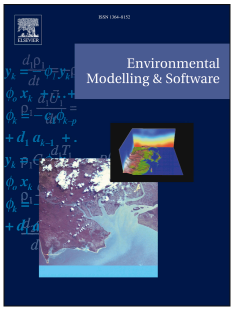
Accepted Manuscript



Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices

Andrea Saltelli, Ksenia Aleksankina, William Becker, Pamela Fennell, Federico Ferretti, Niels Holst, Sushan Li, Qiongli Wu

PII: S1364-8152(18)30282-2

DOI: https://doi.org/10.1016/j.envsoft.2019.01.012

Reference: ENSO 4376

To appear in: *Environmental Modelling and Software*

Received Date: 23 March 2018

Revised Date: 7 January 2019

Accepted Date: 15 January 2019

Please cite this article as: Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices, *Environmental Modelling and Software* (2019), doi: https://doi.org/10.1016/j.envsoft.2019.01.012.

Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices

Andrea Saltelli (1), Ksenia Aleksankina (2), William Becker (3), Pamela Fennell (4), Federico Ferretti (3), Niels Holst (5), Sushan Li (6), Qiongli Wu (7)

(1) University of Bergen, NO & Universitat Oberta de Catalunya, ES
(2) University of Edinburgh & NERC Centre for Ecology and Hydrology, Edinburgh, UK
(3) European Commission, Joint Research Centre, Ispra, IT
(4) University College London, UK
(5) Aarhus University, DK
(6) Technische Universität Darmstadt, Darmstadt, DE
(7) Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan, CHINA

Abstract

Sensitivity analysis provides information on the relative importance of model input parameters and assumptions. It is distinct from uncertainty analysis, which addresses the question 'How uncertain is the prediction?' Uncertainty analysis needs to map what a model does when selected input assumptions and parameters are left free to vary over their range of existence, and this is equally true of a sensitivity analysis. Despite this, many uncertainty and sensitivity analyses still explore the input space moving along one-dimensional corridors leaving space of the input factors mostly unexplored. Our extensive systematic literature review shows that many highly cited papers (42% in the present analysis) fail the elementary requirement to properly explore the space of the input factors. The results, while discipline-dependent, point to a worrying lack of standards and recognized good practices. We end by exploring possible reasons for this problem, and suggest some guidelines for proper use of the methods.

1 Introduction

Mathematical models have become increasingly prominent tools in decision-making processes in engineering, science, economics and policy-making, among other applications. Driven by increasing computing power, coupled with the abundance of available data, models have also become increasingly complex—examples include large climate or economic models, which aim to include ever more processes at an ever-higher resolution. However, this increased complexity requires much more information to be specified as model inputs (parameters and other assumptions used in the model construction), and typically this information is not well-known. It is therefore essential to understand the impact of these uncertainties on the model output, if the model is to be used effectively and responsibly in any decision-making process. *Sensitivity analysis* (SA) and *uncertainty analysis* (UA) are the two main tools used in exploring the uncertainty of such models.

One definition of sensitivity analysis is "the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" (Saltelli, 2002). As such it is very much related to – but distinct from – uncertainty analysis (UA), which, as we define it here, characterizes the uncertainty in model prediction, without identifying which assumptions are primarily responsible. Uncertainty analysis can include a broad range of applications relating to uncertainty—a very thorough reference can be found in (Ghanem, Higdon, & Owhadi, 2017). Ideally, an uncertainty analysis precedes a sensitivity analysis: before uncertainty can be apportioned it needs to be estimated. However, this is not necessarily the case, and applications involving model calibration/optimisation may not require the quantification of uncertainty. Other taxonomies are also possible relating UA to SA, see e.g. (Razavi, Sheikholeslami, Gupta, & Haghnegahdar, 2019), although for the purpose of the present work we remain with the definitions above.

Before proceeding, let us clarify terminology. In building a model, a number of things must be specified, including the type and structure of model, parameters, resolution, calibration data and so forth (see Figure 1). Each of these has an associated uncertainty, and is therefore an *assumption*. In a quantitative analysis of uncertainty, we can only investigate (vary) a subset of these assumptions. This subset we call the *input factors*—note that this includes all items varied in a SA or UA, i.e. model parameters, as well as any other types of assumption that will be varied. In performing any uncertainty and sensitivity analysis, it is crucial to keep in mind that the uncertainty in the assumptions that are outside the set of input factors will not be explored (Nearing & Gupta, 2018; Saltelli, Stark, Becker, & Stano, 2015). The results of the model for any values of the input factors, we call the *model output*.

Focusing now on the uncertainty in the input factors alone, if the model is deterministic, then assessing the uncertainty in the output boils down to propagating the uncertainty from the input factors to the output, for example by repeatedly running the model using different values for the uncertain inputs within their plausible ranges. This can be done with a Monte Carlo simulation, or with some ad hoc design, to generate a distribution of possible model results (the grey area in Figure 1).

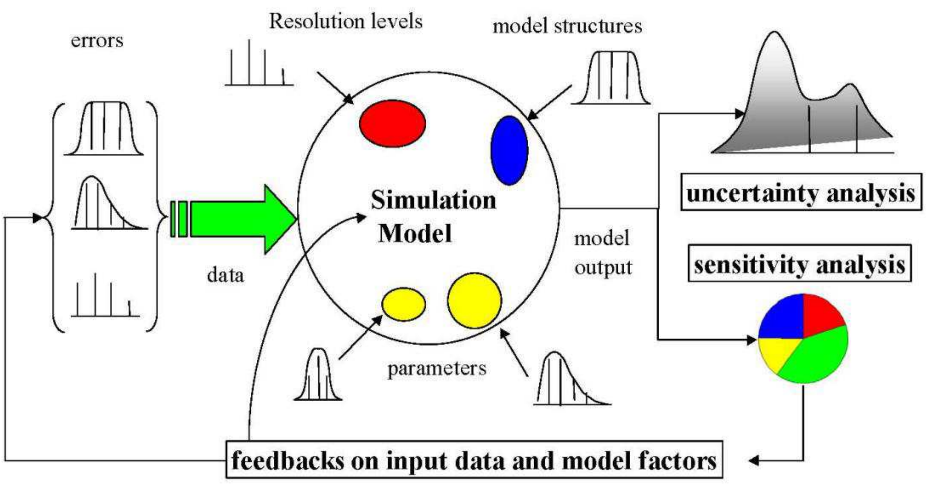


Figure 1: ***Idealized uncertainty and sensitivity analysis. Uncertainty coming from heterogeneous sources is propagated through the model to generate an empirical distribution of the output of interest (grey curve). The uncertainty in the model output, captured e.g. by its variance, is then decomposed according to source, thus producing a sensitivity analysis.***

Characterising the output distribution – e.g. by constructing it empirically from the output data points, constitutes an uncertainty analysis. The UA may also involve extracting summary statistics, such as the mean, median, and variance, from this distribution and possibly by assigning confidence bounds, e.g. on the mean.

Once this is done, the next step could be to use sensitivity analysis to assign this uncertainty to the input factors. Sensitivity analysis allows us to infer that, for example, "this factor alone is responsible for 70% of the uncertainty in the output".

Sensitivity analysis is used for many purposes. Primarily it is used as a tool to quantify the contributions of model inputs, or sub-groups of inputs, to the uncertainty in the model output—examples of such applications include (Eisenhower, O'Neill, Narayanan, Fonoberov, & Mezić, 2012) and (Becker et al., 2012). This use of sensitivity analysis will be the focus of the present paper. In this uncertainty setting, typical objectives are to identify which input factors contribute the most to model uncertainty ("factor prioritisation") so that further information might be collected about these parameters to reduce model uncertainty, or to identify factors which contribute very little and can potentially be fixed ("factor fixing") (Saltelli & Tarantola, 2002).

Other applications that are not necessarily related to uncertainty are for example in engineering design, where "design sensitivity analysis" is used as a tool for structural optimisation (Allaire, Jouve, & Toader, 2004). Sensitivity analysis can also be used to better understand processes within models, and thereby, the natural systems on which they are based (Becker et al., 2011), or as a quality assurance tool: an unexpected strong dependence of the output upon an input deemed irrelevant might either illuminate the analyst on an unexpected feature of the system or reveal a conceptual or coding error.

The importance of sensitivity analysis is widely acknowledged. Sensitivity analysis is prescribed in national and international guidelines in the context of impact assessment (e.g. (European Commission, 2009; Office of Management and Budget, 2006; U.S. Environmental Protection Agency (EPA), 2009). When the output of a model feeds into policy prescription and planning, a sensitivity analysis would appear as an essential element of due diligence.

Despite the clear importance of sensitivity analysis, there are a number of problems observed in practical sensitivity analysis and uncertainty analysis, which can be found in all fields of research. These problems range from confusions in terminology to statistically inaccurate techniques which can (perhaps dangerously) underestimate model uncertainty. Specifically:

- While most practitioners of SA distinguish it from UA, modellers overall tend to conflate the two terms, e.g. performing an uncertainty analysis and calling it a sensitivity analysis.
- The sensitivity analysis methodology often relies on so-called *local* techniques which are invalid for nonlinear models.

One of the main aims of this paper is to back up these assertions with evidence. Demonstrating that there is a systematic problem in practical sensitivity analysis might be a first step towards improving the situation. Some reviews of sensitivity analysis practice do already exist: in (Ferretti, Saltelli, & Tarantola, 2016), an assessment of the state of sensitivity analysis was performed using a bibliometric approach. (Shin, Guillaume, Croke, & Jakeman, 2013) review the state of sensitivity analysis (or lack thereof) in hydrological modelling. However, to the authors' knowledge, there is no detailed cross-disciplinary assessment of the state of sensitivity analysis, as practised by modellers.

Accordingly, this paper has the following objectives:

- To assess the “state” of sensitivity analysis across a range of academic disciplines. We do this by a systematic review of a large number of highly cited papers in which sensitivity analysis is the focus in some respect.
- To discuss – based on this review - known problems and misinterpretations of sensitivity analysis, why these might occur, and propose some ideas for how these problems might be addressed.

Following these objectives, in Section 2 we outline in more detail what we consider to be the basic requirements of a valid sensitivity analysis, as well as explaining commonly-observed problems. In Section 3 we outline a procedure for systematically selecting highly cited sensitivity analysis papers across a range of disciplines, and criteria for review. The results of this systematic review are presented in Section 4, which is followed by a discussion on the root of the problems observed, with some suggestions to improve the situation. Section 6 reports our main conclusions.

2 Common pitfalls of sensitivity analysis

There are a range of practical problems and methodological difficulties associated with sensitivity analysis. Here, we highlight two particular issues which we believe are particularly prevalent and could be addressed.

The first is a simple issue of terminology—many scientists conflate the meaning of SA and UA. In a large class of instances (e.g. in economics) SA is understood as an analysis of the robustness of the prediction (UA). This is perhaps due to an influential econometric paper (Leamer, 1985), entitled “Sensitivity analysis would help”, whose problem setting and motivation were to ensure the robustness of a regression analysis with respect to various modelling choices, e.g. in the selection of regressors. As a result, in economics and finance, it is common to see the expression ‘sensitivity analysis’ used to mean what we have defined here as uncertainty analysis. Clearly, this can have an impact on the quality of an uncertainty and sensitivity analysis, if the objectives are not even clear.

The second issue is that modellers tend to change factors one at a time (instead of globally), possibly as a result of their training and methodological disposition to think in terms of derivatives. Here we explore this technical issue in more depth.

Many practitioners accept a taxonomy of sensitivity analysis based on distinguishing between local and global methods (Saltelli et al., 2008). Let f be a generic black-box representation of a model, which has input factors $\boldsymbol{x} = \{x_1, x_2, \ldots, x_k\}$ and a scalar output y, such that $y = f(\boldsymbol{x})$. A local method in its simplest form yields the partial derivative of the model with respect to one of its input factors, i.e. $\partial y / \partial x_i$. Two notable deficiencies of this definition of sensitivity are that first, if f is nonlinear with respect to x_i, then its partial derivative will change depending on where in the range of x_i you choose to measure. Second, and more generally, if there are interactions between model inputs, then $\partial y / \partial x_i$ will change depending on the values of the remaining input factors as well. In short, first partial derivatives are only a valid measure of sensitivity when the model is linear, in which case $\partial y / \partial x_i$ will remain constant for any $\boldsymbol{x}$.

A common variation of the first partial derivative is usually referred to as the *one at a time* (OAT) approach. Let x_i^* be the nominal value of the *i*th input factor. Now define $y_i^{\max} = f(x_1^*, x_2^*, \ldots, x_i^{\max}, \ldots, x_k^*)$ as the model output where all input factors are at nominal values except the *i*th, which is set to its maximum. An OAT sensitivity measure is e.g. $\Delta_i = (y_i^{\max} - y_i^{\min})/(x_i^{\max} - x_i^{\min})$, where $y_i^{\min}$ follows a similar definition.

The OAT approach, and partial derivatives (which are a type of OAT approach), keep all other input factors fixed except the one that is being perturbed. From here on, we use the term "OAT" to refer to both local sensitivity analysis approaches and OAT of the type discussed in the preceding paragraph.

A global sensitivity analysis method, at the other extreme, could be an analysis of variance (ANOVA) as usually taught in experimental design, which informs the analyst about factors' *global* influence in terms of their contribution to the variance of the model output, including the effect of interactions among factors (Box, Hunter, & Hunter, 2005). Perhaps the most prevalent example of a global measure is the *first-order sensitivity index* (Sobol', 1993),

$$S_i = \frac{V_{x_i}\left(E_{x_{\sim i}}(y|x_i)\right)}{V(y)}$$

where $V(y)$ is the unconditional variance of y, obtained when all factors x_i are allowed to vary, and $E_{x_{\sim i}}(y|x_i)$ is the mean of y when one factor is fixed. Incidentally, this measure was originally proposed by Karl Pearson to measure nonlinear dependence between random variables (Pearson, 1905). The first-order sensitivity index is part of a class of sensitivity measures which are called 'variance-based'. Its meaning (under the assumption of independence between input factors) can be expressed in plain

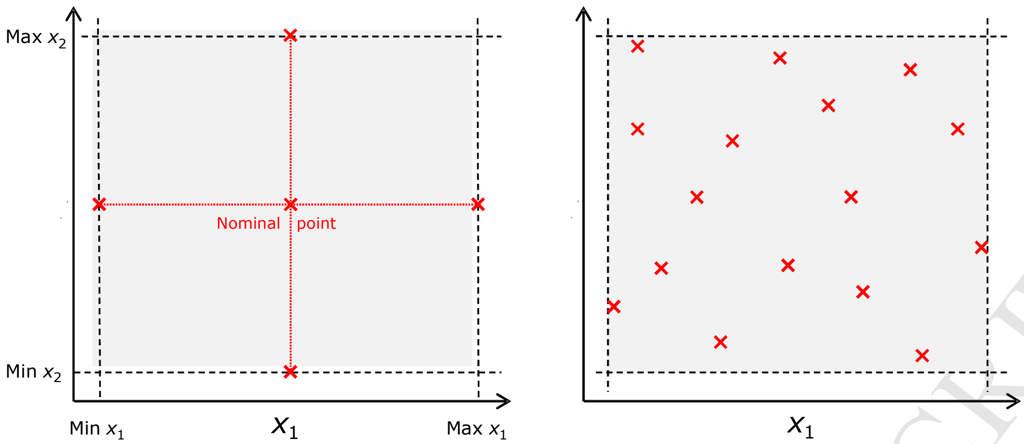


Figure 1 OAT design (left) contrasted against global design (right)

could be fixed. $S_i = 1$ implies that all of the variance of y is driven by x_i, and hence that fixing it also uniquely determines y.

Other global approaches to sensitivity analysis include the elementary effects approach (Morris, 1991), global derivative-based measures (Sobol' & Kucherenko, 2009), moment-independent methods (Da Veiga, 2015), variogram-based approaches (Razavi et al., 2019), and many others. A further discussion of the theory of sensitivity indices is beyond the scope of this paper and the reader is referred e.g. to (Saltelli et al., 2008) and (Ghanem et al., 2017).

Global approaches are requisite to performing a valid sensitivity analysis when models feature nonlinearities and interactions. To understand the issue, it is helpful to think of the set of all possible combinations of input factors as an "input space". For example, with two model inputs, any combination of values could be marked as a point on a two-dimensional plane, with the range of factor 1 on one axis, and the range of factor 2 on the other. In the case of three input factors the input space would be a cube, and for higher numbers, a hypercube. Figure 2 (left) illustrates an OAT design with two input factors, and a corresponding global design (right) that might be used to estimate the global measures discussed in the previous section.

Evidently, OAT designs cannot effectively explore a multidimensional space. We can further illustrate this with a simple example, taken from (Saltelli & Annoni, 2010). Imagine that the input space is a three-dimensional cube of side one. Moving one factor at a time by a distance of ½ away from the centre of the cube generates points on the faces of the cube, but never on its corners. All these points are in fact on the surface of a sphere internal and tangent to the cube, as illustrated in Figure 3. The volume of the sphere divided by the volume of the cube is about ½. If we increase the number of dimensions this ratio goes towards zero very quickly. In ten dimensions, the volume of the hypersphere divided by the volume of the hypercube is 0.0025, one-fourth of one percent. In practice, it is even more restrictive than that because the OAT design does not even explore inside the hypersphere, and is limited to a "hypercross". In other words, moving factors OAT in ten dimensions leaves over 99.75% of the input space totally unexplored. This under-exploration of the input space directly translates into a deficient sensitivity analysis, and is but one of the many incarnations of the so-called "curse of dimensionality", and the reason why an OAT SA is perfunctory, unless the model is proven to be linear.

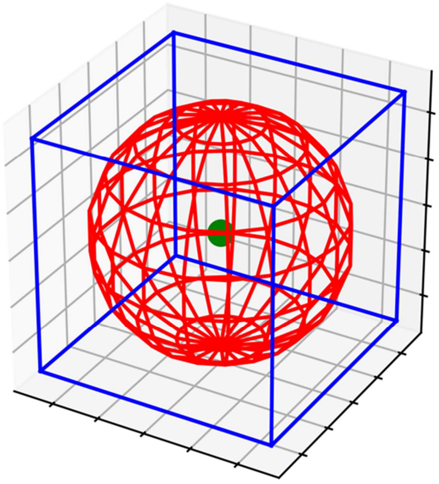


Figure 3: A sphere included in a cube (three-dimensional case) and tangent to its faces. The volume of the sphere divided that of the cube is roughly 1/2. If the dimension were ten instead of three the same ratio would be 0.0025.

Statisticians are well acquainted with this problem. This is why, in the theory of experimental design (Box et al., 2005) factors are moved in groups, rather than OAT, to optimize the exploration of the space of the factors. In sensitivity analysis, global designs are either based on random, quasi-random or space-filling designs (see Figure 2, right); or on OAT designs that are repeated in multiple locations of the input space—the latter are used for e.g. global derivative based measures, Monte Carlo estimation of variance-based sensitivity indices, and elementary effects, among others.

3 Meta-analysis

In order to understand the prevalence and type of sensitivity analysis across different fields, and to understand the extent of the issues discussed in the previous section, an extensive literature review (a meta-study) was carried out. The review was based on highly cited articles that have a focus on sensitivity analysis. The reasoning here was that the most highly cited articles should represent, on average, "commonest practice" relative to that field. Therefore, by analysing these papers, we should be able to conclude, with reasonable confidence, that the rigour of sensitivity analysis in a given field is at, or below, the level of its top-cited papers.

3.1 Selection procedure

The literature search was conducted on the Scopus database. In order to identify relevant papers, the following search criteria were used (after a few iterations of analysis and refinement)[c] . First, the strings "sensitivity analysis" *and* "model/modelling", *and* "uncertainty" were required to be present in the title, abstract or keywords. This ensures that the paper has a significant focus on sensitivity analysis, that it is related to mathematical models, and concerns uncertainty (as opposed to e.g. design sensitivity analysis and optimisation, which is a separate topic). Second, the papers were restricted to the years 2012-2017, in order to provide a sample of recent research. Finally, the results were required to be journal articles, and in English (the latter for ease of reviewing).

This search resulted in around 6000 articles. The search query is deliberately restrictive, in that sensitivity analysis articles exist that do not mention "model" in the abstract, title or keywords, for example. However, it was considered to be an unbiased way of automatically selecting sensitivity analysis papers across fields. Preliminary attempts indicated that simply mentioning "sensitivity analysis" yielded far too many irrelevant articles (around 47,000). The sample here, therefore, can be considered as representative, but the numbers of papers returned are significantly below the true number of sensitivity analysis papers in the literature.

Each paper returned by the search is tagged using one or more subject identifiers. Subject areas with less than 100 articles meeting the search criteria (of which there were eight) were not examined in this study. The resulting 19 subject areas are as follows:

- AgrBioSci (Agricultural and Biological Sciences)
- BiochemGenMBio (Biochemistry, Genetics and Molecular Biology)
- BusManAcc (Business, Management and Accounting)
- Chemi (Chemistry)
- ChemEng (Chemical Engineering)
- CompSci (Computer Science)

[c] Exact query specifications available in the Additional Online Material. Retrieved from https://www.scopus.com between March and May 2017

- DecSci (Decisional Science)
- EarthSci (Earth and Planetary Sciences)
- EconFin (Economy and Finance)
- Energy (Energy)
- Engineering (Engineering)
- EnvSci (Environmental Science)
- ImmunMicrobio (Immunology and Microbiology)
- MatSci (Material Science)
- Math (Math)
- Medicine (Medicine)
- PharTox (Pharmacology and Toxicology)
- PhysAstro (Physics and Astronomy)
- SocSci (Social Science)

In order to provide a manageable sample of articles for review, the top twenty most-cited papers from each field were selected. Since most papers include more than one subject identifier, some papers featured in more than one of the top-twenty lists. The reviewing was distributed between the authors of the present article. Even though the initial search criteria had been refined to focus on model-related sensitivity analysis, a total of 44 papers had to be discarded as not including a sensitivity analysis, nor an uncertainty analysis, or because they reported an analysis of the dependence of the output upon just one factor (which does not constitute a sensitivity analysis). A total of 280 papers were finally retained for the analysis, though in total 324 papers were reviewed.

A limitation of this selection procedure is that older papers are more likely to be well-cited, see e.g. (Davis & Cochran, 2015), therefore the distribution of papers reviewed will be biased towards older articles (our results confirm this bias). However, our reasoning is that first, it is only after a few years that it is possible to reliably identify "influential" (well-cited) papers from less influential ones, so it would be very difficult to identify influential papers only from 2017, for example. Moreover, we believe that highly cited older papers will be used as a benchmark by many researchers to guide their methodology. So highly cited papers, even if a few years old, can still be used as an indicator of the state of sensitivity analysis in a given field.

3.2 Review criteria

Each paper was reviewed against a set of simple criteria, as follows.

1. Was an uncertainty analysis performed? If so, was a global or local approach used?
2. Was a sensitivity analysis performed? If so, was a global or local approach used?
3. Was the paper primarily focused on the *method* of sensitivity analysis, or on the *model* (application)?
4. Was the model used linear, nonlinear, or was it unclear?

These criteria are explained in more detail below. Additional to these criteria, some general notes on each paper were taken.

3.2.1 OAT/global uncertainty and sensitivity analysis

The identification of OAT and global sensitivity analyses is one of the focal points of this study. In reviewing each paper, we noted whether an uncertainty analysis or sensitivity analysis had been performed, or both. For both the uncertainty and sensitivity analysis, we checked to see if the results had been generated using global or OAT methods, as discussed in Section 3.2.

As discussed, we define OAT methods as all approaches where factors are moved only one at a time, even when derivatives are computed efficiently, such as when using the adjoint method (Cacuci, 2005). Note that some methods, such as that in (Sobol' & Kucherenko, 2009) or in (Morris, 1991) *are* based on derivatives but are classified as global methods because they sample partial derivatives or incremental ratios at multiple locations in the input space.

We have defined as global any approach that is based on moving factors together, such as in Design of Experiment (DoE). A Monte Carlo analysis followed by an analysis of the scatterplots of y versus the various input factors x_i is also classified as global (albeit qualitative), as well as approaches based on regression coefficients of y versus the x_i, the use of Sobol' sensitivity indices - independently of the way these are computed, screening methods such as the method of Morris, Monte Carlo filtering, various methods known as 'moment-independent' and so on, see (Saltelli et al., 2008) for a description, and the additional online material for the methods met in the papers reviewed. Useful recent reviews are (Norton, 2015)(Pianosi et al., 2016).

One might wonder what an OAT uncertainty analysis looks like. In fact, some papers quantify uncertainty by observing $y_i^{\max}$ and $y_i^{\min}$ for each input factor during an OAT experiment, and assign the range of uncertainty on y as $[y^{\min}, y^{\max}]$, where $y^{\min} = \min_i(y_i^{\min})$, and similarly for $y_i^{\max}$. Clearly, this ignores the additional uncertainty in y when more than one factor at a time is set to its maximum or minimum values.

3.2.2 Method/model

It is useful to make a distinction between method and model-focused papers.

Model-focused papers are defined as those which focus on a model, and use sensitivity analysis as a tool to investigate uncertainty or other aspects of the model. The primary conclusions of the paper are therefore related to the model. These types of paper will often have a greater impact on the application (which is ultimately the outcome of concern), for example in assessing the uncertainty/sensitivity of climate models or other models used in decision-making.

Method-focused papers are those that introduce sensitivity analysis methodology, and use a model as a case study to demonstrate the new approach. Conclusions are therefore focused on the performance of the method, and results relating to the model are of secondary interest. Typically, the authors are familiar with sensitivity analysis techniques, which allows them to propose new approaches. These papers are more likely to feature high-quality sensitivity analysis techniques.

3.2.3 Model linearity

Finally, since OAT approaches are only valid in the case of a linear model, each paper was assessed to see if the application model was demonstrably linear or not. In many cases this was unclear, but where it was possible to ascertain linearity, this was recorded.

4 Results

The full results of this study, including the scoring matrix, as well as the authors' review notes, are given in the Additional Online Material, and a summary table is given in the Appendix.

4.1 Prevalence across disciplines

Figure 4 shows the distribution of sensitivity analysis papers across research fields, by density (number of SA papers divided by the total number in the search period) and by number. Given that model use is pervasive in the disciplines investigated these densities are very low, even accounting for the fact that not all sensitivity analysis papers will have been picked up by the search. This observation is indeed supported in investigations focusing on one discipline, such as hydrology (Shin et al., 2013). The greatest density of papers is found in decision science, as well as model-intensive subjects such as earth sciences, environmental science and energy. The greatest raw numbers are found in environmental science, engineering, and medicine, although the latter does not have a high density due to the very large overall research output. Note that articles can be tagged with more than one subject identifier.

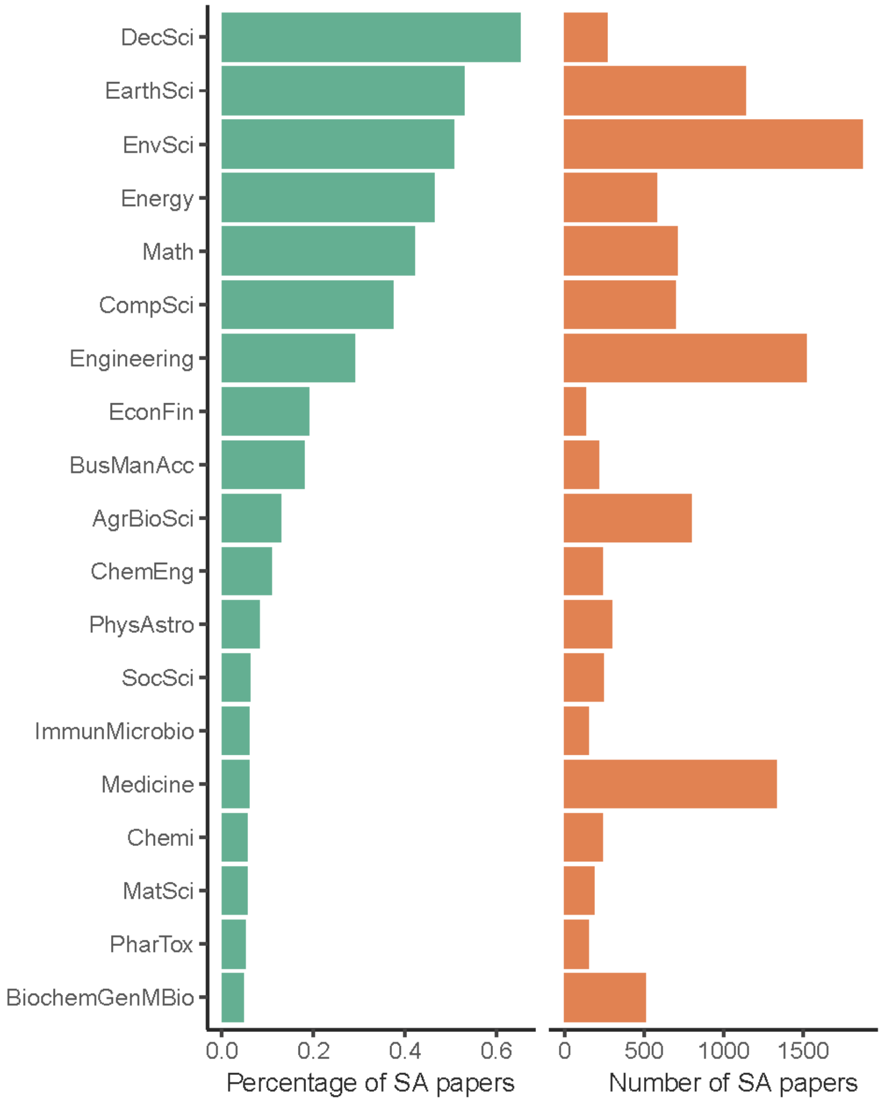


Figure 4: Density and number of sensitivity analysis articles returned by search criteria, by subject

4.2 Uncertainty analysis

Paper focus	Method	10%
	Model	90%
Model linearity	Linear	7%

	Nonlinear	61%
	Unclear	32%
Uncertainty analysis type	One at a time	7%
	Global	21%
	Unclear/absent	72%
Sensitivity analysis type	One at a time	34%
	Global	41%
	Unclear/absent	25%

Table 1: Percentages of reviewed papers based on focus, model linearity, uncertainty and sensitivity analysis type.

Although, as discussed, uncertainty analysis and sensitivity analysis are distinct (but related) disciplines, in the literature the term “sensitivity analysis” is sometimes used to describe both terms. As a result, the set of papers reviewed also included number of papers that were concerned with pure UA. Indeed, of the 280 papers reviewed, 24 did not contain any kind of sensitivity analysis and instead only concerned uncertainty analysis: these represent clear conflations of sensitivity and uncertainty analysis.

Table 1 reports the occurrence of UA found in the literature review. In about ¾ of papers, there was either no UA present, or the methodology was not clearly specified. The former is due to the fact that our search query specifically targeted sensitivity analysis papers, so it is unsurprising that there are a large proportion of papers with little attention given to the UA part. On the other hand, about ¾ of the UAs that were observed *were* global in nature. This is most likely because a Monte Carlo analysis (randomly sampling from input distributions) is fairly intuitive and accessible to most researchers, whereas an “OAT uncertainty analysis” is arguably less intuitive.

The same analysis can be applied by subject area: see Figure 5. Here we see that uncertainty analysis was found much more commonly in Pharmacology and Toxicology and Medicine (within the papers that we reviewed) than Social Sciences and Computer Science, for example. This should not be taken as an overall indication of the quantity of uncertainty analysis, because our sample has overwhelmingly targeted sensitivity analysis papers. However, it indicates that in Pharmacology and Toxicology and Medicine, either it is particularly common to perform UA simultaneously with SA, or the terms are confused. Taking the case of Pharmacology and Toxicology, we find that of the papers reviewed, only four had a sensitivity analysis, whereas ten had an uncertainty analysis. This flags that sensitivity analysis may often refer to uncertainty analysis within this field.

On the other hand, a quite prevalent trend in some fields is the practice of performing a global UA (i.e. via a Monte Carlo analysis) side by side with an OAT SA: this was observed in particular, in Medicine, and in Economics & Finance. In Medicine, for example, it seems to be common to perform an OAT sensitivity analysis, presenting the results in a tornado plot (a bar chart which shows the effect on the output of varying each assumption by a fixed amount in either direction). We speculate that the authors involved were unaware of the chance to use elementary scatterplots of the output versus the input to rank the factors by importance – or simply they did not find this kind of analysis relevant or useful. In any case, once a certain practice becomes established within a given field (i.e. found in highly cited papers), it sets a strong precedent which is difficult to supersede. Researchers and reviewers (not unreasonably) assume that if a method is found in influential articles then it must be correct.

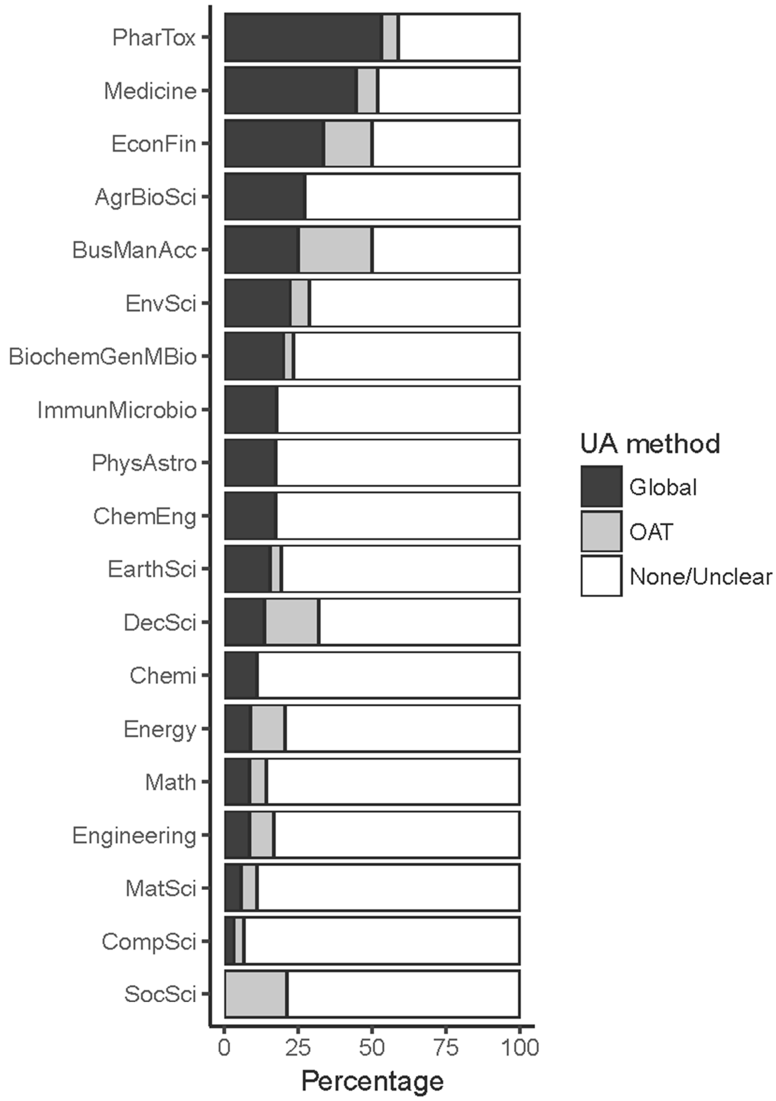


Figure 5: Classification of uncertainty analysis by subject identifier, sorted by proportion of global methods

4.3 Global vs local SA

Turning now to sensitivity analysis, Table 1 shows that 41% of sensitivity analyses use global methods, with 34% using OAT methods, and 25% having an unclear method type or no sensitivity analysis present. This is encouraging, in that nearly half of studies use global methods. Still, at least one-third of highly cited papers, matching our search criteria, use deficient OAT methods.

Figure 6 shows that the distribution of global methods varies widely across disciplines. Immunology and Microbiology show more than 70% of papers featuring global methods. This is followed by disciplines that are fairly model-intensive, such as Material Science, Biochemistry, Computer Science, and Engineering. At the other end of the spectrum, Pharmacology and Toxicology; and Business, Management and Accounting have very low proportions of global SA—about 10% and 20% respectively. Perhaps surprisingly, some disciplines that tend to rely heavily on large computer models, such as Earth Science and Environmental Science, still feature quite low rates of global sensitivity analysis. This is a concern, particularly when large-budget models are used for making significant decisions, such as climate models in policy-making—see a discussion in (Saltelli et al., 2015). On the other hand, other model-heavy subjects such as Engineering and Materials Science have higher ratios. Yet it is worth recalling that even Engineering has only around a half of confirmed global approaches, and these are the most highly cited articles.

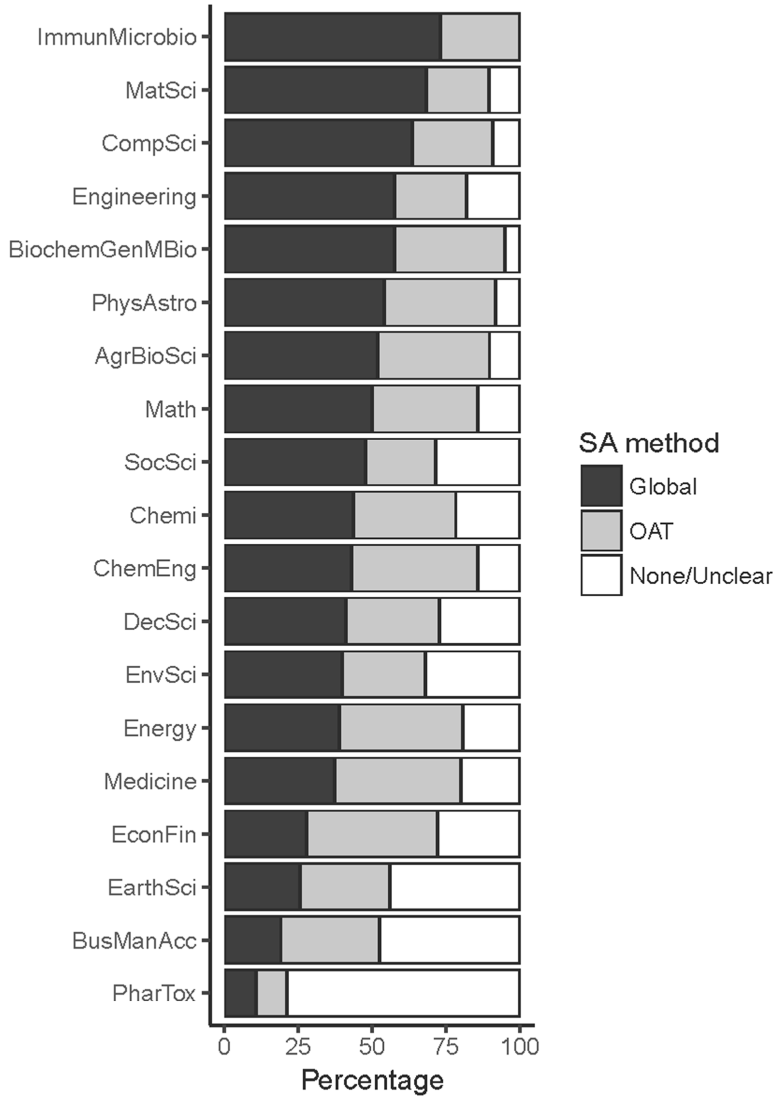


Figure 6: Classification of sensitivity analysis by subject identifier, sorted by proportion of global methods.

As a complement to the manual literature review, we also investigated the prevalence of UA and SA methods based purely on text mining, by identifying at least one known global sensitivity analysis technique (i.e. variance-based, metamodeling, elementary effects etc.), in keeping with the methodology of a previous paper from some of the present authors (Ferretti et al., 2016). Figure 7 shows the results of that paper as extended to 2015 and 2016 (the original analysis stopped at 2014). This is a rougher approach but allows the inclusion of a much larger number of papers. Here it would seem that an even smaller fraction of papers that feature sensitivity analysis adopts a global SA approach.

At least three reasons explain the difference with the results in the present paper. First, as has been well-established here, “sensitivity analysis” is often also used to indicate uncertainty analysis, so that the upper curve in Figure 7 shows a mixture of UA and SA, as well as an inevitable share of papers not pertaining to mathematical modelling. Secondly, the estimation of the number of global SA papers is likely an underestimate because papers may apply simpler global methods, e.g. a scatterplot-based analysis, but not necessarily refer to the articles or techniques listed. Finally, in the manual literature review we focus only on highly cited papers, which should (ideally) be of a higher standard than the average in a given field.

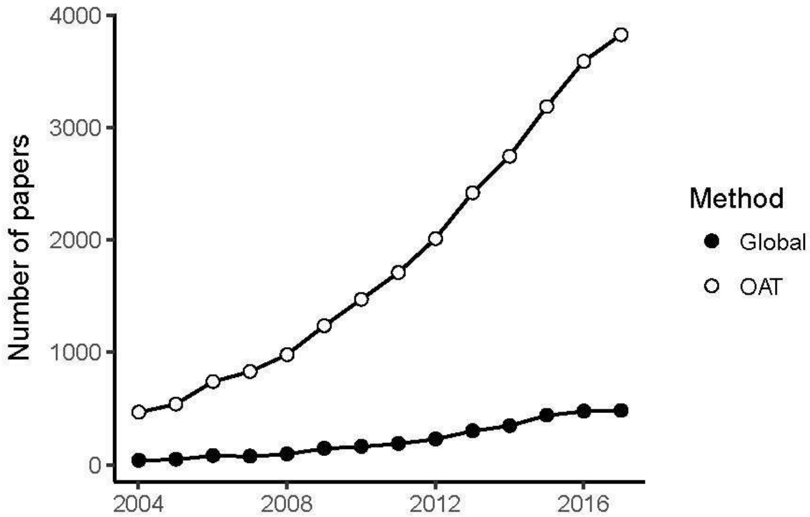


Figure 7: Results from Ferretti et al., extended to 2016 (present paper)

4.4 Method and model focus

Table 1 shows that most papers are unsurprisingly focused on the application, i.e. on the model at hand, and not on the methods. Of the total of 280 papers, 35 were methodological, i.e. having SA/UA methods as their subject. Of these, 24 advocate the use of global methods. On the one hand, this is encouraging because it shows that global methods are being promoted. On the other hand, a small but significant fraction of methodological papers are still advising statistically-incorrect OAT methods.

We note among the method papers a marked preference for variance-based measures of sensitivity – such as the sensitivity indices of which the Pearson correlation ratio discussed previously is a special case. We also see an active line of research in moment-independent methods (Borgonovo, Castaings, & Tarantola, 2012).

4.5 Model linearity

As discussed, if a model is linear, an OAT or derivative based approach is adequate. However, the linearity or nonlinearity of the model is rarely evident, at least from the manuscripts. Table 1 shows the proportions of linear and nonlinear models. Only in 8% of the cases were we able to conclude that the model was definitely linear, whereas over half of papers included clearly nonlinear models, with the remainder being unclear. This demonstrates that first, researchers tend to work with nonlinear models. Second, in the large majority of cases, global methods are essential to perform a methodologically-sound sensitivity analysis.

5 Discussion

5.1 Reasons for bad practice

The results of this study clearly show that there are serious methodological deficiencies in highly cited papers in most if not all disciplines. Why is this so often the case? We speculate that this is due to at least five reasons, which we outline here.

- First, sensitivity analysis is intrinsically attached to modelling, which itself is not a unified subject. Indeed, modelling typically requires a set of skills learned through experience and hence includes elements of craft as much as of science (Rosen, 1991); as such every discipline goes about modelling following local disciplinary standards and practices (Padilla, Diallo, Lynch, & Gore, 2018). Similarly, sensitivity analysis practice is found in largely isolated pockets attached to each modelling discipline. This fragmentation hinders development of the subject and spreading of good practice, while simultaneously allowing malpractice to survive relatively unchallenged. This issue is discussed in more depth in the following section.
- A second point is that most scientists conflate the meaning of SA and UA. If the meaning of sensitivity analysis is not even understood, it is unsurprising that the quality of sensitivity analysis is sometimes lacking.
- Third, global sensitivity analysis unavoidably requires a good background in statistics to implement and to interpret results. Some researchers simply haven't enough knowledge and training in statistics and consequently, the cost in time and money required to learn and understand the necessary techniques may be considered prohibitive. More generally, researchers may not even be aware that global sensitivity analysis techniques exist. Under these

circumstances, it seems that researchers often revert to the more intuitive OAT approach. Among other things, it offers an ease of interpretation: in moving just one input factor, the change observed in the model output must come from that input alone. Moreover, global methods may be discouraging in that the more factors that are moved, the higher the chance that the model will crash or misbehave. Note that this is precisely the reason why a global SA is a good instrument of model verification: it is unusual to run a global SA without detecting model errors – modellers call this jokingly Lubarsky's Law of Cybernetic Entomology, according to which 'there is always one more bug'.

- Fourth, although mature global sensitivity analysis methods have been around for more than 25 years, this still may not be enough time for established good practice to filter down into the many research fields in which modelling is used. This may be partly due to a lack of comparative examples across a range of fields. Moreover, researchers tend to emulate methods found in highly cited papers (assuming that they are best practice), which as this study has demonstrated, are often methodologically deficient.
- Finally, as noted in (Leamer, 2010), the reluctance to take up these methods may be due to their candour. A proper method, by honestly propagating all of the input uncertainty, may lead to an inconveniently wide distribution of the output of interest. For example, a cost-benefit analysis reporting a distribution encompassing possible large losses as well as large gains may not be what the owner of the problem wishes to hear. This is the same as to say that the volatility of the inference is exposed, and thus is the insufficiency of the evidence. According to (Leamer, 2010), as well as to (Funtowicz & Ravetz, 1990), this situation may induce modellers to 'massage' the uncertainty in the input factors so that the output falls in a more desirable zone. For cases where a considerable asymmetry exists between model developers and users (Jakeman, Letcher, & Norton, 2006) it might be advisable to resort to *sensitivity auditing*, an extension of sensitivity analysis beyond parametric analysis to include an assessment of the entire knowledge- and model-generating process for policy-related cases, (Saltelli, Guimaraes Pereira, van der Sluijs, & Funtowicz, 2013), to assess the credibility of degree of uncertainty attributed to each input factor, and to make sure that the uncertainty has been neither inflated nor deflated to achieve a desired end. Inflation and deflation of uncertainty are quite common in e.g. regulatory controversies; typically, the 'regulated' tend to inflate uncertainty so as to deter regulation, while the opposite is the case for regulators (Michaels, 2008). Sensitivity

auditing's seven point checklist is recommended by the European Commission guidelines for impact assessment (European Commission, 2009), p.393.

5.2 Isolated communities

The scattered state of sensitivity analysis practice merits some further discussion. If modelling is a non-standardised discipline (Padilla et al., 2018), the same holds a fortiori for uncertainty and sensitivity analysis, hence the difficulty for good practices to establish themselves. Researchers from different fields have difficulties to communicate with one another in a transversal topic, such as SA, that is practised across a wide range of scientific and modelling disciplines) .

Robert Rosen, a system ecologist, tackles the specificities of modelling in the scientific method in his work 'Life Itself'(Rosen, 1991). Here he suggests that when a model is built to represent a natural system, we should look at the play of causality. The argument is that the natural system is kept together – Rosen uses the word 'entailed' - by *material*, *efficient* and *final* causality. In contrast, the formal system, i.e. the model, is only internally entailed by *formal* causality. Rosen uses here the four causality categories of Aristotle, on which we will not dwell here, to highlight that no arrow of causality flows from the natural system to the formal one. In other words, the act of encoding (Figure 8) is not driven by causality, which would fix the model specification, but is driven by the needs and the craft of the modeller. The implication is that different modelling teams, given the same data, can produce altogether different models and inference (Refsgaard, van der Sluijs, Brown, & van der Keur, 2006).

Thus, the success of the modelling operation is judged by the usefulness – or otherwise - of the insights made possible by the operation of decoding, which is another way of saying that all models are wrong but some are useful – according to an aphorism attributed to George Box.

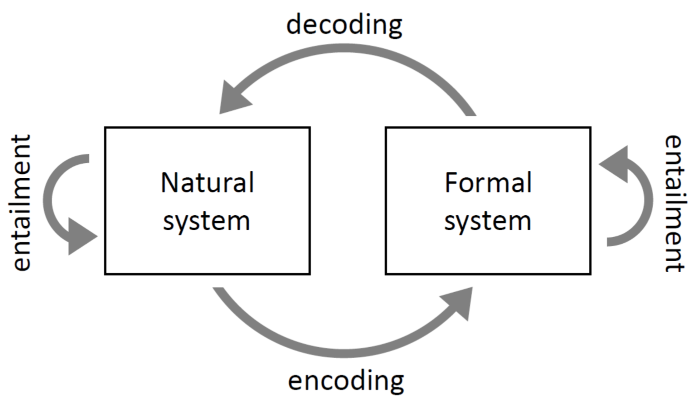


Figure 8: The modelling relation following Rosen (1991). For a discussion see (Saltelli et al., 2008)**.**

Models thus depend crucially upon craftmanship of the modellers. This, together with the diversity of modelling applications, motives, and constraints, explain why modelling never became an independent discipline. In our opinion this contributes to explaining why modelling is so discipline-specific, as noted by (Padilla et al., 2018). The spread in modelling practices and cultures may be one of the reasons why methodologies which are ancillary to modelling, such as uncertainty and sensitivity analysis, are not part of a standardized syllabus being taught across disciplines, and are at times ignored even in communities proficient in modelling, such as for example hydrology (Shin et al., 2013).

Despite the fragmentation of sensitivity and uncertainty analysis, some cross-disciplinary networks exist. One such community might be said to have formed around a series of SAMO conferences (for sensitivity analysis of model output, see http://samo2016.univ-reunion.fr/). SAMO has been held every three years since 1995. This community is active in training and dissemination. However, SAMO by no means captures the full spectrum of practitioners interested in uncertainty and sensitivity analysis. For example, in the United States, SA-related activities are under the heading of 'Verification, Validation and Uncertainty Quantification' (VVUQ), for which a journal of the American Society of Mechanical Engineers is available (http://verification.asmedigitalcollection.asme.org/journal.aspx). Other sensitivity analysis related gatherings include the *Conference on Uncertainty Quantification* organised by the Society for Industrial and Applied Mathematics, the *International Conference on Uncertainty Quantification in Computational Sciences and Engineering* organised by the European Community on Computational Methods in Applied Sciences, and sessions in thematic conferences such as the *Uncertainty in Structural Dynamics* conference organised by Department of Mechanical Engineering of the KU Leuven, or the session on *Advances in Diagnostics, Sensitivity, and Uncertainty Analysis of Earth and Environmental Systems Models* organised annually at the European Geosciences Union conference in Vienna.

Despite these communities, the majority of practitioners remain scattered in isolated pockets, and sensitivity analysis is hence not part of a recognized syllabus. Who or what scientific forum can then decide if a method is a good or a bad practice? To make an example, in (Nearing & Gupta, 2018; Stark & Saltelli, 2018), who can authoritatively discourage modellers from over interpreting the results from multi-model ensembles as if they were a random sample from a distribution? This question remains - for the time being, unanswered. A possible solution to this unsatisfactory state of affairs would be that statistics as a discipline takes responsibility for statistical methods for model validation and verification. This would not make modelling into a discipline but would go a long way toward improving modelling practice. Additionally, most if not all the tools of sensitivity analysis are statistical in nature. This thesis has been suggested in a discussion paper entitled ‘Should statistics rescue mathematical modelling?’ (Saltelli, 2018).

5.3 Parallels with the p-value

The systematic problems observed in sensitivity analysis share similarities with the recent crisis in statistics over the p-value. A paper published in 2005 (Ioannidis, 2005) warned about the poor quality of most published research results. The paper was taken up by the media, and the periodical “The Economist” devoted its cover to the issue in 2013 (“How science goes wrong,” 2013), with a full article describing the subtleties of use and misuse of statistics in deciding about the significance of scientific results. The specific subject of concern was the use of the p-value, “the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value” (Wasserstein & Lazar, 2016). The p-value is used as a fundamental tool by researchers to decide if a given result is just the result of chance or indeed an effect worth publishing.

In 2016, the pressure surrounding the statistical community was so high that the American Statistical Association felt the need to intervene with a statement (Wasserstein & Lazar, 2016) to clarify how the test should be used. Useful reading on the topic are (Colquhoun, 2014; Gigerenzer & Marewski, 2014; Stark & Saltelli, 2018). These articles show a complex mix of causes – from poor training to bad incentives – which result in the generalized failure in the use of the p-value, evidenced by attempts to repeat published results, see e.g. (Shanks et al., 2015).

The problem is seen as a combination of confirmation bias - authors looking for the effect they presume will be there (confirmation bias), or authors desperate to publish a positive result (publish or perish), of p-hacking – changing the setup of the study or the composition of the sample till an effect emerges, and HARKing, formulating the research Hypothesis After the Results are Known, (Kerr, 1998). The latter involves repeatedly running comparison tests between different combinations of variables until a “significant” result is found, which violates the conditions of applicability of the P-test.

Overall, it is clear that the consequences of bad statistics can be dramatic – for example when wrong cures for cancer are identified at the pre-clinical stage of research, and are then passed on to the clinical trial phase (Begley & Ellis, 2012). Similarly, it is not difficult to imagine the consequences of a wrong or missing uncertainty and sensitivity analyses given the pervasive role of models. In risk analysis this can lead to ignoring dangerous operating conditions for a facility, in decision analysis, this can lead to wrong investments or policies. A simple sensitivity analysis run on the formula used for the pricing of the complex derivative products at the root of the sub-prime mortgage crisis would have revealed the fragility of the formula (Salmon, 2009; Wilmott & Orrell, 2017). Whether the 'quants' – the experts in charge of these mathematical constructs – wanted to know this fragility is of course another story. Finally, a missing uncertainty analysis allows audacious risk or cost-benefit analysis to be run over centennial time scales while a proper UA would show clearly that the uncertainties are too big to conclude anything. An example discussed in (Saltelli et al., 2015) was the computing the increased crime rate due to increased temperature at the year 2100.

5.4 Recommendations for best practice

It is outside of the scope of this paper to give a detailed guide to sensitivity analysis—for thorough references, readers are referred to (Saltelli et al., 2008) or (Ghanem et al., 2017). Nevertheless, and although considerable differences exist in the use of sensitivity analysis among disciplines, all fields would benefit from the adoption of good practices. Our personal list of preferences, which agrees with the methodological papers seen in this review, would include the following recommendations:

- Both uncertainty and sensitivity analysis should be based on a *global* exploration of the space of input factors, be it using an experimental design, Monte Carlo or other ad-hoc designs. The discussion in this paper has demonstrated that local/OAT methods do not adequately represent models with nonlinearities.
- With some exceptions, it is advisable to perform both uncertainty and sensitivity analysis. Once an analyst has performed an uncertainty analysis and is informed of the robustness of the inference, it would appear natural to ascertain where volatility/uncertainty is coming from. At the other extreme, a sensitivity analysis without uncertainty analysis is usually illogical – the relative importance of a factor on the model output has a different relevance depending on whether the output has a small or large variance. However, there are cases – for instance, studies to identify the dominant effects on the output for a subsequent model reduction or calibration analysis – where the analyst may be satisfied with a pure SA.
- Sensitivity and uncertainty analysis should be focused on a question. Most models have many outputs, and these outputs can be used to answer a range of different questions. The relationship (sensitivity) between the input factors and each different model output can be very different. For this reason, it is essential to focus the sensitivity analysis on the question addressed by the model rather than more generally on the model.

- When sensitivity analysis is performed, it should allow the relative importance of input factors and combinations of factors, to be assessed, either visually (scatterplots) or quantitatively (regression coefficients, sensitivity measures or other).
- Sensitivity and uncertainty analysis are themselves uncertain, because there is considerable uncertainty in quantifying the uncertainty in input factors, and modellers should be frank about how they arrived at the supposed uncertainties (Saltelli et al., 2013). This should be kept in mind and efforts made to capture the uncertainty of input assumptions as accurately as possible.
- Even an apparently perfect uncertainty and sensitivity analysis is no assurance against error. As noted by (Pilkey & Pilkey-Jarvis, 2009) "It is important to recognize that the sensitivity of the parameter in the equation is what is being determined, not the sensitivity of the parameter in nature. […] If the model is wrong or if it is a poor representation of reality, determining the sensitivity of an individual parameter in the model is a meaningless pursuit."

As regards what method should be used, our preference is for methods which are exploratory, model-independent, able to capture interactions and to treat a group of factors. A carefully performed uncertainty analysis, followed by sensitivity analysis, is an important ingredient of the quality assurance of a model as well as a necessary condition for any model-based analysis or inference.

6 Conclusions

The main message of the present work is that a carefully performed sensitivity analysis is an important ingredient of the quality assurance of a model as well as a necessary condition for any model-based analysis or inference. However, such analyses are not common enough and often inaccurate, indicating that action is urgent on the front of quality assurance procedures for mathematical models. In particular, a significant fraction of papers investigated use sensitivity analysis approaches which fail elementary considerations of experimental design and do not properly explore the space of the input factors, with the result that uncertainty is generally underestimated and sensitivity is wrongly estimated. Up to 65% of the reviewed (highly cited) papers are based on inadequate methods (i.e. varying one input factor at a time), although even in the most generous interpretation, where all models with unclear linearity are assumed linear, still over 20% of papers contain inadequate methodology. Further, a significant number of papers confuse sensitivity and uncertainty analysis, which is likely to exacerbate the problem with spreading good practice.

The fact that these figures concern highly cited papers has two implications: first, if we assume that highly cited papers represent the upper end of methodological rigour in a given field, then the overall problem may be even worse. Second, these are some of the most visible papers in their field, and are used as guides for best practice. Therefore, they can promote continued deficient methodology.

In our opinion, the problem with sensitivity analysis is partly attributable to the fact that mathematical modelling is not a discipline in its own right, and every branch of science and technology approaches modelling following its own culture and practice. Uncertainty and sensitivity analyses are likewise orphans of a disciplinary home. One can also note that signals of distress as to the quality of mathematical modelling are heard from different disciplines: from economics (Reinert, 2000; Romer, 2015) to natural sciences (Oreskes, 2000; Oreskes, Shrader-Frechette, & Belitz, 1994; Pilkey & Pilkey-Jarvis, 2009). The situation has worrying analogies with what we have witnessed in data analysis, where misuse of the p-value (Colquhoun, 2014) has been singled out as one of the reasons of the present reproducibility crisis affecting science (Ioannidis, 2005; Saltelli & Funtowicz, 2017). The importance of this analogy is in the warning it sounds for the credibility of science if such pervasive weaknesses in methodology are not addressed. The need to heed this warning in the case of sensitivity and uncertainty analysis is becoming increasingly urgent.

References

Allaire, G., Jouve, F., & Toader, A. M. (2004). Structural optimization using sensitivity analysis and a level-set method. *Journal of Computational Physics*, *194*(1), 363–393. https://doi.org/10.1016/j.jcp.2003.09.032

Becker, W., Oakley, J. E., Surace, C., Gili, P., Rowson, J., & Worden, K. (2012). Bayesian sensitivity analysis of a nonlinear finite element model. *Mechanical Systems and Signal Processing*, *32*, 18–31. https://doi.org/10.1016/j.ymssp.2012.03.009

Becker, W., Rowson, J., Oakley, J. E., Yoxall, A., Manson, G., & Worden, K. (2011). Bayesian sensitivity analysis of a model of the aortic valve. *Journal of Biomechanics*, *44*(8), 1499–1506. https://doi.org/10.1016/j.jbiomech.2011.03.008

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. Retrieved from http://dx.doi.org/10.1038/483531a

Borgonovo, E., Castaings, W., & Tarantola, S. (2012). Model emulation and moment-independent sensitivity analysis: An application to environmental modelling, *34*, 105–115. https://doi.org/10.1016/j.envsoft.2011.06.006

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: design, innovation, and* *discovery* (2nd ed). Hoboken, N.J: Wiley-Interscience.

Cacuci, D. G. (2005). *Sensitivity and Uncertainty Analysis Theory*. *Analysis*. https://doi.org/10.1201/9780203911396.ch10

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society Open Science*, *1*, 140216. https://doi.org/10.1098/rsos.140216

Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, *85*(7), 1283–1305. https://doi.org/10.1080/00949655.2014.945932

Davis, P. M., & Cochran, A. (2015). Cited Half-Life of the Journal Literature. *ArXiv.Org*, 1–15.

Eisenhower, B., O'Neill, Z., Narayanan, S., Fonoberov, V. A., & Mezić, I. (2012). A methodology for meta-model based optimization in building energy models. *Energy and Buildings*, *47*, 292–301. https://doi.org/10.1016/j.enbuild.2011.12.001

European Commission. (2009). *European Commission IMPACT ASSESSMENT GUIDELINES*. Retrieved from http://ec.europa.eu/smart-regulation/impact/commission_guidelines/docs/iag_2009_en.pdf

Ferretti, F., Saltelli, A., & Tarantola, S. (2016). Trends in sensitivity analysis practice in the last decade. *Science of The Total Environment*, *568*, 666–670. https://doi.org/10.1016/j.scitotenv.2016.02.133

Funtowicz, S., & Ravetz, J. R. (1990). *Uncertainty and Quality in Science for Policy*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-0621-1_3

Ghanem, R., Higdon, D., & Owhadi, H. (2017). *Handbook of uncertainty quantification*. (Springer, Ed.).

Gigerenzer, G., & Marewski, J. N. (2014). Surrogate Science: The Idol of a Universal Method for Scientific Inference. *Journal of Management*, (September), 0149206314547522-. https://doi.org/10.1177/0149206314547522

How science goes wrong. (2013, October). *The Economist*. Retrieved from https://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8). https://doi.org/10.1371/journal.pmed.0020124

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models,. *Environmental Modelling & Software*, *21*(5), 602–614.

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Leamer, E. E. (1985). Sensitivity Analyses Would Help. *The American Economic Review*, *75*(3), 308–313.

Leamer, E. E. (2010). Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives*, *24*(2), 31–46. https://doi.org/10.1257/jep.24.2.31

Michaels, D. (2008). *Doubt is Their Product: How Industry's Assault on Science Threatens Your Health*. Oxford University Press. Retrieved from https://books.google.es/books?id=J0P3IdSYO_MC

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, *33*(2), 161. https://doi.org/10.2307/1269043

Nearing, G. S., & Gupta, H. V. (2018). Ensembles vs. information theory: supporting science under uncertainty. *Frontiers of Earth Science*, 1–8. https://doi.org/10.1007/s11707-018-0709-9

Norton, J. P. (2015). An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software*, *69*(C), 166–174. https://doi.org/10.1016/j.envsoft.2015.03.020

Office of Management and Budget. (2006). *Proposed Risk Assessment Bulletin*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/omb/inforeg/proposed_risk_assessment_bulletin_010906.pdf

Oreskes, N. (2000). Why Predict? Historical Perspectives on Prediction in Earth Science. In *Prediction: Science, Decision Making, and the Future of Nature* (pp. 23–40).

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, *263*(5147). Retrieved from http://science.sciencemag.org/content/263/5147/641

Padilla, J. J., Diallo, S. Y., Lynch, C. J., & Gore, R. (2018). Observations on the practice and profession of modeling and simulation: A survey approach. *SIMULATION*, *94*(6), 493–506. https://doi.org/10.1177/0037549717737159

Pearson, K. (1905). *On the general theory of skew correlation and non-linear regression*. London: Dulau and co.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/J.ENVSOFT.2016.02.008

Pilkey, O. H., & Pilkey-Jarvis, L. (2009). *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future*. Columbia University Press.

Razavi, S., Sheikholeslami, R., Gupta, H. V., & Haghnegahdar, A. (2019). VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. *Environmental Modelling & Software*, *112*, 95–107. https://doi.org/10.1016/J.ENVSOFT.2018.10.005

Refsgaard, J. C., van der Sluijs, J. P., Brown, J., & van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, *29*(11), 1586–1597. Retrieved from http://www.jstor.org/stable/2225764

Reinert, E. S. (2000). Full circle: economics from scholasticism through innovation and back into mathematical scholasticism. *Journal of Economic Studies*, *27*(4/5), 364–376. https://doi.org/10.1108/01443580010341862

Romer, P. (2015). Mathiness in the Theory of Economic Growth. *American Economic Review*, *105*(5), 89–93. https://doi.org/10.1257/aer.p20151066

Rosen, R. (1991). *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*.

Columbia University Press. Retrieved from https://books.google.es/books?id=DR8L4snDnkIC

Salmon, F. (2009, February). Recipe for Disaster: The Formula That Killed Wall Street. *Wired* . Retrieved from https://www.wired.com/2009/02/wp-quant/

Saltelli, A. (2002). Sensitivity analysis for importance assessment. In *Risk Analysis* (Vol. 22, pp. 579–590). https://doi.org/10.1111/0272-4332.00040

Saltelli, A. (2018). Should statistics rescue mathematical modelling? *ArXiv*, *arXiv:1712*(06457).

Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, *25*(12), 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Saltelli, A., & Funtowicz, S. (2017). What is science's crisis really about? *Futures*, *91*, 5–11. https://doi.org/10.1016/j.futures.2017.05.010

Saltelli, A., Guimaraes Pereira, Â., van der Sluijs, J. P. ., & Funtowicz, S. (2013). What do I make of your latinorumc Sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy*, *9*(2/3/4), 213–234. https://doi.org/10.1504/IJFIP.2013.058610

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., … Tarantola, S. (2008). *Global sensitivity analysis : the primer*. John Wiley.

Saltelli, A., Stark, P. B., Becker, W., & Stano, P. (2015). Climate Models as Economic Guides - Scientific Challenge of Quixotic Quest? *Issues in Science and Technology*, *31*(3), 1–8. Retrieved from https://www.stat.berkeley.edu/~aldous/157/Papers/saltelliEtal15

Saltelli, A., & Tarantola, S. (2002). On the Relative Importance of Input Factors in Mathematical Models. *Journal of the American Statistical Association*, *97*(459), 702–709. https://doi.org/10.1198/016214502388618447

Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., … Puhlmann, L. M. C. (2015). Romance, Risk, and Replication: Can Consumer Choices and Risk-Taking Be Primed by Mating Motives? *Journal of Experimental Psychology: General*, *144*(6), 142–158. https://doi.org/10.1037/xge0000116

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. *Journal of Hydrology*, *503*, 135–152. https://doi.org/10.1016/J.JHYDROL.2013.08.047

Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models,. *Mathematical Modelling and Computational Experiment (Translated from Russian: I.M. Sobol', Sensitivity Estimates for Nonlinear Mathematical Models, Matematicheskoe Modelirovanie 2 (1990) 112–118).*, 407–414.

Sobol', I. M., & Kucherenko, S. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, *79*(10), 3009–3017. https://doi.org/10.1016/j.matcom.2009.01.023

Stark, P. B., & Saltelli, A. (2018, July). Cargo-cult statistics and scientific crisis. *Significance*. Retrieved from https://www.significancemagazine.com/2-uncategorised/593-cargo-cult-statistics-and-scientific-crisis

U.S. Environmental Protection Agency (EPA). (2009). *Guidance on the Development, Evaluation, and Application of Environmental Models*. Retrieved from http://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1003E4R.PDF

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement onp-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wilmott, P., & Orrell, D. (2017). *The Money Formula*. Wiley & Sons.

Annex

Table 1 shows the results of the reviews in a condensed form. The meaning of the headings is given in Section 3.

Category	METHOD					MODEL LINEARITY			PAPER FOCUS		Total
	Global SA	OAT SA	Global UA	OAT UA	Other/Unclear	Linear	Nonlinear	Unclear	Method	Model	reviewed
AgrBioSci	15	11	6	0	6	1	22	4	3	24	27
BiochemGenMBio	23	15	6	1	7	2	19	15	0	36	36
BusManAcc	4	7	5	5	1	1	18	2	3	18	21
Chemi	10	8	2	0	5	0	17	5	1	21	22
ChemEng	12	12	4	0	5	0	16	12	1	27	28
CompSci	21	9	1	1	2	8	16	6	11	22	33
DecSci	9	7	3	4	0	2	20	1	7	15	22
EarthSci	11	13	4	1	17	5	13	24	2	41	43
EconFin	5	8	6	3	0	1	16	1	0	18	18
Energy	14	15	3	4	2	3	17	16	0	36	36
Engineering	38	16	5	5	5	3	51	11	3	62	65
EnvSci	31	22	14	4	16	6	44	24	11	67	78
ImmunMicrobio	19	7	3	0	5	2	6	13	0	21	21
Math	21	15	3	2	6	4	24	13	11	29	40
MatSci	13	4	1	1	0	0	16	2	0	18	18
Medicine	26	30	25	4	13	2	24	37	2	62	64
PharTox	2	2	9	1	3	1	11	5	1	18	19
PhysAstro	13	9	4	0	0	1	20	2	2	21	23
SocSci	10	5	0	4	2	1	14	5	6	15	21

Table 2: Summary of results by subject identifier.

Highlights of “Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices”

- A systematic review of scientific papers mentioning sensitivity analysis has been performed
- The analysis addresses the use of SA in the context of mathematical modelling, focusing on highly cited works.
- In total 324 papers were reviewed. After cleaning the sample280 papers were retained for the analysis.
- Many highly-cited papers (42% in the present analysis) present a SA of poor quality.
- The results, while discipline-dependent, point to a worrying lack of standards and recognized good practices.
- Some guidelines for proper use of the methods are suggested.