Sensitivity auditing

Book or Report Section

Published Version

Lo Piano, S. ORCID: https://orcid.org/0000-0002-2625-483X, Sheikholeslami, R., Puy, A. and Saltelli, A. (2023) Sensitivity auditing. In: Saltelli, A. and Di Fiore, M. (eds.) The politics of modelling. Oxford University Press, Oxford, pp. 121-136. ISBN 9780198872412 doi: https://doi.org/10.1093/oso/9780198872412.003.0008 Available at https://centaur.reading.ac.uk/119501/

To link to this article DOI: http://dx.doi.org/10.1093/oso/9780198872412.003.0008

Publisher: Oxford University Press

PART III

THE RULES IN PRACTICE

8

Sensitivity auditing

A practical checklist for auditing decision-relevant models

Samuele Lo Piano, Razi Sheikholeslami, Arnald Puy, and Andrea Saltelli

Introduction

The *Manifesto* that offers the occasion for the present volume (Saltelli et al. 2020a) is based on five rules to improve the way models are used for policy-making and advocate for a better reciprocal domestication between models and society overall. The rules are themselves the outcome of the sedimentation of earlier checklists for model quality, foremost that provided by sensitivity auditing, with contributions from other strands of scholarship.

It would be natural at this point to ask 'What is sensitivity auditing?' The short answer would be that it is an extension of sensitivity analysis (see chapter 5). In turn, sensitivity analysis is the logical complement of an uncertainty analysis, also known as uncertainty quantification. Some definitions will be helpful here.

Uncertainty analysis: By assigning a range of variation and a distribution to uncertain model inputs and assumptions, one can generate an empirical distribution function for the output(s) of interest by running the model on samples from these distributions. The expressions 'error propagation analysis' and 'uncertainty cascade' are also used (Christie et al. 2011: 86). See also chapter 2.

Sensitivity analysis is a methodology used to ascertain which of the uncertain inputs is more influential in generating the uncertainty in the output.

These two analyses 'talk' to one another. If the output has little or no uncertainty, there is no point in dissecting its uncertainty to discover 'the culprits'. Even the extreme opposite scenario, one where uncertainty is large enough to impair the meaningful application of the model (see chapter 4), does not offer a chance for any meaningful inference. In all cases in between, discovering the responsible inputs may help alleviate the issue.

These analyses imply many choices on top of the usual assumptions linked to the construction of the model as such: one has to decide what inputs are to be taken to be uncertain, how to choose ranges and distributions (often a very expensive and time-consuming step), and what method to employ to select 'sensitive' factors.

Samuele Lo Piano et al., *Sensitivity auditing*. In: *The politics of modelling*. Edited by: Andrea Saltelli and Monica Di Fiore, Oxford University Press. © Oxford University Press (2023). DOI: 10.1093/oso/9780198872412.003.0008

It is not difficult to imagine a situation where a model is used in decision- or policy-making, and where different actors have different visions of what should be modelled and how. In this case, the entire modelling process and its conclusions, including its technical uncertainty and sensitivity analyses, could be deconstructed by contesting the choices just mentioned.

The examples provided in this chapter take this approach by exploring and bringing into the open subjective or normative elements of a modelling process. In other words, a 'technical' sensitivity analysis cannot be the end of the story when the model undergoes a regulatory audit or becomes the subject of a public debate—as we have seen for the case of models related to COVID-19 (Saltelli et al. 2020a).

This approach goes by the name of *sensitivity auditing* (Saltelli and Funtowicz 2014), which is recommended *inter alia* by the European Commission and the SAPEA (2018). Sensitivity auditing consists of a seven-rule checklist (Box 8.1):

Box 8.1 The seven-rule checklist of sensitivity auditing

1. Check against a rhetorical use of mathematics.
2. Adopt an 'assumption hunting' attitude.
3. Detect Garbage In Garbage Out.
4. Find sensitive assumptions before they find you.
5. Aim for transparency.
6. Do the right sums, not just the sums right.
7. Perform thorough, state-of-the-art uncertainty and sensitivity analyses.

The first rule is *Check against a rhetorical use of mathematics*. As already discussed in chapter 4, larger models command more epistemic authority and discourage criticism. This rule invites us to appreciate the dimension of a model in relation to both its context and purpose, and the evidence that has entered the model's construction. *Adopt an 'assumption hunting' attitude* is the second rule. Models are based on assumptions, including interpretations of the underlying systems' behaviour. Some of these assumptions come with the 'tag' of the model, and are explicit—for example, when we say, 'This is an equilibrium model.' More often, assumptions are implicit. Since the construction of a model may unfold over an extended period, the same modellers may forget them.[1] The next step of sensitivity auditing is *Detect Garbage In Garbage Out (GIGO)*. This recommendation points out the circumstances where the uncertainty associated with a mathematical prediction has been overstated (magnifying uncertainty) or

[1] From Millgram (2015: 29): 'normally even the people who produced the model will not remember many of the assumptions incorporated into it, short of redoing their work, which means that the client cannot simply ask then what went into it'.

understated (hiding uncertainty). The latter case is perhaps the most frequent. When minimizing and/or simplifying uncertainties, a modeller aims to show that the prediction of the model is 'crisp'. For instance, nobody is interested in a cost–benefit analysis whose distribution output spans from a high loss to a high gain with equal probabilities. This strategic behaviour—'where the uncertainty of the inputs must be suppressed, lest they render its outputs totally indeterminate'—was named GIGO by Funtowicz and Ravetz (1990: 6). Reluctance to face uncertainty is a well-known issue in evidence-based policy (Scoones and Stirling 2020). The opposite gamble may be embraced to increase the uncertainty and, in doing so, defeat the assessment of, say, a regulatory agency (Saltelli 2018, Saltelli et al. 2022).

The fourth rule is *Find sensitive assumptions before they find you*. Modellers expecting a public debate around the inference produced by their models ought to better prepare in advance by running a proper sensitivity analysis. An interesting case relating to a dispute over the cost of climate change is described in Saltelli and D'Hombres (2010). Here, one of the parties in the dispute, Nicholas Stern, resorted to a sensitivity analysis after his impact assessment had been contested. However, his sensitivity analysis appeared weak when seen through the lens of sensitivity auditing (ibid.).

Aim for transparency is the fifth rule. Black-box models may simply not play well in a public debate. The open science movement strongly advocates for transparency and availability of the model source code, and making this intelligible through appropriate documentation and comments.

The sixth rule is *Do the right sums, not just the sums right*. Though this may appear the same as rule (2), it has more to do with major political or ideological worldviews. It reflects sociologists' thinking about the so-called technologies of humility (Jasanoff 2003), whereby in the production of evidence for policy, one should be careful to identify possible winners and losers, and make sure that the concerns of the latter are not missed in the debate. This rule also reflects on how quantification can determine the policy of what is being quantified (Salais 2010) in a way that is not immediately apparent due to the purported neutrality of mathematical models. Finally, the last rule is *Perform thorough, state-of-the-art uncertainty and sensitivity analyses*. Leaving undiscussed the results of a sensitivity analysis, or presenting one that is based on an inadequate design, is still fairly common practice in modelling studies across disciplinary fields (Ferretti, Saltelli, and Tarantola 2016, Saltelli et al. 2019, Lo Piano and Benini 2022); see also chapter 5).

The epistemic background of sensitivity auditing

Sensitivity auditing takes inspiration from post-normal science (PNS) (Funtowicz and Ravetz 1993), a paradigm in science for policy that applies when facts are uncertain, values in dispute, stakes high, and decisions urgent (for example in the presence of a pandemic caused by a new virus). While PNS is not easy to synthesize, its main feature is an emphasis on the quality of the assessment process, rather than the pursuit of a truth that might be elusive in conflicted issues. In a similar spirit, PNS does not appeal to the neutrality of a unitary science, and accepts that different disciplines and accredited experts may uphold different legitimate views (Saltelli et al. 2020b).

Under this paradigm, the decision-making community needs to involve more than just the experts; it should also include investigative journalists, whistleblowers, and lay citizens affected by or simply interested in the issue under analysis. Ideally, sensitivity auditing itself is also meant for use in participatory settings, where one negotiates the worth of a model, as well as the order of worth of the various sides of a problem. By the latter it means the relative position on the issue under consideration of concerns about, for example, social justice, environmental quality, or respect for existing values and traditions (Thévenot 2022). Like sensitivity analysis, sensitivity auditing functions in both construction and deconstruction, either in building a defensible, plausible model, or in demonstrating the irrelevance or contradictions of a model-based inference one is trying to contextualize, in the spirit of statactivism (Bruno, Didier, and Prévieux 2014)—for example, when one strives to bring to the surface issues or categories that an existing quantification has made invisible.

Sensitivity auditing has similarities to an older scheme of practices for the quality of numerical information known as Numeral Unit Spread Assessment Pedigree (NUSAP), which was also inspired by the PNS paradigm. NUSAP was introduced by Funtowicz and Ravetz (1990) as a way of shedding light on the quality of numbers being used in a political setting. The first two categories are self-evident: Spread reviews the error in the value of the Numeral, and Assessment and Pedigree are supposedly the results of a process of negotiation among interested parties. Assessment is a summary judgement of the quality of the numerical evidence, and might be as simple as 'conservative' or 'optimist'. Pedigree is a judgement of the quality of the process leading to the numeral, and of the numeral's policy implications. What is important in NUSAP is not so much the scores assigned to Assessment and Pedigree, but the occasion these categories offer us to engage in a reflection on the worth of the numeral (van der Sluijs et al. 2005).

The present volume has already mentioned the expanding discipline of sociology of quantification (Mennicken and Espeland 2019, Popp Berman and Hirschman 2018, Mennicken and Salais 2022). Sensitivity auditing was not developed by sociologists, yet the approach reflects many tenets of this discipline: sociologists of quantification investigate how different mechanisms of quantification are performative in shaping the reality that they should supposedly just be measuring. For this reason, one should assess a given quantification both in terms of its technical quality and its fairness. Disparate and important dimensions such as higher education (Espeland and Sauder 2016), employment (Salais 2010), and inequality (De Leonardis 2022) come to be as a result of a digitization to a large extent subtracted from democratic deliberation. While we cannot enter into this discussion in more depth here, we note that sensitivity auditing addresses this 'fairness' dimension of quantification for the case of mathematical models, and forms of quantification in general.

Finally, sensitivity auditing is one possible approach to the 'modelling of the modelling process' (Lo Piano et al. 2022, Puy, Lo Piano, and Saltelli 2020). To illustrate what this means, we need to call upon novelist Jorge Luis Borges and his short story, 'The Garden of Forking Paths' (Borges 1941). The story has been taken up in statistics (Gelman and Loken 2013) to allude to the many degrees of freedom offered to an analyst marching towards an inference. Additionally, feeding the same data to different teams can return varied statistical results (Breznau et al. 2021), and it is only natural that the effect is also seen in mathematical modelling (Refsgaard et al. 2006). Modelling of the modelling process implies, like in Borges' short story, taking all the paths instead of just one, and exploring the consequences of these different trajectories on the inference.

The next section illustrates how sensitivity auditing can be applied to models, indicators, and metrics. We will target instances of quantification used in policy-making and/or those that have received vast media coverage. Although our illustration proceeds by taking one rule at a time for each case, it should be clear that the rules can be recursive, that their order can change, and that to some extent rules overlap with one another.

The Programme for International Student Assessment (PISA)

The PISA test was designed by the Organisation for Economic Co-operation and Development (OECD) to measure the problem-solving skills of 15-year-old school pupils. The test has been undertaken every three years since 2000. Test performances on a country basis are ranked and benchmarked against a standard. In the 2018 round of the test, 79 countries participated. Criticisms of the methodological and ideological stances of the test, among other aspects, were raised in the press (see Araujo, Saltelli, and Schnepf 2017) and quoted references can be seen in chapter 9.

Several dimensions emerged after applying the sensitivity auditing checklist to the PISA test score.

Check against a rhetorical use of mathematics: A reading of the test results that emphasizes a causal relation from the test scores to economic growth has been offered by prominent political bodies, including the European Commission. A document released stated that if European Union (EU) countries could significantly increase their PISA score, this would lead to a quantifiable growth in Gross Domestic Product (GDP) for the EU (Woessmann 2016).

Adopt an 'assumption hunting' attitude: The PISA test builds on the postulate that the skills students require to succeed in current knowledge societies can be benchmarked against a one-size-fits-all international standard. Differences in curricula between countries are suppressed and excluded from the test. However, countries' wellbeing and success may emerge from these very curricular differences.

Detect pseudo-science: When communicating the test results, the survey organizers report only the standard error of the countries' test scores, neglecting important sources of volatility in country rankings. These sources of volatility may emerge, for instance, by excluding students with special educational needs or newly arrived immigrants. In past editions of the test, excluding these groups resulted in discarding more than 5% of the potential participants from some countries. However, this 5% is above the threshold imposed by the test designers as an assurance of its representativeness (Wuttke 2007). Even the tendency of less capable students to refrain from participating in the test may result in a non-representative participating cohort, thus potentially producing a bias that goes well beyond the standard error of the country's rank (Micklewright, Schnepf, and Skinner 2012).

Find sensitive assumptions before they find you: Calls for testing the sensitivity of the PISA rankings and the volatility they bring to modelling assumptions, data collection method, and use of the data items have gone unanswered (Micklewright and Schnepf 2006).

Aim for transparency: Lack of data availability represents one of the main limitations of the PISA test. This impedes an analysis of the sensitivity of the achievement scores to the modelling choices and data items used in attributing scores to countries.

Do the right sums, not just the sums right: The PISA test was conceived for the purpose of facilitating a measurement of the degree to which the teaching that students receive is useful in terms of the life challenges they may encounter in today's knowledge societies. In this sense, the test *de facto* makes the 'economic' case for education, which is framed exclusively as a means to economic growth rather than to *Bildung* and emancipation.

Perform thorough, state-of-the-art uncertainty and sensitivity analyses: Thoroughly assessing all the sources of uncertainty in the PISA rankings would require their simultaneous activation. The resulting uncertainty could produce more volatile rankings, which would require adequate communication: for example, 'uncertainties being taken into consideration, the rank of country X could vary between five and twenty' (Araujo, Saltelli, and Schnepf 2017). However, this analysis is missing from the communication on the significance of the produced countries' rank.

Nutrition and public health economic evaluations

Non-communicable diseases (NCDs) are epidemiologically affected by lifestyle habits, including diet, smoking, and physical activity level. For this reason, it is of primary importance that we evaluate how policies aimed at triggering changes in these factors may produce societal consequences in terms of disease likelihood. These policies are, in turn, informed by the available body of evidence and modelling activities. Lo Piano and Robinson (2019) identified the following criticalities:

Check against a rhetorical use of mathematics: In the field of nutrition and public health economic evaluations, an existing tendency is to resort to overly complex models. This pattern may be driven by researchers' love for their craft, which motivates them to prioritize the full use of available computational resources rather than addressing the policy issue. This translates into, for instance, systematically resorting to Markov chain models despite the availability of representations that are less computationally demanding (Clarke et al. 2005). Additionally, cross-comparing different modelling typologies is a practice that is scarcely explored in the field.

Adopt an 'assumption hunting' attitude: A study's conclusion may depend on an assumption whose impact is untested. In the NCD domain, the most critical assumptions include: the modelling of the dose responses adopted in terms of risk factor estimates; how the dose intake varies upon policy implementation, especially across sociodemographic cohorts and geographical areas; the timeframe of the interventions and their (diminishing/increasing) returns; and the actual NCDs taken into account in the modelling exercise, as well as their change over the timeframe considered (Lo Piano and Robinson 2019).

Detect Garbage In Garbage Out: The strategy of reducing uncertainty in terms of nutrient intake has been adopted to downplay the overall output uncertainty in some modelling activities (Lo Piano and Robinson 2019). Some modellers also produced estimates of the incidence and prevalence of diseases by resorting to educated guesses when information of sufficient quality was not available (European Society of Cardiology 2012).

Find sensitive assumptions before they find you: Uncertainty and sensitivity analyses should be used more widely in the field, especially in light of the number of assumptions made. For instance, the impact of the method used to impute missing data in terms of dose intake on output uncertainty (using other countries' figures or correcting algorithms) could be tested in these settings.

Aim for transparency: Transparent modelling should be the standard for enabling scrutiny from peers *a fortiori* in a decision-making context when models are to be used to inform policies that will eventually influence citizens' lives. For example, the impact of policies to compare the growing prevalence of overweight and obesity in children in the EU was tested in the proprietary model Joint Action on Nutrition and Physical Activity (JANPA), which was not available for public scrutiny (Lo Piano and Robinson 2019).

Do the right sums, not just the sums right: Health evaluations emphasize the economic dimension by glossing over social and cultural aspects of lifestyle and nutrition choice. The latter, however, could be crucial for citizens when prioritizing options. Health and quality of life are captured as per the normative dimensions of citizens' values, which are not necessarily captured through monetary proxies.

Perform thorough, state-of-the-art uncertainty and sensitivity analyses: Sensitivity analysis is known and applied in several nutrition and public health economic evaluations. However, in the vast majority of the cases, these are implemented by varying one factor at a time. As these models are likely to be non-additive, this approach fails to capture interactions between factors. More seriously, in this kind of sensitivity analysis, the vast majority of the output uncertainty space is left unexplored (Saltelli and Annoni 2010). Hence, uncertainty is not adequately characterized or apportioned.

Sociohydrology

The field of sociohydrology has emerged in response to the failure of the traditional, anthropocentric paradigm in water resource management, where human activity is deemed a mere boundary condition of the hydrologic model (Sivapalan, Savenije, and Blöschl 2012a). Sociohydrology seeks to understand how the hydrological cycle changes according to interactions among natural and human forces. Sociohydrology uses Coupled Human and Water Systems (CHAWS) models to address the complex water-related problems currently facing human societies. The key features of CHAWS, which include complexity, cross-scale dynamics, and uncertainty (Sivapalan 2015, Sivapalan, Savenije, and Blöschl 2012b, Wheater and Gober 2013, Liu et al. 2007), make sociohydrology a domain in which the adoption of the PNS paradigm and sensitivity auditing is fully justified.

Check against a rhetorical use of mathematics: CHAWS rest on the assumption that uncertainty reduction and minimization in dynamic and emergent human–water systems can be achieved by adding complexity to the models. This approach is also resorted to in order to rectify the discrepancies between model outputs and observed values (see, for example, Pan et al. (2018)). As a case in point, more parameters are added to justify the social processes included. As things stand in terms of the development of sociohydrology, however, sophistication of the social processes in CHAWS models cannot match the high level of detail seen in hydrologic models due to the lack of knowledge.

Adopt an 'assumption hunting' attitude: CHAWS models contain several assumptions regarding human values, beliefs, and norms related to water use, livelihoods, and the environment (Alonso Vicario et al. 2020, Hemmati et al. 2021). These cultural norms and values drive human behaviour with respect to water resources. They have typically been conceptualized in a 'black box' manner and represented by proxy data (Roobavannan et al. 2018). However, to make a judgement regarding the quality of this parameterization of cultural values, more information would be needed than is currently available in CHAWS.

Detect Garbage In Garbage Out: Risks of future flood damage are underestimated in CHAWS due to issues such as short collective memory, excessive trust in flood protection structures, and a high level of risk-taking in models (Viglione et al. 2014).

Find sensitive assumptions before they find you: The pitfalls of one-factor-at-a-time sensitivity analysis are well known in the hydrology community. However, this approach has been widely used in sociohydrology (see, for example, Liu et al. 2015 and Srinivasan 2015).

Aim for transparency: CHAWS models are not typically open access and are only available for limited case studies. This prevents the community from verifying prior results. In this context, agent-based modelling, for example, is one of the commonly used modelling tools in CHAWS (Shafiee and Zechman 2013, Zhao, Cai, and Wang 2013, Pouladi et al. 2019). However, these models are usually not well documented and lacking in descriptions (Grimm et al. 2006), hence leading to scarcely reproducible results.

Do the right sums, not just the sums right: CHAWS may be attempting to answer the wrong question. It asks 'What is the most likely future?' instead of 'What kind of future do we want and what are the consequences of different policy decisions relative to that desired future?' (Gober and Wheater 2015).

Perform thorough, state-of-the-art uncertainty and sensitivity analyses: In this domain, a systematic global sensitivity analysis can avoid implausible results by exploring and apportioning the output uncertainty to its driving factors and their interactions. It may do this by, for instance, simultaneously considering the uncertainties related to the hydrologic environment (for example, non-stationarity deriving from anthropogenic factors, such as changes in land use, climate, and water use) and the social aspect of CHAWS (socioeconomic development, demography, and agent behaviour) (Sheikholeslami et al. 2019). However, global sensitivity analysis has rarely been performed. Two notable examples are Elshafei et al. (2016) and Ghoreishi, Razavi, and Elshorbagy (2021). These authors investigated the sensitivity of CHAWS to model parameters, which govern the internal dynamics of the system and determine the external sociopolitical context. Ghoreishi et al. (2021) performed a global sensitivity analysis on their agent-based agricultural water demand model to determine the most important model parameters to be calibrated.

Food security

Evaluating the outcomes of humanity's use of the Earth's resources is tightly linked to estimating its overall level of appropriation, particularly in terms of determining the nature of human activities on Earth and the resources needed to sustain them. In this context, a prominent challenge is food security—that is, meeting the nutritional needs of a growing global population, as also recognized in the United Nations agenda for 2030 (United Nations 2021), including a Zero Hunger Strategy cutting across several Sustainable Development Goals. Bahadur KC et al. (2016) proposed their own recipe for achieving healthy and sustainable food provision based on innovative agricultural techniques and dietary re-adaptation. The contribution narrates a successful story from the perspective of a person in 2050 looking back to the past. Saltelli and Lo Piano (2017) used this work as a test benchmark for a deconstruction along the lines of sensitivity auditing.

Check against a rhetorical use of mathematics: The package of policies proposed to assure more sustainable farming and healthy diets consisted of the following: consumer education; increasing the cost of unhealthy food; capturing the environmental costs associated with farming; reducing corn subsidy in the US; and enhancing storage and processing facilities in the developing world. This mixture of policies raises some questions of viability, particularly in relation to potential unintended consequences, as discussed below.

Adopt an 'assumption hunting' attitude: The proposed assessment rests on a number of assumptions that are poorly explored. One is neglecting the principle of diminishing return when projecting the yield increase in cultivation to the year 2050—a constant increase in yield over decades hits against known phenomena of topsoil erosion and exhaustion. Additionally, the study assumes a less caloric diet for an increased share of the adult population in the year 2050, and a lower extension of cultivated land globally due to higher yields being obtained with less impactful agricultural techniques. All of these assumptions appear to err on the side of optimism—for example, they do not consider the adjustment costs of the transformation or the realism of popular acceptance of the suggested policies.

Detect Garbage In Garbage Out: Reductions in global cultivated land area are estimated at 438 million hectares for the year 2050, with three significant digits. However, uncertainty in the current yearly global cultivated land extensions amounts to around 20% (around 1000 million hectares according to one of the reference databases (Lo Piano 2017)), making an accuracy of three significant digits for projections to 2050 implausible.

Find sensitive assumptions before they find you: The sensitivity of the output estimates to the input assumptions was left unaddressed in the study published by Bahadur KC et al. (2016).

Aim for transparency: The fact that the model underpinning the quantitative scenario is only available upon request hampers scrutiny by peers and policy makers. Models should be made available along with their quantitative outcomes so as to fully foster their replication and scrutiny.

Do the right sums, not just the sums right: The analysis proposed adopts primarily the standpoint of developed countries in pursuing food security with technical solutions and a policy package tailored to this area of the world. Nevertheless, there remain significant political issues of power asymmetry for developing countries in the context of the international food commodity trade. This translates into an unequal caloric exchange with regard to food crops (Falconí, Ramos-Martin, and Cango 2017) that is not explored in the study of food security. Briefly, a political problem has been reframed into a technical one, while the policy proposals are designed to meet the needs of a minority.

Perform thorough, state-of-the-art uncertainty and sensitivity analyses: The uncertainty space has not been explored, because all the information is conveyed with crisp, uncertainty-free figures.

Conclusion

The case studies examined show how sensitivity analysis can be used in practical terms. These applications open the quantifications to inspection by other disciplines, including social sciences studies. The relation of sensitivity auditing with the loose community of practitioners of PNS is also an element favouring the take-up of its rules, whose spirit—as discussed—is for practical purposes similar to that of the *Manifesto*. The relation between modelling and society has been made more intense and at the same time more conflicted by the COVID-19 pandemic (Pielke 2020, Rhodes and Lancaster 2020). The present crisis of trust in expertise also affects mathematical modelling, and has links to the political crisis affecting several mature democracies under the paradigm that 'solutions to the problem of knowledge are solutions to the problem of social order' (Shapin and Schaffer 2011: 387). Sociology of quantification is actively mapping this territory of conflict for the case of statistics, yet more work is expected for mathematical modelling.

Criticizing the political use of statistics in what he calls '[g]overnance driven quantification', Salais (2010) talks about the reversal of the statistical pyramid, whereby instead of statistics generating concepts and categories useful for collective learning, indicators are produced to demonstrate that preselected policies are efficiently achieved. This is the known nemesis of evidence-based policy into policy-based evidence; among scholars, there is an expectation that these instrumental uses of evidence might come under increasing criticism (van Zwanenberg 2020) (see chapter 7).

References

Alonso Vicario, S., M. Mazzoleni, S. Bhamidipati, M. Gharesifard, E. Ridolfi, C. Pandolfo, and L. Alfonso. 2020. 'Unravelling the Influence of Human Behaviour on Reducing Casualties during Flood Evacuation'. *Hydrological Sciences Journal*, *65*(*14*), 2359–2375, doi: 10.1080/02626667.2020.1810254.

Araujo, L., A. Saltelli, and S.V. Schnepf. 2017. 'Do PISA Data Justify PISA-Based Education Policy?' *International Journal of Comparative Education and Development*, *19*(*1*), 20–34, doi: 10.1108/ IJCED-12-2016–0023.

Bahadur KC, K., E.F. Krishna, G. Dias, T. Zundel, and S. Pascoal. 2016. 'Pathways Leading to a More Sustainable and Healthy Global Food System'. *The Solutions Journal*, *7*(*5*), 10–12.

Borges, J. 1941. *El Jardin de los Senderos que se Bifurcan*, SUR, Buenos Aires.

Breznau, N., E.M. Rinke, A. Wuttke, H.H.V. Nguyen, M. Adem, J. Adriaans, A. Alvarez-Benjumea, H.K. Andersen, D. Auer, F. Azevedo, O. Bahnsen, D. Balzer, G. Bauer, P.C. Bauer, M. Baumann, S. Baute, V. Benoit, J. Bernauer, C. Berning, … T. Żółtak. 2022. 'Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty', *Proceedings of the National Academy of Sciences*, *119*(*44*), e2203150119, doi: 10.1073/pnas.2203150119.

Bruno, I., E. Didier, and J. Prévieux. 2014. *Statactivisme. Comment lutter avec des nombres*, Zones, La Découverte, Paris.

Christie, M., A. Cliffe, P. Dawid, and S.S. Senn. 2011. *Simplicity, Complexity and Modelling*, Wiley, New York.

Clarke, P.M., A.M. Gray, A. Briggs, R.J. Stevens, D.R. Matthews, R.R. Holman, and on behalf of the UK Prospective Diabetes Study (UKPDS). 2005. 'Cost-Utility Analyses of Intensive Blood Glucose and Tight Blood Pressure Control in Type 2 Diabetes (UKPDS 72)'. *Diabetologia*, *48*(*5*), 868–77, doi: 10.1007/s00125-005-1717-3.

De Leonardis, O. 2022. 'Quantifying Inequality: From Contentious Politics to the Dream of an Indifferent Power'. In *The New Politics of Numbers: Utopia, Evidence and Democracy*, edited by A. Mennicken and R. Salais, Executive Politics and Governance. Springer International Publishing, Cham, pp. 135–66, doi: 10.1007/978-3-030-78201-6_5.

Elshafei, Y., M. Tonts, M. Sivapalan, and M.R. Hipsey. 2016. 'Sensitivity of Emergent Sociohydrologic Dynamics to Internal System Properties and External Sociopolitical Factors: Implications for Water Management'. *Water Resources Research* *52*(*6*), 4944–66, doi: 10.1002/2015WR017944.

Espeland, W.N. and M. Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*, Russell Sage Foundation, New York.

European Society of Cardiology. 2012. *European Cardiovascular Disease Statistics 2012*, Tech. Rep. European Heart Network & European Society of Cardiology, Sophia Antipolis, France.

Falconí, F., J. Ramos-Martin, and P. Cango. 2017. 'Caloric Unequal Exchange in Latin America and the Caribbean'. *Ecological Economics*, *134*, 140–9, doi: 10.1016/j.ecolecon.2017.01.009.

Ferretti, F., A. Saltelli, and S. Tarantola. 2016. 'Trends in Sensitivity Analysis Practice in the Last Decade'. *Science of The Total Environment*, *568*, 666–70, doi: 10.1016/j.scitotenv.2016. 02.133.

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

Funtowicz, S. and J.R. Ravetz. 1993. 'Science for the Post-Normal Age'. *Futures*, *25*(*7*), 739–55, doi: 10.1016/0016-3287(93)90022- L.

Gelman, A. and E. Loken. 2013. 'The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "p-Hacking" and the Research Hypothesis Was Posited Ahead of Time', *Statistical Modeling, Causal Inference, and Social Science* [Blog], https://statmodeling. stat.columbia.edu/2021/03/16/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-2/ (last accessed 14 April 2023).

Ghoreishi, M., S. Razavi, and A. Elshorbagy. 2021. 'Understanding Human Adaptation to Drought: Agent-Based Agricultural Water Demand Modeling in the Bow River Basin, Canada'. *Hydrological Sciences Journal*, *66*(*97*), 389–407, doi: 10.1080/02626667.2021.1873344.

Gober, P. and H.S. Wheater. 2015. 'Debates—Perspectives on Socio-Hydrology: Modeling Flood Risk as a Public Policy Problem'. *Water Resources Research*, *51*(*6*), 4782–8, doi: 10.1002/2015WR016945.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S.K. Heinz, G. Huse, A. Huth, J.U. Jepsen, C. Jørgensen, W.M. Mooij, B. Müller, G. Pe'er, C. Piou, S.F. Railsback, A.M. Robbins, M.M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R.A. Stillman, R. Vabø, U. Visser, and D.L. DeAngelis. 2006. 'A Standard Protocol for Describing Individual-Based and Agent-Based Models'. *Ecological Modelling*, *198*(*1*), 115–26, doi: 10.1016/j.ecolmodel.2006.04.023.

Hemmati, M., H.N. Mahmoud, B.R. Ellingwood, and A.T. Crooks. 2021. 'Unraveling the Complexity of Human Behavior and Urbanization on Community Vulnerability to Floods'. *Scientific Reports*, *11*(*1*), 20085, doi: 10.1038/s41598-021-99587-0.

Jasanoff, S. 2003. 'Technologies of Humility: Citizen Participation in Governing Science'. *41*(*3*), 223–44, doi: 10.1023/A:1025557512320.

Liu, D., F. Tian, M. Lin, and M. Sivapalan. 2015. 'A Conceptual Socio-Hydrological Model of the Coevolution of Humans and Water: Case Study of the Tarim River Basin, Western China'. *Hydrology and Earth System Sciences*, *19*(*2*), 1035–54, doi: https://doi.org/10.5194/hess-19-1035–2015.

Liu, J., S.R.T. Dietz, M.A. Carpenter, C. Folke, E. Moran, A.N. Pell, P. Deadman, T. Kratz, J. Lubchenco, E. Ostrom, Z. Ouyang, W. Provencher, C.L. Redman, S.H. Schneider, and W.W. Taylor. 2007. 'Complexity of Coupled Human and Natural Systems'. *Science 317*(*5844*), 1513–16, doi: 10.1126/science.1144004.

Lo Piano, S. 2017. 'Quantitative Story-Telling and Sensitivity-Auditing Appraisal of Expert Solutions to the Issue of Food Security' [Presentation], Conference: Post-Normal Science Symposium PNS3—University of Tuebingen, doi: 10.13140/RG.2.2.36288.43525.

Lo Piano, S. and L. Benini. 2022. 'A Critical Perspective on Uncertainty Appraisal and Sensitivity Analysis in Life Cycle Assessment'. *Journal of Industrial Ecology*, *26*(*3*), 763–81, doi: 10.1111/jiec.13237.

Lo Piano, S., M. Janos Lorincz, A. Puy, S. Pye, A. Saltelli, S. Thor Smith, and J.P. van der Sluijs. 2022. *Unpacking the Modelling Process for Energy Policy Making* [Preprint], doi: 10.48550/arXiv. 2111.00782.

Lo Piano, S. and M. Robinson. 2019. 'Nutrition and Public Health Economic Evaluations under the Lenses of Post Normal Science'. *Futures*, *112*, 102436, doi: 10.1016/J.FUTURES.2019.06.008.

Mennicken, A. and W.N. Espeland. 2019. 'What's New with Numbers? Sociological Approaches to the Study of Quantification'. *Annual Review of Sociology*, *45*(*1*), 223–45, doi: 10.1146/annurev-soc–073117–041343.

Mennicken, A. and R. Salais, eds. 2022. *The New Politics of Numbers*, Palgrave Macmillan, London.

Micklewright, J. and S.V. Schnepf. 2006. *Inequality of Learning in Industrialised Countries*, 2517, Institute of Labor Economics (IZA), Oxford University Press, Oxford.

Micklewright, J., S.V. Schnepf, and C. Skinner. 2012. 'Non-Response Biases in Surveys of Schoolchildren: The Case of the English Programme for International Student Assessment (PISA) Samples'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(*4*), 915–38, doi: 10.1111/j.1467-985X.2012.01036.x.

Millgram, E. 2015. *The Great Endarkenment: Philosophy for an Age of Hyperspecialization*, Oxford University Press, Oxford.

Pan, H., B. Deal, G. Destouni, Y. Zhang, and Z. Kalantari. 2018. 'Sociohydrology Modeling for Complex Urban Environments in Support of Integrated Land and Water Resource Management Practices'. *Land Degradation and Development*, *29*(*10*), 3639–52, doi: 10.1002/ldr.3106.

Pielke, R. Jr. 2020. 'The Mudfight Over "Wild-Ass" Covid Numbers Is Pathological'. *Wired*, 22 April 2020 [Preprint]. https://www.wired.com/story/the-mudfight-over-wild-ass-covid-numbers-is-pathological (last accessed 20 March 2023).

Popp Berman, E. and D. Hirschman. 2018. 'The Sociology of Quantification: Where Are We Now?' *Contemporary Sociology*, *47*(*3*), 257–66.

Pouladi, P., A. Afshar, M.H. Afshar, A. Molajou, and H. Farahmand. 2019. 'Agent-Based Socio-Hydrological Modeling for Restoration of Urmia Lake: Application of Theory of Planned Behavior'. *Journal of Hydrology*, *576*, 736–48, doi: 10.1016/j.jhydrol.2019.06.080.

Puy, A., S. Lo Piano, and A. Saltelli. 2020. 'Current Models Underestimate Future Irrigated Areas'. *Geophysical Research Letters*, *47*(*8*), e2020GL087360, doi: 10.1029/2020GL087360.

Refsgaard, J.C., J.P. van der Sluijs, J. Brown, and P. van der Keur. 2006. 'A Framework for Dealing with Uncertainty Due to Model Structure Error'. *Advances in Water Resources*, *29*(*11*), 1586–97.

Rhodes, T. and K. Lancaster. 2020. 'Mathematical Models as Public Troubles in COVID-19 Infection Control: Following the Numbers'. *Health Sociology Review*, *29*, 1–18, doi: 10.1080/14461242.2020.1764376.

Roobavannan, M., T.H.M. van Emmerik, Y. Elshafei, J. Kandasamy, M.R. Sanderson, S. Vigneswaran, S. Pande, and M. Sivapalan. 2018. 'Norms and Values in Socio-hydrological Models'. *Hydrology and Earth System Sciences*, *22*(*2*), 1337–49, doi: 10.5194/hess-22-1337-2018.

Salais, R. 2010. 'La donnée n'est pas un donné. Pour une analyse critique de l'évaluation chiffrée de la performance'. *Revue française d'administration publique*, *135*(*3*), 497–515, doi: 0.3917/rfap.135.049.

Saltelli, A. 2018. 'Why Science's Crisis Should Not Become a Political Battling Ground'. *Futures*, *104*, 85–90.

Saltelli, A., K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu. 2019. 'Why So Many Published Sensitivity Analyses Are False: A Systematic Review of Sensitivity Analysis Practices'. *Environmental Modelling & Software*, *114*, 29–39, doi: 10.1016/J.ENVSOFT.2019.01.012.

Saltelli, A. and P. Annoni. 2010. 'How to Avoid a Perfunctory Sensitivity Analysis'. *Environmental Modelling & Software*, *25*(*12*), 1508–17, doi: 10.1016/j.envsoft.2010.04.012.

Saltelli, A., G. Bammer, I. Bruno, E. Charters, M. Di Fiore, E. Didier, W.N. Espeland, J. Kay, S. Lo Piano, D. Mayo, R. Pielke Jr, T. Portaluri, T.M. Porter, A. Puy, I. Rafols, J.R. Ravetz, E. Reinert, D. Sarewitz, P.B. Stark, A. Stirling, J. van der Sluijs, and P. Vineis. 2020a. 'Five Ways to Ensure That Models Serve Society: A Manifesto'. *Nature*, *582*(*7813*), 482–4, doi: 10.1038/d41586-020-01812-9.

Saltelli, A., L. Benini, S. Funtowicz, M. Giampietro, M. Kaiser, E.S. Reinert, and J.P. van der Sluijs. 2020b. 'The Technique Is Never Neutral. How Methodological Choices Condition the Generation of Narratives for Sustainability'. *Environmental Science and Policy*, *106*, 87–98.

Saltelli, A. and B. D'Hombres. 2010. 'Sensitivity Analysis Didn't Help. A Practitioner's Critique of the Stern Review'. *Global Environmental Change*, *20*(*2*), 298–302, doi: 10.1016/j.gloenvcha.2009.12. 003.

Saltelli, A., D.J. Dankel, M. Di Fiore, N. Holland, and M. Pigeon. 2022. 'Science, the Endless Frontier of Regulatory Capture'. *Futures*, *135*, 102860, doi: 10.1016/j.futures.2021.102860.

Saltelli, A. and S. Funtowicz. 2014. 'When All Models Are Wrong'. *Issues in Science and Technology*, *30*(*2*), 79–85.

Saltelli, A. and S. Lo Piano. 2017. 'Problematic Quantifications: A Critical Appraisal of Scenario Making for a Global "Sustainable" Food Production'. *Food Ethics*, *1*(*2*), 173–9.

SAPEA. 2018. 'SAPEA Participates in Scientific Integrity Conference', *SAPEA*, 20 February 2018, https://sapea.info/sapea-participates-in-scientific-integrity-conference/ (last accessed 14 April 2023).

Scoones, I. and A. Stirling. 2020. *The Politics of Uncertainty*, edited by I. Scoones and A. Stirling, Routledge, Abingdon, New York, doi: 10.4324/9781003023845.

Shafiee, E. and E.Z. Berglund. 2013. 'An Agent-Based Modeling Framework for Sociotechnical Simulation of Water Distribution Contamination Events'. *Journal of Hydroinformatics*, *15*(*3*), 862–80, doi: 10.2166/hydro.2013.158.

Shapin, S. and S. Schaffer. 2011. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*, Princeton Classics, Princeton University Press, Princeton, NJ.

Sheikholeslami, R., S. Razavi, H.V. Gupta, W. Becker, and A. Haghnegahdar. 2019. 'Global Sensitivity Analysis for High-Dimensional Problems: How to Objectively Group Factors and Measure Robustness and Convergence While Reducing Computational Cost'. *Environmental Modelling & Software*, *111*, 282–99, doi: 10.1016/j.envsoft.2018.09.002.

Sivapalan, M. 2015. 'Debates—Perspectives on Socio-Hydrology: Changing Water Systems and the "Tyranny of Small Problems"—Socio-Hydrology'. *Water Resources Research*, *51*(*6*), 4795–805, doi: 10.1002/2015WR017080.

Sivapalan, M., H.H.G. Savenije, and G. Blöschl. 2012a. 'Socio-Hydrology: A New Science of People and Water'. *Hydrological Processes*, *26*(*8*), 1270–6, doi: 10.1002/hyp.8426.

Sivapalan, M., H.H.G. Savenije, and G. Blöschl. 2012b. 'Socio-Hydrology: A New Science of People and Water'. *Hydrological Processes*, *26*(*8*), 1270–6, doi: 10.1002/hyp.8426.

Srinivasan, V. 2015. 'Reimagining the Past – Use of Counterfactual Trajectories in Socio-Hydrological Modelling: The Case of Chennai, India'. *Hydrology and Earth System Sciences*, *19*(*2*), 785–801, doi: 10.5194/hess-19-785-2015.

Thévenot, L. 2022. 'A New Calculable Global World in the Making: Governing through Transnational Certification Standards'. In *The New Politics of Numbers: Utopia, Evidence and Democracy*, edited by A. Mennicken and R. Salais, Executive Politics and Governance, Springer International Publishing, Cham, pp. 197–252, doi: 10.1007/978-3-030-78201-6_7.

United Nations. 2021. 'Transforming Our World: The 2030 Agenda for Sustainable Development', Department of Economic and Social Affairs. https://sdgs.un.org/2030agenda (last accessed 14 April 2023).

van der Sluijs, J.P., M. Craye, S. Funtowicz, P. Kloprogge, J.R. Ravetz, and J. Risbey. May 2005. 'Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System'. *Risk Analysis*, *25*(*2*), 481–92, doi: 10.1111/j.1539-6924.2005.00604.x.

van Zwanenberg, P. 2020. 'The Unravelling of Technocratic Orthodoxy'. In *The Politics of Uncertainty*, edited by I. Scoones and A. Stirling, Routledge, Abingdon, pp. 58–72.

Viglione, A., G. Di Baldassarre, L. Brandimarte, L. Kuil, G. Carr, J. Luis Salinas, A. Scolobig, and G. Blöschl. 2014. 'Insights from Socio-Hydrology Modelling on Dealing with Flood risk—Roles of Collective Memory, Risk-Taking Attitude and Trust'. *Journal of Hydrology*, *518*, 71–82, doi: 10.1016/j. jhydrol.2014.01.018.

Wheater, H. and P. Gober. 2013. 'Water Security in the Canadian Prairies: Science and Management Challenges'. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*, 20120409, doi: 10.1098/rsta.2012.0409.

Woessmann, L. 2016. 'The Economic Case for Education'. *Education Economics*, *24*(*1*), 3–32, doi:10.1080/09645292.2015.1059801.

Wuttke, J. 2007. *Uncertainty and Bias in PISA*, SSRN Scholarly Paper ID 1159042, Social Science Research Network, Rochester, NY.

Zhao, J., X. Cai, and Z. Wang. 2013. 'Comparing Administered and Market-Based Water Allocation Systems through a Consistent Agent-Based Modeling Framework'. *Journal of Environmental Management*, *123*, 120–30, doi: 10.1016/j.jenvman.2013.03.005.