# What do I make of your latinorum? Sensitivity auditing of mathematical modelling

Andrea Saltelli* and
Ângela Guimarães Pereira

European Commission,
Joint Research Centre,
Institute for the Protection and Security of the Citizen,
Via E. Fermi, 2749, 21027 Ispra (VA), Italy
E-mail: andrea.saltelli@jrc.ec.europa.eu
E-mail: angela.pereira@jrc.ec.europa.eu
*Corresponding author

Jeroen P. Van der Sluijs

Copernicus Institute of Sustainable Development,
Utrecht University (NL),
Heidelberglaan 2, 3584 CS Utrecht, The Netherlands
E-mail: j.p.vandersluijs@uu.nl

Silvio Funtowicz

Centre for the Study of the Sciences and the Humanities (SVT)
Allegaten, University of Bergen (NO),
34 – Postboks 7805 5020 Bergen, Norway
E-mail: silvio.funtowicz@svt.uib.no

**Abstract:** Sensitivity analysis, mandated by existing guidelines as a good practice to use in conjunction to mathematical modelling, is as such insufficient to ensure quality in the treatment of *uncertainty of science for policy*. If one accepts that policy-related science calls for an extension of the traditional internal, peer review-based methods of quality assurance to higher levels of supervision, where extended participation and explicit value judgments are necessary, then by the same token sensitivity analysis must extend beyond the technical exploration of the space of uncertain assumptions when the inference being sought via mathematical modelling is subject to relevant uncertainties and stakes. We thus provide seven rules to extend the use of sensitivity analysis (or how to apportion uncertainty in model-based inference among input factors) in a process of sensitivity auditing of models used in a policy context. Each rule will be illustrated by examples.

**Keywords:** sensitivity analysis; sensitivity auditing; NUSAP; post normal science; PNS; knowledge quality assessment.

**Reference** to this paper should be made as follows: Saltelli, A., Guimarães Pereira, Â., Van der Sluijs, J.P. and Funtowicz, S. (2013) 'What do I make of your latinorum? Sensitivity auditing of mathematical modelling', *Int. J. Foresight and Innovation Policy*, Vol. 9, Nos. 2/3/4, pp.213–234.

**Biographical notes:** Andrea Saltelli has worked on physical chemistry, environmental sciences and applied statistics, publishing over 80 peer reviewed papers and three books. His main disciplinary focus is on *sensitivity analysis* of model output, a discipline where statistical tools are used to interpret the output from mathematical or computational models, and on *sensitivity auditing*, an extension of sensitivity analysis to the entire evidence-generating process in a policy context. A second focus is the construction of *composite indicators* or indices. Presently, he leads the Econometric and Applied Statistics Unit of the European Commission at the Joint Research Centre in Ispra (I). The unit, with a staff of 30, develops econometric and statistic applications, mostly in support to the services of the European Commission, in fields such as lifelong learning, inequality, employment, competitiveness and innovation. He participates in the training of European Commission staff on impact assessment.

Ângela Guimarães Pereira is a Scientific Officer of the European Commission at the Joint Research Centre. She holds a PhD in Environmental Science. Her research focuses on multiple forms in which interfaces between science and society occur, covering aspects of public engagement, ethics dialogues and communication about techno-science developments. Through European projects, she has conducted research on anticipation and knowledge assessment in connection to environmental governance and technology development. She has published widely and is co-editor of a number of books, namely *Interfaces between Science and Society with Greenleaf and Science for Policy* with Oxford University Press.

Jeroen P. Van der Sluijs is an Associate Professor at the Copernicus Institute, Utrecht University. His work focuses on those situations where scientific assessment is used as a basis for policymaking on environmental risks before conclusive scientific evidence and scientific consensus are available on the causal relationships, the magnitude, and the probabilities of these risks. His work seeks to understand and improve the science-policy interface in a context of scientific controversy and uncertainty by contributing and applying deliberative methods and tools for knowledge quality assessment. He has published 72 journal articles and contributed 25 chapters to edited volumes.

Silvio Funtowicz taught mathematics, logic and research methodology in Buenos Aires, Argentina. During the decade of 1980, he was a Research Fellow at the University of Leeds, England. Since 1989 and until his retirement in 2011, he was a member of the European Commission's Joint Research Centre. Since then, he is a Professor at the Centre for the Study of the Sciences and the Humanities (SVT) of the University of Bergen, Norway. He is the author of *Uncertainty and Quality in Science for Policy* (1990, in collaboration with Jerry Ravetz) and numerous papers in the field of environmental and technological risks and policy-related research. He has lectured extensively and he is a member of the editorial board of several publications and the scientific committee of many projects and international conferences.

## 1 Introduction

In this paper we argue that the quality assessment of mathematical or simulation models that underpin current policy-making requires a process which transcends the mere assessment of the model uncertainties and parametric sensitivities (sensitivity analysis) up to including a practice of organised scepticism toward the inference provided by mathematical models (sensitivity auditing). This process will also implicitly look into plausibility assertions embedded in the creation and the use of those models. Sensitivity tools need to be adopted which ensure a complete exploration of the space of the input uncertainties, but at the same time the boundaries of such a space need to be questioned, if needed, also by a process of extended peer-review which cuts across disciplines as well as across the fence separating practitioners from stakeholders. Sensitivity auditing also needs to cope with non-quantifiable uncertainties, eschewing the hubris of quantification at all cost.

We consider this enhancement of sensitivity analysis as necessary and urgent. Examples of the strategic use of mathematical modelling are not rare to find, as will be shown in the present work. We call strategic – for the purpose of the present work – those instances where mathematics and large-sized models are used more to obfuscate than to illustrate, not unlike the use of Latin by the elites in the classic age, whereby the language of the law and of the church was used as a vulgus-baffling device. "What do I make of your Latinorum?" which in Italian rends into the satirical iambic cadence of *Che vuol ch'io faccia del suo latinorum*, from Alessandro Manzoni's The Betrothed, is the retort of Renzo – a character in the novel, to the rector of his parish who robs Renzo of his planned marriage, and tries to confuse him with ecclesiastic Latin.

Further to this, as practitioners, we are puzzled to see that even when an appraisal of model sensitivities is attempted by modellers, this is often of poor or perfunctory quality (Saltelli and d'Hombres, 2010; Saltelli and Annoni, 2010). In this sense our work feeds into that current of thought, which takes issue at the poor quality of existing modelling practices (Taleb, 2007; Pilkey and Pilkey-Jarvis, 2007; Savage, 2009).

Note that we do not deal here with strategies for model updating, data assimilation, model optimisation, model averaging, ensemble modelling or any of the tools which may be used to perfect the predictive capacity of a model. We take the model at the moment of its 'use' in a policy context, where all previous knowledge and information has been already processed and included in the proposed model representation, i.e., where the model is already in its best available formulation, be it in the form of an ensemble or an average of models. It is at this point that sensitivity analysis must be performed, either by the modellers themselves or by a third party, to inform the users of what drives the model predictions and their uncertainty. Likewise it is at the moment of model use, e.g., in an ex-ante impact assessment, that the extension of sensitivity analysis to sensitivity auditing finds its use. In this sense tools such as model averaging become the substance of the analysis provided by sensitivity auditing. For example a model's pedigree (see Section 2.2) may include the history of its calibration.

Writing in *NATURE*, Stainforth and Smith (2012) suggest that "*the public-image problem of current models stems partly from scientists' failures to identify the limitations openly*". Targeting especially those models that probe possibilities, they suggest that "*the role of science is to probe on the plausibility and relevance of such possibilities*". There appears to be a lack of systematic endeavour on model-specific studies which would question a model's plausibility, the credibility and legitimacy of its implicit rhetoric, and how its assumptions are established and justified. As shown later, these doubts about the worth of modelling are not confined to the scientific literature, or to specialised blogs, but are echoed in the press wherever the issue at hand commands public attention.

We suggest an organised critical appraisal of model quality that we call 'sensitivity auditing', which scrutinises to what extent a model can fit the purpose of informing and justifying policy analysis or intervention. The focus of the auditing transcends the model itself and encompasses the entire modelling process. Sensitivity auditing borrows ideas and strategies from sensitivity analysis (Saltelli et al., 2000, 2008, 2010), from the NUSAP system for multidimensional uncertainty assessment (Funtowicz and Ravetz, 1990; Van der Sluijs et al., 2005) and from post-normal science (Funtowicz and Ravetz, 1993). In sensitivity auditing 'what to look for' is as important as 'how to look for' and 'who should do the looking'.

We start by providing some background of the intellectual context from which we move; we then introduce sensitivity auditing as a continuous process of 'vigilance'; we discuss what rules this process should have and how these should be implemented. Finally we link this endeavour to the central concept of the present special issue: plausibility.

## 2 Background

In this section we introduce the frameworks and experiences that inspire sensitivity auditing, and how these can be used to build a theory for the analysis of a model's plausibility.

### *2.1 Post normal science*

In their book 'uncertainty and quality of science for policy' (1990), Funtowicz and Ravetz developed new conceptual and practical tools for coping with uncertainty and the assurance of quality of quantitative information in policy-related research. Complementing this effort, Funtowicz and Ravetz (1993) also introduced a novel mode of scientific problem-solving suitable to policy issues where facts are uncertain, values in dispute, stakes high and decisions urgent. They called it 'Post Normal Science' (PNS), to relate it to Tomas Kuhn's (1962) book defining normal science and to distinguish it from Stephen Toulmin's (1985) postmodern science.

PNS is a quest for the appropriate management of quality in the presence of irreducible uncertainty (Knight, 1921); it comprises an awareness of the role of values and the acceptance of a plurality of commitments and perspectives. These are expressed through an extended peer community, involving all relevant scientific disciplines, as well as concerned citizens (as in Feyerabend, 1975) and a plurality of stakeholders in the tasks of problem framing, assessment and quality control.

PNS emerges from the realisation that major societal issues involving risk and uncertainty are poorly dealt with by modern science rigidly organised along disciplinary lines (see chapter 9 in Toulmin, 2001), and under the paradigm of 'sound science'. PNS embraces complexity (including in the set of norms and values) and fosters a new system of scientific governance. The purpose is to enable a plurality of different though legitimate perspectives to be brought to bear on the debate in a reflexive fashion. In this way, PNS calls for attention to the so-called type III error (Raifa, 1968; Dunn, 1997), which manifests itself when an issue is framed at the exclusion of one or more relevant and legitimate constituencies.

## *2.2 NUSAP*

In the tradition of PNS, several new multidimensional and reflective approaches have been developed to systematically address unquantifiable dimensions of uncertainty. The most widely known is the NUSAP system for multidimensional uncertainty assessment (Funtowicz and Ravetz, 1990; Van der Sluijs et al., 2005). NUSAP is a tool and notational system for the analysis and diagnosis of uncertainty in science for policy. It responds to the limited used of e.g. Monte Carlo techniques since these only address quantitative dimensions of uncertainty, whereas what we usually face in policy relevant models is a very complex mass of uncertainties involving technical, methodological, epistemological and societal dimensions. With NUSAP one aims to qualify quantities - especially those that feed into the policy process - by using the five qualifiers of the NUSAP acronym:

- *Numeral*, the numerical value of the claimed quantity
- *Unit*, its units
- *Spread*, a measure of (statistical or measurement) error
- *Assessment*, an assessment of the reliability of the claim made by experts
- *Pedigree*, which conveys an evaluative account of the production process of the quantity, and indicates different aspects of its underpinning and scientific status. Pedigree is expressed by means of a set of pedigree criteria to assess these different aspects in models, such as empirical basis, methodological rigour, degree of validation and the usage of proxy representations [see for example, Van der Sluijs (2005, 2010) in relation to climate models].

The NUSAP approach is adopted in the Netherlands as part of the Guidance on Uncertainty Assessment and Communication of the Netherlands Environmental Assessment Agency (Van der Sluijs et al., 2008; Petersen et al., 2011).

## *2.3 Sensitivity analysis*

There is a consensus among practitioners from a plurality of disciplines (Kennedy, 2007; Leamer, 1990; Pilkey and Pilkey-Jarvis, 2007; Saltelli et al., 2008, 2010; Santner et al., 2003; Oakley and O'Hagan, 2004; Saisana et al., 2005) as well as among guidelines devoted to modelling and impact assessment (EC, 2009; EPA, 2009; OMB, 2006), that sensitivity analysis is indispensable to judge the quality of inference based on mathematical models.

A definition of global sensitivity analysis is "The study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs" (Saltelli, 2002).

Good practices in sensitivity analysis (see a recent review at Saltelli et al., 2012) prescribe that the uncertainty in the inference be quantified by a simultaneous activation of all possible uncertainties in the assumptions underlying the analysis, followed by an identification of those assumptions chiefly responsible for the uncertainty in the inference. Being numerical experiments, these analyses should be implemented following a statistical design, as one expects for a physical or biological experiment. Assumptions become then factors, whose effect is explored using techniques partly derived from experimental design, a branch of applied statistics.

EPA (2009) describes well what an ideal sensitivity analysis must do:

> "[SA] methods should preferably be able to deal with a model regardless of assumptions about a model's linearity and additivity, consider interaction effects among input uncertainties, […], and evaluate the effect of an input while all other inputs are allowed to vary as well."

A class of methods which fulfil EPA's technical requirements is based on decomposing the variance of the inference according to bits which can be attributed to either input factors or combinations of factors: the so-called interactions. This kind of analysis is only successful to the extent that all sources of uncertainties have been identified, which is in most cases impossible to prove (see rule 7 – Section 4 of this paper), and that the model is relevant to the issue being analysed. These vast limitations of a technical sensitivity analysis should not justify omitting it or performing it in a perfunctory way. As discussed in Saltelli and Annoni (2010), notwithstanding existing guidelines, most sensitivity analyses seen in the literature tend to display a cavalier attitude with respect to statistical design, model non-linearity and model non-additivity issues (see rule 7 – Section 4 of this paper). For instance, doing sensitivity analysis by varying one factor at a time leaves the investigator in the dark as to what happens in the largest part of the input space, as well as to the effect of factors' interaction. As noted in The Flaw of Averages (Savage, 2009) when a scaffold is made by coupling several folding ladders with planks, one cannot 'shake' one ladder at a time to test the safety of the scaffold. A better idea of the stability of the scaffold is obtained by shaking all ladders simultaneously. The fact that books are written to make this self-evident point suggests that it is not necessarily shared by all practitioners.

## 3 Sensitivity auditing

Sensitivity auditing aims to extend sensitivity analysis to contexts where mathematical modelling feeds into a policy context. More generally, sensitivity auditing gauges the quality of scientific information in all cases where models are at play and their outcome feeds into the public discourse, be it in the context of a policy assessment (*ex ante* or *ex post*), or in the public arenas where policies are contested. Sensitivity auditing starts from the awareness that in an adversarial context, not only the nature of the evidence, but also the degree of certainty and uncertainty associated to the evidence, will be the subject of partisan interests. It encompasses the ideas of PNS exposed earlier and its associated concept of quality assurance by an *extended peer community*. An *extended peer community* consists not merely of persons with some form or other of institutional accreditation, but of all those with a desire and/or interest to participate in *extended peer review* processes for the resolution of a specific issue. Sensitivity auditing thus demands spaces where relevant social actors are enabled and invited to scrutinise modelling activities and their policy applications. The consideration and inclusion of actors' specific knowledge ultimately adds to the plausibility of model-based inference. As noted by Feyerabend (2010, p.262).

> "[…] in a democracy local populations not only will, but also should, use the sciences in ways most suitable to them, The objections that citizens do not have the expertise to judge scientific matters overlooks that important problems often lie across the boundaries of various sciences so that scientists within these sciences don't have the needed expertise either. Moreover doubtful cases always produce experts from one side, experts for the other side, and experts in between. But the competence of the general public could be vastly improved by an education that exposes expert fallibility instead of acting as if it did not exist."

The set of rules presented here for sensitivity auditing presupposes that an interested 'extended peer community' is identified, and eventually involved in the auditing of the mathematical modelling. As will be discussed later, a virtual community already exists which debates all that happens in the field of modelling of climate change, see, e.g., http://www.wattsupwiththat.com, and http://allmodelsarewrong.com/.

Useful recipes for sensitivity auditing proposed here are:

1. check against rhetoric use of mathematical modelling
2. adopt an 'assumption hunting' attitude
3. detect garbage in garbage out (GIGO), in the extended definition of Funtowicz and Ravetz (1990)
4. find sensitive assumptions before these find you
5. aim for transparency
6. do the right sums
7. focus the analysis on the key question answered by the model, exploring holistically the entire space of the assumptions.

Note that while some of these rules are a pointer to good versus bad practices (e.g., rule 7) some other call for a shift in the stance of the analysts (e.g., rule 1).

Before going into the rules in detail, we would like to motivate the need for their introduction with some examples in which modelling used to underpin policy has shown evident symptoms of dysfunction.

### *3.1 The financial crisis and the modelling of collateralised debt obligations*

This is a quite well known story which concerns the formula of David X. Li used in the pricing of collateralised debt obligations (the infamous CDO's). The story is popularised in an article on the magazine *Wired* (Salmon, 2009), where one learns how the toxicity of these securities (which packed as many as two thousand individual mortgages into a single obligation) was elegantly overlooked by applying a modelling approach (Gaussian Copula) whereby the probability of joint default of any couple of individual mortgages in the bundle was described by a correlation coefficient estimated on historic data. Unfortunately the 'history' on which this parameter was estimated was a short one, and only relative to a period of housing market upswing; thus the probability of the joint failure of two mortgages was very low in the world of the model. The story changed when the housing bubble exploded, whereby Li's formula lost any predictive power in the real world. Of course this accident could not have been overlooked by the 'quants', the mathematicians who are employed in the world of finance. 'Anything that relies on correlation is charlatanism', noted Nassim N. Taleb (cited *ibidem*). The point of the anecdote is that when important stakes are at play, the normative stance of all actors – including scientists, must be questioned openly. Yet it would be unfair to finger the quants as the sole modellers with a responsibility for the crisis. As amply debated in the press and the specialised literature, it was the entire macro-economic modelling fabric that was found wanting: "*The standard macroeconomic models have failed, by all the most important tests of scientific theory. They did not predict that the financial crisis would happen; and when it did, they understated its effects*" [Stiglitz, (2011), p.591]. We shall return to the issue in Section 4.

### *3.2 Dutch overhead power lines cause 0.5 cases of child leukaemia per year, model says*

In 2000, the Health Council of the Netherlands reviewed the epidemiological state of knowledge on health risks of extreme low frequency electro magnetic fields (ELF EMFs) and concluded that a 'relatively consistent association between the occurrence of childhood leukaemia and living in the vicinity of overhead power lines' exists. In response, the Ministry asked the Netherlands Institute for Public Health and the Environment (RIVM) to quantify what the risk of overhead power lines for the Netherlands population would be if one would assume that the association is causal. Making use of estimations on numbers of dwellings in different (magnetic) zones close to overhead power lines, (RIVM) translated the *relative risks* found in international pooled analyses into an annual number of extra cases of childhood leukaemia (Van der Plas et al, 2001; Pruppers, 2003). Their chain of calculations resulted in the claim that overhead power lines add 0.4-0.5 extra cases leukaemia annually in NL (to a total of 110 cases per year). To enable the quantification requested by the Ministry, RIVM had to make a vast amount of assumptions, both prior to and in the model calculation chain. Not all of these were stated explicitly in the report. de Jong et al. (2012) applied the 'assumption hunting' approach (see rule 2, Section 4 of this paper) to deconstruct RIVM's model calculation. In a first step, 35 assumptions were identified. In an expert workshop that included RIVM experts involved, the list of assumptions was reviewed, completed, and ranked according to (estimated) ordinal importance with regard to influence on the outcome of the calculation. The top five assumptions are listed in Table 1. Next, the pedigree of each was assessed. The assumptions with the highest expected impact on the number of extra cases of child leukaemia turned out to be also the ones with the lowest pedigree: many of these assumptions are difficult to underpin and highly value-laden, i.e. dependent upon the normative framework espoused by the observer. Moreover, the assumption-hunting workshop found that the assumptions which are regarded to be most problematic *are prior to the model calculation chain developed by the RIVM*: they are hidden in numbers that are taken from disciplines other than those purportedly involved in the analysis, such as the *relative risk* factors established in the pooled analysis of epidemiological studies. This finding highlights the key importance of a wide extension of the peer community engaged in model quality control.

**Table 1** Top 5 of assumptions in overhead power lines health risk study

| *Rank* | *Assumption* |
|---|---|
| 1 | A causal relationship exists between exposure to electromagnetic fields of overhead power lines and the occurrence of childhood leukaemia |
| 2 | Overhead power lines are the main differentiating source of exposure to electromagnetic fields for children |
| 3 | The height of the (prolonged) average of exposure causes the effect |
| 4 | A threshold value exists |
| 5 | The current in the year prior to determining the incidence of childhood leukaemia is representative for the average current during the development of childhood leukaemia |

*Source:* de Jong et al. (2012)

### *3.3 AIDS*

The next example comes from the excellent book by Orrin H. Pilkey and Linda Pilkey-Jarvis (2007, pp.36–38), called Useless Arithmetic: Why Environmental Scientists Can't Predict the Future, whose title is eloquent enough. The book describes many instances of dysfunctional policy advice supported by mathematical modelling. Each of the stories is compelling, and the book is, in our view, a must-read for modellers, regulators, and any stakeholder who desires to learn some unpleasant truths about how models are built and used. The story we report here regards the use of mathematical modelling by a UN agency concerned with AIDS (UNAIDS). The authors note that 250,000 would die in 1999 according to Epi Model (an epidemiological model, Chin and Lwanga, 1991). But that year 375,000 died of all causes. The number of AIDS victims is far too large a proportion, 2/3 of the total deaths. Another model predicted 143000 deaths of AIDS. In 2001, the much advanced ASSA 2000 model predicted that there must have been 92000 AIDS deaths. Noting that the estimates are quite arduous due to the poor quality of the input, Pilkey and Pilkey-Jarvis (2007) make the point that this overestimate of AIDS death numbers is not without consequences: it might result in under-spending on fighting malaria, an even more serious killer. "*Malaria experts say that 900 000 deaths for Malaria occur every year in sub-Saharan Africa. Seventy per cent of the dead are children under five years of age*." The stern conclusion is: "*The possibility that a true global* [AIDS] *disaster is just around the corner unfortunately provides an unparalleled opportunity for the modelling that jacks-up the numbers to draw attention and funding. Failure to make a simple reality check allowed the results to become accepted 'facts' […] The models were polluted by a huge sympathy bias*".

Other examples and quotes from the book of Pilkey and Pilkey-Jarvis (2007)will be seen in Section 4.

## 4 Rules for sensitivity auditing

### *4.1 Rule 1: check against rhetoric use of mathematical modelling*

As noted by Hornberger and Spear (1981):

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

This sober view of modelling was popularised by novelist Douglas Adams in one of his classic novels (1987, p.69):

> "Well, Gordon's great insight was to design a program which allowed you to specify in advance what decision you wished it to reach, and only then to give it all the facts. The program's task, […], was to construct a plausible series of logical-sounding steps to connect the premises with the conclusion."

Adam's irony is cogent. Mathematical modelling is an apt tool to transform evidence-based policy in reverse. The abundance of parameters and assumptions makes the task of mapping the facts to the desired inference trivially easy. This does not apply only to the over-parameterised models addressed by Hornberger and Spear (1981), but also to the relatively parsimonious models used in applied econometrics, as vividly illustrated by Edward E. Leamer (2010).

This use of mathematical modelling (a technique, a language) in a scantly disguised normative (or advocacy) mode can be termed rhetoric, or strategic (Boulanger et al., 2007), like the use of Latin in Manzoni's novel.

There is a vast literature sounding the alarm on instances of corruption in the use of mathematical models, with the earliest warnings coming from Saunders Mac Lane (1988a, 1988b) in an exchange of letters in the journal SCIENCE on the subject of system analysis as practiced at International Institute for Applied Systems Analysis (IIASA):

> "[…] this type of 'systems analysis' consists of the construction of massive imaginary future 'scenarios' with elaborate equations for quantitative 'models' which combine to provide predictions or projections […] which cannot be verified by checking against objective facts. Instead IIASA studies often proceed by combining in series a number of such unverified models, feeding the output of one such model as input into another equally unverified model."

The *mediatic* aspect of the issue is investigated by philosopher Jean Baudrillard (1999, p.92), according to whom modelling, when used not in '*controlled scientific conditions'* but "*in mass communication, […] assumes the force of reality, abolishing and volatilizing the latter in favour of that neo-reality of a model materialized by the medium itself"*.

Along similar lines, one of the authors of this paper (Van der Sluijs et al., 1998) observed that 'Once environmental numbers are thrown over the disciplinary fence, important caveats tend to be ignored, uncertainties compressed and numbers used at face value'.

Coming to non-specialist books on the subject, beside the work of Pilkey and Pilkey-Jarvis already cited, we recall the work of Nassim Nichola Taleb (The Black Swan, 2007), where issue is taken against modellers' attempt to *Platonify* reality, meaning by this the attempt to stick to elegant formal structures to describe facts which are too stubborn to be subdued by such simplifications. Taleb's call recalls to the mind Stephen Toulmin's (2001) plea for reasonableness as opposed to overstretched rationality.

Apparently one is never vigilant enough as in spite of the cited warnings, many have, unfortunately *a posteriori,* seen the links between the 2008 credit crunch and the mathematical models disingenuously used to price the financial products at the heart of the crisis. According to Leamer (2010), "*With the ashes of the mathematical models used to rate mortgage-backed securities still smoldering on Wall Street, now is an ideal time to revisit the sensitivity issues*", which incidentally is also the scope of the present work. Joseph Stiglitz, in a chapter aptly named "*Complexity – going beyond transparency*" notes: "*[…] Part of the agenda of computer models was to maximize the fraction of, say, a lousy sub-prime mortgage that could get an AAA rating, then an AA rating, and so forth, […]*", (2010, p.161), thereby linking '*Perverse incentives*' to '*flawed models*' (ibidem:92). Finally Jerome Ravetz (2009), in discussing the ethics of scientists, muses "*Yet we now know that the collective endeavour of these […] very nice entrepreneurial scientists* [the '*quants*', mathematicians employed in finance] *has resulted in the creation of a mountain of toxic fake securities*".

In summary, Rule 1 prescribes that the prospective sensitivity auditors maintain open eyes and '*organised scepticism*' toward technical and normative hurdles limiting the plausibility of a model-based inference.

### *4.2 Rule 2: adopt an 'assumption hunting' attitude*

Models are full of more or less implicit assumptions, which once made explicit, pose varying challenges to the belief of the beholder. These assumptions may have sedimented into the interstices of a model, or they may have been pondered in the pre-analytic phase of the work and henceforth forgotten by the same users of the model. As an example, Laes et al. (2011: …) notes that in relation to the "*[…] calculation of the external costs of a potential large-scale nuclear accident* […] *[*an analysis*] resulted in a list of 30 calculation steps and assumptions*".

Hence the need for assumption hunting or – to use Toulmin's elegant words. "[..]*the hunt for suppressed premises*" (1985, p.43). The expression 'assumptions hunting' is due to Scriven (1976), and its current usage in relation to models has been made popular mostly by Dutch investigators.

Kloprogge et al. (2011) suggest to structure the evaluation of a model-based inference into a series of steps covering analysis, revision and communication. The analysis focuses on identifying explicit and implicit assumptions in the calculation chain and the potential value-ladenness of key assumptions. The revision includes a sensitivity analysis and a possible diversification of the assumptions, while communication aims at making explicit the entire process, inclusive of elements of value-ladenness and possible alternatives. The degree of value-ladenness is estimated via the use of pedigree matrices following the NUSAP methodology (Funtowicz and Ravetz, 1990). The revision phase combines information from the pedigree analysis and the sensitivity analysis. The assumptions with a weak pedigree and a strong sensitivity on the inference are those which deserve more scrutiny, revision, and which are at the top of the communication effort. These assumptions are the ones for which the implicit plausibility assertions need to be carefully examined.

In the work of Laes et. al. (2011) already cited, the communication phase led to a substantial rejection by stakeholders of the model as a relevant tool for policy, since the model's assumptions were judged either implausible or contentious. Along similar lines a participatory approach known as 'Coproduction of knowledge model, CPM', Lane et al.,

(2011) was applied to a case of flood risk. The team was composed of experts, both certified (academic natural and social scientists) and non-certified (local people affected by flooding), and ended up discarding the off-shelf model already available for the hydro geological and hydraulic analyses and developing their own model, using an alternative framing of the problem and producing alternative solutions.

An application of these approaches to microbial contamination risk analysis is described by Boone et al. (2010), while the case of electromagnetic fields is investigated in de Jong et al. (2012) as previously discussed. Rules for mapping sources of uncertainties and hence assumptions in the field of hydrology are also discussed in Beven et al. (2010).

Just to dispel the impression that assumptions-related scruples are the preserve of a restricted circle of practitioners, one can read the Financial Times (2011) economist John Kay elaborating on the '*making up*' of the missing data needed to operate models:

> "You assume the future will be like the past, or you extrapolate a trend. Whatever you do, no cell on the spreadsheet may be left unfilled. If necessary, you put a finger in the air. This may lead to extravagant flights of fantasy. To use Britain's Department of Transport scheme for assessing projects, you have to impute values of time in 13 different activities, not just today, but in 2053. […] What will be average car occupancy rates, differentiated by time of day, in 2035?"

The future being unlike the past (and this being the source of many explicit or implicit assumptions) is an old problem. In the words of Frank Knight (1921, p.313):

> "We live in a world of contradiction and paradox, a fact of which perhaps the most fundamental illustration is this: that the existence of a problem of knowledge depends on the future being different from the past, while the possibility of the solution of the problem depends on the future being like the past."

### *4.3 Rule 3: detect GIGO*

"Garbage in garbage out (GIGO), is the strategic minimisation of uncertainty in order to inflate certainty in the inference, as defined both by econometricians (Edward Leamer, Peter Kennedy – see below) and epistemologists [Funtowicz and Ravetz, (1990), p.6]. According to the latter, GIGO-science – or pseudo-science, is "*where uncertainties in inputs must be suppressed lest outputs become indeterminate*". This implies artificially deflating the uncertainty in the assumptions to avoid that the distribution of the inference becomes so flat as to be useless. Saltelli and D'Hombres (2010) use sensitivity auditing to argue that this is the case for the cost benefit analysis proposed by various parties in relation to climate change action or inaction. A very similar standpoint – only turned in an affirmative/positive version, is from Leamer's (1990) work, where he states:

> "I have proposed a form of organised sensitivity analysis that I call 'global sensitivity analysis' in which a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified."
>
> "Conclusions are judged to be sturdy only if the neighborhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful."

This is after all not a new idea. A related trade-off is in Imre Lakatos (1976, p.57): '*when increasing certainty, you decrease content*', meaning by this that the more one tries to make a theorem 'refutation-proof', the more the theorem's range of application is reduced. Leamer's viewpoint is upheld in standard econometrics textbooks; see, e.g., Kennedy (2007) – see rule 4, Section 4 of this paper.

In a policy context, uncertainty can be amplified as well as minimised according to convenience. Oreskes and Conway (2010) describe a famous case of uncertainty amplification for the health effect of tobacco. They compare the narrative of the tobacco companies fighting to deny the health effects of smoking, to those of climate sceptics, who – according to these authors – amplified uncertainty about anthropogenic climate change. Naomi Oreskes is well known to modellers for having extensively written against the concept of model validation or verification. According to Oreskes et al. (1994), models can be evaluated, or corroborated, but never proven true. Earlier the issue of model indeterminacy had been hotly debated in hydrology where a much cited paper proclaimed in its title: "Ground-water models cannot be validated" (Konikow and Bredehoeft, 1992).

In a later work Oreskes (2000, p.36) articulated her critique by noting that

> "[M]odels area complex amalgam of theoretical and phenomenological laws (and the governing equations and algorithms that represent them), empirical input parameters, and a model conceptualization. When a model generates a prediction, of what precisely is the prediction a test? The laws? The input data? The conceptualization? Any part (or several parts) of the model might be in error, and there is no simple way to determine which one it is."

Oreskes's point is linked to the parallel often made between a logical proposition – a theory-based statement – and a model prediction. Although models share the scientific flavour of postulated laws or theories they are not laws, in that the making of a model is substantially more fraught with assumptions than crisp theories or agile laws ordinarily are. She notes "[…] *to be of value in theory testing, the predictions involved must be capable of refuting the theory that generated them.*" What happens when the 'theory' is not a law but a mathematical model? "*This is where predictions [...] become particularly sticky.*" The crux of the matter is that model-based inferences are very delicate artefacts.

Another interesting story about uncertainty manipulation is told by David Michaels (2005) – a former EPA employee, on the battles between industry and regulators over the US data quality act and the standard for exposure to beryllium. This is where industry fought hard to amplify uncertainty, according to the author, to prevent regulators from imposing more stringent standards. The same debate in the US surrounded the Proposed Risk Assessment Bulletin (OMB, 2006), which was received by some as an attempt "*to bog the* [regulatory] *process down, in the name of transparency*" (Robert Shull cited in Macilwain, 2006: …). In the same article one reads "*[...] the proposed bulletin resembles several earlier efforts, including rules on 'information quality*' *and requirements for cost–benefit analyses, that make use of the OMB's* [Office for Management and Budget] *extensive powers to weaken all forms of regulation.*"

An important consequence of Rule 3 is that one should be particularly severe against spurious accuracy, e.g., when a result is given with a number of digits exceeding (at times ludicrously) a plausible estimate of the associated uncertainty.

### *4.4 Rule 4: find sensitive assumptions before they find you*

One of the ten commandments of applied econometrics according to Peter Kennedy's popular Econometrics textbook on Applied Econometrics (2007, p.367) is: "*Thou shall* *confess in the presence of sensitivity. Corollary: Thou shall anticipate criticism*". The wisdom of this principle is evident, in that when unwanted or unexpected model sensitivity is exposed by a third party, it becomes arduous for the proponent modellers to reinstate a just-falsified inference. Thus sensitivity analysis, or better sensitivity auditing, can be used to anticipate a critique. This is the application to modelling of Robert K. Merton '*organised scepticism*'. According to Merton (1942) *Communalism, Universalism, Disinterestedness and Organised Skepticism* are the operating principles (norms) of the scientific method. Also for Pilkey and Pilkey-Jarvis (2007, p.189) '*Scientific mathematical modelling should involve constant efforts to falsify the model*'.

A typical illustration of this rule is the case of the Stern review – by Sir Nicholas Stern (2006). The purpose of the review, commissioned by the UK government, was to study the '*Economics of climate change*', and was considered by the UK prime minister of the time "*The most important report on the future ever published by this government*" (Times, 2006). As the review attracted considerable criticism, especially in relation to its chapter dealing with cost benefit analysis of taking action today to offset climate damage tomorrow (see, e.g., Nordhaus, 2007), the Stern team published a sensitivity analysis a posteriori, to offset these critiques, in a document known as "Technical Annex to postscript" of the Stern (2006) review. The story is detailed in Saltelli and D'Hombres (2010) who raise the following criticisms to both the authors of the Stern review and their opponents (Nordhaus, 2007 in this case):

1. a rigorous design-based sensitivity analysis is lacking
2. highly uncertain numbers (e.g., discount rates) are used at face value extending the scope of a cost benefit analysis to cover two centuries from the present time
3. mathematical models appear to be used expediently, for the sake of projecting pre-established normative stances.

A sensitivity analysis performed (by the Stern team) together with the cost benefit analysis would have shown the extreme volatility of the inference or, at least, would have allowed a better argument to be developed.

An implication of this rule for users of a sensitivity auditing is that a model-based inference unaccompanied by a technically sound sensitivity analysis (see rule 7) should be regarded as suspicious. Model developers and proponents of the inference should explain on what basis they have considered such an analysis as dispensable.

## *4.5 Rule 5: aim to for transparency*

The discussion of rule 3 above about the data quality act and the risk assessment bulletin has shown how the issue of transparency can be the subject of dispute. While transparency can in general be seen as an element of quality, it can at times be denounced as pretext to '*bog the process down*'. While keeping this caveat in mind, we shall mostly embrace transparency as useful in the context of mathematical modelling when this has to feed into the policy process.

According to the OMB (2002), models should be made available to a third party so that it can "use the same data, computer model or statistical methods to replicate the analytic results reported in the original study. […] The more important benefit of transparency is that the public will be able to assess how much an agency's analytic result hinges on the specific analytic choices made by the agency".

The OMB suggestion hence is that reproducibility is a necessary condition to transparency. Our suggestion is that transparency is in turn useful to defend the legitimacy and epistemic authority of the institutions making use of mathematical modelling in the context of a policy assessment.

Often the same model used within an organisation to paddle through the analysis, the workhorse pulling the cart of the daily 'what if', '*caeteris paribus*' work, might be used in an adversarial context, simply because it is expected that external stakeholders will accept the house's wisdom and its model. This may well be the case, but one would be wise not to bank on it. The problem is that in real life *caeteris* are never *paribus*.

In the words of Joseph Stiglitz (2011): "Models by their nature are like blinders. In leaving out certain things, they focus our attention on other things. They provide a frame through which we see the world".

Imagine a government agency which has made a considerable investment in developing its own assessment model. We suggest complementing this model (*the frame through which we see the world*) with something more agile and proportionate that can stand in court, possibly but not necessarily side by side with the agency's model.

Real life examples show that model use may even become counterproductive in a policy debate when this type of simplification is not operated. In the case already discussed of Lane et al. (2011), a model simplification/reformulation led to a reframing of the issue and to possibly alternative solutions. In the context of climate this point is made by Pilkey and Pilkey-Jarvis (2007, p.86), where it is argued that the climate-sceptics' work would be harder if

> "[…] the global change modeling community would firmly and publicly recognize that its efforts to truly quantify the future are an academic exercise and that existing field data on atmospheric temperatures, melting glaciers, […] and other evidence should be relied on to a much greater degree to convince politicians that we have a problem. Let the models point to a trend and answer 'what-if' questions. A serious societal debate about 'solutions' can never occur as long as modellers hold out the probability, just around the corner, of accurate projections of future climates and seal-level position".

Five years after the publication of Pilkey and Pilkey-Jarvis' book we see comforting signs that the point has been driven home; it is now admitted that the more one understands climate, the more model predictions may become uncertain (Maslin and Austin, 2012), and more and more means and standard deviations (e.g., of temperature) populate the discourse on climate (Hansen et al., 2012). Still policy makers associate a 50% certainty to limit temperature increase to 2 degree centigrade, a climate policy target, with a greenhouse gas concentration at 450 ppm CO2-equivalent (Meinshausen, 2005). Given that these three numbers (0.5, 450, 2) are model-generated, some more circumspection would befit the prospective sensitivity auditor.

Finally one has to admit that at present there is in general little scientific incentive to reproduce a model, and more generally, to replicate scientific papers. Reproducibility has been the subject of articles in the scientific press, and even editorials in Biostatistics (Peng, 2009) and in Nature Biotechnology (2012): nobody tries to reproduce the results because, among other reasons, it takes resources that are never justified. The only arena in which it is done is in an already contested and controversial issue and where the stakes are no longer in the context of 'normal' science but in the context of use.

### 4.6 Rule 6: do the right sums

In 'Return to Reason' (2001) Stephen Toulmin vividly recalls the dangers of precision: '*Doing the sum right*' is a far lesser challenge than '*Doing the right sums*'.

In modelling as in life, framing error, or Type III errors, are the most dangerous. When performing an uncertainty and sensitivity analysis, one falls easily into that which Taleb (2007) calls '*The delusion of uncertainty*' and which is known in Dutch as '*Lampposting*', whereby '*The uncertainties which are more carefully scrutinised are usually those which are the least relevant*" (Van der Sluijs, seen on www.nusap.net).

Lamp posting refers to the joke of the drunkard looking for his lost keys not in his house's garden, where he lost them, but in the street under a lamp as '*there is more light here*'.

A Type III error is illustrated in the work of Pilkey and Pilkey-Jarvis cited earlier. The examples concern the Yucca Mountain repository for radioactive waste. A model named Total System Performance Assessment (TSPA) has been used for the safety analysis computations. TSPA is composed of 286 sub-models. A key assumption in TSPA is the range of permitted values for the permeability of the geological formation. A low permeability is crucial to ensure that water will take a long time to percolate from surface to disposal. For the Yucca Mountain test disposal site, a range of 0.02 to 1 millimetre per year was used for the percolation rate. The confidence of the stakeholders in TSPA was not helped when evidence was produced which led to an upward revision of four orders of magnitude of this parameter (of the order of three metres per year). The evidence in question was the presence at the repository level of an isotope of chlorine 36Cl associated to atomic bomb detonations in the atmosphere. According to the authors the error was due to the modelling of the granite formation as a homogeneous medium, while a fissures and faults model of the same formation would have been more realistic. Of course these types of errors can be corrected by more work and comparison with additional evidence, but they were not corrected at the time of the assessment.

Sensitivity analysis is not immune to Type III errors, and neither is sensitivity auditing [Pilkey and Pilkey-Jarvis, (2007), p.25]:

> "It is important […] to recognize that the sensitivity of the parameter in the equation is what is being determined, not the sensitivity of the parameter in nature. […] If the model is wrong or if it is a poor representation of reality, determining the sensitivity of an individual parameter in the model is a meaningless pursuit."

Type III errors which are most common are likely those associated to neglecting part of the views and expectations surrounding the issue at stake.

### 4.7 Rule 7: focus the analysis

Sensitivity is often omitted in modelling studies, or it is executed in a perfunctory fashion. According to Leamer (2010) "*One reason these methods* [global sensitivity analysis] *are rarely used is their honesty seems destructive*". Most sensitivity analyses published in the highest ranked journals such *SCIENCE* and *Nature* are perfunctory. This may sound a surprising claim. Still the analyses of sensitivity reviewed in Saltelli and Annoni (2010) were run without a statistical design, moving just one input factor at a time. Neglecting the existence if interactions among factors, this approach bumps against the curse of dimensionality, as at already moderate dimensionality moving one factor at a time explores only a tiny fraction of the space of the input.

In general a sensitivity analysis performed without a statistical design and without an estimate of the error is poor. In the context of modelling studies used in support to policy, sensitivity analysis should also be parsimonious and possible cogent, i.e., it should focus on a single target variable, this being the relevant inference that the modelling study is trying to underpin. The analysis should be one and not many, covering the entire evidential chain, as opposed to covering one sub-model at a time. Again this is needed to ensure that all interactions among factors in different compartments are being captured. Following this view, rather than diluting the sensitivity analysis showing its results for different scenarios, the scenario should be one of the variables investigated in the frame of the analysis.

### *4.7.1 Who should apply the rules?*

As argued earlier, the process by which sensitivity auditing is carried out should be consistent with the ideas of post-normal science embraced here. A participatory process where *relevant* members of the extended community of peers are first identified and subsequently involved in this process is hence fundamental to the successful implementation of the present prescriptions. The identification of this community can be done in many ways, the most obvious being through institutional analysis and stakeholder analysis. The spaces where such scrutiny occurs can be a myriad according to the communities involved (Guimarães Pereira et al., 2009). One may have focus groups, juries, consensus workshops, or in-depth interviews with relevant individuals. It is obvious that the less specialised the community involved is, the better the unfolding of the modelling ‘black boxes’ has to be prepared. We argue that “progressive disclosure of information” (Guimarães Pereira et al., 2002, 2003) is a key principle of design of communication in participatory settings due to the support it provides for mixed *expertise*. Information is supplied in layers of increasing specialisation, depending on actors’ interests and necessary specialised information. In other words, in the course of a process of systematic analysis of the model-based inference such as the one suggested in this paper, the modelling ‘black box’ needs to be progressively and intelligibly open to those who participate in the exercise. The community involved the space and the fairness with which the object of scrutiny is looked at gives legitimacy to the whole exercise, and helps with examining the plausibility of audited models.

## 5 Conclusions

Throughout this paper we have discussed mathematical modelling in general, without distinguishing between data-driven and principle-driven models, between micro and macro, or between the natural sciences and social sciences styles of modelling. It is clear that the arguments developed in this paper are general to forms of evidence that demand statistical, mathematical or otherwise disciplinary elaboration. We would hence prescribe similar recipes when the model is in fact a statistical indicator, whose construction customarily demands several modelling steps (Boulanger et al., 2007; Paruolo et al., 2012).

Since no recipe is universal, many words of caution should be added to our list of prescriptions, however sensible these may appear to a benevolent reader. In a value-laden context the correctness of the prescriptions may well be a casualty in the power game, as the first casualty in a battle is the battle plan.

Indeed the quality of the rules can only be judged in relation to their fitness for an assigned purpose in a specified case, and there is no guarantee that major blunders will be avoided by a diligent application. The rules are a minimum, due-diligence requirement for the use of model-based inference in a policy discourse, and there seems to be little justification, given the stakes involved in any policy, in omitting these simple well-meaning steps. At the very least, sensitivity auditing will ensure that the recipients of the analysis are fully aware of the conditionality of the predictions, and a notch more sceptical of model-based evidence when this is presented on the basis of an authority principle.

It is possible that the practitioners' community is becoming more sympathetic with the 'uncertainty exploration' concerns raised in the present work, even in hotly debated areas such as climate. "*Many of those of us who spend our working hours, and other hours, thinking about uncertainty, strongly believe the climate modelling community must not put resolution and processes (to improve the simulator) above generating multiple predictions (to improve our estimates of how wrong the simulator is)*". Our optimism remains nevertheless tempered by the fact that the quote just offered originates from a blog [allmodelsarewrong.com, (Edwards, 2012)] whose title puts it squarely in the field of sympathisers to sensitivity auditing.

As we can see from other contributions to this special issue the concept of plausibility encapsulates ideas of respected opinion, and *applauded* (credible and relevant and legitimate) assertions.

Our suggestion is that sensitivity auditing is to a model's plausibility what sensitivity analysis is to a model's technical relevance. In other words a well-run sensitivity analysis makes a model-based inference worth considering. Whether the model can then 'stand in court', survives and possibly be applauded needs the stronger medicine of sensitivity auditing.

## Acknowledgements

Helpful suggestions and corrections were offered by Bruna De Marchi, University of Bergen (NO), Centre for the Study of the Sciences and the Humanities (SVT) and by Sjoerd Hardeman, European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen. The remaining errors are due to the authors.

## References

Adams, D. (1987) *Dirk Gently's Holistic Detective Agency*, Pocket Books, New York.

Baudrillard, J. (1999) *Revenge of the Crystal: Selected Writings on the Modern Object and Its Destiny*, pp.1968–1983, Pluto Classics, London.

Beven, K., Leedal, D. and Alcock, R. (2010) 'Uncertainty and good practice in hydrological prediction', *VATTEN*, Vol. 66, Nos. 3–4, pp.159–163.

Boone, I., Van der Stede, Y., Dewulf, J., Messens, W., Aerts, M., Daube, G. and Mintiens, K. (2010) 'NUSAP: a method to evaluate the quality of assumptions in quantitative microbial risk assessment', *Journal of Risk Research*, Vol. 13, No. 3, pp.337–352.

Boulanger, P-M. (2007) 'Political uses of social indicators: overview and application to sustainable development indicators', *International Journal of Sustainable Development*, Vol. 10, Nos. 1–2, pp.14–32.

Chin, J. and Lwanga, S.K. (1991) 'Estimation and projection of adult AIDS cases: a simple epidemiological model', *Bulletin of the World Health Organization*, Vol. 69, No. 4, pp.399–406.

de Jong, A., Wardekker, J.A. and Van der Sluijs, J.P. (2012) 'Assumptions in quantitative analyses of health risks of overhead power lines', *Environmental Science & Policy*, Vol. 16, No. 1, pp.114–121.

Dunn, W.N. (1997) *Cognitive Impairment and Social Problem Solving: Some Tests for Type III Errors in Policy Analysis*, Graduate School of Public and International Affairs, University of Pittsburgh, Pittsburgh.

EC (2009) *Impact Assessment Guidelines*, 15 January, p.25, Technical Report 92, SEC [online] http://ec.europa.eu/governance/impact/commission_guidelines/docs/iag_2009_en.pdf (accessed December 2012).

Edwards, T. (2012) *Limitless Possibilities*, 13 June [online] http://allmodelsarewrong.com/limitless-possibilities/ (accessed 17 August 2013).

EPA (2009) *Guidance on the Development, Evaluation, and Application of Environmental Models*, March, p.26, Technical Report EPA/100/K-09/003, Office of the Science Advisor, Council for Regulatory Environmental Modeling [online] http://web.archive.org/web/20130817202323/http://allmodelsarewrong.com/limitless-possibilities/ (accessed December 2012).

Feyerabend, P. (1975, 2010) *Against Method*, Verso publisher, London.

Funtowicz, S. and Ravetz, J. (1990) *Uncertainty and Quality in Science for Policy*, Kluwer Academic Publishers, The Netherlands.

Funtowicz, S.O. and Ravetz, J.R. (1993) 'Science for the post-normal age', *Futures*, Vol. 25, No. 7, pp.739–755.

Guimarães Pereira, A. and Corral Quintana, S. (2002) 'From technocratic to participatory decision support systems: responding to the new governance initiatives', *Journal of Geographic Information and Decision Analysis*, Vol. 6, No. 2, pp.95–107.

Guimarães Pereira, A., Rinaudo, J-D., Jeffrey, P., Blasques, J., Corral Quintana, S., Courtois, N., Funtowicz, S. and Petit, V. (2003) 'ICT tools to support public participation in water resources governance & planning: experiences from the design and testing of a multi-media platform', *Journal of Environmental Assessment Policy and Management*, Vol. 05, No. 3, pp.495–420.

Guimarães Pereira, A., Raes, F., Pedrosa, T.D.S., Rosa, P., Brodersen, S., Jørgensen, M.S., Ferreira, F., Querol, X. and Rea, J. (2009) 'Atmospheric composition change research: time to go post-normal?', *Atmospheric Environment*, Vol. 43, No. 33, pp.5423–5432.

Hansen, J., Sato, M. and Ruedy, R. (2012) 'Perception of climate change', *Proc. Natl. Acad. Sci.*, doi:10.1073/pnas.1205276109, Early draft posted as *Public Perception of Climate Change And The New Climate Dice*, arXiv.org:1204.1286 [online] http://www.giss.nasa.gov/research/briefs/hansen_17/ (accessed 17 August 2013).

Hornberger, G.M. and Spear, R.C. (1981) 'An approach to the preliminary analysis of environmental systems', *Journal of Environmental Management*, Vol. 12, No. 1, pp.7–18.

Kay, J. (2011) 'A wise man knows one thing – the limits of his knowledge', *Financial Times*, 29 November.

Kennedy, P. (2007) A *Guide to Econometrics*, 5th ed., p.396, Blackwell Publishing, Oxford.

Kloprogge, P., Van der Sluijs, J.P. and Petersen, A.C. (2011) 'A method for the analysis of assumptions in model-based environmental assessments', *Environmental Modelling & Software*, Vol. 26, No. 3, pp.289–301.

Knight, F.H. (1921) *Risk, Uncertainty, and Profit*, Hart, Schaffner & Marx; Houghton Mifflin Co., Boston, MA.

Konikow, L.F. and Bredehoeft, J.D. (1992) 'Ground-water models cannot be validated', *Advances in Water Resources*, Vol. 15, No. 1, pp.75–83.

Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*, Chicago University Press, Chicago.

Laes, E., Meskens, G. and Van der Sluijs, J.P. (2011) 'On the contribution of external cost calculations to energy system governance: the case of a potential large-scale nuclear accident', *Energy Policy*, Vol. 39, No. 9, pp.5664–5673.

Lakatos, I. (1976) *Proofs and Refutations, The Logic of Mathematical Discovery*, 2005 ed., Cambridge University Press, Cambridge.

Lane, S.N., Odoni, N., Landström, Whatmore, S.J., Ward, N. and Bradley, S. (2011) 'Doing flood risk science differently: an experiment in radical scientific method', *Transactions of the Institute of British Geographers*, Vol. 36, No. 1, pp.15–36.

Leamer, E. (1990) 'Let's take the con out of econometrics, and sensitivity analysis would help', In: *Modelling Economic Series*, Clarendon Press, Oxford, also in Leamer, E. (1990) 'Let's take the con out of econometrics', *American Economics Review*, March 1983, Vol. 73, pp.31–43.

Leamer, E.E. (2010) 'Tantalus on the road to Asymptopia', *Journal of Economic Perspectives*, Vol. 24, No. 2, pp.31–46.

Mac Lane, S. (1988a) 'IIASA's credibility', *Science, Letters*, Vol. 241, No. 4870, p.1144.

Mac Lane, S. (1988b) 'IIASA and modeling', *Science, Letters*, Vol. 242, No. 4886, pp.1623–1624.

Macilwain, C. (2006) 'Safe and sound?', *Nature*, Vol. 442, No. 7100, pp.242–243.

Manzoni, A. (1827) *The Betrothed*, English translation from Project Gutenberg [online] http://www.gutenberg.org/ebooks/35155 (accessed December 2012).

Maslin M. and Austin, P. (2012) 'Climate models at their limit?', *Nature*, Vol. 486, No. 7402, pp.183–184.

Meinshausen, M. (2005) 'On the risk of overshooting 2°C', paper presented at *Scientific Symposium 'Avoiding Dangerous Climate Change'*, 1–3 February, MetOffice, Exeter [online] http://www.pik-potsdam.de/~mmalte/simcap/publications/meinshausenm_risk_of_overshooting_final_webversion.pdf (accessed December 2012).

Merton, R.K. (1942) 'The normative structure of science', in Merton, R.K. (1973) (Eds.): *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press, Chicago.

Michaels, D. (2005) 'Doubt is their product', *Scientific American*, Vol. 292, No. 6, pp.96–101.

Nature Biotechnology (2012) 'Further confirmation needed', *Editorial*, Vol. 30, No. 9, p.806.

Nordhaus, W. (2007) 'Critical assumptions in the Stern review on climate change', *Science*, Vol. 317, No. 5835, pp.201–202.

Oakley, J.E. and O'Hagan, A. (2004) 'Probabilistic sensitivity analysis of complex models: a Bayesian approach', *Journal of the Royal Statistical Society B*, Vol. 66, No. 3, pp.751–769.

OMB: Office of Management and Budget (2002) *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, Friday, February 22, Notice, Republication [online] http://www.whitehouse.gov/omb/fedreg_final_information_quality_guidelines/ (accessed December 2012).

OMB: Office of Management and Budget's (OMB) Office of Information and Regulatory Affairs (OIRA) (2006) *Proposed Risk Assessment Bulletin*, 9 January [online] http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/proposed_risk_assessment_bulletin_010906.pdf (accessed December 2012).

Oreskes, N. (2000) 'Why predict? Historical perspectives on prediction in Earth Science', in Sarewitz et al. (Eds.): *Prediction, Science, Decision Making and the Future of Nature*, Island Press, Washington DC.

Oreskes, N. and Conway, E.M. (2010) *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, Bloomsbury Press, New York.

Oreskes, N., Shrader-Frechette, K. and Belitz, K. (1994) 'Verification, validation, and confirmation of numerical models in the Earth Sciences', *Science*, Vol. 263, No. 5147, pp.641–646.

Paruolo, P., Saisana, M. and Saltelli, A. (2012) 'Ratings and rankings: voodoo or science?', *Journal Royal Statistical Society A*, Vol. 176, No. 2, pp.1–26.

Peng, R.D. (2009) 'Reproducible research and biostatistics', *Biostatistics*, July, Vol. 10, No. 3, pp.405–408.

Petersen, A.C., Cath, A., Hage, M., Kunseler, E. and Van derSluijs, J.P. (2011) 'Post-normal science in practice at the Netherlands environmental assessment agency', *Science Technology & Human Values*, Vol. 36, No. 3, pp.362–388.

Pilkey, O.H. and Pilkey-Jarvis, L. (2007) *Useless Arithmetic. Why Environmental Scientists Can't Predict the Future*, Columbia University Press, New York.

Pruppers, M.J.M. (2003) *Blootstellingaanextreemlaagfrequentelektromagnetischevelden van hoogspanningslijnen – Herberekeningnaaraanleiding van het KEMA/RIVM-onderzoeknaar de kosten en baten van maatregelenterbeperking van magnetischeveldenbijhoogspanningslijnen*, RIVM Briefrapport 032/2003.Bilthoven.RIVM.

Raifa, H. (1968) *Decision Analysis*, Addison-Wesley, Reading, MA.

Ravetz, J. (2009) 'Morals and manners in modern science', *Nature*, 5 February, Vol. 457.

Saisana, M., Saltelli, A. and Tarantola, S. (2005) 'Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators', *Journal of the Royal Statistical Society – A*, Vol. 168, No. 2, pp.307–323.

Salmon, F. (2009) 'Recipe for disaster, the formula that killed Wall Street', *Wired*, February [online] http://www.wired.com/techbiz/it/magazine/17-03/wp_quant (accessed 17 August 2013).

Saltelli, A. (2002) 'Sensitivity analysis for importance assessment', *Risk Analysis*, Vol. 22, No. 3, pp.579–590.

Saltelli, A. and Annoni, P. (2010) 'How to avoid a perfunctory sensitivity analysis', *Environmental Modeling and Software*, Vol. 25, No. 12, pp.1508–1517.

Saltelli, A. and D'Hombres, (2010) 'Sensitivity analysis didn't help. A practitioner's critique of the Stern review', *Global Environmental Change*, Vol. 20, No. 2, pp.298–302.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S. (2010) 'Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index', *Computer Physics Communications*, Vol. 181, No. 2, pp.259–270.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008) *Global Sensitivity Analysis*, The Primer, John Wiley and Sons, Chichester.

Saltelli, A., Ratto, M., Tarantola, S. and Campolongo, F. (2012) 'Sensitivity analysis for chemical models', *Chemical Reviews*, Vol. 112, No. 5, pp.PR1–PR21, Perennial Review of the 2005 paper, DOI: 10.1021/cr200301u, Publication Date (Web): April 19, 2012.

Saltelli, A., Tarantola, S. and Campolongo, F. (2000) 'Sensitivity analysis as an ingredient of modeling', *Statistical Science*, Vol. 15, No. 4, pp.377–395.

Santner, T., Williams, B. and Notz, W. (2003) *Design and Analysis of Computer Experiments*, Springer-Verlag, New York.

Savage, S.L. (2009) *The Flaw of Averages*, Wiley, Hoboken.

Scriven, M. (1976) *Reasoning*, McGraw Hill, Columbus.

Stainforth, D.A. and Smith, L.A. (2012) 'Policy: clarify the limits of climate models', *NATURE*, 13th September, No. 7415.

Stern, N. (2006) *Stern Review on the Economics of Climate Change*, UK Government Economic Service, London [online] http://www.sternreview.org.uk, Technical Annex to the Postscript [online] http://webarchive.nationalarchives.gov.uk/+/http://www.hm-treasury.gov.uk/sternreview_index.htm (accessed December 2012).

Stiglitz, J. (2010) *Freefall, Free Markets and the Sinking of the Global Economy*, Penguin, London.

Stiglitz, J. (2011) 'Rethinking macroeconomics: what failed, and how to repair it', *Journal of the European Economic Association*, Vol. 9, No. 4, pp.591–645.

Taleb, N.N. (2007) *The Black Swan: The Impact of the Highly Improbable*, Random House, New York.

Times (2006) 'Global warming report calls for immediate action', *Times*, 30 October.

Toulmin, S. (1985) *The Return to Cosmology: Post Modern Science and the Theology of Nature*, University of California Press, Berkeley.

Toulmin, S. (2001) *Return to Reason*, 2003 ed., Harvard University Press.

Van der Plas, M., Houthuijs, D.J.M., Dusseldorp, A., Pennders, R.M.J. and Pruppers, M.J.M. (2001) *Magnetische velden van hoogspanningslijnen en leukemie bij kinderen*, RIVM rapport nr. 610050 007, Bilthoven, RIVM.

Van der Sluijs, J.P. (2005) 'Uncertainty as a monster in the science policy interface: four coping strategies', *Water Science and Technology*, Vol. 52, No. 6, pp.87–92.

Van der Sluijs, J.P. (2010) 'Uncertainty and complexity: the need for new ways of interfacing climate science and climate policy', in Driessen, P., Leroy, P. and Van Vierssen, W. (Eds.): *From Climate Change to Social Change: Perspectives on Science – Policy Interactions*, pp.31–49, International Books, Utrecht.

Van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J. (2005) 'Combining quantitative and qualitative measures of uncertainty in model based environmental assessment: the NUSAP system', *Risk Analysis*, Vol. 25, No. 2, pp.481–492.

Van der Sluijs, J.P., Petersen, A.C., Janssen, P.H.M., Risbey, J.S. and Ravetz, J.R. (2008) 'Exploring the quality of evidence for complex and contested policy decisions', *Environmental Research Letters*, Vol. 3, No. 2, 024008 (9pp).

Van der Sluijs, J.P., Van Eijndhoven, J., Shackley, S. and Wynne, B. (1998) 'Anchoring devices in science for policy: the case of consensus around climate sensitivity', *Social Studies of Science*, Vol. 28, No. 2, pp.291–323.