variance decomposition for linear additive models is provided by regression analysis, using the standardized regression coefficients. In

particular, for linear models, $\beta_j^2 = V_i/V$ (McKay, 1996, Saltelli and Bolado, 1998). The first-order terms in (3) can be estimated for the purpose of sensitivity analysis using methods such as the Fourier amplitude sensitivity test (FAST Cukier et al., 1973; see Saltelli, Tarantola and Chan, 1999, for a review), the method of Sobol' (1990) and others (Iman and Hora, 1990; McKay, 1995, 1996). All of these methods are referred to as variance-based in our following discussion.

It is a fact known to both modelers and experimentalists that the number and importance of the interaction terms usually grows (a) with the number of factors and (b) with the range of variation of the factors. This means that if we were able to compute all of the V_i terms, then most likely $\sum_{i=1}^{k} V_i$ would still be lower than the total variance of Y. The difference $V(Y) - \sum_{i=1}^{k} V_i$ is a measure of the impact of the interactions. Liepmann and Stephanopoulos (1985) have argued that when the summation of the first-order terms is as high as 65% of the total variance V, one should consider the analysis as satisfactory. This rule of thumb was forced into the practice of sensitivity analysis by the difficulty of computing the interaction terms, even when using straightforward design of experiment (DOE) methods. If a model has k factors, the total number of terms in (3) (including the first-order ones) is as high as $2^k - 1$. This problem has been referred to as the curse of dimensionality by Rabitz et al. (1999). This curse can affect both experimentalists preparing an experimental design and modelers, although for the latter the problem is more acute. In models one wants to explore more factors than one can control in field or laboratory experiments. Factors are varied on a wider scale in a numerical experiment than they are in a physical one. Unfortunately, increasing the number of factors, and their ranges of variation, commonly results in more and larger interactions. All this should suggest that, to be of practical use, a global sensitivity analysis method should be able to cope with the dimensionality problem.

In fact such methods have been recently developed. Among these, the extended FAST method (Saltelli, Tarantola and Chan, 1999) and the method of Sobol' (Sobol', 1990; Homma and Saltelli, 1996) are capable of estimating the total sensitivity index S_{Ti}, defined to be the sum of all effects (first- and higher-order) involving factor X_i. For example, in a three-factor model, the three total effect terms are

$$\begin{aligned} S_{T1} &= S_1 + S_{12} + S_{13} + S_{123}, \\ S_{T2} &= S_2 + S_{12} + S_{23} + S_{123}, \\ S_{T3} &= S_3 + S_{13} + S_{23} + S_{123}, \end{aligned} \tag{4}$$

where now each $S_{i_1, i_2, \ldots, i_s}$ is simply $V_{i_1, i_2, \ldots, i_s}/V$. Direct estimates of S_{Ti} can be obtained, which do not require the estimation of the individual terms in (4), thus making the computation affordable.

Whereas the S_i can be defined as the expected fractional reduction in variance that would be achieved if X_i were known, the S_{Ti} can be thought of as the expected fraction of variance that would be left if only X_i were to stay undetermined. Let X_{-i} be the vector made up of all x_j, $j \neq i$, and let x^*_{-i} be a specified value of X_{-i}. It can be proven that $S_{Ti} = E[V(Y|X_{-i} = x^*_{-i})]/V$. The terms "bottom marginal variance" and "inclusive variance" have been used for S_{Ti}; see Jansen et al. (1994).

Whereas the sum of the individual effect terms of all orders add up to 1, the sum of the S_{Ti}'s is in general larger than 1, because each interaction of order r is counted r times in it.

2.2 Computational Methods

We now offer some computational details on the extended FAST and Sobol' methods.

In extended FAST (Saltelli, Tarantola and Chan, 1999), the S_{Ti}'s are evaluated by a search curve that scans the space of the input factors, in such a way that each factor is explored with a selected integer frequency. Each uncertain input factor X_i is associated with a frequency ω_i, and a set of standardized parametric equations

$$\begin{aligned} X_i(s) &= G(\sin \omega_i s) \\ &= \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin \omega_i s) \end{aligned} \tag{5}$$

allows each factor to be explored globally across its range of variation, as the parameter s is varied over $(-\pi, \pi)$. Equation (5) describes a piecewise linear wave function that systematically explores the unit hypercube

$$\Omega \equiv (X|0 \leq X_i \leq 1; i = 1, \ldots, k) \tag{6}$$

from which standard samples of noncorrelated input factors that are uniformly distributed in the range $[0, 1)$ can be generated. To compute a given S_{Ti}, the factor X_i is assigned a high scanning frequency ω_i, while the factors in the complementary vector X_{-i} are assigned low frequencies in the range $(1, \omega_i/2)$. A Fourier analysis allows the sum of all the $V_{i_1, i_2, \ldots, i_s}$ terms that do not include the factor X_i (let us indicate it as V_{-i}) to be recovered at the low end of the spectrum (see equation (10)). The quantity $V - V_{-i}$ is clearly the sum of all effects involving X_i, that is, those effects needed to compute S_{Ti}. From the spectrum at higher frequencies [see equation (9)] one can recover V_i, and hence the first order term S_i.

In practice, the output $Y = f(X_1(s), X_2(s), \ldots, X_k(s))$ is evaluated along the curve and is considered as a function $f(s)$ of s. The output Y shows different periodicities combined with the different frequencies ω_i, whatever the model $f(\cdot)$ is. If the factor X_i is influential, the oscillations of Y at the frequency ω_i, and its integer multiples $p\omega_i$, will be of high amplitude. Such amplitudes are quantified by the spectrum $\Lambda^2(\omega)$ evaluated at those frequencies, which represents the basis for computing the sensitivity. The spectrum $\Lambda^2(\omega)$ of $f(s)$ at each frequency ω is computed as $\Lambda^2(\omega) = A^2(\omega) + B^2(\omega)$, where

$$A(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\cos\omega s\, ds, \tag{7}$$

$$B(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\sin\omega s\, ds \tag{8}$$

are the Fourier coefficients of $f(s)$. Such integrals are computed using quadrature formulae that operate on a set of N points that are selected along the curve and equally spaced; N is related to the frequency ω_i through the Nyquist theorem, $N = 2M\omega_i + 1$ (see Saltelli, Tarantola and Chan, 1999, Appendix), where M is the number of harmonics considered in the analysis (usually M is set to 4). Hence, upon fixing ω_i, the number of points along the curve is determined. With N points on a curve, it is possible to derive estimates for V_i, V_{-i} and V:

$$\widehat{V}_i = 2\sum_{p=1}^{M}\Lambda^2(p\omega_i), \tag{9}$$

$$\widehat{V}_{-i} = 2\left[\Lambda^2(1) + \Lambda^2(2) + \cdots + \Lambda^2\left(\frac{\omega_i}{2}\right)\right], \tag{10}$$

$$\widehat{V} = 2[\Lambda^2(1) + \Lambda^2(2) + \cdots + \Lambda^2(M\omega_i)]. \tag{11}$$

Finally, the $\widehat{S}_i$ and $\widehat{S}_{Ti}$ for the factor X_i are obtained, respectively, by $\widehat{V}_i/\widehat{V}$ and $1 - \widehat{V}_{-i}/\widehat{V}$. If we are interested in computing $\widehat{S}_j$ and $\widehat{S}_{Tj}$ for a different factor X_j, we need to assign a high scanning frequency to X_j, and low frequencies to the complementary set. This has the effect of creating a new sampling curve, which explores the space Ω with different parametric equations. The model has to be executed again N times over the new sample points and formulae (7)–(11) have to be evaluated for obtaining the new sensitivity estimates.

The extended FAST is a generalization of the FAST method, which was introduced in the 1970s (Cukier et al., 1973; Schaibly and Shuler, 1973; Cukier, Schaibly and Shuler, 1975; Cukier, Levine and Schuler, 1978) and computationally upgraded by Koda, McRae and Seinfeld (1979). The main limitation of FAST, in comparison to its extended version, is in the impossibility of computing the indices S_{Ti}. The main peculiarity of the extended FAST is in the way the frequencies ω_i are associated with the factors X_i, and in the choice of the function $G(\cdot)$. In plain FAST, other functions $G(\cdot)$ are employed, which do not guarantee that the sample is uniformly distributed in Ω. Full technical details on FAST and its extended version are given in Saltelli, Tarantola and Chan (1999).

Using the method of Sobol' (1990), the domain Ω [equation (6)] is explored by multidimensional integrals computed via Monte Carlo. In this case, V_{-i} is given by

$$V_{-i} = E\Big(f(x_1^1, x_2^1, \ldots, x_{i-1}^1, x_i^1, x_{i+1}^1, \ldots, x_k^1) \cdot f(x_1^1, x_2^1, \ldots, x_{i-1}^1, x_i^2, x_{i+1}^1, \ldots, x_k^1)\Big) - f_0^2, \tag{12}$$

where f_0 is the mean, and the superscript index (1 or 2) denotes different samples being used; similarly the first-order term is computed as

$$V_i = E\Big(f(x_1^1, x_2^1, \ldots, x_{i-1}^1, x_i^1, x_{i+1}^1, \ldots, x_k^1) \cdot f(x_1^2, x_2^2, \ldots, x_{i-1}^2, x_i^1, x_{i+1}^2, \ldots, x_k^2)\Big) - f_0^2. \tag{13}$$

In both cases the expectation is computed via Monte Carlo using quasirandom numbers; that is, the multidimensional integration is reduced to a plain sum (Sobol', 1990). A different set of simulations is needed for each factor. The cost of these computations can be expressed in terms of the number of model evaluations required (i.e., number of times the model needs to be executed). The cost depends upon the number of factors k and on the sample size N used to compute S_i or S_{Ti} (the cost is N per factor per sensitivity index using Sobol'). The extended FAST method is numerically more efficient than the method of Sobol'. FAST can compute for the same cost N both S_i and S_{Ti}, that is, the first order plus the total index for a given factor. The computational cost per factor, N, may be of the order of 500 model runs or greater; that is, obtaining a set of S_i, S_{Ti} is more expensive than computing a set of β_i's using regression analysis (see also examples in Section 3).

For both the FAST and Sobol' methods f must be square-integrable. Both methods only apply for uncorrelated factors. The extension of variance-based tests to correlated inputs is not an easy one, because for nonorthogonal samples the variance decomposition loses its uniqueness. Orthogonalization procedures are discussed in Bedford (1998) whereas McKay (1995) suggests a partial sensitivity index that is the model-free generalization of the partial regression coefficient. As in the

stepwise regression analysis case, the order in which the factors are entered into the analysis is material.

A further step can be made to complete the process of making the analysis of model sensitivity more agile and easy to interpret. This is based on the observation that in complex models with several submodels, or modules, or simply with sets of variables pertaining to different logical levels, one might desire to decompose the prediction variance according to subgroups of factors (i.e., the input factors might be multivariate). For example, let us assume that the set $\mathbf{X}$ of the input factors is partitioned in two groups $\mathbf{W}$ and $\mathbf{Z}$ such that $\mathbf{X} = \mathbf{W} \cup \mathbf{Z}$. The variance V of Y is then decomposed as $V = V_{\mathbf{W}} + V_{\mathbf{Z}} + V_{\mathbf{WZ}}$, similarly to (3). For the computation, via the extended FAST, of the first order ($S_{\mathbf{W}}$, $S_{\mathbf{Z}}$) and the total effect indices ($S_{T\mathbf{W}}$, $S_{T\mathbf{Z}}$), the procedure shown to select the frequencies for X_i and $X_{\sim i}$ is here applied to $\mathbf{W}$ and $\mathbf{Z}$. A straightforward extension of (12) and (13) allows the computation of the indices for groups of factors using the method of Sobol'. The extension to more than two subsets is trivial. An application example of an analysis by groups is given in Figure 4e for the level E model.

2.3 Comparison with Other Approaches

The variance-based methods discussed so far are not the most widely used in engineering, physics or chemistry; neither are the regression methods described in Section 2.1. What one meets most often in articles of physics or chemistry is the so-called one factor at a time (OAT) approach, where inference about the behavior of Y is made by changing one factor at a time and investigating the change in Y or by computing, for example, via the so-called direct method, sets of partial derivatives of Y with respect to X_j, such as

$$S_j = \frac{\partial Y}{\partial X_j}. \tag{14}$$

These are computed at a point $\mathbf{x}^0 = (x_1^0, x_2^0, \ldots, x_k^0)$, where all the X_j's are fixed to their nominal values x_j^0. In other words, the quantity S_j is the local sensitivity index measuring the effect on Y of perturbing X_j around the nominal value x_j^0. Popular alternatives to (14) are

$$S_j = \frac{x_j^0}{Y^0}\frac{\partial Y}{\partial X_j}, \tag{15}$$

measuring the effect on Y of perturbing X_j by a fixed fraction of X_j's mean value, and

$$S_j = \frac{\sigma(X_j)}{\sigma(Y)}\frac{\partial Y}{\partial X_j}, \tag{16}$$

measuring the effect on Y of perturbing X_j by a fixed fraction of X_j's standard deviation. Local sensitivity analysis is typically encountered in the solution of inverse problems, that is, a class of estimation problems where the model is generally noninvertible and the parameters to be estimated are not directly observable; an example is the determination of quantum mechanical potentials from yield rates of chemical reactions (Rabitz, 1989). Local approaches have been used in chemical kinetics, to determine kinetic coefficients of complex systems from measured data (Turanyi, 1990), or in hydrogeology (Smidts and Devooght, 1997). For a review of local methods, showing how these can be coupled with multivariate statistical methods, such as principal component analysis, see Turanyi (1990). When contrasted with these other methods, the variance-based measures of Sections 2.1 and 2.2 display a number of useful properties for sensitivity analysis:

1. These measures take into consideration the whole range of input variation and the form of its probability density function (pdf). Local methods only look at a neighborhood of $\mathbf{x}^0$.
2. The effect of X_j is averaged over the entire range of X_j, as well over $\mathbf{X}_{\sim j}$; that is, the space of all factors but X_j. By contrast, in the local approach the effect of the variation of X_j is measured when all other X_r, $r \neq j$, are kept constant at their nominal values.
3. These measures can identify interaction effects (as in standard ANOVA) and take them into account when assessing the overall influence of X_j on Y.
4. These measures are model independent in the sense that, unlike linear or rank regression analysis, their performance is not conditional on the additivity or linearity of the model.
5. They can cope with the curse of dimensionality, thanks to the possibility of estimating the S_{Ti}'s.

We would like to end this section devoted to the methods by pointing the reader to alternative original approaches to sensitivity analysis that are not treated in the present work:

- A very elegant method for sensitivity analysis, used in reliability problems, is the first-order reliability method (FORM; see Cawlfield and Wu, 1993). This method focuses on extreme percentiles of the distribution of Y that are of relevance in risk analysis, and identifies factors that are most responsible for Y assuming values in a specified forbidden region.

• Sacks, Welch, Mitchel and Wynn (1989) suggest a method of sensitivity analysis that is based on decomposing Y itself into terms (functions) of increasing dimensionality, and look at these as measures of sensitivity (see also an application in Welch et al., 1992). This method has much in common with the work of Sobol' (1990), described in Section 2.2. A comparison of analysis of variance in classic factorial design with the method of Sobol' is given in Archer, Saltelli and Sobol' (1997).

• A decomposition of Y into finite differences is proposed in Rabitz et al. (1999). The procedure relies on assumptions on the order of the highest nonzero terms in (3). An approximation to Y is built by knowing its value on lines, planes and hyperplanes that pass through a given selected point in the space of the input factors.

• The projection pursuit regression is a nonparametric regression technique developed by Friedman and Stuetzle (1981). The usual regression model is generalized by replacing the linear manner in which the factors X_i enter the prediction process by arbitrary nonlinear functions of the X_i determined nonparametrically. The technique permits discovery of interaction and highly nonlinear relationships between Y and the X_i. However, the user must choose the number of the nonlinear functions through a compromise between parsimony, interpretability and explanatory power. An application is given in Draper, Saltelli, Tarantola and Prado (2000).

• In generalized sensitivity analysis (see Hornberger and Spear, 1981) the output realizations are classified as "behavior" (A) or "nonbehavior" (B). The "behavior" is often stipulated in terms of how well model predictions fit available data. For each input factor X_i, two subsamples are identified: those that lead to acceptable behavior, X_i^A, and those that lead to unacceptable behavior, X_i^B. For factor X_i the Kolmogorov–Smirnov two-sample hypothesis test is then applied to check if the two subsamples can be considered as generated from the same statistical distribution (null hypothesis). The factor X_i is regarded as important when the null hypothesis is rejected. The statistical test is then repeated for the remaining factors.

• Bayesian sensitivity analysis is concerned with decision-making under uncertainty, where the interest is in selecting the optimal action from a feasible set of alternatives (see Rios Insua, 1990). Sensitivity analysis is used in this context to test the robustness of the dominating alternative with respect to uncertainties in the prior beliefs, which are modeled with prior probability distributions, and with respect to the process of modeling the decision maker's preferences among consequences.

• The term "Bayesian sensitivity analysis" is used by O' Hagan, Kennedy and Oakley (1999) to point to the approach where the model output Y is treated as a random process, and hence estimated by (hierarchical) modeling given a set of points sampled from $\mathbf{X}$. Then the same approach as Sacks et al. (1989) is used for sensitivity analysis.

Most of the material touched upon above is described in a recent handbook for sensitivity analysis (Saltelli, Chan and Scott, 2000).

3. APPLYING THE METHODS TO THE EXAMPLES

This section treats seven worked examples. Of these, the first two (Legendre polynomial and Bateman equations) can be reproduced by the reader based on the information provided here. For the other models, the reader is referred to the original references.

LEGENDRE POLYNOMIAL. This example is introduced to illustrate the importance of accounting for interactions in a model (McKay, 1996). Strongly nonadditive models are not necessarily complex ones. This is a Legendre polynomial of order d,

$$L_d(x) = \frac{1}{2^d} \sum_{m=0}^{d/2} \left[(-1)^m \frac{d!}{m!(d-m)!} \times \frac{(2d-2m)!}{d!(d-2m)!} x^{d-2m} \right] \tag{17}$$

taken as a function of x and of the integer number d, with $x \in [-1, 1]$ and $d = \{1, 2, 3, 4, 5\}$. The input factors x and d are assumed independent and uniformly distributed. This model has only two nonzero indices (analytically computed): $S_x = 0.2$ and the second-order term $S_{xd} = 0.8$. Thus, $S_d = 0$, whereas $S_{Td} = S_d + S_{xd} = 0.8$. This simple test case shows that a factor may have zero first-order term and nonzero interactions.

BATEMAN EQUATIONS. This example introduces time-dependent outputs and shows how the results can be displayed in an intuitive fashion. The Bateman equations describe a simple chemical or radioactive chain where each element's growth rate is proportional to the father concentration, and each element's decay rate λ_i is proportional to the concentration C_i of the element itself. The

set of Bateman equations is given in Figure 5. Its analytical solution is

$$C_i(t) = \sum_{m=1}^{i} \left[C_m^0 \prod_{r=m,\, m\neq i}^{i-1} \lambda_r \times \sum_{n=m}^{i} \frac{e^{-\lambda_n t}}{\prod_{l=m,\, m\neq i,\, l\neq n}^{i} \lambda_l - \lambda_n} \right]. \quad (18)$$

Here λ_i and C_i^0 are, respectively, the decay rate and the initial concentration of element i, and the product $\prod$ equals 1 when one of its indices becomes zero or negative.

Four elements are considered in this example. The model output is taken as the concentration of the fourth element across time ($Y(t) = C_4(t)$) and is driven by eight, mutually independent, uncertain factors: the four initial concentrations C_i^0 and the four decay (or reaction) coefficients λ_i. Their ranges and distributions are given in Table 1.

Figure 5 displays the results as a function of time. The curves displayed in (a) are the results of 8,200 model evaluations and give a global view of the uncertainty affecting the output at different time points. The output dynamics occurs in the time range from 10^4 to 10^8 s where the output shows a peak and then drops down to zero. The values in (a) have been used to estimate the model coefficient of determination R_y^2 [panel (b)], which is evaluated at each time point. The noncomputable region (shaded area) is due to singularities in matrix inversions. It can be noted that the regression explains approximately 80% of the output uncertainty up to 10^6 s. At later times R_y^2 decreases rapidly, indicating the presence of nonlinearities. The analysis at times beyond 10^8 s is of no interest because the output goes to zero, and correspondingly, $R_y^2 = 0$. The

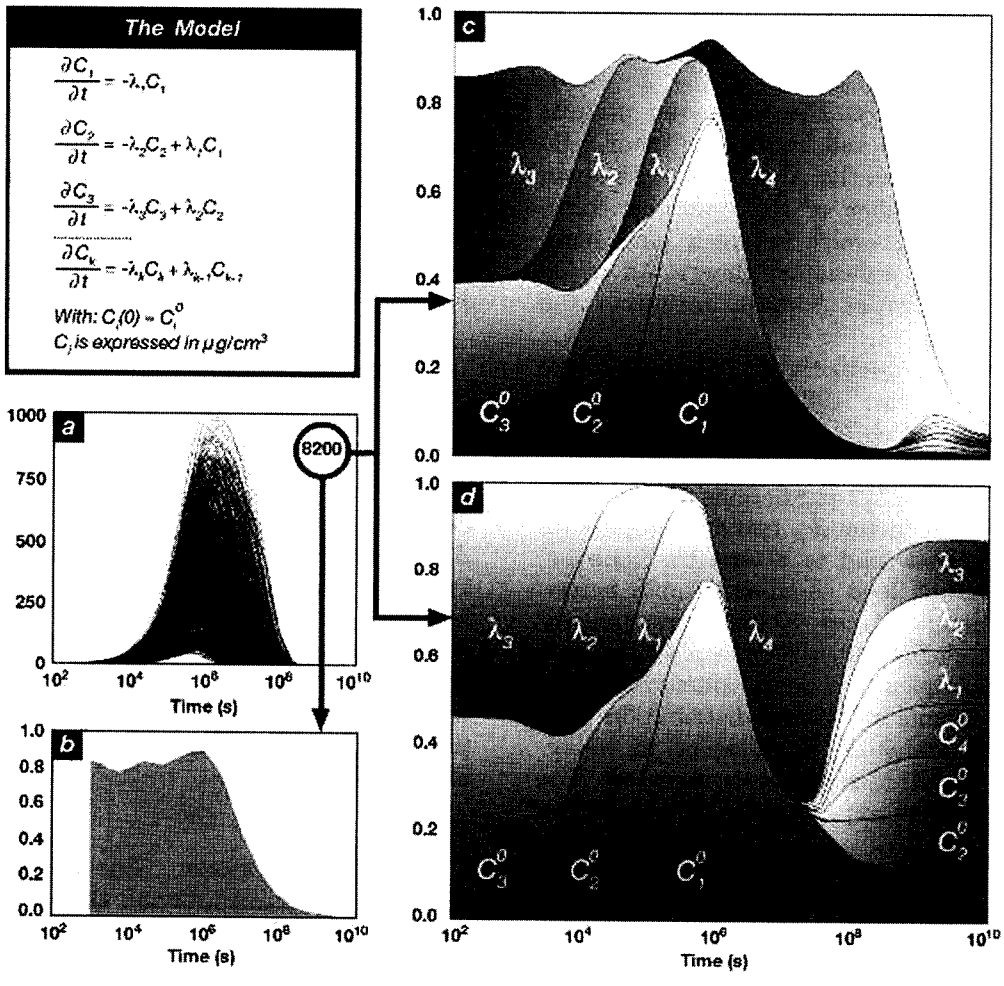

FIG. 5. *Bateman test case: the panels display the results as functions of time (seconds).* (a) *The uncertainty analysis on the concentration of the fourth element of the chain, performed on a set of* 8,200 *model evaluations.* (b) *The* R_y^2 *estimate as obtained from the set of model runs.* (c) *and* (d) *Time-dependent pie charts of the* S_i *and the* S_{Ti}^* *for the* 8 *factors of the model, as obtained from the set of* 8,200 *model runs.*

TABLE 1
Characterization of the input factors for the Bateman equations case

Notation	Definition	Range	Units
C_1^0	Initial concentration 1	0–1,000	moles
C_2^0	Initial concentration 2	0–100	moles
C_3^0	Initial concentration 3	0–10	moles
C_4^0	Initial concentration 4	0–0.01	moles
λ_1	Decay rate 1	1×10^{-6}–5×10^{-6}	s^{-1}
λ_2	Decay rate 2	1×10^{-5}–1×10^{-4}	s^{-1}
λ_3	Decay rate 3	1×10^{-4}–1×10^{-3}	s^{-1}
λ_4	Decay rate 4	2×10^{-8}–2×10^{-6}	s^{-1}

8,200 time-dependent curves displayed in (a) have also been employed to run the extended FAST. Each run of the extended FAST provides first-order (S_i) and total indices (S_{Ti}) at a given time point. For presentation purposes, we have divided each S_{Ti} by the summation $\sum_i S_{Ti}$, indicating the resulting index, scaled between 0 and 1, as S^*_{Ti}. The sensitivity indices are displayed as a time-dependent pie chart. Any given vertical cut made either through panel (c) or (d) of Figure 5 at a given time point corresponds to a pie chart for that time point. The length of the segments identified by the cut are proportional to the areas of the sectors of the corresponding pie chart. The colored region in (c), that is, the sum of the S_i, is an index of the model additivity. The model can be considered nearly additive (up to 10^8 s), because $\sum S_i > 0.8$. Panel (c) shows that, in correspondence to the maximum of Y (at time = 10^6 s), the driving factor is C_1^0. At later times (10^7 < time < 10^8 s), the most important factor becomes λ_4. Comparing inserts (c) and (d) helps to identify the factors involved in the interactions. For example, the whole output variance at 10^7 s is explained exclusively by λ_4 and C_1^0 [Figure 5(c)], apart from a missing bit lower than 20%. Looking at Figure 5(d) we see that for the same time point all the variation (inclusive of interactions) is due to exactly the same two factors. This tells us that there must be a single interaction effect involving λ_4 and C_1^0 at this time point.

This analysis helps us understanding how sensitive the model prediction is to its input factors: output uncertainty could efficiently be reduced by reducing uncertainty on the most important factors.

LEVEL E MODEL. The level E model involves a system of PDE's describing mass transfer with decay and chemical reaction in a one-dimensional multilayered geosphere. The output variable considered in this study is the total annual dose to man due to all the migrating radionuclides. The reader can replicate this test case by looking at the OECD reference (OECD/NEA PSAC User Group, 1989).

The factors for the level E exercise and their distributions (Table 2) were decided by a panel of experts, who designed this exercise for a benchmark of Monte Carlo methods (OECD, 1989). Also for this test case the factors are assumed independent.

The rationale for examining this model is that it has been also used for a benchmark of SA methods (OECD, 1993). Further, it is a rather difficult example for SA due to its nonlinearity and nonadditivity.

A standard numerical technique (Crank–Nicholson) is used to integrate the system of PDE's. In Figure 4, the panels display the results as functions of time (years). The two curves in Figure 4a are the result of 514 and 3,084 model evaluations, respectively. We discuss first the set of size 3,084. This has been used for estimating R_y^2 [panel (b)], the first-order S_i and total indices S_{Ti} [panels (c) and (d), respectively] for all the 12 factors of the model. Figure 4(b) shows that the underlying model is strongly nonlinear, as the R_y^2 value is always below 0.2. The additive component of the model, measured by the sum of the S_i's, is shown by the colored region in (c), which is below 0.6 everywhere. This means that more than 40% of the output uncertainty is due to interactions among factors. A cumulative plot of the total indices for the 12 factors is given in panel (d). The sum of the total indices oscillates between 4 and 7 along the time interval considered in the analysis, indicating a large contribution of interactions to the output variance (this sum should equal 1 for a perfectly additive model). The most important factors can be identified:

- $V(1)$ = water velocity in the geosphere's first layer;
- $L(1)$ = length of geosphere's first layer;
- $R(1)$ = factor to compute the retention coefficient for Np (first layer);
- W = stream flow rate.

Even for this test model, the analysis reinforces our confidence in the understanding of the behavior of the model. Note how, for this nonadditive test

TABLE 2
Characterization of the input factors for the level E test case

Notation	Definition	Distribution	Range	Units
T	Containment time	Uniform	100–1,000	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 velocity in geosphere's 1st layer	Log-uniform	10^{-3}–10^{-1}	m yr^{-1}
$l^{(1)}$	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 velocity in geosphere's 2nd layer	Log-uniform	10^{-2}–10^{-1}	m yr^{-1}
$l^{(2)}$	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^{-1}

case, several factors are simultaneously influential. This derives from how the test case was designed, as the experts in OECD (1989) were aiming for a model "balanced" in all parameters. A practical consequence is that, unlike the Bateman case, it is not so easy to reduce the variance in the prediction by just reducing uncertainty in one or two factors.

An example of the analysis by groups of factors is given in Figure 4(e), where the 12 factors of the model have been partitioned into just two groups: the indices S^*_{Ti} are estimated for natural and engineered barriers. The analysis by groups of factors is cheaper than the analysis for single factors: in this case, for instance, two pairs of indices (S_i and S_{Ti} for the two groups) are computed instead of 12 pairs. Using a base sample size of 257 the analysis by groups requires 514 model evaluations, whereas the analysis for each single factor involves 3,084 model runs. In Figure 4(e), the result obtained for the analysis by groups highlights the modest role played by engineered barriers in comparison to natural barriers, confirming the beliefs of risk assessment practitioners involved in nuclear waste disposal studies. In Figure 4(f), two pie diagrams illustrate the fractional contribution of each group to the output variance at a given time point.

LEVEL E/G MODEL. This example is introduced for its approach to "scenario uncertainty." Level E/G is an upgraded version of the level E model described above (Draper et al., 1999; Draper et al., 2000). This more general model allows the exploration of alternative scenarios that are selected at runtime; that is, in each Monte Carlo run a different scenario may be selected based on a Russian roulette trigger. In this way, it was possible to perform a "model uncertainty" audit on the basis of the variance decomposition

TABLE 3
Calibration analysis: average values $\overline{S}(i)$ and $\overline{S}_T(i)$ as computed over the 128 estimates of $S(i)$ and $S_T(i)$ plus or minus their standard deviations σ_S and σ_{T_S}

	TiN nanocrystalline film	
	$\overline{S}(i) \pm \sigma_S$	$\overline{S}_T(i) \pm \sigma_{S_T}$
E(Gpa)	0.70 ± 0.10	0.88 ± 0.03
ν	0.01 ± 0.01	0.01 ± 0.01
ρ(g/cm^3)	0.10 ± 0.04	0.28 ± 0.09
t(nm)	0.00 ± 0.00	0.01 ± 0.01

$$\begin{aligned}\hat{V}(y) &= V_S\left[\hat{E}(Y|S)\right] + E_S\left[\hat{V}(Y|S)\right] \\ &= \sum_{i=1}^{l} p_i(\hat{\mu}_i - \hat{\mu})^2 + \sum_{i=1}^{l} p_i\hat{\sigma}_i^2\end{aligned} \tag{19}$$

that permits us to explore what fraction of the variance $V(Y)$ is due to scenarios (the term $V_S[\hat{E}(Y|S)]$, also called between-scenarios variance) and how much to the uncertain model parameters (the term $E_S[\hat{V}(Y|S)]$, also called within-scenario variance).

In (19), p_i is the probability of the ith scenario, l is the number of scenarios considered and $\hat{\mu}_i$ and $\hat{\sigma}_i$ are the means and standard deviations of Y within the ith scenario. The model involves 54 uncertain factors (poorly known physical quantities and intrinsic parametric uncertainties, such as the time of occurrence of a geological event) as well as six different scenarios. The probability distributions for the factors and the scenarios are selected by expert judgement. An interesting result

of this exercise, run taking as output of interest $Z(t) = \max Y(s)$, $s \in [0;t]$, where $Y(s)$ is the dose absorbed at time s, is that the percentage of variance due to scenarios remained stable at about 30% of the total variance even though the probabilities p_i were allowed to vary within a range assigned by the experts (Draper et al., 1999).

Another interesting result of the same work, obtained taking $Y(s)$ directly as the output of interest, is the following: if the first-order sensitivity index is computed for the factor "trigger" that drives the selection of the scenario, one obtains a quite low sensitivity (Figure 6), whereas the total sensitivity index for the same factor is very large. This tells us that, unless at some stage we are able to drop this trigger by fixing the scenario (e.g., by better field data), this trigger factor will tend to increase the overall variance because of its cooperative effect with other factors.

ENVIRONMENTAL INDICATORS CASE. In this example the model is an indicator, that is, an aggregation of data and expert opinion that is used in decision making. Tarantola, Jesinghaus and Puolamaa (2000), carried out this test case based on data for Austria in the year 1994. Atmospheric emissions (e.g., Kilograms of CO_2 emitted per year), emission factors (e.g., Kilograms of CO_2 emitted per ton of waste incinerated) and production rates (e.g., Kilograms of municipal waste incinerated) are available in the CORINAIR database, maintained by the European Environment Agency (EEA, 1996), for several pollutants at different spatial resolutions.

Sources of uncertainty in input data include roughly estimated emission factors and production rates, together with the choice of aggregation levels for data (see caption in Figure 2). Data in CORINAIR are often provided with quality labels representing the confidence levels at which data are known (i.e., label A is attached to "best quality" data, while E is given to "worst quality" data).

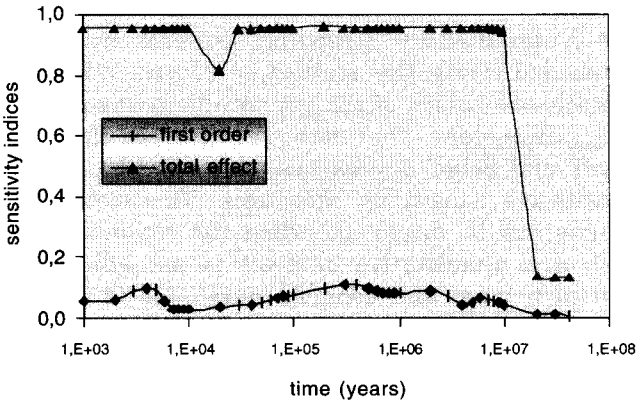

FIG. 6. *Level E/G test case: sensitivity analysis for the trigger factor. The S_i and the S_{Ti} for the factor responsible for "scenario uncertainty," estimated using the extended FAST, are shown across time (years).*

For the two policy options considered for the disposal of the Austrian solid waste (incineration and landfill), production rates and emission factors are combined to evaluate atmospheric pollutant emissions.

Such emissions are then aggregated to a set of pressure indicators using two different approaches, one suggested by the Finnish statistical office and the second by EUROSTAT. The set of Finnish indicators (Puolamaa, Kaplas and Reinikainen, 1996) includes those for greenhouse and acidification. The greenhouse indicator, for instance, is obtained through a weighted sum of the effects of carbon dioxide, methane and nitrous oxide emissions (i.e., N_2O and NO_X). Global warming potentials are used as weights (Houghton et al., 1996). Alternatively, the set of EUROSTAT indicators (EUROSTAT, 1999) is used. This includes air pollution, climate change and ozone layer depletion. Climate change, for instance, is obtained by adding up emissions of CO_2, CH_4, N_2O and NO_X. Each term in the sum is weighted by weights obtained from surveys among natural scientists. The two indicator systems can be selected at runtime according to the value sampled for the trigger factor [E/F]. This factor is responsible for the uncertainty in the selection of the indicator set and is related to the variability due to the preference of the expert (in reality a given expert would either select the Finnish or the EUROSTAT system).

The pressure indicators are finally aggregated (as weighted sums) to form environmental pressure indices (PI's), one for incineration and one for landfill. The PI for a course of action is intended to quantify the total hazard to the environment of that action. The model output Y is defined as the logarithm of the ratio between the PI for incineration and the PI for landfill.

The weights used in the aggregation of the indicators (e.g., the weights needed to add up greenhouse with acidification, air pollution with climate change etc.) are called inter weights. They quantify the perceived relative importance of the problem covered by the indicators (e.g., whether acidification is more or less important than greenhouse effect and so on). The interweights can be formulated in two ways. In the first, the interweight for a given indicator is defined as the target reduction that we want to achieve for that indicator within a

defined time frame (Adriaanse, 1993). In the second, the interweights are obtained as the result of a procedure, called analytic hierarchy process (AHP), which relies on the rankings provided by stakeholders from various domains (environmentalists, politicians, public etc.) upon the problems covered by the indicators (Puolamaa, Kaplas and Reinikainen, 1996).

The model contains 196 uncertain factors, which have been partitioned into 7 groups (see the caption for Figure 2). The total sensitivity indices have been estimated for each group using the extended FAST technique. In all, 7,182 model evaluations are performed. The bimodal histogram displayed in Figure 1 represents the outcome of the uncertainty analysis: the left-hand region, where incineration is preferred, encompasses approximately 40% of the total area. The pie chart in Figure 2 shows the total sensitivity indices as obtained by the extended FAST (even here we use the normalized $S^*_{Ti} = S_{Ti}/\sum_i S_{Ti}$ for ease of presentation): these results confirm that the choice of the set of indicators has an overwhelming influence and reveal that the scientific community must work to find a consensus on the proper system of indicators. Figures 1 and 2 together tell us that at present there is no scientific basis to prefer incineration versus landfill. Here sensitivity analysis does not help to decide whether one should use the Finnish or the EUROSTAT indicators, but rather tells us about the relative importance of the various types of uncertainty. Other case studies of the same model are in Tarantola, Jesinghaus and Puolamma (2000).

THE KIM MODEL. The aim of this example is to show how sensitivity analysis can be used for model-building purposes. Also illustrated in this example is a two-stage approach that can be used for many-factor models. Uncertainty and sensitivity analyses played a central role in the construction of the KIM model, a kinetic model for OH-initiated oxidation of dimethyl sulphide (DMS; CH_3SCH_3) that is relevant to climate change studies (Charlson et al., 1992). KIM is a zero-dimensional model that incorporates a description of the reaction pathways for the formation of sulphur-containing molecules, such as sulphur dioxide (SO_2) and methane sulphonic acid (MSA, CH_3SO_3H) from DMS (Figure 3). The KIM model is relevant to climate change studies, because of the important contribution of DMS emissions to the formation of climatically active atmospheric aerosols and, in particular, the hypothesised feedback mechanism linking the biogenic sulphur cycle to the greenhouse effect (Charlson et al., 1992). In brief, the accumulation of greenhouse gases such as CO_2 would increase sea temperature, thereby increasing the fluxes of DMS to the atmosphere; DMS, being a precursor of aerosols, would lead to increased cloudiness (and cloud brightness) that would shield solar radiation. This would ultimately lower earth's temperature, counteracting the greenhouse effect.

As discussed in Section 1, the building of the KIM model proceeded by running the model in a Monte Carlo framework. Global sensitivity analysis was used systematically by the chemists to verify their understanding of the model.

Figure 7 relates to a version of KIM that includes multiphase (droplets–air) transport and chemistry and dry deposition for SO_2 (Campolongo et al., 1999). Of interest in the study are α, the concentration ratio in marine aerosol between MSA and non-sea-salt sulphates ($\alpha = \mathrm{MSA}/(SO_2 + H_2SO_4)$), and its temperature dependency. The predicted values of α are compared with field observations of MSA–to–non-sea-salt-sulphate ratios (Bates, Calhoun and Quinn, 1992). The sensitivity analysis procedure, tested on a recent version of KIM that includes 68 uncertain factors, underlines the importance of the kinetic coefficients k_1, involved in the reaction between OH and DMS, and k_{21}, which controls the gas phase yield of $DMSO_2$ from the oxidation of DMSO. The strong dependence of α upon k_{21}, as highlighted by the analysis, has forced the chemists to consider the need for further experimental determination of k_{21}. It is unlikely that a similar conclusion could have been reached safely using a one-factor-at-a-time approach.

Because of the large number of factors to be sampled, in a subsequent work (Campolongo, Tarantola and Saltelli, 1999) a two-step procedure was implemented:

- A preliminary screening exercise was first conducted using the method of Morris (1991) to identify the subset of the potentially most explanatory parameters. The method is cheaper than the FAST and Sobol' approaches. However, it provides qualitative sensitivity measures. The method of Morris can be suggested when the computational cost of a quantitative analysis is not affordable, due to a large and complex model (as is true in this case).
- A quantitative method (the extended FAST) was consequently applied to the subset of preselected inputs. Results of this analysis confirmed the large importance of the kinetic coefficients k_1 and k_{21}, the former of which was found responsible for about 90% of the total output variance.

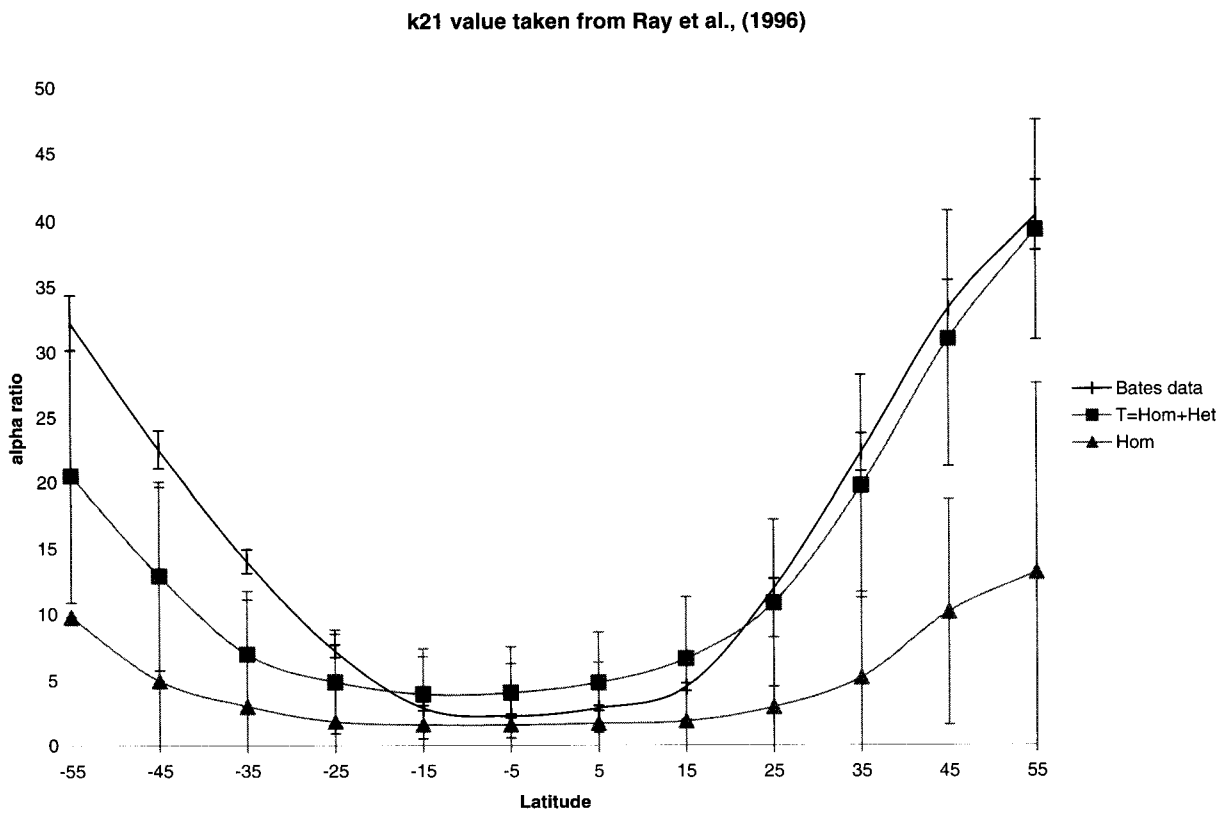

FIG. 7. *KIM, latitude dependence of the α ratio: experimental data and two different model predictions are compared across latitude.*

AN EXAMPLE OF CALIBRATION. This example illustrates the use of SA for the calibration of an experimental setup. The term calibration is used here by analogy with the physical experiment world, where an instrument is calibrated against a known standard. In Tarantola, Pastorelli, Beghi and Bottani (2000), SA is employed with the objective of investigating the performance of an estimation procedure for the determination of a set of physical parameters from a set of experimental measurements and computer simulations before the measurements themselves are made. Specifically, the objective here is to indicate if the parameters of interest have a chance of being properly estimated, thus saving extensive laboratory time and costs.

This example is of interest in the semiconductors industry: a computational model (Hardouin Duparc, Sanz Velasco and Velasco, 1984) simulates the underlying physical process. The physical system consists of a thin homogeneous film of uniform thickness (e.g., amorphous carbon), which is deposited on a substrate (typically silicon, Si) whose properties are fully known. The answer being sought is how accurately the film's elastic properties (i.e., Young's modulus E and the Poisson ratio ν) will be estimated in the presence of simultaneous uncertainty in the design parameters of the film (i.e., its mass density ρ and its thickness t).

In the usual estimation procedure, the elastic properties E and ν are determined via a standard least-squares procedure

$$\text{LS}(E,\nu) = \sum_i \left[\frac{v_c^i(E,\nu) - v_{\text{exp}}^i}{\sigma_{\text{exp}}^i}\right]^2, \tag{20}$$

where the v_{exp}^i are the experimental acoustic wave velocities observed at a set of incidence angles i onto the film, $v_c^i(E,\nu)$ are the model predictions calculated for a mesh over the plane of points (E,ν) and the σ_{exp}^i are related to the measurement errors at a given angle. The solution is identified as the point $(\hat{E},\hat{\nu})$ that corresponds to the minimum of the function LS $(\cdot,\cdot)$. The calibration does not usually account for uncertainties in ρ and t; that is, the minimum of LS $(\cdot,\cdot)$ is sought over all values (E,ν) in the plane. In the present example we try to overcome this limitation.

In our calibration study, the four factors (E,ν,ρ,t) are all allowed to vary. The target function of our analysis is

$$\text{LS}_j(E,\nu,\rho,t) = \sum_i \left[v_c^i(E,\nu,\rho,t) - v_{cj}^i(E_j,\nu_j,\rho_j,t_j)\right]^2. \tag{21}$$

In (21), we have dropped the σ^i_{exp}, assuming that the measurement error is independent of the angle, that is, that σ^i_{exp} does not depend on i. Further, assuming that no measurements have been taken yet, we have replaced the experimental velocity v^i_{exp} with one of the model output value v^i_{cj}, calculated for some given point $(E_j, \nu_j, \rho_j, t_j)$. Clearly the target functions LS_j now depend on the point j selected in the space of the four factors (E, ν, ρ, t). To compute sensitivity indices for LS_j we now need to explore systematically the space of the four factors (E, ν, ρ, t). We do this using the method of Sobol'. A base sample, of size 512, is used for estimating the indices for each of the four factors. Two indices are computed for each factor, the first-order index and the total index. Hence the total cost of the analysis is $(2 \times 4 + 2) \times 512 = 5{,}120$ model executions. At this point, using the same sample of 5,120 model executions, we can compute as many sensitivity indices as we like for LS_j, simply by selecting different points $v^i_{cj}(E_j, \nu_j, \rho_j, t_j)$ in (21) and using the sample to estimate the sensitivity coefficients. The computational cost of the sensitivity analysis depends essentially on the computer time needed to perform the 5,120 runs of the model that provide the values (at different angles i) of $v^i_c(E, \nu, \rho, t)$ for different values of (E, ν, ρ, t), while the effort needed to estimate the sensitivity indices given the set of model outputs is negligible.

We have actually performed 128 sensitivity analyses by selecting the sample points $(E_j, \nu_j, \rho_j, t_j)$, $j = 1, \ldots, 128$, to uniformly cover the domain of the input factors. The 128 points in the space of (E, ν, ρ, t) are simply a subset of the 5,120 already explored. The average values of the first order and the total effective indices, respectively $\overline{S}(i)$ and $\overline{S}_T(i)$, over the 128 estimates, are given in Table 3 for one of the films analysed.

The $\overline{S}_T(i)$ for E is always very high, indicating that in all three cases estimation will only be guaranteed for E, whereas ν will remain completely undetermined. Thus, we know in advance of the calibration that subsequent experimental measurements will be capable of estimating E but not ν.

The small values of σ_{S_T} [i.e., the standard deviation of the 128 estimates of $S_T(i)$] for parameter E indicate that the sensitivity estimates are robust to (poorly influenced by) the choice of the points $(E_j, \nu_j, \rho_j, t_j)$.

3.1 On the Generation of the Input

We end this section on the worked examples by discussing the generation of the input probability density function for the input factors. Usually sensitivity starts from pdf's given by the experts (see the level E example). The analyst moves on from the assigned pdf's and attempts to characterize how Y depends upon $\mathbf{X}$. In the case of the KIM model, chemists postulate values of the kinetic coefficients that are physically reasonable, often by analogy with similar compounds, and use uncertainty and sensitivity analysis to study their plausibility (Saltelli and Hjorth, 1995). In the calibration example, the film producer provides ρ and t, along with their related uncertainties. Distributions for the input factors could also derive from available data, physical bounding considerations or, finally, results (in the form of posterior pdf's) from previous procedures. An application to time series analysis is given in Planas and Depoutot (2000). See also an application to graphical methods (scatterplots of residuals) to the same problem setting in Young (1999).

4. DISCUSSION

The purpose of this section is to see if some general conclusions on the use of global quantitative sensitivity analysis can be inferred from the examples presented so far.

Sensitivity analysis can be an important element of the model building process. It allows the impact of different factors on Y to be analyzed. It helps to elucidate the impact of different model structures (Environmental indicators test case), mechanisms (KIM) or scenarios (Level E/G) on the models' responses.

The example of the environmental indicators shows that SA is particularly useful in treating structural uncertainty and providing guidance for the identification of the weak links of a scientific assessment chain.

The KIM model shows that in the presence of uncertain evidence and poorly understood mechanisms, global SA can provide assistance in the model-building process. One does not explore a multidimensional space moving a step at a time; global SA provides effective ways for such an exploration.

Sensitivity analysis can also be used to ascertain which subset of input factors (if any) accounts for most of the output variance (and in what percentage). Those factors with a small percentage can be frozen to any value within their range. This contributes to model simplification and was indeed the original motivation of the work of Sobol' (1990). For instance, in the Bateman model, all the factors but C_1^0 and λ_4 could be fixed at $t = 10^7$ s, as shown by the sensitivity pattern of Figure 5. What we mean here is that assigning particular values to factors

other than C_1^0 and λ_4 does not add useful information to the effect of determining the output (see below).

Global SA can be used for mechanism reduction (dropping or fixing nonrelevant parts of the model, as was done routinely in the process of building the KIM model) and for model lumping (building or extracting a model from a more complex one). This has some epistemological implications, as touches upon the "relevance" of a model. It has been argued that often the complexity of models largely exceeds the requirements for which they are used. Especially if one adopts the viewpoint that models are heuristic constructs, built for a task (see, e.g., Oreskes, Shrader-Frechette and Belitz, 1994), then they should not be more complex than they need to be. For example, a complex submodel for the engineered barriers in the level E test case would not be justified. A model is "relevant" when its input factors actually cause variation in the model response. A discussion of the concept of relevance can be found in Beck, Ravetz, Mulkey and Barnwel (1997).

Global SA can be of use as a quality assurance tool, to make sure that the assumed dependence of the output on the input factors in the model makes physical sense and can be reconciled with the analyst's understanding of the system. We have not shown any example of this type of application, because modeling and coding errors that are identified via SA are usually corrected at once. Yet, in our experience, it never happens that a sensitivity analysis is run on a virgin model without some inadequacy of the model being identified.

Another application that was not illustrated here is to aggregation analysis, whereby one treats as an uncertain factor some internal or computational characteristic of a computer program (e.g., number of nodes, of elements, of layers etc.; an example is in Helton, Iman, Johnson and Leigh, 1989). In this case SA helps to tune the degree of detail of the model to the task at hand.

Global quantitative SA could be used prior to and within model identification and parameter estimation (see the example on calibration). In particular, before a given data acquisition activity that entails laboratory experiments or field data is undertaken, the candidate models put forward to describe the system should undergo quantitative SA. The convenience of this approach is in the low cost of the computational experiments compared to the physical ones. The same kind of analysis can serve to ascertain if it is possible to discriminate among competing models.

We have mentioned in Section 2 that local, one at-a-time sensitivity analysis (in forms such as equations (14)–(16)) is today the most commonly used brand of sensitivity analysis. Yet this approach is not appropriate when the problem is to identify the relative influence of different factors on the output in the presence of finite variation in the factors.

Modelers are usually faced with inputs at different levels of uncertainty, often so severe as to cover orders of magnitude. In these instances global tests are needed that cover the entire space of existence of the input factors and all the factors must be varied simultaneously.

In the present article we have never explicitly computed derivative-based sensitivity measures; it would have been foolish to do so for the Legendre polynomial (strong interaction present), and noninformative for the level E case (nonlinear, as shown by the model coefficient of determination R_y^2). We could in theory have used derivative-based sensitivity measures for the Bateman test cases within those time regions where $R_y^2 \geq 0.8$. However, we would trust more a plain standard regression coefficient, which intrinsically includes the normalization of the factors's uncertainty ranges. The technique to choose in any given study should work regardless of the degree of linearity and monotonicity of the model, unless the analyst chooses to proceed by increasing the complexity of the analysis one step at a time. For instance, one could start using the simplest test $\partial Y/\partial X_j$, but this would perform well only if the model is linear. A drawback of this approach is that, to find out if this is the case, the analyst needs to use the more expensive linear regression analysis. If the linear regression analysis were to perform poorly (low R_y^2), then the analyst would have to decide on whether to attempt a nonlinear regression or use the variance-based measures described in this paper. If the model turns out to be additive, the first-order sensitivity coefficients can be used. Even in this case, however, the additivity or otherwise of the model becomes known a posteriori, that is, after the first-order indices have been computed. It is clear that computing from the start the full set of first-order plus total sensitivity indices would provide complete information while saving an analyst's time.

The full potential of sensitivity analysis, especially of the new quantitative measures discussed in this work, is still to be discovered by most modelers. Worse, the same can be said for regression-based methods that are by no mean new, or for the classic "design of experiments" when it is applied to numerical experiments.

In other words, modelers tend to lag behind experimentalists when it comes to planning an efficient and informative design. This might be due in part to an overconfident attitude on the part of the analysts. In experimental physics, overconfidence in the treatment of measurement error, and underestimation of uncertainty, has been discussed among others by Henrion and Fischhoff (1986). This overconfidence might be at play also in computational experiments, with the aggravating circumstance that in these latter experiments more factors are at play and their variation is larger.

A possible cause of difficulty for SA is when the model consists of a large set (k) of input factors and, simultaneously, has many output variables (e.g., m). In such a case, a complete analysis would require the estimation of the sensitivity of each output to every input, thus returning $m \times k$ indices. We believe that the analysis can be made more effective by focussing not on the model output *per se* but on the problem that such output is supposed to solve. To this end, "model use" should be declared before uncertainty and sensitivity analyses are performed. For example, we have not analyzed the full set of outputs of the environmental indicators test case, but only the "top statement" (landfill or incineration) that was the object of the analysis. This approach has the merit of increasing the transparency of the analysis.

The authors reviewed in Oreskes, Shrader-Frechette and Belitz (1994) argue that models are heuristic constructs, built for a task, rather than either true or false representations of the world, and that model evidence does not have the same role as mathematical proof. Instead, a model can provide generic evidence to defend or disprove a thesis. In this context, sensitivity analysis can be seen to be an important element of evidence building.

ACKNOWLEDGMENT

This work has been funded in part by Statistical Office of the European Union (EUROSTAT), Contract 13592-1998-02 A1CA ISP LU, Lot 14, SC97.

REFERENCES

ADRIAANSE, A. (1993). *Environmental Policy Performance Indicators.* Sdu Uitgeverij Koninginnegracht, The Hague.

ARCHER, G., SALTELLI, A. and SOBOL', I. M. (1997). Sensitivity measures, ANOVA like techniques and the use of bootstrap. *J. Statist. Comput. Simul.* **58** 99–120.

BATES, T. S., CALHOUN, J. A. and QUINN, P. K. (1992). Variations in the methanesulfonate to sulfate molar ratio in submicrometer marine areosol particles over the South Pacific Ocean. *J. Geophys. Res.* **97** 9859–9865.

BECK, M. B., RAVETZ, J. R., MULKEY, L. A. and BARNWELL, T. O. (1997). On the problem of model validation for predictive exposure assessments. *Stochastic Hydrology and Hydraulics* **11** 229–254.

BEDFORD, T. (1998). Sensitivity indices for (tree-) dependent variables. In *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output (SAMO98)* (K. Chan, S. Tarantola and F. Campolongo, eds.) 17–20. EUR Report 17758 EN, Luxembourg.

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978). *Statistics for Experimenters.* Wiley, New York.

CAMPOLONGO, F., SALTELLI, A., JENSEN, N. R., WILSON, J. and HJORTH, J. (1999). The role of multiphase chemistry in the oxidation of dimethylsulphide (DMS). A latitude dependent analysis. *J. Atmospheric Chem.* **32** 327–356.

CAMPOLONGO, F., TARANTOLA, S. and SALTELLI, A. (1999). Tackling quantitatively large dimensionality problems. *Comput. Phys. Comm.* **117** 75–85.

CAWLFIELD, J. D. and WU, M.-C. (1993). Probabilistic sensitivity analysis for one-dimensional reactive transport in porous media. *Water Resources Research* **29**, 661–672.

CHARLSON, R. J., SCHWARTZ, S. E., HALES, J. M., CESS, R. D., COAKLEY, J. A. JR., HANSEN, J. E. and HOFMANN, D. J. (1992). Climate forcing by anthropogenic aerosols. *Science* **255** 423–430.

CUKIER, R. I., FORTUIN, C. M., SHULER, K. E., PETSCHEK A. G. and SCHAIBLY, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. *J. Chem. Phys.* **59** 3873–3878.

CUKIER, R. I., LEVINE, H. B. and SHULER, K. E. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *J. Comput. Phys.* **26** 1–42.

CUKIER, R.I., SCHAIBLY, J. H. and SHULER, K. E. (1975). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations. *J. Chem. Phys.* **63** 1140–1149.

DRAPER, D., PEREIRA, A., PRADO, P., SALTELLI, A., CHEAL, R., EGUILIOR, S., MENDES, B. and TARANTOLA, S. (1999). Scenario and parametric uncertainty in GESAMAC: a methodological study in nuclear waste disposal risk assessment. *Comput. Phys. Comm.* **117** 142–155.

DRAPER, D., SALTELLI, A., TARANTOLA, S. and PRADO, P. (2000). Scenario and parametric sensitivity and uncertainty analyses in nuclear waste disposal risk assessment: the case of GESAMAC. In *Sensitivity Analysis* (A. Saltelli, K. Chan and M. Scott, eds.) 275–292. Wiley, New York.

DRAPER, N. R., and SMITH, H. (1981). *Applied Regression Analysis.* Wiley, New York.

European Environment Agency (EEA) (1996). *Atmospheric Emission Inventory Guidebook by EMEP/CORINAIR.* European Environment Agency, Copenhagen.

EUROSTAT (1999). *Towards Environmental Pressure Indicators for the EU.* EUROSTAT, Luxembourg.

FRIEDMAN, J. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 580–619.

HARDOUIN DUPARC, O., SANZ VELASCO, E. and VELASCO, V. R. (1984). Elastic surface waves in crystals with overlayers: cubic symmetry. *Phys. Rev. B* **30** 2042–2048.

HELTON, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engrg. System Safety* **42** 327–367.

HELTON, J. C., IMAN, R. L., JOHNSON, J. D. and LEIGH, C. D. (1989). Uncertainty and sensitivity analysis of a dry containment test problem for the MAEROS aerosol model. *Nuclear Sci. Engrg.* **102** 22–42.

HENRION, M. and FISCHHOFF, B. (1986). Assessing uncertainty in physical constants. *Amer. J. Phys.* **54** 791–799.

HOMMA, T. and SALTELLI, A. (1996). Importance measures in global sensitivity analysis of model output. *Reliability Engrg. System Safety* **52** 1–17.

HORNBERGER, G. M. and SPEAR, R. C. (1981). An approach to the preliminary analysis of environmental systems. *J. Environ. Management* **12** 7–18.

HOUGHTON, J. T., MEIRA FILHO, L. G., CALLANDER, B. A., HARRIS, N., KATTENBERG, A. and MASKELL, K., eds. (1996). *Climate Change 1995. The Science of Climate Change. Contribution of WGI to the Second Assessment Report of the Intergovernmental Panel on Climate Change.* IPCC, Cambridge.

IMAN, R. L. and HELTON, J. C. (1988). An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal.* **8** 71–90.

IMAN, R. L. and HORA, S. C. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Anal.* **10** 401–406.

JANSEN, M. J. W., ROSSING, W. A. H. and DAAMEN, R. A. (1994). Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In *Predictability and Nonlinear Modeling in Natural Sciences and Economics* (Gasman and van Straten eds.) 334–343.

KLEIJNEN, J. P. C. (1998). Experimental design for sensitivity analysis, optimization, and validation of simulation models. In *Handbook of Simulation* (J. Banks, ed.) 173–224. Wiley, New York.

KODA, M., MCRAE, G. J. and SEINFELD, J. H. M. (1979). Automatic sensitivity analysis of kinetic mechanisms. *Internat. J. Chem. Kinetics* **11** 427–444.

LIEPMANN, D. and STEPHANOPOULOS, G. (1985). Development and global sensitivity analysis of a closed ecosystem model. *Ecological Modeling* **30** 13–47.

MCKAY, M. D. (1995). Evaluating prediction uncertainty. Los Alamos National Laboratories Report NUREG/CR-6311, LA-12915-MS.

MCKAY, M. D. (1996). Variance-based methods for assessing uncertainty importance in NUREG-1150 analyses. LA-UR-96-2695.

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

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). OECD–NEA, Paris.

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. OECD–NEA, Paris.

O'HAGAN, A., KENNEDY, M. C. and OAKLEY, J. E. (1999). Uncertainty analysis and other inference tools for complex computer codes (with discussion). *Bayesian Statist.* **6** 503–524.

ORESKES, N., SHRADER-FRECHETTE, K. and BELITZ, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **263** 641–646.

PLANAS, C. and DEPOUTOT, R. (2000). Sensitivity analysis as a tool for time series modeling. In *Sensitivity Analysis* (A. Saltelli, K. Chan and M. Scott, eds.) 293–310. Wiley, New York.

PUOLAMAA, M., KAPLAS, M. and REINIKAINEN, T. (1996). *Index of Environmental Friendliness—A Methodological Study.* Statistics Finland, Helsinki.

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

RABITZ, H., ALI, Ö. F., SHORTER, J. and SHIM, K. (1999). Efficient input–output model representations. *Comput. Phys. Comm.* **117** 11–20.

RIOS INSUA, D. (1990). *Sensitivity Analysis in Multi-Objective Decision Making.* Springer, Berlin.

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435.

SALTELLI, A. and BOLADO, R. (1998). An alternative way to compute Fourier amplitude sensitivity test (FAST). *Comput. Statist. Data Anal.* **26** 445–460.

SALTELLI, A., CHAN, K. and SCOTT, M., eds. (2000). *Sensitivity Analysis.* Wiley, New York.

SALTELLI, A. and HJORTH, J. (1995). Uncertainty and sensitivity analyses of OH-initiated dimethylsulphide (DMS) oxidation kinetics. *J. Atmospheric Chem.* **21** 187–221.

SALTELLI, A. and SOBOL', I. M. (1995). About the use of rank transformation in sensitivity analysis of model output. *Reliability Engrg. System Safety* **50** 225–239.

SALTELLI, A., TARANTOLA, S. and CHAN, K. (1999). A quantitative, model independent method for global sensitivity analysis of model output. *Technometrics* **41** 39–56.

SCHAIBLY, J. H. and SHULER, K. E. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. Part II. Applications. *J. Chem. Phys.* **59** 3879–3888.

SMIDTS, O. F. and DEVOOGHT, J. (1997). A variational method for determining uncertain parameters and geometry in hydrogeology. *Reliability Engrg. System Safety* **57** 5–19.

SOBOL', I. M. (1990). Sensitivity estimates for nonlinear mathematical models. *Mathematicheskoe Modelirovanie.* **2** 112–118. [Translated as Sensitivity analysis for nonlinear mathematical models. *Math. Modeling Comput. Experiment* **1** (1993) 407–414.]

TARANTOLA, S., JESINGHAUS, J. and PUOLAMAA, M. (2000). Global sensitivity analysis: a quality assurance tool in environmental policy modeling. In *Sensitivity Analysis* (A. Saltelli, K. Chan and M. Scott, eds.) 385–397. Wiley, New York.

TARANTOLA, S., PASTORELLI, R., BEGHI, M. G. and BOTTANI, C. E. (2000). A dataless pre-calibration analysis in solid state physics. In *Sensitivity Analysis* (A. Saltelli, K. Chan and M. Scott, eds.) 311–316. Wiley, New York.

TURANYI, T. (1990). Sensitivity analysis of complex kinetic systems. Tools and applications. *J. Math. Chem.* **5** 203–248.

YOUNG, P. (1999). Data-based mechanistic modeling, generalised sensitivity and dominant mode analysis. *Comput. Phys. Comm.* **117** 113–129.

WELCH, W. J., BUCK, R. J., SACKS, J., WYNN, H. P., MITCHELL, T. J. and MORRIS, M. D. (1992). Screening, predicting, and computer experiments. *Technometrics* **34** 15–47.