Futures 144 (2022) 103041

Contents lists available at ScienceDirect

Futures

journal homepage: www.elsevier.com/locate/futures

# Unpacking the modelling process via sensitivity auditing

Samuele Lo Piano [a,*], Razi Sheikholeslami [b], Arnald Puy [c,d,e], Andrea Saltelli [f]

[a] School of the Built Environment, University of Reading, Reading, United Kingdom
[b] Department of Civil Engineering, Sharif University of Technology, Tehran, Iran
[c] School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom
[d] Department of Ecology and Evolutionary Biology, Princeton University, NJ, USA
[e] Centre for the Study of the Sciences and the Humanities (SVT), University of Bergen, Bergen, Norway
[f] Barcelona School of Management, Universitat Pompeu Fabra, Barcelona, Spain

ARTICLE INFO

*Keywords:*
Model quality assessment
Science-policy interface
Ethics of quantification
Sensitivity analysis
Uncertainty analysis
Transparency

ABSTRACT

Acknowledging the conditionality of model-based evidence facilitates the dialogue between model developers and model users, especially when models are used to guide decisions at the science-policy interface. In general, model users have limited access to verify the realism of a model, being only exposed to model plausibility and trustworthiness; instead, modellers have an an array of validation and verification techniques available. In the end, model credibility is what both developers and users aim for, also in the interest of shielding from the possible pitfall of over-interpreting the model results. To this end, in this contribution we discuss sensitivity auditing, an extension of sensitivity analysis, that can help model developers and users to overcome communication barriers and foster dialogue around modelling activities. The use of sensitivity auditing is not limited to models in a restricted sense, but it can be applied to any policy-relevant instance of quantification, including metrics, rankings and indicators. We present six real-world applications of sensitivity auditing to instances of quantification in a range of socio-environmental systems, including public health, education, and the water-food nexus. These examples reveal the usefulness of sensitivity auditing in facilitating the proper use of numbers and models at the science-policy-society interface and in avoiding uncertainty laundering.

## 1. Introduction

Society is increasingly exposed to evidence from mathematical models. As in the case of COVID-19, it has become apparent that the level of reciprocal domestication between models and society has scope for improvement (Saltelli, Bammer et al., 2020), lest models themselves become mired in controversy (Rhodes & Lancaster, 2020).

Models are typically controlled by large numbers of interacting factors. The construction of a model involves modelling choices and assumptions that tend to stratify over the lifetime of the model. This calls for the careful and continuous monitoring of uncertainty at all stages of model construction and operation, especially when models are called on to inform the adoption of policy decisions or the drafting of regulations (Pontius & Millones, 2011).

Uncertainties may be characterised using two distinct but related methods of analysis. The first is uncertainty analysis (UA) (Saltelli et al., 2008), wherein several uncertain model inputs are assigned a probability distribution that reflects their measurement error,

* Corresponding author.
*E-mail address:* s.lopiano@reading.ac.uk (S. Lo Piano).

https://doi.org/10.1016/j.futures.2022.103041
Received 29 June 2021; Received in revised form 20 July 2022; Accepted 27 September 2022
Available online 29 September 2022
0016-3287/

natural variation, inherent randomness, and/or the disagreement of experts. By sampling from these distributions and running the model repeatedly, it is possible to propagate these uncertainties and generate an empirical distribution function for the output(s) of interest. This is also known as an error propagation exercise or forward uncertainty analysis. The available methods for performing this analysis are often classified as *Monte Carlo methods.* Their purpose is to explore the space of the input factors and to generate a full spectrum of model behaviours.

A second useful tool is sensitivity analysis (SA) (Saltelli et al., 2008), in which the goal is to ascertain the most influential uncertain inputs. In other words, SA determines the sensitivity of the model's outputs to individual inputs (or groups of inputs) by systematically exploring their uncertainty space.

Both UA and SA imply a degree of discretion and judgement. This is especially the case when setting the boundary of the modelled system in terms of choosing which physical and/or social processes to include, what temporal resolution and/or spatial scale to adopt, and so on. Additionally, performing UA and SA requires choices as to what inputs are to be taken as uncertain, the selection of the probability distributions to describe the input uncertainty, and the question of what type of sensitivity (i.e., importance) measure to use.

The description so far has been of a 'technical' nature. However, the choices described might be conditioned by individual bias – whether disciplinary or normative – or by plain interests. It is not difficult to imagine a situation in which a model is used to assist in a decision related to policy or regulation and where different actors have divergent visions of what should be modelled and how. In other words, mathematical models reflect the cultural and normative biases of their developers. As a result, the entire modelling process and its conclusions, including its technical UA and SA, can be deconstructed by first revealing and then contesting the modeller's choices and views. These may include the model's purpose and the frames adopted.

The value of deconstructing the modelling process is especially significant in situations characterised by high stakes, urgent decisions, uncertain facts, and contrasting values. This is the realm of Post-normal Science (PNS) (Funtowicz & Ravetz, 1993, 1994; Ravetz, 1997) for settings where the use of science falls beyond the traditional puzzle-solving domain of normal science (Kuhn, 1962).

The key ingredients of PNS, which are especially valuable in relation to forms of quantification of various natures, are:

- The non-separability of facts from values in the framing of the problem;
- The consideration of the position to the observer (i.e., an invitation to reflexivity);
- The reference to an extended peer community, intended as deliberative, contributing both disciplinary and non-disciplinary knowing to the solution of a problem. This community may include academics as well as whistleblowers, investigative journalists, and lay citizens involved in or interested by the problem at hand.

That is to say, technical SA is never the end of the story when the model must be audited or scrutinised; something more is needed. We propose here *sensitivity auditing*, an approach to extend uncertainty and sensitivity analyses to PNS-like settings (Saltelli et al., 2013). The name *sensitivity auditing* reflects the ambition to render the model accountable to an audience beyond its creators (Saltelli & Funtowicz, 2014).

The audience for the approach would ideally consist of the extended peer community that has a stake or interest in the issue being quantified. Here, sensitivity auditing would be useful to negotiate the nature of the problem, the framing of the story being told by a quantification, and – in general – its underpinning technical and political assumptions. It is not rare that in a conflicted issue, in what have been termed 'wicked problems' (Rittel & Webber, 1973), what constitutes a solution for a party is the problem for the opposing one.

## 2. Sensitivity auditing

Sensitivity auditing has been recommended in guidelines for impact assessment, including those of the Science Advice for Policy by European Academies (2019) and those revised by the European Commission (2021).

Sensitivity auditing is based on a seven-point checklist shown as follows:

- **Rhetorical use:** Check against a rhetorical use of mathematics – Are large models being used where simpler ones would suffice? Are model results and scope extrapolated beyond their intended range/settings of applicability?
- **Assumption hunting:** What assumptions were made? Were these explicit or implicit?
- **Detect Garbage In, Garbage Out (GIGO):** Was the uncertainty in the input artificially constrained to boost the model's certainty? Or, conversely, was it bloated so as to, for example, prevent regulation in a case of harmful products?
- **Anticipate criticism:** Find sensitive assumptions before they find you – It is better to anticipate criticism by undertaking robust uncertainty and sensitivity analyses before publishing one's results.
- **Aim for transparency:** Black box models do not play well within a public debate.
- **Do the right sums:** Do the right sums, not just the sums right - Is the issue properly identified or does the model address the 'wrong' problem (or a closed definition of what the problem might be), instead of including multiple perspectives?
- **Perform UA, SA:** Perform thorough and state-of-the-art uncertainty and sensitivity analyses.

All parties in a dispute in the context of conflicting scientific evidence could in principle use sensitivity auditing. For instance, to build a defensible, plausible model. An opposing party could use the approach instead to demonstrate the irrelevance, bias or strategic use of a model-based inference to serve one own's agenda produced by the 'antagonist'. The valuable aspect of this situation is that all parties need to engage with the evidence suggested by their opponents instead of just dismissing it.

In this contribution, we present for the first time a systematic mapping of sensitivity auditing onto other theoretical frameworks recently proposed, as well as positioning the methodology into the emerging field of ethics of quantification.

The tenets of sensitivity auditing were stressed in a recent manifesto, *Five ways to ensure that models serve society* (Saltelli, Bammer et al., 2020). In the document, the modelling of the COVID-19 pandemic was taken as the point of departure to discuss the social nature of modelling and the need for better societal negotiation of what models can achieve. Sensitivity auditing has also an affinity with the Numeral Unit Spread Assessment Pedigree (NUSAP) system in evaluating the quality of quantitative information (Funtowicz & Ravetz, 1990; van der Sluijs et al., 2005), as well as the technologies of humility of Jasanoff (2007) (Table 1).

The next section illustrates selected instances of quantification (models, indicators, and metrics) where sensitivity auditing was applied. Metrics, ranks and indicators are often presented as crisp numbers, even when based on sophisticated modelling activities, and built under the assumption that they can capture complex socio-economic phenomena. For this reason, the application of sensitivity auditing to these forms of quantification is relevant, especially when the numbers form the basis for policies and decisions.

We chose six case studies from various socio-environmental fields to encourage the uptake of sensitivity auditing across domains having different modelling practices. In this way we aim to raise interest from a wider audience corresponding to different schools and cultures of modelling. Focusing on a single model/modelling activity would certainly offer the advantage of a more detailed close up on the approach, yet readers outside the specific domain investigated would miss the potential of sensitivity auditing.

To offer an idea of the versatility of the approach, selected examples are given relevant to policymaking and/or that have received vast media coverage. The pool of instances of quantification covered ranges from modelling practices in terms of several modelling activities within a particular domain to single indicators/ranks. We apply systematically all rules to all cases by expanding on the findings of the previous publications in which we have deconstructed modelling activities from our research field or through collaborations with modellers from those fields. This obeys a didactic function: in real life, the rules will need to meet the demands of the users; additionally the rules are - in our experience - to be applied recursively.

## 3. Case studies of quantifications on socio-environmental systems

In this section, we present six examples for investigation (Table 2) with the sensitivity auditing seven-point checklist. The two cases from the domain of hydrology (sociohydrology and global hydrological models) are new. These examples, which come together under the umbrella term of socio-environmental systems, were typically designed to address major societally relevant questions, including the following: How much of the biological capacity of the planet is required by a given human activity or population? What are the impacts of future hydro-meteorological conditions on water availability? How will the hydrological cycle respond and evolve over time under a human-dominated biosphere? What are the available agro-ecological strategies to improve rural livelihood and food security? The inherent complexity of these sustainability-related questions, together with severe uncertainties imposed by anthropogenic socio-economic and climatic changes, make these case studies suitable candidates for performing sensitivity auditing. A brief description of the case studies is given in the rest of this section.

- *Nutrition and public health economic evaluations*: Lifestyle habits such as diet, smoking, and physical activity level are epidemiologically linked to non-communicable diseases (NCD). For this reason, it is important to evaluate how policies aimed at triggering changes in these factors may produce societal consequences in terms of disease likelihood. The policies are underpinned by the available body of evidence, in which modelling activity plays a key role (Lo Piano & Robinson, 2019). Modelling activities are

**Table 1**
Comparison of sensitivity auditing with the manifesto for responsible modelling (Saltelli, Bammer et al., 2020) and Jasanoff's technologies of humility (Jasanoff, 2007).

| Sensitivity auditing | Manifesto for responsible modelling | Jasanoff's technologies of humility |
|---|---|---|
| Rhetorical use of models | Modelling hubris | Technologies of hubris (including risk assessment, cost-benefit analysis, and climate modelling) |
| Assumption hunting | Mind the assumptions | 'Predictive technologies are limited in their capacity to internalise challenges that arise outside their framing assumptions' |
| Inflating or deflating uncertainty | Mind the unknown | 'These technologies show peripheral blindness towards uncertainty and ambiguity' |
| Anticipate criticism | – | – |
| Transparency | 'Predictions need to be transparent and humble' | Transparency implies participation |
| Mind the framing | Mind the framing | Mind the framing.<br>Reflect on vulnerabilities.<br>Reflect on winners and losers.<br>Reflect on learning opportunities. |
| Perform proper sensitivity analysis | '…perform global uncertainty and sensitivity analyses … allowing all that is uncertain – variables, mathematical relationships, and boundary conditions – to vary simultaneously as runs of the model produce its range of predictions' | – |

**Table 2**
Example applications of quantification sensitivity auditing. The name of the quantification has been spelled out when the study has focused on a single approach.

| Field, quantification name | Type of quantification | Model scope | Ref. |
|---|---|---|---|
| Nutrition and Public Health | Several modelling activities | Produce economic evaluations and policy comparisons | Lo Piano and Robinson (2019) |
| Sustainability, Ecological Footprint (EF) | Indicator | Quantify and benchmark humanity's impact | Giampietro and Saltelli (2014a, b) |
| Global Hydrological Models | Several modelling activities | Inform policies on irrigated agriculture | Puy et al. (2022a) |
| Sociohydrology, Coupled Human and Water Systems (CHAWS) | Several modelling activities | Understand human-driven impacts on the hydrological cycle | Ghoreishi et al. (2021) |
| Food Security | Model | Showcase a sustainable food system for the year 2050 | Lo Piano (2017), Saltelli and Lo Piano (2017) |
| Education, Programme for International Student Assessment (PISA) test | Metrics | Evaluate the literacy level across countries | Araujo et al. (2017) |

aimed at assessing the effects of the exposures on stress factors and gauging how policies aimed at reducing these factors may trigger economic benefits.

- *Ecological footprint*: The EF is a successful sustainability indicator proposed by the *Global Footprint Network*. EF measures human demand on natural capital, which is understood as the quantity of natural land (expressed in *global hectares equivalents*) required to support a given local area, a specific country, or globally. The concept of *Earth overshoot day* is the dimension of EF that has received the widest press coverage. This quantity is assessed annually and represents the date by which humanity is expected to have used all available natural resources from the Earth's yearly natural budget. The systematic anticipation of the yearly *Earth overshoot day* is widely recognised as a sign of humanity's unsustainable pattern of economic development.
- *Global estimates of irrigation water requirements*: The impact of humanity on Earth's water resources is a crucial dimension of sustainability, with the largest water withdrawals typically allocated to irrigate agricultural lands. The progressive increase in computing power and the growing awareness of humanity's impact on global water resources led to the creation of the first global hydrological models at the end of the 1980s (Bierkens, 2015). Nowadays, several global models produce spatially-distributed estimates of irrigation water requirements (Doll and Siebert (2002), Hanasaki et al. (2008), Sutanudjaja et al. (2018)). Regulatory agencies such as the United Nations (UN), the World Bank, and the IPCC use these models to inform policies on water use, climate change, and food security (Veldkamp et al., 2018).
- *Sociohydrology*: The field of sociohydrology emerged in response to the perceived failure of the traditional paradigm in water resource management, which regarded human activity merely as a boundary condition of the hydrologic system (Sivapalan et al., 2012). Sociohydrology seeks a more thorough understanding of how the hydrological cycle changes according to interactions among natural and human forces. CHAWS models play a prominent role in understanding the complex water-related problems that human societies are currently facing. The key features of these models are complexity, cross-scale dynamics, and uncertainty (Liu et al., 2007; Sivapalan, 2015; Sivapalan et al., 2012; Wheater & Gober, 2013).
- *Food security*: Food security is concerned with the global system's capacity to meet the nutritional needs of a growing global population. The UN established a *Zero Hunger Strategy* cutting across several *Sustainable Development Goals*, as identified in the agenda for 2030 (United Nations). We focus on a work by Bahadur et al. (2016) wherein a trajectory was proposed for achieving healthy and sustainable food provision in terms of better agricultural techniques and dietary readaptation.
- *PISA*: The PISA test was designed by the Organisation for Economic Co-operation and Development (OECD) to measure the problem-solving skills of fifteen-year-old school pupils (https://www.oecd.org/pisa/). The test has been run every three years since 2000, producing periodic country rankings. 79 countries participated in the 2018 round of tests. Criticisms of the methodological and ideological stances of the test, among other aspects, were raised with the publication of an open letter in *The Guardian* in 2014. In the course of the dispute, academics and non-governmental organisations called for a moratorium on and rethinking of the test (Araujo et al., 2017).

## 4. What aspects emerge by applying the sensitivity auditing checklist?

In this section, we offer key highlights from applying the seven rules to the six case studies.

- *Rhetorical use:* Using sensitivity auditing, Lo Piano and Robinson (2019) concluded that in the field of nutrition and public health economic evaluations, most authors adopt Markov chain models despite the availability of less-computational demanding representations (Clarke et al., 2005).

The same issue emerges in the domain of global hydrological models, which are based on large-scale algorithms including a plurality of sub-models and dozens of factors, pushing models at the cutting-edge of computer power (Smith, 2015; Sood & Smakhtin, 2015). Sensitivity auditing has shown that the irrigation water withdrawal values that these models output can be nicely approximated using a simple linear regression with the extension of irrigation as the only predictor variable (Puy, Borgonovo, Lo Piano, Levin, & Saltelli, 2021). An excess of model detail also dominates the field of sociohydrology, where several factors/components are usually included in the structure of models to capture human reactions to hydrological variability. As many aspects of hydro-social processes are unknown or contradictory, and universally accepted laws for human behaviour related to economic growth, politics, history, culture, and other factors do not exist, modellers often tend to compensate for their lack of knowledge by adding more and more factors for which scarce evidence is available (Levy et al., 2016; Sivapalan & Blöschl, 2015).

An interesting case of overinterpretation of a quantification is offered by the EF accounting scheme (Galli et al., 2016; Giampietro & Saltelli, 2014a). What is presented as a measure of the planet's ecological limits is merely a balancing of $CO_2$ emissions with available land acting as a sink, on which considerable disagreement exists (Galli et al., 2016; Giampietro & Saltelli, 2014a).

Prevailing narratives interpret the PISA test by suggesting a causal relationship between test scores and economic growth. This is, for instance, the case in a document from the European Commission, which states that if the countries of the European Union (EU) could significantly increase their PISA scores, this would directly lead to a quantifiable growth in Gross Domestic Product (GDP) (Woessmann, 2016).

- *Assumption hunting:* In the field of nutrition and public health economic evaluations, Lo Piano and Robinson (2019) noted a series of unexplored assumptions related to:
  - The modelling of the dose-responses adopted in terms of risk factor estimates;
  - The dose-intake variation upon policy implementation and how this would affect different socio-demographic cohorts and geographical areas;
  - The timeframe of the interventions and their diminishing/increasing returns;
  - The actual pool of NCDs taken into account in the modelling exercise and their consistency with the investigated timeframe.

These unexplored assumptions may result in a biased picture of the space of the policy options, which may lead to the selection of sub-optimal policies or unintended effects.

In the EF assessment (Giampietro & Saltelli, 2014a), unexplored assumptions include the issue of neglecting the decrease in $CO_2$ absorbing capacity with the ageing of forests, or the paradox that replacing natural ecosystems with more productive human-made vegetation would lead to an improvement in the planet's biocapacity rather than impoverishment due to loss of biodiversity and natural habitats.

In the food security assessment (Saltelli & Lo Piano, 2017), available projections to 2050 neglected the diminishing return phenomenon for yields when increasing the cultivated area. The authors also made several claims concerning the possibility of reducing the extension of globally cultivated land thanks to higher yields, while neglecting various forms of pressure on the ecosystem. The joint effect of these assumptions is at odds with the plausibility of achieving food security in the year 2050 by cultivating a lower global land extension.

Regarding hydrology, sensitivity auditing led to the conclusion that global hydrological models are implicitly grounded in an engineering vision of irrigation, wherein irrigation water withdrawals are defined by quantifiable physical and biological processes. This vision ignores farmers' practices and know-how, which significantly influence crop water demand and agricultural water management (Puy et al., 2022a). Even CHAWS models are based on 'black-box' assumptions (Roobavannan et al., 2018) regarding human values, beliefs, livelihoods, and the environment, which are believed to drive human behaviour with respect to water resources usage.

In terms of political assumptions, the PISA test builds on the postulate that it is possible to benchmark the skills students need in order to succeed in present-day 'knowledge societies' against a one-size-fits-all international standard – typically literacy, math, and science. Other important subjects, as well as differences in curricula across countries, are excluded from the test. However, countries' well-being and success may emerge from these very curricula differences, and all countries cannot be assumed to necessarily pursue a knowledge economy trajectory (Araujo et al., 2017). Failure to capture this aspect could hamper the capacity to design and promote an education system that acknowledges the values and identities of countries.

- *Detect GIGO:* Some studies in the field of nutrition and public health made use of artificially constrained figures in terms of the uncertainty in nutrient intake (Lo Piano & Robinson, 2019).

Even in the EF accounting, no error in terms of biocapacity is considered, nor is the accuracy of the variable discussed at the local, national, and global levels. A data quality score is the only proxy included at the country level.

In the case of CHAWS, sensitivity auditing notes that underplaying uncertainties in human-water systems causes considerable biases in the projected future changes and acceptable decisions. For example, in flood risk assessment, the underestimation of future flood damage can severely limit the economic development of an area due to short collective memory, excessive trust in flood protection structure, and a modelled high risk-taking attitude (Viglione et al., 2014). As another example, Zarekarizi et al. (2020) showed that ignoring deep uncertainties surrounding flood hazards, economic factors, and the house lifetime can drastically change the 'optimal building elevation' to meet the US Federal Emergency Management Agency regulations for mitigation decisions.

In the food security assessment, the figures for the reduction of the global cultivated land area are produced with three significant digits, estimated at 438 million hectares for 2050 (Badur et al., 2016). However, uncertainty figures in the current yearly global cultivated land extensions are around 20% (i.e., 1000 million hectares (Lo Piano, 2017)).

When communicating the PISA test results, the survey organisers report only the standard error of the countries' test scores. Other factors of uncertainty that could significantly contribute to volatility in country rankings are neglected. This is the case, for instance, in the potential bias arising from the exclusion of students with special educational needs or newly arrived immigrants. This has led to the figures for some countries in certain editions of the test falling below the representativeness threshold imposed by the test designers as an assurance of its reliability (Wuttke, 2007). The same issue may emerge from the tendency of the less capable students to refrain from participating in the test, which may eventually result in a significant bias between the 'actual' and 'estimated' scores of a country, well beyond what is captured in the standard error of the country's rank (Micklewright et al., 2012).

- *Anticipate criticism:* In the field of nutrition and public health evaluation, the methods used to impute missing data regarding dose intake (i.e., using other countries' figures or running correcting algorithms) were not tested using UA or SA.

Even in the case of PISA, calls from practitioners have remained unheard to test the sensitivity of the rankings and their volatility to modelling assumptions, data collection methods, and the use of the data items (Micklewright & Schnepf, 2006). Leaving uncertainty unaddressed is still a common practice, as in the case of the EF. The same holds for the 2050 assessment of food security, where reporting on the assessment merely specifies a single number and, in this way, overlooks its possible range of variability and its causes.

- *Aim for transparency:* Variable levels of transparency are observable in the case studies. For instance, the global irrigation water withdrawal estimates outputted by various large-scale models are available at the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) (Warszawski et al., 2014). Some of these models are open-source, where the code is available in existing repositories (e.g., GitHub LPJmL 4.0, VIC, PCR-GLOBWB), whereas others are still in the process of opening up (e.g., WaterGap (Müller Schmied et al., 2020)). However, the underpinning code is rarely commented and the repositories do not include a user manual (Kauffeldt et al., 2016). Even CHAWS models are typically not open-access or they are made only available for limited case studies (Pouladi et al., 2019; Shafiee & Zechman, 2013; Zhao et al., 2013), often with wanting documentation.

In the case of the EF, the documentation on the accounting is available but relevant technical coefficients are not openly traceable. This is the case for the *equivalence factors*, which reflect the relative productivity of world average hectares of different types of land use. The question of how these quantities were arrived at can only be answered by using a satellite workbook of the *footprint accounts*, which is accessible upon request; as such, the resources are not directly available in an open repository. The same applies to the model that underpins the findings of the food security study presented in Badur et al. (2016).

In the field of nutrition and public health evaluation, a notable example is that of Joint Action on Nutrition and Physical Activity (JANPA), whereby a model has been used to explore policy impacts and compare the growing prevalence of overweight and obesity in children. The model is privately owned and has not been made available for scrutiny by the public and other modellers.

- *Do the right sum:* Flattening health evaluations along a mere economic dimension runs the risk of overlooking the social and cultural aspects of lifestyle and nutrition, which – in reality – may be the most relevant aspects for citizens when making choices. The notion of health and quality of life is to be captured as per the normative dimensions of citizens' values, which are not necessarily represented by monetary proxies.

Even for global hydrology modelling, most studies do not address the geopolitics of water resources nor important policy dimensions such as resource ownership (public vs. private) and access. Emphasising efficiency and productivity hampers the perception of significant problems of inequalities in access to resources. The concentration of substantial data and computational resources within international organisations may exacerbate North-South power asymmetries and foster data colonialism (Thatcher et al., 2016).

In the case of food security, the analysis proposed is framed from a developed-world perspective, primarily to be tackled with technical solutions and a policy package aimed at addressing the issues in this area of the world. However, the political issue of power asymmetry in the international food commodity trade (e.g., in terms of unequal caloric exchange in food crops (Falconi et al., 2017)), as well as how this affects developing countries, are left entirely unexplored in the study. In other words, a political problem has been reframed into a technical one, while privileging the view of a minority.

The risk of addressing the wrong question has also been raised for CHAWS that focus on the question of "What is the most likely future?" rather than "What kind of future do we want and what are the consequences of different policy decisions relative to that desired future?" (Gober & Wheater, 2015). Closer collaborations among sociohydrologists, practitioners, and other real-world actors have been advocated to overcome this issue (Lang et al., 2012; Rokaya et al., 2017).

The PISA test was conceived under the assumption that it would facilitate measurement of the effectiveness of students' preparation. This is conceptualised as the degree to which the teaching that students have received is useful regarding the life challenges they will encounter in today's knowledge societies. The test *de facto* makes the 'economic' case for education, which is exclusively presented and discussed as a means for economic growth rather than for *Bildung*, self-cultivation, and emancipation. Incidentally, it is possible to see the same narrative underpinning PISA at play in the pervasive rankings of universities, which have been instrumental in creating a global market for higher education.

- *Perform UA, SA:* We identified several levels of characterisation of uncertainty in the studies considered. These ranged from an almost full omission (in the case of the EF, the food security assessment, and the PISA test scores) to various levels of apportioning. In the field of nutrition and public health economic evaluations, the vast majority of the studies implemented SA by varying one-factor-at-a-time (OAT). However, the approach is wanting for non-additive models because it fails to capture interactions across factors and leaves the vast portion of the output uncertainty space unexplored (Saltelli & Annoni, 2010).

In global hydrological models, it is also notable that previous research have mostly assessed parametric and structural uncertainties in a piecewise manner. This has involved examining the effect derived from changing specific structures (e.g., evapotranspiration equation (Vörösmarty et al., 1998) and climate forcing (Müller Schmied et al., 2014)), mainly through OAT or other non-global techniques. Another adopted sub-optimal strategy has been to explore uncertainties through model ensembles, as in Schewe et al. (2014) or Wada et al. (2013), which do not sample all model formulations (Parker, 2013). A limitation of model ensembles is that they do not systematically sample the uncertainty space (Puy et al., 2022a). It is also difficult to know whether the convergence of the ensemble is due to the robustness of the estimation or the failure to capture a significant source of uncertainty.

Better practices are found in the domain of sociohydrology, where uncertainties related to the social aspect of CHAWS (e.g., demography, socio-economic development, and agent behaviour), as well as those in the hydrologic environment (e.g., non-stationarity deriving from anthropogenic factors such as changes in land use, climate, and water use), have been simultaneously addressed. Yet only two studies ((Elshafei et al., 2016; Ghoreishi et al., 2021)) performed a global SA by simultaneously investigating CHAWS' sensitivity to those model factors that (i) govern the internal dynamics of the system, and (ii) determine the external sociopolitical context.

## 5. Discussion and conclusions

To what extent did sensitivity auditing help to make sense of the various quantifications discussed in the present work? How can it contribute to better use of mathematical models, or to the process of reciprocal domestication between models and society?

An important element of this analysis is *the scope of a quantification*. It is possible to extrapolate models and statistical indicators alike beyond the function they were originally conceived for. Sometimes, they may become the very target of policies, like in the case of the PISA test. In so doing, they cease to serve as good metrics (Goodhart, 1981). The same applies to mathematical models when the model replaces the modelled system as the *locus* of attention (Rayner, 2012). For example, if EF is taken at face value, ecologically unsound measures may be adopted by replacing natural ecosystems with more productive human-made vegetation. As noted in Page (2018), using a single model may correspond to hubris; moreover, different forms of quantification, as well as qualitative analyses, may better protect against the risk of poor decisions.

Another important element that has emerged from our analysis is the *balancing of the level of complexity and the model's purpose*. As far as mathematical models are concerned, the examples discussed here indicate a risk of excess in model details, as discussed for global hydrological models (Puy, Borgonovo, Lo Piano, Levin, & Saltelli, 2021) and models used in nutrition and public health economic evaluations (Lo Piano & Robinson, 2019). Developers often pursue complexity in the hope that the model will eventually be 'right', neglecting the trade-off between model complexity and propagation error, which is a phenomenon known as the uncertainty cascade (Christie et al., 2011) or O'Neil's conjecture (O'Neill, 1989). At the same time, a complex model is perceived as conferring epistemic authority. However, the appropriate level of model complexity depends on the model's purpose and function, such as whether it aims to understand dynamics, manage resources, predict variables, or inform policy-making (Puy et al., 2022b).

A trade-off that modellers often face is *exploring the model uncertainty space versus adding complexity*. Leaving unexplored assumptions may correspond to an instrumental attempt to save the model from irrelevance; after all, no market exists for a model whose output is too uncertain to inform a policy direction. Additionally, model overcomplexification may be used as a justification to leave the model uncertainty space unexplored, such as when one may already have used all the computational power available and more simulations may be unaffordable. As such, decision-makers could be presented with spuriously accurate figures that narrow the possible course of action, limiting the policy options to a single or restricted pool of outcomes. Our recipe here is to allocate resources to explore the uncertainty space of assumptions, especially when attempting to develop a model in a direction of greater detail and complexity. Increasing complexity without simultaneously exploring the implications of this addition results in unexplored complexity, which is ultimately detrimental to trust in the modelling work.

For instance, Sheikholeslami et al. (2019) showed that for a complex coupled land-surface hydrology model developed by Environment and Climate Change Canada (MESH) (Pietroniro et al., 2007), the model output variability was primarily determined by 4 of the 100 input factors.

There is a substantial amount to gain by heeding *practices for transparent quantification*. The ingredients of *open science* (Fecher & Friesike, 2014), which include open models, data, and code, are welcome foundations for strategies to alleviate the problems highlighted by our sensitivity auditing. Progress has indeed been made, for instance, in some global hydrological models (Kauffeldt et al., 2016). However, existing initiatives for open and responsible modelling need to gain traction. On the same grounds that research institutions and journals now consider pre-registration and availability of data mandatory, this should be extended to properly commented source code (Walters, 2020).

*Assumptions and model framing* also represent a crucial issue. Tacit assumptions may be at the basis of a model, as discussed in the case of food security. The methodological and normative stances on which these assumptions rest may appear obvious or 'standard' to the modellers, but heretical in the eyes of an observer with a different orientation. By bringing them under the spotlight, sensitivity auditing allows to disclose what choices of inclusion and omission a modelling activity is based on, i.e. what it is assumed in and what is assumed out in the representation, which relations of causality have been proposed and their rationale. The impact of some of these can quantitatively be assessed through uncertainty and sensitivity analyses (rule 4 and rule 7 of the checklist). Transparency (rule 6) about both quantifiable and unquantifiable assumptions is also essential to promote scrutiny and reflexivity from the extended peer community.

As noted by Beck, "*It is not uncommon for political programmes to be decided in advance simply by the choice of what expert representatives are included in the circle of advisers*" (Beck, 1992). Accepting that the technique is never neutral (Saltelli, Benini et al., 2020) implies that spelling out the adopted standpoint to all interested parties is necessary for the healthy use of quantification.

Finally, what role can modelling play in the context of *extended peer communities?*. Knowledge and model quality appraisal tools have proven to be quite effective to shed light on modelling studies and practices. An example not discussed in the present work but relevant to our discussion is that of extended peer communities in action, as investigated in Laes et al. (2011). Here, the community involved in the *ExternE* project attempted to estimate the external costs of a potential large-scale nuclear accident at a Belgian reactor. In the course of a veritable 'hunt for assumptions', the workshop participants identified as many as 30 assumptions underpinning the calculations. The methodology adopted was the NUSAP scheme for knowledge-quality assessment (van der Sluijs et al., 2005), which allowed the experts and stakeholders to bring to the fore the implicit meanings and value-laden nature of the assumptions underpinning the proposed quantification. The participants also managed to identify the six most critical calculation steps and concluded that most of them were questionable, and ultimately unreliable to the effect of producing a plausible estimate of the cost of an accident.

Making models and their use more responsible can serve as an important element of an overall *ethics of quantification* (Saltelli & Di Fiore, 2020; Saltelli et al., 2021), both for visible numbers (mathematical and statistical models, rating, and ranking) and invisible ones (e.g., deep learning algorithms used to make decisions in business, policy, and everyday life (Lo Piano, 2020)). Recent studies on the sociology of quantification (Mennicken & Espeland, 2019; Popp Berman & Hirschman, 2018) offer rich material for modellers to consider in their work. Established good practices or guidelines (see the *Supplementary Material* for a review) are not lacking in the field of quantification. The problem with modelling proper is that it is not an established discipline, and different academic communities pursue it in their own separate ways (Saltelli, 2019a). Thus, the dialogue among modellers is less structured than that among, for example, statisticians debating significance or official statisticians with their codes of good practice.

Many researchers have insisted – as we do in sensitivity auditing – on the need to pay attention to context and purpose, and to revalidate models whenever they change or are applied to a different case (Badham et al., 2019; Eker et al., 2018; Hamilton et al., 2019; Jakeman et al., 2006; Little et al., 2019; Padilla et al., 2018). Several prior works recommend that modellers should heed the advice from the social sciences, but two problems arise here.

One problem is the existence of a crisis in scientific practice (see Saltelli and Funtowicz (2017) or references in Saltelli, Bammer et al. (2020)[supplementary]). The crisis has led some authors to discuss a "Darwinian fitness" of bad science (Smaldino & McElreath, 2016), which is a comment also voiced by statisticians combating malpractices such as P-hacking and HARKing (Kerr, 1998). The same types of malpractice are also present in mathematical modelling (Rhodes & Lancaster, 2020), which are aggravated by the lack of disciplinary oversight mentioned above (Saltelli, 2019b).

A second problem is that the two great families of science – the natural and social sciences – are still scarcely permeable relative to one another. For Crowe (1969), there are two "insular scientific communities – the natural and the social – between which there is very little communication and a great deal of envy, suspicion, disdain, and competition for scarce resources". As a result, according to the sociologist of science Pinch (1992), social studies appear to have little traction on the disciplines they study; the sociology of science and the scientific disciplines themselves do not communicate as much as they could.

For these reasons, it is possible that the adoption of approaches such as sensitivity auditing will be driven more by necessity – the danger of a collapse of trust in numbers – than by a voluntary movement of all modelling communities. We see the structured use of sensitivity auditing at the stage of model building, where it is applied to anticipate criticism, as a promising route in the direction of a positive change for responsible modelling. Sensitivity auditing could serve as a powerful analytical lens in the present blossoming of forms of activism to scrutinise different instances of quantification (see Supporting Materials), as reflected by the examples discussed in this paper.

## Conflicts of interest

The authors declare no conflicts of interest.

## Acknowledgements

This work has been funded by the European Commission (Marie Sklodowska-Curie Global Fellowship to Arnald Puy, Grant number 792178).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.futures.2022.103041.

## References

Araujo, L., Saltelli, A., & Schnepf, S. V. (2017). Do PISA data justify PISA-based education policy? *International Journal of Comparative Education and Development, 19* (1), 20–34. https://doi.org/10.1108/IJCED-12-2016-0023

Badham, J., Elsawah, S., Guillaume, J. H., Hamilton, S. H., Hunt, R. J., Jakeman, A. J., Pierce, S. A., Snow, V. O., Babbar-Sebens, M., Fu, B., Gober, P., Hill, M. C., Iwanaga, T., Loucks, D. P., Merritt, W. S., Peckham, S. D., Richmond, A. K., Zare, F., Ames, D., & Bammer, G. (2019). Effective modeling for Integrated water resource management: A guide to contextual practices by phases and steps and future opportunities. *Environmental Modelling and Software, 116*, 40–56. https://doi.org/10.1016/j.envsoft.2019.02.013

Badur, K. K., Fraser, E. D., Pascoal, S., Dias, G., & Zundel, T. (2016). Pathways leading to a more sustainable and healthy global food system. *The Solutions Journal*. https://www.thesolutionsjournal.com/article/pathways-leading-sustainable-healthy-global-food-system/.

Bahadur KC, K., Fraser, E., Dias, G., Zundel, T., & Pascoal, S. (2016). Pathways leading to a more sustainable and healthy global food system. *The Solutions Journal, 7*(5), 10–12.

Beck, P. U. (1992). *Risk society: Towards a new modernity*. Published in Association with Theory, Culture & Society, Sage Publications (CA).

Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. arXiv: 2014WR016527[10.1002] ISBN: 1944-7973 tex.arxivid: 10.1002/2014WR016527 *Water Resources Research, 51*(7), 4923–4947. https://doi.org/10.1002/2015WR017173.

Christie, M., Cliffe, A., Dawid, P., & Senn, S. S. (2011). *Simplicity, complexity and modelling*. Wiley. https://www.wiley.com/en-us/Simplicity%2C+Complexity+and+Modelling-p-9780470740026.

Clarke, P. M., Gray, A. M., Briggs, A., Stevens, R. J., Matthews, D. R., Holman, R. R., & 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), 868–877. https://doi.org/10.1007/s00125-005-1717-3

Crowe, B. L. (1969). The tragedy of the commons revisited. *Science, 166*(3909), 1103–1107.

Döll, P., & Siebert, S. (2002). Global modeling of irrigation water requirements, 8-1-8-10 *Water Resources Research, 38*(4). https://doi.org/10.1029/2001WR000355.

Eker, S., Rovenskaya, E., Obersteiner, M., & Langan, S. (2018). Practice and perspectives in the validation of resource management models. *Nature Communications, 9*(1), 5359. https://doi.org/10.1038/s41467-018-07811-9

Elshafei, Y., Tonts, M., Sivapalan, M., & Hipsey, M. R. (2016). Sensitivity of emergent sociohydrologic dynamics to internal system properties and external sociopolitical factors: Implications for water management. *Water Resources Research, 52*(6), 4944–4966. https://doi.org/10.1002/2015WR017944

European Commission, 2021. Better regulation toolbox. https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better4regulation-guidelines-and-toolbox_en.

Falconí, F., Ramos-Martin, J., & Cango, P. (2017). Caloric unequal exchange in Latin America and the Caribbean. *Ecological Economics, 134*, 140–149. https://doi.org/10.1016/j.ecolecon.2017.01.009

Fecher, B., & Friesike, S. (2014). Open science: one term, five schools of thought. In S. Bartling, & S. Friesike (Eds.), *Opening science* (pp. 17–47). Springer.

Funtowicz, S., & Ravetz, J. R. (1990). *Uncertainty and quality in science for policy*. https://doi.org/10.1007/978-94-009-0621-1_3

Funtowicz, S., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures, 25*(7), 739–755. https://doi.org/10.1016/0016-3287(93)90022-L

Funtowicz, S., & Ravetz, J. R. (1994). The worth of a songbird: Ecological economics as a post-normal science. *Ecological Economics, 10*(3), 197–207. https://doi.org/10.1016/0921-8009(94)90108-2

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

Ghoreishi, M., Razavi, S., & Elshorbagy, A. (2021). Understanding human adaptation to drought: agent-based agricultural water demand modeling in the Bow River Basin, Canada. *Hydrological Sciences Journal, 0*(ja). https://doi.org/10.1080/02626667.2021.1873344

Ghoreishi, M., Sheikholeslami, R., Elshorbagy, A., Razavi, S., & Belcher, K. (2021). Peering into agricultural rebound phenomenon using a global sensitivity analysis approach. *Journal of Hydrology, 602*, Article 126739. https://doi.org/10.1016/j.jhydrol.2021.126739

Giampietro, M., & Saltelli, A. (2014a). Footprints to nowhere. *Ecological Indicators, 46*, 610–621. https://doi.org/10.1016/j.ecolind.2014.01.030

Giampietro, M., & Saltelli, A. (2014b). Footworking in circles. *Ecological Indicators, 46*, 260–263. https://doi.org/10.1016/j.ecolind.2014.06.019

Gober, P., & Wheater, H. S. (2015). Debates–perspectives on socio-hydrology: Modeling flood risk as a public policy problem. *Water Resources Research, 51*(6), 4782–4788. https://doi.org/10.1002/2015WR016945

Goodhart, C. (1981). Problems of monetary management: The U.K. experience. In A. S. Courakis (Ed.), *Inflation, depression, and economic policy in the West* (pp. 111–146). Rowman & Littlefield.

Hamilton, S. H., Fu, B., Guillaume, J. H., Badham, J., Elsawah, S., Gober, P., Hunt, R. J., Iwanaga, T., Jakeman, A. J., Ames, D. P., Curtis, A., Hill, M. C., Pierce, S. A., & Zare, F. (2019). A framework for characterising and evaluating the effectiveness of environmental modelling. *Environmental Modelling and Software, 118*, 83–98. https://doi.org/10.1016/j.envsoft.2019.04.008

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., & Tanaka, K. (2008). An integrated model for the assessment of global water resources - part 2: Applications and assessments. *Hydrology and Earth System Sciences, 12*(4), 1027–1037. https://doi.org/10.5194/hess-12-1027-2008

Jakeman, A., Letcher, R., & Norton, J. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling & Software, 21*(5), 602–614.

Jasanoff, S. (2007). Technologies of humility, 33-33 *Nature, 450*(7166). https://doi.org/10.1038/450033a.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., & Thielen, J. (2016). Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environmental Modelling and Software, 75*, 68–76. https://doi.org/10.1016/j.envsoft.2015.09.009

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Laes, E., Meskens, G., & 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, 39*(9), 5664–5673. https://doi.org/10.1016/j.enpol.2011.04.016

Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Swilling, M., & Thomas, C. J. (2012). Transdisciplinary research in sustainability science: practice, principles, and challenges. *Sustainability Science, 7*(1), 25–43. https://doi.org/10.1007/s11625-011-0149-x

Levy, M. C., Garcia, M., Blair, P., Chen, X., Gomes, S. L., Gower, D. B., Grames, J., Kuil, L., Liu, Y., Marston, L., McCord, P. F., Roobavannan, M., & Zeng, R. (2016). Wicked but worth it: student perspectives on socio-hydrology. *Hydrological Processes, 30*(9), 1467–1472. https://doi.org/10.1002/hyp.10791

Little, J. C., Hester, E. T., Elsawah, S., Filz, G. M., Sandu, A., Carey, C. C., Iwanaga, T., & Jakeman, A. J. (2019). A tiered, system-of-systems modeling framework for resolving complex socio-environmental policy issues. *Environmental Modelling and Software, 112*, 82–94. https://doi.org/10.1016/j.envsoft.2018.11.011

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H., & Taylor, W. W. (2007). Complexity of coupled human and natural systems. *Science, 317*(5844), 1513–1516. https://doi.org/10.1126/science.1144004

Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and Social Sciences Communications, 7*(1), 1–7. https://doi.org/10.1057/s41599-020-0501-9

Lo Piano, S., & Robinson, M. (2019). Nutrition and public health economic evaluations under the lenses of post normal science. *Futures, 112*, Article 102436. https://doi.org/10.1016/J.FUTURES.2019.06.008

Lo Piano, S. (2017). Quantitative story-telling and sensitivity-auditing appraisal of expert solutions to the issue of food security. doi: 10.13140/RG.2.2.36288.43525.

Mennicken, A., & Espeland, W. N. (2019). What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology, 45*(1), 223–245. https://doi.org/10.1146/annurev-soc-073117-041343

Micklewright, J., Schnepf, S. V., & Skinner, C. (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), 915–938. https://doi.org/10.1111/j.1467-985X.2012.01036.x

Micklewright, J. & Schnepf, S. V. (2006). Inequality of learning in industrialised countries. *Technical report 2517*. Institute of Labor Economics (IZA). ⟨https://ideas.repec.org/p/iza/izadps/dp2517.html⟩.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences, 18*(9), 3511–3538. https://doi.org/10.5194/hess-18-3511-2014

MüllerSchmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., & Döll, P. (2020). The global water resources and use model WaterGAP v2.2d: Model description and evaluation. *Geoscientific Model Development*, 1–69. https://doi.org/10.5194/gmd-2020-225

O'Neill, R. V. (1989). Perspectives in hierarchy and scale. In J. Roughgarden, R. M. May, & S. A. Levin (Eds.), *Perspectives in ecological theory*. https://doi.org/10.1515/9781400860180

Padilla, J. J., Diallo, S. Y., Lynch, C. J., & Gore, R. (2018). Observations on the practice and profession of modeling and simulation: A survey approach. *Simulation, 94*(6), 493–506. https://doi.org/10.1177/0037549717737159

Page, S. E. (2018). *The model thinker* (first edition). New York: Basic Books. https://www.basicbooks.com/titles/scott-e-page/the-model-thinker/9780465094639/.

Parker, W. S. (2013). Ensemble modeling, uncertainty and robust predictions. *WIREs Climate Change, 4*(3), 213–223. https://doi.org/10.1002/wcc.220

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E. D., Caldwell, R., Evora, N., & Pellerin, P. (2007). Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences, 11*(4), 1279–1294. https://doi.org/10.5194/hess-11-1279-2007

Pinch, T. J. (1992). Opening black boxes: Science, technology and society. *Social Studies of Science, 22*(3), 487–510. https://doi.org/10.1177/0306312792022003003

Pontius, R. G., & Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing, 32*(15), 4407–4429. https://doi.org/10.1080/01431161.2011.552923

Popp Berman, E., & Hirschman, D. (2018). The sociology of quantification: Where are we now? *Contemporary Sociology, 47*(3), 257–266.

Pouladi, P., Afshar, A., Afshar, M. H., Molajou, A., & Farahmand, H. (2019). Agent-based socio-hydrological modeling for restoration of Urmia Lake: Application of theory of planned behavior. *Journal of Hydrology, 576*, 736–748. https://doi.org/10.1016/j.jhydrol.2019.06.080

Puy, A., Beneventano, P., Levin, S. A., Lo Piano, S., Portaluri, T., & Saltelli, A. (2022b). Models with higher effective dimensions tend to produce more uncertain estimates. *Science Advances*. , Article eabn9450. https://doi.org/10.1126/sciadv.abn9450. In press.

Puy, A., Borgonovo, E., Lo Piano, S., Levin, S. A., & Saltelli, A. (2021). Irrigated areas drive irrigation water withdrawals. *Nature Communications, 25*, 4525. https://doi.org/10.1038/s41467-021-24508-8

Puy, A., Sheikholeslami, R., Gupta, H. V., Hall, J. W., Lankford, B., LoPiano, S., Meier, J., Pappenberger, F., Porporato, A., Vico, G., & Saltelli, A. (2022a). The delusive accuracy of global irrigation water withdrawal estimates. *Nature Communications, 13*(1), 3183. https://doi.org/10.1038/s41467-022-30731-8

United Nations (2021). Pathways to zero hunger — zero hunger challenge. ⟨https://www.un.org/zerohunger/node/20⟩.

Ravetz, J. R. (1997). Integrated environmental assessment forum, developing guidelines for "good practice"Ĭ, project ULYSSES. *Technical report*. Darmstadt University of Technology. ⟨http://www.jvds.nl/ulysses/eWP97-1.pdf⟩.

Rayner, S. (2012). Uncomfortable knowledge: The social construction of ignorance in science and environmental policy discourses. *Economy and Society, 41*(1), 107–125. https://doi.org/10.1080/03085147.2011.637335

Rhodes, T., & Lancaster, K. (2020). Mathematical models as public troubles in COVID-19 infection control: Following the numbers. *Health Sociology Review*, 1–18. https://doi.org/10.1080/14461242.2020.1764376

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169. https://doi.org/10.1007/BF01405730

Rokaya, P., Sheikholeslami, R., Kurkute, S., Nazarbakhsh, M., Zhang, F., & Reed, M. G. (2017). Multiple factors that shaped sustainability science journal: A 10-year review. *Sustainability Science, 12*(6), 855–868. https://doi.org/10.1007/s11625-017-0495-4

Roobavannan, M., van Emmerik, T. H. M., Elshafei, Y., Kandasamy, J., Sanderson, M. R., Vigneswaran, S., Pande, S., & Sivapalan, M. (2018). Norms and values in sociohydrological models. *Hydrology and Earth System Sciences, 22*(2), 1337–1349. https://doi.org/10.5194/hess-22-1337-2018

Saltelli, A. (2019a). What can we learn from the debate on statistical significance? ⟨https://errorstatistics.com/2019/11/22/a-saltelli-guest-post-what-can-we-learn-from-the-debate-on-statistical-significance/⟩.

Saltelli, A. (2019b). Statistical versus mathematical modelling: A short comment. *Nature Communications, 10*, 1–3. https://doi.org/10.1038/s41467-019-11865-8

Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software, 25*(12), 1508–1517. https://doi.org/10.1016/j.envsoft.2010.04.012

Saltelli, A., Bammer, G., Bruno, I., Charters, E., Fiore, M. D., Didier, E., Espeland, W. N., Kay, J., LoPiano, S., Mayo, D., Jr, R. P., Portaluri, T., Porter, T. M., Puy, A., Rafols, I., Ravetz, J. R., Reinert, E., Sarewitz, D., Stark, P. B., … Vineis, P. (2020). Five ways to ensure that models serve society: A manifesto. *Nature, 582*(7813), 482–484. https://doi.org/10.1038/d41586-020-01812-9

Saltelli, A., Benini, L., Funtowicz, S., Giampietro, M., Kaiser, M., Reinert, E. S., & van der Sluijs, J. P. (2020). The technique is never neutral. How methodological choices condition the generation of narratives for sustainability. *Environmental science and Policy, 106*, 87–98.

Saltelli, A., & Di Fiore, M. (2020). From sociology of quantification to ethics of quantification. *Humanities and Social Sciences Communications, 7*, Article 69. https://doi.org/10.1057/s41599-020-00557-0

Saltelli, A., & Funtowicz, S. (2014). When all models are wrong. *Issues in Science and Technology, 30*(2), 79–85. https://www.jstor.org/stable/43315849.

Saltelli, A., & Funtowicz, S. (2017). What is scienceas crisis really about? *Futures, 91*, 5–11. https://doi.org/10.1016/j.futures.2017.05.010

Saltelli, A., Guimaraes Pereira, Â., Sluijs, J. P. V.d., & Funtowicz, S. (2013). What do I make of your latinorum sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy, 9*(2/3/4), 213. https://doi.org/10.1504/IJFIP.2013.058610

Saltelli, A., & Lo Piano, S. (2017). Problematic quantifications: a critical appraisal of scenario making for a global 'sustainable' food production. *Food Ethics, 1*(2), 173–179.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: The primer*. Hoboken, NJ: Wiley.

Saltelli, A., Andreoni, A., Drechsler, W., Ghosh, J., Kattel, R., Kvangraven, I. H., Rafols, I., Reinert, E. S., Stirling, A., & Xu, T. (2021). Why ethics of quantification is needed now. *Technical report*. London.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., … Kabat, P. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences of the United States of America, 111*(9), 3245–3250. https://doi.org/10.1073/pnas.1222460110

Science Advice for Policy by European Academies (2019). Making sense of science for policy under conditions of complexity and uncertainty. *Technical report*. SAPEA, Berlin. ⟨https://www.sapea.info/topics/making-sense-of-science/⟩.

Shafiee, M. E., & Zechman, E. M. (2013). An agent-based modeling framework for sociotechnical simulation of water distribution contamination events. *Journal of Hydroinformatics, 15*(3), 862–880. https://doi.org/10.2166/hydro.2013.158

Sheikholeslami, R., Razavi, S., Gupta, H. V., Becker, W., & Haghnegahdar, A. (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*, 282–299. https://doi.org/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), 4795–4805. https://doi.org/10.1002/2015WR017080

Sivapalan, M., & Blöschl, G. (2015). Time scale interactions and the coevolution of humans and water. *Water Resources Research, 51*(9), 6988–7022. https://doi.org/10.1002/2015WR017896

Sivapalan, M., Savenije, H. H. G., & Blöschl, G. (2012). Socio-hydrology: A new science of people and water. *Hydrological Processes, 26*(8), 1270–1276. https://doi.org/10.1002/hyp.8426

Smaldino, P., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science, 3*(160384).

Smith, K. A. (2015). Investigating uncertainty in global hydrology modelling (Doctoral dissertation). University of Nottingham). ⟨10.13140/RG.2.2.22810.95685⟩.

Sood, A., & Smakhtin, V. (2015). Global hydrological models: A review. *Hydrological Sciences Journal, 60*(4), 549–565. https://doi.org/10.1080/02626667.2014.950580

Sutanudjaja, E. H., van Beek, L. P. H., Drost, N., de Graaf, I. E. M., de Jong, K., Peßenteiner, S., Straatsma, M. W., Wada, Y., Wanders, N., Wisser, D., & Bierkens, M. F. P. (2018). PCR-GLOBWB 2.0: A 5 arc-minute global hydrological and water resources model. *Geoscientific Model Development, 11*(6), 2429–2453.

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space, 34*(6), 990–1006. https://doi.org/10.1177/0263775816633195

van der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. R., & Risbey, J. (2005). Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: The NUSAP system. *Risk Analysis, 25*(2), 481–492. https://doi.org/10.1111/j.1539-6924.2005.00604.x

Veldkamp, T. I., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., & Wada, Y. (2018). Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: A multi-model validation study. *Environmental Research Letters, 13*(5). https://doi.org/10.1088/1748-9326/aab96f

Viglione, A., Di Baldassarre, G., Brandimarte, L., Kuil, L., Carr, G., Salinas, J. L., Scolobig, A., & Blöschl, G. (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. https://doi.org/10.1016/j.jhydrol.2014.01.018

Vörösmarty, C. J., Federer, C. A., & Schloss, A. L. (1998). Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling. *Journal of Hydrology, 207*(3-4), 147–169. https://doi.org/10.1016/S0022-1694(98)00109-7

Wada, Y., Wisser, D., Eisner, S., Flörke, M., Gerten, D., Haddeland, I., Hanasaki, N., Masaki, Y., Portmann, F. T., Stacke, T., Tessler, Z., & Schewe, J. (2013). Multimodel projections and uncertainties of irrigation water demand under climate change. *Geophysical Research Letters, 40*(17), 4626–4632. https://doi.org/10.1002/grl.50686

Walters, W. P. (2020). Code sharing in the open science era. *Journal of Chemical Information and Modeling, 60*(10), 4417–4420. https://doi.org/10.1021/acs.jcim.0c01000

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The inter-sectoral impact model intercomparison project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences, 111*(9), 3228–3232. https://doi.org/10.1073/pnas.1312330110

Wheater, H., & Gober, P. (2013). Water security in the Canadian Prairies: Science and management challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371*(2002), Article 20120409. https://doi.org/10.1098/rsta.2012.0409

Woessmann, L. (2016). The economic case for education. *Education Economics, 24*(1), 3–32. https://doi.org/10.1080/09645292.2015.1059801

Wuttke, J. (2007). Uncertainty and bias in PISA (SSRN Scholarly Paper No. ID 1159042). Social Science Research Network. Rochester, NY. ⟨https://papers.ssrn.com/abstract=1159042⟩.

Zarekarizi, M., Srikrishnan, V., & Keller, K. (2020). Neglecting uncertainties biases house-elevation decisions to manage riverine flood risks. *Nature Communications, 11*(1), 5361. https://doi.org/10.1038/s41467-020-19188-9

Zhao, J., Cai, X., & Wang, Z. (2013). Comparing administered and market-based water allocation systems through a consistent agent-based modeling framework. *Journal of Environmental Management, 123*, 120–130. https://doi.org/10.1016/j.jenvman.2013.03.005