Sensitivity Analysis for Chemical Models

Andrea Saltelli,* Marco Ratto, Stefano Tarantola, and Francesca Campolongo

European Commission, Joint Research Centre of Ispra (I), Institute for the Protection and Security of the Citizen, TP 361, Via Enrico Fermi 1, 21020 Ispra (VA), Italy

Received November 22, 2004

Contents

1. Introduction

Chemists routinely create models of reaction systems to understand reaction mechanisms, kinetic properties, process yields under various operating conditions, or the impact of chemicals on man and the environment. As opposed to concise physical laws, these models are attempts to mimic the system by hypothesizing, extracting, and encoding system features (e.g. a potentially relevant reaction pathway versus another plausible one), within a process that can hardly be formalized scientifically.[1] The model will hopefully help to corroborate or falsify a given description of reality, e.g. by validating a reaction scheme for a photochemical process in the atmosphere, and possibly to influence it, e.g. by allowing the identification of optimal operating conditions for an industrial process or suggesting mitigating strategies for an undesired environmental impact.

These models are customarily built in the presence of uncertainties of various levels, in the pathway, in the order of the kinetics associated to the pathway, in the numerical value of the kinetic and thermodynamic constants for that pathway, and so on.

Propagating via the model all these uncertainties onto the model output of interest, e.g. the yield of a process, is the job of uncertainty analysis. Determining the strength of the relation between a given uncertain input and the output is the job of sensitivity analysis.[2]

Mathematical sensitivities (in the form of model output derivatives) are a straightforward implementation of this sensitivity concept. If the model output of interest is Y, its sensitivity to an input factor X_i is simply $Y'_{X_i} = \partial Y/\partial X_i$. This measure tells how sensitive the output is to a perturbation of the input. If a measure independent from the units used for Y and X_i is needed, $S^r_{X_i} = (\bar{X}_i/\bar{Y})(\partial Y/\partial X_i)$ can be used, where $\bar{X}_i$ is the nominal (or central, if a range is known) value of factor X_i and $\bar{Y}$ is the value taken by Y when all input factors are at their nominal value.

If factors are uncertain within a known or hypothesized range, then the measure $S^{\sigma}_{X_i} = (\sigma_{X_i}/\sigma_Y)(\partial Y/\partial X_i)$ can be of use, where the standard deviations σ_{X_i}, σ_Y are uncertainty analysis' input and output, respectively, in the sense that σ_{X_i} comes from the available knowledge on X_i, while σ_Y must be inferred using the model.

These sensitivity measures can be efficiently computed by an array of techniques, ranging from automated differentiation (where the computer program that implements the model is modified so that the sensitivities are computed with a modicum of extra execution time[3]) to direct methods (where the differential equations describing the model are solved directly in terms of species concentrations and their derivatives[4]). There is a vast amount of literature on these sensitivity measures,[5−11] which shall be referred to as *local* in the following. The majority of sensitivity analyses met with in chemistry and physics are local and derivative-based. Local sensitivities are useful for a variety of applications, such as the solution of inverse problems, e.g. relating macroscopic observables of a system, such as kinetic constants, to the quantum mechanics properties of the system,[6] or the analysis of runaway and parametric sensitivity of various types of chemical reactors.[8] Contexts where local sensitivity has been widely used are as follows: (1) to understand the reaction path, mechanism, or rate-determining steps in a detailed kinetic model with a large number of elementary reactions, e.g. in photochemistry or in combustion chemistry,[4,7,9] (see ref 12 for an alternative approach in this context), (2) to extract important elementary reactions from a complex kinetic model to obtain a reduced model (e.g. a minimal reaction scheme) with equivalent predictive power[7] or to select important reactions for further analysis,[13,14] (3) to estimate the output of a

* Corresponding author. Telephone: +39 0332 789686. Fax: +39 0332 785733. E-mail: andrea.saltelli@jrc.it.

10.1021/cr040659d CCC: $53.50
Published on Web 00/00/0000 PAGE EST: 18

Andrea Saltelli, born in 1953, graduated in Chemistry in 1976, first of his course, winning prizes from the University of Rome (La Sapienza) and the Italian Nuclear Authority. He produced a steady flow of publications in disciplinary journals over the last 30 years, on topics from Chemistry to Environmental Sciences to Applied Statistics and Econometrics. His focus was sensitivity analysis, where he tried to foster the use of quantitative methods in all settings where mathematical or computational models are used. Presently, he leads the Econometric and Applied Statistics Unit of the Joint Research Centre in Ispra. The Unit, with a staff of 25, develops models, indicators, and econometric–statistic applications, mostly in support of the Services of the European Commission. Andrea Saltelli is active in the organization of international courses and summer schools on sensitivity analysis.

Stefano Tarantola, born in 1967, is a Scientific Officer of the Joint Research Centre of the European Commission with an M.Sc. in engineering and a Ph.D. in Science and Technologies for engineering from the Polytechnic of Milan. He carries out and coordinates scientific R&D tasks in the field of economic and innovation indicators. He has conducted statistical work on indicators and composite indicators for the structural indicators initiative. He has experience in systems analysis and modeling, including an evaluation of the assumptions behind decision-making on environmental and social issues. He combines sensitivity analysis and participatory methods for the construction of composite indicators and develops methodologies for sensitivity analysis. He is the author of papers in the peer reviewed literature, coauthor of two books on sensitivity analysis, and an organizer of summer schools and other events on sensitivity analysis.

Marco Ratto, born in 1970, graduated in Chemical Engineering from the University of Genova in 1994. He received his Ph.D. in Chemical Engineering in 1998 with a dissertation on the "Sensitivity of dynamic predictions to parameter uncertainty". Until 2000, he carried out his research activity at the University of Genova. In the period 1998–2000, he also cooperated with the Institute for Physical Chemistry of Materials (ICFAM) of the Italian National Research Council (CNR) in Genova, carrying out experimental and theoretical studies of the dynamics of surface properties and transport phenomena at the interface molten metals–surrounding atmosphere. Since 2000, he has been working at the Joint Research Centre of the European Commission. His main research interests concern modeling, simulation, and estimation; global sensitivity and uncertainty analysis; and nonlinear dynamics, in a variety of applied disciplines—from chemical and process engineering systems, to environmental sciences, to macroeconomic modeling. He is author of papers in peer review journals and coauthor of a book on sensitivity analysis.

Francesca Campolongo, born in 1970, graduated in Applied Mathemtics from the University of Pisa, Italy, in 1993. In 1998, she completed her Ph.D. in modeling and sensitivity analysis at the Faculty of Environmental Sciences of Griffith University, QLD, Australia. Two Australian scholarships, OPRS and GUPRS, covered her Ph.D. In 1995 she was selected to participate in a 1-month summer school on ecological modeling at Cornell Univeristy (Ithaca, NY). Since March 1998, she has been working at the Joint Research Centre of the European Commission (Ispra, Italy), where she also obtained in 2002 a prize as "best JRC young scientist of the year". Her main research interests focus on modeling, uncertainty and sensitivity analysis, and risk assessment, applied to a variety of settings. During her Ph.D. and postdoctoral studies, she focused on environmental problems, with specific attention to the climate change issue. Since March 2000, she has developed expertise in the field of financial mathematics and financial risk analysis. She has over 20 scientific publications on refereed journals and is coauthor of two books on sensitivity analysis published by John Wiley & Sons.

computer program in the neighborhood of a given set of boundary conditions (BCs) without rerunning the program [This is often the rationale for using automated differentiation software.[3,15] This approach may be effective for deviations of up to 10−20% away from the baseline BC.], and, (4) for data assimilation, to reconcile model parameters with observations.[6,16,17]

In these local sensitivity measures the effect of X_i is taken while assuming all other factors fixed. This approach falls, hence, in the class of the one-factor-at-a-time (OAT) methods. A cruder approach often met in the literature is when, instead of derivatives, incremental ratios are taken by moving factors one

at a time away from their baseline value by a fixed (e.g. 5%) fraction, irrespective of the factor's presumed uncertainty range.[18]

Using derivative or otherwise OAT methods for the purpose of assessing the relative importance of input factors, e.g. to decide which factor mostly deserves better measurements, can only be effective if the model is linear in all its factors, unless some form of averaging of the system derivatives over the space of the input factors is performed. The same reservation holds if the purpose of the sensitivity analysis is to screen the factors as to fix those which are less influential on the output, although when the factors' derivatives differ by orders of magnitudes from one another, their use in screening might be safe enough. In general, though, a single measure such as $S^{\sigma}_{X_i} = (\sigma_{X_i}/\sigma_Y)(\partial Y/\partial X_i)$ can take care of the fact that different factors have different ranges of uncertainty but not of model nonlinearities due to the same factors.

Second-order derivatives can improve the sensitivity analysis offered by the first-order methods[14,19] and are useful for variational data assimilation.[16]

The chemists who in the 1970s applied Fourier transforms to sensitivity analysis were motivated by the realization that most models met in chemistry are of a rather nonlinear nature. Nonmonotonic, nonadditive features, to which we shall return later in this review, are also not uncommon. For these models, OAT methods can be of limited use if not outright misleading when the analysis aims to assess the relative importance of uncertain factors, and model-free measures are needed, possibly independent of assumptions about the model structure. To this effect, Cukier, Schaibly, and co-workers developed the Fourier amplitude sensitivity test (FAST;[20−23] see also ref 24, in which is given a variation of the basic scheme of FAST, defined as the Walsh amplitude sensitivity procedure (WASP), a method for discrete models where the factor variation is intrinsically two-valued) later made computationally available by Koda, McRae, and others.[25,26]

What FAST does, in a nutshell, is to decompose the variance $V = \sigma_Y^2$ of Y using spectral analysis, so that $V = V_1 + V_2 + ... + V_k + R$, where V_i is that part of the variance of Y that can be attributed to X_i alone, k is the number of uncertain factors, and R is a residual. Thus, $S_i = V_i/V$ can be taken as a measure of the sensitivity of Y with respect to X_i. We will offer a precise definition of V_i and R in section 2.

Although FAST is a sound approach to the problem, it has seen little use in the scientific community at large, including among chemists, and few applications of FAST are available in the literature.[27−30] Further chemical applications of FAST are cited in ref 4. FAST is mentioned in some reviews of sensitivity analysis methods[4,7,8,31−35] but ignored in others.[10,36]

At the time of its development, FAST was laborious to implement and computationally expensive. Some researchers were uncomfortable with moving away from the mathematical appeal of derivatives. [The review System Analysis at Molecular Scale by H. Rabitz[6] states, "Generally the emphasis in sensitivity analysis is not on the explicit introduction of any particular variation [in the input] but rather on the computation and examination of the sensitivity coefficients [the derivatives], as they are independent of any special variation in the parameters". The same author later contributed considerably to the development of global sensitivity analysis by introducing a new class of high-dimensional model representation (section 2).]

In the 1990s, several investigators, sometimes without realizing it, developed Monte Carlo-based estimates of the FAST sensitivity measure; see refs 37−44. In these approaches, the space of the input factors was explored using Monte Carlo-based techniques (such as Latin hypercube sampling, random or quasi-random numbers), rather than by the Fourier trajectories, but in general the same sensitivity measure $S_i = V_i/V$ was computed. The best formalization of this approach is due to Sobol'.[42]

At the same time, practitioners in risk analysis were using for the purpose of sensitivity analysis various kinds of Monte Carlo (MC)-based linear regression.[31] In these approaches, the space of the input factors is sampled via a MC method and a regression model is built (e.g. by a least-squares calculation) from the estimates of Y produced by the model.

We anticipate here that the standardized regression coefficients β_i thus generated, where $\beta_i = (\sigma_{X_i}/\sigma_Y)b_i$ and b_i is the raw regression coefficient, are related to the FAST measure, as well as to $S^{\sigma}_{X_i}$. In fact, for linear models $\beta_i^2 = S_i = (S^{\sigma}_{X_i})^2$. We show in ref 45 how $S^{\sigma}_{X_i}$ is an effective measure for linear models, β_i is an effective measure for moderately nonlinear models, for which an effective linear regression model can be built, and S_i is the model-free extension that works even for strongly nonlinear models.

Rabitz and co-workers[46,47] developed alternative strategies for global sensitivity analysis that are inspired by the work of Sobol', and, hence, homologous to FAST, but are based on decomposing the model Y of interest on the basis of finite differences built along lines, planes, and hyperplanes that pass through a given selected point in the space of the input factors. Rabitz's approach—which can be seen as "in between" derivative-based methods and variance-based ones—can also be used to compute the variance-based sensitivity indices S_i.

Finally, it is important to note that using variance-based techniques in numerical experiments is the same as applying ANOVA [analysis of variance[33,48] (here and in the following we shall list references in chronological order)] in experimental design, as the same variance decomposition scheme holds in the two cases. One could hence say that modelers are converging with experimentalists treating Y, the outcome of a numerical experiment, as an experimental outcome whose relation to the control variables, the input factors, can be assessed on the basis of statistical inference. Sacks, Welch, and others[49,50] were among the first to think along these lines. Morris[51] developed an effective experimental design scheme for numerical experiments which has similarities to a variance-based measure.

This convergence of FAST-based and sampling-based strategies for sensitivity analysis and also a convergence between these and experimental design theory, which escaped even recent reviews,[8] vindicate the original intuition of the FAST developers that S_i was a good measure for chemical models. Both FAST and MC computation schemes have been upgraded in recent years,[2,32,45,52−54] becoming less expensive and easier to apply, as we shall see in the next sections.

As already mentioned, OAT approaches still dominate the chemical literature even when the purpose of the analysis is to assess the relative importance of input factors in the presence of factors uncertainty.

To contrast this practice, we shall select and review those which we consider as the most promising modern approaches to sensitivity analysis, with some emphasis on the methods in the FAST family, comparing, with the help of worked examples, the performances of variance-based methods with different kinds of local, regression, or screening-based measures.

2. Methods

2.1. A Simple Example, Using Local Sensitivity Analysis

It would be impossible to describe all sensitivity analysis methods within the purpose of the present work. Available reviews are refs 4−8, 10, 11, 18, 31, 34−36, and 55−57. Those of Rabitz,[5,6] Turányi and Tomlin,[4,7] and Varma and Morbidelli[8] are of particular interest for chemists. References 31 and 34 cover well Monte Carlo-based regression methods, ref 10 focuses on local strategies, and ref 56 focuses on experimental design methods.

A wide spectrum of different perspectives and approaches can be found in ref 2, a multiauthor book with input from practitioners such as Rabitz, Turányi, Helton, Sobol', the authors of the present review, and others. Our plan here is to offer a selection of sensitivity analysis methods, with emphasis on global methods, which might be of relevance to and applicable by the *Chemical Review* readership. The methods are illustrated by examples. We start with a simple reversible chemical reaction A ↔ B, with reaction rates k_1 and k_{-1} for the direct and inverse reactions, respectively, whose solution, for the initial conditions (ICs)

$$[\mathrm{A}](t=0) = [\mathrm{A}]_0$$

$$[\mathrm{B}](t=0) = 0$$

is

$$[\mathrm{A}] = \frac{[\mathrm{A}]_0}{k_1 + k_{-1}}(k_1 \mathrm{e}^{-(k_1+k_{-1})t} + k_{-1}), \qquad [\mathrm{B}] = [\mathrm{A}]_0 - [\mathrm{A}] \quad (1)$$

This "model" is so simple as to allow a characterization of the system sensitivity by analytic methods, but we shall work it out by pretending it is a system of partial differential equations for a large reaction model, as tackled by chemists using solvers such as

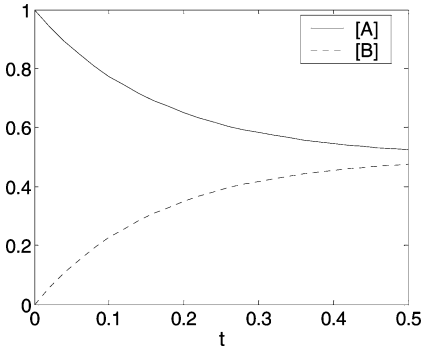

Figure 1. Time evolution of [A] and [B] from eq 1.

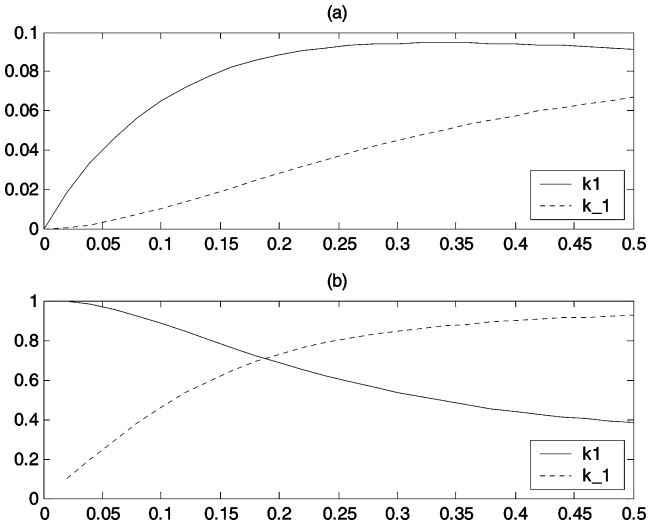

Figure 2. Absolute values of (a) $[\mathrm{A}]'_{k_1} = \partial[\mathrm{A}]/\partial k_1$ and $[\mathrm{A}]'_{k_{-1}} = \partial[\mathrm{A}]/\partial k_{-1}$ and (b) $S^{\sigma}_{k_1} = (\sigma_{k_1}/\sigma_{[\mathrm{A}]})(\partial[\mathrm{A}]/\partial k_1)$ and $S^{\sigma}_{k_{-1}} = (\sigma_{k_{-1}}/\sigma_{[\mathrm{A}]})(\partial[\mathrm{A}]/\partial k_{-1})$.

FACSIMILE,[58] CHEMKIN,[59] or others, where the relative effect of the uncertain inputs in determining the uncertainty on the output of interest is unknown. We assume

$$k_1 \sim N(3,0.3)$$
$$k_{-1} \sim N(3,1) \quad (2)$$

where the symbol ∼ stands for "distributed as" and N stands for normal distribution. Thus, both uncertain factors are normally distributed with mean 3. The standard deviation is 0.3 for k_1 and 1 for k_{-1}.

Figure 1 gives the time evolution of A, B while Figure 2 gives the absolute values of the pure local sensitivity of A with respect to the factors, i.e.

$$[\mathrm{A}]'_{k_1} = \frac{\partial[\mathrm{A}]}{\partial k_1}$$

$$[\mathrm{A}]'_{k_{-1}} = \frac{\partial[\mathrm{A}]}{\partial k_{-1}}$$

computed at the "nominal" value $k_1 = k_{-1} = 3$, as well as the absolute values of the pseudoglobal sensitivities:

$$S^{\sigma}_{k_1} = \frac{\sigma_{k_1}}{\sigma_{[\mathrm{A}]}}\frac{\partial[\mathrm{A}]}{\partial k_1}, \quad S^{\sigma}_{k_{-1}} = \frac{\sigma_{k_{-1}}}{\sigma_{[\mathrm{A}]}}\frac{\partial[\mathrm{A}]}{\partial k_{-1}}$$

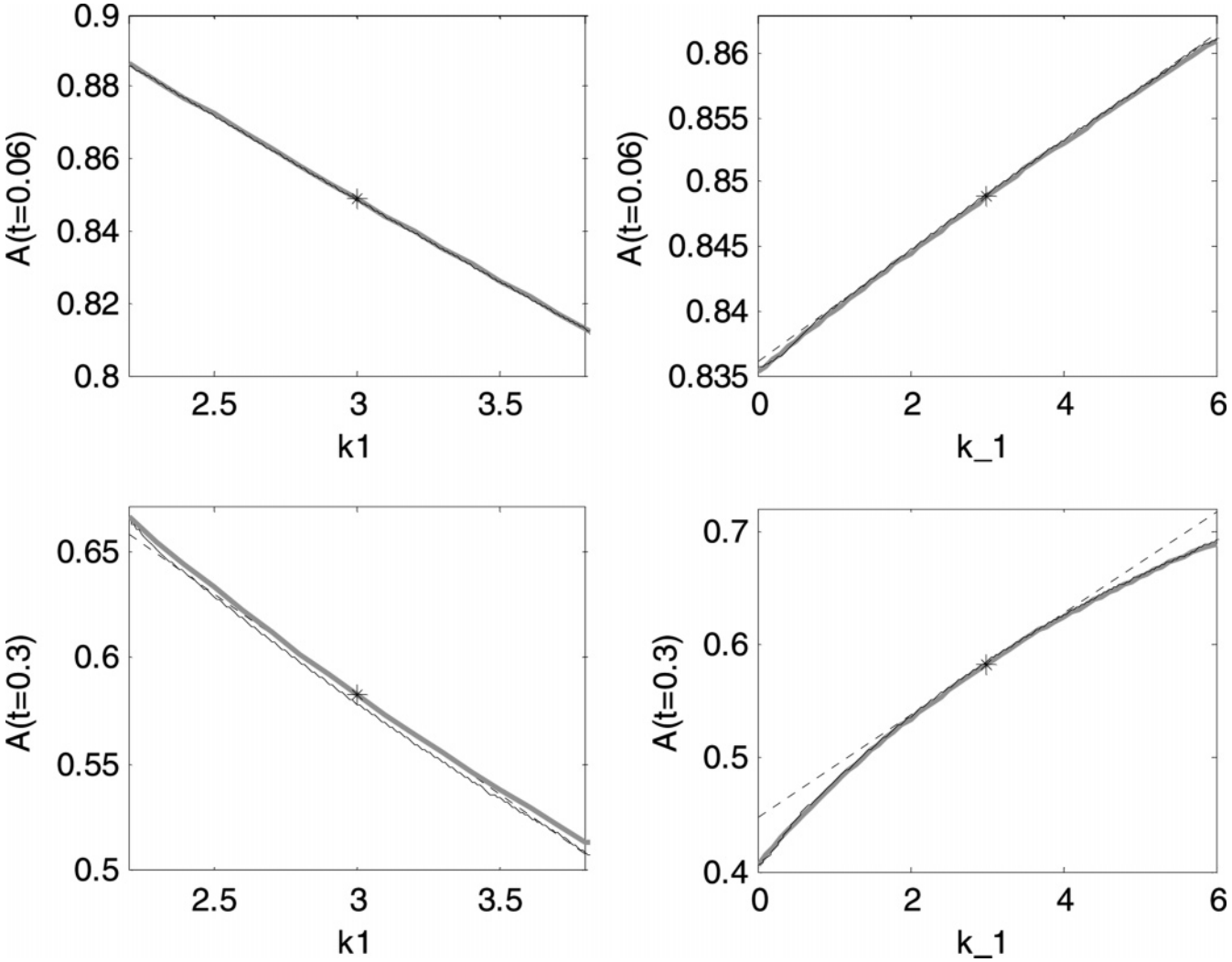

Figure 3. First-order Taylor expansion (dotted lines); first cut-HDMR decomposition (bold gray lines); and first-order HDMR decomposition (solid black lines). The star represents the base value, at which the Taylor and cut-HDMR expansions are constructed.

The value of $\sigma_{[A]}$ used in Figure 2 is not the exact one but has been computed using the approximation $\sigma_Y^2 \cong \sum_{i=1}^{k} \sigma_{X_i}^2 (\partial Y/\partial X_i)^2$ to model $Y = f(X_1,X_2,...,X_k)$, which for our simple two-factor model of eq 1 gives

$$\sigma_{[A]}^2 \cong \sigma_{k_1}^2 \left(\frac{\partial [A]}{\partial k_1}\right)^2 + \sigma_{k_{-1}}^2 \left(\frac{\partial [A]}{\partial k_{-1}}\right)^2, \text{ or equivalently}$$

$$(S_{k_1}^{\sigma})^2 + (S_{k_{-1}}^{\sigma})^2 = 1 \qquad (3)$$

We discuss the applicability of this approximation in a moment. It is evident from the plot that $S_{k_1}^{\sigma}$ and $S_{k_{-1}}^{\sigma}$ offer a more realistic picture of the relative importance of k_1 and k_{-1} in determining the uncertainty of [A] than $[A]'_{k_1}$ and $[A]'_{k_{-1}}$, as the sigma-normalized sensitivity measures are capable of weighting the larger role of k_{-1} which descends from its larger standard deviation. This is not to say that $[A]'_{k_1}$ and $[A]'_{k_{-1}}$ are useless. We have used them, for instance, to compute (Figure 3) an approximate map of [A] as a function of k_1 and k_{-1} using a simple Taylor expansion. This kind of approximation[60] becomes very valuable when the model under analysis is expensive to run. [Throughout this work, computation cost must be understood as the number of times one needs to run the model that computes $Y = f(X_1,X_2,...,X_k)$. The time needed to compute the sensitivity tests is usually negligible by comparison.] More accurate tools for this kind of model approximation, also shown in Figure 3, are discussed in section 2.4. Note that the entire Taylor representation in Figure 3 was built using for each plot just the function value at $k_1 = k_{-1} = 3$ plus the two derivatives $[A]'_{k_1}$ and $[A]'_{k_{-1}}$ at the same point.

Derivative-based measures such as $\partial[A]/\partial k_1$ and $S_{k_1}^{\sigma}$ are members of a large class of local measures used in chemistry, which includes, e.g., functional sensitivities, feature sensitivities, and others. These are extensively dealt with in the literature. There are furthermore a variety of methods to compute these derivatives for large systems of differential equations, such as the Green functions method, the direct method, the decoupled direct method, the adjoint method, and others.[7,8,11] Given that the emphasis of the present review is on global methods, we will not describe them here. Automated differentiation methods, whereby the modeler modifies the simulation model so that it can compute derivatives with a

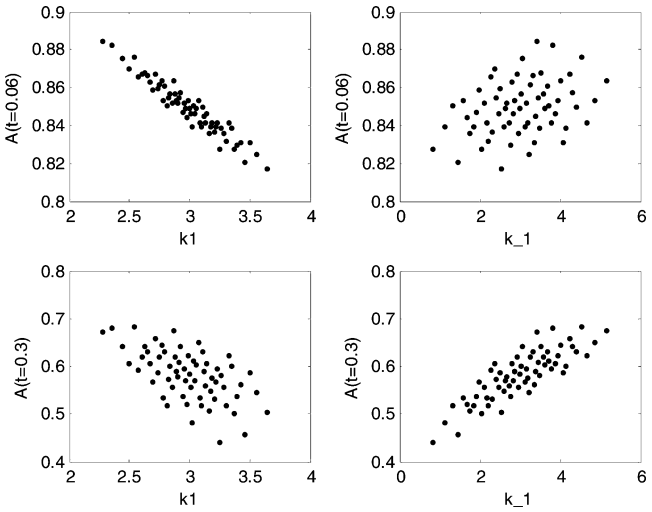

Figure 4. Scatter plots of [A] versus k_1,k_{-1} at time t = 0.3 and t = 0.06. k_1 appears more influential than k_{-1} at t = 0.06, while the reverse is true at t = 0.3.

minimum of extra computing time, are also extensively used in chemistry.[3]

The $S^{\sigma}_{k_1}$ and $S^{\sigma}_{k_{-1}}$ curves in Figure 2 are an example of quantitative, albeit approximate, sensitivity analysis. Imagine that k_1 and k_{-1} are two poorly known rates of a more complex system and that we need to know which of them—once properly determined in the laboratory—would give us better chances to reduce the uncertainty in the output [A] of interest. Figure 2 would allow us to say that if we need to reduce the uncertainty of [A] at the beginning of the process (t = 0.05), then k_1 is a better candidate than k_{-1}. The reverse is true if we are interested in [A] near equilibrium (t = 0.4). Considerations of this nature are also relevant for the purpose of model calibration (see section 2.6). The reader will believe Figure 2 conditionally upon the truth of eq 3, which tells us that the variance of [A] can be partitioned in bits proportional to the squared $S^{\sigma}_{k_1}$ and $S^{\sigma}_{k_{-1}}$. We reconsider the validity of this approximation in the next section.

2.2. Monte Carlo Methods on the Simple Example

We move now into Monte Carlo simulation, drawing independent samples from the distribution in eq 2 and running the computer program that evaluates the solutions to eq 1. This would be a slow step if eq 1 were a system of many partial differential equations. Imagine we run eq 1 just 64 times, obtaining 64 different estimates $[A]_i$, with i = 1, 2, ..., 64, each corresponding to a sample k_{1i},k_{-1i}, where the two factors have been drawn independently from one another but in such a way that each of them is a sample from its distribution (eq 2).

One way to use these estimates in sensitivity analysis is to make scatter plots (Figure 4) which allow a visual impression of the system degree of linearity.[31,61] For systems with tens of uncertain factors, scatter plots become impractical, and a more concise description of the system sensitivities can be obtained by feeding $[A]_i$, k_{1i}, and k_{-1i}, into a regression algorithm, searching for a linear model of the form

$$[A] = b_0 + b_1k_1 + b_{-1}k_{-1} \quad (4)$$

The b_i coefficients are dimensioned, and the common use is to standardize them, rewriting eq 4 as

$$[\tilde{A}] = \beta_1\tilde{k}_1 + \beta_{-1}\tilde{k}_{-1} \quad (5)$$

where $[\tilde{A}] = ([A] - \mu_{[A]})/\sigma_{[A]}$ and $\tilde{k}_i = (k_i - \mu_{k_i})/\sigma_{k_i}$ are standardized variables, $\mu_{[A]}$ and μ_{k_i} are the mean of [A] and k_i, respectively, $\sigma_{[A]}$ and σ_{k_i} are the standard deviations, $\beta_i = (\sigma_{[A]}/\sigma_{k_i})b_i$ are the so-called standardized regression coefficients (SRC), and we have used k_i to indicate either k_1 or k_{-1}. It is a known result of linear regression analysis[62] that if the factors are independent and the model is linear, then for the model in eq 1:

$$\beta_1^2 + \beta_{-1}^2 = 1 \quad (6)$$

If the model, as in our case, deviates from linearity, then the sum of the squared β's will quantify the deviation. More often, this statistics is computed directly from the simulation data and the regression model results:

$$R^2_{[A]} = \sum_{j=1}^{N} \frac{([A]_i^* - \mu_{[A]})^2}{([A]_i - \mu_{[A]})^2} \quad (7)$$

where N is the number of simulations, 64 in this case, $[A]_i$ are the simulation results, and $[A]_i^*$ are the values of [A] provided by the regression model (eq 4). $R^2_{[A]}$, known as the model coefficient of determination, is a positive number in [0,1] which indicates which fraction of the original model variance is explained by the regression model. When this number is high, e.g., 0.7 or higher, then we can use the standardized regression coefficients for sensitivity analysis, albeit at the price of remaining ignorant about that fraction of the model variance not explained by the SRCs. An application of this strategy to a model for tropospheric oxidation of dimethyl sulfide is in ref 63, where a rather high value of R^2 allowed factors to be ranked confidently in a system with about 50 temperature-dependent chemical reactions.

Note that an improvement of the performance of the regression-based approach, i.e., a higher value of R^2_Y, can be obtained by transforming both the input and the output sample vectors to ranks, e.g., rank N for the highest Y and rank 1 for the lowest, and the same for the input factors.[64] Rank transformation can substantially linearize a nonlinear, albeit monotonic, function. The problem with this approach is that the conclusions of the sensitivity analysis apply to the rank version of the model rather than to the model itself.[65]

The identity $\beta^2_{k_i} = (S^{\sigma}_{k_i})^2$ for a linear model is evident (eqs 3 and 6). Yet, when the model is even partly nonlinear, the standardized regression coefficients are superior to the normalized derivatives, first, because they allow the degree of nonlinearity of the model to be estimated and, second, as they offer a measure of the effect of, e.g., k_1 on [A] which is

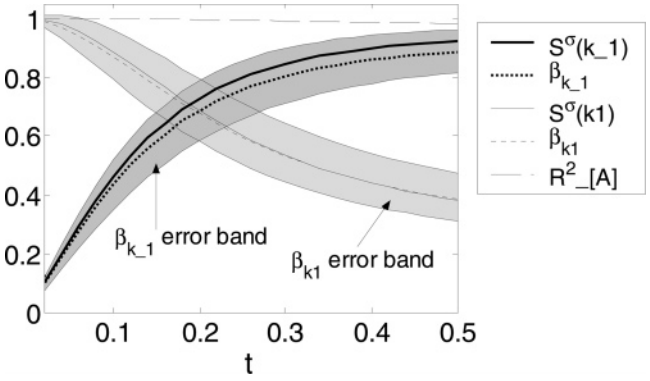

Figure 5. (i) $R^2_{[A]}$ versus time; (ii) $S^{\sigma}_{k_1}$ and β_{k_1} with confidence bounds versus time (descending lines); (iii) $S^{\sigma}_{k_{-1}}$ and $\beta_{k_{-1}}$ with confidence bounds versus time (ascending lines). Error bounds are based on a bootstrap analysis of size 100 of the 64 original data points.

averaged over a sample of possible values for k_{-1}, as opposed to being computed at the fixed point $k_{-1} = 3$, as was the case for $S^{\sigma}_{k_i}$.

In Figure 5 we have plotted the $S^{\sigma}_{k_i}$ and β_{k_i} for both k_1 and k_{-1}. We see that for this test model (eq 1) the two measures are equivalent. The model coefficient of determination for these data is very high and ranges from $R^2_{[A]} = 1$ at the initial time points to $R^2_{[A]} = 0.98$. To stay with our pretension that the model in eq 1 is something expensive to estimate, we have performed only 64 model evaluations, and we have, hence, computed the error bounds (2 standard deviations) on the β_i's using bootstrap; i.e., we have re-estimated the regression (eq 4) by drawing 100 different samples (with replacement) of size 64 of the original 64 simulations. This procedure helps us to decide whether N should be increased beyond the present value of 64, e.g. when the uncertainty bounds are too wide for our purposes.

On the basis of the value of $R^2_{[A]}$ and on the basis of Figure 5, we can say that the multidimensional averaging provided by the β_{k_i}'s with respect to the k_i's does not make much of a difference, so that eq 3 is a valid approximation for this model. Yet what would happen if we had a model with a low R^2_Y? Obtaining a measure of the average local sensitivity, e.g., $\langle \partial c_i/\partial k_l \rangle$, the effect of kinetic rate l over the concentration of species i averaged over all the space of the uncertain input k's, was the main motivation of Cukier and co-workers in the 1970s to develop a new method, the Fourier amplitude sensitivity test (FAST), suited to nonlinear problems.

Before moving into a discussion of it, let us make the point of the preceding discussion. Derivatives can provide detailed information about system behavior. Yet, usual practice to estimate them at a single point in the space of the input limits their applicability away from that point, which is something needed in sensitivity analysis when there are sensible uncertainties in the input factors. Equation 3 can help, by allowing a combination of local derivatives and input factors variation. Yet to trust eq 3 on its own, without a check of the linearity of the system, would be unwarranted. A quick glance at the space of the input, even with a handful of data points as done by our regression analysis of size 64, is straightforward and safer. The model coefficient of determination would help to identify problems, such as, e.g., a nonmonotonicity between input and output (Figure 6), which would result in low or zero β_i and S^{σ}_i even for an influent factor X_i. In this case, in order to build a regression-based sensitivity analysis, one would have to look at various trial nonlinear regression models, e.g., via a brute force search for the most convenient regression variables (and combinations of). Software packages are available to do that, but the search may be computationally expensive.

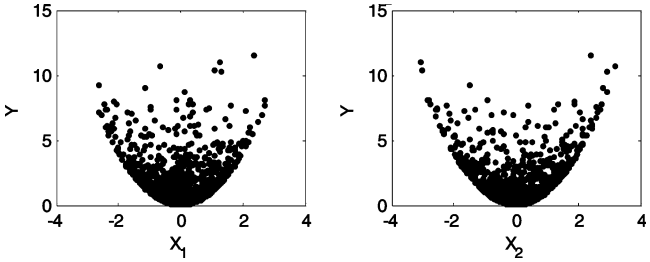

Figure 6. Scatter plot for model $Y = \sum_{i=1}^{k} \alpha_i X_i^2$ and $X_i \sim N(0,1)$ for all i's. β_i's are zero for this model ($k = 2$; $\alpha_i = 1$ in the plots). FAST indices are instead $S_i = \alpha_i^2/\sum \alpha_i^2$.

2.3. Fourier Amplitude Sensitivity Test (FAST)

FAST[20−27] is a sensitivity analysis method which works irrespective of the degree of linearity or additivity of the model. Let us call S_i the FAST-based sensitivity index for a generic factor X_i which feeds into a model $Y = f(X_1,X_2,...,X_k)$ with k independent uncertain factors. We can assume that all factors are uniformly distributed in [0,1] (This is standard practice.[2] Input factors can then be mapped from the unit hypercube to their actual distribution.), so that the space of definition of f is a unit hypercube in k dimensions.

We plan to illustrate that S_i is a model-free extension of the standard regression coefficients β_i, in the sense that $S_i = \beta_i^2$ for linear models.

Before showing how S_i can be arrived at using Fourier analysis, a possibly intuitive description is now given. We ask ourselves if we can reduce the uncertainty in Y by removing the uncertainty (e.g. by measuring directly) in some of the X_i's. There will surely be factors more promising than others for this kind of analysis. Let us call $V(Y)$ the variance of Y, and $V_{\mathbf{X}_{-i}}(Y|X_i = X_i^*)$ the variance that we would obtain if we could fix X_i to some value. The subscript $\mathbf{X}_{-i}$ of V is to remind us that this variance is taken over all factors other than X_i, which is fixed.

X_i^* could be the true value of X_i determined with a measurement. If we could compute $V_{\mathbf{X}_{-i}}(Y|X_i = X_i^*)$ for all factors at all points, we would surely find the one with the smallest $V_{\mathbf{X}_{-i}}(Y|X_i = X_i^*)$, but at this point, we would be past sensitivity analysis, having determined all uncertain factors. What we can do before having actually taken the measure, i.e., before knowing X_i^*, is to look at what would be obtained if we took the average of $V_{\mathbf{X}_{-i}}(Y|X_i = X_i^*)$ over the possible values of X_i^*. We would write this as $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$. We have dropped the dependency from

X_i^*, as the quantity $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$ is independent of any particular point in the distribution of X_i. This quantity is relevant to the solution of the problem just posed; i.e., the factor with the smallest $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$ would appear to be the desired most promising candidate for measurement, in terms of expected reduction of the variance of Y. A known algebraic result is that

$$V(Y) = E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i)) + V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) \quad (8)$$

so that the smallest $E_{X_i}(V_{\mathbf{X}_{-i}}(Y|X_i))$ will correspond to the largest $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$. The FAST-based sensitivity index is simply

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))}{V(Y)} \quad (9)$$

S_i is hence the educated answer to the question: "If one wants to reduce the variance of the output, which factors should be fixed first?" We have called this question "factor prioritization setting" and have argued that such a framing of the sensitivity analysis with respect to a setting allows an otherwise vague concept of factor importance to be clarified.[45,54] Not following this practice could result in several tests being thrown at the problem and in several rankings of factor importance being obtained, without a basis to decide which one to believe.

The values of S_{k_1} and $S_{k_{-1}}$ for the model in eq 1 would practically coincide with the β_i of Figure 5, due to model's quasi-linearity, and hence, we do not plot them here. Before introducing the next model, which will make these measures worth using, we briefly illustrate the theory behind FAST and the related method of Sobol'.[42]

When using FAST, S_i is computed by exploring the k-dimensional space of the input factors with a search curve defined by a set of parametric equations

$$X_i = G_i \sin(\omega_i s) \quad (10)$$

with $i = 1, 2, ..., k$, where s is a scalar varying in $(-\infty,+\infty)$, the ω_i are a set of different angular frequencies associated with each factor, and the G_i are properly selected transformation functions. Scanning eq 10 for different values of s results in a curve in the k-dimensional hypercube whereby each dimension is explored with a different frequency ω_i. Fourier analysis allows, then, the computation of $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$ on the basis of the signal at ω_i and its harmonics. The implementation of the method requires care, mostly in avoiding interferences, based on accurate selection of the set of k frequencies ω_i.[26,27] Extensions of the FAST method are described in refs 53 and 66.

Cukier and co-workers had noted[23] that the FAST indices could be seen as the first-order terms in the decomposition of the unconditional variance, which for independent factors can be written as

$$V(Y) = \sum_i V_i + \sum_{i<j} V_{ij} + \sum_{i<j<l} V_{ijl} + ... + V_{123...k} \quad (11)$$

where

$$V_i = V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$$

$$V_{ij} = V_{X_iX_j}(E_{\mathbf{X}_{-ij}}(Y|X_i,X_j)) - V_i - V_j$$

$$V_{ijl} = V_{X_iX_jX_l}(E_{\mathbf{X}_{-ijl}}(Y|X_i,X_j,X_l)) - V_{ij} - V_{jl} - V_{il} - V_i - V_j - V_l$$

$$... \quad (12)$$

Note that in writing, e.g., $V_{X_iX_j}(E_{\mathbf{X}_{-ij}}(Y|X_i,X_j))$ we mean that the inner expectation is over all factors but X_i,X_j and the outer variance is over X_i,X_j.

In classic FAST, only the main effect terms S_i are computed, and the success of a given analysis is empirically evaluated by the sum of these terms: if this is high, as a rule of the thumb greater than 0.6,[30] then the analysis is successful.

The V_i describe the so-called "additive" part of a model. In turn, additive models are defined as those for which $\sum_i S_i = 1$. Extended FAST[53] allows the computation of higher order terms.

It is easy to verify[45] that for linear models both relations $\sum_i \beta_i^2 = 1$ and $\sum_i S_i = 1$ hold. Yet the second relationship holds also for models that are nonlinear albeit additive. To make a trivial example, $Y = \sum_{i=1}^{k} \alpha_i X_i^2$ is nonlinear and nonmonotonic if the distribution function of the X_i's is centered at zero (Figure 6). Yet, this is an additive model, for which $V(Y) = \sum_i V_i$ and $\sum_i S_i = 1$. For our almost linear model (eq 1), $\sum_i S_i = S_{k_1} + S_{k_{-1}} = 0.995$ at $t = 0.06$ and $=0.991$ at $t = 0.3$, while the remaining bit of variance, $S_{k_1k_{-1}} = 0.005/0.009$ for $t \in [0,0.3]$, describes the insignificant interaction effect of k_1 and k_{-1}. In experimental design, $S_{k_1k_{-1}}$ is also known as a second-order or two-way effect.

Unlike the case of local sensitivity analysis, where the cost of computing a second-order term is, in general, affordable,[3] terms of higher order in eq 11 are seldom used in global sensitivity analysis, because of their number and computational cost. There are in fact

$$\binom{k}{2}$$

terms of the second-order (V_{ij}),

$$\binom{k}{3}$$

terms of the V_{ijl} type, and so on, for a total of $2^k - 1$ terms in eq 11. This problem is known among practitioners as "the curse of dimensionality". It has been argued[46,47] that terms above the second-order ones should not be too frequent in sound models of physical systems, but we find this assumption unsafe.

2.4. Monte Carlo-Based Version of FAST and the Work of Sobol'

Several Monte Carlo-based estimates of the first-order terms V_i have been proposed.[37−44] It will be sufficient here to consider the work of Sobol'.[42] Sobol' noted that the function itself can be decomposed into

terms of increasing dimensionality, i.e.,

$$Y = f(X_1,X_2,...,X_k) = f_0 + \sum_i f_i(X_i) + \sum_{\substack{ij \\ i<j}} f_{ij}(X_i,X_j) + ... + f_{12...k} \quad (13)$$

There are infinite ways of decomposing f, but for independent factors there is a unique decomposition in orthogonal terms which ensures [As described in ref 42, $f_i = E(Y|X_i) - E(Y)$ and $f_{ij} = E(Y|X_i,X_j) - E(Y|X_i) - E(Y|X_j) + E(Y)$, where $E(Y) = f_0$.]

$$V_i = V(f_i)$$

$$V_{ij} = V(f_{ij}) \quad (14)$$

and so on. The f_i's, f_{ij}'s, ... are known as ANOVA-HDMR,[46] where HDMR stands for high-dimensional model representation and ANOVA refers to the analysis of variance from experimental design.[48]

Sobol' offered a Monte Carlo strategy to compute indices of any order, that is based on a Monte Carlo exploration of the input space. To make an example, to estimate V_i, the following algorithm is used

$$\tilde{V}_i = \sum_{j=1}^{N} f(x^a_{j1},x^a_{j2},\cdots,x^a_{ji},\cdots,x^a_{jk}) \times f(x^b_{j1},x^b_{j2},\cdots,x^a_{ji},\cdots,x^b_{jk}) - f_0^{\,2} \quad (15)$$

where N is the sample size of a MC simulation, k is the number of independent factors, and the superscripts a and b stand to indicate that different independent input matrixes have been used:

$$\mathbf{A} = \begin{pmatrix} x^a_{11} & \cdots & x^a_{1k} \\ \cdots & \cdots & \cdots \\ x^a_{N1} & \cdots & x^a_{Nk} \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} x^b_{11} & \cdots & x^b_{1k} \\ \cdots & \cdots & \cdots \\ x^b_{N1} & \cdots & x^b_{Nk} \end{pmatrix} \quad (16)$$

Equation 15 says that in order to compute $\tilde{V}_i$ one has to resample all factors but X_i. While it takes some reasoning[67] to demonstrate that $\tilde{V}_i$ is an estimate of the partial variance $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i))$, it is intuitive that $\tilde{V}_i$ is large when X_i is influent. If X_i controls the output, then large values of f will be multiplied with one another in eq 15, and the same is true for low values. If X_i is non-influent, low and high values of f will be randomly combined, resulting in a lower value of $\tilde{V}_i$.

Sobol' noted that an important objective of the sensitivity analysis is to identify those factors that have no sensible effect on the output. To tackle the problem, he rewrote eq 11 for sets of factors as

$$V(Y) = V_{\mathbf{U}}(E_{\mathbf{Z}}(Y|\mathbf{U})) + V_{\mathbf{Z}}(E_{\mathbf{U}}(Y|\mathbf{Z})) + V_{\mathbf{UZ}} \quad (17)$$

where all factors in $\mathbf{X}$ have been partitioned in two sets, (1) a trial set $\mathbf{U}$ of supposedly noninfluential factors and (2) the remaining factors $\mathbf{Z}$. $V_{\mathbf{UZ}}$ is the pure interaction effect between the two sets and can be easily computed by difference. If $V(Y) = V_{\mathbf{Z}}(E_{\mathbf{U}}(Y|\mathbf{Z}))$, then one can conclude that the set $\mathbf{U}$ is truly noninfluential. Note that the condition of noninfluence implies $V_{\mathbf{U}}(E_{\mathbf{Z}}(Y|\mathbf{U})) + V_{\mathbf{UZ}} = 0$, and not simply $V_{\mathbf{U}}(E_{\mathbf{Z}}(Y|\mathbf{U})) = 0$. We shall go back to this in a moment.

As mentioned in the Introduction, decompositions such as in eqs 11 and 13 are common in experimental design, where one varies control variables; e.g., for a chemical experiment one would vary temperature, catalyst, and concentrations among a set of preestablished levels (often just two, high and low, for each variable) as to maximize the number of effects (first order, second order, ...) one can determine for a given cost in terms of number of experiments.[48] More precisely, if we had sampled the points in the k-dimensional hypercube with a n^k factorial design as $\{0, 1/(n-1), ..., (n-2)/(n-1), 1\}$, then the $f_i(X_i)$ of eq 13 would be identical to the ANOVA estimate of the main effect in a complete factorial.[33,48] It has also been suggested to use the functions in eq 13 directly for the purpose of sensitivity analysis.[49]

We have plotted in Figure 3 f_{k_1} and $f_{k_{-1}}$ for the model in eq 1. No other functions are needed for this model, as $[A] = f_{k_1} + f_{k_{-1}} + f_{k_1k_{-1}}$ and the term $f_{k_1k_{-1}}$ is almost flat at zero.

This direct use of functional decompositions such as f_{k_i} for sensitivity analysis is elegant and informative but can become impractical when the number of factors and/or of their nonzero interactions increases.

Representations such as eq 13 of multivariate functions by superposition of simpler functions such as projection pursuit, radial basis functions, and others are common in Mathematics[46] and have a long history.[47,68]

Rabitz and co-workers[46,47] further proposed an alternative model representation that is based on knowing the model values on lines, planes, and hyperplanes that pass through a selected point in the space of the input factors. He calls these cut-HDMR, to distinguish them from the ANOVA-HDMR of eq 11.

An example of cut-HDMR for the model in eq 1 is also in Figure 3, together with the Taylor expansion (based on local derivatives at the midpoint) of the same factors.

The three model representations are substantially equivalent for this quasi-linear model.

The cut-HDMR has been applied to chemical problems.[69−72] An interesting property of the cut-HDMR is that it can be used as a basis for efficiently computing the ANOVA-HDMR (the terms in eq 11). We shall discuss the merits of the HDMR with an application in the next section. Yet, the method still depends on the assumption that in eq 13 terms of order higher than two or three are negligible.

2.5. The Second Test Case

We move now to show that it is better to have global sensitivity analysis in terms of FAST or Monte Carlo estimates of S_i, rather than in terms of β_i or derivatives. [Although, as discussed in section 2.2,

S_i^σ and β_i can provide an approximate estimate of S_i.] To do this, we introduce our second example.

This is a thermal runaway analysis of a batch reactor, with exothermic reaction A → B

$$\frac{d[A]}{dt} = -k(T)[A]^n \tag{18}$$

where n is the order of the reaction, and

$$\rho c_v \frac{dT}{dt} = (-\Delta H)k(T)[A]^n - s_v u(T - T_a) \tag{19}$$

where ρ is the density of the fluid mixture [kg/m^3], c_v is the mean specific heat capacity of the reaction mixture [J/(K mol)], ΔH is the molar reaction enthalpy [J/mol], s_v is the surface area per unit volume [m^2/m^3], u is the overall heat transfer coefficient [J/(m^2 s K)], and T_a is the ambient temperature.

The initial conditions are $[A] = [A]_0$, $T = T_0$, and $t = 0$.

This is customarily rewritten in dimensionless form:

$$\frac{dx}{d\tau} = \exp\left(\frac{\theta}{1 + \theta/\gamma}\right)(1 - x)^n = F_1(x,\theta)$$

$$\frac{d\theta}{d\tau} = B \exp\left(\frac{\theta}{1 + \theta/\gamma}\right)(1 - x)^n - \frac{B}{\psi}(\theta - \theta_a) = F_2(x,\theta)$$

with initial conditions (IC's) $x = 0$ and $\theta = 0$, at $\tau = 0$, and the dimensionless variables

$$x = \frac{[A]_0 - [A]}{[A]_0} \quad \theta = \frac{T_0 - T}{T_0}\gamma \qquad \tau = tk(T_0)([A]_0)^{n-1}$$

and dimensionless parameters

$$B = \frac{(-\Delta H)[A]_0}{\rho c_v T_0}\gamma\text{: dimensionless heat of reaction}$$

$$\gamma = \frac{E}{R_g T_0}\text{: dimensionless activation energy}$$

$$\psi = \frac{(-\Delta H)k(T_0)([A]_0)^n}{s_v u T_0}\gamma\text{:}$$

Semenov number = (heat release potential)/(heat removal potential)

This system has been widely analyzed in the last century to characterize thermal runaway at varying operating conditions.[8] At given rate constant and ambient temperature, the system is completely determined by the parameters B and ψ, and critical conditions are usually illustrated in the $B - \psi$ parameter plane.

A reactor under explosive conditions is sensitive to small variations in, e.g., the initial temperature, while, under nonexplosive conditions, the system remains insensitive to such variations. Thus, boundaries between runaway (explosive) and nonrunaway (nonexplosive) conditions can be identified on the basis of its sensitivity to small changes in the operating parameters. The system can also be characterized by the derivative of the maximum temperature reached in the reactor versus the initial temperature, i.e.,[73] $S(\theta^*,\theta_0) = d\theta^*/d\theta_0$.

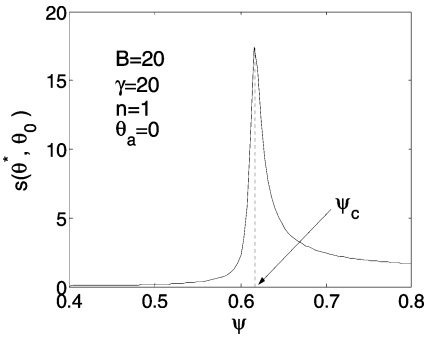

Figure 7. $S(\theta^*,\theta_i) = d\theta^*/d\theta_i$ versus Semenov number ψ for model (18, 19).

The runaway boundary is defined as the critical value of each parameter for which the sensitivity to the initial condition is maximum; e.g., for the Semenov number ψ, we have the results in Figure 7.

For ψ values smaller than ψ_c, the system is in nonrunaway conditions; i.e., the maximum temperature reached in the reactor is not very high, and this maximum is insensitive to small variations in the inlet temperature. With an increase in ψ, both the maximum temperature and its sensitivity to T_0 smoothly increase until, in proximity to ψ_c, there is a sharp rise for both of them that rapidly brings the reactor to a strong temperature increase. For ψ values higher than ψ_c, the sensitivity goes back to smaller values, leaving unchanged the extreme temperature rise reached at ψ_c. From this, fixing reaction kinetic (n,γ) and ambient temperature (θ_a), the curve in the $B - \psi$ plane can be obtained (Figure 8).

Let us consider the case of a system with nominal parameter design $B = 20$, $\gamma = 20$, $n = 1$, $\theta_a = 0$, and $\psi = 0.5$. Under these conditions, the system should be within the nonrunaway region. The system, however, is characterized by uncertainties. So, let us assume the following uncertainty distributions for model parameters:

$$B \sim N(20,4)$$

$$\gamma \sim N(20,2)$$

$$\theta_a \sim N(0,0.2)$$

$$\theta_0 \sim N(0,0.2)$$

$$\text{and } \psi \sim U(0.4,0.6) \tag{20}$$

where U indicates a uniform distribution.

Under the chosen operating conditions ($\gamma = 20$), a 0.02 standard deviation for the ambient and initial dimensionless temperatures corresponds to an about 3 K standard deviation in the absolute temperature scale.

We perform a Monte Carlo simulation, whose total cost is 6144 model evaluations, and analyze the behavior of the temperature maximum. In Figure 9 we can see that, even if the nominal conditions of the reactor are stable, there are threshold values for B, θ_a, and ψ for which the maximum temperature in the

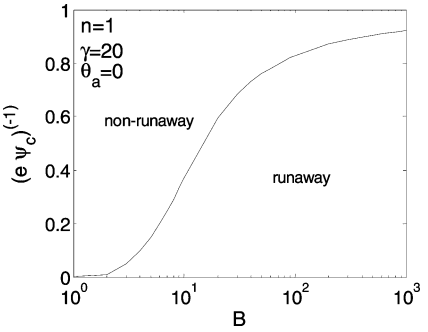

Figure 8. Runaway versus nonrunaway in the plane B,1/$e\psi$ for fixed (n,γ) and (θ_a) (e is the number of Neper).

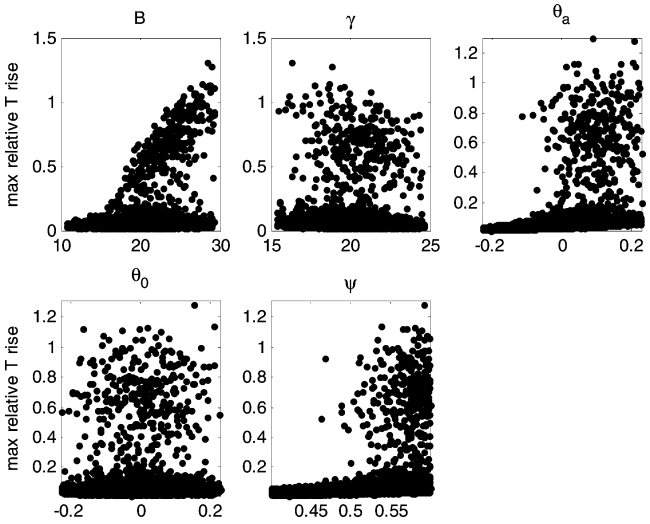

Figure 9. Relative temperature change at the maximum $[(T^{max} - T_0)/T_0]$ versus the uncertain model parameters. A rise in the ordinate to about 1.2 corresponds to a temperature shift of about 300 K for the operating conditions assumed.

reactor can have a sharp rise (absolute temperature can double, with a rise of e.g. 300 K). This striking result vindicates the use of global (e.g. Monte Carlo) exploration methods for uncertainty analysis even in the presence of moderate factors uncertainty. This result is due to the nonlinear and nonadditive nature of the problem, as shown by the different sensitivity measures for Y.

Let us look at the β_i^2 and S_i in Table 1 first. The sensitivity based on the β_i's can only capture 40% of the variation of the maximum temperature. Considering the variance-based main effects, we can arrive at 43%. This implies a 57% interaction between the model parameters. We could stop the analysis at this point, or we might pursue our investigation to achieve a full mapping of the input/output relationship. Stopping here would mean that we are happy with having learned that the parameter which offers a better chance of reducing the variance in the maximum temperature is ψ. Yet, this factor only accounts for ~18% of the variance, and the large unknown interactions might suggest that a much larger reduction in variance can be achieved if one could identify the interacting factors and try to fix them.

One avenue to do that would be to compute individual interaction terms.[74,75] [In this example all

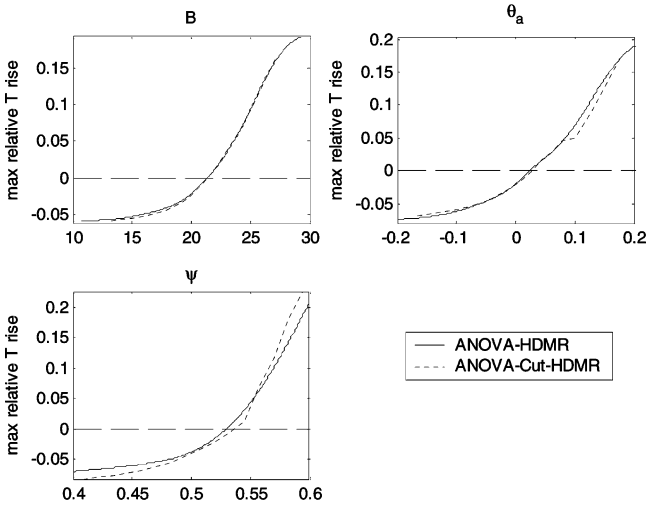

Figure 10. First-order ANOVA-HDMR (solid lines) and its approximation through the cut-HDMR expansion of only the subset of factors (B,θ_a,ψ, dotted lines) of maximum relative temperature rise $(T^{max} - T_0)/T_0$ (mean value of $(T^{max} - T_0)/T_0$ is 0.1). Plain cut-HDMR terms are also shown (dash−dot lines). All the lines have zero mean; i.e., HDMR functions plot the change with respect to the overall mean (which is about 0.1).

the second- and third-order interaction terms could be computed at no extra cost; see also section 3.[74] The full variance decomposition in this example could be obtained with a total cost of 6144 model runs. If only first-order indices were of interest, only a single shot of 500 runs would be sufficient, applying the method of Ratto et al.[75] To compute only first and total effects (see section 2.7 next), leaving out second-order effects from the analysis, 3000 runs would have been sufficient.] We get that the most significant second-order interaction terms $S_{ij} = V_{ij}/V(Y)$ are $S_{B\psi} = 0.17$, $S_{\theta_a\psi} = 0.17$, and $S_{B\theta_a} = 0.166$. Given that B, θ_a, and ψ seem to be the factors that interact the most, we may further compute the overall effect of these three factors. This comes out to be 0.961 [The sum of all effects of the factors B, θ_a, and ψ is made of their first-order terms, plus the three second-order terms, plus the single third-order term. This sum can also be written as $V_{B\theta_a\psi}(E_{\gamma\theta_1}(Y|B,\theta_a,\psi))/V(Y)$ and computed as such, i.e., without computing all the terms.], i.e., almost the total variance. If we measure the third-order interaction term, we obtain $S_{B\theta_a\psi} = 0.032$.

This example points to the importance of identifying interactions in sensitivity analysis.

We now apply to the example both ANOVA- and cut-HDMR. In Figure 10 we show the first-order ANOVA-HDMR terms (the f_i terms in eq 11) of the maximum relative temperature rise $(T^{max} - T_0)/T_0$ for the three most important parameters.

These show that the first-order relationships are monotonic, which explains why β_i^2 gives an acceptable estimate of the first order sensitivities in Table 1, with an overall error of about 0.03.

In Figure 10 we also show the approximated ANOVA-HDMR terms f_i, obtained by passing through the third-order cut-HDMR expansion of the most important factors (B,θ_a,ψ). The approximation is fairly good and illustrates the standard usage of cut-HDMR, i.e., as an efficient way to estimate the ANOVA-HDMR, such as f_i.

Table 1. Sensitivity Measures for Model (16−18)

	β_i^2	S_i	S_i^T
ψ	0.1761	0.1781	0.6738
θ_a	0.117	0.1641	0.556
B	0.1043	0.08	0.4692
γ	0.0028	0.0019	0.0322
θ_i	0.0016	0.0015	0.0128
sum	0.40	0.43	

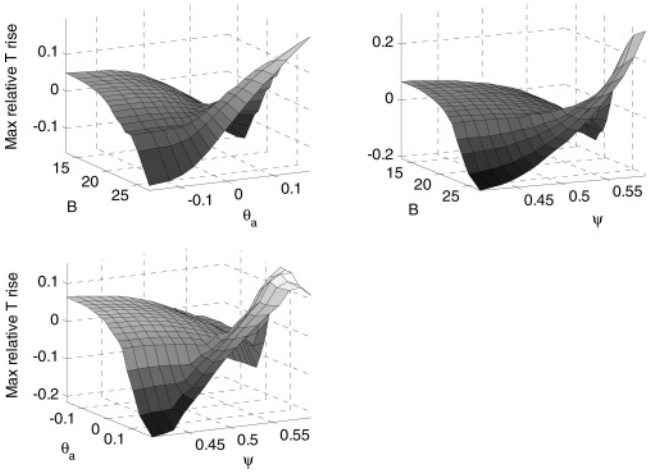

Figure 11. Second-order ANOVA-HDMR terms of the maximum relative temperature rise, approximated through the cut-HDMR of the subset of factors (B,θ_a,ψ). The grid of points for the cut-HDMR expansion up to the third order was of size 16 for each factor, for a total cost of 4912 runs.

In Figure 11 we show the approximated second-order ANOVA-HDMR terms, again obtained through the cut-HDMR expansion of the subset (B,θ_a,ψ). The grid of points for the cut-HDMR expansion up to the third order was of size 16 for each factor, for a total cost of 4912 runs.

2.6. Monte Carlo Filtering—An Overview

Sensitivity analysis plays an important role in the verification of the formal correctness of models. It is rare to perform a SA without identifying model formal or coding errors, whose correction is thus made possible. For a discussion of the motivation for sensitivity analysis in the scientific method, see the last chapter in ref 45. We focus in this work on two possible objectives for sensitivity analysis, i.e., factors mapping and factor importance analysis. The cut-HDMR is effective for both, as it allows for an efficient though approximate estimate of the f_i terms (mapping), on which a sensitivity measure such as S_i can be computed (importance). At times, one is interested in a particular form of mapping, i.e., when the objective of the analysis is to measure what fraction of the model realizations falls within established bounds or regions. This objective can be pursued using a Monte Carlo method known as Monte Carlo filtering (MCF[76−78]). In MCF, one samples the space of the input factors as in the plain MC method and then categorizes the corresponding model output as either within or without the target region (the terms *behavior* (B) or *nonbehavior* (NB), are used). This categorization is then mapped back into the input factors, each of which is thus also partitioned into a behavioral and a nonbehavioral subsample. When the two B, NB samples for a factor are statistically different, then the factor is an influential one. This approach to SA is also known as regionalized sensitivity analysis (RSA). MCF is often used in calibration, as it can successfully point to the existence of alternative behavioral regions in the multidimensional space of the input. In this case, a combination of MCF and variance-based sensitivity analysis of the likelihood (the probability of the data given the model) can be helpful. This and the pros and cons of RSA and its extensions are reviewed in ref 45.

2.7. Total Sensitivity Indices

Is there a more compact way to analyze the model in eqs 18−20 without computing all $2^5 - 1$ terms of our 5-factor model? Surely for larger dimensionalities of the input factors space, a more compact measure would be useful. This is offered by another variance-based measure,[43] which was implicit in our discussion of the Sobol' "group" sensitivity (eq 17). Imagine that the set $\mathbf{U}$ in eq 17 contains only one factor, X_i, and that as a result $\mathbf{Z} = \mathbf{X}_{-i}$. Hence, eq 17 becomes

$$V(Y) = V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) + V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i})) + V_{X_i\mathbf{X}_{-i}} \quad (21)$$

The condition for X_i to be truly noninfluential is that $V_{X_i}(E_{\mathbf{X}_{-i}}(Y|X_i)) + V_{X_i\mathbf{X}_{-i}} = 0$, which is the same as $V(Y) - V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i})) = 0$. Because of eq 8, this is the same as to say that $E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i})) = 0$. In summary, if X_i is noninfluential, then $S_{T_i} = E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i}))/V(Y) = 0$. We call S_{T_i} the total sensitivity index of factor X_i. It is easy to prove[45] that the condition S_{T_i} is necessary and sufficient for X_i to be noninfluential. The descriptive power of this measure is evident by looking at the last column of Table 1. Even if we had not computed all second- and third-order interaction effects, it would now be evident from the difference between the S_i and S_{Ti} values for each factor that B, θ_a, and ψ are involved in significant interactions. From the total indices we can also see that all the interaction terms of factor γ with (B,θ_a,ψ) cover most of the 3.9% of total variance unexplained by the group (B,θ_a,ψ). Other advantages of the S_{Ti} measure are as follows:

(i) With S_{Ti}, we no longer have to limit our analysis to additive models.

(ii) It dispels the curse of dimensionality. One does not need to calculate all the 2^k terms in eq 11 but just the $2k$ measures S_i and S_{Ti}, to obtain a good characterization of the system.

(iii) S_{Ti} can be easily computed using extended FAST[53] or extended Sobol' methods.[74]

(iv) When eq 11 holds, e.g. when the input factors are independent, S_{Ti} can be easily seen to be equal to the sum of all terms (first-order plus interactions) that include factor X_i. For a simple three-factor model, this would imply that $S_{T1} = S_1 + S_{12} + S_{13} + S_{123}$. Even when the factors are not independent, S_{Ti} is an effective measure to use, e.g., if one wants to reduce the variance of the output acting on a subgroup of factors.[54] It is intuitive that when interac-

tions are present, a reduction in the variance of the output can be achieved by determining simultaneously the true values of two or more interacting factors.

Note that when one wants to criticize the use of S_i as a sensitivity measure, one usually builds a test case where a factor has a zero first-order term and important nonzero higher order terms.[44,79] These criticisms are, in our opinion, unfounded,[45] as S_i is the right measure to use for factors prioritization, as we discussed above. If one factor has a zero first-order term, no variance reduction can be expected by determining the true value of just that factor. On the other hand, if the analyst intends to identify noninfluential factors in order to remove them from the variance propagation analysis, then a broader concept of importance must be invoked, which corresponds to the S_{Ti} measure. $S_{Ti} = E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i}))/V(Y)$ is, in fact, the expected variance that would be left if all factors but X_i were determined and provides the educated answer to the question, "Which factor can be fixed anywhere in its range of variability without affecting the output?", the answer being all those factors whose S_{Ti} is zero. We call this the "factors fixing" setting.[45]

Computing S_i and S_{Ti} for each factor X_i, while still being far from a full factors mapping, gives a fairly instructive description of the system.

2.8. The Method of Morris

Another useful sensitivity measure, which is computationally less expensive than the variance-based methods, is the measure of Morris,[51] which is particularly suited when the number of uncertain factors is high and/or the model is expensive to compute. It belongs, thus, to the family of screening sensitivity analysis methods.[80]

To illustrate it, we go back to the dimethyl sulfide (DMS) example already mentioned when discussing regression-based sensitivity analysis. The model KIM[12,63] describes temperature-dependent tropospheric air and droplet chemistry for DMS. DMS chemistry is extensively investigated for its climatic implications.[81] In a work published in 1999, the KIM model included about 50 chemical reactions and 68 uncertain input factors, mostly kinetic and Henry law constants, which could be screened down to the 10 most important ones using the method of Morris.[82] The analysis was then completed by applying extended FAST to the 10 most important factors for a quantitative analysis using the S_i and S_{Ti} indices. Here, an updated version of KIM[83] is considered where the number of uncertain input factors is cut down to 56, and the sensitivity of the model is investigated via the Morris method. In an explanatory fashion, results are also compared with what would be obtained by a derivative-based analysis.

The method of Morris[51] varies one factor at a time across a certain number of levels selected in the space of the input factors. For each variation, a factor's elementary effect is computed, which is an incremental ratio for that factor:

$$u_i = \frac{Y(x_1,x_2,...,x_i+\Delta x_i,...,x_k) - Y(x_1,x_2,...,x_i,...,x_k)}{\Delta x_i} \quad (22)$$

A set of stepwise curves scan the factors levels, as to generate for each factor r different estimates of elementary effects u_i. The mean μ_i and the standard deviation σ_i of the elementary effects u_i over the r repetitions are used to assess the factors' importance. A high value of μ_i flags a high linear effect for a given factor, while a high value of σ_i flags either nonlinear or nonadditive factor behavior. It is useful also to compute the modulus version μ_i^* of the Morris method, i.e., the average of the $|u_i|$, and the importance of input factors is often assessed by plotting factors on the (μ_i^*, σ_i) axes (Figure 12). The factors closest to the origin are less influential.

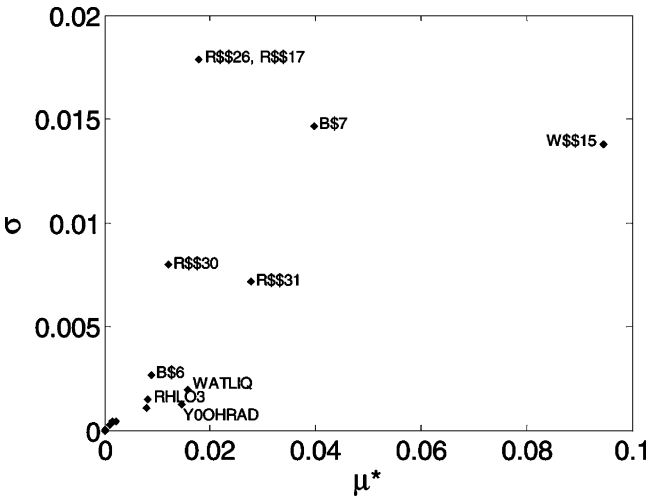

Figure 12. Screening of input factors based on the method of Morris. Factors away from the origin are the most important, while the factors clustered in (0,0) can be fixed. RHLO3 is the Henry's law constant for ozone (O_3). WATLIQ is the water liquid content. Y0OHRAD is the initial concentration of the OH radical. The R and B parameters represent rate constants of gas-phase reactions, while the W parameters indicate rate constants of liquid-phase reactions. A more detailed factors description is in ref 82.

μ_i^* has similarities with the S_{Ti} index, in the sense that it tends to produce a ranking of the factors very similar or identical to that based on the S_{Ti} indices.[45]

Looking at Figure 13 (comparative ranking for the input factors using $\partial Y/\partial X_i$, S_i^σ, and μ_i^*) for the KIM model suggests that ranking of factors is strongly altered when using derivative-based measures. If the rankings of the different measures were equal, we would have points in the left panel of Figure 13 aligned on a monotonic curve, while in the right panel points would stay on a straight line of unit slope. Conversely, both high and low importance factors are completely shifted if derivative-based measures are used. Only the least important factors seem to have a similar ranking, even if significant changes are detected also in this case (see points in the gray oval in the right panel of Figure 13).

The lack of reliability of the derivative-based measures is due to the nonlinearity of the present

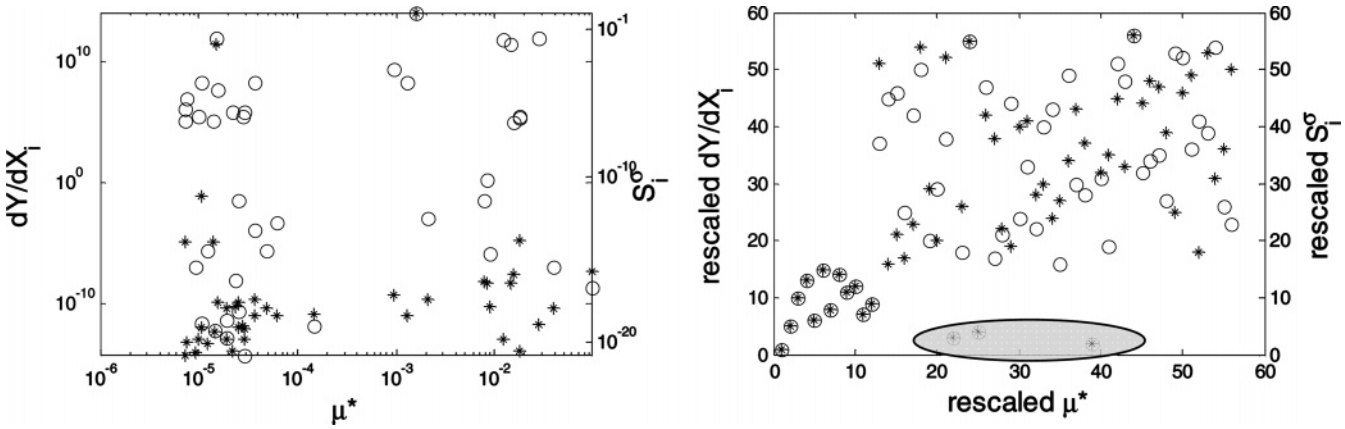

Figure 13. $\partial Y/\partial X_i$ (circles) and S_i^{σ} (stars) versus μ_i^*. Left panel: logarithmic plots. Right panel: rank-rescaled measures (high importance = high rank; i.e., the most important factor has rescaled measure = 56).

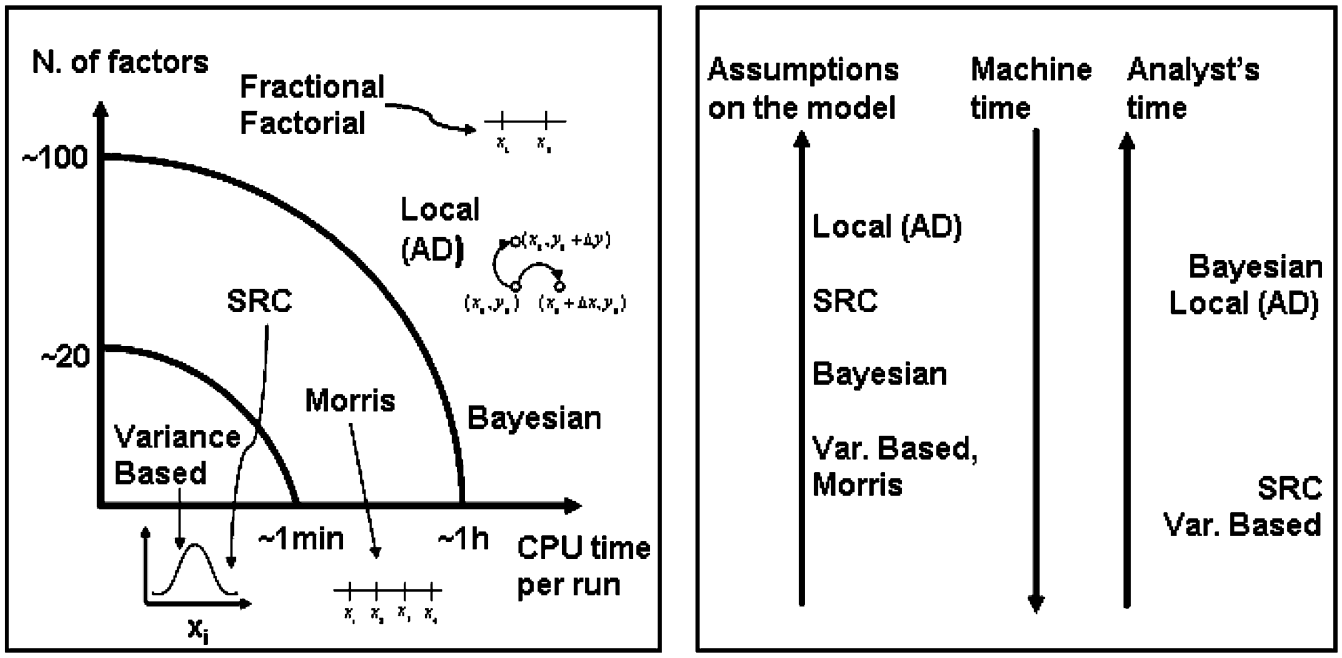

Figure 14. Sketch of the various techniques available and their use as a function of computational cost of the model, complexity of the model, dimensionality of the input space, and analyst time. AD means "automated differentiation".

version of the KIM model, as confirmed by performing a regression-based analysis through Monte Carlo simulation which generates a R^2 value of 0.57.

3. When to Use What

We discuss now the choice of the proper sensitivity analysis technique depending on considerations such as (i) the computational cost of running the model, (ii) the number of input factors, (iii) the degree of complexity of the model coding, (iv) the amount of analyst's time involved in the sensitivity analysis, and (v) the setting for the analysis. Figure 14 supports graphically the explanations using these characteristics as discriminating criteria.

For models that require a modest amount of CPU time (i.e. up to the order of 1 min per run), and with a number of input factors which does not exceed, say, 20, the class of the variance-based techniques yields the more accurate pattern of sensitivity. Both the method of Sobol' (very easy to code[42,45,74]) and the extended FAST (less easy[53]) provide all the pairs of first-order and total indices at a cost of $(k + 2)N$[74] model runs for Sobol' and $\approx kN$ model runs for the extended FAST, where k is the number of factors and N is the number of rows of the matrixes **A** and **B** in eq 16. Typically, $N \approx 500/1000$. To give an order of magnitude of the computational requirement, for a model with 10 factors and 0.5 min of CPU time per run, a good characterization of the system via S_i and S_{Ti} can be obtained at the cost of ~42/84 h of CPU time. [Note that FAST is impractical if the input factors are sampled from discrete distributions.[53]]

With the method of Sobol', in addition to the first-order and total indices computed with $(k + 2)N$ model runs, all the interaction terms of order $(k - 2)$ can be obtained at no extra cost.[74] At the additional cost of kN model runs, *double* estimates of all the first-order, second-order, $(k - 2)$-th order, and total indices can be obtained[74]. [This sampling design scheme has been applied in the test case of section 2.5, with $N = 512$, $k = 5$, and an overall cost $(2k + 2)N = 6144$.] Finally, any other interaction term between the third

and the $(k - 3)$-th can be estimated at the further additional cost of N model runs each.

When using the method of Sobol', we make use of quasi-random numbers to generate the sample matrixes **A** and **B** (eq 16) for the analysis. These are sequences of multidimensional points characterized by "optimal" space-filling properties.[84,85]

A very popular form of sampling that also aims to scan efficiently the input factors space is the latin hypercube sampling (LHS), considered by some as the most effective strategy when the model is expensive to evaluate.[86] The space-filling properties of LHS can be enhanced by optimization algorithms.[49]

When the input factors are correlated, an ad hoc computational scheme must be adopted. An efficient and unbiased estimation procedure is available for first-order indices and is based on *replicated* LHS.[87,88] This is also easy to code,[45] and the cost to estimate all the first-order indices is Nr model runs, where r is the number of replicates needed (usually around 10) and the cost is independent of the number of factors.

For higher order indices as well as total indices in the case of correlated input, one has to apply a brute force approach whereby the operators V and E (eq 9) are to be written in explicit form (i.e. as the variance of a mean, involving a double computing loop). The computational cost is thus Nr model runs per index.

All these techniques are implemented in SIMLAB, a free software package.[89]

Less expensive alternatives to the variance-based methods are the standardized regression coefficients, SRCs. With a single batch of N sampled points (say, $N \approx 500/1000$ LHS points or less depending on the cost of the model), the SRCs and their rank-transformed version can be estimated for all the input factors.

As mentioned, the SRCs are only effective for linear or quasi-linear models, i.e., for $R^2 \geq 0.7/0.9$. Regression methods, also implemented in SIMLAB, are always useful to be looked at in order to investigate the degree of linearity of the model.

When the CPU time increases (say, up to 10 min per run) or the number of factors increases (say, up to 100), the method of Morris,[51] of which an extension is implemented in SIMLAB, offers the best result.

The number of sampled points required is $N_{\text{Morris}} = r(k + 1)$, where r is generally set to $r \approx 4/8$ and k is the number of input factors. To make an example, with 80 factors and 5 min of CPU time per run, all the model outputs can be ready in 27 h if $r = 4$ is taken. The main drawback of this method is that samples are taken from levels, while both the SRC and the variance-based methods take samples from distributions.

When the number of input factors and/or the CPU time is even as large as to preclude the use of the method of Morris, then supersaturated fractional factorial designs, where factors are iteratively perturbed in batches, can be used.[41,80] However, these methods preclude an effective exploration of the space of the inputs, as they mostly operate at very few factor levels and require strong assumptions on the model behavior. To make an example, Bettonvill, reviewed in ref 80, assumes a known monotonic relationship between the output and each of the inputs.

Automatic differentiation techniques[3] can also be used when CPU time is very large. They are inherently local. In addition, they require intervention of the analyst in the computer code that implements the model. However, for expensive models, these methods may offer an approximate solution for factors importance assessment and are very informative for factors mapping, as well as for data assimilation applications.[17] If higher order derivatives are computed,[16] these give information about multifactor curvature effects and could be seen as a bridge between local and global methods; e.g., a second-order term of the type $\partial^2 Y/\partial X_i \partial X_j$ gives information about a possible interaction effect between X_i and X_j, although the S_{ij} variance-based measure will include an element of averaging over the entire space of the factors. The advantages of higher order (second, third) local sensitivity analysis in the presence of nonlinear outputs (e.g. an ozone peak concentration) are discussed in ref 19. In a Taylor-expansion frame, higher order terms allow a better exploration further away from the baseline. According to ref 14, while first-order sensitivities can predict ozone concentration at about a 25% factors variation away from their baseline, with second-order terms the prediction is good up to 50% variations away from the baseline values.

As an alternative, a Monte Carlo based approach to estimation has also been tried in chemistry, which includes a quantitative sensitivity analysis step.[45,90,91]

Sensitivity analysis is also driven by the setting.[45] When the purpose of the analysis is to prioritize factors, the first-order sensitivity indices S_i have a strong motivation for use. If the objective is to fix noninfluential factors, then the total sensitivity indices S_{Ti}, or equivalently the measure of Morris, come into use. If a particular region in the space of the output (e.g. above or below a given threshold) is of interest, then Monte Carlo filtering and associated methods can be tried as an alternative or complement to the measures just mentioned. If the purpose of the analysis is a diagnostic mapping of the input/output relationship, then various types of HDMR can be tested (see Figure 10). At all of these settings, the computation of derivatives, especially if achieved with a modicum of extra computing, is advisable for a general understanding of the model.

Additional software sources are given in ref 3. Among these, we would like to recommend as a suggested practice the KINALC package.[92] KINALC is a postprocessor to CHEMKIN, a widespread simulation program, and carries out processing sensitivity analysis including principal component analysis of the sensitivity matrix. As argued in ref 7, a principal component analysis (PCA) is a useful postprocessing tool to local sensitivity analysis. PCA can extract at no additional cost relevant features of the chemical mechanism and can assist in the setup of the parameter estimation step.

4. Other Methods and Ongoing Research

Several methods have been proposed for sensitivity analysis in various settings, and a wealth of special issues where these are applied is available in the literature.[93−96] Some of these methods are domain specific, such as for instance methods for use in reliability engineering,[97] and we have not treated them in the present review.

With the exception of the automated differentiation approach,[3] which requires extensive manipulation of the computer program where the model is implemented, we have also left out in the present review methods that require a direct intervention of the analyst on the model. In fact, all methods described so far treat the model as a black box.

In Bayesian sensitivity analysis,[98,99] the analyst needs to implement, by Bayesian updating, the algorithm to estimate the model itself at "untried" points.[49] Once the appropriate sample is generated, model values and good estimates of S_i can be very cheaply generated. Bayesian methods can be recommended when CPU time is very large. They are superior to factorial designs, in that they operate on distributions, not on levels, for the input factors. Note that the Bayesian approach can be unsuccessful for particularly stiff models.

Approaches that demand extensive analysis of the model also require that the model remains stable in time, as each model revision, especially in the Bayesian approach, will call for a new analysis of the model prior to sensitivity calculation.

Present research in sensitivity analysis focuses on how to accelerate the computation of the sensitivity indices (S_i and higher order). The Bayesian method already cited[98,99] is a possible avenue, as well as the cut-HDMR-based approach illustrated in Figure 10. Another strategy, easier to code, is based on random balance designs and uses Fourier analysis to estimate all the first-order indices at a total cost of N model runs (i.e. the same cost of SRCs).[66] State-dependent parameter (SDP) modeling, a nonparametric model estimation approach based on recursive filtering and smoothing estimation, is also being applied successfully[75] to produce both the ANOVA-HDMR f_i terms and the relative S_i at the same cost of SRCs.

5. References

(1) Rosen, R. *A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*; Columbia University Press: New York, 1991.
(2) Saltelli, A. In *Sensitivity Analysis*; Chan, K., Scott, M., Eds.; John Wiley & Sons: Chichester, 2000.
(3) Grievank, A. In *Evaluating derivatives, Principles and techniques of algorithmic differentiation*; SIAM, 2000.
(4) Turányi, T. *J. Math. Chem.* **1990**, *5*, 203.
(5) Rabitz, H.; Kramer, M.; Dacol, D. *Annu. Rev. Phys. Chem.* **1983**, *34*, 419.
(6) Rabitz, H. *Science* **1989**, *246*, 221.
(7) Tomlin, A. S.; Turányi, T. In *Low-temperature Combustion and Autoignition*; Pilling M. J., Hancock, G., Eds.; Elsevier: 1997; p 293.
(8) Varma, A.; Morbidelli, M.; Wu, H. *Parametric Sensitivity in Chemical Systems*; Cambridge: 1999.
(9) Turányi, T.; Rabitz, H. In *Sensitivity Analysis*; Chan, K., Scott, M., Eds.; John Wiley & Sons: Chichester, 2000; p 81.
(10) Cacuci, D. G. In *Sensitivity & Uncertainty Analysis, Volume 1: Theory*; Chapman & Hall: 2003.
(11) Ionesco-Bujor, M.; Cacuci, D. G. *Nucl. Sci. Eng.* **2004**, *147*, 189.
(12) Saltelli, A.; Hjorth, J. *J. Atmos. Chem.* **1995**, *21*, 187.
(13) Pandis, S. N.; Seinfeld, J. H. *J. Geophys. Res.* **1989**, *94*, 1105.
(14) Vuilieumier, L.; Harley, R.; Brown, N. J. *Environ. Sci. Technol.* **1997**, *31*, 1206.
(15) Kioutsioukis, I.; Melas, D.; Zerefos, C.; Ziomas, I. *Comput. Phys. Commun.* **2005**, *167*, 23.
(16) Le Dimet, F.-X.; Navon, I. M.; Daescu, D. N. *Mon. Weather Rev.* **2002**, *130*, 629.
(17) Mallet, V.; Sportisse, B. *Atmos. Chem. Phys. Discuss.* **2004**, *4*, 1371.
(18) Saltelli, A. *J. Geophys. Res.* **1999**, *104*, 3789; **1999**, *104*, 24013.
(19) Hakami, A.; Odman M. T.; Russel, A. G. *Environ. Sci. Technol.* **2003**, *37*, 2442.
(20) Cukier, R. I.; Fortuin, C. M.; Shuler, K. E.; Petschek, A. G.; Schaibly, J. H. *J. Chem. Phys.* **1973**, *59*, 3873.
(21) Schaibly, J. H.; Shuler, K. E. *J. Chem. Phys.* **1973**, *59*, 3879.
(22) Cukier, R. I.; Schaibly, J. H.; Shuler, K. E. *J. Chem. Phys.* **1975**, *63*, 1140.
(23) Cukier, R. I.; Levine, H. B.; Shuler, K. E. *J. Comput. Phys.* **1978**, *26*, 1.
(24) Pierce, T. H.; Cukier, R. I. *J. Comput. Phys.* **1981**, *41*, 427.
(25) Koda, M.; McRae, G. J.; Seinfeld, J. H. *Int. J. Chem. Kinet.* **1979**, *11*, 427.
(26) McRae, G. J.; Tilden, J. W.; Seinfeld, J. H. *Comput. Chem. Eng.* **1982**, *6*, 15.
(27) Koda, M.; Dogru, A. H.; Seinfeld, J. H. *J. Comput. Phys.* **1979**, *30*, 259.
(28) Koda, M. *Atmos. Environ.* **1982**, *16*, 2595.
(29) Koda, M.; Seinfeld, J. H. *IEEE Trans. Autom. Control* **1982**, *27*, 951.
(30) Liepmann, D.; Stephanopoulos, G. *Ecol. Modell.* **1985**, *30*, 13.
(31) Helton, J. C. *Reliab. Eng. Syst. Saf.* **1993**, *42*, 327.
(32) Saltelli, A.; Tarantola, S.; Campolongo, F. *Stat. Sci.* **2000**, *15*, 377.
(33) Santner, T. J.; Williams, B. J.; Notz, W. I. *Design and Analysis of Computer Experiments*; Springer-Verlag: 2003.
(34) Helton, J. C.; Davis, F. J. *Reliab. Eng. Syst. Saf.* **2003**, *81*, 23.
(35) Cacuci D. G.; Ionesco-Bujor, M. *Nucl. Sci. Eng.* **2004**, *147*, 204.
(36) Goldsmith, C. H. In *Encyclopaedia of Biostatistics*; Armitage, P., Colton, T., Eds.; Wiley: New York, 1998.
(37) Hora, S. C.; Iman, R. L. *Sandia Lab. Rep.* **1986**, SAND85-2839.
(38) Ishigami, T.; Homma, T. In *Proceedings of the ISUMA '90*, First International Symposium on Uncertainty Modelling and Analysis; University of Maryland: December 3−5; 1990; p 398.
(39) Iman, R. L.; Hora, S. C. *Risk Anal.* **1990**, *10*, 401.
(40) Krzykacz-Hausmann, B. Technical Report GRS-A-1700; Gesellschaft fuer Reaktor Sicherheit (GRS) MbH: Garching, 1990.
(41) Saltelli, A.; Andres, T. H.; Homma, T. *Comput. Stat. Data Anal.* **1993**, *15*, 211.
(42) Sobol', I. M. *Math. Modell. Comput. Exp. (Engl. Transl.)* **1993**, *1*, 407.
(43) Homma, T.; Saltelli, A. *Reliab. Eng. Syst. Saf.* **1996**, *52*, 1.
(44) McKay, M. D. *LA-UR-96-2695* **1996**, 1.
(45) Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M. *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models*; John Wiley & Sons Publishers: New York, 2004.
(46) Rabitz, H.; Aliş, Ö. F.; Shorter, J.; Shim, K. *Comput. Phys. Commun.* **1999**, *117*, 11.
(47) Rabitz, H.; Aliş, Ö. F. In *Sensitivity Analysis*; Chan, K., Scott, M., Eds.; John Wiley & Sons: Chichester, 2000; p 199.
(48) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for experimenters*; John Wiley and Sons: New York, 1978.
(49) Sacks, J.; Welch, W. J.; Mitchell, T. J.; Wynn, H. P. *Stat. Sci.* **1989**, *4*, 409.
(50) Welch, W. J.; Buck, R. J.; Sacks, J.; Wynn, H. P.; Mitchell, T. J.; Morris, M. D. *Technometrics* **1992**, *34*, 15.
(51) Morris, M. D. *Technometrics* **1991**, *33*, 161.
(52) Saltelli, A.; Bolado, R. *Comput. Stat. Data Anal.* **1998**, *26*, 445.
(53) Saltelli, A.; Tarantola S.; Chan, K. *Technometrics* **1999**, *41*, 39.
(54) Saltelli A.; Tarantola, S. *J. Am. Stat. Assoc.* **2002**, *97*, 702.
(55) Hamby, D. M. *Environ. Monit. Assess.* **1994**, *32*, 135.
(56) Kleijnen, J. P. C. In *Handbook of Simulation*; Banks, J., Ed.; Wiley: New York, 1998.
(57) Frey, H. C., Ed. *Risk Anal.* **2002**, 22.
(58) Chance, E.; Curtis, A.; Jones, I.; Kirby, C. Report AERE-B8775; Harwell: 1977.
(59) http://www.reactiondesign.com/; see also CHEMKIN page at SANDIA Labs. http://www.ca.sandia.gov/chemkin/.
(60) Turányi, T.; Zalotai, L.; Dóbé, S.; Bercé, T. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2568.
(61) Kleijnen, J. P.; Helton, J. C. *Reliab. Eng. Syst. Saf.* **1999**, *65*, 147.
(62) Draper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, 1981.
(63) Campolongo, F.; Saltelli, A.; Jensen, N. R.; Wilson, J.; Hjorth, J. *J. Atmos. Chem.* **1999**, *32*, 327.
(64) Iman, R. L.; Conover, W. J. *Technometrics* **1979**, *21*, 499.
(65) Saltelli, A.; Sobol', I. M. *Reliab. Eng. Syst. Saf.* **1995**, *50*, 225.
(66) Tarantola, S.; Gatelli, D.; Mara, T. *Reliab. Eng. Syst. Saf.* **2005**, in press.

(67) Chan, K.; Tarantola, S.; Saltelli, A.; Sobol', I. M. In *Sensitivity Analysis*; Chan, K., Scott, M., Eds.; John Wiley & Sons: Chichester, 2000; p 167.
(68) Archer, G.; Saltelli, A.; Sobol', I. M. *J. Stat. Comput. Simul.* **1997**, *58*, 99.
(69) Shim, K.; Rabitz, H. *Phys. Rev. B* **1998**, *58*, 1940.
(70) Rabitz, H.; Alis, Ö. *J. Math. Chem.* **1999**, *25*, 197.
(71) Shorter, J. A.; Ip, P. C.; Rabitz, H. *J. Phys. Chem. A* **1999**, *103*, 7192.
(72) Shorter, J. A.; Ip, P. C.; Rabitz, H. *Geophys. Res. Lett.* **2000**, *27*, 3485.
(73) Morbidelli, M.; Varma, A. *Chem. Eng. Sci.* **1988**, *43*, 91.
(74) Saltelli, A. *Comput. Phys. Commun.* **2002**, *145*, 280.
(75) Ratto, M.; Saltelli, A.; Tarantola, S.; Young, P. *J. R. Stat. Soc., B* **2004**, submitted.
(76) Hornberger, G. M.; Spear, R. C. *J. Environ. Manage.* **1981**, *12*, 7.
(77) Young, P. C.; Parkinson, S. D.; Lees, M. *J. Appl. Stat.* **1996**, *23*, 165.
(78) Young, P. *Comput. Phys. Commun.* **1999**, *117*, 113.
(79) Krzykacz-Hausmann, B. In *Proceedings of SAMO2001*; Prado, P., Bolado, R., Eds.; CIEMAT: Madrid, 2001; p 31.
(80) Campolongo, F.; Kleijnen, J.; Andres, T. In *Sensitivity Analysis*; Chan, K., Scott, M., Eds.; John Wiley & Sons: Chichester, 2000; p 65.
(81) Capaldo, K. P.; Pandis, S. N. *J. Geophys. Res.* **1997**, *102*, 23/251.
(82) Campolongo, F.; Tarantola, S.; Saltelli, A. *Comput. Phys. Commun.* **1999**, *117*, 75.
(83) Le Bras, G.; Barneş I.; Hjorth, J.; Zetzsch, C.; Martinez, E.; Mihalopoulos, N. *Report EUR 19569 EN*; European Commission: Luxembourg; 2000.
(84) Sobol', I. M. *USSR Comput. Math. Math. Phys.* **1976**, *16* (5), 1332.
(85) Bratley, P.; Bennet, L. F. *ACM Trans. Math. Software* **1988**, *14*, 88.
(86) Iman R. L.; Conover W. J. *Commun. Stat.: Theory Methods* **1980**, *A9*, 1749.
(87) McKay, M. D.; Beckman, R. J.; Conover, W. J. *Technometrics* **1979**, *21*, 239.
(88) McKay, M. D. *Tech. Rep. NUREG/CR−6311, LA-12915-MS*; U.S. Nuclear Regulatory Commission and Los Alamos National Laboratory: 1995.
(89) www.jrc.cec.eu.int/uasa/prj-sa-soft.asp. This software has been maintained since 1990 by the authors at the Joint Research Centre of the European Commission.
(90) Ratto, M.; Tarantola, S.; Saltelli, A. *Comput. Phys. Commun.* **2001**, *136*, 212.
(91) Scire, J. J.; Dryer, F. L.; Yetter, R. A. *Int. J. Chem. Kinet.* **2001**, *33*, 784.
(92) KINALC is freely available at http://www.chem.leeds.ac.uk/Combustion/kinalc.htm. See also: Turányi, T. *Comput. Chem.* **1990**, *14*, 253.
(93) Scott, M., Saltelli, A., Eds. *J. Stat. Comput. Simul.* **1997**, 57.
(94) Saltelli, A., Chan, K., Scott, M., Eds. *Comput. Phys. Commun.* **1999**, 117.
(95) Tarantola, S., Saltelli, A., Eds. *Reliab. Eng. Syst. Saf.* **2003**, 79.
(96) Helton, J. C.; McKay, M. D.; Cooke, R.; Saltelli, A. *Reliab. Eng. Syst. Saf.,* to appear in spring 2005.
(97) Cawlfield, J. D. In *Sensitivity Analysis*; Chan, K., Scott, M., Eds.; John Wiley & Sons: Chichester, 2000; p 155.
(98) O'Hagan, A.; Kennedy, M. C.; Oakley J. E. *Bayesian Stat.* **1999**, *6*, 503.
(99) Oakley, J.; O'Hagan, A. *J. R. Stat. Soc., Ser. B* **2004**, *66*, 751.

CR040659D