Review

Sensitivity analysis practices: Strategies for model-based inference

Andrea Saltelli*, Marco Ratto, Stefano Tarantola, Francesca Campolongo,
European Commission, Joint Research Centre of Ispra (I).

Institute for the Protection and Security of the Citizen (IPSC) The European Commission, Joint Research Centre, TP 361, 21020 Ispra (VA), Italy

Abstract

Fourteen years after *Science's* review of sensitivity analysis (SA) methods in 1989 (*System analysis at molecular scale, by H. Rabitz*) we search *Science Online* to identify and then review all recent articles having "sensitivity analysis" as a keyword. In spite of the considerable developments which have taken place in this discipline, of the good practices which have emerged, and of existing guidelines for SA issued on both sides of the Atlantic, we could not find in our review other than very primitive SA tools, based on "one-factor-at-a-time" (OAT) approaches. In the context of model corroboration or falsification, we demonstrate that this use of OAT methods is illicit and unjustified, unless the model under analysis is proved to be linear. We show that available good practices, such as variance based measures and others, are able to overcome OAT shortcomings and easy to implement. These methods also allow the concept of factors importance to be defined rigorously, thus making the factors importance ranking univocal. We analyse the requirements of SA in the context of modelling, and present best available practices on the basis of an elementary model. We also point the reader to available recipes for a rigorous SA.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Global sensitivity analysis; Morris method; Variance based methods; Monte Carlo filtering

Contents

*Corresponding author. Tel.: +39 332 78 9686; fax: +39 332 78 5733.
  E-mail address: andrea.saltelli@jrc.it (A. Saltelli).

1. Sensitivity analysis in the scientific method

Fourteen years after *Science*'s review on sensitivity analysis (SA) [1], considerable developments have taken place in this discipline. As an indicator of the practices currently used in SA, we have searched *Science Online* to identify and review all papers (33) published between 1997 and 2003 having SA as a keyword. In these, we could not detect application of available good practices (Appendix A). The works reviewed highlight the importance of SA in corroborating or falsifying a model-based analysis. Yet all sensitivity analyses were performed using a one-factor-at-a-time (OAT) approach, so called as each factor is perturbed in turn while keeping all other factors fixed at their nominal value. When the purpose of SA is to assess the relative importance of input factors in the presence of factors uncertainty this approach is only justified if the model is proven to be linear.

The present work aims to discuss the merit and correct use of model-free SA approaches. Although problem setting met in modelling are disparate, we demonstrate that available good practices in SA can be of considerable general use. The best SA practices are well established among SA practitioners, among applied statisticians, as well as in some disciplinary communities (see [2,3] for chemical engineering; [4] for bio-statistics; [5] for risk analysis), though not among the scientific community at large.

Different understanding of SA are used in different modelling communities, see, e.g. [2,3], [4] and [5] addressed, respectively, to chemical engineers, bio-statisticians, and practitioners of risk analysis. Quite often though SA is identified almost as a mathematical definition, with a differentiation of the output with respect to the input. This definition is in fact coherent with a vast set of applications of SA to, e.g. the solution of inverse problems [1] or the optimisation of the computation of mass transfer with chemical reaction problems. This approach to sensitivity has prevailed in the modelling community, also when the objective of the analysis was to ascertain the relative importance of input factors in the presence of finite ranges of uncertainties. For this reason quantitative measures of sensitivity, such as $S_{X_i} = V_{X_i}(E_{X_{-i}}(Y|X_i))/V_Y$ discussed in this paper, have been referred to in the literature *importance measures*, e.g. [6].

Importance measures such as S_{X_i} have been since long identified as a best practice, in the model-free sense. In the literature, the FAST (Fourier Amplitude Sensitivity Test, see http://users.frii.com/uliasz/modeling/ref/sens_bib.htm for a bibliography), first introduced in the 1970s, the

Sobol' method [7], the measures of importance of Iman and Hora [6], and that of Sacks et al. [8] all coincide with S_{X_i} (see bibliography at http://sensitivity-analysis.jrc.cec.eu.int/). The bibliographies quoted point the reader to a wealth of applications of variance-based measures for SA of complex models in disparate fields.

1.1. A gallery of applications of the best SA practices

All models have use in SA. Applications worked by the authors include atmospheric chemistry [9,10], transport emissions [11], fish population dynamics [12] composite indicators [13], portfolios [14], oil basins models [15], radioactive waste management [16], geographic information systems [17], solid-state physics [18]. Applications from several practitioners can be found in [19,20], and in several special issues in the specialised literature, e.g. [21,22].

It is worth making a few remarks on some of the above-mentioned applications [13,15]. Most existing *composite indicators* are simple weighted averages of selected sub-indicators [23,86], e.g. as our model (1). The analysis of the robustness of a composite indicator can involve an SA with respect to

(I)　changes in the selection of the underlying indicators,
(II)　error in the underling indicators,
(III)　changes in the scaling method,
(IV)　changes in the aggregation weights [24].

In the main text, (II) and (III) have been investigated via model (1). An analysis of (I, III, IV) has been run on the Internal Market Index, a European benchmark of country performance towards the Single Market [25].

All point (I–IV) have been tackled in [13], to illustrate how SA can put an environmental debate into track by showing that the uncertainty in the decision on whether to burn or dispose solid urban waste depends on the system of indicators chosen and not on the quality of the available data (e.g. emission factors). In this example, a hypothetical Austrian decision maker must take a decision on the issue of solid waste management, based on an analysis of the environmental impact of the available options, i.e. landfill or incineration. The model reads a set of input data (e.g. waste inventories, emission factors for various compounds) and generates for each option a composite indicator CI. $CI(x)$ quantifies how much the option (x) would impact on the environment. The target function Y, defined as the logarithm of the ratio between the CI for incineration and that for landfill, suggests incineration for negative values of
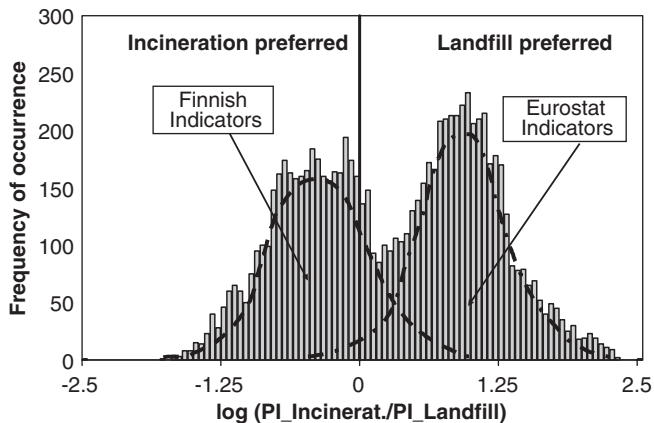
Fig. 1. Uncertainty distribution of the target output function *Y*. The bimodal structure of this distribution nicely shows how the choice of what composite indicator to use drives almost completely the answer of the model.

Y and landfill otherwise. Most of the input factors to the model are uncertain. What makes the application instructive is that one of the "factors" of the model is a trigger between two possible composite indicators: the first proposed by Statistics Finland and the second by the European Statistical Office (Eurostat). The analysis shows that the choice of what composite indicator to use drives almost completely the answer of the model (see Fig. 1). At the present state of knowledge, the waste management issue is non-decidable. Resources should not be allocated for a better definition of the input data (e.g. emission factors or inventories) but to reach a consensus among the different groups of experts on an acceptable composite indicator of environmental impact for solid waste.

The use of trigger variables to select one versus another conceptualisation of a given system has also been used in [15], where the trigger drives the choice of an input data set versus another, each set representing internally consistent but mutually exclusive parameterisations of the system.

1.2. A role for sensitivity analysis in regulatory prescriptions

Practices addressed in this review also meet a need for reliable SA which is also acknowledged in recent regulatory documents. Prescriptions have been issued for SA of models when these are used for policy analysis. In Europe, the European Commission recommends SA in the context of the extended impact assessment guidelines and handbook [26].

The EC handbook for extended impact assessment, a working document by the European Commission, 2002, states: "A good SA should conduct analyses over the full range of plausible values of key parameters and their interactions, to assess how impacts change in response to changes in key parameters".

The Intergovernmental Panel on Climate Change (IPCC) has issued a report on Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories [27] to the request from the United Nations Framework Convention on Climate Change (UNFCCC). Although the report mentions the existence of "…sophisticated computational techniques for determining the sensitivity of a model output to input quantities…", the methods employed are merely local. One of these is the derivative normalised by the input–output standard deviations discussed later. Although the IPCC background papers [28] advise the reader that […the sensitivity is a local approach and is not valid for large deviations in non-linear functions…], they do not provide any prescription for non-linear models.

The best set of prescriptions on the use of SA in modelling is likely the forthcoming *Draft* Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models, prepared by The EPA Council for Regulatory Environmental Modeling (CREM), where one reads, in a section entitled "Which Method to Use?" "Methods should preferably be able to (a) deal with a model regardless of assumptions about a model's linearity and additivity; (b) consider interaction effects among input uncertainties; and (c) cope with differences in the scale and shape of input PDFs; (d) cope with differences in input spatial and temporal dimensions; and (e) evaluate the effect of an input while all other inputs are allowed to vary as well [29], see also [30]. Of the various methods discussed above, only those based on variance […] are characterized by these attributes. When one or more of the criteria are not important, the other tools discussed in this section will provide a reasonable sensitivity assessment."

These official prescriptions seem to confirm that SA, as part of the modelling process, has become all the more urgent due to an increased awareness, among practitioners and the general public, of the need for quality controls on the use of scientific models. We touch this issue briefly next.

1.2.1. The critique of modelling

Last decade has witnessed a change in the role of science in society. This has led to the emergence of the issue of legitimacy in science, the end of scientists' purported neutrality and the need to cope with plurality of frames of reference and value judgements.

A framework for the production of scientific knowledge that has policy, as opposed to academia, as an interlocutor, has been studied by Funtowicz and Ravetz [31–34] and Gibbons et al. [35]. When models are used for policy analysis, one must acknowledge that today's role of scientists in society is not that of revealing truth, but rather of providing evidence, be it "crisp" or circumstantial, based on incomplete knowledge, sometimes in the form of probability, before and within systems of conflicting stakes and beliefs [34]. This is often referred to as the post-normal science setting where "facts are uncertain, values in dispute, stakes high and decisions urgent" [32]. In these contexts, SA can become part of a quality framework

for model-based assessment becoming an element of assessment pedigree, see www.nusap.net and [36].

Models, as key element of the scientific method, could not escape this critique. Scientists frequently caution themselves against blind reliance in the use of models [37]. As an example, *Science* hosted a blunt debate on subject of system analysis and the issue of quality in the modelling process [38,39]. Noticeable was also Konikov and Bredehoeft's work [40], entitled "Groundwater models cannot be validated", reviewed on *Science* in "Verification, Validation and Confirmation of numerical models in the earth sciences", by Oreskes et al. [41]. Both papers focused on the impossibility of model validation. According to Oreskes et al., natural systems are never closed, and models put forward as description of these are never unique. Hence, models can never be "verified" or "validated", but only "confirmed" or "corroborated" by the demonstration of agreement (non-contradiction) between observation and prediction. Scepticism about instrumental use of models is not confined to academia, see the debate between RIVM laboratories and the Dutch press [42] concerning the use of environmental models in the absence of proper and reliable quality criteria for the assessment of model uncertainties. What role should uncertainty and SA play to alleviate these problems?

1.3. Uncertainty and sensitivity analysis

Starting from the critique of modelling, Hornberger and Spear [37] argue that "…most simulation models will be complex, with many parameters, state-variables and non-linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be made to produce virtually any desired behaviour, often with both plausible structure and parameter values."

As a way out of the impasse, the concept of inference's robustness has been elegantly expressed by Edward E. Leamer [43]: "I have proposed a form of organised SA that I call "global SA" in which a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighborhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful."

Furthermore according to Lemons et al. [44] not recognising the "value laden" nature of the framing assumptions used in modelling, results in studies appearing "more factual and value-neutral than warranted".

All this shows that a possible use of models is therefore to map assumptions into inferences, where we use "inference" sensu lato, to indicate a model-based prediction or statement that is relevant to the analysis served by the model. According to good practice in modelling, neighbourhoods of alternative assumptions should be selected and the corresponding interval of inferences identified, rather than mapping a single set of assumptions into a single inference. Thus model-based inference would

come in the form of an empirically generated distribution of values for the prediction of interest. The production of this distribution is commonly known as uncertainty analysis, while the process of linking the uncertainty in the inference to the uncertainty in the assumptions is known as SA.[1]

SA can be seen as the extension to "numerical" experiments of experimental design tools [45]. One avenue of investigation to extend experimental design to numerical experiments was that of Sacks et al. [8] and Welch et al. [46], who focused on designing optimal sampling points to estimate Y at untried points using Bayesian analysis, thus minimising the number of model evaluations. This line of investigation is presently pursued by O'Hagan, Oakley and co-workers [47,48].

Uncertainty and sensitivity analyses are most often run in tandem, and customarily lead to an iterative revision of the model structure. The process may happen to falsify the model based analysis, e.g. demonstrating that the inference offered by the model is too wide to be of use for decision [43,44]. In this case SA may offer guidance as to which of the input assumptions is a better candidate for further analysis aimed to reduce the uncertainty in the inference. We detail next some essential requirements for SA to meet these tasks.

1.3.1. Essential requirements for sensitivity analysis

Computational models of real or manmade systems, as opposed to concise physical laws, are attempts to mimic systems by extracting and encoding system features, within a process that cannot be scientifically formalised [49]. The practice is motivated by the hope that the model will produce information that has a bearing (via a decoding exercise) on the system under investigation. One would hence expect that an important element of model-based analysis would be a justification of the encoding process, e.g. of what was willingly left out. Next, as models do not lend themselves to an intuitive understanding of the relationship between what goes into the model, in terms of factors, laws and structures, and the prediction that comes out of it, one would expect a mapping of model assumptions into model inferences (uncertainty analysis). Finally the modeller should investigate the relative role of the various assumptions in shaping the inference (SA). Uncertainty and sensitivity analyses should be performed iteratively, thus corroborating the encoding process, e.g. by

---

[1] A useful example on the meaning of this linking process can be taken from IPCC [28]: CO_2 emissions from energy production will, in most countries, contribute to a large fraction of total national emissions and in this respect they can be a key source. However, the uncertainty in emission factors and activity data is usually low. That means that the contribution to total uncertainty in CO_2 equivalents is low and little is normally gained by improving the methodology with respect to reducing the total inventory uncertainty. In most inventories, emissions of N_2O from agriculture will constitute a smaller fraction of total emissions of N_2O, but will contribute significantly to the total uncertainty on CO_2 equivalents. Much would be gained by reducing the uncertainty in this source.

showing that elements left out of the model were non-relevant (model simplification), as well as providing guidance for research, e.g. by showing what factors deserve further analysis or measurement (factors prioritisation). What requirements should we impose on our SA so that it can be up to these tasks?

First it is that SA should not be concerned with the model output per se, but with the question that the model has been called to answer. To make an example, if a model predicts contaminant distribution over space and time, the output of interest for SA should be the total area where a given threshold is exceeded at a given time, or the total health effects per unit time. This would depend on which question the model is trying to answer (about assessing a pesticide, the feasibility of a chemical plant, setting an emission threshold, etc.) and on the regulatory requirements at hand. Plots of factors importance computed at different points in space and time would be too many to look at, making the SA irrelevant or perfunctory. An implication of this is that models must change as the question put to them changes. The optimality of a model must be weighted with respect to the task. According to Beck [50], a model is "relevant" when its input factors actually cause variation in the model response that is the object of the analysis. Model "un-relevance" could flag a bad model, or a model unnecessarily complex, used to fend off criticism from stakeholders (e.g. in environmental assessment studies). As an alternative, empirical model adequacy should be sought, especially when the model must be audited. An implication of this is that the merit of a proposed policy could be challenged in the space of the policy options. Different possible emission thresholds could be shown to induce no appreciable variation in the output of concern, e.g. health effect to population, once all other sources of uncertainty have been weighted in.

A *second* requirement of a SA is that all known sources of uncertainty are properly acknowledged, and that the analysis acts on them simultaneously, to ensure that the space of the input uncertainties is thoroughly explored and that possible interactions (Appendix B) are captured by the analysis. Some of the uncertainties might be the result of a parametric estimation, but others may be linked to alternative formulations of the problems, or different framing assumptions which might reflect different views of reality, as well as different value judgements posed on it. When there are observations available to compute posterior probabilities on different plausible models, then SA would plug into a Bayesian model averaging (BMA).

Mechanistic models used in many scientific contexts (e.g. environmental sciences), based on traditional scientific descriptions of component processes, almost always contain ill-defined parameters and are thus referred to as over-parameterised models (e.g. [51], p. 487). Accordingly it is often concluded that the estimation of a unique set of parameters, optimising goodness of fit criteria given the observations, is not possible. Moreover, different competing model structures (different constitutive equations, different types of process considered, spatial/temporal resolution, etc.) are generally available that are compatible with the same empirical evidence [37]. This implies the unfeasibility of the traditional estimation approach.

The analyst is then referred to the broader concept of calibration and acceptability, by allowing for, e.g. the use of qualitative definitions expressed in terms of thresholds, based on "theoretical" (physical, chemical, biological, economical, etc.) constraints, expert opinions, legislation, etc.

In practice, we give up any attempt to find a well-defined optimum, but try to characterise in a compact and readable way the combinations of all model parameters/hypotheses/structures that drive the model to a "good" behaviour.

This implies the analysis of a multidimensional function with, possibly, multiple optima and high order interactions.

Calibration procedures involve Monte Carlo simulation analyses, which can be divided into two big classes: Monte Carlo filtering (MCF) and Bayesian analysis.

Both approaches entail an uncertainty analysis followed by a SA, which assumes now a peculiar and critical value. In fact, the scope of SA is not only to quantify and rank in order of importance the sources of prediction uncertainty, but, which is much more relevant to calibration, to identify the elements (parameters, assumptions, structures, etc.) that are mostly responsible for the model realisations in the acceptable range [52].

Bayesian analysis is usually implemented by the so-called BMA [53,54] is an approach to modelling in which all possible sources of uncertainty are taken into account (model structures, model parameters and data uncertainty) based on Bayesian theory.

SA can be of great help in characterising the properties of the posterior distribution in a compact way, allowing the identification of which input factors (model parameters, structures or hypotheses) or which combinations of them are mostly controlled by data and hence are mostly responsible for good model behaviour. An example is given next. A combination of BMA and SA applied to time series modelling is in [55].

Another way for a SA to become irrelevant is to have different tests thrown at a problem, and different factors importance rankings produced without clue as to what to believe. To avoid this, a *third* requirement for SA is that the concept of importance be defined rigorously before the analysis. In this article we show how this can be achieved by referring to the "model simplification" and "factor prioritisation" tasks just described.

It is also important that uncertainty and SA be used in the process of model development, prior and within model use in analysis. Once an analysis has been produced, its revision via SA by a third party is not something most modellers would willingly submit to.

Properly executed, a SA can gauge model adequacy and relevance, identify critical regions in the space of the inputs, discover factors' interactions, establish priorities for research, and simplify models. These analyses may be part

of a model's pedigree [26,36]. What tools would meet the requirements outlined so far?

2. The methods

We present the methods via a simple example, with a possibly self-evident sensitivity pattern, in order to allow a comparison between methods' prediction and the reader's expectation. The example is a simple linear model[2]

$$Y = \sum_{i=1}^{r} \Omega_i Z_i, \qquad (1)$$

where Y is the output of interest, Z_i are the uncertain input factors and Ω_i are constant coefficients. Zero-centred normal (see Appendix B for definitions) distributions are assumed for the Z_i's, independent from one another:

$$Z_i \sim N(\bar{z}_i, \sigma_{Z_i}), \quad \bar{z}_i = 0, \quad i = 1, 2, \ldots, r. \qquad (2)$$

We also assume $\sigma_{Z_1} < \sigma_{Z_2} < \cdots < \sigma_{Z_r}$, and $\Omega_1 > \Omega_2 > \cdots > \Omega_r$. From (1) and (2), Y results normally distributed with parameters $\bar{y} = \sum_{i=1}^{r} \Omega_i \bar{z}_i$, $\sigma_Y = \sqrt{\sum_{i=1}^{r} \Omega_i^2 \sigma_{Z_i}^2}$.

An SA of model (1–2) should tell us something about the relative importance of the uncertain factors Z_i in determining the output of interest Y.

According to the most widespread practice (Appendix A), the way to do this is by computing derivatives, i.e. $Y'_{Z_i} = \partial Y / \partial Z_i$. Y'_{Z_i} can be computed using an array of different analytic, numeric or coding techniques [2]. For model (1–2) $Y'_{Z_i} = \Omega_i$, independently of σ_{Z_i}. The order of importance of our factors based on Y'_{Z_i} would then be $Z_1 > Z_2 > \cdots > Z_r$, which is at odd with our expectation that the factors' standard deviation should also play a role in the uncertainty in Y. We would suspect that factor with a very large σ_{Z_i} could happen to be the most important factor even if its Ω_i were not the largest.

An available practice is a normalisation of the derivatives by the standard deviations, i.e. $S_{Z_i}^{\sigma} = (\sigma_{Z_i} / \sigma_Y)(\partial Y / \partial Z_i)$. For model (1–2) $S_{Z_i}^{\sigma} = \Omega_i(\sigma_{Z_i} / \sigma_Y)$. Note that while Y'_{Z_i} is truly local in nature, as it needs no assumption on the range of variation of factor Z_i, $S_{Z_i}^{\sigma}$ needs such assumption, so that $S_{Z_i}^{\sigma}$ is a hybrid local–global measure. Recalling that for model (1–2) $\sigma_Y = \sqrt{\sum_{i=1}^{r} \Omega_i^2 \sigma_{Z_i}^2}$, i.e. $\sum_{i=1}^{r} (\Omega_i^2 \sigma_{Z_i}^2 / \sigma_Y^2) = 1$, then for this model $\sum_{i=1}^{r} \left(S_{Z_i}^{\sigma} \right)^2 = 1$, i.e. the squared $S_{Z_i}^{\sigma}$ give how much each individual factor contributes to the variance of the output of interest. If one is trying to assess how much the uncertainty in each of the input factors will affect the uncertainty in the model output Y, and if one accepts the variance of Y to be a good measure of this uncertainty, then the squared $S_{Z_i}^{\sigma}$ seem to be a good measure, that reconciles the SA with our expectation. The factor with a large σ_{Z_i} could this time end up as the most important. The $S_{Z_i}^{\sigma}$

---

[2] In [52] this example was presented as a portfolio model, but it could as well be seen as a simple physical model, e.g. a composite indicators made up of a linear combination of weighed standardised variables [23,24].

measure is also recommended by some existing guidelines, albeit with caveats [27]. The relation $\sigma_Y^2 = \sum_{i=1}^{r} \left(S_{Z_i}^{\sigma} \right)^2$ is not general; it only holds for linear models. For this reason we find unwarranted any use of OAT approaches with models other than strictly linear (see also the examples in this section and in Appendix F). The almost totality of sensitivity analyses met in the literature, not only *Science*'s ones (Appendix A), are of an OAT type [59].

To treat non-linear models we must abandon derivatives. A Monte Carlo experiment on our model demands the generation of a sample matrix

$$\mathbf{M} = \begin{pmatrix} z_1^{(1)} & z_2^{(1)} & \cdots & z_r^{(1)} \\ z_1^{(2)} & z_2^{(2)} & \cdots & z_r^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ z_1^{(N)} & z_2^{(N)} & \cdots & z_r^{(N)} \end{pmatrix}.$$

$\mathbf{M}$ is composed of N rows, each row being a trial set for the evaluation of Y. Being the factors independent, each column can be generated independently from the marginal distributions specified in (2) above. Computing Y for each row in $\mathbf{M}$ results in the output vector $\mathbf{y} = [y^{(1)}, y^{(2)}, \ldots, y^{(N)}]^{\mathrm{T}}$. Feeding both $\mathbf{M}$ and $\mathbf{y}$ into a least-squares algorithm, the analyst will obtain a model of the form $y^{(l)} = b_0 + \sum_{i=1}^{r} b_{Z_i} z_i^{(l)}$. Comparing this with (1) it is easy to see that $b_0 = 0$, $b_{Z_i} = \Omega_i$, $i = 1, 2, \ldots, r$ (if $N \geqslant r + 1$). Being dimensioned, the b_{Z_i} coefficients are not used for SA. The practice is to compute the standardised regression coefficients (SRCs), defined as $\beta_{Z_i} = b_{Z_i} \sigma_{Z_i} / \sigma_Y$. Hence for model (1–2), $\beta_{Z_i} = S_{Z_i}^{\sigma}$. For linear models and independent factors, $(\beta_{Z_i})^2$ provide the fraction of the variance of the model due to each factor. For non-linear models, one has to consider the model coefficient of determination $R_y^2 \in [0,1]$ (see Appendix B for definitions), which represents the fraction of the model's variance accounted for by the regression equation. If this is not too low, e.g. $R_y^2 \geqslant 0.7$, we can still use the β_{Z_i} for SA with the understanding that it will only explain 70% of the model's variance.

While $S_{Z_i}^{\sigma}$ corresponds to a variation of factor Z_i all other factors being held constant, the β_{Z_i} offers a measure of the effect of factor Z_i that is averaged over the N possible values of the other factors. This averaging is essential for non-linear models. Note that for a model non-monotonic in Z_i we could have zero β_{Z_i} even if Z_i were to be an influent factor (Appendix B), as we shall show in a moment.

In order to treat models of an unknown degree of linearity and/or monotonicity, imagine to fix factor Z_i at its midpoint $\bar{z}_i$, making it a constant. How much would this change the variance of Y? We indicate the conditioned variance as $V_{\mathbf{Z}_{-i}}(Y|Z_i = \bar{z}_i)$, where the variance is taken over $\mathbf{Z}_{-i}$, a $(r-1)$ dimensional vector of all factors but Z_i. Comparing the input factors based on $V_{\mathbf{Z}_{-i}}(Y|Z_i = \bar{z}_i)$'s, e.g. by saying that the smaller $V_{\mathbf{Z}_{-i}}(Y|Z_i = \bar{z}_i)$, the most important Z_i, would make the comparison dependant upon

where the factors are fixed. Furthermore for non-linear models, fixing a factor might actually increase the variance instead of reducing it, depending upon where it is fixed. To avoid this, the practice is to average $V_{\mathbf{Z}_{-i}}(Y|Z_i = \bar{z}_i)$ over all possible values of Z_i obtaining $E_{Z_{-i}}(V_{\mathbf{Z}_{-i}}(Y|Z_i))$. A known result from textbook algebra is that:

$$V_Y = E_{Z_i}(V_{\mathbf{Z}_{-i}}(Y|Z_i)) + V_{Z_i}(E_{\mathbf{Z}_{-i}}(Y|Z_i)), \qquad (3)$$

where $V_{Z_i}(E_{\mathbf{Z}_{-i}}(Y|Z_i))$ is called the *main effect* of Z_i on Y, and $E_{Z_i}(V_{\mathbf{Z}_{-i}}(Y|Z_i))$ the *residual*. Given that $V_{Z_i}(E_{\mathbf{Z}_{-i}}(Y|Z_i))$ is large if Z_i is influential, its ratio to the unconditional variance V_Y is used as a measure of sensitivity,

$$S_{Z_i} = \frac{V_{Z_i}(E_{\mathbf{Z}_{-i}}(Y|Z_i))}{V_Y}, \qquad (4)$$

which is nicely scaled in [0,1]. For model (1–2), one finds $S_{Z_i} = (S_{Z_i}^{\sigma})^2 = \beta_{Z_i}^2$, due to the fact that the model is linear and the factors independent [52]. A nice property of the S_{Z_i}'s when applied to (1–2) is that $\sum_{i=1}^{r} S_{Z_i} = 1$. The same was true for the $\beta_{Z_i}^2$ when applied to (1–2). Yet property $\sum_{i=1}^{r} S_{z_i} = 1$ holds for *additive* models (Appendix B), while $\sum_{i=1}^{r} \beta_{Z_i}^2 = 1$ only holds for linear models. S_{Z_i} is a good model-free sensitivity measure, and for all models it gives the expected reduction in the variance of the output that one would obtain if one could fix an individual factor Z_i. For non-additive models $\sum_{i=1}^{r} S_{Z_i} < 1$, as we illustrate next. To this effect, we now let the coefficients Ω_i in model (1) become uncertain as well, e.g. [3]

$$\Omega_i \sim N(\bar{\omega}_i, \sigma_{\omega_i}), \quad \text{with } \bar{\omega}_i \neq 0, \quad i = 1, 2, \ldots, r. \qquad (5)$$

The model (1,5) has now $k = 2r$ input factors $\mathbf{X} \equiv (\Omega_1, \Omega_2, \ldots, \Omega_r, Z_1, Z_2, \ldots, Z_r)$, and we shall use the symbol S_j when we do not need to distinguish between S_{Z_i}, S_{Ω_i}.

Fig. 2 and Tables 1 and 2 show a SA for $r = 3$ and the following parameter values:

$$
\begin{array}{cccc}
r = 3 & \sigma_{Z_1} = 1 & \bar{\omega}_1 = 3 & \sigma_{\Omega_1} = 3 \\
k = 6 & \sigma_{Z_2} = 3 & \bar{\omega}_2 = 1.5 & \sigma_{\Omega_2} = 2 \\
& \sigma_{Z_3} = 5 & \bar{\omega}_3 = 1 & \sigma_{\Omega_3} = 1.5
\end{array}
$$

The S_{Z_i}, S_{Ω_i}, $i = 1, 2, \ldots, r$ coefficients have been estimated using Monte Carlo based methods.

The S_i and higher order terms have all been computed using [15], that is an extension and computational improvement of the method of Sobol' [7]. In general, the Sobol' method has a computational cost for each index of $N = 1024$. In [15], which uses in an optimal way some symmetry properties of the Sobol' sample design, the total cost for the result in Figs. 2 and 3 is reduced to about 10,000 runs. For the test case in (Appendix F) next the cost was about 3 times as much. The cost for the Morris method was of 70 runs. A free software to perform all calculations reported in this paper can be freely downloaded from
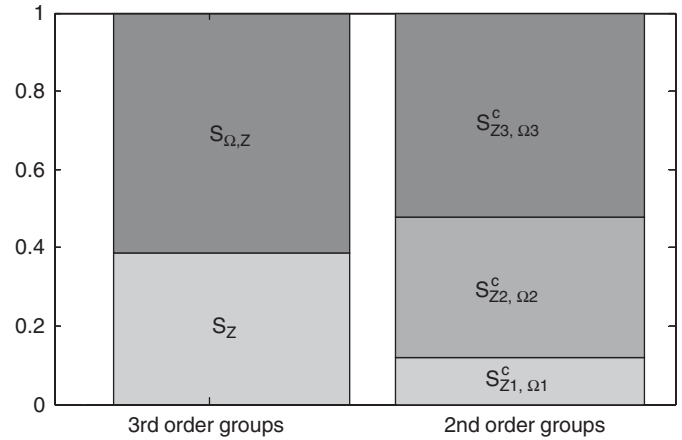

Fig. 2. Sensitivity analysis for second and third order groups of model (1).

www.jrc.cec.eu.int/uasa/. Ongoing research aims to accelerate the computation of the measures [15,48,56]. The measures can also be computed via High Dimensional Model Representations, an approach to model approximation and SA by Rabitz and co-workers ([19, p. 199; 21]; see also http://www.princeton.edu/~hrabitz/pubsubjects.html#hdmr).

The columns relative to the Y_j' and OAT measures vindicate our statements about the unreliability of these methods even for the simple example at hand. Both measures mistake Z_1 for the most important factor. $(S_j^{\sigma})^2$ and β_j^2 agree with one another and with the S_j by giving zero importance to the Ω_i coefficients. The agreement between $[(S_j^{\sigma})^2, \beta_j^2]$ with S_j is not general, as a slight modification of the model (1,5) can easily show (see result s of the modulus version (9) below in Fig. 3 and Tables 3 and 4).

The S_j's no longer add up to one, as the model has become non-additive. We can recover the missing fraction of variance by computing the sensitivity measure on groups of factors, such as, e.g. $S_{Z_2 Z_3}^{c} \equiv V_{Z_2 Z_3}(E_{\mathbf{X}_{-Z_2 Z_3}}(Y|Z_2, Z_3))/V_Y$. We see in Fig. 2 that $S_{Z_2 Z_3}^{c} \cong 0.32 \cong S_{Z_2} + S_{Z_3}$. Computing instead $S_{\Omega_3 Z_3}^{c}$ yields $S_{\Omega_3 Z_3}^{c} = 0.519 > S_{\Omega_3} + S_{Z_3} = 0.177$. We now indicate as $S_{\Omega_3 Z_3}$ (no superscript c) the difference $S_{\Omega_3 Z_3} = S_{\Omega_3 Z_3}^{c} - S_{\Omega_3} - S_{Z_3}$ and so on for other combinations of factors (Fig. 2). All effects of the type $S_{Z_i Z_l}, S_{\Omega_i \Omega_l}, S_{Z_i \Omega_l}, i \neq l$ are null for model (1,5). When the combined effect of two factors, i.e. $S_{\Omega_3 Z_3}^{c}$, is greater than the sum of the individual effects S_{Ω_3} and S_{Z_3}, then this extra effect $S_{\Omega_3 Z_3}$ is the *interaction* (or two-way, or second order) effect of Ω_3, Z_3 on Y. $S_{\Omega_3 Z_3}^{c}$ captures instead all effects including factors Ω_3, Z_3. We see from Fig. 2 that if we sum all first order effects with all second order ones we obtain 1, i.e. all the variance of Y is accounted for, because model (1,5) has interactions only up to the second order. Note that the effect of the Ω_i is only via interactions, to which the S_j coefficients are blind.

For a system with k factors there may be interaction terms up to the order k, i.e. [7]:

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \sum_i \sum_{j>i} \sum_{l>j} S_{ijl} + \cdots + S_{12\ldots k} = 1.$$

$$(6)$$

---

[3] The choice $\bar{\omega}_i = 0$, $i = 1, 2, \ldots, r$ would have generated a model without additive component, i.e. $S_j = 0$, $\forall j$.

Table 1
Main effects and total effects for model (1,5), compared to other sensitivity measures

	FP setting					FF setting			
	S_j	β_j^2 ($R^2 = 0.35$)	OAT	Y_j'	$(S_j^\sigma)^2$	S_{Tj} + ranking		μ_j + ranking	
Z_1	0.066	0.061	3	3	0.060	0.1198	5	7.3482	5
Z_2	0.147	0.136	1.5	1.5	0.135	0.3618	2	13.7051	2
Z_3	0.173	0.164	1	1	0.167	0.5234	1	16.9067	1
Ω_1	0.001	0	0	0	0	0.0569	6	7.1508	6
Ω_2	0	0	0	0	0	0.2202	4	9.2131	4
Ω_3	0.004	0	0	0	0	0.3413	3	12.1134	3

Table 2
Relevant higher order sensitivity indices for model (1,5)

$S_{Z_1\Omega_1}^c$	0.1188	$S_{Z_1Z_2}^c$	0.214	$S_Z = S_{Z_1Z_2Z_3}^c$	0.3865
$S_{Z_2\Omega_2}^c$	0.3610	$S_{Z_1Z_3}^c$	0.239	$S_\Omega = S_{\Omega_1\Omega_2\Omega_3}^c$	0.0054
$S_{Z_3\Omega_3}^c$	0.5193	$S_{Z_2Z_3}^c$	0.320	$S_{Z\Omega} = S_{Z_1\Omega_1} + S_{Z_2\Omega_2} + S_{Z_3\Omega_3}$	0.608
Sum	0.9991	Sum	0.773	Sum	1

The fact that for model (1,5), all terms above the second order are zero is convenient. Already for the present moderate value of $k = 6$, the summands in (6) are very numerous to look at. One would have 6 first order terms, $\binom{6}{2} = 15$ second order, $\binom{6}{3} = 20$ third order, $\binom{6}{4} = 15$ fourth order, $\binom{6}{5} = 6$ firth order, and last term of order k. This makes as many as $2^k - 1 = 2^6 - 1 = 63$. To overcome this difficulty, several practitioners have since long suggested [19,52] to use the *total effect* terms. Imagine that we substitute X_j with $\mathbf{X}_{-j}$ in the formula for the first order effect of X_j, obtaining $V_{\mathbf{X}_{-j}}(E_{X_j}(Y|\mathbf{X}_{-j}))/V_Y$. By definition this is the first order effect of $\mathbf{X}_{-j}$, which can be easily demonstrated (Appendix C) to equal the sum of all terms in development (6) that do not include X_j. Hence $S_{Tj} = 1 - V_{\mathbf{X}_{-j}}(E_{X_j}(Y|\mathbf{X}_{-j}))/V_Y$ equals the sum of all terms that do include X_j. We can exemplify this with factor Ω_i. $S_{T\Omega_i} \equiv 1 - V_{\mathbf{X}_{-\Omega_i}}(E_{\Omega_i}(Y|\mathbf{X}_{-\Omega_i})/V_Y$ reduces for model (1,5) to $S_{T\Omega_i} = S_{\Omega_i} + S_{\Omega_i Z_i}$, i.e. the sum of all non-zero terms that do include Ω_i. Because of (3) we can also write: $S_{Tj} = E_{X_{-j}}(V_{X_j}(Y|\mathbf{X}_{-j}))/V_Y$.

S_j, S_{Tj} for model (1,5) are given in Table 1. As one might expect the sum of the first order terms is less than one, the sum of total effects is higher than 1. If one can compute all the k S_j terms plus all the k S_{Tj} ones, then a fairly complete and parsimonious description of the model in terms of its global SA properties is obtained.

A further simplification of the analysis can be achieved by partitioning the model inputs in groups. In our model (1), for instance, it would be fairly natural to write:

$$S_\mathbf{\Omega} + S_\mathbf{Z} + S_{\mathbf{\Omega},\mathbf{Z}} = 1, \tag{7}$$

where $\mathbf{\Omega} = \Omega_1, \Omega_2, \ldots, \Omega_r$, $\mathbf{Z} = Z_1, Z_2, \ldots, Z_r$ (Fig. 2, Table 2). Or, we could show the results from model (1) as

$$\sum_{i=1}^r S_{\mathbf{A}_i} = 1, \tag{8}$$

where $\mathbf{A}_i = (\Omega_i, Z_i)$. Different ways of grouping the factors might give different insights to the owner of the problem, e.g. formulation (8), if model (1,5) were a composite indicator, would give the overall uncertainty brought by a sub-indicator, including its value Z_i and weight Ω_i (Appendix C).

The indices S_j, S_{Tj} have an interpretation in terms of framing of the SA. As mentioned in Section 1, if one may obtain different ordering of factors' importance using different methods, why bothering doing it? Importance is not per se a mathematical concept. Hence "importance" should be defined at the stage of framing the analysis, possibly in a way that is meaningful to its users. In [16,52] we discuss settings applicable to different contexts for SA.

2.1. Factors' prioritisation (FP) setting

The objective of SA is here to identify the factor which, if determined (i.e. fixed to its true, albeit unknown, value) would lead to the greatest reduction in the variance of the target output Y, and so on for the second most important factor etc., till all factors are ranked.

The concept of importance is thus precised, and linked to a reduction of the variance of the target function. In general, one would not be able to meet the objective of Setting FP, as this would imply knowing what the true value of a factor is. The purpose of Setting FP is to allow a rational choice under uncertainty. It is clear from the discussion so far that S_j is the appropriate measure for this setting. Note that this approach is blind at model's interactions [16,52]. The implication of this for model (1,5) is that only the Z_i factors can be candidate for factor prioritisation.

2.2. Factors' fixing (FF) setting

This is concerned with model simplification, by fixing non-influential factors [7]. The objective of this setting,

Table 3
Main effects and total effects for model (9), compared to other sensitivity measures

	FP setting							FF setting			
	S_j + ranking		β_j^2 + ranking ($R^2 = 0.12$)		OAT	Y_j' [a]	$(S_j^\sigma)^2$ [a]		S_{Tj} + ranking		μ_j + ranking (5th–6th shifted)	
Z_1	−0.001	6	0	6	0	—	—		0.173	5	5.67	6
Z_2	0.094	3	0	5	0	—	—		0.408	3	8.06	3
Z_3	0.108	2	0	4	0	—	—		0.549	1	11.73	1
Ω_1	0.013	5	0.007	3	0	—	—		0.104	6	6.59	5
Ω_2	0.088	4	0.045	2	0	—	—		0.331	4	7.81	4
Ω_3	0.145	1	0.065	1	0	—	—		0.457	2	9.54	2

[a] Local derivatives undefined for absolute values when $\bar{z}_j = 0$.

Table 4
Higher order relevant sensitivity indices for model (9)

$S_{Z_1\Omega_1}^c$	0.036	$S_{Z_1Z_2}^c$	0.107	$S_Z = S_{Z_1Z_2Z_3}^c$	0.268
$S_{Z_2\Omega_2}^c$	0.227	$S_{Z_1Z_3}^c$	0.128	$S_\Omega = S_{\Omega_1\Omega_2\Omega_3}^c$	0.222
$S_{Z_3\Omega_3}^c$	0.406	$S_{Z_2Z_3}^c$	0.236	$S_{Z\Omega}$	0.510
Sum	0.669	Sum	0.471	Sum	1

which could also be labelled as "screening", is to identify the factor or the subset of input factors that we can fix at any given value over their range of uncertainty without reducing significantly the output variance. If such set is identified, this means that the remaining factors explain practically all the unconditional variance.

The non-influential factors can be fixed anywhere in their range of variation without significant loss of information in the model. If one has prior beliefs about the importance of input factors, this setting can be used to prove or disprove a given model representation. It can be seen that $S_{Tj} = 0$ is a necessary and sufficient condition for X_j to be totally non-influent (Appendix C). The method of Morris [57] for factors' screening, which provides a factors' ranking very close to S_{Tj}, can also be used for this setting [52] when the model is expensive to evaluate, as this method demands a smaller sample size (Appendix D). Setting FF is sensitive to factors' interactions, e.g. it shows (Table 1), that *only* ω_1 can be reasonably fixed in model (1,5).

To better clarify the merit of the model-free measures S_j, S_{Tj} within the FP and FF settings, let us now consider the slightly more complicated modulus version of (1):

$$Y = \left| \sum_{j=1}^{r} Z_j \Omega_j \right| \tag{9}$$

with parameter values:

$$r = 3 \quad \sigma_{Z_1} = 1 \quad \bar{\omega}_1 = 3 \quad \sigma_{\Omega_1} = 3$$
$$k = 6 \quad \sigma_{Z_2} = 3 \quad \bar{\omega}_2 = 1.5 \quad \sigma_{\Omega_2} = 2$$
$$\sigma_{Z_3} = 5 \quad \bar{\omega}_3 = 1 \quad \sigma_{\Omega_3} = 1.5$$

We can see in Fig. 3 and Tables 3 and 4 that now the difference between variance based indices on one hand and OAT, local or β_j^2 indices on the other, becomes crucial. While in the original model version (1,5) the β_j^2's and the $(S_j^\sigma)^2$'s gave the same information as the S_j's, in this case the FP setting can be correctly addressed by using the S_j's. The FF setting can only be dealt using the S_{Tj}'s or the μ_j's.

Sometimes practitioners want to analyse input factors with respect to their capacity to produce realisation of the model output Y within a given region, e.g. between bounds, or above a threshold. This leads to a *Factors' mapping (FM) setting,* whose question is "which factor is mostly responsible for producing realisations of Y in the region of interest?" This can be studied by MCF [58], whereby realisations of Y produced by plain Monte Carlo are classified, e.g. as acceptable or non-acceptable. In this way $\mathbf{y} = [y^{(1)}, y^{(2)}, \ldots, y^{(N)}]^T$ is split into two sub-samples, and likewise each of the input samples. Comparing the empirical distribution of the unfiltered realisation versus the filtered ones for each input factor, the factors importance in determining the realisations of interest can be gauged (Appendix E). The example treated so far is of a "prognostic" nature. We use information in the model to make inference on reality. For a "diagnostic" use of models, where we use reality to make inference about the
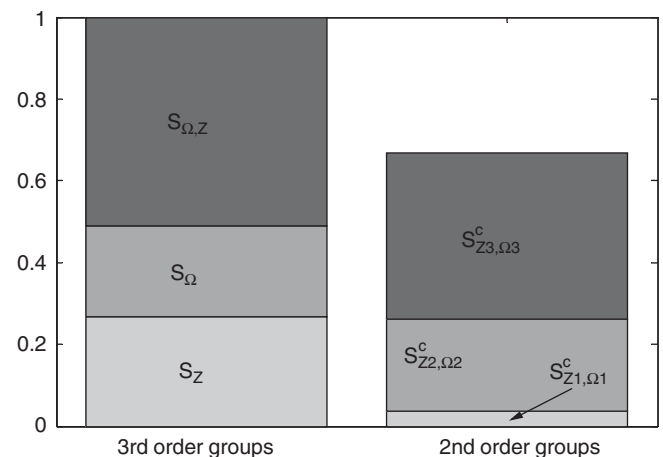

Fig. 3. Sensitivity analysis for second and third order groups of model (9).

structure of the model, such as, e.g. in calibration settings, SA can likewise be valuable, especially when largely overparametrised principle-based models are compared against available evidence. An application is in Appendix F, where we show how an overparametrised six-factor model can be nailed down to a two-factor model using SA. We are unaware of alternative, more succinct approaches to this problem.

3. Conclusions

We have presented a few good practices for a defensible SA,[4] based on requirements such as the choice of a synthetic output variable, the simultaneous analysis of all uncertainties, and a rigorous definition of importance. In summary, if a scalar objective function Y is available, whose variance can be taken as the subject of the analysis, then variance based measures offer a coherent strategy, that is agile (the owner of the problem can decide on if and how to group the factors), model free (also works for non-monotonic, non-additive models), and reduces to SRCs for linear models. This practice lends itself to intuitive interpretations of the analysis, such as that in terms of factors' prioritisation or fixing. None of these properties is shared by the OAT methods. We have also linked the emergence of SA as a discipline to the critique of modelling. In most controversial debates where science plays a role, the negotiation will take place in the space of the uncertainties [37,43,45]. In these contexts SA helps in providing evidence that is defensible, transparent in its assumptions and comparable against and across different framing assumptions [31–35].

Appendix A. Sensitivity analysis in *Science*: a review

Even if there are examples of good practices in the specialised SA literature (discussed in the introduction), as an indicator of the status quo of the practices still used in SA in the modelling community, we have reviewed all papers published in the last 6 years on *Science* where SA is used to complement model-based analysis (33 papers or letters with a keyword search "sensitivity analysis/analyses"). In general, it appears that the need of SA to support and validate model inference is universally acknowledged, and in most cases partially addressed, but available best SA practices are not applied [59], with the partial exception of [60]. The Monte Carlo approach allowing for the simultaneous propagation of the entire input distributions is used only for uncertainty analysis purposes, while for SA, the methods applied are the local derivatives or the one-at-a-time approach (OAT), which is sometimes wrongly applied also for uncertainty analysis purposes. Another problem is that SA is usually performed

only for a subset of parameters (sometimes only one as in [61–63]) selected on the basis of the modeller's prior knowledge of the model. This impedes, for example, some possible misspecifications to emerge, e.g. when a high sensitivity is detected for a parameter which, according to the modeller's intentions, should be irrelevant for the model behaviour. More specifically:

(I) The link between sensitivity and robustness is well established, i.e. modellers agree that an inference becomes stronger if it can be demonstrated that it is insensitive to uncertainties. If the analysis is limited to an Uncertainty Analysis [61–63,64–75], the implementation of a plain error propagation of all uncertainty sources simultaneously with a Monte Carlo approach is intuitive and straightforward and should not imply methodological problems [64,69,71]. However, in the majority of cases [61–68,70,72,73], the analysis is still done perturbing by fixed intervals and OAT the model parameters which is unacceptable because OAT does not allow for co-operative effect of the various uncertainty sources to appear, making the robustness analysis indefensible, apart from the simple case of linear models.

(II) When, after the error propagation in robustness analysis, the analyst is interested in ranking the importance of various sources of uncertainty, i.e. when the true SA is performed [65,71,76–81], the OAT approach is always applied. This would only be justified for purely linear models and, as we have argued in the text this approach can be grossly misleading, and the model free methods should be used instead. The only exception is [60], where partial ranked correlation coefficients (PRCCs) are used to assess relative importance of input factors. Even with the limitation of correlation/regression based methods, the application of a correlation/regression based technique is here quite appropriate, and at least this is the only one example of a global SA method applied. In two cases [78,79], the error propagation implies a falsification of the inference (i.e. uncertainty impede a clear answer and robustness cannot be demonstrated), and parameters are ranked by importance to identify which sources of uncertainty have to be reduced in future work to allow a clear answer. These two latter papers show a conceptually sound use of SA for analysing the robustness of an inference and subsequently identifying the main sources of uncertainty. However, also in such papers the practice is wanting, as a simple OAT approach has been taken.

(III) In some studies sensitivity is a property of a model [66,68,82,83]. A model is defined "sensitive" when the behaviour of a state variable (e.g. a concentration of a chemical or biological element) is highly affected by small variations in other state variables (e.g. temperature, pressure, etc.). Partial derivatives, i.e. local SA, are used to assess this property. We can include in this

---

[4] We have eluded for brevity the treatment of non-independent input factors. In [52] we show how both settings FP, FF can be dealt with using S_j, S_{Tj} when the factors are not independent.

framework the "sensitivity" studies in the Global Change community [75]. Sensitivity in the context of inverse problems in dynamical models, to track dynamical state variables in the input that can cause specified variations in the dynamical output state variables, can also be found in [84].

(IV) SA is sometimes considered in estimation/calibration problems, in which the model fit is evaluated in an OAT fashion [61,65,85]. In [85] SA is used to calibrate an over-parameterised model, where the amount of data is insufficient to estimate the whole set of model parameters (kinetic constants). In such a case, a Monte Carlo study is performed to identify the ranges of parameter values allowing a good fit. A SA to adding explanatory variables can be found [86] as well as a case of scenario analysis, which is referred to as SA [87].

(V) SA is also used as an argument to confute or criticise published model based inferences [86,88–90] or support the defence of model results in technical comments [91,92]. Also in such cases the SA is done using the OAT approach or SA is just an uncertainty analysis.

Appendix B. Some useful definitions.

Normal distribution. If X is normally distributed with parameters μ and σ the probability of X taking a value between x and $x + dx$ is $(1/\sqrt{2\pi}\sigma)e^{-1/2(x-\mu/\sigma)^2}\,dx$.

Additive model. A model $Y = f(Z_1, Z_2, \ldots, Z_k)$ is additive if f can be decomposed as a sum of k functions f_j, each of which is only function of the relative factor Z_j. As trivial examples, $Y = \sum_i Z_i^2$ is non-linear, but additive, while $Y = \prod_i Z_i$ is non-linear and non-additive.

Monotonicity. The relationship between Y and an input factor Z is monotonic if the curve $Y = f(Z)$ is non-decreasing or non-increasing over all the interval of definition of Z. A model with k factors is monotonic if the same rule applies for all factors. In SA of numerical models, this can be verified by Monte Carlo simulation followed by scatter-plots of Y versus each factor at a time. Note that the β_{Z_i}'s can fail grossly for non-monotonic models, such as, e.g. $Y = \sum_i e^{-Z_i^2}$ or $Y = \prod(1/Z_i^2 - 1)$. For these models, if we kept the factors' distribution as from (2), the β_{Z_i} would be zero for all factors Z_i, while the factors would all be influent instead.

Model coefficient of determination. The model coefficient of determination is $R_y^2 = \sum_{i=1}^{N}(\hat{y}^{(i)} - \bar{y})^2 / \sum_{i=1}^{N}(y^{(i)} - \bar{y})^2$, where $\hat{y}^{(i)}$ is the regression model prediction.

Standardised regression coefficients. The $\hat{\beta}_{Z_i}$ provide a regression model in terms of standardised variables $\tilde{y} = (\hat{y} - \bar{y})/\sigma_Y$; $\tilde{z}_i = (z_i - \bar{z}_i)/\sigma_{Z_i}$, i.e. $\tilde{y} = \sum_{i=1}^{r}\hat{\beta}_{Z_i}\tilde{z}_i$, where $\hat{y}$ is the vector of regression model predictions. The β_{Z_i}'s tell us how much this fraction of the variance can be decomposed according to the linear functions of the input factors, leaving us ignorant about the rest. The β_{Z_i}'s are a

progress with respect to the $S_{Z_i}^{\sigma}$; they can be always computed, also for non-linear models, as well as for models with no analytic representation (e.g. a computer programme that computes Y).

Appendix C. Sensitivity analysis by groups and total sensitivity indices

As discussed in the text, the variance of model (1), can be decomposed as $S_{\mathbf{\Omega}} + S_{\mathbf{Z}} + S_{\mathbf{\Omega},\mathbf{Z}} = 1$, where, $\mathbf{Z} = Z_1, Z_2, \ldots, Z_r$. The information we obtain in this way is clearly lower than that provided by the S_j and S_{Tj}. We see that the effect of the $\mathbf{\Omega}$ set at the first order is zero, while the second order term $S_{\mathbf{\Omega},\mathbf{Z}}$ is 0,61, so it is not surprising that the sum of the total effects is 1,61 (the 0,61 is counted twice): $S_{T\mathbf{\Omega}} = S_{\mathbf{\Omega}} + S_{\mathbf{\Omega},\mathbf{Z}}$, $S_{T\mathbf{Z}} = S_{\mathbf{Z}} + S_{\mathbf{\Omega},\mathbf{Z}}$. Now all that we know about the sensitivity pattern of model (1) is the combined effect of all the coefficients $\mathbf{\Omega} = \Omega_1, \Omega_2, \ldots, \Omega_r$, that of all the factors $\mathbf{Z} = Z_1, Z_2, \ldots, Z_r$, and their interaction. Note that if we decompose the variance of model (1) as $\sum_{i=1}^{r}S_{\mathbf{A}_i} = 1$, where $\mathbf{A}_i = (\Omega_i, Z_i)$, we do not need summands such as $\sum_{\substack{j=i \\ j>i}}^{r}S_{\mathbf{A}_i\mathbf{A}_j}$ because the problem has returned additive, given that the interactions are "within" the groups $\mathbf{A}_i$.

The total effect term can also be understood in terms of groups. Imagine we group factors into X_i and $\mathbf{X}_{-i}$. Then $S_i + S_{-i} + S_{-i,i} = 1$, where $S_{-i} = V_{\mathbf{X}_{-i}}(E_{X_i}(Y|\mathbf{X}_{-i}))/V_Y$, and

$$S_{Ti} = 1 - S_{-i} = S_i + S_{-i,i} = E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i}))/V_Y$$

equals the total effect of X_i, i.e. its first order term plus all interactions between X_i and $\mathbf{X}_{-i}$.

As said in the main text, the condition $E_{\mathbf{X}_{-i}}(V_{X_i}(Y|\mathbf{X}_{-i})) = 0$ is both necessary and sufficient for factor X_i to be non-influent, which makes the index S_{Ti} suited for fixing non-influential factors. A demonstration is the following. If X_i is non-influent, then $E_{\mathbf{X}_{-i}}(V_{X_i}((Y|\mathbf{X}_{-i})) = 0$ because fixing "all but X_i" results in the inner variance over X_i to be zero (under that hypothesis the variance of Y is driven only by non-X_i), and this remains zero if we take the average over all possible values of non-X_i. As a result S_{Ti} is zero if X_i is totally non-influential. Conversely, $E_{\mathbf{X}_{-i}}(V_{X_i}((Y|\mathbf{X}_{-i})) = 0$, implies that $V_{X_i}(Y|\mathbf{X}_{-i})$ is identically zero for any possible value of $\mathbf{X}_{-i}$, as a variances can only take positive values and hence we cannot achieve a zero mean by averaging $V_{X_i}(Y|\mathbf{X}_{-i})$ of opposite signs at different points. This implies that Y is either a constant or only depends from $\mathbf{X}_{-i}$. This completes our demonstrations.

To dispel the fear that S_i and S_i^T are too complex statistics to estimate, we explain below how they can be computed when the input factors are independent.

1. Choose a base sample dimension N.
2. Generate a Monte Carlo sample of dimension $2N$ of the input factors and define two matrices of

data, each containing half of the sample (ordered by column):

$$A = \begin{bmatrix} x_1^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\ x_1^{(2)} & \cdots & x_i^{(2)} & \cdots & x_2^{(1)} \\ \vdots & & & \ddots & \vdots \\ x_1^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)} \end{bmatrix} \qquad B = \begin{bmatrix} x_1^{(N+1)} & \cdots & x_i^{(N+1)} & \cdots & x_k^{(N+1)} \\ x_1^{(N+2)} & \cdots & x_i^{(N+2)} & & x_2^{(N+1)} \\ \vdots & & & \ddots & \vdots \\ x_1^{(2N)} & \cdots & x_i^{(2N)} & \cdots & x_k^{(2N)} \end{bmatrix}$$

3. Define a matrix M_i formed by all columns of A, except the ith column which is taken from B and a matrix M_{Ti}, complementary to M_i, formed with the ith column of A and with all the remaining columns of B:

$$M_i = \begin{bmatrix} x_1^{(1)} & \cdots & x_i^{(N+1)} & \cdots & x_k^{(1)} \\ x_1^{(2)} & \cdots & x_i^{(N+2)} & \cdots & x_2^{(1)} \\ \vdots & & & \ddots & \vdots \\ x_1^{(N)} & \cdots & x_i^{(2N)} & \cdots & x_k^{(N)} \end{bmatrix} \qquad M_{Ti} = \begin{bmatrix} x_1^{(N+1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(N+1)} \\ x_1^{(N+2)} & \cdots & x_i^{(2)} & \cdots & x_2^{(N+1)} \\ \vdots & & & \ddots & \vdots \\ x_1^{(2N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(2N)} \end{bmatrix}$$

4. Compute the model output for all the input values in the sample matrices A, M_i and M_{Ti}, obtaining three column vectors of model outputs of dimension $N \times 1$: $\mathbf{y} = f(A)$, $\mathbf{y}' = f(M_i)$, $\mathbf{y}'_T = f(M_{Ti})$.

5. The sensitivity indices are hence computed based on scalar products of the above defined vectors of model outputs. First define the quantities:

6. $$U = \frac{1}{N}\sum_{j=1}^{N} y^{(j)} y'^{(j)}, \qquad U_T = \frac{1}{N}\sum_{j=1}^{N} y^{(j)} y'^{(j)}_T,$$

7. $$f_0 = (1/N)\sum_{j=1}^{N} y^{(j)} \quad \text{(The mean)},$$

$$V = (1/N)\sum_{j=1}^{N} y^{(j)2} - f_0^2 \quad \text{(The total variance)},$$

then,

8. $$V_i = V_{X_i}[E_{\mathbf{X}_{-i}}(Y|X_i)] = U - f_0^2,$$
$$V_{-i} = V_{\mathbf{X}_{-i}}[E_{X_i}(Y|\mathbf{X}_{-i})] = U_T - f_0^2$$

and finally

9. $$S_i = \frac{V_i}{V}, \qquad S_{Ti} = 1 - \frac{V_{-i}}{V}.$$

As shown, the computation of S_i and S_i^T involves very easy matrix algebra (scalar products).

Appendix D. The method of Morris

The method of Morris is an efficient and easy to implement screening tool. It operates at lower sample size then the variance based measures. Like in experimental design, when using Morris each factor is sampled at a small number (e.g. 2, 3, 4) of selected values, called levels. The distance between two consecutive levels is Δ, i.e. $l_{j+1} = l_j + \Delta$. For each factor, a number of one-step differences are estimated along the various axes. For instance, for factor X_j one computes at different sample points $\mathbf{X}^i = (X_1^i, X_2^i, ..., X_j^i, ..., X_k^i)$, $i = 1, ..., N'$, differences such as

$$EE_{X_j}^i = (Y(\mathbf{X}^i + \Delta\mathbf{e}_j) - Y(\mathbf{X}^i))/\Delta,$$

where $(\mathbf{X}^i + \Delta\mathbf{e}_j)$ is equal to vector $\mathbf{X}^i$ apart for its jth component that has been increased by Δ, i.e. which has been sampled at the successive level. Values of $EE_{X_j}^i$ over different points $i = 1, 2, ..., N'$ are averaged, to produce a Morris μ_{X_j} and its related standard deviation σ_{X_j}. Hence although an individual $EE_{X_j}^i$ term can be seen as an OAT approach, the summary statistics μ_{X_j} and σ_{X_j} are global measures of sensitivity.

The value of N' for Morris is as a rule quite lower than N for variance based methods, say $N' \approx (\frac{1}{100}, \frac{1}{10})N$. The total number of model evaluations needed to estimate all μ_{X_j} and σ_{X_j} is a linear function of the number of factors k, i.e. $r \times (k + 1)$.

In problem settings such as that of Factor Fixing, we recommend the use of the modulus difference $EE_{X_j}^i$ above because, as explained in [52,93] the resulting modified

measure μ_{X_j} can be proved to be *equivalent* (in terms of ranking important factors) to S_{Tj}. The equivalence is intuitive. If moving X_j of one step along its axis at all points in the input space results in no modulus change in Y, then surely X_j is non-influent. Values of the modified (i.e. taking the absolute value in (1)) Morris measure μ for the case with $k = 2r$ factors are given in the last columns of Tables 1–3. One can see that the ordering of the factors based on $S_{Tj}\mu_j$, which is in general very similar, in this case is even identical.

Appendix E. The test of Smirnov in sensitivity analysis

The Smirnov two-sample test (two-sided version) is used on each factor, independently [37]. The Smirnov test is applicable when a qualitative definition for the "good" or "acceptable" behaviour of a model can be defined, e.g. through a set of constraints, thresholds, ceilings, time bounds based on available information on the system. The steps for the analysis are as follows.

- Define a range for k input factors X_i $(1 < i < k)$, reflecting uncertainties in the model and make a number of Monte Carlo simulations. Each Monte Carlo simulation is associated to a vector of values of the input factors.
- Classify model outputs, according to the specification of the "acceptable" model behaviour (qualify a simulation as behaviour (B) if the model output lies within constraints, non-behaviour $(\bar{B})$ otherwise).
- Classifying simulations as either B or $\bar{B}$, a set of binary elements is defined allowing to distinguish two sub-sets for each X_i: $(X_i|B)$ of m elements and $(X_i|\bar{B})$ of n elements (where $n + m = N$, the total number of Monte Carlo runs performed).
- The Smirnov two-sample test (two-sided version) is performed for each factor *independently*, analysing the maximum distance d_{max} between the cumulative distributions of the B and $\bar{B}$ sets (see Fig. 4).

In [52], MCF, BMA and related topics are discussed, including limitations of pure MCF and the combination of global SA and BMA.

Appendix F. The use of sensitivity analysis in a diagnostic setting—calibrating an hyper-parametrised model

Consider a calibration problem for a computational model. We do not know how the model is done—imagine it is a computer code. The output of interest Y is a measure of likelihood, obtained after comparing the model prediction m with data d, e.g. $Y = \text{const.} \times \exp(-[\text{sum of squared residuals of } m \text{ versus } d])$. In many cases, the log-likelihood scale is considered, which, following the definition above, is equivalent to the negative of the sum of scores.

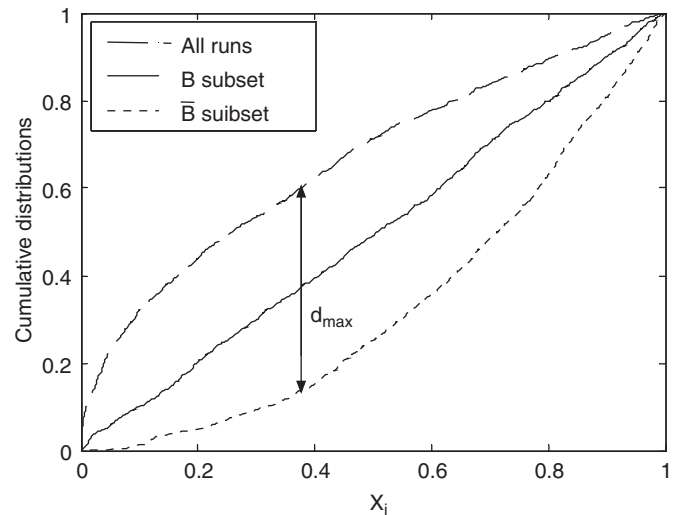Suppose we model some phenomenon varying with time (e.g. a chemical reaction), and we have a set of observations

Fig. 4. Graphical representation of the Smirnov test for checking the separation of the distributions under the B and $\bar{B}$ sets.

$t = 1, \ldots, T$, then:

$$Y = \text{const.} \times \exp\left(-\sum_{t=1}^{T}(m_t - d_t)^2\right).$$

How can we characterise the good parameter set for model calibration? Under a classical estimation framework one would feed the likelihood function Y to some maximisation routine and try to get the maximum in the k-dimensional model parameter space. This is in principle a straightforward route but in practice can be extremely complex and cumbersome, e.g. in the case of over-parameterised models.

The present example illustrates this kind of circumstances, for a model with six parameters. We will proceed "blindly", step by step in our search, and only in the end we will uncover the interaction structure underlying the calibration.

Some graphical inspection is, probably, the first step one could perform. So, assume we did some thousands of Monte Carlo runs and we wanted to analyse the mapping between the model parameters and the likelihood. Parameters are sampled from univariate normal distributions.

We can start with scatter plots of the log-likelihood vs. parameters. This is not very informative (Fig. 5a): no signs of nicely shaped "parabola" patterns with clearly identifiable maxima. One clear information is only available: the origin is the "worst" location in the model parameter space. Interesting…but the opposite of what we are looking for. We could then proceed with a categorisation, by selecting the model runs providing a log-likelihood larger than a given threshold. Doing again scatter plots with the filtered sample leaves us in the dark (Fig. 5b). This confirms that there is no way in characterising an optimum simply with a univariate analysis.

Let us increase the dimensionality and analyse two-dimensional (2-D) structures. This can still be done
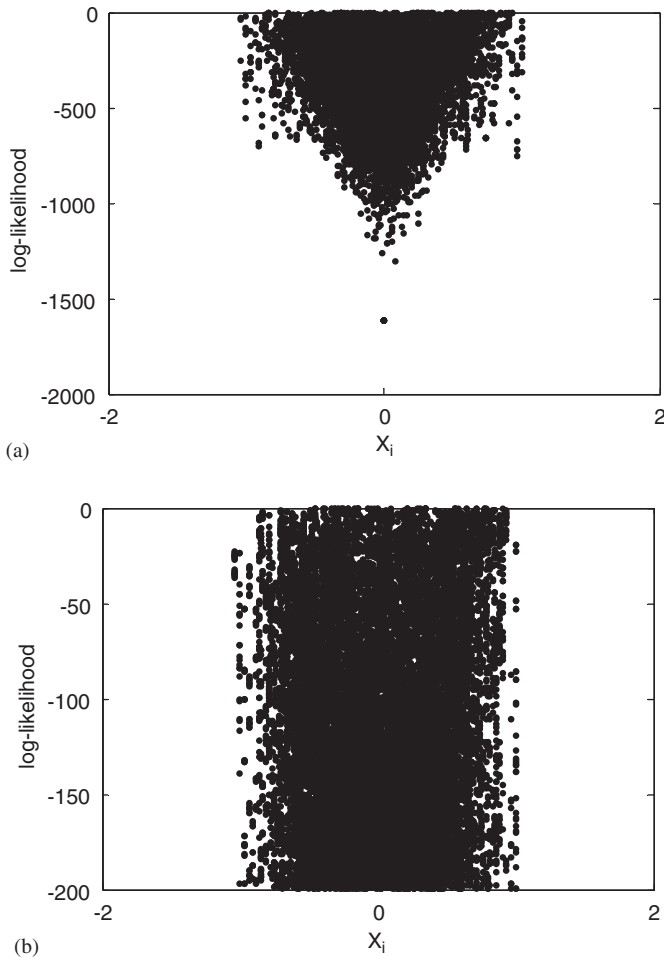
Fig. 5. (a) Log-likelihood vs. a generic input factor (patterns for all input factors are indistinguishable). (b) Same as (a) but filtering the sample for values of log-likelihood > -200 (patterns for all input factors are indistinguishable).
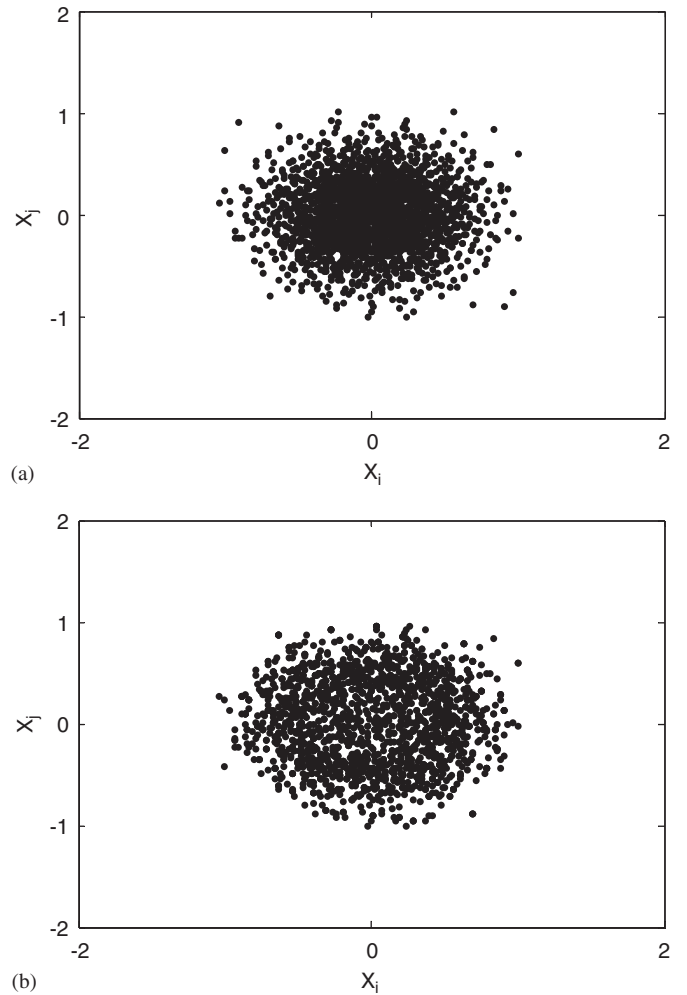

Fig. 6. (a) Pairwise scatterplots of the input sample for a generic couple of input factors (patterns for all couples are indistinguishable) (b) same as (a), but using the filtered sample for values of log-likelihood > -200 (patterns for all couples are again indistinguishable).

graphically analysing 2-D projections of the Monte Carlo sample onto planes defined by couples of model parameters (i.e. again under the form of scatter plots). Comparing the 2-D projections of the original Monte Carlo sample (Fig. 6a) to those of the filtered sample (Fig. 6b) is likewise non-informative: no difference can be noticed after the categorisation. Note that even if we computed on the filtered input factors (Fig. 6b) the pairwise correlation coefficients we would obtain zeros. Also Principal Component Analysis would not be informative as applied to the filtered input sample, as there are no correlations among the filtered factors. So, it seems that there does not exist any 2-D elementary structure allowing characterising somehow the interaction structure binding parameters under "good" behaviour. We should then move to higher dimensions. Graphically? This becomes cumbersome: how to visualise scatter plots in 3-D? Correlation analysis is clearly limited to 2-D structures, so, what can we do?

SA is able to quantify the effect of model parameters under interaction structures of any order. In this context, we extend the meaning of SA, which is not only to quantify and rank in order of importance the sources of prediction

uncertainty, but, which is much more relevant to calibration, to identify the elements (parameters, assumptions, structures, etc.) that are mostly responsible for the model realisations in the acceptable range.

Computing the first order sensitivity indices for the log-likelihood and the second order ones (Fig. 7), a story starts to emerge; there are non-zero second order effects, but only within the closed groups involving factors (X_1, X_2, X_3) and (X_4, X_5, X_6).

Computing the third order effects (Fig. 8) only $S_{123}, S_{456} \neq 0$. Regrouping and adding the terms up gives an interesting result:

$$S_{123}^c = S_1 + S_2 + S_3 + S_{12} + S_{13} + S_{23} + S_{123} = 0.5,$$
$$S_{456}^c = S_4 + S_5 + S_6 + S_{45} + S_{46} + S_{56} + S_{456} = 0.5,$$

where we have used the superscript c symbol to denote the effects closed within the indices. The variance of the problem is characterised by two groups of three factors. Higher term orders are zero.
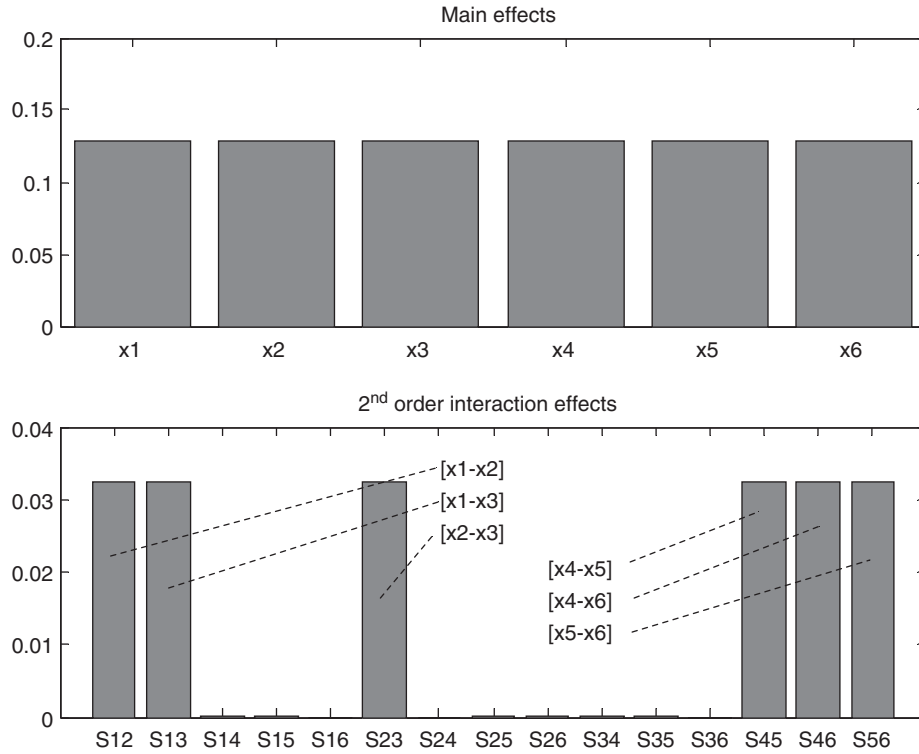
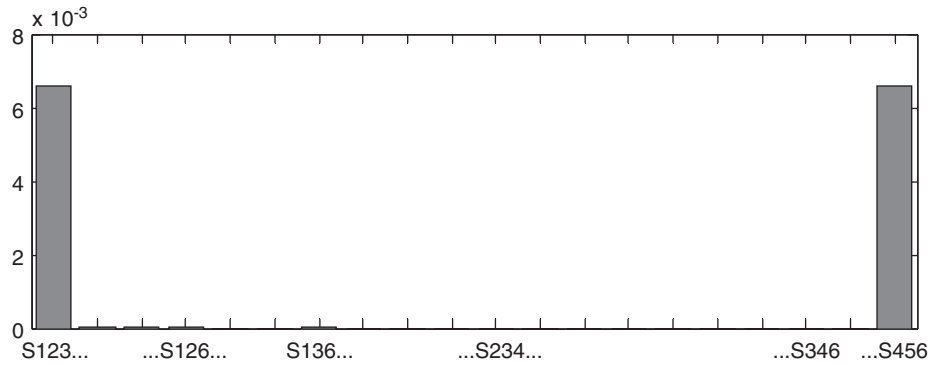Fig. 7. First and second order sensitivity indices for the log-likelihood.

Fig. 8. Third order sensitivity indices for the log-likelihood.

This leads the investigator to conclude that what could be reasonably estimated are two unknown functions of two parameter sub-sets. We can now reveal that the unknown log-likelihood function to optimise was the sum of two spheres.

$$f(X_1, \ldots, X_6) = -\left(\sqrt{X_1^2 + X_2^2 + X_3^2} - R_1\right)^2 / A_1$$
$$- \left(\sqrt{X_4^2 + X_5^2 + X_6^2} - R_2\right)^2 / A_2.$$

Were the investigator to identify this structure, by trial and error, he/she would conclude that all that estimation can provide are the two radiuses.

To this goal, the information of the "worst" location could have been useful. The easiest structure leaving out the origin is a spherical symmetry. But in which dimensions? There are as many as $\binom{6}{2} = 15$ 2-D; $\binom{6}{3} = 20$ 3-D, $\binom{6}{4} = 15$ 4-D; $\binom{6}{5} = 6$ 5-D spheres. May be too many for a blind search. SA has been helpful in driving the analysis to the correct solution!

In conclusion, where a classical estimation approach is impractical and model factors cannot be defined, nor the clear definition of a well defined model structure or set of hypotheses can be established, applying, e.g. standard statistical testing procedures, SA becomes an essential tool.

Model factors can be classified, e.g. as "important/unimportant" according to their capability of driving the model behaviour. Such capability is clearly highlighted by the SA, which plays a similar role, e.g. of a *t*-test on a least-squares estimate of a linear model. To exemplify the relevance of SA in this context, one could say that "sensitivity indices are to calibration, what standard statistical tests are to estimation" [52].

References

[1] Rabitz H. System analysis at molecular scale. Science 1989;246:221–6.

[2] Turanyi T, Rabitz H. Local methods and their applications. In: Saltelli A, Chan K, Scott M, editors. Sensitivity analysis. New York: Wiley; 2000. p. 367–83. For automated differentiation see Grievank A. Evaluating derivatives. Principles and techniques of algorithmic differentiation. Philadelphia, PA: SIAM; 2000.

[3] Varma A, Morbidelli M, Wu H. Parametric sensitivity in chemical systems. Cambridge Series in Chemical Engineering, 1999.

[4] Goldsmith CH. Sensitivity analysis. In: Armitage P, Theodore C, editors. Encyclopedia of biostatistics. New York: Wiley; 1998.

[5] Helton JC. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab Eng Syst Saf 1993;42:327–67.

[6] Iman R L, Hora S C. A robust measure of uncertainty importance for use in fault tree system analysis. Risk Anal 1990;10(3):401–6.

[7] Sobol' IM. Sensitivity estimates for nonlinear mathematical models. Mat. Model. 1990;2:112–8 (in Russian, translated in Math Model Comput Exp 1993; 1: 407–414).

[8] Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. Stat Sci 1989;4:409–35.

[9] Campolongo F, Saltelli A, Jensen NR, Wilson J, Hjorth J. The role of multiphase chemistry in the oxidation of dimethylsulphide (DMS). A latitude dependent analysis. J Atmos Chem 1999;32:327–56.

[10] Campolongo F, Tarantola S, Saltelli A. Tackling quantitatively large dimensionality problems. Comput Phys Commun 1999;117:75–85.

[11] Kioutsioukis I, Tarantola S, Saltelli A, Gatelli D. Uncertainty and global sensitivity analysis of road transport emission estimates. Atmos Environ 2004;38(38):6609–20.

[12] Zaldivar J-M, Campolongo F. An application of sensitivity analysis to fish population dynamics. In: Saltelli et al., editors. Sensitivity analysis. New York: Wiley; 2000. p. 367–83.

[13] Tarantola S, Jesinghaus J, Puolamaa M. In: Saltelli et al., editors. Sensitivity analysis. New York: Wiley; 2000. p. 385–97.

[14] Campolongo F, Rossi A. In: Proceedings of the PSAM6, sixth international conference on probabilistic safety assessment and management, Puerto Rico, June 23–28, 2002. See also Chapter 6 in Saltelli et al. Sensitivity analysis in practice. A guide to assessing scientific models. New York: Wiley; 2004.

[15] Saltelli A. Making best use of model evaluations to compute sensitivity indices. Comput Phys Commun 2002;145:280–97.

[16] Saltelli A, Tarantola S. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. J Am Stat Assoc 2002;97(459):702–9.

[17] Crosetto M, Tarantola S. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. Int J Geogr Inform Sci 2001;15(4):415–37.

[18] Pastorelli R, Tarantola S, Beghi M G, Bottani C E, Saltelli A. Design of surface Brillouin scattering experiments by sensitivity analysis. Surf Sci 2000;468:37–50.

[19] Saltelli A, Chan K, Scott M, editors. Sensitivity analysis. New York: Wiley; 2000.

[20] Saltelli A, Tarantola S, Campolongo F. Sensitivity analysis as an ingredient of modelling. Stat Sci 2000;15(4):377–95.

[21] Saltelli A, Chan K, Scott M, editors. Special Issue on sensitivity analysis. Comput Phys Commun 1999; 117 (1,2).

[22] Frey HC, Patil SR. Identification and review of sensitivity analysis methods. Risk Anal 2002;22(3):553–78.

[23] Saisana M, Tarantola S. State-of-the-art report on current methodologies and practices for composite indicator development. EUR 20408 EN, European Commission, Joint Research Centre, 2002.

[24] Saisana M, Saltelli A, Tarantola S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. J R Stat Soc A 2005;168(2):1–17.

[25] Tarantola S, Saisana M, Saltelli A, Schmiedel F, Leapman N. Statistical Techniques and participatory approaches for the composition of the European Internal Market Index 1992-2001, EUR 20547 EN, 2002.

[26] EC, 2000, European Commission's communication on extended impact assessment Brussels, 05/06/2002 COM(2002) 276 final. Guidelines for implementing the directive are available at the Governance page of the EC http://europa.eu.int/comm/governance/docs/index_en.htm

[27] IPCC, 2000, Good practice guidance and uncertainty management in National Greenhouse Gas Inventories, http://www.ipcc-nggip.iges.or.jp/public/gp/gpgaum.htm

[28] IPCC, 1999, Background papers—IPCC expert meetings on good practice guidance and uncertainty management in National Greenhouse Gas Inventories, http://www.ipcc-nggip.iges.or.jp/public/gp/gpg-bgp.htm

[29] Frey HC. Guest Editorial. Risk Anal 2002; 22(3): 539–46.

[30] Saltelli A. Sensitivity analysis for importance assessment. Risk Anal 2002;22:579–90.

[31] Funtowicz S, Ravetz JR. Science for the post-normal age. Futures 1993;25:735–55.

[32] Funtowicz S, Ravetz JR. Post-normal science—an insight now maturing. Futures 1999;31:641–6.

[33] Funtowicz S, Ravetz JR. A new scientific methodology for global environmental issues. In: Costanza R, editor. Ecological economics. New York: Columbia; 1991. p. 137–52.

[34] Funtowicz S, Ravetz JR. Three types of risk assessment and the emergence of post-normal science. In: Krimsky S, Golding D, editors. Social theories of risk. Westport, CT: Greenwood; 1992. p. 251–73.

[35] Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, Trow M. The new production of knowledge. The dynamics of science and research in contemporary societies. London: Sage; 1994.

[36] Saltelli A, Funtowicz S. The precautionary principle: implications for risk management strategies. Eur J Oncol 2003;2:67–80.

[37] Hornberger GM, Spear RC. An approach to the preliminary analysis of environmental systems. J Env Manage 1981;12:7–18.

[38] Mac Lane S. IIASA's credibility. Science 1988;241:1144.

[39] Mac Lane S. IIASA and modeling. Science 1988;242:1623–4.

[40] Konikov LF, Bredehoeft JD. Groundwater models cannot be validated. Adv Water Res 1992;15(1):75–83.

[41] Oreskes N, Shrader-Frechette K, Belitz K. Verification, validation, and confirmation of numerical models in the earth sciences. Science 1994;263:641–6.

[42] Van der Sluijs JP. A way out of the credibility crisis around model—use in integrated environmental assessment. Futures 2002;34:133–46.

[43] Leamer EE. Sensitivity analysis would help. In: Granger CWJ, editor. Modelling economic series. Oxford: Clarendon Press; 1990.

[44] Lemons J, Shrader-Frechette K, Cranor C. The precautionary principle: scientific uncertainty and type I and type II errors. Found Sci 1997;2:207–36.

[45] Box GEP, Hunter WG, Hunter JS. Statistics for experimenters. New York: Wiley; 1978.

[46] Welch WJ, Buck RJ, Sacks J, Wynn HP, Mitchell TJ, Morris MD. Screening, predicting, and computer experiments. Technometrics 1992;34:15–47.

[47] O'Hagan A, Kennedy MC, Oakley JE. Uncertainty analysis and other inference tools for complex computer codes (with discussion). Bayesian Stat 1999;6:503–24.

[48] Oakley J, O'Hagan A. Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc Ser B 2004;66:751–69.

[49] Rosen R. A comprehensive inquiry into the nature, origin, and fabrication of life. New York: Columbia University Press; 1991.

[50] Beck MB, Ravetz JR, Mulkey LA, Barnwell TO. On the problem of model validation for predictive exposure assessments. Stoch Hydrol Hydraul 1997;11:229–54.

[51] Draper NR, Smith H. Applied regression analysis. New York: Wiley; 1981.

[52] Saltelli A, Tarantola S, Campolongo F, Ratto M. Sensitivity analysis in practice. A guide to assessing scientific models. New York: Wiley; 2004.

[53] Hoeting JA, Madigan D, Rafetry AE, Volinsky CT. Bayesian model averaging: a tutorial. Stat Sci 1999;14:382–417.

[54] Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc 1995; 90(430):773–95.

[55] Saltelli A, Planas C, Tarantola S, Campolongo F. The concept of sensitivity with applications to official statistics. Res Off Stat 1999; 2: 47–77. See also Planas and Depoutot. In: Saltelli et al., editors. Sensitivity analysis. New York: Wiley; 2000. p. 367–83.

[56] Ratto M, Tarantola S, Saltelli A, Young PC. J R Stat Soc Ser B, 2004, submitted for publication.

[57] Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics 1991;33(2):161–74.

[58] Spear RC, Hornberger GM. Eutrohpication in the peel inlet II: identification of critical uncertainties via generalized sensitivity analysis. Water Res 1980;14:43–9.

[59] Saltelli A. Sensitivity analysis. Could better methods be used? J Geophys Res 1999;104(D3):3789–93 [plus 1999; 104(D19): 24,013].

[60] Blower SM, Gershengorn HB, Grant RM. A tale of two futures: HIV and antiretroviral therapy in San Francisco. Science 2000;287:650–4.

[61] Valleron A-J, Boelle P-Y, Will R, Cesbron J-Y. Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. Science 2001;294:1726–8 [in Reports].

[62] Halloran ME, Longini Jr IM, Nizam A, Yang Y. Containing bioterrorist smallpox. Science 2002;298:1428–32 [in Reports].

[63] Wilmoth JR, Deegan LJ, Lundström H, Horiuchi S. Increase of maximum life-span in Sweden, 1861–1999. Science 2000;289:2366–8 [in Reports].

[64] Neutel A-M, Heesterbeek JAP, De Ruiter PC. Stability in real food webs: weak links in long loops. Science 2002;296:1120–3 [in Reports].

[65] Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, et al. A kinetic framework for a mammalian RNA polymerase in vivo. Science 2002;298:1623–6 [in Reports].

[66] Kitano H. Systems biology: a brief overview. Science 2002;295: 1662–4 [in Review].

[67] Brookmeyer R, Blades N. Prevention of inhalational anthrax in the US outbreak. Science 2002;295:1861 [in Brevia].

[68] Royer DL, Wing SL, Beerling DJ, Jolley DW, Koch PL, Hickey LJ, et al. Paleobotanical evidence for near present-day levels of atmospheric CO_2 during part of the tertiary. Science 2001;292: 2310–3 [in Reports].

[69] Duarte CM, Agustí S. The CO_2 balance of unproductive aquatic ecosystems. Science 1998;281:234–6 [in Reports].

[70] Berger A, Loutre MF, McIntyre A. Intertropical latitudes and precessional and half-precessional cycles. Science 1997;278:1476–8 [in Technical Comments].

[71] Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome. Science 2003;300:1966–70 [Published online May 23, 2003].

[72] Gu L, Baldocchi DD, Wofsy SC, Munger JW, Michalsky JJ, Urbanski SP, et al. A response of a deciduous forest to the Mount Pinatubo eruption: enhanced photosynthesis. Science 2003;299: 2035–8.

[73] Balmford A, Bruner A, Cooper P, Costanza R, Farber S, Green RE, et al. Economic reasons for conserving wild nature. Science 2002;297:950–3.

[74] Medley GF. Predicting the unpredictable. Science 2001;294:1663–4.

[75] Reilly J, Stone PH, Forest CE, Webster MD, Jacoby HD, Prinn RG. Uncertainty and climate change assessments. Science 2001;293: 430–3.

[76] Smith AE, Slepchenko BM, Schaff JC, Loew LM, Macara IG. Systems analysis of ran transport. Science 2002;295:488–91 [in Reports].

[77] Kao RR, Gravenor MB, Baylis M, Bostock CJ, Chihota CM, Evans JC, et al. The potential size and duration of an epidemic of bovine spongiform encephalopathy in British sheep. Science 2001;295:332–5 [Published online November 22, 2001].

[78] Rojstaczer S, Sterling SM, Moore NJ. Human appropriation of photosynthesis products. Science 2001;294:2549–52 [in Reports].

[79] Sala OE, Chapin III FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, et al. Global biodiversity scenarios for the year 2100. Science 2000;287:1770–4 [in Review].

[80] Berner RA, Petsch ST, Lake JA, Beerling DJ, Popp BN, Lane RS, et al. Isotope fractionation and atmospheric oxygen: implications for phanerozoic O_2 evolution. Science 2000;287:1630–3 [in Reports].

[81] Berner RA. Geochemistry and geophysics: the rise of plants and their effect on weathering and atmospheric CO_2. Science 1997;276:544–6 [in Perspectives].

[82] Gumbsch P. Materials science: enhanced: a dislocation crash test. Science 1998;279:1489–90 [in Research].

[83] Kremen C, Niles JO, Dalton MG, Daily GC, Ehrlich PR, Fay JP, et al. Economic incentives for rain forest conservation across scales. Science 2000;288:1828–32.

[84] Druffel ERM. Pulses of rapid ventilation in the North Atlantic surface ocean during the past century. Science 1997;275: 1454–7.

[85] Eisenmesser EZ, Bosco DA, Akke M, Kern D. Enzyme dynamics during catalysis. Science 2002;295:1520–3 [in Reports].

[86] Jamison DT, Sandbu ME. Global health: who ranking of health system performance. Science 2001;293:1595–6 [in Policy Forum].

[87] Jolly D, Haxeltine A. Effect of low glacial atmospheric CO_2 on tropical African Montane vegetation. Science 1997;276:786–8 [in Reports].

[88] Lewison JGG, May RM. Government funding of research and development. Science 1997;278:878–80 [in Technical Comments].

[89] Moolgavkar SH. Consideration of copollutants. Science 2000; 290:453.

[90] Apted MJ, Langmuir D, Moeller DW, Ahn J, Waltar AE, von Winterfeldt D, et al. Yucca Mountain: should we delay? Science 2002;296:2333–5.

[91] Chesler EJ, Rodriguez-Zas SL, Mogil JS, Darvasi A, Usuka J, Grupe A, et al. In silico mapping of mouse quantitative trait loci. Science 2001;294:2423 [in Technical Comments].

[92] Davenport MP, Blower SM, Gershengorn HB, Grant RM. Reversibility of HIV drug resistance. Science 2000;288:1299.

[93] Campolongo F, Cariboni J, Saltelli A. 2004, in preparation.