

Discussion Paper: Should statistics rescue mathematical modelling?

Andrea Saltelli

Centre for the Study of the Sciences and the Humanities (SVT),
University of Bergen (UIB, Bergen, Norway), and

Open Evidence Research,
Universitat Oberta de Catalunya (UOC, Barcelona, Spain)

Abstract

Statistics experiences a storm around the perceived misuse and possible abuse of its methods in the context of the so-called reproducibility crisis. The methods and styles of quantification practiced in mathematical modelling rarely make it to the headlines, though modelling practitioners writing in disciplinary journals flag a host of problems in the field. Technical, cultural and ethical dimensions are simultaneously at play in the current predicaments of both statistics and mathematical modelling.

Since mathematical modelling is not a discipline like statistics, its shortcomings risk remaining untreated longer. We suggest that the tools of statistics and its disciplinary organisation might offer a remedial contribution to mathematical modelling, standardising methodologies and disseminating good practices. Statistics could provide scientists and engineers from all disciplines with a point of anchorage for sound modelling work. This is a vast and long-term undertaking. A step in the proposed direction is offered here by focusing on the use of statistical tools for quality assurance of mathematical models. By way of illustration, techniques for uncertainty quantification, sensitivity analysis and sensitivity auditing are suggested for incorporation in statistical syllabuses and practices.

Keywords: Mathematical modelling; statistical modelling; sensitivity analysis; uncertainty analysis; sensitivity auditing; sociology of quantification;

1 INTRODUCTION

The scope, the narrative and the order of exposition of the paper are given in the present introduction.

1.1 DIFFERENT CRISES

Mediatic attention (Editors, 2013b) has surrounded the emergence of a crisis in science's quality control apparatus (Benessia et al., 2016; Harris, 2017; Ioannidis, Chen, Kodell, Haug, & Hoey, 2005). Prima facie, the crisis concerns the non-reproducibility of scientific findings in a variety of fields such as medicine (Begley & Ellis, 2012; Ioannidis et al., 2016), psychology (Open Science Collaboration (OSC), 2015; Shanks et al., 2015), economics (Ioannidis, Stanley, & Doucouliagos, 2017), and many others (Saltelli & Funtowicz, 2017), be it with important differences among disciplines (Fanelli, 2010). Statistics has been especially targeted as bearing responsibility. Common accusations - to which statistics has been quick to respond (Wasserstein & Lazar, 2016), include poor statistical methods, poor teaching of the same, and disagreements in the profession. The angst toward statistics should not surprise; statistics is the coal face of scientific work, and hardly any scientific activity is performed without its results being couched in its language. Of course, evidence of sloppy practices is not the preserve of statistics, but affect the handling of aspects of research from cell lines to human patients (Harris, 2017).

When interrogated as to the causes of the crisis in statistics, statisticians are quick to point out that this is more than just a technical issue. A wrong system of incentives (Leek et al., 2017), the existence of statistical rituals (Gigerenzer, 2018; Gigerenzer & Marewski, 2015), and the delicacy of statistical work (Gelman & Loken, 2013) have all been singled out as relevant contributors.

What is the situation in mathematical modelling? Mathematical modelling is not a discipline, and different research fields go about modelling following disciplinary habits and fashions. Modellers from various disciplines have warned that costly, dangerous, and persistent problems with mathematical modelling are connected to poor quality control and lack of appropriate standards (Jakeman, Letcher, & Norton, 2006; Padilla, Diallo, Lynch, & Gore, 2018). Additionally mathematical model may obfuscate the normative choice made in their construction and calibration, as is being now discussed in the controversial use of algorithms in many aspects of everyday life (Brauneis & Goodman, 2018). Without an acknowledged core, modelling lacks appropriate internal antibodies – in the form of quality standards, disciplinary associations & disciplinary journals, and recognized leaders, which would allow modellers as a constituency to issue authoritative statements such as those issued by statisticians.

An important issue seen in mathematical modelling is in the management of uncertainty. Uncertainty quantification should be at the hearth of the scientific method, and a fortiori in the use of science for policy (Funtowicz & Ravetz, 1990). In statistics the p-test can be misused as to overestimate the probability of having found a true effect (Colquhoun, 2014). Likewise in modelling studies certainty may be overestimated, thus producing crisp numbers precise to the third decimal digit even in situations of pervasive uncertainty or

ignorance (Saltelli, Stark, Becker, & Stano, 2015), including in important cases where science needs to inform policy (Pilkey & Pilkey-Jarvis, 2009). Normative and cultural problems naturally compound the problem as for the case of statistics. It is an old refrain in mathematical modelling – first noted among hydro-geologists, that since models are often over-parametrized, they can be made to conclude everything (Hornberger & Spear, 1981).

1.2 STATISTICS AT THE RESCUE?

Comparing the situation of mathematical modelling with that of statistics it is evident that while the problems are likely to be more serious with the former – isn't it is after all easier to cheat with computer code than with actual samples? - the prospects for a solution appear better for the latter. Additionally, statistics is unique both in its authority and scope: modellers of all extractions are more likely to listen to statistics than to one another; next, statistical and mathematical modelling are proximate by both content – both are avenues for quantification - and by pathologies. Finally, according to the nascent field of sociology of quantification, the merging of algorithms with big data blurs many existing distinctions: “*what qualities are specific to rankings, or indicators, or models, or algorithms?*” (Popp Berman & Hirschman, 2018).

1.3 WHERE TO START

Mapping the route of statistics' colonization of the infinite field of mathematical modelling is beyond the remit of a single paper. But a first track can be illustrated here. If we accepts the practitioners' hunch that numerical hubris and poor accounting of the uncertainty are a key problem, then statistics can step in providing a bedrock of methodologies for uncertainty quantification and sensitivity analysis (Saltelli et al., 2008). Most of the best practices in uncertainty and sensitivity analysis came from - or can be framed as belonging to - statistics. Additionally, there are at the forefront of research in sensitivity analysis a large class of statistical problem in classification, variable selection and machine learning where sensitivity analysis can make an inroad which would justify and reinforce the concept that sensitivity analysis needs to be part of a statistical syllabus.

1.4 SENSITIVITY AUDITING

Just as per the case of statistics, no solution is possible without careful appraisal of the normative and cultural dimensions of the problem. For this reason, technical uncertainty and sensitivity analyses have been complemented by sensitivity auditing (Saltelli & Funtowicz, 2014; Saltelli, Guimaraes Pereira, van der Sluijs, & Funtowicz, 2013) – where concern focuses on the modelling process, on the underlying implicit assumptions and norms, as well as on the power relationships and interests underpinning a particular model development. Though relatively young, sensitivity auditing is already mentioned in existing impact assessment guidelines (European Commission, 2015). Sensitivity auditing borrows some of its concepts from postnormal science (PNS) (Funtowicz & Ravetz, 1993), whose contribution to a reflexive and defensible use of mathematical modelling (Ravetz, 1997) is known to practitioners such as (Jakeman et al., 2006). The use of postnormal science in and for statistics would deserve a separate essay, perhaps as a continuation of the piece from (Zidek,

2006) on this journal which notes how in adversarial settings, where the evidence is held in balance, “there is plenty of scope [...] for a postnormal statistical scientist”, suggesting that statisticians should be taught PNS as an essential ingredient of their training.

Given the proximity already mentioned among different kinds of quantifications, the present discussion of norms in models feeds into past (Bird et al., 2005) and present-day (Muller, 2018) debates on the ethical and practical implications of using metrics and algorithms (O’Neil, 2016).

2 STATISTICS AND ITS CRISIS

The last decade has brought to the attention of the public a crisis in science’s governance and quality-control system. Here models rarely appear to forefront. The media talk about ‘wrong science (Editors, 2013b)’ and ‘trouble at the lab’ (Editors, 2013a), while some academicians declare most published research findings false (Ioannidis et al., 2005). The crisis is multifaceted, and manifests itself as the non-reproducibility of scientific findings in many fields (Baker, 2015; Begley, 2013; Begley & Ioannidis, 2015) and with a surge in retractions of scientific papers (McCook, 2017). A crisis of expertise and the advent of a post-truth society have become dominant themes of discourse (Flood, 2016). A new aspect of these phenomena is that entire fields and subfields are seen as compromised - the field of empirical economics being a recent addition (Ioannidis et al., 2017). Other observers deny the existence of a crisis (Fanelli, 2018), while the situation is considerably complicated by the political tensions surrounding the use of science in regulation, especially in the US (Saltelli, 2018).

In statistics abuse or misuse of the p-test (Colquhoun, 2014; Gigerenzer & Marewski, 2015; Leek et al., 2017; Singh Chawla, 2017; Wasserstein & Lazar, 2016) has been singled out by many observers as one of the causes of the reproducibility crisis currently affecting science (Editors, 2013b; Ioannidis et al., 2016, 2005; Ioannidis, Forstman, Boutron, Yu, & Cook, 2014; Open Science Collaboration (OSC), 2015). Misuse of p-values includes p-hacking (fishing for favourable p-values (Shanks et al., 2015)) and HARKing (formulating the research Hypothesis After the Results are Known (Kerr, 1998)). Here the desire to achieve a sought for - or simply publishable - result may lead to more or less disingenuous fiddling with the data points, the modelling assumptions, the statistical analysis, or the research hypotheses themselves (Leamer, 2010). It has also been noted that the issue with p-values is just the tip of the iceberg, and that all steps of statistical data analysis may suffer from poor quality statistical work (Leek & Peng, 2015).

Naturally, there are considerable differences among disciplines in terms of cultures and records of quality control (Fanelli, Anderson, Vries, Rohanizadegan, & Evangelou, 2010), with Physics often invoked as a model to be adopted also in relation to its master of statistical techniques (Horton, 2015; Nuzzo, 2015).

The community of statistics exhibits a considerable latitude of opinions on the crisis – the ASA statement (Wasserstein & Lazar, 2016) was complemented by as many as twenty-one commentaries by top statisticians of different orientation. Statisticians also disagree as to the balance of blame – is it the publish or perish zeitgeist or is it a technical problems to be overcome with better methods (Leek et al., 2017; Saltelli & Stark, 2018; Stark & Saltelli, 2018). Some point to a deeply rooted cultural problem, linked to the existence of statistical rituals and consolidated through the history of the profession (Gigerenzer, 2018). Reflexive thinkers draw attention to the delicacy and complexity of statistical work, whereby different outcome can be produced even without statisticians having embarked in ‘fishing expedition’ to secure a desired outcome (Gelman & Loken, 2013; Ravetz, 2018).

The disagreements among statisticians is perhaps more apparent than real. There appear to be overall a prevailing sentiment that reliance on null hypothesis statistical testing (NHST) should be tempered by other methods as well as by a more holistic and integrative view of what constitutes evidence in each and every individual analysis (McShane & Gal, 2017).

The problems remain severe as the crisis of science’s quality control has deep roots (Saltelli & Funtowicz, 2017). A science which cuts corner is unfortunately ecologically fit (Smaldino & McElreath, 2016) to the existing governance arrangement of scientific production (Mirowski, 2011), meaning by this that it tends to reproduce itself better than the virtuous sort. Additionally it appear resilient to attempts to reform (Banobi, Branch, & Hilborn, 2011; M. A. Edwards & Roy, 2017). Still statistics appears to have taken good note of the problem and be committed to identifying solutions.

3 MALPRACTICES IN MATHEMATICAL MODELLING?

We have been told that in the field of clinical medical research the percentage of non-reproducible studies could reach 85% (Chalmers & Glasziou, 2009). Which is that of mathematical modelling? How can the case be made that modelling needs attention, and that its predicaments are possibly more serious than those of statistics? An attempt is made here by reflecting on the vulnerability of modelling work, the existing of rituals, the opinion of practitioners, and examples of problematic use of mathematical modelling.

3.1 VULNERABILITY OF MODELLING

Historian Naomi Oreskes (2000) has observed that model-based predictions tend to be treated as logical inferences as met in the classic hypothetic-deductive model of science’s operation: laws are tested by making predictions which are then tested by experiments. If the prediction is not confirmed, then the law is ‘falsified’ – in the sense of proven false. For Oreskes a major problem in our use of mathematical models is in assimilating them to physical laws, and hence treating their prediction with the same dignity. This is wrong, according to Oreskes, because in order to be of value in theory testing, the model-based predictions “must be capable of refuting the theory that generated them”. Since models are “complex amalgam of theoretical and phenomenological laws (and the governing equations and algorithms that represent them), empirical input

parameters, and a model conceptualization” then when a model prediction fails what part of all this construct was falsified? The underlying theory? The calibration data? The system’s formalization or choice of boundaries? Some of the algorithms used in the model?

The relation between models and data is often more symbiotic than adversarial. This has been noted e.g. in climate studies, where the relation between model and data has been defined ‘incestuous’, exactly to make the point that in modelling studies using data to prove a model wrong is not straightforward (P. N. Edwards, 1999).

According to Robert Rosen (Rosen, 1991), a theoretical biologist known for his contributions to complex system biology and the definition of ‘anticipatory systems’ (Fuller, 2017), modelling is not a science but rather a craft. For Rosen it is impossible to link ‘model’ and ‘system modelled’ within an unambiguous causal framework. In other words, given the same system, multiple representations of the system are possible, even in the context of a closed set of specifications (Refsgaard, van der Sluijs, Brown, & van der Keur, 2006), depending on the expectations and styles of the modellers.

3.2 RITUALS

An important analogy between statistical and mathematical modelling is in the ‘ritual’ use of methods. Existence of rituals in statistics has been discussed extensively by Gigerenzer (Gigerenzer, 2018; Gigerenzer & Marewski, 2015); we similarly call cargo-cult statistics the use of statistical terms and procedures as incantations, with scant understanding of the assumptions or relevance of the calculations (Stark & Saltelli, 2018). As per rituals in modelling the best anecdote is perhaps that offered by Kenneth Arrow. During the Second World War he was a weather officer in the US Army Air Corps working on the production of month-ahead weather forecasts, and this is how he tells the story (Szenberg, 1992):

The statisticians among us subjected these forecasts to verification and they differed in no way from chance. The forecasters themselves were convinced and requested that the forecasts be discontinued. The reply read approximately like this: “The commanding general is well aware that the forecasts are no good. However, he needs them for planning purposes”.

Social scientist Niklas Luhmann, discussed in (Moeller, 2006), uses the terms ‘deparadoxification’ to indicate the use of scientific knowledge to give a pretention of objectivity, to show that policy decisions are based on a publicly verifiable process, rather than on expert’s whim. Along these lines it has been argued that models universally known to be wrong continue to play a role in economic policy decisions (Martin, 2018), which relates to our discussion of algorithms: authorities may willingly or inadvertently recur to algorithms to implements rules which - if applied by human subjects - would be either legislated or made through administrative rulemaking. Big data prediction models can be built instead and used without policy decisions ever having been the object of a political or at least traceable administrative procedure, as noted by (Brauneis & Goodman, 2018).

3.3 PRACTITIONERS' TAKE

Its state of fragmentation has led to mathematical modelling being a field in which the quality control arrangements in place would offer room for improvement. A book on modelling read by all disciplines would go some way toward the enhancing the communication of good practices among fields. Unfortunately, such a book does not exist. A good work is that of Foster Morrison (Morrison, 1991), though the book is silent about verification and validation procedures. These are instead well treated in (Santner, Williams, & Notz, 2003), and in (Oberkampf & Roy, 2010). Rosen's book (Rosen, 1991) contains deep insights from system biology - including on causality, and on 'modelling as a craft'. Still the problem is that none of these works - independently of their merit - can command the necessary epistemic authority in the absence of a disciplinary house for modelling.

Padilla and co-workers, having run a survey involving 283 responding modellers, note that the heterogeneous nature of the modelling and simulation community prevents the emergence of consolidated paradigms, and that as a result verification and verification procedures are a rather trial and error business (Padilla et al., 2018). These authors advocate a "more structured, generalized and standardized approach to verification". Jakeman and co-workers - after reviewing several existing checklist and approaches for model quality - note that poorly informed non-modeller risk remaining "unaware of limitations, uncertainties, omissions and subjective choices in models", with consequent over-reliance in the quality of model-based inference. On their side model developers may oversimplify or overelaborate with the result of obfuscating model use (Jakeman et al., 2006). These authors propose a 10 points checklist for model development. Noticeably, the list starts with the steps of definition of the model purpose and context, of conceptualization of the task, of choice of the family on models & uncertainty specification, before moving to the more usual nitty gritty issues of estimation, calibration and verification. The authors make use of model-relevant elements (Ravetz, 1997) from the tradition of postnormal science (Funtowicz & Ravetz, 1993), including NUSAP, a system of pedigree for quantitative information (Funtowicz & Ravetz, 1990; van der Sluijs et al., 2005). The list also insists on the need for a participatory approach to the construction of model, in the same epistemological tradition.

3.4 REAL LIFE EXAMPLES

There are fields where the suffering in mathematical modelling has been more visibly flagged. "Useless Arithmetics" (Pilkey & Pilkey-Jarvis, 2009), argues that quantitative mathematical models policy makers and government administrators use to form environmental policies are seriously flawed, and offers a host of example from AIDS prevention to the stock of fisheries, from mill tailing to costal erosion. An excerpt is offered here to give the flavour of the work (p.xiii):

The discredited Brunn rule model predicts how much shoreline erosion will be created by sea-level rise, and since no other model claims to do this, the Brunn rule remains in widespread use. The maximum sustainable yield is a concept that fishery models are still using as a mean to preserve fish populations despite the fact that the concept was discredited thirty-five years ago.

3.4.1 Modelling of a nuclear waste disposal

One case treated in the book just mentioned is that of the nuclear waste repository at Yucca Mountain. Here a key number used in the assessment, the percolation rate of water to the unsaturated repository level - remained underestimated by 4 orders of magnitude for more than a decade (Metlay, 2000). Measurements made public in 1996 and later confirmed (Campbell, Wolfsberg, Fabryka-Martin, & Sweetkind, 2003) revealed the presence of Chlorine-36, a bomb-pulse isotope associated to nuclear explosion at the Bikini Atoll in South Pacific 1963. Such a presence implies a travel speed of 3,000 millimetres per years, rather than the range 0.01 to 1 millimetre used in the Total System Performance Assessment model, a mathematical computer program made of 13 models in turn comprised of 286 individual modules. The modelling activity around TSPA has continued (Helton, Hansen, & Swift, 2014), tracking the fate of radionuclides thousands of years into the future. The author of the present work – likewise guilty - has computed radiological doses to population one hundred thousand years into the future (<http://www.oecd-nea.org/tools/abstract/detail/nea-0860/>). Today he would consider that this kind of model-based analyses reliant on the physics, the chemistry, the geology and the radio-isotopic properties of the waste are a fragile input to the safety case of a nuclear disposal, and that the surrounding institutional and social settings and power relations should be the focus of our attention. This would appear to be one of those cases where a narrowly defined scientific knowledge would give prominence to a restricted agenda of defined uncertainties. (Wynne, 1992) in particular discuss the concept of indeterminacy – as distinguished from uncertainty - with the following example:

will the high quality of maintenance, inspection, operation, etc, of a risky technology be sustained in future, multiplied over replications, possibly many all over the world?

If one contrasts the mathematical precision of TSPA with the confusion and neglect surrounding present day's nuclear waste in dump sites in the US (Diaz-Maurin, 2018) it easy to grasps the meaning of Wynne's intuition.

3.4.2 How much will climate cost?

The community of modellers involved in climate change is not exempt from practices which statistics would censor. Treating estimates based on an ad hoc collection of related numerical models with unknown, potentially large systematic errors as if they were a random sample from a distribution centred at the parameter is an example (Stark & Saltelli, 2018). While modelling the future sea level and temperature appear with all their difficulties a meaningful pursuit, that of translating this in percentage of GDP point – the so called 'costing' of climate change, is by the admission of one key expert in the domain a 'terra incognita' (Nordhaus, 1991). In this terra incognita modellers venture to investigate - using large mathematical model - the increase in crime rate at county level as a result of climate change one hundred years from now (Kopp, Rasmussen, & Mastrandrea, 2014; Saltelli et al., 2016, 2015). Modellers from the same community argues for more efforts and resources in this venture (Stern, 2015). Both costing climate one hundred years from now and assessing the fate of nuclear waste tens of thousands of years from now using quantitative predictive models resonate with what Alvin Weinberg called trans-science (Weinberg, 1972), i.e. a practice which lends itself to the language and formalism of science but where science cannot provide answers. In the specific case of attaching

a cost to climate change, an uncertainty and sensitivity analysis can be shown to support this conclusion: the uncertainty of the prediction is so wide as to be useless for reaching a policy conclusion (Saltelli & d’Hombres, 2010), however we might think that reaching one particular conclusion on this issue amounts to a moral imperative toward the planet.

3.4.3 Finance and economics

As per economics, conspicuous model failure has been associated with finance (Wilmott & Orrell, 2017), relating to models used in the pricing of opaque financial products held partly responsible for the onset of the last recession. Philip Mirowski devotes a full chapter in his ‘Never Let a Serious Crisis Go to Waste’ (Mirowski, 2013) to disparage the over-reliance of economists on dynamic stochastic general equilibrium (DSGE) models used for policy simulation. This theme hit the media when both the US senate and Queen Elisabeth asked their economists for clarification about these models’ apparent incapacity to anticipate the crisis (Pierce, 2008). Incidentally, DSGE are among the solution proposed to ‘cost’ climate (Stern, 2015). Economist Paul Romer has recently coined the neologism ‘mathiness’ (Romer, 2015b), taking issue against ‘freshwater economists’ (an allusion to the Chicago school in the great lakes region) for their use of mathematics as Latin, in the sense that mathematics would be used to distract from underlying ideological stances. Romer draws unflattering parallels between macroeconomic and string theory in physics for their invocation to “imaginary causal forces” and excessive deference to authority of the profession’s leaders (Romer, 2018), and invites (Romer, 2015a) his fellow economists to appreciate the importance of intellectual honesty by reading Richard Feynman’s ‘Cargo Cult’ lecture, given at Caltex in 1974 (Feynman, Leighton, & Hutchings, 1985). The lecture famously argued that scientific integrity translates into ‘utter honesty’ and ‘leaning over backwards’ to share all details that could throw doubt on our own analysis: all the information should be given to other investigators for them to judge the value of our own contribution. We shall revisit this theme in our conclusions.

3.4.4 Algorithms of ‘math destruction’?

Pathologies have been noted in the domain of algorithms – themselves models of a sort, where the use of opaquely calibrated models embedding unscrutinised rules is growing (Mancha & Ali, 2017). In algorithms - as well as in modelling, trade secrecy may prevent the dissemination of precisely those models which are most widely used (Brauneis & Goodman, 2018). Algorithms today decide upon an ever-increasing list of cases, such as recruiting (Alexander, 2016), deciding on carriers - including of researchers (Abraham C., 2016), prison sentencing, paroling, custody of minors (Brauneis & Goodman, 2018), with a popular book talking of “Weapons of Math Destruction” (O’Neil, 2016). In New York, where algorithms are used by the administration for a large array of decisions, the mayor has decided to pursue legislation for “algorithmic audits” (Dwyer, 2014) and agendas are set to tackle algorithmic governance (Danaher et al., 2017)¹.

¹ Further references on bias in automated systems can be found at the site <https://abebabirhane.wordpress.com/>.

3.4.5 How many planets do we need?

The mathematical protocol developed by the Global Footprint Network (GFN) aims to assess man's impact on the planet. Simplifying, this is done by aggregating across scales and compartments different impacts due to our use of natural resource into a single number: how much of the planet's surface we use. Thus, the result of the analysis can be communicated as an overall measure of man's impact, such as the 'Earth overshoot day', e.g. 'August 2 is Earth Overshoot Day 2017' (www.overshootday.org) signifies that in less than 8 months, humanity exhausts Earth's budget for the year. This number is problematic. To give an example, the EF protocol suggests that we are overshooting the capacity of the planet by about 50%, but the quantity of nitrogen used in agriculture alone would already require 2.5 planets to be fixed by natural processes – not just 1.6. The 1.6 number could become 16 or 160 or infinity, depending on what impact of man is considered – infinity corresponding to irreversible processes, e.g. man extinguishing a species by overfishing or by pesticides. Several practitioners have objected to the EF; a wealth of references can be found starting from a joint paper written by EF proponents and opponent, (Galli et al., 2016). Also noticeable is the negative opinion about the EF in the Report by the Commission on the Measurement of Economic Performance and Social Progress led by J. Stiglitz, JP Fitoussi and A. Sen (www.stiglitz-sen-fitoussi.fr) (Stiglitz, Sen, & Fitoussi, 2009). This has not affected the success of the Ecological Footprint, and its take up by countries, organizations, and media. Perhaps part of the resilience of the EF is due to its complex algorithmic nature – more complex to be proven false than a non-reproducible p-value in data analysis, in line with our diagnosis that the situation with mathematical modelling is worse than in data analysis. Again, the ethical issue comes into the picture, as in the experience of the author even informed scientists may speak in favour of the EF for the desirability of the message it conveys and the mediatic effect of a crisp, single number capturing all of humanity's impact on the planet.

4 A MODEST BEGINNING

A step in the direction of statistics injecting some structure and standards into mathematical modelling would be by situating in the house of statistics tools for quality assurance of mathematical models. This is not a radically new idea. In the UK a group known as MUCM, for Managing Uncertainty in Complex Models, has been active in 2006-2012 on issues of uncertainty in simulation models addressing uncertainty quantification, uncertainty propagation, uncertainty analysis, sensitivity analysis, calibration (or tuning, history matching, etc.) and ensemble analysis. A quick glance at the project homepage² shows a strong statistical presence among its participants.

² <http://www.mucm.ac.uk/>

In this section an illustration is provided of how this approach might work for the specific case of sensitivity analysis and sensitivity auditing. This is just an example; it is not the scope of this paper to suggest that these methodologies can by themselves lead a thorough reformation of the field. Reforming modelling will involve both the adoption of methods in the practice of research and their insertion into teaching syllabuses. If nothing else, the present proposal could represent a low hanging fruit; sensitivity analysis (SA) already uses tools from statistics, such as Hoeffding decompositions (Archer, Saltelli, & Sobol', 1997), ANOVA, design of experiment (Box, Hunter, Hunter, & J. Stuart, 1978), Bayesian formalism for partial variances and moment independent methods, and many others. Yet statistics does not normally use tools from SA. This needs to change as SA can make a good contribution to problems routinely with in statistics as discussed below.

In the following two sections we shall present sensitivity analysis and auditing in turn.

4.1 SENSITIVITY ANALYSIS

Sensitivity analysis (SA) finds use in a large class of applications, such as model selection, calibration, optimisation, quality assurance and many others (Saltelli, 2002b, 2002a; Saltelli et al., 2010; Saltelli, Ratto, Tarantola, & Campolongo, 2012; Saltelli, Tarantola, & Chan, 1999; Saltelli, Tarantola, & Campolongo, 2000). Sensitivity analysis answers the question 'Which uncertain input factors are responsible for the uncertainty in the prediction?' SA is distinct from uncertainty analysis (UA), answering instead the question 'How uncertain is the prediction?' Different disciplines use these terms differently, though e.g. in economics, sensitivity analysis is often used to mean what we would call an uncertainty analysis.

Recent reviews of sensitivity analysis are (Emanuele Borgonovo & Plischke, 2016; Norton & John, 2015; Wei, Lu, & Song, 2015), while useful textbooks for sensitivity analysis are (Emanuele Borgonovo, 2017; Cacuci, 2003; Rocquigny, Devictor, & Tarantola, 2008; Saltelli et al., 2008; Saltelli, Tarantola, Campolongo, & Ratto, 2004; Saltelli, Chan, & Scott, 2000). With time good practices have become consolidated such as the variance-based measures of sensitivity. These methods decompose the variance of the output into terms of the first order and terms of higher orders describing interactions. Another emerging good practice is that of in moment-independent methods (E. Borgonovo, 2007; E. Borgonovo, Castaings, & Tarantola, 2012; Wei et al., 2015). Instead of looking at how factors affect the output variance, these methods look at how they affect the empirical probability distribution of the output as generated by the uncertainty analysis. As they look at the entire distribution and not just a moment (such as the variance) they are called moment-independent.

Sensitivity analysis is acknowledged as a useful practice in model development and application. Its use in regulatory setting – e.g. in impact assessment studies – is prescribed in guidelines both in Europe and the United States (European Commission(European Commission, 2015), 2015, p. 390-393; Office for the Management and Budget(Office of Management and Budget, 2006), 2006, p. 17-18; Environmental Protection Agency(U.S. Environmental Protection Agency (EPA), 2009), 2009, p.26).

Regrettably most modelling work does not include a sensitivity analysis, and that the majority of sensitivity analyses published in the literature are flawed to the point of irrelevance (Ferretti, Saltelli, & Tarantola, 2016; Saltelli et al., 2017).

A proper uncertainty analysis of the output of a mathematical model needs to map what the model does when the input factors are left free to vary over their range of existence. *A fortiori* this is true of a sensitivity analysis. Most UA and SA still explore the input space moving along mono-dimensional corridors, which leaves the space of variation of the input factors mostly unscathed. This approach is known as OAT, for moving One factor-At-a-Time, and is the one mostly preferred by modellers for a variety of practical reasons, including the ease of interpretation of the results (Saltelli & Annoni, 2010). Statisticians would suggest using instead the theory of design of experiment (DOE (Box et al., 1978)), also because OAT ignores the existence of interactions (factors behaving synergistically), and thus the results based on OAT could dramatically underestimate the uncertainty. In a risk analysis this could imply overlooking the conditions for an accident, such as e.g. missing the conditions for a runaway reaction in a chemical reactor. This is why in the theory of experimental design factors are moved in groups, rather than OAT, to optimise the exploration of the space of the factors. Thus, most published SA fail the elementary requirement to properly explore the space of the input factors and their interplay, though this finding appears - from the personal experience of the author, discipline-dependent (Ferretti et al., 2016; Saltelli et al., 2017).

One problem in sensitivity analysis which is clear to the author from the requests he receives is what method among the many available should be used. There are many measures for the importance of a factor, so that there can then be a different sensitivity analysis from which to arbitrarily pick. This gap urgently needs closing with a clear line of argumentation about what should be used when, and foremost why. The sensitivity measure adopted must be itself defensible.

Regrettably, techniques for proper uncertainty and sensitivity analysis of model-based inference may tend to be eschewed because, in their candour, these techniques may show that the inference itself is too uncertain to be of any practical use. An estimate of an investment's pay-off that gives a range from a large loss to a large gain is not what the problem owner may wish to hear. It has been observed by practitioners of different disciplines (Funtowicz & Ravetz, 1990; Leamer, 1985, 2010) that, under these circumstances, analysts may be tempted to 'adjust' the uncertainty in the input until the output range is narrower and conveniently located in

friendly territory. The opposite could also be the case, where the owner of the analysis wishes uncertainties to be amplified, e.g. to the effect to deter regulatory interventions. Uncertainty can be used strategically or instrumentally, as noted in (Saltelli et al., 2013), and a sound sensitivity analysis can help to make up one's mind about the merits of a case.

It has been noted that doing modelling without sensitivity analysis is like practicing orthopaedics without x-rays. To a practitioner of uncertainty or sensitivity analysis, the idea of running a mathematical model 'just once' sounds just as weird. Why would one wish to do so when we all know that at least part of the information feeding into the model is uncertain?

As uncertainty is omnipresent in modelling, putting uncertainty at the core of model quality has a potentially unifying potential. Present computing and software capacity allows for most models to be executed repeatedly - even and especially at the model development stage. Hence one of the suggestion of the present paper is that, whenever feasible, the modelling work takes place within an ideal Monte Carlo driver for uncertainty and sensitivity analysis - e.g. the model is never run just once; instead it is systematically fed a sample of input values and returns a sample of model output for statistical treatment. If the model runs within a Monte Carlo driver then all sources of uncertainty - framing uncertainties, parametric uncertainties, and so on, can be activated simultaneously, allowing rapid and useful inferences as e.g. to what contribution each set of uncertainties does to the uncertainty in the inference. In other words: since models offers statements that are conditional on their input, this conditionality should be made explicit every time the model is used, including and foremost at the stage of model construction. This is a change with respect to the ten-step model building procedure suggested by Jakeman and co-workers, where "Quantification of uncertainty" is step N. 9 in the checklist. Instead, all steps in the procedure need an uncertainty quantification - e.g. step N.1, "Definition of the purposes for modelling" is already a good subject for analysis of uncertainty - e.g. when disagreement exist among stakeholders about what scope the analysis should serve. Likewise, we suggest that "Identification of model structure and parameter value" (step N.7) might itself part of an uncertainty and sensitivity analysis when for example two or more models were left competing for the task and a decision were to be needed.

4.2 WHAT CAN SENSITIVITY ANALYSIS CONTRIBUTE TO STATISTICS?

One particular measure used in sensitivity analysis - known as Total Sensitivity Indices T_i - can have useful applications in the field of statistics. A contribution to variable selection in regression is a recent example (Becker, Paruolo, & Saltelli, 2014).

While statisticians are familiar with Pearson correlation ratio (which is called Sobol' first order sensitivity coefficient in sensitivity analysis), they are unfamiliar with T_i . These indices capture the influence of all effects (in an ANOVA sense) involving a given factor (Homma & Saltelli, 1996). To make an example, in a system

with three independent factors – a simple deterministic model of the form $Y = f(X_1, X_2, X_3)$ - the total variance can be normalized to unit and decomposed as

$$S_1 + S_2 + S_3 + S_{12} + S_{13} + S_{23} + S_{123} = 1$$

Where S_1 is the first order index or Pearson correlation ratio:

$$S_i = \frac{V_{X \sim i}(E_{X_i}(Y|X_i))}{V(Y)}$$

The total sensitivity index for factor X_1 is simply $T_1 = S_1 + S_{12} + S_{13} + S_{123}$. This index can be computed as

$$T_i = \frac{E_{X_i}(V_{X \sim i}(Y|X_i))}{V(Y)}$$

T_i can be used - e.g. in variable selection in regression – as a trigger to either choose or not chose a particular regressor for the case where the output of interest Y is a measure of fit of a regression model. Thus T_i captures the overall effectiveness of that particular regressor, inclusive of its interactions with the choice/not choice of other regressors.

The extension of this approach to problems commonly dealt with in statistics appears promising. E.g. in problems of classification. A new type of classification score can be developed using T_i to drive the selection of the variables. Likewise, when studying features importance with forests of tree; instead or randomizing individual columns (this is called permutation feature importance and corresponds to exploring one variable at a time – and hence would ignore its interactions with other variables) one can use T_i as a trigger to select columns. The effectiveness of this approach is in principle superior, as we now consider the importance of more columns simultaneously – i.e. with their interactions. This can open the door to all a series of applications in model selection and validation, including to machine learning, see Table 1. This is just an example of a possible applications of a good practice in sensitivity analysis to class of problems dealt with in statistics.

The Journal of the Royal Statistical Society-A has already published articles presenting SA techniques to improve the quality of composite indicators, themselves models of sort. While one application is a standard application of SA to the building of an indicator (Saisana, Saltelli, & Tarantola, 2005) the other shows how one can reason about the importance of a variable transcending the inevitable contradiction associated with the use of weighted linearly aggregated variables (Paruolo, Saisana, & Saltelli, 2013).

Table 1. The dark line separating computable from unknown model needs to be broken by showing how modern SA methods can help in a large class of statistical problems

input	Model	output	Concerned discipline	Role of sensitivity analysis
The x_i can be sampled from	Perfectly known	y is in known deterministically	Applied mathematics	Demonstrated

given distribution for factor X_i	Computable	\tilde{y} may or may not include a stochastic component	Engineering, Computer science	
The \tilde{x}_i are data points, observations	Unknown	\tilde{y} are data points, observations	Statistics	Not-applied

4.3 SENSITIVITY AUDITING

Milton Friedman has famously noted that the most significant theories - those most precious for modelling - are based on the most unrealistic of assumptions (Friedman, 1953; Reinert, Endresen, Ianos, & Saltelli, 2016). This leaves open the questions of deciding what counts as a significant theory, e.g. significant to whom? Joseph Stiglitz has more soberly proposed that the beauty of models is in their acting as blinders, which by leaving a number of things out allow us to see clearly what happens with those elements which are left in (Stiglitz, 2011). Even here there must be criteria, both technical and ethical, to decide that the ‘leaving out’ has been done on transparently formulated normative premises and that the model is still both relevant and reasonable.

The author has heard often modellers claim the ‘neutrality’ of their computer codes. “Computers are impervious to the lure of power”, writes a clever statisticians using computers to fight the practice of gerrymandering (Cho, 2018). The confusion is here between the worth of one’s cause and the neutrality of one’s tool. The opinion of this article is that the technique is never neutral, and hence the use of evidence – including quantitative evidence – must be conceived being prepared for an adversarial setting, unless one’s work is of very solitary academic confinement. In adversarial settings neutrality should gently morph into the relative dependence or independence from arbitrary or implausible assumptions – e.g. in the gerrymandering example, to prove that the district’s boundary configuration serves partisan interests.

Thus, an important element to improve the quality of mathematical modelling should be the extension of the technical dimension of uncertainty to the epistemic and normative dimensions. For this we suggest sensitivity auditing (Saltelli & Funtowicz, 2014; Saltelli et al., 2013). This approach is very recent and few applications are available (Araujo, Saltelli, & Schnepf, 2017; Saltelli & Lo Piano, 2017), though its inclusion into the European Commission’s guidelines for impact assessment (European Commission, 2015) appears promising. Sensitivity auditing is an extension of sensitivity analysis to cover the entire modelling process, inclusive of motivation, power relations, hidden assumptions and normative frameworks. It is based on a seven-point checklist:

Rule 1: ‘Check against rhetorical use of mathematical modelling’; this rule tests if the model elucidates an issue or rather obfuscates it under a veil of math and computing power;

Rule 2: ‘Adopt an “assumption hunting” attitude’; the issues here is: what was ‘assumed out’? What are the tacit, pre-analytical, possibly normative assumptions underlying the analysis?

Rule 3: ‘Detect pseudo-science’; the question here is to detect if the magnitude of model input uncertainties has been instrumentally downplayed (e.g. to obtain a positive inference such as e.g. “the policy will yield a benefit”) or inflated (e.g. to deter action: “the impact of the policy is unclear”).

Rule 4: ‘Find sensitive assumptions before these find you’; this is a plea to anticipate criticism and a reminder that before publishing one’s results a sensitivity analysis should be run and made available.

Rule 5: ‘Aim for transparency’; stakeholders should be able to make sense of, and possibly replicate, the results of the analysis;

Rule 6: ‘Do the right sums’; the analysis should not solve the wrong problem - doing the right sums is more important than doing the sums right. Here the focus is the identity and the legitimacy of the storyteller, and whether other relevant stories could or should be given.

Rule 7: ‘Focus the analysis on the key question answered by the model, exploring holistically the entire space of the assumptions’, see our previous discussion of sensitivity analysis.

Not all rules apply to all models. When a scientific analysis is destined to inform a policy process, the rules become a sensible guide to be developed and implemented with the collaboration of the interested stakeholder and an extended peer-community (Jakeman et al., 2006; Lane et al., 2011).

Rule six on checking the narratives can be extended to a quantitative analysis of the existing frames, using a modicum of quantification to check whether some of the frames can be discarded (Saltelli & Giampietro, 2017). Examples of this approach are to the Programme for International Student Assessment (PISA) by the OECD, (Araujo et al., 2017), to nutrition (Saltelli & Lo Piano, 2017), and to the Ecological Footprint (Giampietro & Saltelli, 2014). Issues of trust and legitimacy in quantification as discussed e.g. in (Porter, 1996) are also brought to bear in this class of approaches.

In relation to the call for transparency one can note that now several journals demand to ‘see the data’ before accepting a paper (Morey et al., 2016) and the open-data scheme is mandatory for all the research projects financed under the H2020 framework. Could the same arrangement be achieved in mathematical modelling? Requirements to make the model available are not unheard of, see e.g. the US regulatory recommendation (Office of Management and Budget, 2006) and the EU guidelines for impact assessment (European Commission, 2015). ‘Seeing the model’ does not only mean making the model available. Journals, government agencies or regulatory body could consider asking for any modelling work a proof of uncertainty and sensitivity analysis, and - when relevant - a discussion of the model’s purpose, funding, validation,

assumptions, process/variable included or excluded, data used in its calibration, and so on (Brauneis & Goodman, 2018; Jakeman et al., 2006).

5 FINAL CONSIDERATIONS

Most modellers I spoke with consent that while complex calculations are now being used to support important decisions, important decisions cannot be justified on the basis of calculated results without an understanding of the associated uncertainties.

At the same time – when confronted with the considerations collected in the present text – some elements of uneasiness do emerge.

One is a deep-rooted resistance to idea of non-neutrality of mathematical modelling. There is here an important cultural and ritual element at play. That a model can be an avenue of possibly instrumental ‘displacement’, in the sense of moving the attention from the system to its model, as discussed in (Rayner, 2012), is still too radical for most practitioners to contemplate. Similarly rule one of sensitivity analysis against excessive or rhetorical complexity is met with a defence of the complex models on the basis of their capacity to possibly ‘surprise’ the analyst. Overall it is noted that modelling is too vast an enterprise to be boxed into a single quality assurance framework. I believe that material exists to give this process a good start (Oberkampf & Roy, 2010; Santner et al., 2003).

A second source of discomfort is the candour – perceived as excessive, of the methodologies advocate here. Discovering that one has to arbitrarily compress the uncertainty in the assumptions in order to obtain a useful inference (Leamer, 1985) is something best left untried. Worse still if the sensitivity analysis were to reveal that the source of the poor (e.g. diffuse) inference is an assumption hard or impossible to nail down or to compress to a lower uncertainty. Along the same discomfort axis is that the invasive probing of sensitivity auditing may invalidate the use of models by showing the futility of, e.g., too ambitious a cost benefit analysis, or by opening the door to relativism – whereby any frame can be upheld given some sort of evidence. These doubts are countered here by rule four of sensitivity auditing – that it is better to deconstruct oneself systematically than to be deconstructed in the field. This consideration is at the core of good scientific practice, as per Robert K. Merton’s principle of ‘Organized Scepticism’ (Merton, 1973), whereby all ideas must be tested and subjected to rigorous, structured community scrutiny. An exposition of this principle was given by Richard Feynman in his Cargo Cult lecture already mentioned, whereby a rigorous analyst ‘bends backward’ to offer to her potential inquisitors all the elements to possibly deconstruct her argument.

In econometrics Mertonian concerns are known as the ‘honesty is the best policy’ approach (Leamer, 2010), and is the 10th law of applied econometrics as formulated by Peter Kennedy: ‘Thou shalt confess in the presence of sensitivity. Corollary: thou shalt anticipate criticism’ (Kennedy, 2008). This reference to sensitivity analysis reconnects us to the topic chosen as an illustration of statistics’ possible rescue of mathematical modelling.

How needed is this rescue? Philosopher Jerome R. Ravetz's prophesized in 1971 that entire research fields might become diseased, (p. 179 of the second edition of (Ravetz, 1971)), and noted: "*reforming a diseased field, or arresting the incipient decline of a healthy one, is a task of great delicacy. It requires a sense of integrity, and a commitment to good work, among a significant section of the members of the field; and committed leaders with scientific ability and political skill.*" It has been argued in the present work that while statistics has been seen to possess the disciplinary arrangements and committed leaders to react to a crisis, mathematical modelling lacks one. In a recent contribution to the Royal Statistical Society magazine 'Significance' (Stark & Saltelli, 2018) a to-do list is suggested for statisticians eager to help a process of reformation of statistical practice and teaching. The present proposal adds a potential item to the list.

6 ACKNOWLEDGMENTS

Philip Stark from the University of Berkeley, Silvio Funtowicz from the University of Bergen (NO), and William Becker from the European Commission's Joint Research Centre offered useful comments and suggestions. The remaining errors are due to the author. This work was partially funded by a Peder Sather grant of the University of Berkeley "Mainstreaming Sensitivity Analysis and Uncertainty Auditing", awarded in June 2016.

7 REFERENCES

- Abraham C. (2016, August 1). Turmoil rocks Canadian biomedical research community. *STAT*. Retrieved from <https://www.statnews.com/2016/08/01/cihr-canada-research/>
- Alexander, L. (2016, August 3). Is an algorithm any less racist than a human? *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2016/aug/03/algorithm-racist-human-employers-work>
- Araujo, L., Saltelli, A., & Schnepf, S. V. (2017). Do PISA data justify PISA-based education policy? *International Journal of Comparative Education and Development*, 19(1), 20–34. <https://doi.org/10.1108/IJCED-12-2016-0023>
- Archer, G. E. B., Saltelli, A., & Sobol', I. M. (1997). Sensitivity measures, anova-like Techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2), 99–120. <https://doi.org/10.1080/00949659708811825>
- Baker, M. (2015). Over half of psychology studies fail reproducibility test. *Nature*. <https://doi.org/10.1038/nature.2015.18248>
- Banobi, J. A., Branch, T. A., & Hilborn, R. (2011). Do rebuttals affect future science? *Ecosphere*, 2(3), 1–11. <https://doi.org/10.1890/ES10-00142.1>
- Becker, W., Paruolo, P., & Saltelli, A. (2014). Exploring Hoover and Perez's experimental designs using global sensitivity analysis. Retrieved from <http://arxiv.org/abs/1401.5617>
- Begley, C. G. (2013). Reproducibility: Six red flags for suspect work. *Nature*, 497(7450), 433–434. Retrieved from <http://dx.doi.org/10.1038/497433a>

- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531–533. Retrieved from <http://dx.doi.org/10.1038/483531a>
- Begley, C. G., & Ioannidis, J. P. A. (2015). Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circulation Research*, 116(1), 116–126. <https://doi.org/10.1161/CIRCRESAHA.114.303819>
- Benessia, A., Funtowicz, S., Giampietro, M., Saltelli, A., Pereira, Â. G., Ravetz, J. R., ... van der Sluijs, J. P. (2016). *Science on the Verge*. Arizona State University. Retrieved from <https://books.google.es/books?id=bjGVjwEACAAJ>
- Bird, S. M., Sir David, C., Farewell, V. T., Harvey, G., Tim, H., & Peter C., S. (2005). Performance indicators: good, bad, and ugly. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1), 1–27. <https://doi.org/10.1111/j.1467-985X.2004.00333.x>
- Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6), 771–784. <https://doi.org/10.1016/J.RESS.2006.04.015>
- Borgonovo, E. (2017). Sensitivity Analysis. An Introduction for the Management Scientist. Springer. https://doi.org/10.1007/978-3-319-52259-3_1
- Borgonovo, E., Castaings, W., & Tarantola, S. (2012). Model emulation and moment-independent sensitivity analysis: An application to environmental modelling, 34, 105–115. <https://doi.org/10.1016/j.envsoft.2011.06.006>
- Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3), 869–887. <https://doi.org/10.1016/J.EJOR.2015.06.032>
- Box, G. E. P., Hunter, W. G., Hunter, & J. Stuart. (1978). *Statistics for experimenters*. New York: Wiley & Sons. <https://doi.org/10.1002/aic.690250233>
- Brauneis, R., & Goodman, E. P. (2018). Algorithmic Transparency for the Smart City. *Yale Journal of Law & Technology*, 20, 103–176. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3012499
- Cacuci, D. G. (2003). *Sensitivity and uncertainty analysis. Volume I, Theory*. Chapman & Hall/CRC Press.
- Campbell, K., Wolfsberg, A., Fabryka-Martin, J., & Sweetkind, D. (2003). Chlorine-36 data at Yucca Mountain: statistical tests of conceptual models for unsaturated-zone flow. *Journal of Contaminant Hydrology*, 62–63, 43–61. [https://doi.org/10.1016/S0169-7722\(02\)00176-6](https://doi.org/10.1016/S0169-7722(02)00176-6)
- Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, 374(9683), 86–89. [https://doi.org/10.1016/S0140-6736\(09\)60329-9](https://doi.org/10.1016/S0140-6736(09)60329-9)
- Cho, W. K. T. (2018). Algorithms can foster a more democratic society. *Nature*, 558(7711), 487–487. <https://doi.org/10.1038/d41586-018-05498-y>
- 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>
- Danaher, J., Hogan, M. ;, Noone, C., Kennedy, R., Behan, A. ;, De Paor, A., ... Schafer, Burkhard; Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data and Society*, 4(2), 1–21.
- Diaz-Maurin, F. (2018, June). Atomic Homefront: a film about struggling to live with Manhattan Project radioactive waste. *Bulletin of the Atomic Scientists*. Retrieved from <https://thebulletin.org/2018/06/atomic-homefront-a-film-about-struggling-to-live-with-manhattan-project-radioactive-waste/>
- Dwyer, J. (2014, August 24). Showing the Algorithms Behind New York City Services. *The New York Times*.

Retrieved from <https://www.nytimes.com/2017/08/24/nyregion/showing-the-algorithms-behind-new-york-city-services.html?nytmobile=0>

- Editors. (2013a). Trouble at the Lab, Scientists Like to Think of Science as Self-Correcting. To an Alarming Degree, It Is Not. *The Economist*. Retrieved from <https://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble>
- Editors. (2013b, October). How science goes wrong. *The Economist*. Retrieved from <https://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong>
- 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(1), 51–61. <https://doi.org/10.1089/ees.2016.0223>
- Edwards, P. N. (1999). Global climate science, uncertainty and politics: Data-laden models, model-filtered data. *Science as Culture*, 8(4), 437–472.
- European Commission. (2015). Guidelines on Impact Assessment - European Commission. Retrieved August 11, 2017, from http://ec.europa.eu/smart-regulation/guidelines/ug_chap3_en.htm
- Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLoS ONE*, 5(4), e10068. <https://doi.org/10.1371/journal.pone.0010068>
- Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2628–2631. <https://doi.org/10.1073/pnas.1708272114>
- Fanelli, D., Anderson, M., Vries, R. de, Rohanizadegan, M., & Evangelou, E. (2010). Do Pressures to Publish Increase Scientists’ Bias? An Empirical Support from US States Data. *PLoS ONE*, 5(4), e10271. <https://doi.org/10.1371/journal.pone.0010271>
- 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>
- Feynman, R. P. (Richard P., Leighton, R., & Hutchings, E. (1985). *Surely you’re joking, Mr. Feynman!* W.W. Norton and Company.
- Flood, A. (2016, November 15). “Post-truth” named word of the year by Oxford Dictionaries. *The Guardian*. Retrieved from <https://www.theguardian.com/books/2016/nov/15/post-truth-named-word-of-the-year-by-oxford-dictionaries>
- Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press.
- Fuller, T. (2017). Anxious relationships: The unmarked futures for post-normal scenarios in anticipatory systems. *Technological Forecasting and Social Change*, 124, 41–50. <https://doi.org/10.1016/j.techfore.2016.07.045>
- Funtowicz, S., & Ravetz, J. R. (1990). *Uncertainty and Quality in Science for Policy*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-0621-1_3
- Funtowicz, S., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures*, 25(7), 739–755. [https://doi.org/10.1016/0016-3287\(93\)90022-L](https://doi.org/10.1016/0016-3287(93)90022-L)
- Galli, A., Giampietro, M., Goldfinger, S., Lazarus, E., Lin, D., Saltelli, A., ... Müller, F. (2016). Questioning the Ecological Footprint. *Ecological Indicators*, 69, 224–232. <https://doi.org/10.1016/j.ecolind.2016.04.014>
- Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem,*

even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Working Paper Department of Statistics, Columbia University. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

- Giampietro, M., & Saltelli, A. (2014). Footprints to nowhere. *Ecological Indicators*, *46*, 610–621. <https://doi.org/10.1016/j.ecolind.2014.01.030>
- Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218. <https://doi.org/10.1177/2515245918771329>
- Gigerenzer, G., & Marewski, J. N. (2015). Surrogate Science. *Journal of Management*, *41*(2), 421–440. <https://doi.org/10.1177/0149206314547522>
- Harris, R. F. (2017). *Rigor mortis : how sloppy science creates worthless cures, crushes hope, and wastes billions*. Basic Books.
- Helton, J. C., Hansen, C. W., & Swift, P. N. (2014). Special Issue: Performance Assessment for the Proposed High-Level Radioactive Waste Repository at Yucca Mountain, Nevada. *Reliability Engineering & System Safety*, *122*, 1–456. Retrieved from <https://www.sciencedirect.com/journal/reliability-engineering-and-system-safety/vol/122>
- Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, *52*(1), 1–17. [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6)
- Hornberger, G. M., & Spear, R. C. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, *12*(1).
- Horton, R. (2015). Offline: What is medicine’s 5 sigma? *The Lancet*, *385*, 1380. [https://doi.org/10.1016/S0140-6736\(15\)60696-1](https://doi.org/10.1016/S0140-6736(15)60696-1)
- Ioannidis, J. P. A., Califf, R., Califf, R., Platt, R., Selby, J., Greenberg, D., & Neumann, P. (2016). Why Most Clinical Research Is Not Useful. *PLOS Medicine*, *13*(6), e1002049. <https://doi.org/10.1371/journal.pmed.1002049>
- Ioannidis, J. P. A., Chen, J., Kodell, R., Haug, C., & Hoey, J. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P. A., Forstman, B., Boutron, I., Yu, L., & Cook, J. (2014). How to Make More Published Research True. *PLoS Medicine*, *11*(10), e1001747. <https://doi.org/10.1371/journal.pmed.1001747>
- Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, *127*, F236–F265.
- Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software*, *21*(5), 602–614.
- Kennedy, P. (2008). *A guide to econometrics*. Wiley-Blackwell; 6 edition.
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Kopp, R., Rasmussen, D., & Mastrandrea, M. (2014). *American Climate Prospectus: Economic Risks in the United States* (Vol. 1.2). Retrieved from https://gspp.berkeley.edu/assets/uploads/research/pdf/American_Climate_Prospectus.pdf
- Lane, S. N., Odoni, N., Landström, C., Whatmore, S. J., Ward, N., & Bradley, S. (2011). Doing flood risk science differently: an experiment in radical scientific method. *Transactions of the Institute of British Geographers*, *36*(1), 15–36. <https://doi.org/10.1111/j.1475-5661.2010.00410.x>

- Leamer, E. E. (1985). Sensitivity Analyses Would Help. *The American Economic Review*, 75(3), 308–313.
- Leamer, E. E. (2010). Tantalus on the Road to Asymptopia. *Journal of Economic Perspectives*, 24(2), 31–46. <https://doi.org/10.1257/jep.24.2.31>
- 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.
- Leek, J., & Peng, R. D. (2015). P values are just the tip of the iceberg. *Nature*, 520, 612.
- Mancha, R., & Ali, H. (2017). The Unexamined Algorithm Is Not Worth Using. *Stanford Social Innovation Review*, (October 11). Retrieved from https://ssir.org/articles/entry/the_unexamined_algorithm_is_not_worth_using
- Martin, F. (2018, June 13). Unelected power: banking’s biggest dilemma. *New Statesmen*. Retrieved from <https://www.newstatesman.com/culture/books/2018/06/unelected-power-banking-s-biggest-dilemma>
- McCook, A. (2017, April). A new record: Major publisher retracting more than 100 studies from cancer journal over fake peer reviews. *Retraction Watch*. Retrieved from <http://retractionwatch.com/2017/04/20/new-record-major-publisher-retracting-100-studies-cancer-journal-fake-peer-reviews/>
- McShane, B. B., & Gal, D. (2017). Statistical Significance and the Dichotomization of Evidence. *Journal of the American Statistical Association*, 112(519), 885–895. <https://doi.org/10.1080/01621459.2017.1289846>
- Merton, R. (1973). *The sociology of science: Theoretical and empirical investigations*. (University of Chicago Press., Ed.). Retrieved from <https://books.google.com/books?hl=en&lr=&id=zPvcHuJMEMwC&oi=fnd&pg=PR9&ots=x6LLWhk3rR&sig=UI8K57WfiB-ksqNzrB-hn7R0hYU>
- Metlay, D. (2000). From Tin Roof to Torn Wet Blanket: Predicting and Observing Ground Water Movement at a Proposed Nuclear Waste Site. In D. Sarewitz, R. A. J. Pielke, & R. Byerly (Eds.), *Prediction: Science, Decision Making, and the Future of Nature*. Island Press.
- Mirowski, P. (2011). *Science-Mart, Privatizing American Science*. Harvard University Press. Retrieved from <https://books.google.es/books?id=wBGw2Z8bnAgC>
- Mirowski, P. (2013). *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown*. Verso. Retrieved from <https://books.google.es/books?id=5mdelx-86jwC>
- Moeller, H. G. (2006). *Luhmann explained*. Open Court Publishing Company.
- Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., ... Zwaan, R. A. (2016). The Peer Reviewers’ Openness Initiative: incentivizing open research practices through peer review. *Royal Society Open Science*, 3, 150547. Retrieved from <http://rsos.royalsocietypublishing.org/content/3/1/150547>
- Morrison, F. (1991). *The art of modeling dynamic systems : forecasting for chaos, randomness, and determinism*. Dover Publications.
- Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press .
- Nordhaus, W. D. (1991). To Slow or Not to Slow: The Economics of The Greenhouse Effect. *The Economic Journal*, 101(407), 920–937. <https://doi.org/10.2307/2233864>
- Norton, J., & John. (2015). An introduction to sensitivity assessment of simulation models. *Environmental Modelling & Software*, 69(C), 166–174. <https://doi.org/10.1016/j.envsoft.2015.03.020>

- Nuzzo, R. (2015). How scientists fool themselves – and how they can stop. *Nature*, 526(7572), 182–185. <https://doi.org/10.1038/526182a>
- O’Neil, C. (2016). *Weapons of math destruction : how big data increases inequality and threatens democracy*. Random House Publishing Group.
- Oberkampf, W. L., & Roy, C. J. (2010). *Verification and validation in scientific computing*. Cambridge University Press.
- Office of Management and Budget. (2006). *Proposed Risk Assessment Bulletin*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/omb/inforeg/proposed_risk_assessment_bulletin_010906.pdf
- Open Science Collaboration (OSC). (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716-aac4716. <https://doi.org/10.1126/science.aac4716>
- Padilla, J. J., Diallo, S. Y., Lynch, C. J., & Gore, R. (2018). Observations on the practice and profession of modeling and simulation: A survey approach. *SIMULATION*, 94(6), 493–506. <https://doi.org/10.1177/0037549717737159>
- Paruolo, P., Saisana, M., & Saltelli, A. (2013). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3), 609–634. <https://doi.org/10.1111/j.1467-985X.2012.01059.x>
- Pierce, A. (2008, November 5). The Queen asks why no one saw the credit crunch coming. *The Telegraph*.
- Pilkey, O. H., & Pilkey-Jarvis, L. (2009). *Useless Arithmetic: Why Environmental Scientists Can’t Predict the Future*. Columbia University Press.
- Popp Berman, E., & Hirschman, D. (2018). The Sociology of Quantification: Where Are We Now? *Contemporary Sociology*, 47(3), 257–266.
- Porter, T. M. (1996). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press. Retrieved from <https://books.google.es/books?id=oK0QpgVfINOC>
- Ravetz, J. R. (1971). *Scientific knowledge and its social problems*. Oxford University Press.
- Ravetz, J. R. (1997). *Integrated Environmental Assessment Forum, developing guidelines for “good practice”, Project ULYSSES*. Retrieved from <http://www.jvds.nl/ulysses/eWP97-1.pdf>
- Ravetz, J. R. (2018). Integrity must underpin quality of statistics. *Nature*, 553(7688), 281–281. <https://doi.org/10.1038/d41586-018-00648-8>
- Rayner, S. (2012). Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses. *Economy and Society*, 41(1), 107–125. <https://doi.org/10.1080/03085147.2011.637335>
- Refsgaard, J. C., van der Sluijs, J. P., Brown, J., & van der Keur, P. (2006). A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources*, 29(11), 1586–1597. Retrieved from <http://www.jstor.org/stable/2225764>
- Reinert, E. S., Endresen, S., Ianos, I., & Saltelli, A. (2016). Epilogue: the future of economic development between utopias and dystopias. In *Handbook of Alternative Theories of Economic Development* (pp. 738–786). Edward Elgar Publishing. Retrieved from https://ideas.repec.org/h/elg/eechap/15311_40.html
- Rocquigny, E. de., Devictor, N., & Tarantola, S. (2008). *Uncertainty in industrial practice : a guide to quantitative uncertainty management*. John Wiley & Sons.
- Romer, P. (2015a). Feynman Integrity. Retrieved from <https://paulromer.net/feynman-integrity/>

- Romer, P. (2015b). Mathiness in the Theory of Economic Growth. *American Economic Review*, 105(5), 89–93. <https://doi.org/10.1257/aer.p20151066>
- Romer, P. (2018). The trouble with macroeconomics. *The American Economist*, *forthcomin*. Retrieved from <https://paulromer.net/wp-content/uploads/2016/09/WP-Trouble.pdf>
- Rosen, R. (1991). *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life*. Columbia University Press. Retrieved from <https://books.google.es/books?id=DR8L4snDnkIC>
- Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2), 307–323. <https://doi.org/10.1111/j.1467-985X.2005.00350.x>
- Saltelli, A. (2002a). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2), 280–297. [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1)
- Saltelli, A. (2002b). Sensitivity Analysis for Importance Assessment. *Risk Analysis*, 22(3), 579–590. <https://doi.org/10.1111/0272-4332.00040>
- Saltelli, A. (2018). Why science’s crisis should not become a political battling ground. *Futures*. <https://doi.org/doi.org/10.1016/j.futures.2018.07.006>
- Saltelli, A., Aleksankina, K., William, B., Fennell, P., Ferretti, F., Holst, N., ... Wu, Q. (2017). Why So Many Published Sensitivity Analyses Are False. A Systematic Review of Sensitivity Analysis Practices. *ArXiv*, 1711.11359.
- Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12), 1508–1517. <https://doi.org/10.1016/j.envsoft.2010.04.012>
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259–270. <https://doi.org/10.1016/j.cpc.2009.09.018>
- Saltelli, A., Chan, K., & Scott, M. (2000). *Sensitivity analysis*. Wiley.
- Saltelli, A., & d’Hombres, B. (2010). Sensitivity analysis didn’t help. A practitioner’s critique of the Stern review. *Global Environmental Change*, 20(2), 298–302. <https://doi.org/10.1016/j.gloenvcha.2009.12.003>
- Saltelli, A., & Funtowicz, S. (2014). When All Models Are Wrong. *Issues in Science and Technology*, 30(2), 79–85. Retrieved from <http://www.jstor.org/stable/43315849>
- Saltelli, A., & Funtowicz, S. (2017). What is science’s crisis really about? *Futures*, 91, 5–11. <https://doi.org/10.1016/j.futures.2017.05.010>
- Saltelli, A., Funtowicz, S., Giampietro, M., Sarewitz, D., Stark, P. B., & van der Sluijs, Jeroen, P. (2016). Modelling: Climate costing is politics not science. *Nature*, 532(7598), 177–177. <https://doi.org/10.1038/532177a>
- Saltelli, A., & Giampietro, M. (2017). What is wrong with evidence based policy, and how can it be improved? *Futures*, 91, 62–71. <https://doi.org/10.1016/j.futures.2016.11.012>
- Saltelli, A., Guimaraes Pereira, Â., van der Sluijs, J. P. ., & Funtowicz, S. (2013). What do I make of your latinorumc Sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy*, 9(2/3/4), 213–234. <https://doi.org/10.1504/IJFIP.2013.058610>
- Saltelli, A., & Lo Piano, S. (2017). Problematic quantifications: a critical appraisal of scenario making for a global ‘sustainable’ food production. *Food Ethics*, 1(2), 173–179.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global*

sensitivity analysis : the primer. John Wiley.

- Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2012). Update 1 of: Sensitivity Analysis for Chemical Models. *Chemical Reviews*, 112(5), PR1-PR21. <https://doi.org/10.1021/cr200301u>
- Saltelli, A., & Stark, P. B. (2018). Fixing statistics is more than a technical issue. *Nature*, 553(7688), 281–281. <https://doi.org/10.1038/d41586-018-00647-9>
- Saltelli, A., Stark, P. B., Becker, W., & Stano, P. (2015). Climate Models as Economic Guides - Scientific Challenge of Quixotic Quest? *Issues in Science and Technology*, 31(3), 1–8. Retrieved from <https://www.stat.berkeley.edu/~aldous/157/Papers/saltelliEtal15>
- Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*, 15(4), 377–395. <https://doi.org/10.1214/ss/1009213004>
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470870958>
- Saltelli, A., Tarantola, S., & Chan, K. P.-S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1), 39–56. <https://doi.org/10.1080/00401706.1999.10485594>
- Santner, T. J., Williams, B. J., & Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.
- Shanks, D. R., Vellido, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., ... Puhmann, L. M. C. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*, 144(6), e142–e158. <https://doi.org/10.1037/xge0000116>
- Singh Chawla, D. (2017). Big names in statistics want to shake up much-maligned P value. *Nature*, 548(7665), 16–17. <https://doi.org/10.1038/nature.2017.22375>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(160384).
- Stark, P. B., & Saltelli, A. (2018, July). Cargo-cult statistics and scientific crisis. *Significance*. Retrieved from <https://www.significancemagazine.com/2-uncategorised/593-cargo-cult-statistics-and-scientific-crisis>
- Stern, N. H. (Nicholas H. (2015). *Why are we waiting? : the logic, urgency, and promise of tackling climate change*. MIT Press.
- Stiglitz, J. E. (2011). Rethinking Macroeconomics: What Failed, and How to Repair It. *Journal of the European Economic Association*, 9(4), 591–645. <https://doi.org/10.7916/D8WS9420>
- Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. *Sustainable Development*, 12, 292. <https://doi.org/10.2139/ssrn.1714428>
- Szenberg, M. (1992). *Eminent economists : their life philosophies*. Cambridge University Press. Retrieved from <http://admin.cambridge.org/gb/academic/subjects/economics/history-economic-thought-and-methodology/eminent-economists-their-life-philosophies#dkQwZVJ4RazyzwHC.97>
- U.S. Environmental Protection Agency (EPA). (2009). *Guidance on the Development, Evaluation, and Application of Environmental Models*. Retrieved from <http://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P1003E4R.PDF>
- van der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. R., & Risbey, J. (2005). Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System. *Risk Analysis*, 25(2), 481–492. <https://doi.org/10.1111/j.1539-6924.2005.00604.x>

- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 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>
- Weinberg, A. (1972). Science and trans-science. *Minerva*, 10, 209–222.
- Wilmott, P., & Orrell, D. (2017). *The Money Formula*. Wiley & Sons.
- Wynne, B. (1992). Uncertainty and environmental learning 1, 2Reconceiving science and policy in the preventive paradigm. *Global Environmental Change*, 2(2), 111–127. [https://doi.org/10.1016/0959-3780\(92\)90017-2](https://doi.org/10.1016/0959-3780(92)90017-2)
- Zidek, J. V. (2006). Editorial: (Post-normal) statistical science. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(1), 1–4. <https://doi.org/10.1111/j.1467-985X.2005.00393.x>