# *The future of sensitivity analysis: an essential discipline for systems modeling and policy support*

# The Future of Sensitivity Analysis: An Essential Discipline for Systems Modeling and Policy Support

Saman Razavi[1*], Anthony Jakeman[2], Andrea Saltelli[3], Clémentine Prieur[4], Bertrand Iooss[5], Emanuele Borgonovo[6], Elmar Plischke[7], Samuele Lo Piano[8], Takuya Iwanaga[2,] William Becker[9], Stefano Tarantola[10], Joseph H.A. Guillaume[2], John Jakeman[11], Hoshin Gupta[12], Nicola Melillo[13], Giovanni Rabitti[14], Vincent Chabridon[5], Qingyun Duan[15], Xifu Sun[2], Stefán Smith[16], Razi Sheikholeslami[1,17], Nasim Hosseini[1], Masoud Asadzadeh[18], Arnald Puy[19,20], Sergei Kucherenko[21], Holger R. Maier[22]

1. School of Environment and Sustainability, Global Institute for Water Security, University of Saskatchewan, Canada (saman.razavi@usask.ca)
2. Fenner School of Environment and Society, Institute for Water Futures, The Australian National University, Australia
3. Open Evidence Research, Universitat Oberta de Catalunya (UOC), Barcelona
4. Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LJK, 38000 Grenoble, France *Institute of Engineering Univ. Grenoble Alpes
5. EDF R&D, Département PRISME, Chatou, France and SINCLAIR AI Laboratory, Saclay, France
6. Department of Decision Sciences, Bocconi University, Milano 20136, Italy
7. Institute of Disposal Research, TU Clausthal, Clausthal-Zellerfeld, Germany
8. School of the Built Environment, University of Reading, United Kingdom
9. Freelance consultant and researcher, Ispra, Italy
10. European Commission, Joint Research Centre (JRC), Ispra, Italy
11. Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque
12. Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, Arizona, 85721, USA
13. Department of Electrical, Computer and Biomedical Engineering, Università degli Studi di Pavia, Pavia, Italy
14. Department of Actuarial Mathematics and Statistics, Maxwell Institute for Mathematical Sciences, Heriot-Watt University, United Kingdom
15. College of Hydrology and Water resources, Hohai University, Nanjing 210098, China
16. School of the Built Environment, University of Reading, United Kingdom
17. Environmental Change Institute, School of Geography and the Environment, University of Oxford, South Parks Road, OX13QY Oxford, United Kingdom
18. Department of Civil Engineering, University of Manitoba, Winnipeg, Manitoba, Canada
19. Department of Ecology and Evolutionary Biology, M31 Guyot Hall, Princeton University, New Jersey 08544, USA
20. Center for the Study of the Sciences and the Humanities, Parkveien 9, University of Bergen, 5020 Bergen, Norway
21. Imperial College London, UK
22. School of Civil, Environmental and Mining Engineering, The University of Adelaide, Australia

## Abstract

Sensitivity analysis (SA) is en route to becoming an integral part of mathematical modeling. The tremendous potential benefits of SA are, however, yet to be fully realized, both for advancing mechanistic and data-driven modeling of human and natural systems, and in support of decision making. In this perspective paper, a multidisciplinary group of researchers and practitioners revisit the current status of SA, and outline research challenges in regard to both theoretical frameworks and their applications to solve real-world problems. Six areas are discussed that warrant further attention, including (1) structuring and standardizing SA as a discipline, (2) realizing the untapped potential of SA for systems modeling, (3) addressing the computational burden of SA, (4) progressing SA in the context of machine learning, (5) clarifying the relationship and role of SA to uncertainty quantification, and (6) evolving the use of SA in support of decision making. An outlook for the future of SA is provided that underlines how SA must underpin a wide variety of activities to better serve science and society.

## Keywords

Sensitivity Analysis, Mathematical Modeling, Machine Learning, Uncertainty Quantification, Decision Making, Model Validation and Verification, Model Robustness, Policy Support

## Highlights

- Sensitivity analysis (SA) should be promoted as an independent discipline
- Several grand challenges hinder full realization of the benefits of SA
- The potential of SA for systems modeling & machine learning is untapped
- New prospects exist for SA to support uncertainty quantification & decision making
- Coordination rather than consensus is key to cross-fertilize new ideas

## Table of Contents

# 1. Introduction

## 1.1. The whats and whys of Sensitivity Analysis

Sensitivity analysis (SA), in the most general sense, is the study of how the 'outputs' of a 'system' are related to, and are influenced by, its 'inputs'. In many applications, the system in question involves a single or a set of mathematical models, encoded using computer software, that simulates the functioning of a real-world system of interest. Such mathematical models can be data-driven (also called statistical), directly mapping inputs to outputs (Engelbrecht et al., 1995; Rodriguez et al., 2010), or mechanistic (also called process-based), solving a set of differential or other mathematical equations governing the (possibly) spatio-temporal behaviors of the underlying processes (Maxwell and Miller, 2005; Haghnegahdar et al., 2018). Inputs of interest, commonly referred to as 'factors' in SA, may include model parameters, forcing variables, boundary and initial conditions, choices of model structural configurations, assumptions and constraints. Outputs may include any functions of model responses, including those that may vary over a spatio-temporal domain, objective functions such as a production or cost function in cost-benefit analysis, or an error function in model calibration.

Why is SA useful? In short, it addresses several fundamental overarching purposes of systems analysis and modeling: (a) *scientific discovery* to explore causalities and how different processes, hypotheses, parameters, scales and their combinations and interactions affect a system (e.g., Gupta and Razavi, 2018); (b) *dimensionality reduction* to identify uninfluential factors in a system that may be redundant and fixed or removed in subsequent analyses (e.g., Sobol' et al., 2007); (c) *data worth assessment* to identify processes, parameters and scales that dominantly control a system, for which new data acquisition reduces targeted uncertainty the most (e.g., Guillaume et al., 2019; Partington et al., 2020); and (d) *decision support* to quantify the sensitivity of an expected outcome to different decision options, constraints, assumptions and/or uncertainties (e.g., Tarantola et al., 2002). SA is now considered a requirement for good modeling practice as indicated by some existing guidelines (European Commission, 2015; Saltelli et al., 2020). In general, and regardless of any specific purpose, SA aims to exploit the 'sparsity of factors' principle, a heuristic stating that, very often, only a small subset of factors in a system have a significant influence on a specific system output (Box and Meyer, 1986).

## 1.2. Why this position paper

SA is a relatively new area of research. It has roots in 'design of experiments' (DOE) which is a broad family of statistical methods conceived in the early 20th century for designing efficient experiments to acquire representative information about the existence of effects of one or multiple variables on another variable in a system (Fisher, 1953). DOE primarily worked in the context of costly, noise-prone lab or field environments. The field of SA started to materialize in the 1970s and 80s with the beginning of the widespread availability of computers for mathematical modeling (e.g., Cukier et al, 1973), and the extension of DOE to the 'design and analysis of computer experiments' (DACE), which are typically noise-free or deterministic in the sense that replicating a computer experiment with the same inputs results in identical model responses (Sacks et al., 1989). More broadly, the notion of sensitivity has historically, but informally, been a building block in various types of study, particularly in decision making

where what-if scenarios or policy effectiveness are assessed by changing one or multiple factors at a time (Tarantola et al., 2000).

What is the status quo in SA? We believe that SA is en route to becoming a mature and independent, but interdisciplinary and enabling, field of science. Tremendous advances in both theory and application of SA have been accomplished, as documented in the recent reviews by Norton (2015), Iooss and Lemaître (2015), Wei et al. (2015), Razavi and Gupta (2015), Borgonovo and Plischke (2016), Pianosi et al. (2016), Borgonovo et al. (2017), Borgonovo (2017), Ghanem et al. (2017), Gupta and Razavi (2017), Saltelli et al. (2019), and Da Veiga et al. (under review). Particularly in recent years, research and practice in SA have gained significant momentum, with many researchers from a variety of backgrounds contributing to a variety of theoretical frameworks, based on the type of applications in their respective disciplines.

Despite these advances, realization of the benefits and true potential of SA across the sciences has been hampered by several challenges. Amongst others, the most pressing challenge is that SA is still a paradigm defined largely by method rather than purpose. Various methods have been developed, rooted in different philosophies towards SA (Razavi and Gupta, 2015). But often, the purpose has been defined on the basis of how a given method works and its capabilities, as well as its authors' disciplinary research focus. Narrow views, lack of communication among scientists across disciplines, and ignoring uncertainties in models can conceal the benefits of SA to researchers and practitioners, leading to an underuse of SA in many branches of modeling and consequently in support of model-based policy making (Saltelli et al., 2019).

In this perspective paper, we synthesize key aspects of the state of the art and, by taking a forward-looking approach, outline some grand challenges facing the new frontiers and opportunities for SA. We draw from the multidisciplinary views and expertise of the authorship team, which includes natural scientists, engineers, decision scientists, computer scientists, systems analysts and mathematicians, and provide opinions on the possible and desirable future evolutions of SA science and practice. Our overarching goal is to contribute to the establishment of SA as a distinct and essential interdisciplinary, enabling science. We believe that SA science must formally become an integral part of systems analysis in general, and mathematical modeling in particular, to unleash the capabilities of SA for addressing emerging problems in engineering, technology, the natural and social sciences, and human-natural systems. The following sections are structured such that one could directly read a section of interest without attention to the other sections.

## 2. An Overview of the State of the Art

To many, Sensitivity Analysis (SA) simply means a process in which one or multiple factors in a problem are changed to evaluate their effects on some outcome or quantity of interest. Such a process has a long history of application, perhaps in all areas of science. Examples include: assessment of the effectiveness of a decision option in a policy-making problem; the impact of a problem constraint on the optimality of a cost or benefit function via shadow prices; or the role and function of a model parameter in generating a model output. Such analyses are generally referred to as 'Local Sensitivity Analysis (LSA)' because the sensitivity of the

problem is assessed *only* around a 'nominal point' in the problem space. LSA is simple, intuitive and appropriate in very specific circumstances. It has, however, been commonly used much more broadly, in circumstances where it has been criticized for providing only a localized view of the problem space, especially when used in the context of investigating parameter importance in mathematical modeling (Saltelli and Annoni, 2010; Saltelli et al., 2019).

The modern era of SA has focused on a notion that is commonly referred to as 'Global Sensitivity Analysis (GSA)' (Saltelli et al., 2000), as it attempts to provide a 'global' representation of how the different factors work and interact across the full problem space to influence some function of the system output - see **Figure 1**. In this section, we briefly summarise four categories of GSA: derivative-based; distribution-based; variogram-based; and regression-based. We also provide overviews of SA when enabled with response surface methods, progress in SA with correlated factors and software tools available for SA.

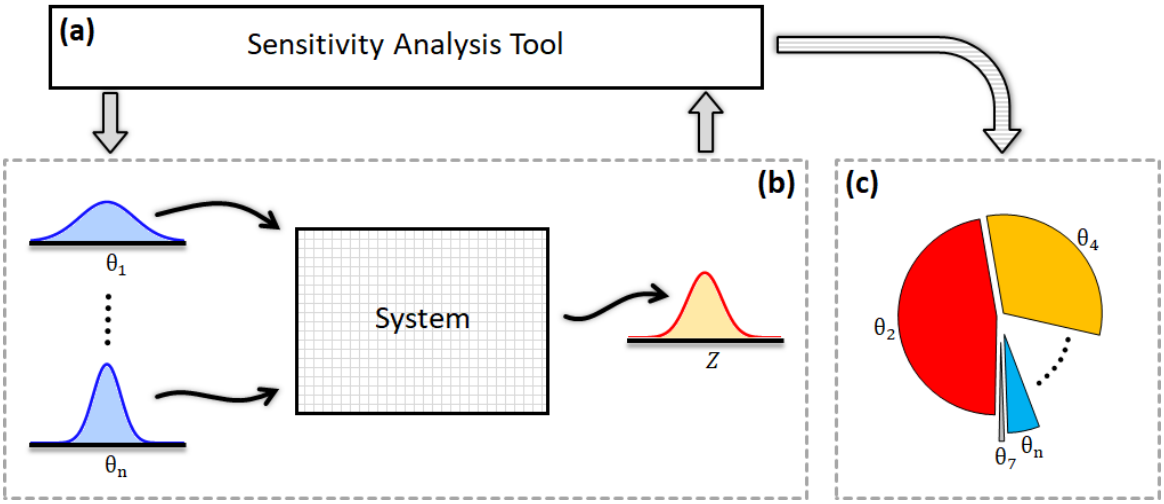


**Figure 1.** A high-level workflow of typical methods for SA. Box (a) represents an SA tool that generates inputs to the system of interest, $\{\theta_1, ..., \theta_n\}$ that can be continuous or discrete variables, or triggers that activate different modeling choices, and receives outputs Z. Box (b) represents the system of interest. Box (c) represents a classic outcome of SA where the contribution of each input's variability on the variability of (some function of) the output is quantified. The outcome of SA can also include information about interactions between inputs, statistical variability of the results, etc., which are not shown here.

## 2.1. Derivative-based approach

Derivative-based methods are a natural and intuitive extension of LSA, where measures of local sensitivity are computed at many 'base points' across factor space and somehow averaged to provide a global assessment of sensitivity. Such measures are said to be 'derivative-based' as they either analytically compute derivatives or numerically quantify the change in output when factors of interest (continuous or discrete) are perturbed around a point. Perturbations typically occur one at a time and by some specific 'perturbation size' (Campolongo et al., 2011). The different derivative-based methods differ in the ways that they choose the base points, the perturbation size, and the distributional properties of the sampled derivatives (e.g., first or second moment) to provide an average global measure of sensitivity

(see e.g., Morris, 1991; Campolongo et al., 2007; Sobol' and Kucherenko, 2009; Campolongo et al., 2011, Lamboni et al., 2013; Rakovec et al., 2014). The outcome of these methods is 'sensitive' to these choices, among which the sensitivity to perturbation size is generally overlooked even though it can be profound (Shin et al., 2013; Haghnegahdar and Razavi, 2017).

### 2.2. Distribution-based approach

Distribution-based methods adopt a different philosophy that bases the analysis on the distributional properties of the output itself, and attempts to quantify how the different inputs contribute to forming those properties. The most common distribution-based method is based on the analysis of output variance, decomposing that variance into portions attributed to individuals or groups of inputs (Sobol', 1993; Owen, 1994; Homma and Saltelli, 1996). Such a 'variance-based' SA was first conceived in the context of non-linear dependence as far back as 1905 (Pearson, 1905), and later in terms of a Fourier analysis in the 70s (Cukier et al, 1978). The full variance-based SA framework was laid down by Ilya Sobol' in 1993 (Sobol', 1993), then linked to the derivative-based SA via Poincaré inequalities by Sobol' and Kucherenko (2009; see also Roustant et al., 2017).

Some distribution-based methods go beyond variance and investigate how higher-order moments of the output depend on the inputs. For example, the method of Dell'Oca et al. (2017) particularly focuses on skewness and kurtosis. Some other distribution-based methods are, however, 'moment-independent' in that they measure the difference between the unconditional distribution of the output and its conditional counterparts when one or more inputs are fixed. For example, the method of Borgonovo (2007) measures this difference via the Borgonovo index, while the methods of Krzykacz-Hausmann (2001) and Pianosi and Wagener (2015, PAWN) use the mutual information and Kolmogorov-Smirnov test, respectively. Another example is the commonly called 'Regional Sensitivity Analysis' (RSA), which, rather than fixing inputs, defines conditional distributions based on thresholds for the model response (Spear et al., 1994; Hamby, 1994).

### 2.3. Variogram-based approach

More recently, a third category has emerged based on the theory of variograms that bridges derivative and distribution-based methods (Razavi and Gupta, 2016a; 2016b; Sheikholeslami and Razavi, 2020). The 'variogram-based' approach recognizes that model outputs are *not* always randomly distributed and they possess, as do their partial derivatives, a spatially-ordered (covariance) structure in the input space. Anisotropic variograms can characterize this structure by quantifying the variance of change in the output as a function of perturbation size in individual inputs. Variogram-based sensitivity measures can be considered more comprehensive than other approaches in the sense that they integrate global sensitivity information across a range of perturbation scales. Derivative-based and variance-based sensitivity measures are also produced as a side product of calculating 'variogram effects'. The efficiency and applicability of the variogram-based approach are demonstrated in Razavi et al. (2019), Becker (2020) and Puy et al. (2020a).

### 2.4. Regression-based approach

Regression-based SA has a long history, traditionally referring to methods that infer sensitivity information from coefficients of a typically linear regression model fitted to a sample of model response surface points (Kleijnen, 1995). Those early methods, from a GSA point of view, have been criticized for their heavy reliance on the prior assumption regarding model response form (e.g., linear or polynomial equation), and the fact that if the quality of fit is poor, the sensitivity estimates are not reliable (Razavi and Gupta, 2015). From an LSA point of view, however, they have proven useful for dimensionality reduction via orthogonal decompositions from parameter samples (Kambhatla and Leen, 1997) or locally approximated sensitivity matrices (Tonkin and Doherty, 2005). Also, such methods when using quadratic regression allow characterization of parameter interactions in the inverse problem (e.g., Shin et al., 2015).

More recently, regression-based SA has witnessed a new generation of methods arising from the machine learning community. The goal of these methods typically is to provide the commonly called 'variable importance' measures, following two general approaches. In one, they assess the importance of each or a sub-set of inputs in constructing a response surface, via for example Multivariate Adaptive Regression Splines (MARS; Friedman, 1991). If the inclusion of an input or set of inputs significantly improves the quality of fit, they are deemed important (Gan et al., 2014). The other approach, conversely, first fits a response surface model using all inputs, and then assesses how the quality of fit degrades when the sample points for each input (or set of inputs) are permuted (Breiman, 2001; Lakshmanan et al., 2015). While these approaches are typically restricted to importance in fitting data, they do have the advantage of also extending well to classification models, for example, using random forests (e.g., Hapfelmeier et al. 2014), allowing for sensitivity measures that apply to both discrete and continuous variables.

### 2.5. Response surface-assisted SA

In the early 2000s, applied mathematicians formally working on 'design and analysis of computer experiments' (DACE) started building linkages with SA (Santner et al., 2003; Fang et al., 2005). Their emphasis has particularly been placed on linking SA and asymptotic statistical theory (Janon et al., 2014a; Gamboa et al., 2016), space-filling designs of experiments (Tissot and Prieur, 2015; Gilquin et al., 2016), structural reliability (Fort et al., 2016), and Bayesian estimation (Pronzato, 2019). Moreover, response surface surrogates rooted in DACE, such as polynomial chaos and Gaussian process regression, have found applications to approximate sensitivity measures in the case of computationally intensive models (Oakley and O'Hagan, 2004; Janon et al., 2014a and 2014b; Wang et al., 2020). A review of these linkages can be found in Ghanem et al. (2017) and Da Veiga et al. (under review).

### 2.6. SA with correlated inputs

One persistent issue in SA is that nearly all applications, regardless of the method used, rest on the assumption that inputs are uncorrelated (Sobol', 1993; Hoeffding, 1948). Inputs, however, can be correlated and their joint distribution can take a variety of forms in real-world problems. Correlation in this context refers to statistical dependency between any subset of inputs, independent of the system under investigation. The correlation effect is different from the 'interaction effect' which refers to the presence of non-additivity of the effects of individual

inputs on the system output (Razavi and Gupta, 2015). It is now being increasingly recognized that ignoring correlation effects and multivariate distributional properties of inputs largely biases, or even falsifies, any SA results (Do and Razavi, 2020). Recently, several methods have been developed to account for such properties, including extensions of the Hoeffding-Sobol' decomposition (Chastaing et al., 2012), regression-based methods (Xu and Gertner, 2008), copula-based methods (Kucherenko, 2012; Do and Razavi, 2020; Sheikholeslami et al., under review) and game-theory concepts (Owen, 2014; Owen and Prieur, 2017; Iooss and Prieur, 2019).

### 2.7. Software tools and applications

To promote and advance the use of SA, there has been tremendous, albeit fragmented, progress in building computer packages. These operationalize the various SA methods via different programming languages, and include Dakota (Adams et al., 2020) in C++, SobolGSA (Kucherenko and Zaccheus, 2020) in C#, MATLAB, and Python, UQLab (Marelli and Sudret, 2014) in MATLAB, OpenTURNS (Baudin et al., 2017) in Python and C++, the 'sensitivity' package (Iooss et al., 2018) in R, SALib (Herman and Usher, 2017) in Python, PSUADE (Tong, 2015) in C, VARS-Tool (Razavi et al., 2019) in MATLAB and C, SAFE (Pianosi et al, 2015; Pianosi et al., 2020) in MATLAB, R and Python, MADS.jl in Julia (Vesselinov et al., 2019) and sensobol (Puy, 2020). As discussed in Douglas-Smith et al. (2020), software programs for SA adopt different design philosophies which reflect different disciplinary foci and vary in terms of usability, including extent of documentation and assumption of users' prior knowledge. In parallel, the generation of test beds for different methods and software packages is receiving increasing attention (Razavi et al., 2019; Becker, 2020).

Applications of SA are widespread across many fields, including earth system modeling (Wagener and Pianosi, 2019), engineering (Guo et al., 2016), biomechanics (Becker et al., 2011), water quality modeling (Koo et al., 2020a and 2020b), hydrology (Shin et al., 2013; Haghnegahdar and Razavi, 2017), water security (Puy et al., 2020c), nuclear safety (Saltelli and Tarantola, 2002; Iooss and Marrel, 2019) and epidemiology (Burgess et al., 2017; VanderWeele and Ding, 2017), to name a few. The most quoted handbook for SA is a primer for global sensitivity analysis (Saltelli et al., 2008) with over 5,000 citations from across the scientific disciplines. A cross-disciplinary review of SA applications can be found in Saltelli et al. (2019). The wide and growing use of SA suggests that a cohesive treatment of SA as a discipline in its own right, inclusive of a well-honed syllabus for teaching, will have a large beneficial impact across the sciences in general.

## 3. Challenges and New Frontiers

Given the significant progress and popularity of sensitivity analysis (SA) in recent years, it is timely to revisit the fundamentals of this relatively young research area, identify its grand challenges and research gaps, and probe into the ways forward. To this end, the multidisciplinary authorship team has identified six major themes of 'challenges and outlook', as outlined in **Figure 2**. In the following, we discuss our perspective on the past, present and future of SA under each theme in a dedicated section. The overarching objective here is to identify possible future avenues that will take SA to the next level, one that is especially

beneficial to meeting the challenges of modeling complex, societal and environmental problems (e.g., Elsawah et al., 2020a).

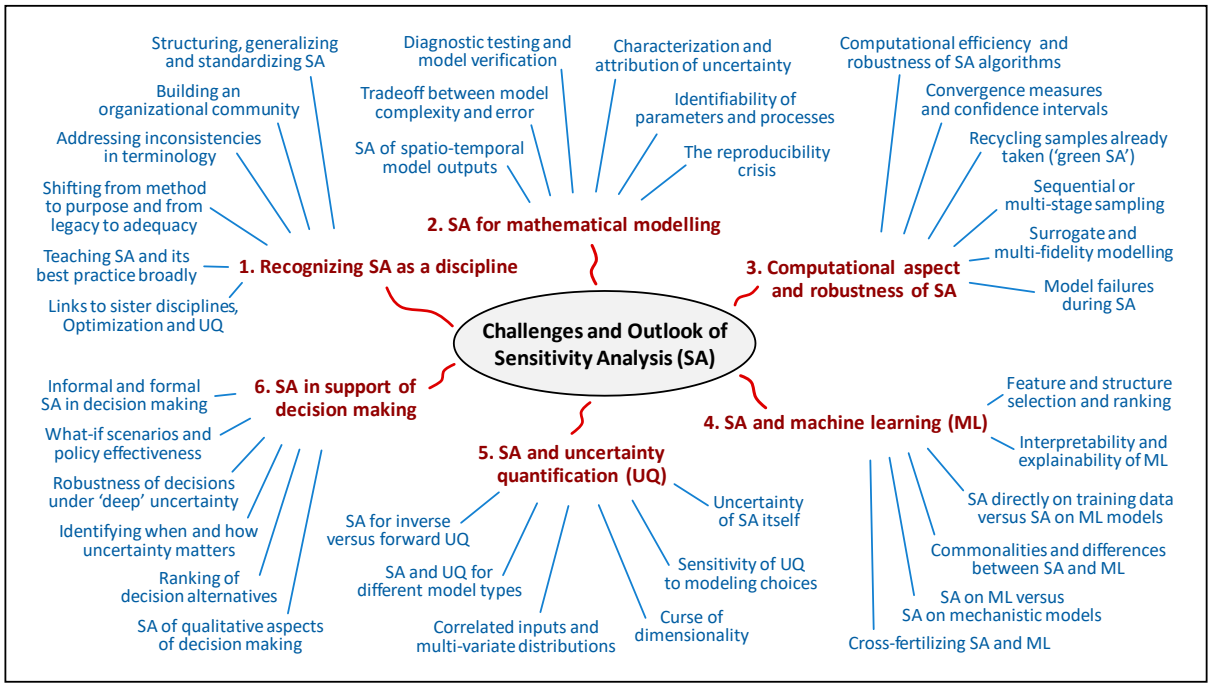


**Figure 2.** The six major themes of 'challenges and outlook' in the theory, methods and application of SA.

### 3.1. Towards a structured, generalized and standardized SA discipline

While SA is now considered by some to be standard practice in modeling (Norton, 2015; Pianosi et al., 2016; Razavi and Gupta, 2015), it is not a formally recognized discipline, nor a coherent subfield of applied mathematics or applied statistics; for example, it is spread across the Mathematics Subject Classification (AMS, 2020). Sociologically, disciplinary fields are communication systems, enabling discourse and the dissemination of knowledge between practitioners (Stichweh, 2001). Additionally, formally recognized disciplines are organizationally distinct, with communities structured around the production of knowledge and training of practitioners (Casetti, 1999; Stichweh, 2001). SA fulfils all of these criteria, with the exception of an organizational community, that make it distinct from related fields of study. For instance, there are no academic institutions nor is there any scientific journal focused on SA.

The only official link binding (part of) the SA family is an international conference called 'Sensitivity Analysis of Model Output' (SAMO), held once every three years since 1995 - the 9th instalment of which was held in 2019 in Barcelona, Spain, with the forthcoming 10th instalment to take place in Tallahassee, Florida (USA) in 2022. Consequently, the current family of SA researchers and practitioners is spread over many disciplines, and there are dramatic differences in the SA capacity and maturity in different contexts.

Despite the lack of a specialized journal for SA, there is a relatively large (and growing) number of publications on the subject. Within the "Water Resources" research area alone, the number of SA publications has grown significantly over the past decades. Based on a search in the Clarivate Analytics Web of Science platform, the yearly publication count is now equal to about one-third of publications in the related, but well-established, field of Optimization (Razavi and Gupta, 2015). The fact that Optimization is an extensive discipline with several dedicated journals suggests that SA should rightfully seek to become an independent discipline, albeit one that interacts effectively with other disciplines.

In the meantime, treatment of SA as a subject is diverse and couched in disciplinary-specific terms and traditions, and taught to varying standards (Saltelli, 2018). Accordingly, the emphasis and importance of SA across the fields in which it is applied are equally as diverse. This state of affairs is not unlike the beginnings of other scientific fields, such as Computer Science which separated from Mathematics and Engineering sometime in the 1940s (Tedre, 2007; Denning et al., 1989), and Hydrology which did not find its footing as a separate discipline until the 1950s or, arguably, the early 90s (McCurley and Jawitz, 2017; Klemeš, 1986). The diverse treatment of SA research is partly responsible for the existing sluggishness to accept SA as a discipline.

### 3.1.1. Recognize SA as a discipline

Recognition of SA as a discipline in its own right requires universal acknowledgement that methodological developments and guidance for application of SA are frequently transferable across application contexts. A major challenge to overcome includes inconsistencies in terminology, methodology and fundamental definitions across the contexts and disciplines in which SA is applied (Saltelli et al., 2019). Arguably, there are examples in the literature where SA practice has been perfunctory or inappropriate, possibly misinforming the users and even a resulting policy (Saltelli and Annoni, 2015; Saltelli et al., 2019). Furthermore, although the links between SA and the well-established field of Uncertainty Quantification (UQ) are clear to SA researchers **(see Section 3.5)**, most SA applications seen in the academic literature do not adequately map the uncertainty in model inference (Saltelli and Annoni, 2010; Ferretti et al., 2016; Saltelli et al., 2019).

The exceptions are typically publications written by SA researchers who pursue SA methodological developments. This has the unfortunate result that SA-related publications fall largely into two classes: proposals for new or refined methods – with illustrative applications – written by SA researchers; or application papers with often poor-quality SA written by non-SA researchers. The scientific journals where these findings are published are largely disconnected for the two classes. In an ideal world, modeling teams should include at least one researcher versed in SA. Given the wide applicability of SA, its practitioners are dispersed across the sciences and their work similarly disseminated. There is a need to establish a publication outlet focused specifically on SA to signify the separate concerns and foci of research. Establishment of an SA-specific journal would strengthen communication of the current state of SA research, particularly for uninitiated modelers.

Another challenge to proper uptake of SA is that its application in some fields might appear under other titles. For example, a recent article in Nature (Adam, 2020) refers to modeling activities "in which thousands of versions of the model are run with a range of assumptions &

inputs, to provide a spread of scenarios with different probabilities" as 'ensemble modeling'. Such activities, however, are typically considered as UQ and possibly SA in the environmental modeling community. SA-type activities are also seen under the title "single-model perturbed physics ensembles" in the climate modeling community (Bellprat et al. 2012). Confusingly, the expression "climate simulation ensemble" is more often used to indicate the case where different models, e.g., developed by different teams, are applied to the same problem (Donev et al., 2016; IPCC, 2016); see a critical discussion in Saltelli, Stark, et al. (2015). The risk of these diverging nomenclatures is that research advances made in UQ and SA may go unnoticed by some communities.

### 3.1.2. Address possible inconsistencies in SA

Different SA methods are based on fundamentally different philosophies, and therefore can result in different, sometimes conflicting, results for the same problem (Tang et al., 2007; Razavi and Gupta, 2015). This inherent nature of the state of the art is unlike many other fields. For example, while the field of Optimization contains a vast variety of approaches, methods and applications, all these may boil down to a common, well-defined philosophy, that is, to find an optimal solution to a formulated problem given certain objective functions and constraints (Maier et al. 2014; Maier et al. 2019). In other words, an optimal solution to a given problem formulation would remain optimal, regardless of the optimization method used, whereas according to current theory the sensitivity assessment for a problem might appear to be quite different, depending on the SA method used.

For improved consistency, two key questions need to be answered before choosing an SA method and carrying out the analysis: (1) why do I need to run SA for a given problem and what is the underlying question that SA is expected to answer? And (2) how should I design the SA experiment to address that underlying question? Thought-out answers are critical, as otherwise most users tend to use methods developed in their own camp (and are therefore most comfortable with), rather than methods that are most suitable for the purpose and problem at hand. As such, focusing on the purpose would facilitate a shift in method selection principles from *legacy to adequacy* (Addor and Melsen, 2019). We emphasize that we do not necessarily advocate the reconciliation of different philosophies and methods, but are pointing out that because SA addresses multiple related problems, it must be made clear why a particular method is the right match for a given research question.

In addition to the purpose and chosen method, SA researchers and practitioners generally need to be more mindful of the subjective, but often *overlooked* decisions they make in the configuration of a method. The 'sensitivity' of SA to such decisions, for example, SA algorithm parameters, may be quite significant for some methods, and ignoring it might result in questionable results. The significance of this issue has been discussed recently by Haghnegahdar and Razavi (2017) in the context of derivative-based methods (e.g., Morris, 1991), and by Puy et al. (2020b) in the context of the PAWN method (Pianosi and Wagener, 2015).

### 3.1.3. Teach SA more broadly and consistently

Formalizing a structured, generalized and standardized SA discipline is attainable in a foreseeable future. Central towards this goal is to invest in systematic and coherent teaching of SA to students across disciplines, who will become the next generation of researchers and

practitioners. SA is currently taught on an ad-hoc basis, mostly via small workshops tailored to specific aspects of SA or as a part of courses related to systems analysis or uncertainty quantification. Perhaps the most formal and systematic effort to teach SA has been a summer school on SA run by European Commission's Joint Research Centre, held ten times between 1999 and 2018. SA needs to become an independent but integral part of the curriculum across the relevant disciplines, alongside other topics such as Optimization, and Validation, Verification and Uncertainty Quantification (VVUQ, see e.g., National Research Council, 2012). New interdisciplinary graduate courses need to be developed to comprehensively cover SA and to teach and promote best SA practice. As discussed in the following sections, the SA discipline has extensive untapped potential for a variety of problems and applications.

### 3.2. Untapped potential of SA for mathematical modeling

Historically, the majority of 'formal' SA applications have been directed towards mathematical models, to better understand how they work and diagnose their deficiencies, and to contribute to their calibration and verification (Saltelli et al., 2000). In this context, a dominant application of SA is for parameter screening, to support model calibration by identifying and fixing non-influential parameters. There is potential, however, for SA to further address several challenges in mathematical modeling through advancements in the management of uncertainty, assessment of model quality through testing and diagnostics, and tackling non-identifiability and model reduction. For example, mathematical modeling could benefit from structure and standards based on statistical principles (Saltelli, 2019), including a systemic appraisal of model uncertainties and sensitivities. In the following, we outline the potential of SA to advance mathematical modeling.

#### 3.2.1. Management of uncertainty

Management of uncertainty through its characterization and attribution should be at the heart of the scientific method and, *a fortiori*, in the use of science for policy (Funtowicz and Ravetz, 1990). The problem of uncertainty management is core to the modeling craft and should be an integral part of any model development and evaluation exercise (Jakeman et al., 2006; Eker et al., 2018). While SA has significant potential, its application often does not adequately map the uncertainty in model inference. In a recent five-point manifesto for responsible modeling, global SA is invoked as essential to the task of mapping the uncertainty in every model assumption (Saltelli et al., 2020).

A major step forward in mathematical modeling will be to better evaluate uncertainty in model predictions and to trace that uncertainty back to its sources across the model components, parameters and inputs. SA is uniquely positioned to do so as it can help to decompose the prediction uncertainty and attribute it to the individual factors and their interactions. Basically, SA can help answer a critical question (Razavi et al., 2019): *when and how does uncertainty matter?*

A major but almost totally neglected issue in mathematical modeling is that, while models are becoming more and more complex, they are treated more and more like a black-box, even by model developers themselves. In real-world applications, those models tend to be used without much attention to their complicated internal mechanics and (not always justified) assumptions. A manifestation of this issue is the fact that many modern, physically-based models include countless numbers of hard-coded parameters (e.g., see Mendoza et al., 2015),

supported by the rationale, explicit or implicit, that scientists can characterize those parameters with absolute certainty. Such practice can render progressive initiatives on 'open science' (Vicente-Saez and Martínez-Fuentes, 2018) and 'open modelling' (e.g., Openmod, 2020) less effective. SA is much needed to prize open and cast light into these black boxes and to illuminate the dominant sources of uncertainty.

Furthermore, the quality of both statistical and mechanistic models struggles with common issues when dealing with uncertainty. In statistics, the p-test can be misused to overestimate the probability of having found a true effect (Colquhoun, 2014; Stark and Saltelli, 2018). Likewise, certainty may be overestimated in modeling studies (Nearing and Gupta, 2018), thus producing unreasonably precise estimates even in situations of pervasive uncertainty or ignorance (Saltelli et al., 2015; Thompson and Smith 2019), including in important cases where science needs to inform policy (Pilkey and Pilkey-Jarvis, 2007). It is an old refrain in mathematical modeling that since models are often over-parameterized with respect to the information that can be extracted from the available data, it can sometimes appear that they can be made to conclude anything we choose (Hornberger and Spear, 1981).

As such, analogous to under- and over-fitting issues in statistical models, mechanistic model development suffers from a trade-off between 'model completeness' and 'propagation error' (Saltelli, 2019) - see **Figure 3**. The former refers to the adequacy of a model in terms of, for instance, how many aspects of the underlying system are included in the model. The latter, also known as the 'uncertainty cascade' effect (Christie et al., 2011), refers to the notion that adding each new aspect to the model, for example, a new parameter which itself is uncertain, potentially increases the overall uncertainty in the output.

Such tradeoffs challenge model developers to calibrate the right level of complexity in the construction of models. SA can facilitate this process by characterizing and attributing the contributions to overall 'model error' so as to identify the 'sweet spot' where uncertainty attains a minimum. As a simple example, consider the case where the uncertainty of concern is with respect to a sought measure of predictive model error around an observational quantity of interest. The sweet spot would be where no significant improvement in model performance occurs by adding more parameters within a given model structure, where significance corresponds to a level commensurate with the errors/noise in the observations of interest (e.g., see Jakeman and Hornberger, 1993). This shares many parallels with the problem of 'model selection' via, for example, stepwise regression and/or use of information criteria in econometrics and elsewhere (Becker et al., under review).

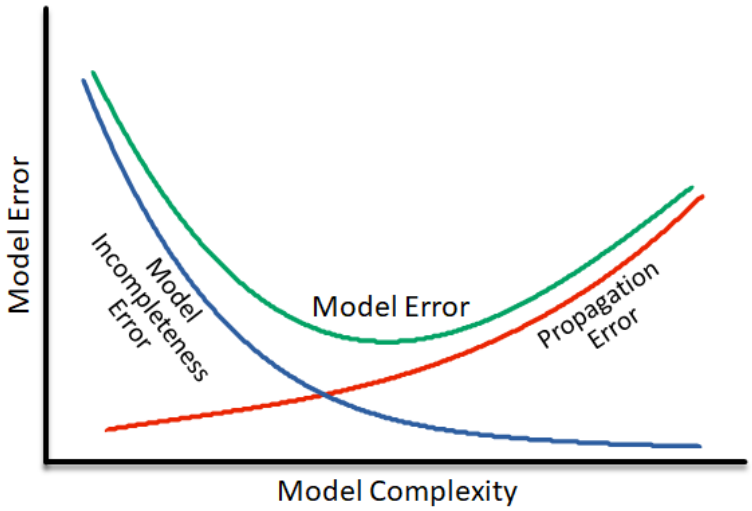


**Figure 3.** Model error as a function of model complexity, adapted from Saltelli (2019). Known as the conjecture of O'Neill (O'Neill, 1973; Turner and Gardner, 2015), this plot hypothesizes that in the initial stages of model construction, the addition of features improves the descriptive and predictive capacity of the model. Beyond a point, the accumulated error brought about by the uncertainty in the description of the features accumulates and propagates to the output, decreasing the descriptive and predictive quality of the model.

### 3.2.2. Diagnostic testing and model verification

SA has significant potential to help in diagnosing the behavior of a mathematical model and for assessing how plausibly the model mimics the system under study for the given application. This capability provides understanding of how a model works, and points to the parts of a model that are deficient. A key to this end is to properly frame the SA problem and articulate that understanding the sensitivity of '*what'* to '*what'* matters for this purpose. As outlined in Gupta and Razavi (2018), the former *'what'* may be chosen from any of the following categories: (a) one or more model performance metrics that quantify the goodness-of-fit of the model responses to observed data (e.g., mean squared errors), (b) a specific targeted aspect of those responses (e.g., extremes or percentiles such as peak flows in a hydrologic model), (c) a compressed set of properties that characterize those responses (e.g., hydrologic signatures such as runoff ratio), or d) the entire spatio-temporally varying trajectory of responses themselves. The latter *'what'* may include continuous or discrete variables describing model parameters, forcings, structural assumptions, etc.

To diagnostically test a model, one may compare SA results with expert knowledge on how the underlying system being modeled works. Example studies based on time-varying and time-aggregate SA results include Wagener et al. (2003), Herman et al. (2013), Haghnegahdar et al. (2017) and Razavi and Gupta (2019). Moreover, the recent emergence of 'given-data' SA methods will provide unprecedented opportunities for model diagnostic testing, as they can directly be applied to observed data as well, in addition to the mathematical models (Plischke et al., 2013; Sheikholeslami and Razavi, 2020). The knowledge gained via SA can be documented in a model's user guide to help practitioners configure and parameterize the

model more effectively. Diagnostic evaluation of models in this manner is analogous to *property-based testing* (Claessen and Hughes, 2000), wherein the logical properties of model behavior are evaluated according to expected behavior. Failure of a model to conform to expected behavior falsifies the assumption that the model is correctly implemented.

In addition, a study using mathematical models may face a diversity of errors and subjectivities. These may stem from process conceptualization and mathematical representation, parameterization, inputs and boundary conditions, discretization choices in space and time, numerical solvers and software coding, up to and including the framing and biases of the modelers themselves (Oreskes, 2000; Iwanaga et al., 2021). Modelers, however, do not subscribe to a unified, reliable and agreed-on code of good practices for testing their models and the quality of the inference that they produce. More work is needed to develop testing strategies based on SA that cover the diversity of subjective factors involved in the process of model development (e.g., Peeters, 2017).

### 3.2.3. Non-identifiability and model reduction

Most models are poorly-identifiable, largely because of over-parameterization relative to the data and information available - see Guillaume et al. (2019) for an overview of identifiability. The assessment of model appropriateness (for a purpose) requires understanding of its identifiability, the sources of any non-identifiability, and the influence of any non-identifiability on the model (Guillaume et al., 2019). SA and identifiability analysis (IA) are different but complementary, primarily because SA is about the properties of a model itself, while IA is more about model properties with respect to observed data. It can be shown that an insensitive parameter is non-identifiable, but the converse is not necessarily true, that is, a sensitive parameter may or may not be identifiable. Therefore, SA can help in part to recognize the non-identifiable components of a model.

Knowledge of identifiability can be used to simplify a model structure by fixing or combining parameters that on their own are ineffective in influencing model outputs. Model reduction, however, should be done with caution, as a parameter that seems non-influential under a particular condition might become quite influential under a new condition (e.g., see Tonkin and Doherty, 2009). For example, a snowmelt parameter in a hydrologic model has no influence in time periods without snow, whereas it becomes dominantly influential in snowmelt seasons. In such cases, fixing the parameter will limit the agility, and therefore, the fidelity of the model in mimicking the underlying system. Also, fixing parameters with small sensitivity indices may result in model variations that cannot be explained in the lower dimensional space (Hart and Gremau, 2019).

### 3.2.4. The reproducibility crisis and SA

The challenges of modeling need to be seen in the broader context of the so-called ‘reproducibility crisis’ (Saltelli and Funtowicz, 2017; Saltelli, 2018) where misuse or abuse of statistics (Stark and Saltelli, 2018; Gigerenzer and Marewski, 2015; Leek et al., 2017; Singh Chawla, 2017; Wasserstein and Lazar, 2016; Gigerenzer, 2018) is often cited as the root cause of the crisis (Ioannidis, 2005). Current non-reproducible science is ecologically fit to the existing science governance arrangements (Smaldino and McElreath, 2016), including its ‘publish or perish’ culture (Banobi, 2011), and is resistent to reform (Chalmers and Glasziou, 2009; Edwards and Roy, 2017). In the field of clinical medical research, for instance, the

percentage of non-reproducible studies may be as high as 85% (Gigerenzer and Marewski, 2015).

The field of mathematical and computational modeling has started grappling with the reproducibility crisis as well (Hutton et al., 2016; Saltelli, 2019; Saltelli, Bammer et al., 2020). Development of research-specific software is at the core of modern modeling efforts. Most modelers, however, are not formally taught software development practices (Hannay et al., 2009), such that models are rarely designed and developed in a manner that supports further use (or reuse) beyond its original research-specific context. The consequent lack of accessible code and data then feeds into issues of reproducibility (Hutton et al., 2016; Hut et al., 2017).

As alluded to above, the 'publish or perish' culture limits the recognition researchers receive for developing, and maintaining, long-lived software, associated data and supporting documentation that underpins reproducibility. In some cases, code and data may not be accessible at all even after contacting authors (Stodden et al., 2018). One observation is that researchers want to perform research, not write software (Crouch et al., 2013; Sletholt et al., 2012). That said, there is increasing recognition of the importance, and benefits of, supporting open and accessible research software.

Support of initiatives towards improving computational reproducibility has been growing (e.g., Ahalt et al., 2014; Crouch et al., 2013), culminating recently with the FAIR (Findable, Accessible, Interoperable, Reusable) principles for open and accessible data management (Wilkinson et al., 2016). Some journals now award "open code badges" (Kidwell et al., 2016) to highlight publications with accessible code and data.

SA and its practitioners can contribute to addressing the aspects of this crisis which directly affect mathematical modeling. Reproducible model-based studies need the kind of transparency that SA can offer, by way of making explicit the conditionality of model-based inference, as well as the conditionality of the associated uncertainties. Essential to this end is to standardize and promote best SA practice, along with the development of SA-related software that can easily be coupled with any model. There has been an increasing number of open software packages which democratize both common and experimental SA techniques and applications (a non-exhaustive list is provided in Section 2).

### 3.3. Computational aspects and robustness of SA algorithms

Computational burden has been a major hindrance to the application of modern SA methods to real-world problems. Many of the existing SA methods have remained under-utilized in various disciplines as they require a large sample size, particularly for models with higher-dimensional spaces. State-of-the-art, spatially distributed models are typically computationally demanding themselves and take minutes, hours or even days for a single run. Although the growth of computing resources is making the application of current algorithms to existing problems more affordable (see, e.g., Prieur et al., 2019), not everyone has access to powerful computing, and there will always be modeling problems where computing power will not be quite enough for existing algorithms.

For example, most (i.e., ~70%) SA applications in earth and environmental systems modeling have been limited to low-dimensional models (i.e., with 20 or fewer factors involved) (Sheikholeslami et al., 2019), whereas there are abundant applications with models that can

have up to hundreds (e.g., ~900 in Borgonovo and Smith, 2011) or even thousands (e.g., ~40,000 in Lu et al., 2020) of parameters. In the machine learning context, the number of model parameters can reach millions (e.g., BERT; Houlsby et al., 2019), even trillions of parameters (e.g., ZeRO; Rajbhandari et al., 2019). The application of SA with machine learning is further complicated because of the fundamental differences between machine learning and other types of models **(see Section 3.4)**.

Computational obstacles need to be properly assessed and addressed so that SA can be applied to any model, particularly those whose results are of immediate significance to society. Inadequate, or non-existent, application of SA leads to models for which society cannot characterize their confidence. Modeling is a social activity, and the acceptance of model prescriptions regarding, for example addressing an industrial risk, financial crisis, hurricane, or a pandemic, calls for mutual trust between model developers and end users (Saltelli et al. 2020). SA can help with building this trust, by providing insights into the internal functioning of models. The future, therefore, needs new generations of algorithms to keep pace with the ever-increasing complexity and dimensionality of the state-of-the-art models. Building on known theoretical and empirical convergence rates for many sampling-based approaches, further theory could identify fundamental limits on existing classes of algorithms to help identify breakthroughs required.

### 3.3.1. Essential definitions and components

A complete assessment of the computational performance of any SA algorithm must be conducted across four aspects: efficiency, convergence, reliability and robustness. *Efficiency* refers to the amount of time/number of computations required to perform SA and is often assessed by the number of model runs (i.e., sample size) required for the algorithm to converge to some specified level. *Convergence* of an SA algorithm is non-trivial to assess, as the answer typically cannot be preordained from theory. It depends on several factors including: the model type and its complexity, the overall objective of the SA (e.g., prioritization, where sample size can generally be smaller, versus screening), the SA method itself, definition of convergence and level of certainty required, choice of time period for the input forcing variables, and the width of parameter ranges and distribution sampled (see Shin et al., 2013).

*Reliability* refers to any measure of *correctness* of SA results and its accurate assessment requires the availability of the 'true' SA results. Reliability of an algorithm may only be assessed when the model is simple (e.g., simple algebraic test functions) or the true results are somehow given. *Robustness*, often used in lieu of *reliability*, measures how consistent an algorithm performance remains when the sample points and algorithm parameters change. For example, an SA algorithm is robust to sampling variability if its performance remains almost 'identical' when applied on two different sample sets taken from the same model. In cases where running multiple replicates of the same experiment is not possible, *bootstrapping* (Efron, 1987) is often used with SA algorithms to estimate robustness in the form of uncertainty distributions on sensitivity indices without requiring additional model evaluations.

Addressing computational challenges requires a proper understanding of the design, functioning and interaction of the three general components of any SA algorithm. These components are: (1) the 'experimental design' that employs a sampling strategy to select sample points in the factor space of interest; (2) the 'function evaluation procedure' to run the

model, collect and store sampled model responses (i.e., obtained by running a model many times on a spatio-temporal and/or other domains); and (3) the 'integration mechanism' to numerically integrate the sampled data to estimate sensitivity indices.

The choice of experimental design is often dictated by the integration mechanism of the SA algorithm of interest. For example, in the method of Morris (1991), the mechanism that integrates the elementary effects across the factor space requires sample points to be taken equally spaced from each other by a given distance, by changing one factor at a time. As a result, a sample taken for one SA algorithm may not be usable by another algorithm or possibly for other purposes. The function evaluation procedure is typically the most computationally intensive component of SA. This is not only because of the computational burden of running the models themselves, but also the *overhead* for storing, retrieving and manipulating the model responses on high-resolution domains (e.g spatio-temporal). The following sections outline possible progress in addressing computational challenges with respect to the above three components.

### 3.3.2. Experimental design and integration

Improving the efficiency of SA is tied to improving experimental designs in conjunction with integration mechanisms. For example, consider that global sensitivity measures are commonly written as an average over the distribution of the input of interest ($X_i$) of an inner statistic (Borgonovo et al., 2016). The brute force application of the definition of global importance measures would lead to an estimation cost of C=N*n*K, where N is the number of runs from the distribution of input $X_i$, n is the number of runs needed for the inner statistic and K is the number of model inputs. If N=1000, n=1000 and K=10, we are already at 10,000,000 model runs. This cost can be reduced to C=N(K+2) by using the sampling and integration mechanism of Saltelli (2002) to estimate first- and total-order variance-based sensitivity indices.

A recently developed approach to sampling and integration is to extract information contained in all *pairs* of sample points rather than the individual points. This is useful because the number of pairs (2-combinations) grows quadratically (~$n^2/2$) with the sample size, n (Razavi and Gupta, 2016a). For example, if n=1000, we get 499,500 pairs, but doubling the sample size to n=2000 results in a fourfold increase to 1,999,000 pairs. Razavi and Gupta (2016b), Becker (2020), and Puy et al. (2020a) have shown the efficiency of this approach in estimating variance-based total-order effects, through the method 'variogram analysis response surfaces' (VARS).

Alternatively, the future of SA may step more towards 'sampling-free' algorithms that can work on any 'given data' (see e.g., Plischke 2010; Plischke et al. 2013; Pianosi and Wagener, 2018; Sheikholeslami and Razavi, 2020). Such approaches may be referred to as 'given-data sensitivity analysis', or alternatively 'green sensitivity analysis' in that they can recycle available samples, for example from previous model runs, allowing for samples to be incrementally obtained and avoid wasting computational budget. The computational cost of the corresponding estimators is then n model evaluations. In addition to being 'green', some of these approaches tend to be computationally much more efficient than other methods, and in certain cases produce robust SA estimates with a very small sample size (Sheikholeslami and Razavi, 2020).

Notably, most algorithms for given-data SA involve a parameter-tuning step that may be non-trivial with a bias-variance compromise perspective, which is to avoid both over-fitting and over-smoothing. Examples of such parameters include the bandwidth for kernel regression based estimators, the number of leaves for random forest based estimators, the truncation argument for spectral procedures, etc. Bootstrapping (Efron, 1987) may be used for the selection of such parameters (see, e.g., Heredia et al., 2020), but it may consume an excessive amount of computation time. Adaptive selection of these parameters or developing parameter-independent algorithms is a challenging issue that needs to be addressed in future. More recently, authors have proposed parameter estimation procedures based on nearest neighbors (Broto, 2020), rank statistics (Gamboa et al., 2020) and robustness-based optimization (Sheikholeslami and Razavi, 2020). These methods are still relatively new and need to be tested across a range of problems with different dimensionalities (see, e.g., Puy et al., 2020a).

Above all, convergence considerations need to be at the heart of the development and application of any SA algorithms. Recently developed convergence criteria (e.g., Sarrazin et al., 2016; Sheikholeslami et al., 2019) and stopping criteria (see, e.g., Gilquin et al., 2016; Rugama et al., 2018) can be useful in this regard. In general, the literature is replete with studies that indicate convergence for a particular model or function and some particular instances of the above other factors. But, these offer only limited guidance. Consequently, many users usually choose the computational budget (i.e., the number of model runs) for SA on an ad-hoc basis, rather than on convergence, reliability and robustness considerations. Rather than relying on guidance from past studies, analysts should be encouraged to adopt methods and software packages that explicitly address this issue. Bootstrapped estimates of sensitivity indices are, for example, now common and enabled by default in R and Python packages.

SA users should assess convergence rates as the sample size increases, based on intermediate results. But a typical hindrance is that many sampling strategies involve one-stage sampling that generates the entire set of sample points at once, requiring the user to specify the sample size a priori (Sheikholeslami and Razavi, 2017). This is a disadvantage as it is unlikely for users to know the optimal sample size that enables the algorithm to converge to robust results. Therefore, there is a need for sequential or multi-stage sampling strategies such as Sobol' sequences (Sobol' 1967, Sobol' et al. 2011) and Progressive Latin Hypercube Sampling (PLHS; Sheikholeslami and Razavi, 2017) that enlarge the sample size during the course of SA while preserving the distributional properties of interest. The superior performance of Sobol' low-discrepancy sequences over random sampling has been demonstrated in several studies (Gilquin et al, 2016; Gilquin et al, 2017b; Sheikholeslami et al. 2017, Rugama et al., 2018; Kucherenko et al., 2011, Kucherenko et al., 2015).

Furthermore, as the value of SA for high-resolution (e.g., spatio-temporal) model outputs is increasingly recognized, innovative strategies to handle the increased storage and retrieval overhead are needed. Currently, all model runs are typically stored first, requiring excessive storage capacity for large models, and sensitivity indices are computed post hoc. Future developments, similar to Jakeman et al. (2020) and Terraz et al. (2017), can helpfully include SA algorithms that merge function evaluation and integration mechanisms such that sensitivity indices are updated as new results are made available.

### 3.3.3. Function evaluations

Much attention has been geared towards the *function evaluation procedure* under the umbrella of 'surrogate modeling'. Surrogate models, also called response surface models, metamodels or emulators, are used in lieu of computationally intensive models and can be statistical (i.e., 'response surface surrogates') or mechanistic (i.e., 'lower-fidelity mechanistic surrogates') (Razavi et al., 2012). SA methods based on response surface surrogates build approximations, such as polynomial chaos expansions (Xiu and Karniadakis, 2002), (Q)RS-HDMR (Zuniga et al., 2013) and Gaussian process kriging (Rasmussen and Williams, 2005), using a limited number of expensive model evaluations. Once built, sensitivity measures can be estimated by sampling the surrogate instead of the original model at negligible cost, or, in some cases, can be estimated analytically (Sudret, 2008; Marrel et al., 2009).

Due to the sheer computational expense of some models, building an accurate surrogate using only data from the most trusted numerical model can be challenging. To reduce the computational burden of building surrogates, multi-fidelity methods combine limited high-fidelity data with larger amounts of lower-fidelity data coming from models with reduced physics or coarser numerical discretization. SA methods using multi-fidelity approximations can produce sensitivity estimates that converge to high-fidelity estimates, but do so at a fraction of the cost (Palar et al., 2018). Moreover, these methods can build upon the aforementioned advances made for single-fidelity models and adaptively allocate samples to resolve uncertainties and sensitivities (Jakeman et al., 2020).

The use of surrogate modeling introduces a new challenge, accounting for the uncertainty arising from the surrogate model itself (e.g., model error) combined with the errors of the estimation procedure. While progress has been made in the assessment of surrogate modeling uncertainty (see e.g., Jones, 2001; Oakley and O'Hagan, 2004; Sobester et al., 2005; Razavi et al., 2012; Janon et al., 2014a and 2014b), and some Bayesian approaches are already capable of incorporating this uncertainty into posterior distributions of sensitivity measures (Oakley and O'Hagan, 2004; Gramacy and Taddy, 2010), further advancements to properly incorporate such uncertainties in SA estimates and respective confidence intervals are likely.

Surrogate modeling strategies, particularly those based on response surfaces, become less effective in high-dimensional problems (Razavi et al., 2012). To address limitations related to high-dimensionality, adaptive and goal-oriented (Buzzard 2012; Jakeman et al. 2020) approaches can be used. These approaches can allocate samples to lower dimensional subspaces in a manner that addresses the curse of dimensionality and results in enormous computational gains.

Lastly, an issue hindering the application of SA to large, complex models is that some models may fail to run properly ('crash') at particular points in the factor space and not produce a response. Simulation failures mainly occur due to non-robust numerical implementations, the violation of numerical stability conditions, or errors in programming. SA algorithms are typically ill-equipped to deal with such failures, as they require running models under many configurations of factors. In addition to improving properties of the original models (e.g., Kavetski and Clark 2010), more research is needed to equip SA algorithms to handle model failures, which is becoming a more pressing issue as the complexity of mathematical models

grows. One of the very first studies addressing this issue in the context of SA is Sheikholeslami et al. (2019), where a surrogate modeling strategy is used to fill in when the original model fails. To handle this issue, strategies can also be adopted from other types of analyses. For example, Bachoc et al. (2016) used a design of experiments to detect computation failures and code instabilities, and Bachoc et al. (2020b) developed a method to classify model parameters to computation-failure or -success groups during optimization.

### 3.4. SA and Machine Learning

Machine Learning (ML) has achieved unprecedented performance in complex tasks typically performed by humans such as image classification (Krizhevsky et al., 2017), natural language processing (Young et al., 2018), and gaming (Silver et al., 2018). This success, combined with the growth in computational power and the increasing availability of big data, has motivated the application of ML to a wide range of problems across many disciplines, including the earth sciences (Reichstein et al., 2019), robotics (Torresen, 2018), medicine (Hosny et al., 2018) and finance (Lee et al., 2019). Research and development with ML are now viewed as a major avenue forward by many industrial sectors such as energy, security, cyber security, transport, defence, aeronautics and aerospace (AVSI, 2020).

Deep Learning (DL) has emerged in recent years as a leading ML approach for a wide variety of regression and classification applications (Goodfellow et al., 2016). DL is a newly-formalized term that refers to the way Artificial Neural Networks (ANNs) with more than one hidden layers learn representations from data. With a rich and long history dating back to the 1980s (e.g., Rumelhart et al., 1986; Hornik et al., 1989), ANNs have become perhaps the most popular tool for ML. Therefore, major portions of this section are primarily focused on ANNs.

A critical challenge facing ML, particularly DL applications is their typical lack of 'interpretability' and 'explainability'. These two terms, usually used interchangeably in the literature, refer to the ability of a model developer to make sense of why the model functions the way it does and to explain that to a user (Rudin, 2019; Samek and Müller, 2019; Roscher et al., 2020). In many real-world applications, the acceptance and use of an ML model's outputs requires an explanation of why and how the model works. In addition, transparency and auditing of ML models can raise legal issues nowadays (Rudin, 2019), especially when personal data are involved. What complicates this further is that ML is reliant on processes that infer correlation rather than causation (Obermeyer and Emanuel, 2016).

SA can offer new opportunities for the development and application of ML. These opportunities are rooted in the fact that SA and ML, in many cases, look at the same problem via two different approaches. In fact, a goal of ML in most application areas, especially in environmental modeling, is to construct a function that maps variables in an input space to those in an output space (Hastie et al., 2002; Razavi and Tolson, 2011). Generally, such functions are purely data-driven, not accounting for any underlying processes, physical or otherwise. Similarly, SA looks at the relationship between the inputs and outputs, but instead of constructing a mapping function, it estimates the relational strength between each single or group of inputs and the outputs, via different sensitivity indices.

Such 'informal' commonalities between tools for SA and ML provide significant potential for each field to benefit from the other. In exploring the potential, however, one must be mindful of the central differences that exist between computer experiments that provide data for SA

and more general experiments, including those in laboratories or in fields, which provide data for ML. These differences are as follows:

(1) Computer experiments are usually deterministic (with the exception of stochastic models such as agent-based simulators), whereas real-world data, commonly used in ML, are usually polluted with observational errors, often with unknown properties.

(2) The linkage between the input and output variables in computer experiments is generally via hypothesized *causal* relationships, while this is not necessarily the case in other types of experiments.

(3) In ML applications, users typically need to have access to very large data sets, but this is typically not possible in the computer modeling context, where physical data acquisition, such as for model verification, may be very expensive.

(4) In computer experiments, users have full control over the experimental design and the way a sample is taken, whereas this is usually not the case in other types of experiments.

The following sub-sections explain how the fields of SA and ML already have and can continue to cross-fertilize.

### 3.4.1. Feature and structure selection in ML

SA can support feature ranking and selection, where the term 'feature' is equivalent to the term 'factor' in common SA literature. The objective is to find the dependency strength between features and targeted labels to enable the user to choose the features that best explain and possibly predict the output of interest.

SA can be used *prior* to ML model design and training to choose only the features that are most statistically associated with the output data (Galelli et al., 2014). The classic (non-SA) approaches to do so include the standard statistical correlation metrics (e.g., Pearson correlation coefficient, Spearman's rank correlation coefficient, and Kendall's tau), information-theoretic metrics (e.g., entropy, mutual information, and dissimilarity measures), and advanced dependence measures such as distance covariance and Hilbert-Schmidt Independence Criterion (Da Veiga, 2015). These classic approaches can be complemented by the advanced SA techniques that work directly on sample data, in the absence of any model. For example, the recently developed 'given-data' SA paradigm (Plischke et al., 2013; Sheikholeslami and Razavi, 2020) can be used on data available for ML to rank features according to their relational strength with the output of interest.

Another approach could be to use the 'target and conditional SA' (Raguet and Marrel, 2018) that enhances feature selection when the underlying phenomenon is under-represented in the dataset (e.g., unbalanced datasets associated with the prediction of rare or extreme output events). Special care, however, needs to be taken in SA on training data, because sample data on features are by and large real-world data (or properties thereof), typically having undefined distributions and unclear correlation structures and spatial dependencies.

SA can also be applied to an ML model *after* training to identify the controlling features and how they interact to generate the model output. A simple way to do so is local SA. Examples include the use of one-factor-at-a-time SA (e.g., Lek et al., 1995; 1996; Maier and Dandy, 1997; Liong et al., 2000) and the calculation of partial derivatives of the model outputs in response to changes in the model inputs (e.g. Dimopoulos et al., 1995; 1999; Tison et al., 2007; Vasilakos et al., 2009; Mount et al., 2013). Such assessments can also be expanded following the concepts of global SA. For example, importance score methods for feature ranking in ML, based on permutation and resampling, have informal roots in SA and strong connections with Sobol' sensitivity theory. Examples of such methods with ANNs and random forests include Breiman (2001), Lakshmanan et al. (2015), Gregorutti (2015), Wei et al. (2015), and Benoumechiara (2019). More formally, SA has been used to rank features into ANNs and random forests, according to their importance in explaining the variation in outputs (Fock, 2014; Fernández-Navarro, 2017; Zhang, 2019).

Thus, SA can point to the most influential features learned by an ML model, its most active parts, and detect interactions between features (Lundberg and Lee, 2017; Lundberg et al., 2020; Ribeiro et al., 2016; Štrumbelj and Kononenko, 2014). In this way, SA can potentially enable the identification of optimal levels of structural complexity of ML models, which is particularly useful in designing deep learning constructions. For example, eFAST and random balance design have been used to prune redundant neurons in ANNs (Lauret et al., 2006; Li, 2017). A similar application to an area adjacent to ML, that of variable selection in regression, has been successfully tested by Becker et al. (under review).

### 3.4.2. Interpretability and explainability of ML

Despite being very successful, ML has been criticized for being a black box, where the reasons for an answer are unknown. This challenge may offset the value of ML in a range of applications, particularly where researchers and decision makers seek transparency. SA can help in peering inside the ML, to improve its explainability and interpretability (e.g., Lundberg and Lee, 2017). The goal here is to produce 'explanations' that are intelligible and meaningful to end users, which aid in improving transparency and building trust, help in identifying the best ML model among several comparably performing models, and enable diagnosis of model errors (Samek et al., 2019).

A significant portion of efforts in the literature to provide explanations have been based on the assessment of feature importance for developed ML models, as described in **Section 3.4.1**. Thus, SA offers new opportunities to provide insights into the general behavior of a model by highlighting how the different features influence the model output. Such insights are particularly important for the 'structural validation' of ML models (Humphrey et al., 2017). If those models are not structurally valid, their behavioural response to different input stimuli can be erratic and counter to physical system understanding, making them difficult to apply in practice with confidence; see Wu et al. (2014) for a discussion. ValidANN (Humphrey et al., 2017) is an example software package using SA for this purpose. In the assessment of feature importance, SA has to deal with two challenges often encountered in ML (as is the case in some other types of modeling): the often high-dimensionality of the feature space and multicollinearity/dependencies between those features. These challenges are discussed in **Sections 3.5.3 and 3.5.4**.

In addition to assessing the sensitivities to the features, SA in principle has potential to be applied to any parts of ML models, including their structure (e.g., the number of layers and neurons in ANNs) and parameters (e.g., weights and biases in ANNs). Such practice, however, can be hampered by fundamental differences between standard SA applications to mechanistic models and those to ML models. These differences, as outlined below, necessitate further research to develop SA strategies particularly tailored for ML.

First, unlike mechanistic models, the structure of many ML models, particularly ANNs, are based on the notion of 'connectionism', meaning their internal operations are massively parallelized. For example, in a mechanistic hydrologic model, the soil parameterization equation may be solely responsible for representing how soil columns store and release water, while other parts of the model may be in charge of other physical processes. In the case of an ANN-based hydrologic model, however, one may not be able to single out what neuron or group of neurons is responsible for representing the same soil processes. In fact, if one re-trains that ANN model with a different parameter initialization, a wholly separate group of neurons might end up being responsible for the soil processes.

Second, the statistical properties of the parameters of ML models (e.g., weights) are usually not process-informed (Mount et al., 2016). In the case of SA of mechanistic models, the inputs considered are typically model parameters, sampled by an experimental design with known statistical properties defined by the user. However, in the ML context, this is not easily doable and, for example, it is non-trivial to assign a range to the weights of an ANN for a given problem (see Kingston et al., 2005a; Razavi and Tolson, 2011). In general, the value of SA for providing insight into and extracting knowledge from ML models can be improved significantly by using state-of-the-art model development practices (Maier et al., 2010; Wu et al., 2014) that improve parameter identifiability (Guillaume et al., 2019), such as input variable selection (see Galelli et al., 2014) and model structure selection (Kingston et al., 2008), and by accounting for physical plausibility (Kingston et al., 2005b) and parameter uncertainty (Kingston et al., 2005a; 2006) explicitly during the model calibration process.

SA can also support interpretability and explainability of ML in the context of classification. In this context, a major problem is with examining and explaining the robustness of decision boundaries for classification with respect to data and/or model hypotheses. SA can provide insights into the robustness with respect to the specification of input distributions. For example, Bachoc et al. (2020a) applied a sensitivity index developed by Lemaître et al. (2015) for robustness analysis of decision boundaries in classification.

Moreover, SA is useful to provide explanations in the context of classification. Robustness of classification, for example, is subject to the decision boundaries that can be identified through ML. The decision boundaries may, for example, be sensitive to the distribution of inputs and/or model hypotheses. Specific SA methods have been successfully applied to explain the influential factors on decision boundaries and robustness analysis (e.g., Lemaître et al., 2015; Sueur et al., 2017; Bachoc et al., 2020a; Gauchy et al., 2020). SA can also identify influential inputs regarding the occurrence of critical events, which are important in the robustness assessment of decision boundaries (Raguet and Marrel, 2018; Spagnol et al., 2019; Marrel and Chabridon, 2020; Molnar, 2019).

### 3.4.3. ML-powered SA

Progress in ML undoubtedly provides fertile ground for new ideas in SA. Most notably, the ML capability to provide efficient data-driven function approximation has provided tremendous opportunities for surrogate modeling in the context of SA, when computer experiments are intensive. Example methods arising from ML that have been used in SA include Gaussian processes (Rasmussen, 2004; Yang et al., 2018), generalized polynomial chaos expansions (Sudret, 2008), reduced basis methods (Hesthaven et al., 2016) and ANNs (Beh et al., 2017). See **Section 3.3.3** for more on this subject.

Moreover, dependence measures and kernel-based indices used in ML (Gretton et al., 2005) have been introduced to the SA community by Da Veiga (2015) and further extended by De Lozzo and Marrel (2016), especially in regard to the Hilbert-Schmidt Independence Criterion (HSIC) which detects features that are non-influential on an output of interest for screening purposes. Real-world applications that nowadays use these SA methods include nuclear safety (Iooss and Marrel, 2019; Marrel and Chabridon, 2020).

The use of Shapley values (Shapley, 1953) to develop importance measures being able to deal with dependent inputs/features have emerged recently but independently in the SA (Owen, 2014) and ML (Lundberg et al., 2017) communities. Cross-fertilization of ideas between ML and SA is expected to continue and grow over time (Broto et al., 2019; Mase et al., 2020; Hoyt and Owen, 2020).

## 3.5. SA and Uncertainty Quantification

SA in the context of uncertainty quantification (UQ) has a long tradition, stemming back to works such as Bier (1982), in which global sensitivity measures were introduced to identify the key drivers of uncertainty in complex risk assessment problems. Since then, there has been a growing synergy between UQ and SA. Generally, UQ is the science of quantitative characterization and reduction of the uncertainty regarding a certain outcome of a system or model, while SA for UQ is focused on identifying the dominant controls of that uncertainty. For brevity, we do not summarize this rich history, referring to Saltelli et al. (2008), Sullivan (2015), Borgonovo (2017) and to the 'handbook of uncertainty quantification' by Ghanem et al. (2017).

Despite significant advances, SA for UQ still faces a number of challenges. These include possible misconceptions in framing an SA problem for a UQ purpose, incompatibility of some SA frameworks for some model types, complications with handling multivariate and correlated input spaces, sensitivity of UQ to problem setup, and uncertainty in the SA results themselves. In the following, we explain these challenges and possible ways to address them.

### 3.5.1. Mind the goal of UQ with respect to SA

When SA is used in a UQ application, the underlying purpose of UQ should dictate the framing of the SA problem and the method used. In general, there are two types of UQ, inverse UQ and forward UQ. The former aims to estimate unknown model parameters from data, while the latter propagates input uncertainties through a model to estimate output uncertainty. In the case of inverse UQ, SA should be used to identify the parameters most informed by data, for example by looking at the sensitivity of the misfit between the data and model predictions **(see Section 3.2.3)**. In the case of forward UQ, however, one needs to identify the factors that

influence the prediction the most - for a discussion, refer to Gupta and Razavi (2018) and Butler et al. (2020).

There is an implicit, possibly flawed, assumption in many applications of SA that the direction in factor space informed by data is *parallel* to the direction which informs predictions. This assumption can yield misleading results, as those two directions can often be *orthogonal*. For example, fixing parameters identified as non-influential by SA in the inverse UQ setting can lead to significant underestimation of prediction uncertainty. This is because those parameters, while being the largest source of uncertainty, have been ignored in forward UQ. In most cases, if the identification of a parameter is informed by data, the uncertainty around it will decrease.

In the context of UQ, a comprehensive SA practice is one that identifies both of those important directions. With such a practice, the utility of SA for UQ can be greatly improved, as it would allow for the efficient estimation of uncertain parameters and quantify predictive uncertainty simultaneously. To do so, for a given problem, SA needs to be applied in the two different settings independently; one to assess the sensitivity of a goodness-of-fit metric to the factors and the other to assess the sensitivity of the predicted quantity itself (Gupta and Razavi, 2018; Shin et al., 2013).

### 3.5.2. No single method for all model types

SA can be used for a wide variety of model types, for example, those expressed in the form of partial differential equations (PDEs, such as contaminant transport models (e.g., Wei et al., 2014), those that are linked to the solution of an optimization problem (e.g., DICE (Lamontagne et al., 2019) or STOCFOR3 (Lu et al., 2020)), or those that are in an agent-based form (Fadikar et al., 2018). Different model types are engineered in different ways, and this challenge demands systematic research that avoids encouraging an ‘apply the same hammer’ attitude with respect to methods.

For example, SA enabled with response surface surrogates **(see Section 3.3.3)** can be useful for lower dimensional problems with smooth parameter-output maps, while sampling-based approaches can be more useful for non-smooth models and higher dimensions (Becker, 2020). Moreover, to date, most SA methods have been developed for deterministic models, that is, the same input always produces the same output, while little attention has been given to models with stochastic responses such as can occur with agent-based models. While the majority of SA applications have been to assess parametric variations in agent-based models (Lee et al., 2015), an open research question is how to include alternative agent-based elements in a comprehensive SA, so that one can assess sensitivity of the response to changes in the rule of an agent simultaneously with changes in a parameter. Moreover, this question should be expanded to models in general where there is a challenge to jointly consider changes in model structure and parameters.

The exploration of global SA for optimization is a subject of recent research (Spagnol et al., 2019). Similarly, optimization problems may call for the use of information-theory-based methods. In fact, early works such as Avriel and Williams (1970) show that the information value is a natural sensitivity measure when a decision support model is cast in the form of an optimization problem. Similarly, Felli and Hazen (1999), Oakley (2009) and Strong and Oakley (2013) suggest using the information value as a sensitivity measure to explicitly compare

decision alternatives. Recently, Borgonovo et al. (2021) discuss the conditions under which global sensitivity measures can be interpreted as information value. Thus, the most important input is also the input that is most informative for the decision problem at hand. For classification, tools for low dimensional visualization of high dimensional data, (e.g., van der Maaten and Hinton, 2008), could be explored for their useability within SA. For stochastic models, we note that the literature in the management sciences has addressed the problem intensively (Rubinstein 1989; Hong 2009; Hong and Liu, 2009), and investigators in other disciplines might benefit from those results.

### 3.5.3. Multivariate and correlated input spaces

One of the most critical assumptions/decisions in UQ is the choice of the multivariate distributional properties of uncertain input variables, which are propagated through the model. In practice, the marginal probability density functions (PDFs) of the inputs are obtained via various means such as direct measurements, statistical inference, design or operation rules, and expert judgment, and can be accompanied by an estimated level of accuracy or confidence. In addition, UQ problems often come with certain *constraints* on the input, for example, when the input space is non-rectangular, and/or when the inputs are dependent. In such cases, some SA methods become handicapped, such as when the functional ANOVA expansion becomes ill-posed (Owen and Prieur 2017). In general, improper multivariate distributional properties, including the correlation structure among inputs, may lead to wrong inferences, even if the most appropriate SA method is used (Do and Razavi, 2020).

The field of SA in terms of methods to handle input constraints and correlation structures is still embryonic. Of the very few studies available, one may refer to the work of Kucherenko et al. (2017) for non-rectangular input spaces and to Kucherenko et al. (2012), Tarantola and Mara (2017), and Do and Razavi (2020) for correlated input spaces. Promising methods seem also to be moment-independent methods (Borgonovo, 2007), dependence measures (Da Veiga, 2015; De Lozzo and Marrel, 2016) and Shapley Values (Owen and Prieur, 2017), whose definitions remain well posed in the presence of input constraints. Nonetheless, the presence of constraints also impacts other aspects of SA, such as the interpretation of interactions and the assessment of direction of change. For these aspects also, further research is needed to identify the most appropriate methods.

### 3.5.4. Curse of dimensionality

The state-of-the-art models that are often encountered in UQ problems are commonly associated with high dimensionality and significant computational burden, as discussed in **Section 3.3**. Higher dimensionality exacerbates the difficulty of assigning multivariate distributions to uncertain inputs, as discussed in the previous section. A second difficulty is with ANOVA-type expansions, where the number of interaction terms is exponential in the number of inputs. Such cases require excessively large sample sizes, often becoming computationally prohibitive. A third difficulty is with the sampling strategies themselves in high-dimensional spaces. Many modern sampling strategies optimize the way samples are taken to allow a parsimonious use of the model and to maximize efficiency given the available computational budget (Pronzato and Müller, 2012; Pázman and Pronzato, 2014; Sheikholeslami and Razavi, 2017; Becker et al., 2018).

However, optimization-based sampling in high-dimensional spaces can become challenging due to the curse of dimensionality. Greedy sampling methods can be used to reduce the computational cost of optimization-based sampling methods (Oakley and O'Hagan, 2004; Maday et al., 2009; Schaback and Wendlend 2006; Jakeman et al., 2019; Harbrecht et al., 2020), but while being efficient in many cases, they can still ultimately suffer from the curse of dimensionality.

Often the lower-dimensional subspaces that impact estimates of uncertainty are efficiently described by linear (or possibly non-linear) combinations of parameters. Where SA is a means to an end, being unable to uniquely identify individual parameters is often inconsequential. By moving beyond identifying directions aligned with the axes of the parameter space, significant dimension reduction can be achieved. Ideally, SA should identify directions that are most influential. Consider a simulator that is a nonlinear function of the equally weighted sum of parameters, *y=(p1+p2)/2*. Each parameter will be found to be important but only one direction will have a non-zero influence on the function (in 2D, the x-y plane). The function will be constant in all orthogonal directions. Recently, great success has been achieved using methods such as active subspaces, which find linear rotations of the parameters which are important (Constantine, 2015).

Finding non-axial directions has been used successfully to reduce the cost of inverse UQ. These methods work by restricting resources to identifying and exploiting subspaces that are informed by data and can reduce the computational cost of the inverse problem by orders of magnitude (Tonkin and Doherty, 2005; Spantini et al., 2015). When estimating data-informed prediction uncertainty (e.g., combining forward and inverse UQ), the optimal approach is to find the directions that are both informed by data and that influence predictions. When quantifying uncertainty for linear models, these directions can be computed exactly using generalized eigenvalue decompositions. Initial work has been carried out for linear models (Lieberman and Willcox, 2014), but further research is needed, especially for non-linear models.

Kucherenko et al. (2011) showed that it is not the model 'nominal dimensions' but 'effective dimensions' that define the model complexity. In this respect, they loosely divided models into three types: (A) models with only a few important variables, (B) models with equally important variables and with dominant low-order interactions terms in their ANOVA decomposition, and (C) models with equally important variables and with dominant high-order interaction terms. They argued that type A and B models have low effective dimensions and, therefore, their handling with SA is relatively easy regardless of their nominal dimensionality.

### 3.5.5. Sensitivity of UQ to modeling choices

The assessment of how uncertainty estimates change with different modeling decisions, such as numerical discretization schemes, is important but often ignored. For example, when using numerical solutions to PDEs between two coupled models, the choice of accuracy and cost of UQ depends on the mesh and timestep size of each model, and the resolution of the coupling between the models in space and time. Identifying how sensitive a prediction is to these choices can significantly reduce the cost of UQ. For example, if the final prediction is relatively insensitive to one model, the resolution of that model and its coupling can be coarse without trading off accuracy.

The other important decision one may make is what output to choose in framing the SA problem. Indeed, various outputs can be considered, such as the mean of the model response, its variance, a probability that the output exceeds a threshold, or a quantile of the output. For example, in the context of risk or reliability analysis (Liu et al., 2006), the SA results would be different if the quantity of interest is related to the tail of the distribution of a model output such as a failure probability, a quantile or a super-quantile (Hong and Liu, 2010; Lemaître et al., 2015), compared to a case where the sensitivity of a measure of central tendency is of interest. The problem of matching the output of interest with the sensitivity measure is discussed in Borgonovo et al. (2016), where several global sensitivity measures (variance-based, moment-independent, quantile-based) are examined from an information value viewpoint. See also **Section 3.2.2** for a discussion on the framing of the SA problem from a model verification point of view. The 'goal-oriented SA' framework (Fort et al., 2016) is relevant in this context as well.

Moreover, in many applications, the probability density functions (PDFs) used to describe the uncertainty in inputs may themselves be highly uncertain (Morio, 2011). This 'second-level' of input uncertainty is often the case where there is no data available regarding the processes that those inputs control. Addressing this type of uncertainty is an essential and fruitful area of research and has recently been attracting attention. 'Input PDF robustness analysis' has been recently defined as a particular setting of SA (Lemaitre et al., 2015; Gauchy et al., 2020; Da Veiga et al., under review). Other example works include Chabridon et al. (2018) for rare-event reliability analysis, Schöbi and Sudret (2019) in the context of probability-boxes, Hart and Gremaud (2019a, 2019b) for variance-based indices, and Meynaoui et al. (2019) for Hilbert-Schmidt Independence Criterion (HSIC).

### 3.5.6. Uncertainty in SA results themselves

A major component of best practice in SA is the assessment of uncertainty in the estimates of sensitivity measures. This uncertainty is directly related to reliability and robustness of SA, as discussed in **Section 3.3**. While well-known methods such as bootstrapping (Efron, 1987) are available to provide an uncertainty estimate, it is notable that a minority of works apply this quantification systematically. Bootstrapping needs to be handled with caution as strictly the samples taken should be random (as with Monte Carlo samples), and it requires smoothness and symmetry of the bootstrap distribution, which is not always attainable. Care is also required to check if the sample size is too small to contain enough information for bootstrapping. More advanced bootstrap procedures are required if the distribution is skewed or multimodal, such as bias-corrected and accelerated bootstrap intervals (Efron, 1987).

Recent progress has seen more general bootstrap-like methods that can work well for different types of samples and sampling strategies, including bootstrapping of samples generated by Quasi-Monte Carlo or Latin hypercube sampling based on multiple independent replicates of an estimator (Owen, 2013). Heuristic approaches such as introducing 'dummy parameters' (Zadeh et al., 2017) and 'model variable augmentation' (Mai and Tolson, 2019) have also shown promise. Furthermore, future work could follow Bayesian methods for the calculation of confidence intervals on the estimates of global sensitivity measures. For example studies, refer to Oakley and O'Hagan (2004), Le Gratiet et al. (2014) and Antoniano-Villalobos et al. (2020). Further research is needed, especially in the presence of the curse of dimensionality. 'Sensitivity analysis of sensitivity analysis' has been also suggested as a way to measure the

influence of the analysis' own design parameters (Haghnegahdar and Razavi, 2017; Puy et al., 2020a; Puy et al. 2020b) and the choice of methods (Razavi and Gupta, 2015; Mora et al., 2019).

### 3.6. SA in support of Decision Making

#### 3.6.1. The deep roots of SA in the field of decision making

SA has historically but informally been a major building block of the decision making process. The notion of SA, for example, has been embedded in the classic and widely used concepts of shadow prices (Dorfman et al., 1987) and scenario analysis (Duinker and Greig, 2007; Elsawah et al., 2020b). The former, used in constrained optimization, quantifies how much more profit (i.e., objective function) one would get by increasing the amount of a resource by one unit (i.e., constraints). This practice can be viewed as one-factor-at-a-time local SA (OAT SA, one form of LSA), often on continuous variables, around an optimal point in the decision variable space. The latter, however, revolves around what-if scenarios, evaluation of policy effectiveness, analysis of causality and robustness analysis, where one or several variables at once are changed around a base case or within a factor space to evaluate change in the outcome.

A what-if scenario evaluates the effect of a change in inputs on a decision outcome, while policy effectiveness evaluates either the effect size (continuous variables) or existence of effect of a policy change (discrete variables). Analysis of causality attributes change in the output to change in inputs, and robustness analysis can either test whether a recommendation changes (Guillaume et al., 2016), or evaluate the effect of changes in factor space (McPhail et al., 2018). Robustness is particularly recognised as useful in addressing uncertainty arising from the existence of multiple plausible futures (Maier et al., 2016; Iwanaga et al., 2020), and other forms of 'deep uncertainty' (Marchau et al., 2019).

The above examples indicate how informal (and often local) SA has contributed and will continue to contribute to a variety of decision-making problems. We note that while LSA has been often criticized for being perfunctory (Saltelli and Annoni, 2010) when used to support mathematical modeling **(see Section 3.2)**, it is an essential means for many decision support systems where the users need to assess the impact of a change in a policy or the environment on the status quo. Consistent with the role and function of SA, these strong ties exist because decision making is in fact fundamentally about identifying how objectives of interest are influenced by possible interventions.

In economics, the '*ceteris paribus'* concept, a Latin phrase meaning 'all else being equal', is on the basis of a one-factor-at-a-time local SA. This concept is used in mainstream economic thinking to measure the effect of a shock to one economic variable (e.g., the price of a commodity or a set of wages) on another, provided all other variables remain the same. Economists know well that this is a crude approximation (Mirowski, 2013), so the point is the use one makes of it. For example, this approach would be good to understand the system, but poor to prescribe a policy response. Similar cautions apply to the 'what-if' scenarios described above. To address these possible limitations, formal SA has recently found its footing in decision science as described below.

### 3.6.2. Modern SA for decision making under uncertainty

More recently, formal approaches to SA, particularly for global SA, have emerged as a means to support *decision making under uncertainty*. SA can decompose uncertainty in the outcome of a decision option and attribute that to different sources of uncertainty in a decision problem. Identifying the dominant controls of uncertainty and how they interact adds transparency to the problem and guides the decision-making process towards minding the uncertainties that matter the most. Tarantola et al. (2002) first laid down a framework on how modern SA can support decision analysis, which has since gained significant momentum, by outlining SA capabilities to:

(1) Understand whether the current state of knowledge on input uncertainty is sufficient to enable a decision to be taken (Maier et al., 2016);

(2) Identify data sources or parameters that require investing resources for knowledge improvement to achieve the desired level of confidence in making a decision (Lamontagne et al., 2019);

(3) In the presence of different policy options, clarify how various uncertainty sources and their extents affect the confidence in the expected outcome of each policy (Marangoni et al., 2017);

(4) Flag models used out of context and to a degree of complexity not supported by available information for the decision problem at hand (Herman et al., 2015); and

(5) Invalidate policy assessments in cases where untested, possibly unjustified, assumptions dominantly control model outputs (Workman et al., 2020; Puy et al., 2020c).

### 3.6.3. SA and robustness of decisions under deep uncertainty

Assessment of the 'robustness' of decision alternatives is becoming increasingly important in light of 'deep uncertainty', which refers to a situation when stakeholders do not know, or cannot agree on, a system model that relates action to consequences, the probability distributions to represent uncertainty in the inputs to the model, and/or how to value the desirability of alternative outcomes (Lempert et al., 2003; Maier et al., 2016). Addressing deep uncertainty requires the identification of options that perform well over a wide range of plausible future conditions. This creates additional challenges and opportunities for the development of SA approaches in support of decision making, especially in terms of how to best perturb model inputs.

Robustness in this context refers to the *insensitivity* of a decision outcome to variation in model inputs and parameters, and in general to the assumptions made in the decision-making process. The robustness of the utility of a particular decision alternative can be quantified with the aid of robustness metrics, which use different ways to combine model outputs from different sensitivity trials, corresponding to different sets of plausible combinations of model inputs, into a single value to quantify different aspects of the performance of a decision alternative over these trials, such as best-case performance, worst-case performance, average performance, and variability in performance (see McPhail et al., 2018).

SA-based assessment of the robustness of decision alternatives under deep uncertainty requires careful consideration of the way the model input space is sampled (see McPhail et al., 2020), depending on the philosophy that underpins the robustness assessment. If the goal is to quantify robustness under as broad a range of future conditions as possible, then a large number of samples that cover the model input space as uniformly as possible is required. However, if the goal is to calculate robustness under "possible future states of the world that represent alternative plausible conditions under different assumptions" (Mahmoud et al., 2009), the number of samples used is quite small (<10), as each sample generally corresponds to a coherent narrative storyline (scenario) of an alternative hypothetical future (van Notten et al., 2005). Scenarios are plausible stories about the future of a system that is too complex to predict (Wiek et al., 2013; Elsawah et al., 2020b) and are often obtained via participatory processes involving a variety of stakeholders (e.g., Riddell et al., 2018; Wada et al., 2019; Razavi et al., 2020). It should also be noted that due to the temporal dimension associated with deep uncertainty, the samples of the model input space often correspond to time series (e.g., Guo et al., 2018; Culley et al., 2019; Riddell et al., 2019), irrespective of which philosophical approach underpins the robustness assessment.

### 3.6.4. SA and ranking of decision alternatives

SA has mainly been used to determine the sensitivity of model outputs to plausible changes in model inputs and parameters. While this can provide useful information to support decision making, it does not assess the sensitivity of the relative ranking, or preference, of different decision alternatives to potential changes in model inputs and parameters. This can be achieved by SA when focused on identifying the smallest combined changes in model inputs and parameters that result in performance / rank equivalence of two decision alternatives, which can be expressed as a distance metric. The smaller this metric, the more robust (insensitive) the relative performance / rank of a particular decision alternative and vice versa. In addition, the model inputs and parameters that have the largest influence on the relative performance / rank of decision alternatives can be identified. While such sensitivity analyses have already been applied to simulation models (e.g., Ravalico et al., 2009; 2010; Marangoni et al., 2017; Lamontagne et al., 2019; Puy et al., 2020c), as well as decision models such as multi-criteria decision analysis (Hyde et al., 2005; 2006; Herman et al., 2015; Ganji et al., 2016), they need to be developed further, especially under conditions of deep uncertainty, to ensure robust decision outcomes are achieved.

### 3.6.5. SA and qualitative aspects of decision making

SA is not only a *quantitative* paradigm but also an *epistemological* one. When used for regulation and policy making, SA must be broadened to include consideration of epistemological aspects linked to the plurality of disciplines and interested actors at play. Different norms and incommensurable values may emerge in this context. Questions such as 'what are the different narratives of a problem?', 'who is telling what story?' and 'which of these narratives are being privileged in the modeling activity carried out to support the decision-making process?' are naturally brought to the fore.

To address these issues, Saltelli et al. (2013) proposed a framework for 'sensitivity auditing'. Sensitivity auditing emphasizes the framing of a decision analysis, its institutional context, and the motivations and interests of the researchers, stakeholders and policy makers involved. An

analyst can scrutinize a mathematical model used to assist a decision-making process against the 'sensitivity auditing' checklist to cast light on potential criticalities such as: (1) rhetorical use of mathematical modeling, (2) identification of the underpinning technical assumptions, (3) uncertainty inflation or deflation, (4) unaddressed uncertainty and sensitivity of the model at the time the results are published, (5) lack of model transparency, (6) frames privileged and frames excluded, and (7) incomplete or lack of SA.

The European Commission (2015) and the European Science Academies (SAPEA, 2019) recommend sensitivity auditing in the context of modeling in support of policy making. Example applications of sensitivity auditing are found in the fields of education (OECD-PISA study, Araujo et al., 2017), food security (Saltelli and Lo Piano, 2017), public health and nutrition (Lo Piano and Robinson, 2019) and sustainability metrics (Galli et al., 2016). The seven points of sensitivity auditing are also substantially subsumed in the manifesto for responsible modeling published in Nature (Saltelli et al., 2020).

Further attempts to thoroughly capture the quality of the knowledge in a modeling activity include the model pedigree concept (Eker et al., 2018) and the 'Numeral, Unit, Spread, Assessment and Pedigree' (NUSAP) framework (Van Der Sluijs et al., 2005) for knowledge quality assessment. Incidentally, both NUSAP and sensitivity auditing are approaches belonging to the tradition of post-normal science, a style of use of science for policy that becomes relevant when facts are uncertain, values are in dispute, stakes are high and decisions are urgent (Funtowicz and Ravetz, 1993).

### 3.6.5. Revisiting the link between SA and decision making

While formal SA is finding its footing in the area of decision making, there is a need to revisit the principles of decision making to identify where and how decision theories and applications have, perhaps informally, been based on the fundamentals of SA. Such efforts could facilitate bridging the two fields and take advantage of recent advances in SA in emerging decision problems across a variety of domains, as well as correspondingly motivate advances in SA methodology for decision making. In this process, one must be mindful of the commonly asked questions by decision makers. Studies with formal SA methods often tend to answer different (often more sophisticated) questions to those related to specific quantities of interest that decision makers care most about. Therefore, to be most useful, decision makers need to be engaged in the process of co-formulating the SA problem to ensure it addresses the right question(s).

## 4. Synthesis and Concluding Remarks

The process of developing the common perspective expressed in this paper across the multidisciplinary team of authors faced interesting challenges, related principally to the various disciplinary and methodological views, as well as experiences across different application areas. That diversity promoted a synergy and more comprehensive coverage of potential opportunities to strengthen the role of SA, as summarized in the following key messages:

(a) Collective efforts are needed to structure, generalize and standardize the state of the art in SA such that it forms a distinct, cross-field discipline **(Section 3.1)**. Such efforts must emphasize: (1) teaching SA as integral to systems analysis and modeling, and

decision making; (2) developing protocols for best SA practice that are transferable across (specific) contexts and applications; and (3) launching scientific journals dedicated to SA.

(b) Much work is needed to realize the tremendous untapped potential of SA for mathematical modeling of socio-environmental and other societal problems which are confounded by uncertainty **(Section 3.2)**. SA can help with the management of uncertainty by (1) characterizing how models and the underlying real-world systems work, (2) identifying the adequate level of model complexity for the problem of interest, and (3) pointing to possible model deficiencies and non-identifiability issues, as well as where to invest to reduce critical uncertainties.

(c) Computational burden is recognized as a major hindrance to the application of SA to cases where SA can be most useful, such as for high-dimensional problems **(Section 3.3)**. Greater efforts should be directed to developing SA algorithms that are (1) more computationally efficient, (2) more statistically robust, (3) able in particular to consume 'recycled' samples, however taken, and (4) able to provide credible confidence measures on their results.

(d) The recent revitalized rise of machine learning (ML), particularly deep learning methods, could be further enhanced by formal theories of SA **(Section 3.4)**. The great potential of SA needs to be discovered for the following purposes and beyond: (1) explainability and interpretability of ML, (2) input variable selection, (3) enabling ML to work with small data, where big data sizes are not available, and (4) building trust in ML models.

(e) SA is a much needed complement and/or building block to most uncertainty quantification (UQ) practices, regardless of whether the aim is forward or inverse UQ **(Section 3.5)**. SA and UQ need to be better combined to support a variety of purposes, including: (1) apportioning uncertainty, (2) handling the curse of dimensionality, (3) addressing unknowns around the distribution of inputs and their correlation structure, and (4) assessing the sensitivity of uncertainty estimates to various choices made in the design of a UQ problem.

(f) Decision and policy making under uncertainty can significantly benefit, formally or informally, from advancements in SA, including from the notion of sensitivity auditing, that is, an extension of SA where systems models are used to support policy **(Section 3.6)**. Conversely, SA can benefit from reflecting on and formalising ways in which decision making and decision makers have previously used SA concepts informally. SA, when used in support of decision making can address critical questions, such as (1) where and how does uncertainty matter?, (2) have the impacts of all important assumptions been treated? (3) where should we invest to increase confidence in the expected outcome of a policy option?, and (4) has the policy uncertainty been artificially inflated or constrained?

All together, the above points call for more cross fertilization of different research and practice streams on SA across a wide range of disciplines. An implication of this broadening of SA is that it should be considered a 'multi-discipline' – it is a subject that is intrinsically of interest to multiple disciplines that will continue to have distinct, but ideally interconnected, literature.

Mathematicians and computer scientists are interested in more efficient ways of calculating measures, decision scientists are interested in identifying different measures, modelers/systems analysts are interested in how they can use those measures in their work, and decision makers are interested in the outputs and implications of the analyses.

SA is a vertically integrated, 'deep' topic. Those at the surface, at the highest level of abstraction, do not want to know, and do not really need to know, what is happening at the bottom - they simply apply a method already developed for their purpose. Conversely, those at the bottom produce fundamental work that does not always need to be directly responsive to immediate demands at the top - for example, the computational inefficiency of an algorithm in practice does not matter much at the development stages of new theories.

From this perspective, SA therefore needs coordination rather than consensus – we expect that multiple views and even definitions of core concepts will continue to co-exist, but the field needs to ensure that cross-fertilization of ideas continues and expands to allow different disciplines and application areas to benefit from one another despite their differences. In order to be of impact to society, it is crucial that this coordination then connects with the needs of planners, policy analysts and decision makers, with active engagement supporting the development of a shared understanding of the questions that they want answered, as well as the questions they do not yet know they want answered.

The authorship team of this perspective invites discussion and collaboration with researchers and practitioners across every area of science interested in theories, developments and applications of SA. In our vision, over the next decade SA will underpin a wide variety of activities around scientific discovery and decision support.

## Acknowledgements

This perspective paper is the outcome of a one-day workshop called "The Future of Sensitivity Analysis", which was held as a satellite event to the Ninth International Conference on Sensitivity Analysis of Model Output (SAMO), October 28-31, 2019 in Barcelona, Spain. We are thankful to the sponsors of this event, including the French research association on stochastic methods for the analysis of numerical codes (MASCOT-NUM), Open Evidence research at Universitat Oberta de Catalunya, the Joint Research Centre of the European Commission, the University of Bergen (Norway), and the French CERFACS, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique. The financial and logistic support to the lead author by the Integrated modeling Program for Canada (IMPC) under the framework of Global Water Futures (GWF) is acknowledged. Furthermore, part of the efforts leading to this paper was supported by The National Socio-Environmental Synthesis Center (SESYNC) of the United States under funding received from the National Science Foundation DBI-1639145. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government. John Jakeman's work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery

through Advanced Computing (SciDAC) program. Joseph Guillaume received funding from an Australian Research Council Discovery Early Career Award (project no. DE190100317). Arnald Puy worked on this paper on a Marie Sklodowska-Curie Global Fellowship, grant number 792178. Takuya Iwanaga is supported through an Australian Government Research Training Program (AGRTP) Scholarship and the ANU Hilda-John Endowment Fund. We would like to thanks Dan Ames, Editor-in-Chief, for insightful comments and encouragement.

## References

Adam, D., 2020. Simulating the pandemic: What COVID forecasters can learn from climate models. Nature 587, 533–534. https://doi.org/10.1038/d41586-020-03208-1

Adams, B.M., Bohnhoff, W.J., Dalbey, K., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hooper, R., Hough, P.D., Hu, K., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A., Seidl, D.T., Stephens, J.A., Swiler, L.P., Winokur, J., 2020. Dakota A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis: Version 6.12 User?s Manual. (No. SAND2020-5001). Sandia National Lab. (SNL-NM), Albuquerque, NM (United States); Sandia National Lab. (SNL-CA), Livermore, CA (United States). https://doi.org/10.2172/1630694

Addor, N., Melsen, L.A., 2019. Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models. Water Resources Research 55, 378–390. https://doi.org/10.1029/2018WR022958

Ahalt, S., Band, L., Christopherson, L., Idaszak, R., Lenhardt, C., Minsker, B., Palmer, M., Shelley, M., Tiemann, M., Zimmerman, A., 2014. Water Science Software Institute: Agile and Open Source Scientific Software Development. Computing in Science & Engineering 16, 18–26. https://doi.org/10.1109/MCSE.2014.5

AMS, 2020. MSC2020 database [WWW Document]. URL https://mathscinet.ams.org/msc/msc2020.html?t=&s=sensitivity&btn=Search&ls=s (accessed 12.7.20).

Antoniano-Villalobos, I., Borgonovo, E., Lu, X., 2020. Nonparametric estimation of probabilistic sensitivity measures. Stat Comput 30, 447–467. https://doi.org/10.1007/s11222-019-09887-9

Araujo, L., Saltelli, A., Schnepf, S.V., 2017. Do PISA data justify PISA-based education policy? International Journal of Comparative Education and Development 19, 20–34. https://doi.org/10.1108/IJCED-12-2016-0023

Avriel, M., Williams, A.C., 1970. The Value of Information and Stochastic Programming. Operations Research 18, 947–954. https://doi.org/10.1287/opre.18.5.947

AVSI, 2020. Final report - AFE 87 - Machine Learning. Aerospace Vehicle Systems Institute / Texas A&M Engineering experiment Station, https://avsi.aero/projects/current-projects/cert-of-ml-systems/afe-87-machine-learning/.

Bachoc, F., Ammar, K., Martinez, J.-M., 2016. Improvement of Code Behavior in a Design of Experiments by Metamodeling. Nuclear Science and Engineering 183. https://doi.org/10.13182/NSE15-108

Bachoc, F., Gamboa, F., Halford, M., Loubes, J.-M., Risser, L., 2020a. Entropic Variable Projection for Explainability and Intepretability. arXiv:1810.07924 [cs, stat].

Bachoc, F., Helbert, C., Picheny, V., 2020b. Gaussian process optimization with failures: classification and convergence proof. J Glob Optim 78, 483–506. https://doi.org/10.1007/s10898-020-00920-0

Banobi, J.A., Branch, T.A., Hilborn, R., 2011. Do rebuttals affect future science? Ecosphere 2, art37. https://doi.org/10.1890/ES10-00142.1

Becker, W., 2020. Metafunctions for benchmarking in sensitivity analysis. Reliability Engineering & System Safety 204, 107189. https://doi.org/10.1016/j.ress.2020.107189

Becker, W., Parulo, P., Saltelli, A., under review. Variable selection in regression models using global sensitivity analysis. Journal of Time Series Econometrics.

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, 1499–1506. https://doi.org/10.1016/j.jbiomech.2011.03.008

Becker, W.E., Tarantola, S., Deman, G., 2018. Sensitivity analysis approaches to high-dimensional screening problems at low sample size. Journal of Statistical Computation and Simulation 88, 2089–2110. https://doi.org/10.1080/00949655.2018.1450876

Beh, E.H.Y., Zheng, F., Dandy, G.C., Maier, H.R., Kapelan, Z., 2017. Robust optimization of water infrastructure planning under deep uncertainty using metamodels. Environmental Modelling & Software 93, 92–105. https://doi.org/10.1016/j.envsoft.2017.03.013

Bellprat, O., Kotlarski, S., Lüthi, D., Schär, C., 2012. Exploring Perturbed Physics Ensembles in a Regional Climate Model. Journal of Climate 25, 4582–4599. https://doi.org/10.1175/JCLI-D-11-00275.1

Benoumechiara, N., 2019. Treatment of dependency in sensitivity analysis for industrial reliability (phdthesis). Sorbonne Université ; EDF R&D.

Bier, V.M., 1982. A measure of uncertainty importance for components in fault trees (Thesis). Massachusetts Institute of Technology.

Borgonovo, E., 2017. Sensitivity Analysis: An Introduction for the Management Scientist, International Series in Operations Research & Management Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-52259-3

Borgonovo, E., 2007. A new uncertainty importance measure. Reliability Engineering & System Safety 92, 771–784. https://doi.org/10.1016/j.ress.2006.04.015

Borgonovo, E., Hazen, G.B., Jose, V.R.R., Plischke, E., 2021. Probabilistic sensitivity measures as information value. European Journal of Operational Research 289, 595–610. https://doi.org/10.1016/j.ejor.2020.07.010

Borgonovo, E., Hazen, G.B., Plischke, E., 2016. A Common Rationale for Global Sensitivity Measures and Their Estimation: A Common Rationale for Global Sensitivity Measures. Risk Analysis 36, 1871–1895. https://doi.org/10.1111/risa.12555

Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., Hill, M.C., 2017. Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. Water Resources Research 53, 7933–7950. https://doi.org/10.1002/2017WR020767

Borgonovo, E., Plischke, E., 2016. Sensitivity analysis: A review of recent advances. European Journal of Operational Research 248, 869–887. https://doi.org/10.1016/j.ejor.2015.06.032

Borgonovo, E., Smith, C.L., 2011. A Study of Interactions in the Risk Assessment of Complex Engineering Systems: An Application to Space PSA. Operations Research 59, 1461–1476.

Box, G.E.P., Meyer, R.D., 1986. An Analysis for Unreplicated Fractional Factorials. Technometrics 28, 11–18. https://doi.org/10.2307/1269599

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Broto, B., Bachoc, F., Depecker, M., 2020. Variance Reduction for Estimation of Shapley Effects and Adaptation to Unknown Input Distribution. SIAM/ASA J. Uncertainty Quantification 8, 693–716. https://doi.org/10.1137/18M1234631

Burgess, S., Bowden, J., Fall, T., Ingelsson, E., Thompson, S.G., 2017. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology 28, 30–42. https://doi.org/10.1097/EDE.0000000000000559

Butler, T., Jakeman, J.D., Wildey, T., 2020. Optimal experimental design for prediction based on push-forward probability measures. Journal of Computational Physics 416, 109518. https://doi.org/10.1016/j.jcp.2020.109518

Buzzard, G.T., 2012. Global sensitivity analysis using sparse grid interpolation and polynomial chaos. Reliability Engineering & System Safety, SAMO 2010 107, 82–89. https://doi.org/10.1016/j.ress.2011.07.011

Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling & Software, Modelling, computer-assisted simulations, and mapping of dangerous phenomena for hazard assessment 22, 1509–1518. https://doi.org/10.1016/j.envsoft.2006.10.004

Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. A unified approach. Computer Physics Communications 182, 978–988. https://doi.org/10.1016/j.cpc.2010.12.039

Casetti, E., 1999. The Evolution of Scientific Disciplines, Mathematical Modeling, and Human Geography. Geographical Analysis 31, 332–339. https://doi.org/10.1111/j.1538-4632.1999.tb00988.x

Chabridon, V., Balesdent, M., Bourinet, J.-M., Morio, J., Gayton, N., 2018. Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty. Reliability Engineering & System Safety 178, 164–178. https://doi.org/10.1016/j.ress.2018.06.008

Chalmers, I., Glasziou, P., 2009. Avoidable waste in the production and reporting of research evidence. The Lancet 374, 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Christie, M., Cliffe, A., Dawid, P., Senn, S. (Eds.), 2011. Simplicity, Complexity and Modelling: Christie/Simplicity, Complexity and Modelling. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9781119951445

Claessen, K., Hughes, J., 2000. QuickCheck: a lightweight tool for random testing of Haskell programs, in: Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming, ICFP '00. Association for Computing Machinery, New York, NY, USA, pp. 268–279. https://doi.org/10.1145/351240.351266

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

Constantine, P.G., 2015. Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies. Society for Industrial and Applied Mathematics, USA.

Crouch, S., Hong, N.C., Hettrick, S., Jackson, M., Pawlik, A., Sufi, S., Carr, L., Roure, D.D., Goble, C., Parsons, M., 2013. The Software Sustainability Institute: Changing Research Software Attitudes and Practices. Computing in Science Engineering 15, 74–80. https://doi.org/10.1109/MCSE.2013.133

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. J. Chem. Phys. 59, 3873–3878. https://doi.org/10.1063/1.1680571

Cukier, R.I., Levine, H.B., Shuler, K.E., 1978. Nonlinear sensitivity analysis of multiparameter model systems. Journal of Computational Physics 26, 1–42. https://doi.org/10.1016/0021-9991(78)90097-9

Culley, S., Bennett, B., Westra, S., Maier, H.R., 2019. Generating realistic perturbed hydrometeorological time series to inform scenario-neutral climate impact assessments. Journal of Hydrology 576, 111–122. https://doi.org/10.1016/j.jhydrol.2019.06.005

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

Da Veiga, S., Gamboa, F., Iooss, B., Prieur, C., under review. Basics and trends in sensitivity analysis. Theory and practice in R. Society for Industrial and Applied Mathematics.

Da Veiga, S., Wahl, F., Gamboa, F., 2009. Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs. Technometrics 51, 452–463. https://doi.org/10.1198/TECH.2009.08124

De Lozzo, M., Marrel, A., 2016. New improvements in the use of dependence measures for sensitivity analysis and screening. Journal of Statistical Computation and Simulation 86, 3038–3058. https://doi.org/10.1080/00949655.2016.1149854

Dell'Oca, A., Riva, M., Guadagnini, A., 2017. Moment-based metrics for global sensitivity analysis of hydrological systems. Hydrology and Earth System Sciences 21, 6219–6234. https://doi.org/10.5194/hess-21-6219-2017

Denning, P.J., Comer, D.E., Gries, D., Mulder, M.C., Tucker, A., Turner, A.J., Young, P.R., 1989. Computing as a discipline. Computer 22, 63–70. https://doi.org/10.1109/2.19833

Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A., Lek, S., 1999. Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). Ecological Modelling 120, 157–165. https://doi.org/10.1016/S0304-3800(99)00099-X

Dimopoulos, Y., Bourret, P., Lek, S., 1995. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process Lett 2, 1–4. https://doi.org/10.1007/BF02309007

Do, N.C., Razavi, S., 2020. Correlation Effects? A Major but Often Neglected Component in Sensitivity and Uncertainty Analysis. Water Resources Research 56, e2019WR025436. https://doi.org/10.1029/2019WR025436

Dorfman, R., Samuelson, P.A., Solow, R.M., 1987. Linear Programming and Economic Analysis. Courier Corporation.

Douglas-Smith, D., Iwanaga, T., Croke, B.F.W., Jakeman, A.J., 2020. Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. Environmental Modelling & Software 124, 104588. https://doi.org/10.1016/j.envsoft.2019.104588

Duinker, P.N., Greig, L.A., 2007. Scenario analysis in environmental impact assessment: Improving explorations of the future. Environmental Impact Assessment Review 27, 206–219. https://doi.org/10.1016/j.eiar.2006.11.001

Edwards, M.A., Roy, S., 2017. Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. Environmental Engineering Science 34, 51–61. https://doi.org/10.1089/ees.2016.0223

Efron, B., 1987. Better Bootstrap Confidence Intervals. Journal of the American Statistical Association 82, 171–185. https://doi.org/10.2307/2289144

Eker, S., Rovenskaya, E., Obersteiner, M., Langan, S., 2018. Practice and perspectives in the validation of resource management models. Nature Communications 9, 5359. https://doi.org/10.1038/s41467-018-07811-9

Elsawah, S., Filatova, T., Jakeman, A.J., Kettner, A.J., Zellner, M.L., Athanasiadis, I.N., Hamilton, S.H., Axtell, R.L., Brown, D.G., Gilligan, J.M., Janssen, M.A., Robinson, D.T., Rozenberg, J., Ullah, I.I.T., Lade, S.J., 2020a. Eight grand challenges in socio-environmental systems modeling. 1 2, 16226–16226. https://doi.org/10.18174/sesmo.2020a16226

Elsawah, S., Hamilton, S.H., Jakeman, A.J., Rothman, D., Schweizer, V., Trutnevyte, E., Carlsen, H., Drakes, C., Frame, B., Fu, B., Guivarch, C., Haasnoot, M., Kemp-Benedict, E., Kok, K., Kosow, H., Ryan, M., van Delden, H., 2020b. Scenario processes for socio-environmental systems analysis of futures: A review of recent efforts and a salient research agenda for supporting decision making. Science of The Total Environment 729, 138393. https://doi.org/10.1016/j.scitotenv.2020.138393

Engelbrecht, A.P., Cloete, I., Zurada, J.M., 1995. Determining the significance of input parameters using sensitivity analysis, in: Mira, J., Sandoval, F. (Eds.), From Natural to Artificial Neural Computation, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 382–388. https://doi.org/10.1007/3-540-59497-3_199

European Commission, 2015. Better regulation toolbox [WWW Document]. European Commission - European Commission. URL https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox/better-regulation-toolbox_en (accessed 12.7.20).

Fadikar, A., Higdon, D., Chen, J., Lewis, B., Venkatramanan, S., Marathe, M., 2018. Calibrating a Stochastic, Agent-Based Model Using Quantile-Based Emulation. SIAM/ASA J. Uncertainty Quantification 6, 1685–1706. https://doi.org/10.1137/17M1161233

Fang, K.-T., Li, R., Sudjianto, A., Li, R., Sudjianto, A., 2005. Design and Modeling for Computer Experiments. Chapman and Hall/CRC. https://doi.org/10.1201/9781420034899

Felli, J.C., Hazen, G.B., 1999. A Bayesian approach to sensitivity analysis. Health Economics 8, 263–268. https://doi.org/10.1002/(SICI)1099-1050(199905)8:3<263::AID-HEC426>3.0.CO;2-S

Fernández-Navarro, F., Carbonero-Ruz, M., Becerra Alonso, D., Torres-Jiménez, M., 2017. Global Sensitivity Estimates for Neural Network Classifiers. IEEE Transactions on Neural Networks and Learning Systems 28, 2592–2604. https://doi.org/10.1109/TNNLS.2016.2598657

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

Fisher, R.A., 1953. The Design of Experiments. Oliver and Boyd.

Fock, E., 2014. Global Sensitivity Analysis Approach for Input Selection and System Identification Purposes—A New Framework for Feedforward Neural Networks. IEEE Transactions on Neural Networks and Learning Systems 25, 1484–1495. https://doi.org/10.1109/TNNLS.2013.2294437

Fort, J.-C., Klein, T., Rachdi, N., 2016. New sensitivity analysis subordinated to a contrast. Communications in Statistics - Theory and Methods 45, 4349–4364. https://doi.org/10.1080/03610926.2014.901369

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. The Annals of Statistics 19, 1–67. https://doi.org/10.1214/aos/1176347963

Funtowicz, S.O., Ravetz, J.R., 1993. Science for the post-normal age. Futures 25, 739–755. https://doi.org/10.1016/0016-3287(93)90022-L

Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-009-0621-1

Galelli, S., Humphrey, G.B., Maier, H.R., Castelletti, A., Dandy, G.C., Gibbs, M.S., 2014. An evaluation framework for input variable selection algorithms for environmental data-driven models. Environmental Modelling & Software 62, 33–51. https://doi.org/10.1016/j.envsoft.2014.08.015

Galli, A., Giampietro, M., Goldfinger, S., Lazarus, E., Lin, D., Saltelli, A., Wackernagel, M., Müller, F., 2016. Questioning the Ecological Footprint. Ecological Indicators 69, 224–232. https://doi.org/10.1016/j.ecolind.2016.04.014

Gamboa, F., Gremaud, P., Klein, T., Lagnoux, A., 2020. Global Sensitivity Analysis: a new generation of mighty estimators based on rank statistics. arXiv:2003.01772 [math, stat].

Gamboa, F., Janon, A., Klein, T., Lagnoux, A., Prieur, C., 2016. Statistical inference for Sobol pick-freeze Monte Carlo method. Statistics 50, 881–902. https://doi.org/10.1080/02331888.2015.1105803

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. Environmental Modelling & Software 51, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Ganji, A., Maier, H.R., Dandy, G.C., 2016. A modified Sobol′ sensitivity analysis method for decision-making in environmental problems. Environmental Modelling & Software 75, 15–27. https://doi.org/10.1016/j.envsoft.2015.10.001

Gauchy, C., Stenger, J., Sueur, R., Iooss, B., 2020. An information geometry approach for robustness analysis in uncertainty quantification of computer codes. hal-02425477v2 https://hal.archives-ouvertes.fr/hal-02425477.

Ghanem, R., 2017. Handbook of uncertainty quantification. Springer Berlin Heidelberg, New York, NY.

Gigerenzer, G., 2018. Statistical Rituals: The Replication Delusion and How We Got There. Advances in Methods and Practices in Psychological Science 1, 198–218. https://doi.org/10.1177/2515245918771329

Gigerenzer, G., Marewski, J.N., 2015. Surrogate Science: The Idol of a Universal Method for Scientific Inference. Journal of Management 41, 421–440. https://doi.org/10.1177/0149206314547522

Gilquin, L., Arnaud, E., Prieur, C., Monod, H., 2016. Recursive estimation procedure of Sobol' indices based on replicated designs.

Gilquin, L., Capelle, T., Arnaud, E., Prieur, C., 2017a. Sensitivity Analysis and Optimisation of a Land Use and Transport Integrated Model | Journal de la Société Française de Statistique. Computer Experiments, Uncertainty and Sensitivity Analysis 158.

Gilquin, L., Jiménez Rugama, L.A., Arnaud, É., Hickernell, F.J., Monod, H., Prieur, C., 2017b. Iterative construction of replicated designs based on Sobol' sequences. Comptes Rendus Mathematique 355, 10–14. https://doi.org/10.1016/j.crma.2016.11.013

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning, Adaptive computation and machine learning. The MIT Press, Cambridge, Massachusetts.

Gramacy, R.B., Taddy, M.A., 2010. Categorical Inputs, Sensitivity Analysis, Optimization and Importance Tempering with tgp Version 2, an R Package for Treed Gaussian Process Models. Journal of Statistical Software 33, 1–48. https://doi.org/10.18637/jss.v033.i06

Gregorutti, B., 2015. Forêts aléatoires et sélection de variables : analyse des données des enregistreurs de vol pour la sécurité aérienne (phdthesis). Université Pierre et Marie Curie - Paris VI.

Gretton, A., Bousquet, O., Smola, A., Schölkopf, B., 2005. Measuring Statistical Dependence with Hilbert-Schmidt Norms, in: Jain, S., Simon, H.U., Tomita, E. (Eds.), Algorithmic Learning Theory, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 63–77. https://doi.org/10.1007/11564089_7

Guillaume, J.H.A., Arshad, M., Jakeman, A.J., Jalava, M., Kummu, M., 2016. Robust discrimination between uncertain management alternatives by iterative reflection on crossover point scenarios: Principles, design and implementations. Environmental Modelling & Software 83, 326–343. https://doi.org/10.1016/j.envsoft.2016.04.005

Guillaume, J.H.A., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., Stigter, J.D., 2019. Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. Environmental Modelling & Software 119, 418–432. https://doi.org/10.1016/j.envsoft.2019.07.007

Guo, D., Westra, S., Maier, H.R., 2018. An inverse approach to perturb historical rainfall data for scenario-neutral climate impact studies. Journal of Hydrology 556, 877–890. https://doi.org/10.1016/j.jhydrol.2016.03.025

Guo, L., Meng, Z., Sun, Y., Wang, L., 2016. Parameter identification and sensitivity analysis of solar cell models with cat swarm optimization algorithm. Energy Conversion and Management 108, 520–528. https://doi.org/10.1016/j.enconman.2015.11.041

Gupta, H., Razavi, S., 2017. Chapter 20 - Challenges and Future Outlook of Sensitivity Analysis, in: Petropoulos, G.P., Srivastava, P.K. (Eds.), Sensitivity Analysis in Earth Observation Modelling. Elsevier, pp. 397–415. https://doi.org/10.1016/B978-0-12-803011-0.00020-3

Gupta, H.V., Razavi, S., 2018. Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models. Water Resources Research 54, 8692–8717. https://doi.org/10.1029/2018WR022668

Haghnegahdar, A., Razavi, S., 2017. Insights into sensitivity analysis of Earth and environmental systems models: On the impact of parameter perturbation scale. Environmental Modelling & Software 95, 115–131. https://doi.org/10.1016/j.envsoft.2017.03.031

Haghnegahdar, A., Razavi, S., Yassin, F., Wheater, H., 2017. Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. Hydrological Processes 31, 4462–4476. https://doi.org/10.1002/hyp.11358

Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. Environ Monit Assess 32, 135–154. https://doi.org/10.1007/BF00547132

Hannay, J.E., MacLeod, C., Singer, J., Langtangen, H.P., Pfahl, D., Wilson, G., 2009. How do scientists develop and use scientific software?, in: 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering. IEEE, pp. 1–8. https://doi.org/10.1109/SECSE.2009.5069155

Hapfelmeier, A., Hothorn, T., Ulm, K., Strobl, C., 2014. A new variable importance measure for random forests with missing data. Stat Comput 24, 21–34. https://doi.org/10.1007/s11222-012-9349-1

Harbrecht, H., Jakeman, J.D., Zaspel, P., 2020. Cholesky-based experimental design for Gaussian process and kernel-based emulation and calibration [WWW Document]. URL https://edoc.unibas.ch/79042/ (accessed 12.7.20).

Hart, J.L., Gremaud, P.A., 2019a. Robustness of the Sobol' Indices to Distributional Uncertainty. IJUQ 9. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019030553

Hart, J.L., Gremaud, P.A., 2019b. Robustness of the Sobol' Indices to Marginal Distribution Uncertainty. SIAM/ASA J. Uncertainty Quantification 7, 1224–1244. https://doi.org/10.1137/18M123387X

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction, 2nd ed. ed, Springer series in statistics. Springer, New York, NY.

Herman, J.D., Reed, P.M., Wagener, T., 2013. Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. Water Resources Research 49, 1400–1414. https://doi.org/10.1002/wrcr.20124

Herman, J.D., Reed, P.M., Zeff, H.B., Characklis, G.W., 2015. How Should Robustness Be Defined for Water Systems Planning under Change? Journal of Water Resources Planning and Management 141, 04015012. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000509

Herman, J.D., Usher, W., 2017. SALib: An open-source Python library for Sensitivity Analysis [WWW Document]. The Journal of Open Source Software. https://doi.org/10.21105/joss.00097

Hesthaven, J.S., Rozza, G., Stamm, B., 2016. Certified Reduced Basis Methods for Parametrized Partial Differential Equations, SpringerBriefs in Mathematics. Springer International Publishing. https://doi.org/10.1007/978-3-319-22470-1

Hoeffding, W., 1948. A Class of Statistics with Asymptotically Normal Distribution. Ann. Math. Statist. 19, 293–325. https://doi.org/10.1214/aoms/1177730196

Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety 52, 1–17. https://doi.org/10.1016/0951-8320(96)00002-6

Hong, L.J., 2009. Estimating Quantile Sensitivities. Operations Research 57, 118–130. https://doi.org/10.1287/opre.1080.0531

Hong, L.J., Liu, G., 2010. Pathwise Estimation of Probability Sensitivities Through Terminating or Steady-State Simulations. Operations Research 58, 357–370. https://doi.org/10.1287/opre.1090.0739

Hong, L.J., Liu, G., 2009. Simulating Sensitivities of Conditional Value at Risk. Management Science 55, 281–293. https://doi.org/10.1287/mnsc.1080.0901

Hornberger, G.M., Spear, R.C., 1981. Approach to the preliminary analysis of environmental systems. J. Environ. Manage.; (United States) 12:1.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S., 2019. Parameter-Efficient Transfer Learning for NLP, in: Chaudhuri, K., Salakhutdinov, R. (Eds.), Proceedings of Machine Learning Research. PMLR, Long Beach, California, USA, pp. 2790–2799.

Hoyt, C., Owen, A.B., 2020. Efficient estimation of the ANOVA mean dimension, with an application to neural net classification. arXiv:2007.01281 [cs, math, stat].

Humphrey, G.B., Maier, H.R., Wu, W., Mount, N.J., Dandy, G.C., Abrahart, R.J., Dawson, C.W., 2017. Improved validation framework and R-package for artificial neural network models. Environmental Modelling & Software 92, 82–106. https://doi.org/10.1016/j.envsoft.2017.01.023

Hut, R.W., van de Giesen, N.C., Drost, N., 2017. Comment on "Most computational hydrology is not reproducible, so is it really science?" by Christopher Hutton et al.: Let hydrologists learn the latest computer science by working with Research Software Engineers (RSEs) and not reinvent the waterwheel our. Water Resources Research. https://doi.org/10.1002/2017WR020665

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science? Water Resources Research 52, 7548–7555. https://doi.org/10.1002/2016WR019285

Hyde, K.M., Maier, H.R., 2006. Distance-based and stochastic uncertainty analysis for multi-criteria decision analysis in Excel using Visual Basic for Applications. Environmental Modelling & Software 21, 1695–1710. https://doi.org/10.1016/j.envsoft.2005.08.004

Hyde, K.M., Maier, H.R., Colby, C.B., 2005. A distance-based uncertainty analysis approach to multi-criteria decision analysis for water resource decision making. Journal of Environmental Management, Integrative modelling for sustainable water allocation 77, 278–290. https://doi.org/10.1016/j.jenvman.2005.06.011

Ioannidis, J.P.A., 2005. Why Most Published Research Findings Are False. PLOS Medicine 2, e124. https://doi.org/10.1371/journal.pmed.0020124

Iooss, B., Da Veiga, S., Janon, A., Pujol, G., 2018. sensitivity: Global Sensitivity Analysis of Model Outputs, R package version 1.22.0. https://CRAN.R-project.org/package=sensitivity.

Iooss, B., Lemaître, P., 2015. A Review on Global Sensitivity Analysis Methods, in: Dellino, G., Meloni, C. (Eds.), Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications, Operations Research/Computer Science Interfaces Series. Springer US, Boston, MA, pp. 101–122. https://doi.org/10.1007/978-1-4899-7547-8_5

Iooss, B., Marrel, A., 2019. Advanced Methodology for Uncertainty Propagation in Computer Experiments with Large Number of Inputs. Nuclear Technology 205, 1588–1606. https://doi.org/10.1080/00295450.2019.1573617

Iooss, B., Prieur, C., 2019. Shapley Effects for Sensitivity Analysis with Correlated Inputs: Comparisons with Sobol' Indices, Numerical Estimation and Applications. IJUQ 9. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019028372

IPCC, 2016. 8.1.2.1 Model Intercomparisons and Ensembles - AR4 WGI Chapter 8: Climate Models and their Evaluation. https://archive.ipcc.ch/publications_and_data/ar4/wg1/en/ch8s8-1-2-1.html.

Iwanaga, T., Partington, D., Ticehurst, J., Croke, B.F.W., Jakeman, A.J., 2020. A socio-environmental model for exploring sustainable water management futures: Participatory and collaborative modelling in the Lower Campaspe catchment. Journal of Hydrology: Regional Studies 28, 100669. https://doi.org/10.1016/j.ejrh.2020.100669

Iwanaga, T., Wang, H.-H., Hamilton, S.H., Grimm, V., Koralewski, T.E., Salado, A., Elsawah, S., Razavi, S., Yang, J., Glynn, P., Badham, J., Voinov, A., Chen, M., Grant, W.E., Peterson, T.R., Frank, K., Shenk, G., Barton, C.M., Jakeman, A.J., Little, J.C., 2021. Socio-technical scales in socio-environmental modeling: managing a system-of-systems modeling approach. Environmental Modelling & Software 104885. https://doi.org/10.1016/j.envsoft.2020.104885

Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? Water Resources Research 29, 2637–2649. https://doi.org/10.1029/93WR00877

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21, 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004

Jakeman, J.D., Eldred, M.S., Geraci, G., Gorodetsky, A., 2020. Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis. International Journal for Numerical Methods in Engineering 121, 1314–1343. https://doi.org/10.1002/nme.6268

Jakeman, J.D., Franzelin, F., Narayan, A., Eldred, M., Plfüger, D., 2019. Polynomial chaos expansions for dependent random variables. Computer Methods in Applied Mechanics and Engineering 351, 643–666. https://doi.org/10.1016/j.cma.2019.03.049

Janon, A., Klein, T., Lagnoux, A., Nodet, M., Prieur, C., 2014a. Asymptotic normality and efficiency of two Sobol index estimators. ESAIM: PS 18, 342–364. https://doi.org/10.1051/ps/2013040

Janon, A., Nodet, M., Prieur, C., 2014b. Uncertainties Assessment in global sensitivity indices estimation from metamodels. International Journal for Uncertainty Quantification 4, 21–36. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2012004291

Jones, D.R., 2001. A Taxonomy of Global Optimization Methods Based on Response Surfaces. Journal of Global Optimization 21, 345–383. https://doi.org/10.1023/A:1012771025575

Kambhatla, N., Leen, T.K., 1997. Dimension Reduction by Local Principal Component Analysis. Neural Computation 9, 1493–1516. https://doi.org/10.1162/neco.1997.9.7.1493

Kavetski, D., Clark, M.P., 2010. Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction. Water Resources Research 46. https://doi.org/10.1029/2009WR008896

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S., Nosek, B.A., 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology 14, e1002456. https://doi.org/10.1371/journal.pbio.1002456

Kingston, G.B., Lambert, M.F., Maier, H.R., 2005a. Bayesian training of artificial neural networks used for water resources modeling. Water Resources Research 41. https://doi.org/10.1029/2005WR004152

Kingston, G.B., Maier, H.R., Lambert, M.F., 2008. Bayesian model selection applied to artificial neural networks used for water resources modeling. Water Resources Research 44. https://doi.org/10.1029/2007WR006155

Kingston, G.B., Maier, H.R., Lambert, M.F., 2006. A probabilistic method for assisting knowledge extraction from artificial neural networks used for hydrological prediction. Mathematical and Computer Modelling, Application of Natural Computing Methods to Water Resources and Environmental Modelling 44, 499–512. https://doi.org/10.1016/j.mcm.2006.01.008

Kingston, G.B., Maier, H.R., Lambert, M.F., 2005b. Calibration and validation of neural networks to ensure physically plausible hydrological modeling. Journal of Hydrology 314, 158–176. https://doi.org/10.1016/j.jhydrol.2005.03.013

Kleijnen, J.P.C., 1995. Sensitivity analysis and optimization of system dynamics models: Regression analysis and statistical design of experiments. System Dynamics Review 11, 275–288. https://doi.org/10.1002/sdr.4260110403

Klemeš, V., 1986. Dilettantism in hydrology: Transition or destiny? Water Resources Research 22, 177S-188S. https://doi.org/10.1029/WR022i09Sp0177S

Koo, H., Chen, M., Jakeman, A.J., Zhang, F., 2020a. A global sensitivity analysis approach for identifying critical sources of uncertainty in non-identifiable, spatially distributed environmental models: A holistic analysis applied to SWAT for input datasets and model parameters. Environmental Modelling & Software 127, 104676. https://doi.org/10.1016/j.envsoft.2020.104676

Koo, H., Iwanaga, T., Croke, B.F.W., Jakeman, A.J., Yang, J., Wang, H.-H., Sun, X., Lü, G., Li, X., Yue, T., Yuan, W., Liu, X., Chen, M., 2020b. Position Paper: Sensitivity Analysis of Spatially Distributed Environmental Models- A pragmatic framework for the exploration of uncertainty sources. Environmental Modelling & Software 104857. https://doi.org/10.1016/j.envsoft.2020.104857

Krzykacz-Hausmann, B., 2001. Epistemic sensitivity analysis based on the concept of entropy, in: Proceedings of the 3rd International Conference on Sensitivity Analysis of Model Output. Presented at the SAMO 2001, Madrid, Spain, pp. 31–35.

Kucherenko, S., Albrecht, D., Saltelli, A., 2015. Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques. arXiv:1505.02350 [stat].

Kucherenko, S., Feil, B., Shah, N., Mauntz, W., 2011. The identification of model effective dimensions using global sensitivity analysis. Reliability Engineering & System Safety 96, 440–449. https://doi.org/10.1016/j.ress.2010.11.003

Kucherenko, S., Klymenko, O.V., Shah, N., 2017. Sobol' indices for problems defined in non-rectangular domains. Reliability Engineering & System Safety 167, 218–231. https://doi.org/10.1016/j.ress.2017.06.001

Kucherenko, S., Tarantola, S., Annoni, P., 2012. Estimation of global sensitivity indices for models with dependent variables. Computer Physics Communications 183, 937–946. https://doi.org/10.1016/j.cpc.2011.12.020

Kucherenko, S., Zaccheus, O., 2020. SobolGSA. http://www.imperial.ac.uk/a-z-research/process-systems-engineering/research/free-software/sobolgsa-software/.

Lakshmanan, V., Karstens, C., Krause, J., Elmore, K., Ryzhkov, A., Berkseth, S., 2015. Which Polarimetric Variables Are Important for Weather/No-Weather Discrimination? Journal of Atmospheric and Oceanic Technology 32, 1209–1223. https://doi.org/10.1175/JTECH-D-13-00205.1

Lamboni, M., Iooss, B., Popelin, A.-L., Gamboa, F., 2013. Derivative-based global sensitivity measures: General links with Sobol' indices and numerical tests. Mathematics and Computers in Simulation 87, 45–54. https://doi.org/10.1016/j.matcom.2013.02.002

Lamontagne, J.R., Reed, P.M., Marangoni, G., Keller, K., Garner, G.G., 2019. Robust abatement pathways to tolerable climate futures require immediate global action. Nature Climate Change 9, 290–294. https://doi.org/10.1038/s41558-019-0426-8

Lauret, P., Fock, E., Mara, T.A., 2006. A Node Pruning Algorithm Based on a Fourier Amplitude Sensitivity Test Method. IEEE Transactions on Neural Networks 17, 273–293. https://doi.org/10.1109/TNN.2006.871707

Le Gratiet, L., Cannamela, C., Iooss, B., 2014. A Bayesian Approach for Global Sensitivity Analysis of (Multifidelity) Computer Codes. SIAM/ASA Journal on Uncertainty Quantification 2, 336–363. https://doi.org/10.1137/130926869

Lee, J., Kim, R., Koh, Y., Kang, J., 2019. Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network. IEEE Access 7, 167260–167277. https://doi.org/10.1109/ACCESS.2019.2953542

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, G., Sun, Z., Parker, D.C., 2015. The Complexities of Agent-Based Modeling Output Analysis. Journal of Artificial Societies and Social Simulation 18, 4. https://doi.org/10.18564/jasss.2897

Leek, J., McShane, B.B., Gelman, A., Colquhoun, D., Nuijten, M.B., Goodman, S.N., 2017. Five ways to fix statistics. Nature 551, 557–559. https://doi.org/10.1038/d41586-017-07522-z

Lek, S., Belaud, A., Dimopoulos, I., Lauga, J., Moreau, J., 1995. Improved estimation, using neural networks, of the food consumption of fish populations. Mar. Freshwater Res. 46, 1229–1236. https://doi.org/10.1071/mf9951229

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecological Modelling 90, 39–52. https://doi.org/10.1016/0304-3800(95)00142-5

Lemaître, P., Sergienko, E., Arnaud, A., Bousquet, N., Gamboa, F., Iooss, B., 2015. Density modification-based reliability sensitivity analysis. Journal of Statistical Computation and Simulation 85, 1200–1223. https://doi.org/10.1080/00949655.2013.873039

Lempert, R.J., Popper, S.W., Bankes, S.C., 2003. Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis.

Lieberman, C., Willcox, K., 2014. Nonlinear Goal-Oriented Bayesian Inference: Application to Carbon Capture and Storage. SIAM Journal on Scientific Computing 36, B427–B449. https://doi.org/10.1137/130928315

Liong, S.-Y., Lim, W.-H., Paudyal, G.N., 2000. River Stage Forecasting in Bangladesh: Neural Network Approach. Journal of Computing in Civil Engineering 14, 1–8. https://doi.org/10.1061/(ASCE)0887-3801(2000)14:1(1)

Liu, H., Chen, W., Sudjianto, A., 2006. Relative Entropy Based Method for Probabilistic Sensitivity Analysis in Engineering Design. J. Mech. Des 128, 326–336. https://doi.org/10.1115/1.2159025

Lo Piano, S., Robinson, M., 2019. Nutrition and public health economic evaluations under the lenses of post normal science. Futures 112, 102436. https://doi.org/10.1016/j.futures.2019.06.008

Lu, X., Rudi, A., Borgonovo, E., Rosasco, L., 2020. Faster Kriging: Facing High-Dimensional Simulators. Operations Research 68, 233–249. https://doi.org/10.1287/opre.2019.1860

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774.

Maaten, L. van der, Hinton, G., 2008. Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579–2605.

Maday, Y., Nguyen, N.C., Patera, A.T., Pau, S.H., 2009. A general multipurpose interpolation procedure: the magic points. Communications on Pure & Applied Analysis 8, 383. https://doi.org/10.3934/cpaa.2009.8.383

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., White, D., Winter, L., 2009. A formal framework for scenario development in support of environmental decision-making. Environmental Modelling & Software 24, 798–808. https://doi.org/10.1016/j.envsoft.2008.11.010

Mai, J., Tolson, B.A., 2019. Model Variable Augmentation (MVA) for Diagnostic Assessment of Sensitivity Analysis Results. Water Resources Research 55, 2631–2651. https://doi.org/10.1029/2018WR023382

Maier, H.R., Dandy, G.C., 1997. Determining Inputs for Neural Network Models of Multivariate Time Series. Computer-Aided Civil and Infrastructure Engineering 12, 353–368. https://doi.org/10.1111/0885-9507.00069

Maier, H.R., Guillaume, J.H.A., van Delden, H., Riddell, G.A., Haasnoot, M., Kwakkel, J.H., 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? Environmental Modelling & Software 81, 154–164. https://doi.org/10.1016/j.envsoft.2016.03.014

Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software 25, 891–909. https://doi.org/10.1016/j.envsoft.2010.02.003

Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D.P., Vrugt, J.A., Zecchin, A.C., Minsker, B.S., Barbour, E.J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., Reed, P.M., 2014. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. Environmental Modelling & Software 62, 271–299. https://doi.org/10.1016/j.envsoft.2014.09.013

Maier, H.R., Razavi, S., Kapelan, Z., Matott, L.S., Kasprzyk, J., Tolson, B.A., 2019. Introductory overview: Optimization using evolutionary algorithms and other metaheuristics. Environmental Modelling & Software 114, 195–213. https://doi.org/10.1016/j.envsoft.2018.11.018

Marangoni, G., Tavoni, M., Bosetti, V., Borgonovo, E., Capros, P., Fricko, O., Gernaat, D.E.H.J., Guivarch, C., Havlik, P., Huppmann, D., Johnson, N., Karkatsoulis, P., Keppo, I., Krey, V., Ó Broin, E., Price, J., van Vuuren, D.P., 2017. Sensitivity of projected long-term CO2 emissions across the Shared Socioeconomic Pathways. Nature Climate Change 7, 113–117. https://doi.org/10.1038/nclimate3199

Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M., Popper, S.W. (Eds.), 2019. Decision Making under Deep Uncertainty: From Theory to Practice. Springer International Publishing. https://doi.org/10.1007/978-3-030-05252-2

Marelli, S., Sudret, B., 2014. UQLab: A Framework for Uncertainty Quantification in Matlab, in: Vulnerability, Uncertainty, and Risk. American Society of Civil Engineers, Liverpool, UK, pp. 2554–2563. https://doi.org/10.1061/9780784413609.257

Marrel, A., Chabridon, V., 2020. Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor.

Marrel, A., Iooss, B., Laurent, B., Roustant, O., 2009. Calculations of Sobol indices for the Gaussian process metamodel. Reliability Engineering & System Safety 94, 742–751. https://doi.org/10.1016/j.ress.2008.07.008

Mase, M., Owen, A.B., Seiler, B., 2020. Explaining black box decisions by Shapley cohort refinement. arXiv:1911.00467 [cs, econ, stat].

Maxwell, R.M., Miller, N.L., 2005. Development of a coupled land surface and groundwater model. Journal of Hydrometeorology 6–233.

McCurley, K.L., Jawitz, J.W., 2017. Hyphenated hydrology: Interdisciplinary evolution of water resource science. Water Resources Research 53, 2972–2982. https://doi.org/10.1002/2016WR019835

McPhail, C., Maier, H.R., Kwakkel, J.H., Giuliani, M., Castelletti, A., Westra, S., 2018. Robustness Metrics: How Are They Calculated, When Should They Be Used and Why Do They Give Different Results? Earth's Future 6, 169–191. https://doi.org/10.1002/2017EF000649

McPhail, C., Maier, H.R., Westra, S., Kwakkel, J.H., Linden, L., 2020. Impact of Scenario Selection on Robustness. Water Resources Research 56. https://doi.org/10.1029/2019WR026515

Mendoza, P.A., Clark, M.P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., Gupta, H., 2015. Are we unnecessarily constraining the agility of complex process-based models? Water Resources Research 51, 716–728. https://doi.org/10.1002/2014WR015820

Meynaoui, A., Marrel, A., Laurent, B., 2019. New statistical methodology for second level global sensitivity analysis.

Mirowski, P., 2013. Never Let a Serious Crisis Go to Waste, 1st edition. ed. Verso Trade, London ; New York.

Molnar, C., 2019. Interpretable Machine Learning. https://christophm.github.io/interpretable-ml-book/.

Mora, E.B., Spelling, J., van der Weijde, A.H., 2019. Benchmarking the PAWN distribution-based method against the variance-based method in global sensitivity analysis: Empirical results. Environmental Modelling & Software 122, 104556. https://doi.org/10.1016/j.envsoft.2019.104556

Morio, J., 2011. Influence of input PDF parameters of a model on a failure probability estimation. Simulation Modelling Practice and Theory 19, 2244–2255. https://doi.org/10.1016/j.simpat.2011.08.003

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33, 161–174. https://doi.org/10.1080/00401706.1991.10484804

Mount, N.J., Dawson, C.W., Abrahart, R.J., 2013. Legitimising data-driven models: exemplification of a new data-driven mechanistic modelling framework. Hydrology and Earth System Sciences 17, 2827–2843. https://doi.org/10.5194/hess-17-2827-2013

Mount, N.J., Maier, H.R., Toth, E., Elshorbagy, A., Solomatine, D., Chang, F.-J., Abrahart, R.J., 2016. Data-driven modelling approaches for socio-hydrology: opportunities and challenges within the Panta Rhei Science Plan. Hydrological Sciences Journal 1–17. https://doi.org/10.1080/02626667.2016.1159683

National Research Council, 2012. Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. https://doi.org/10.17226/13395

Nearing, G.S., Gupta, H.V., 2018. Ensembles vs. information theory: supporting science under uncertainty. Frontiers of Earth Science 12, 653–660. https://doi.org/10.1007/s11707-018-0709-9

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

Oakley, J.E., 2009. Decision-Theoretic Sensitivity Analysis for Complex Computer Models. Technometrics 51, 121–129. https://doi.org/10.1198/TECH.2009.0014

Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66, 751–769. https://doi.org/10.1111/j.1467-9868.2004.05304.x

Obermeyer, Z., Emanuel, E.J., 2016. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 375, 1216–1219. https://doi.org/10.1056/NEJMp1606181

O'Neill, R.V., 1973. Error analysis of ecological models, in: Radionuclides in Ecosystems. National Technical Information Service, Springfield, pp. 898–908.

Openmod, 2020. Open Energy Modelling Initiative [WWW Document]. URL http://openmod-initiative.org/manifesto.html (accessed 12.7.20).

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

Owen, A.B., 2014. Sobol' Indices and Shapley Value. SIAM/ASA Journal on Uncertainty Quantification 2, 245–251. https://doi.org/10.1137/130936233

Owen, A.B., 2013. Monte Carlo theory, methods and examples.

Owen, A.B., Prieur, C., 2017. On Shapley Value for Measuring Importance of Dependent Inputs. SIAM/ASA Journal on Uncertainty Quantification 5, 986–1002. https://doi.org/10.1137/16M1097717

Palar, P.S., Zuhal, L.R., Shimoyama, K., Tsuchiya, T., 2018. Global sensitivity analysis via multi-fidelity polynomial chaos expansion. Reliability Engineering & System Safety 170, 175–190. https://doi.org/10.1016/j.ress.2017.10.013

Pázman, A., Pronzato, L., 2014. Optimum design accounting for the global nonlinear behavior of the model. The Annals of Statistics 42, 1426–1451. https://doi.org/10.1214/14-AOS1232

Pearson, K., 1905. On the General Theory of Skew Correlation and Non-linear Regression. Dulau and Company.

Peeters, L.J.M., 2017. Assumption Hunting in Groundwater Modeling: Find Assumptions Before They Find You: L. Peeters Groundwater xx, no. x: xx-xx. Groundwater 55, 665–669. https://doi.org/10.1111/gwat.12565

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

Pianosi, F., Sarrazin, F., Wagener, T., 2020. How successfully is open-source research software adopted? Results and implications of surveying the users of a sensitivity analysis toolbox. Environmental Modelling & Software 124, 104579. https://doi.org/10.1016/j.envsoft.2019.104579

Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for Global Sensitivity Analysis. Environmental Modelling & Software 70, 80–85. https://doi.org/10.1016/j.envsoft.2015.04.009

Pianosi, F., Wagener, T., 2018. Distribution-based sensitivity analysis from a generic input-output sample. Environmental Modelling & Software 108, 197–207. https://doi.org/10.1016/j.envsoft.2018.07.019

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environmental Modelling & Software 67, 1–11. https://doi.org/10.1016/j.envsoft.2015.01.004

Pilkey, O.H., Pilkey-Jarvis, L., 2007. Useless arithmetic: why environmental scientists can't predict the future. Columbia University Press, New York.

Plischke, E., 2010. An effective algorithm for computing global sensitivity indices (EASI). Reliability Engineering & System Safety 95, 354–360. https://doi.org/10.1016/j.ress.2009.11.005

Plischke, E., Borgonovo, E., Smith, C.L., 2013. Global sensitivity measures from given data. European Journal of Operational Research 226, 536–550. https://doi.org/10.1016/j.ejor.2012.11.047

Prieur, C., Viry, L., Blayo, E., Brankart, J.-M., 2019. A global sensitivity analysis approach for marine biogeochemical modeling. Ocean Modelling 139, 101402. https://doi.org/10.1016/j.ocemod.2019.101402

Pronzato, L., Müller, W.G., 2012. Design of computer experiments: space filling and beyond. Statistics and Computing 22, 681–701. https://doi.org/10.1007/s11222-011-9242-3

Puy, A., 2020. sensobol: Computation of High-Order Sobol' Sensitivity Indices, R Package v0.2.2. https://CRAN.R-project.org/package=sensobol.

Puy, A., Becker, W., Piano, S.L., Saltelli, A., 2020a. The battle of total-order sensitivity estimators. arXiv:2009.01147 [stat].

Puy, A., Lo Piano, S., Saltelli, A., 2020b. A sensitivity analysis of the PAWN sensitivity index. Environmental Modelling & Software 127, 104679. https://doi.org/10.1016/j.envsoft.2020.104679

Puy, A., Lo Piano, S., Saltelli, A., 2020c. Current Models Underestimate Future Irrigated Areas. Geophysical Research Letters 47. https://doi.org/10.1029/2020GL087360 Rabitz, H., Aliş, Ö.F., 1999. General foundations of high-dimensional model representations. Journal of Mathematical Chemistry 25, 197–233. https://doi.org/10.1023/A:1019188517934

Raguet, H., Marrel, A., 2018. Target and Conditional Sensitivity Analysis with Emphasis on Dependence Measures. arXiv:1801.10047 [stat].

Rajbhandari, S., Rasley, J., Ruwase, O., He, Y., 2020. ZeRO: Memory Optimizations Toward Training Trillion Parameter Models. arXiv:1910.02054 [cs, stat].

Rakovec, O., Hill, M.C., Clark, M.P., Weerts, A.H., Teuling, A.J., Uijlenhoet, R., 2014. Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models: Distributed evaluation of local sensitivity analysis. Water Resources Research 50, 409–426. https://doi.org/10.1002/2013WR014063

Rasmussen, C.E., 2004. Gaussian Processes in Machine Learning, in: Bousquet, O., von Luxburg, U., Rätsch, G. (Eds.), Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 63–71. https://doi.org/10.1007/978-3-540-28650-9_4

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press.

Ravalico, J.K., Dandy, G.C., Maier, H.R., 2010. Management Option Rank Equivalence (MORE) – A new method of sensitivity analysis for decision-making. Environmental Modelling & Software 25, 171–181. https://doi.org/10.1016/j.envsoft.2009.06.012

Ravalico, J.K., Maier, H.R., Dandy, G.C., 2009. Sensitivity analysis for decision-making using the MORE method—A Pareto approach. Reliability Engineering & System Safety 94, 1229–1237. https://doi.org/10.1016/j.ress.2009.01.009

Razavi, S., Gober, P., Maier, H.R., Brouwer, R., Wheater, H., 2020. Anthropocene flooding: Challenges for science and society. Hydrological Processes 34, 1996–2000. https://doi.org/10.1002/hyp.13723

Razavi, S., Gupta, H.V., 2019. A multi-method Generalized Global Sensitivity Matrix approach to accounting for the dynamical nature of earth and environmental systems models. Environmental Modelling & Software 114, 1–11. https://doi.org/10.1016/j.envsoft.2018.12.002

Razavi, S., Gupta, H.V., 2016a. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resources Research 52, 423–439. https://doi.org/10.1002/2015WR017558

Razavi, S., Gupta, H.V., 2016b. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. Water Resources Research 52, 440–455. https://doi.org/10.1002/2015WR017559

Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models: A Critical Look at Sensitivity Analysis. Water Resources Research 51, 3070–3092. https://doi.org/10.1002/2014WR016527

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

Razavi, S., Tolson, B.A., 2011. A New Formulation for Feedforward Neural Networks. IEEE Transactions on Neural Networks 22, 1588–1598. https://doi.org/10.1109/TNN.2011.2163169

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources: REVIEW. Water Resources Research 48. https://doi.org/10.1029/2011WR011527

Ribeiro, M., Singh, S., Guestrin, C., 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. Presented at the Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, Association for Computational Linguistics, San Diego, California, pp. 97–101. https://doi.org/10.18653/v1/N16-3020

Riddell, G.A., van Delden, H., Dandy, G.C., Zecchin, A.C., Maier, H.R., 2018. Enhancing the policy relevance of exploratory scenarios: Generic approach and application to disaster risk reduction. Futures 99, 1–15. https://doi.org/10.1016/j.futures.2018.03.006

Riddell, G.A., van Delden, H., Maier, H.R., Zecchin, A.C., 2019. Exploratory scenario analysis for disaster risk reduction: Considering alternative pathways in disaster risk assessment. International Journal of Disaster Risk Reduction 39, 101230. https://doi.org/10.1016/j.ijdrr.2019.101230

Rodriguez, J.D., Perez, A., Lozano, J.A., 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 569–575. https://doi.org/10.1109/TPAMI.2009.187

Roscher, R., Bohn, B., Duarte, M.F., Garcke, J., 2020. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access 8, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

Roustant, O., Barthe, F., Iooss, B., 2017. Poincaré inequalities on intervals – application to sensitivity analysis. Electronic Journal of Statistics 11, 3081–3119. https://doi.org/10.1214/17-EJS1310

Rubinstein, R.Y., 1989. Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models. Operations Research 37, 72–81. https://doi.org/10.1287/opre.37.1.72

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rugama, L.A.J., Gilquin, L., 2018. Reliable error estimation for Sobol' indices. Stat Comput 28, 725–738. https://doi.org/10.1007/s11222-017-9759-1

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and Analysis of Computer Experiments. Statistical Science 4, 409–423. https://doi.org/10.1214/ss/1177012413

Saltelli, A., 2019. A short comment on statistical versus mathematical modelling. Nature Communications. https://doi.org/10.1038/s41467-019-11865-8

Saltelli, Andrea, 2019. Discussion Paper: Should statistics rescue mathematical modelling? arXiv:1712.06457 [stat].

Saltelli, A., 2018. Why science's crisis should not become a political battling ground. Futures 104, 85–90. https://doi.org/10.1016/j.futures.2018.07.006

Saltelli, A., 2002a. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications 145, 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1

Saltelli, A., 2002b. Sensitivity Analysis for Importance Assessment. Risk Analysis 22, 579–590. https://doi.org/10.1111/0272-4332.00040

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., Wu, Q., 2019. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environmental Modelling & Software 114, 29–39. https://doi.org/10.1016/j.envsoft.2019.01.012

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

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Fiore, M.D., Didier, E., Espeland, W.N., Kay, J., Piano, S.L., Mayo, D., Jr, R.P., Portaluri, T., Porter, T.M., Puy, A., Rafols, I., Ravetz, J.R., Reinert, E., Sarewitz, D., Stark, P.B., Stirling, A., Sluijs, J. van der, Vineis, P., 2020. Five ways to ensure that models serve society: a manifesto. Nature 582, 482–484. https://doi.org/10.1038/d41586-020-01812-9

Saltelli, A., Becker, W., Stano, P., Stark, P.B., 2015. Climate Models as Economic Guides: Scientific Challenge or Quixotic Quest? Issues in Science and Technology 31.

Saltelli, A., Funtowicz, S., 2017. What is science's crisis really about? Futures, Post-Normal science in practice 91, 5–11. https://doi.org/10.1016/j.futures.2017.05.010

Saltelli, A., Lo Piano, S., 2017. Problematic Quantifications: a Critical Appraisal of Scenario Making for a Global 'Sustainable' Food Production. Food ethics 1, 173–179. https://doi.org/10.1007/s41055-017-0020-6

Saltelli, A., Pereira, Â.G., van der Sluijs, J.P., Funtowicz, S., 2013. What do I make of your Latinorum? Sensitivity auditing of mathematical modelling. IJFIP 9, 213. https://doi.org/10.1504/IJFIP.2013.058610

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. Wiley, West Sussex, U.K.

Saltelli, A., Tarantola, S., 2002. On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. Journal of the American Statistical Association 97, 702–709. https://doi.org/10.1198/016214502388618447

Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity Anaysis as an Ingredient of Modeling. Statist. Sci. 15, 377–395. https://doi.org/10.1214/ss/1009213004

Samek, W., Müller, K.-R., 2019. Towards Explainable Artificial Intelligence, in: Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 5–22. https://doi.org/10.1007/978-3-030-28954-6_1

Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R. (Eds.), 2019. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Lecture Notes in Computer Science. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-28954-6

Santner, T.J., Williams, B.J., Notz, W.I., 2003. The Design and Analysis of Computer Experiments, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4757-3799-8

SAPEA (Science Advice for Policy by European Academies), 2019. Making sense of science for policy under conditions of complexity and uncertainty. https://doi.org/10.26356/masos.

Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global Sensitivity Analysis of environmental models: Convergence and validation. Environmental Modelling & Software 79, 135–152. https://doi.org/10.1016/j.envsoft.2016.02.005

Schaback, R., Wendland, H., 2006. Kernel techniques: From machine learning to meshless methods. Acta Numerica 15, 543–639. https://doi.org/10.1017/S0962492906270016

Schöbi, R., Sudret, B., 2019. Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions. Reliability Engineering & System Safety 187, 129–141. https://doi.org/10.1016/j.ress.2018.11.021

Seydel, R., 1993. The Art of Modeling Dynamic Systems: Forecasting for Chaos, Randomness, and Determinism (Foster Morrison). SIAM Review 35, 167–168. https://doi.org/10.1137/1035036

Shapley, L.S., 1953. 17. A Value for n-Person Games, in: Kuhn, H.W., Tucker, A.W. (Eds.), Contributions to the Theory of Games (AM-28), Volume II. Princeton University Press, Princeton, pp. 307–318. https://doi.org/10.1515/9781400881970-018

Sheikholeslami, R., Gharari, S., Papalexiou, S.M., Clark, M.P., under review. VISCOUS: A Variance-Based Sensitivity Analysis Using Copulas for Efficient Identification of Dominant Hydrological Processes. Water Resources Research.

Sheikholeslami, R., Razavi, S., 2020. A Fresh Look at Variography: Measuring Dependence and Possible Sensitivities Across Geophysical Systems From Any Given Data.

Geophysical Research Letters 47, e2020GL089829. https://doi.org/10.1029/2020GL089829
Sheikholeslami, R., Razavi, S., 2017. Progressive Latin Hypercube Sampling: An efficient approach for robust sampling-based analysis of environmental models. Environmental Modelling & Software 93, 109–126. https://doi.org/10.1016/j.envsoft.2017.03.010
Sheikholeslami, R., Razavi, S., Gupta, H.V., Becker, W., Haghnegahdar, A., 2019a. Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. Environmental Modelling & Software 111, 282–299. https://doi.org/10.1016/j.envsoft.2018.09.002
Sheikholeslami, R., Razavi, S., Haghnegahdar, A., 2019b. What should we do when a model crashes? Recommendations for global sensitivity analysis of Earth and environmental systems models. Geoscientific Model Development 12, 4275–4296. https://doi.org/10.5194/gmd-12-4275-2019
Shin, M.-J., Guillaume, J.H.A., Croke, B.F.W., Jakeman, A.J., 2015. A review of foundational methods for checking the structural identifiability of models: Results for rainfall-runoff. Journal of Hydrology 520, 1–16. https://doi.org/10.1016/j.jhydrol.2014.11.040
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
Singh Chawla, D., 2017. Big names in statistics want to shake up much-maligned P value. Nature 548, 16–17. https://doi.org/10.1038/nature.2017.22375
Sletholt, M.T., Hannay, J.E., Pfahl, D., Langtangen, H.P., 2012. What do we know about scientific software development's agile practices? 24–36. https://doi.org/10.1109/MCSE.2011.113
Smaldino, P.E., McElreath, R., 2016. The natural selection of bad science. Royal Society Open Science 3, 160384. https://doi.org/10.1098/rsos.160384
Sobol′, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6
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 1, 407–414.
Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7, 86–112. https://doi.org/10.1016/0041-5553(67)90144-9
Sobol', I.M., Asotsky, D., Kreinin, A., Kucherenko, S., 2011. Construction and Comparison of High-Dimensional Sobol' Generators. Wilmott 2011, 64–79. https://doi.org/10.1002/wilm.10056
Sobol', I.M., Kucherenko, S., 2009. Derivative based global sensitivity measures and their link with global sensitivity indices. Mathematics and Computers in Simulation 79, 3009–3017. https://doi.org/10.1016/j.matcom.2009.01.023
Sobol', I.M., Tarantola, S., Gatelli, D., Kucherenko, S.S., Mauntz, W., 2007. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliability Engineering & System Safety 92, 957–960. https://doi.org/10.1016/j.ress.2006.07.001

Spagnol, A., Riche, R.L., Veiga, S.D., 2019. Global Sensitivity Analysis for Optimization with Variable Selection. SIAM/ASA Journal on Uncertainty Quantification 7, 417–443. https://doi.org/10.1137/18M1167978

Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L., Marzouk, Y., 2015. Optimal Low-rank Approximations of Bayesian Linear Inverse Problems. SIAM Journal on Scientific Computing 37, A2451–A2487. https://doi.org/10.1137/140977308

Spear, R.C., Grieb, T.M., Shang, N., 1994. Parameter uncertainty and interaction in complex environmental models. Water Resources Research 30, 3159–3169. https://doi.org/10.1029/94WR01732

Stark, P.B., Saltelli, A., 2018. Cargo-cult statistics and scientific crisis. Significance 15, 40–43. https://doi.org/10.1111/j.1740-9713.2018.01174.x

Stichweh, R., 2001. Scientific Disciplines, History of, in: International Encyclopedia of the Social & Behavioral Sciences. Elsevier, pp. 13727–13731. https://doi.org/10.1016/B0-08-043076-7/03187-9

Strong, M., Oakley, J.E., 2013. An Efficient Method for Computing Single-Parameter Partial Expected Value of Perfect Information. Med Decis Making 33, 755–766. https://doi.org/10.1177/0272989X12465123

Štrumbelj, E., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems 41, 647–665. https://doi.org/10.1007/s10115-013-0679-x

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. Reliability Engineering & System Safety, Bayesian Networks in Dependability 93, 964–979. https://doi.org/10.1016/j.ress.2007.04.002

Sueur, R., Iooss, B., Delage, T., 2017. Sensitivity analysis using perturbed-law based indices for quantiles and application to an industrial case. Presented at the 10th International Conference on Mathematical Methods in Reliability (MMR 2017).

Sullivan, T.J., 2015. Introduction to uncertainty quantification, Texts in applied mathematics. Springer, Cham Heidelberg New York, NY Dordrecht London.

Tang, Y., Reed, P., Wagener, T., van Werkhoven, K., 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Sciences 11, 793–817. https://doi.org/10.5194/hess-11-793-2007

Tarantola, S., Giglioli, N., Jesinghaus, J., Saltelli, A., 2002. Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making? Stochastic Environmental Research and Risk Assessment (SERRA) 16, 63–76. https://doi.org/10.1007/s00477-001-0085-x

Tarantola, S., Jesinghaus, J., Puolamaa, M., 2000. Global sensitivity analysis: a quality assurance tool in environmental policy modelling, in: Saltelli, A., Chan, K., Scott, M. (Eds.), Sensitivity Analysis, Probability and Statistics. John Wiley & Sons publishers, pp. 385–397.

Tarantola, S., Mara, T.A., 2017. Variance-Based Sensitivity Indices of Computer Models with Dependent Inputs: The Fourier Amplitude Sensitivity Test. International Journal for Uncertainty Quantification 7, 511–523. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2017020291

Tedre, M., 2007. Know Your Discipline: Teaching the Philosophy of Computer Science. Journal of Information Technology Education: Research 6, 105–122. https://doi.org/10.28945/204

Terraz, T., Ribes, A., Fournier, Y., Iooss, B., Raffin, B., 2017. Melissa: large scale in transit sensitivity analysis avoiding intermediate files, in: Proceedings of the International

Conference for High Performance Computing, Networking, Storage and Analysis. ACM, Denver Colorado, pp. 1–14. https://doi.org/10.1145/3126908.3126922
Thompson, E.L., Smith, L.A., 2019. Escape from model-land. Economics: The Open-Access, Open-Assessment E-Journal. https://doi.org/10.5018/economics-ejournal.ja.2019-40
Tison, J., Park, Y.-S., Coste, M., Wasson, J.G., Rimet, F., Ector, L., Delmas, F., 2007. Predicting diatom reference communities at the French hydrosystem scale: A first step towards the definition of the good ecological status. Ecological Modelling 203, 99–108. https://doi.org/10.1016/j.ecolmodel.2006.02.047
Tissot, J.-Y., Prieur, C., 2015. A randomized orthogonal array-based procedure for the estimation of first- and second-order Sobol' indices. Journal of Statistical Computation and Simulation 85, 1358–1381. https://doi.org/10.1080/00949655.2014.971799
Tong, C., 2015. PSUADE Reference Manual (v1.7). Lawrence Livermore National Laboratory, Livermore, CA.
Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques: CALIBRATION-CONSTRAINED SSMC ANALYSIS. Water Resources Research 45. https://doi.org/10.1029/2007WR006678
Tonkin, M.J., Doherty, J., 2005. A hybrid regularized inversion methodology for highly parameterized environmental models: HYBRID REGULARIZATION METHODOLOGY. Water Resources Research 41. https://doi.org/10.1029/2005WR003995
Turner, M.G., Gardner, R.H., 2015. Introduction to Models, in: Turner, M.G., Gardner, R.H. (Eds.), Landscape Ecology in Theory and Practice: Pattern and Process. Springer, New York, NY, pp. 63–95. https://doi.org/10.1007/978-1-4939-2794-4_3
Van Der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J., Risbey, J., 2005. Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. Risk Analysis 25, 481–492. https://doi.org/10.1111/j.1539-6924.2005.00604.x
van Notten, Ph.W.F., Sleegers, A.M., van Asselt, M.B.A., 2005. The future shocks: On discontinuity and scenario development. Technological Forecasting and Social Change 72, 175–194. https://doi.org/10.1016/j.techfore.2003.12.003
VanderWeele, T.J., Ding, P., 2017. Sensitivity Analysis in Observational Research: Introducing the E-Value. Annals of Internal Medicine 167, 268. https://doi.org/10.7326/M16-2607
Vasilakos, C., Kalabokidis, K., Hatzopoulos, J., Matsinos, I., 2009. Identifying wildland fire ignition factors through sensitivity analysis of a neural network. Natural Hazards 50, 125–143. https://doi.org/10.1007/s11069-008-9326-3
Vesselinov, V.V., Alexandrov, B.S., O'Malley, D., 2019. Nonnegative tensor factorization for contaminant source identification. Journal of Contaminant Hydrology 220, 66–97. https://doi.org/10.1016/j.jconhyd.2018.11.010
Vicente-Saez, R., Martinez-Fuentes, C., 2018. Open Science now: A systematic literature review for an integrated definition. Journal of Business Research 88, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043
Wada, Y., Vinca, A., Parkinson, S., Willaarts, B.A., Magnuszewski, P., Mochizuki, J., Mayor, B., Wang, Y., Burek, P., Byers, E., Riahi, K., Krey, V., Langan, S., van Dijk, M., Grey, D., Hillers, A., Novak, R., Mukherjee, A., Bhattacharya, A., Bhardwaj, S., Romshoo, S.A., Thambi, S., Muhammad, A., Ilyas, A., Khan, A., Lashari, B.K., Mahar, R.B., Ghulam, R., Siddiqi, A., Wescoat, J., Yogeswara, N., Ashraf, A., Sidhu, B.S., Tong, J., 2019. Co-

designing Indus Water-Energy-Land Futures. One Earth 1, 185–194. https://doi.org/10.1016/j.oneear.2019.10.006

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrological Processes 17, 455–476. https://doi.org/10.1002/hyp.1135

Wagener, T., Pianosi, F., 2019. What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. Earth-Science Reviews 194, 1–18. https://doi.org/10.1016/j.earscirev.2019.04.006

Wang, H., Gong, W., Duan, Q., Di, Z., 2020. Evaluation of parameter interaction effect of hydrological models using the sparse polynomial chaos (SPC) method. Environmental Modelling & Software 125, 104612. https://doi.org/10.1016/j.envsoft.2019.104612

Wasserstein, R.L., Lazar, N.A., 2016. The ASA Statement on *p* -Values: Context, Process, and Purpose. The American Statistician 70, 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wei, P., Lu, Z., Song, J., 2015. Variable importance analysis: A comprehensive review. Reliability Engineering & System Safety 142, 399–432. https://doi.org/10.1016/j.ress.2015.05.018

Wei, P., Lu, Z., Song, J., 2014. Moment-Independent Sensitivity Analysis Using Copula. Risk Analysis 34, 210–222. https://doi.org/10.1111/risa.12110

Wiek, A., Keeler, L.W., Schweizer, V., Lang, D.J., 2013. Plausibility indications in future scenarios. International Journal of Foresight and Innovation Policy 9, 133. https://doi.org/10.1504/IJFIP.2013.058611

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Workman, M., Dooley, K., Lomax, G., Maltby, J., Darch, G., 2020. Decision making in contexts of deep uncertainty - An alternative approach for long-term climate policy. Environmental Science & Policy 103, 77–84. https://doi.org/10.1016/j.envsci.2019.10.002

Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. Environmental Modelling & Software 54, 108–127. https://doi.org/10.1016/j.envsoft.2013.12.016

Xiu, D., Karniadakis, G.E., 2002. The Wiener--Askey Polynomial Chaos for Stochastic Differential Equations. SIAM Journal on Scientific Computing 24, 619–644. https://doi.org/10.1137/S1064827501387826

Xu, C., Gertner, G.Z., 2008. Uncertainty and sensitivity analysis for models with correlated parameters. Reliability Engineering & System Safety 93, 1563–1573. https://doi.org/10.1016/j.ress.2007.06.003

Yang, J., Jakeman, A., Fang, G., Chen, X., 2018. Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator. Environmental Modelling & Software 101, 289–300. https://doi.org/10.1016/j.envsoft.2017.11.037

Zadeh, F.K., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., Bauwens, W., 2017. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. Environmental Modelling & Software 91, 210–222. https://doi.org/10.1016/j.envsoft.2017.02.001

Zhang, P., 2019. A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. Applied Soft Computing 85, 105859. https://doi.org/10.1016/j.asoc.2019.105859

Zuniga, M.M., Kucherenko, S., Shah, N., 2013. Metamodelling with independent and dependent inputs. Computer Physics Communications 184, 1570–1580. https://doi.org/10.1016/j.cpc.2013.02.005