Draft, January 15, 2019, Do not circulate.

Wrong, useful or wasteful? Mathematical modelling's problems, opportunities and solutions

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)

1 QUANTIFICATION BLUES?

The crisis in science's quality control apparatus [1] has been particularly felt in the field of statistics [2]–[7], as well as in the use of algorithms as decision support tools [8], in indicators of sustainability [9], [10], university rankings [11], ease of doing business [12] and many others. What links all these apparently heterogeneous practices? All – from statistics as a discipline downward to the practices and the products - aim to produce quantified knowledge. As noted in the nascent field of sociology of quantification [13], these different instances of quantification blur into one another under the joint pressures of big data and artificial intelligence, so that the question of "what qualities are specific to rankings, or indicators, or models, or algorithms" [13] becomes a legitimate one. Mathematical models are also mentioned, and yet this field – possibly one among the most mighty engines of quantification - has been left unscathed by the quantification blues, with very few voices lamenting the lack of disciplinary standards for quality control [14] or the difficulty for users to understand models due to the growing asymmetry of information among model users and model developers [15].

An important issue in mathematical modelling is in the management of uncertainty. Uncertainty quantification is at the heart of the scientific method, and *a fortiori* in the use of science for policy [16]. In statistics the p-values can be misused as to overestimate the probability of having found a true effect [3]. Likewise in modelling studies certainty may be overestimated, thus producing crisp numbers to three decimal places even in situations of pervasive uncertainty or ignorance [17], including in the use of science for policy [18]. An old refrain in mathematical modelling – first noted among hydro-geologists, is that since models are often over-parametrized, they can be made to conclude everything [19]. We are now told that in the field of clinical medical research the percentage of non-reproducible studies could be as high as 85% [20]. What about those studies that rely on mathematical modelling?

By combining analysis and case studies, we argue that mathematical modelling would deserve urgent attention, and that its problems are considerably worse than those already noted in statistics. An extended treatment of the topic with additional real life examples and discussion is available online [21].

2 MALPRACTICES OR RITUALS?

In a previous article, we criticised the ritual use of statistical methods by those with scant understanding of the assumptions or relevance of their calculations [7]. This same problem of ritual behaviour befalls mathematical modelling, and is best explained by an anecdote from Kenneth Arrow. During the Second World War, Arrow 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 [22]:

> *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 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 experts' whim [23]. Along these lines it has been suggested that models universally known to be wrong continue to play a role in economic policy decisions [24], while the book "Useless Arithmetic" [18] argues that the quantitative mathematical models policy makers and government administrators use to form environmental policies are often seriously flawed, providing a host of examples, from AIDS prevention to the evaluation of stock of fisheries, from mill tailing to costal erosion.

3 A MILLION YEAR OF CERTAINTY

One case treated in Useless Arithmetic is that of Total System Performance Assessment (TSPA), a mathematical computer program made of 13 models in turn comprised of 286 individual modules – running to hundreds of thousands of coding lines, used for the safety of nuclear waste disposal. The story of TSPA is linked to the tormented fate of the nuclear waste repository at Yucca Mountain.

In 2004 a federal court ruled that the model based safety case for the approval of the disposal site was to be guaranteed for up to one million year, thus extending the already ambitious 10,000 previously established by the Environmental Protection Agency [18].

To complicate the matter, a key number used in the assessment, the percolation rate of water to the unsaturated repository level - was underestimated by 4 orders of magnitude [25]. Measurements made public in 1996, and later confirmed [26], revealed the presence of Chlorine-36, a bomb-pulse isotope associated with the nuclear explosion at the Bikini Atoll in the South Pacific in 1963. Such a presence implies a travel speed of 3,000 millimetres per year, rather than the range 0.01 to 1 millimetre per year used in the TSPA. Regrettably, the wrong range was maintained in use in spite of the available evidence for more than a decade [25].

As often the case with modelling study of high political relevance, apportioning the blame for the limits of TSPA between the modelling and the policy communities is beyond the point, as those limits are precisely a result of the technical and institutional context in which the model is developed [27].

Tracking the fate of radionuclides from a geological formation up to man via an array of different pathways, for tens of thousands of years or more into the future is a case of mathematical hubris.

It ignores that increasing the complexity of a model one decreases modelling bias but increases measurement error ([28], Figure 1). This trade-off is known under different names in different disciplines, e.g. as the Zadeh's principle of incompatibility in system analysis, whereby as complexity increases "precision and significance (or relevance) become almost mutually exclusive characteristics" [29].

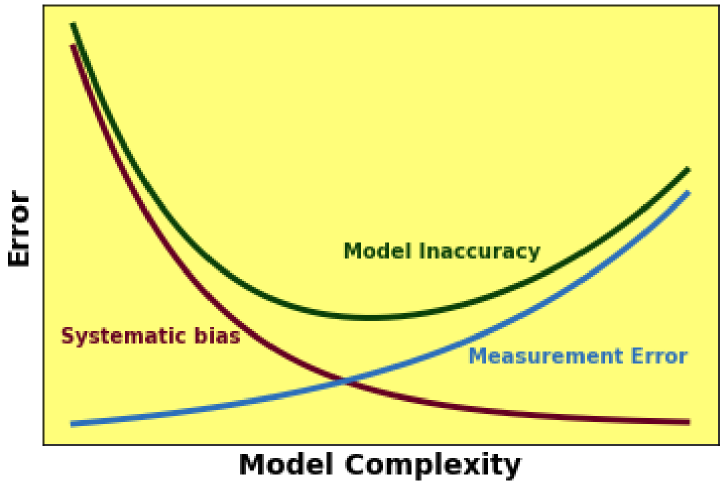

Figure 1 Model error versus model complexity. Adapted from [28].

Additionally, this kind of model-based analyses, reliant on the physics, the chemistry, the geology the radio-isotopic properties of the waste and the surrounding biosphere, gives prominence to an agenda of defined uncertainties, which neglects the surrounding institutional and social settings and power relations. While discussing the difference between uncertainty, ignorance and indeterminacy Wynne [30] exemplify a situation of indeterminacy as follows:

> *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 [31] it easy to grasps the meaning of Wynne's intuition.

In summary the fault with this modelling enterprise is not the lack of technical skills – on the contrary this modelling community spearheaded best practices e.g. in uncertainty and sensitivity analysis techniques already in the eighties [32]. The problem of this still ongoing activity [33] is in the excessive ambition of the modellers to describe a complex natural system, which allows the model institutional users to present this as an issue which can be handled within the familiar categories of risk analysis.

4 CLIMATIC BILLS

One of the winners of the 2018 Nobel prize in economics is Willem Nordhaus, known for his application of model-based economic thought to climate studies. The Noble prize awarded to Nordhaus represents a powerful endorsement of the modelling practices of this particular community. By modelling the impact on the economy of different courses of action or inaction, these modelling studies aim to devise strategies to cope with climate change. Yet computing GDP gains and losses one century ahead is a practice likely to attract some scepticism.

The same Nordhaus noted that while modelling the future sea-level and temperature appear, with all their difficulties, a meaningful pursuit, translating these climate projections into effects on GDP – the so called 'costing' of climate change – is a 'terra incognita' [34]. In this 'unknown land', modellers venture to investigate - using large mathematical models - the increase in crime rate at county level as a result of climate change one hundred years from now [17], [35], [36]. Modellers from the same community argue for more efforts and resources in this venture [37]. 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 [38], 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, the uncertainty of the prediction is so wide as to be useless for reaching a policy conclusion. The range of losses inflicted to GDP growth by climatic change – as predicted by these models - can be shown to range from values so low as to suggest a 'wait and see' attitude to an 'act immediately' one. As noted in [39] both attitudes, when entrusted to this kind of models – are delusional [17].

5 ECONOMIC LATIN

So far, our critique has targeted the use of mathematical models tens of thousands of year into the future for the case of nuclear waste disposal, and decades into the future for the case of climate costing. Is perhaps modelling on a safer footing over shorter horizons? In fact, models' accuracy and relevance does not depend on their time horizon, but on the characteristics and context of the system. The models behind the Rosetta mission managed to place a human artefact on a comet flying past the Sun after a twelve-years voyage and manoeuvring in space, which constitutes an evident success. Incidentally, our discussion should not be read as an indictment of mathematical modelling. As noted in system ecology [40], modelling is the very essence of science and thought. When STS scholars talk about 'invisible science' [41], meaning by this the science and technology products which make our life more comfortable, one chapter should be devoted to the 'invisible models' which underpin all these technologies and functions, from control systems to decision support tools, in spite of the algorithmic dystopias discussed in the extended version of the present work [21].

Yet, complex, reflexive systems may resist attempts at prediction and control even over short time horizon, while the context, the interests and the values of the involved parties are never far away. Ignoring these resistances may render a model irrelevant, or transform it into a fig leave [18].

A much discussed case is that of formulae used in finance [42], where models used in the pricing of opaque financial products are held as partly responsible for the onset of the last recession. To simplify a rather technical matter, the so called 'quants' – a special class of econometricians – chose to model the price evolution of bundles of mortgages by calibrating these based on data for periods when the real estate market was going up. Needless to say, these calibrations did not include – many say did not wish to include - what would happen when the market – as it happened – elected to go down. In this, as in all cases discussed here, the lack of a proper uncertainty quantification – or the artificial compression of the uncertainty itself, was a problem.

Moving to the field of economics proper, the over-reliance of economists on dynamic stochastic general equilibrium (DSGE) models used for policy simulation has been debated [43]. This theme hit the media when both the US senate and Queen Elizabeth asked their economists for clarification about these models' apparent incapacity to anticipate the crisis [44]. Incidentally, DSGE are among the solution proposed to 'cost' climate [37]. Also in this case, the problem is not purely technical: a DSGE can effectively simulate the effects of shocks in prices or wages by propagating them through the overall economy; it is hence a valid simulation tool, be it that it rests on very strong assumptions on the behaviour of markets and of actors in the market. A DSGE works under the 'All the rest being equal' hypothesis, which real life rarely meets, and which is why the use of a DSGE as policy prediction or policy justification tool may lead to disappointment.

Economist Paul Romer – the second winner of the 2018 Nobel prize for Economics, has recently coined the neologism 'mathiness' [45], taking issue against 'freshwater economists' (an allusion to the Chicago school in the great lakes region) for their use of mathematical modelling as Latin, in the sense that mathematics can be used to distract from underlying ideological stances. Romer draws unflattering parallels between macroeconomics and string theory in physics for their invocation to "imaginary causal forces" and excessive deference to authority of the profession's leaders [46], and invites [47] his fellow economists to appreciate the importance of intellectual honesty, pointing to a famous lecture of physicist Richard Feynman as an example [48]. The appeal to physics as a virtuous discipline is frequent in present-day discussions of science crisis [49].

Empirical economics research – a field closer to statistical than to mathematical modelling, appear likewise mired in a crisis of quality control [50], which confirms the general nature of a crisis in quantification alluded to in this work.

6 A MODEST BEGINNING

A step in the right direction for mathematical modelling would be to have statistics inject some structure and standards. This is not a radically new idea. A group of statisticians active in the UK has tackled uncertainty in simulation models by addressing uncertainty quantification, uncertainty propagation, uncertainty analysis, sensitivity analysis, calibration, ensemble analysis and related topics (see http://www.mucm.ac.uk/). Dialogues between the communities of applied mathematics and statistics on uncertainty quantification in modelling have taken place, e.g. by comparing methods used in the two communities for similar settings [51]. Sensitivity analysis offers a convincing example of a tool for model quality which can be seen as statistical in both its language and its tools, and which finds use in a large class of applications, such as model selection, calibration, optimisation, and so on [52]–[57]. Sensitivity analysis answers the question 'Which uncertain input factors are responsible for the uncertainty in the prediction?' An alternative statement of this is "Are the results from a particular model more sensitive to changes in the model and the methods used to estimate its parameters, or to changes in the data? [58]"

The use of sensitivity analysis in regulatory settings – e.g. in impact assessment studies – is prescribed in guidelines both in Europe and in the United States (European Commission[59], 2015, p. 390-393; Office for the Management and Budget[60], 2006, p. 17-18; Environmental Protection Agency[61], 2009, p.26).

Regrettably, most modelling work does not include a sensitivity analysis and the majority of sensitivity analyses published in the literature are flawed to the point of irrelevance, though much variation exists among disciplines [62], [63].

Techniques for proper uncertainty and sensitivity analysis of model-based inference may tend to be eschewed because, in their candour, they may show that the inference itself is too uncertain to be of any practical use. For example, an estimate of an investment's pay-off that gives a range from a large loss to a large gain is not what the client may wish to hear. Therefore, it has been observed by practitioners of different disciplines [16], [64], [65] that, in these sorts of 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 [66], and a sound sensitivity analysis can help to make up one's mind about the merits of a case.

As uncertainty is omnipresent in modelling, putting uncertainty at the core of model quality has a potentially unifying potential. Present computers allow for most models to be executed repeatedly - even and especially at the model development stage. Hence, whenever feasible, the modelling work could take place within an ideal Monte Carlo driver for uncertainty and sensitivity analysis, whereby all sources of uncertainty - framing uncertainties, parametric uncertainties, and so on, can be activated simultaneously, allowing 'on the flight' inference as to what contribution each set of uncertainties makes 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 at the stage of model construction. An extension of this approach to statistics proper would be a modelling of the modelling process itself. This could address – with some imagination and creativity - what has been called ‘the statistical garden of the forking paths’ [67], alluding to the myriad of ways which analysts can take when searching for patterns in data.

7 From sensitivity analysis to sensitivity auditing

It is a common refrain in economics that the most significant theories - those most precious for modelling - are based on the most unrealistic of assumptions [68], [69]. This leaves open the questions of deciding what counts as a significant theory, e.g. significant to whom? An alternative view is 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 [70]. Even here there must be criteria, both technical and ethical, to decide what should be left out, who decides, and if the resulting model is still relevant and plausible to the eyes of those affected by the modelling exercise. The discussion of the Yucca Mountain case has shown that model’s purpose and context cannot be ignored in assessing model quality.

We have often heard modellers claim the ‘neutrality’ of their computer codes. “Computers are impervious to the lure of power”, writes a statisticians using computers to fight the practice of gerrymandering [71]. But there is confusion here between the worth of one’s cause and the neutrality of one’s tool. Our position is that the technique is never neutral, and hence the use of evidence – including quantitative evidence – must be conceived for an adversarial setting – where different interests will frame the issue with different disciplinary lenses, 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, which is the topic of sensitivity auditing [66], [72], also recommended by the European Commission’s guidelines for impact assessment [59]. The ambition of sensitivity auditing is to address the entire modelling process, inclusive of motivation, power relations, hidden assumptions and normative frameworks, and addressing issues of trust and legitimacy. It is based on a seven-point checklist:

> <u>Rule 1</u>: ‘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;
>
> <u>Rule 2</u>: ‘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?
>
> <u>Rule 3</u>: ‘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 within an extended peer-community [15], [73]. 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 put to rest [74]. Examples of these approach are to education, [75], nutrition [76], and indicators [9].

In relation to the call for transparency one can note that now several journals demand to 'see the data' before accepting a paper [77], 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 present in the US [60] and the EU [59]. But '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 [15], [78].

8 WHY NOT?

What obstacles could impede the proposed reformation of modelling practices? For one, modelling is too vast an enterprise to be boxed into a single quality assurance framework, though material exists to give this process a good start [79], [80].

Next comes the deep-rooted resistance among practitioners to the idea of non-neutrality of mathematical modelling, with an important cultural and ritual element at play. An additional source of discomfort is the candour – which can be 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 [64] is something best ignored. Sensitivity analysis may reveal that the source of the poor (e.g. diffuse) inference is an assumption hard or impossible to test. It can be objected that this invasive approach opens 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 the principle of 'Organized Scepticism' [81], whereby all ideas must be tested and subjected to rigorous, structured community scrutiny.

In econometrics, a similar principle holds that 'honesty is the best policy' [65], and is formulated as a commandment: 'Thou shalt confess in the presence of sensitivity. Corollary: thou shalt anticipate criticism' [82].

Finally, does mathematical modelling truly need to be rescued by statistics, while statistics appears to be itself in a crisis of sorts? Philosopher Jerome R. Ravetz's prophesized in 1971 that entire research fields might become diseased, (p. 179 of the second edition of [83]), 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*." While statistics has been seen to possess the disciplinary arrangements and committed leaders to react to such a crisis, mathematical modelling currently lacks one. Statistics could help mathematical modelling to find its way [21].

Acknowledgment: The author gratefully acknowledges a grant from the Peder Sather Centre of Berkeley aimed to mainstream sensitivity analysis and auditing.

9 REFERENCES

[1] A. Saltelli and S. Funtowicz, "What is science's crisis really about?," *Futures*, vol. 91, pp. 5–11, 2017.

[2] D. Singh Chawla, "Big names in statistics want to shake up much-maligned P value," *Nature*, vol. 548, no. 7665, pp. 16–17, Jul. 2017.

[3] D. Colquhoun, "An investigation of the false discovery rate and the misinterpretation of p-values," *R. Soc. Open Sci.*, vol. 1, p. 140216, 2014.

[4] J. Leek, B. B. McShane, A. Gelman, D. Colquhoun, M. B. Nuijten, and S. N. Goodman, "Five ways to fix statistics," *Nature*, vol. 551, pp. 557–559, 2017.

[5] G. Gigerenzer and J. N. Marewski, "Surrogate Science," *J. Manage.*, vol. 41, no. 2, pp. 421–440, Feb. 2015.

[6] R. L. Wasserstein and N. A. Lazar, "The ASA's Statement onp-Values: Context, Process, and Purpose," *Am. Stat.*, vol. 70, no. 2, pp. 129–133, 2016.

[7] P. B. Stark and A. Saltelli, "Cargo-cult statistics and scientific crisis," *Significance*, Jul-2018.

[8] C. O'Neil, *Weapons of math destruction : how big data increases inequality and threatens democracy*. Random House Publishing Group, 2016.

[9] M. Giampietro and A. Saltelli, "Footprints to nowhere," *Ecol. Indic.*, vol. 46, pp. 610–621, Nov. 2014.

[10] L. Blomqvist, B. W. Brook, E. C. Ellis, P. M. Kareiva, T. Nordhaus, and M. Shellenberger, "Does the

Shoe Fit? Real versus Imagined Ecological Footprints," *PLoS Biol.*, vol. 11, no. 11, p. e1001700, Nov. 2013.

[11] P. Paruolo, M. Saisana, and A. Saltelli, "Ratings and rankings: voodoo or science?," *J. R. Stat. Soc. Ser. A (Statistics Soc.*, vol. 176, no. 3, pp. 609–634, Jun. 2013.

[12] D. Lawder and L. Wroughton, "World Bank economist Paul Romer quits after Chile comments," *Reuters*, 2018. [Online]. Available: https://www.reuters.com/article/us-worldbank-economist-romer/world-bank-economist-paul-romer-quits-after-chile-comments-idUSKBN1FD38Y. [Accessed: 21-Dec-2018].

[13] E. Popp Berman and D. Hirschman, "The Sociology of Quantification: Where Are We Now?," *Contemp. Sociol.*, vol. 47, no. 3, pp. 257–266, 2018.

[14] J. J. Padilla, S. Y. Diallo, C. J. Lynch, and R. Gore, "Observations on the practice and profession of modeling and simulation: A survey approach," *Simulation*, vol. 94, no. 6, pp. 493–506, Oct. 2018.

[15] A. J. Jakeman, R. A. Letcher, and J. P. Norton, "Ten iterative steps in development and evaluation of environmental models," *Environ. Model. Softw.*, vol. 21, no. 5, pp. 602–614, 2006.

[16] S. Funtowicz and J. R. Ravetz, *Uncertainty and Quality in Science for Policy*. Dordrecht: Kluwer, 1990.

[17] A. Saltelli, P. B. Stark, W. Becker, and P. Stano, "Climate Models as Economic Guides - Scientific Challenge of Quixotic Quest?," *Issues Sci. Technol.*, vol. 31, no. 3, pp. 1–8, 2015.

[18] O. H. Pilkey and L. Pilkey-Jarvis, *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future*. Columbia University Press, 2009.

[19] G. M. Hornberger and R. C. Spear, "An approach to the preliminary analysis of environmental systems," *J. Environ. Manage.*, vol. 12, no. 1, 1981.

[20] I. Chalmers and P. Glasziou, "Avoidable waste in the production and reporting of research evidence," *Lancet*, vol. 374, no. 9683, pp. 86–89, 2009.

[21] A. Saltelli, "Should statistics rescue mathematical modelling?," *arXiv*, vol. arXiv:1712, no. 06457, 2018.

[22] M. Szenberg, *Eminent economists : their life philosophies*. Cambridge University Press, 1992.

[23] H. G. Moeller, *Luhmann explained*. Open Court Publishing Company, 2006.

[24] F. Martin, "Unelected power: banking's biggest dilemma," *New Statemen*, 13-Jun-2018.

[25] D. Metlay, "From Tin Roof to Torn Wet Blanket: Predicting and Observing Ground Water Movement at a Proposed Nuclear Waste Site," in *Prediction: Science, Decision Making, and the Future of Nature*, D. Sarewitz, R. A. J. Pielke, and R. Byerly, Eds. Island Press, 2000.

[26] K. Campbell, A. Wolfsberg, J. Fabryka-Martin, and D. Sweetkind, "Chlorine-36 data at Yucca Mountain: statistical tests of conceptual models for unsaturated-zone flow," *J. Contam. Hydrol.*, vol. 62–63, pp. 43–61, Apr. 2003.

[27] W. C. Clark and G. Majone, "The Critical Appraisal of Scientific Inquiries with Policy Implications," *Sci. Technol. Hum. Values*, vol. 10, no. 3, pp. 6–19, Jul. 1985.

[28] R. V. O'Neill, "Perspectives in Hierarchy and Scale," in *Perspectives in Ecological Theory*, J. Roughgarden, R. M. May, and S. A. Levin, Eds. Princeton: Princeton University Press, 1989.

[29] L. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Trans. Syst. Man. Cybern.*, vol. 3, no. 1, pp. 28–44, 1973.

[30] B. Wynne, "Uncertainty and environmental learning 1, 2Reconceiving science and policy in the

preventive paradigm," *Glob. Environ. Chang.*, vol. 2, no. 2, pp. 111–127, Jun. 1992.

[31] F. Diaz-Maurin, "Atomic Homefront: a film about struggling to live with Manhattan Project radioactive waste," *Bulletin of the Atomic Scientists*, Jun-2018.

[32] OECD (Organisation for Economic Co-operation and Development), "PSACOIN Level S intercomparison," 1993.

[33] J. C. Helton, C. W. Hansen, and P. N. Swift, "Special Issue: Performance Assessment for the Proposed High-Level Radioactive Waste Repository at Yucca Mountain, Nevada," *Reliab. Eng. Syst. Saf.*, vol. 122, pp. 1–456, 2014.

[34] W. D. Nordhaus, "To Slow or Not to Slow: The Economics of The Greenhouse Effect," *Econ. J.*, vol. 101, no. 407, pp. 920–937, Jul. 1991.

[35] R. Kopp, D. Rasmussen, and M. Mastrandrea, "American Climate Prospectus: Economic Risks in the United States," 2014.

[36] A. Saltelli, S. Funtowicz, M. Giampietro, D. Sarewitz, P. B. Stark, and P. van der Slujis, Jeroen, "Modelling: Climate costing is politics not science," *Nature*, vol. 532, no. 7598, pp. 177–177, Apr. 2016.

[37] N. H. (Nicholas H. Stern, *Why are we waiting? : the logic, urgency, and promise of tackling climate change*. MIT Press, 2015.

[38] A. Weinberg, "Science and trans-science," *Minerva*, vol. 10, pp. 209–222, 1972.

[39] A. Saltelli and B. d'Hombres, "Sensitivity analysis didn't help. A practitioner's critique of the Stern review," *Glob. Environ. Chang.*, vol. 20, no. 2, pp. 298–302, May 2010.

[40] A. H. Louie, "Robert Rosen's anticipatory systems," *Foresight*, vol. 12, no. 3, pp. 18–29, Jun. 2010.

[41] S. Shapin, "Invisible Science," *Hedgehog Rev.*, vol. 18, no. 3, 2016.

[42] P. Wilmott and D. Orrell, *The Money Formula*. Wiley & Sons, 2017.

[43] P. Mirowski, *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown*. Verso, 2013.

[44] A. Pierce, "The Queen asks why no one saw the credit crunch coming," *The Telegraph*, 05-Nov-2008.

[45] P. Romer, "Mathiness in the Theory of Economic Growth," *Am. Econ. Rev.*, vol. 105, no. 5, pp. 89–93, May 2015.

[46] P. Romer, "The trouble with macroeconomics," *Am. Econ.*, vol. forthcomin, 2018.

[47] P. Romer, "Feynman Integrity," *www.paulromer.net*, 2015. [Online]. Available: https://paulromer.net/feynman-integrity/.

[48] R. P. (Richard P. Feynman, R. Leighton, and E. Hutchings, *Surely you're joking, Mr. Feynman!* W.W. Norton and Company, 1985.

[49] R. Nuzzo, "How scientists fool themselves – and how they can stop," *Nature*, vol. 526, no. 7572, pp. 182–185, Oct. 2015.

[50] J. P. A. Ioannidis, T. D. Stanley, and H. Doucouliagos, "The Power of Bias in Economics Research," *Econ. J.*, vol. 127, no. 605, pp. F236–F265, 2017.

[51] A. O'Hagan, "Polynomial Chaos: A Tutorial and Critique from a Statisticianís Perspective," 2013.

[52] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Update 1 of: Sensitivity Analysis for Chemical Models," *Chem. Rev.*, vol. 112, no. 5, pp. PR1-PR21, May 2012.

[53] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," *Comput. Phys. Commun.*, vol. 181, no. 2, pp. 259–270, Feb. 2010.

[54] A. Saltelli, S. Tarantola, and F. Campolongo, "Sensitivity Anaysis as an Ingredient of Modeling," *Stat. Sci.*, vol. 15, no. 4, pp. 377–395, Nov. 2000.

[55] A. Saltelli, S. Tarantola, and K. P.-S. Chan, "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output," *Technometrics*, vol. 41, no. 1, pp. 39–56, Feb. 1999.

[56] A. Saltelli, "Sensitivity Analysis for Importance Assessment," *Risk Anal.*, vol. 22, no. 3, pp. 579–590, Jun. 2002.

[57] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Comput. Phys. Commun.*, vol. 145, no. 2, pp. 280–297, May 2002.

[58] G. Majone, *Evidence, argument, and persuasion in the policy process*. Yale University Press, 1989.

[59] European Commission, "Guidelines on Impact Assessment - European Commission," 2015. .

[60] Office of Management and Budget, "Proposed Risk Assessment Bulletin," 2006.

[61] U.S. Environmental Protection Agency (EPA), "Guidance on the Development, Evaluation, and Application of Environmental Models," 2009.

[62] F. Ferretti, A. Saltelli, and S. Tarantola, "Trends in sensitivity analysis practice in the last decade," *Sci. Total Environ.*, vol. 568, pp. 666–670, Oct. 2016.

[63] A. Saltelli *et al.*, "Why So Many Published Sensitivity Analyses Are False. A Systematic Review of Sensitivity Analysis Practices," *ArXiv*, vol. 1711.11359, 2017.

[64] E. E. Leamer, "Sensitivity Analyses Would Help," *Am. Econ. Rev.*, vol. 75, no. 3, pp. 308–313, 1985.

[65] E. E. Leamer, "Tantalus on the Road to Asymptopia," *J. Econ. Perspect.*, vol. 24, no. 2, pp. 31–46, May 2010.

[66] A. Saltelli, Â. Guimaraes Pereira, J. P. . van der Sluijs, and S. Funtowicz, "What do I make of your latinorumc Sensitivity auditing of mathematical modelling," *Int. J. Foresight Innov. Policy*, vol. 9, no. 2/3/4, pp. 213–234, 2013.

[67] A. Gelman and E. Loken, "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," 2013.

[68] M. Friedman, *Essays in Positive Economics*. University of Chicago Press, 1953.

[69] E. S. Reinert, S. Endresen, I. Ianos, and A. Saltelli, "Epilogue: the future of economic development between utopias and dystopias," in *Handbook of Alternative Theories of Economic Development*, Edward Elgar Publishing, 2016, pp. 738–786.

[70] J. E. Stiglitz, "Rethinking Macroeconomics: What Failed, and How to Repair It," *J. Eur. Econ. Assoc.*, vol. 9, no. 4, pp. 591–645, Jan. 2011.

[71] W. K. T. Cho, "Algorithms can foster a more democratic society," *Nature*, vol. 558, no. 7711, pp. 487–487, Jun. 2018.

[72] A. Saltelli and S. Funtowicz, "When All Models Are Wrong," *Issues Sci. Technol.*, vol. 30, no. 2, pp. 79–85, 2014.

[73] S. N. Lane, N. Odoni, C. Landström, S. J. Whatmore, N. Ward, and S. Bradley, "Doing flood risk science differently: an experiment in radical scientific method," *Trans. Inst. Br. Geogr.*, vol. 36, no. 1,

pp. 15–36, Jan. 2011.

[74] A. Saltelli and M. Giampietro, “What is wrong with evidence based policy, and how can it be improved?,” *Futures*, vol. 91, pp. 62–71, Feb. 2017.

[75] L. Araujo, A. Saltelli, and S. V. Schnepf, “Do PISA data justify PISA-based education policy?,” *Int. J. Comp. Educ. Dev.*, vol. 19, no. 1, pp. 20–34, 2017.

[76] A. Saltelli and S. Lo Piano, “Problematic quantifications: a critical appraisal of scenario making for a global ‘sustainable’ food production,” *Food Ethics*, vol. 1, no. 2, pp. 173–179, 2017.

[77] R. D. Morey *et al.*, “The Peer Reviewers’ Openness Initiative: incentivizing open research practices through peer review,” *R. Soc. Open Sci.*, vol. 3, p. 150547, 2016.

[78] R. Brauneis and E. P. Goodman, “Algorithmic Transparency for the Smart City,” *Yale J. Law Technol.*, vol. 20, pp. 103–176, 2018.

[79] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. Springer-Verlag, 2003.

[80] W. L. Oberkampf and C. J. Roy, *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[81] R. Merton, *The sociology of science: Theoretical and empirical investigations*. 1973.

[82] P. Kennedy, *A guide to econometrics*. Wiley-Blackwell; 6 edition, 2008.

[83] J. R. Ravetz, *Scientific knowledge and its social problems*. Oxford University Press, 1971.