Seminario extra corso: “Global sensitivity analysis or Multiverse Analysis?”

Andrea Saltelli

June 11, 2026

Aula 203, secondo piano, Centro Didattico Morgagni

Università degli Studi di Firenze

Responsabili scientifiche: Michela Baccini, Chiara Marziscuola

https://www.sismec.info/notizie/masterclass-environmental-exposures-and-health-measuring-impacts-and-understanding-uncertainties/

Andrea Saltelli

HOME ABOUT ME PUBLICATIONS NEWS & VIDEOS RESOURCES

Latest papers

Models and the common good

A short analysis of norms and counter-norms in modelling, discussing how models may misuse their epistemic authority for reasons of occasion, opportunity and interest.

Outline:
Narratives and counter-narratives,
State of exception,
Seeking Cinderella,
Reformation

Ranking the Rankers

Ranking the Rankers? We investigate using global sensitivity analysis the properties of the science-wide author databases' standardised citation indicators produced since 2015 by a team of researchers led by John P. A. Ioannidis.

Digital twins

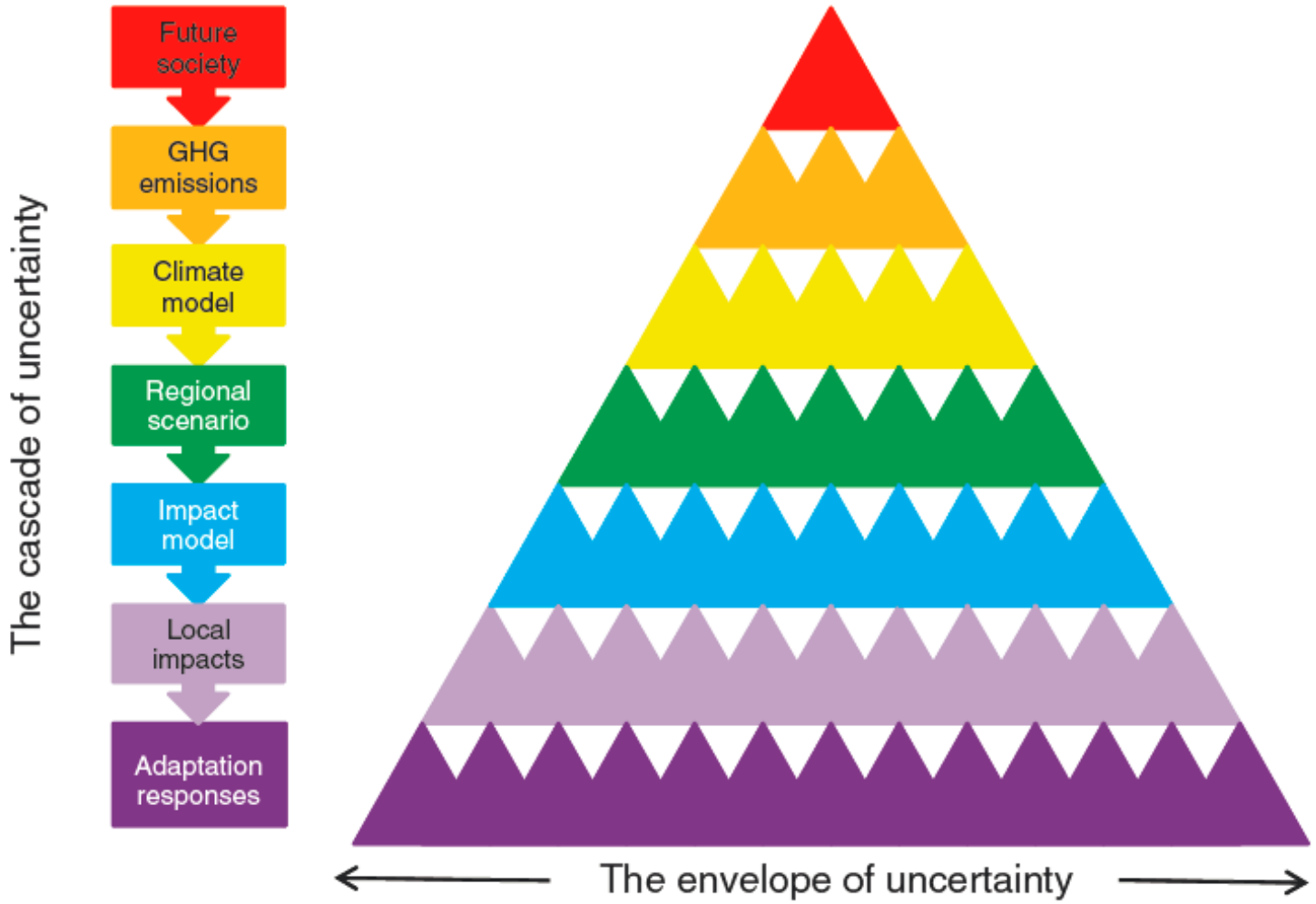
We reflect on the development of digital twins of

Where to find this talk:

www.andreasaltelli.eu

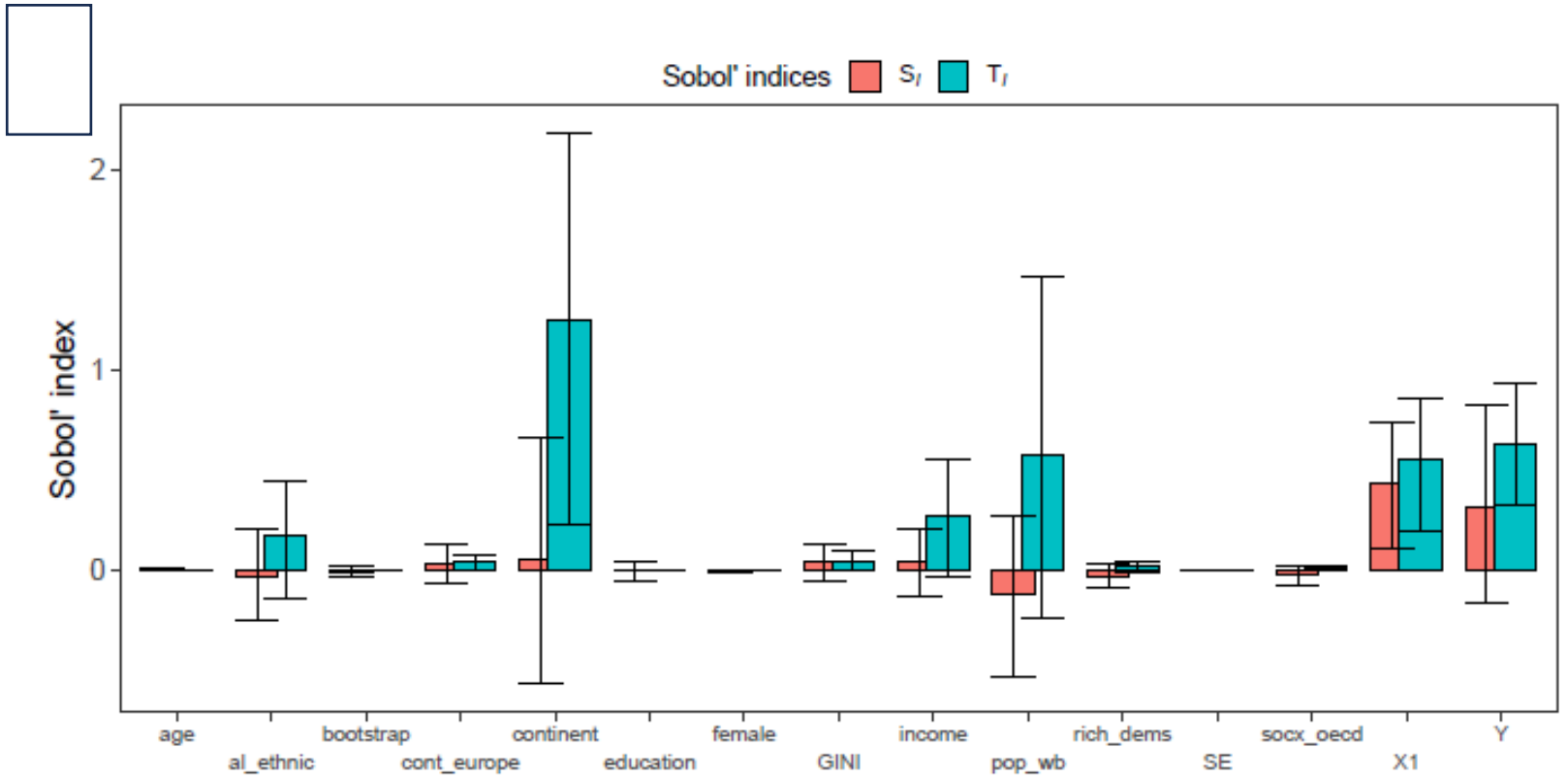
Part I - What are uncertainty analysis and global sensitivity analysis?

Uncertainty analysis: the study of the uncertainty in model output

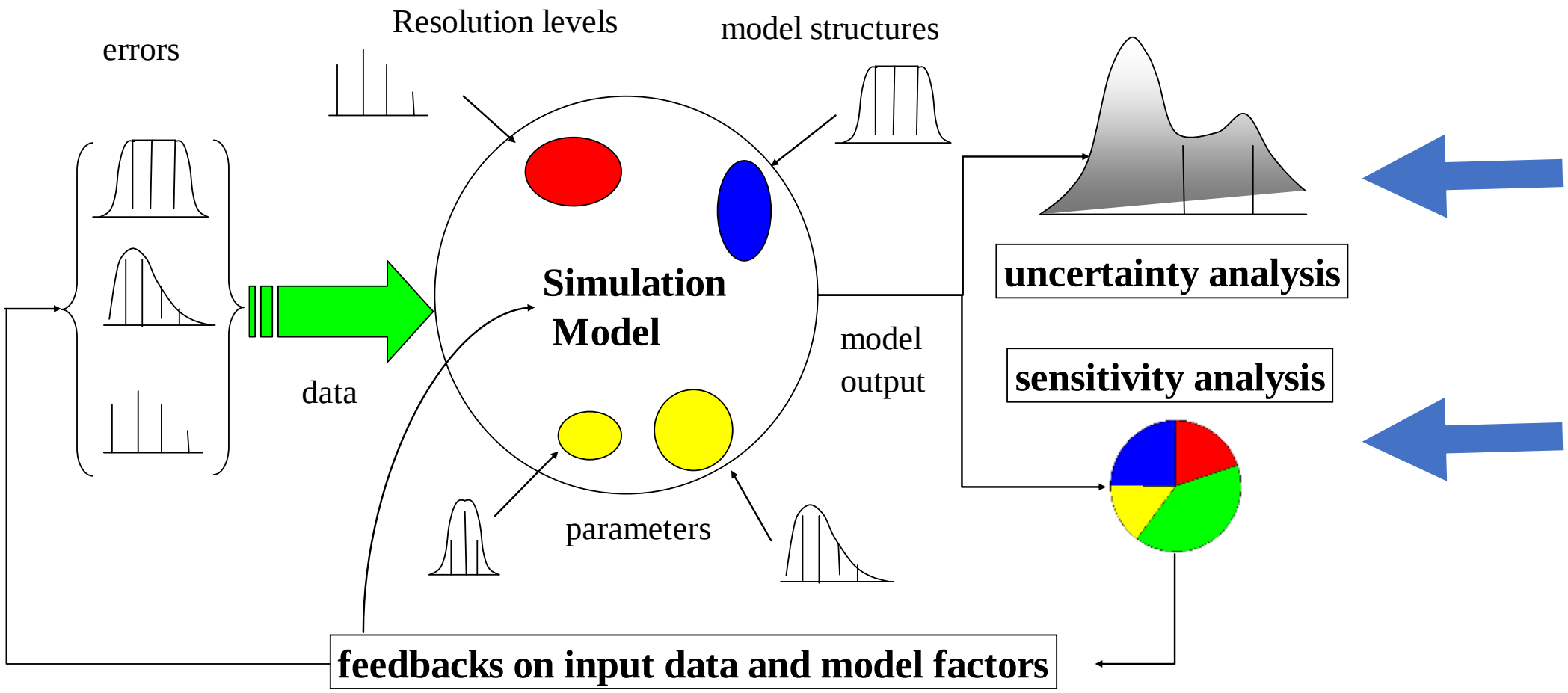


Source: https://www.climate-lab-book.ac.uk/2014/cascade-of-uncertainty/

Sensitivity analysis: the study of the relative importance of different input factors on the model output



Stefano Tarantola, circa 2008



Why global? Why don't we simply move one factor At a Time (OAT)

A geometric proof

Environmental Modelling & Software 25 (2010) 1508–1517

Contents lists available at ScienceDirect

Environmental Modelling & Software

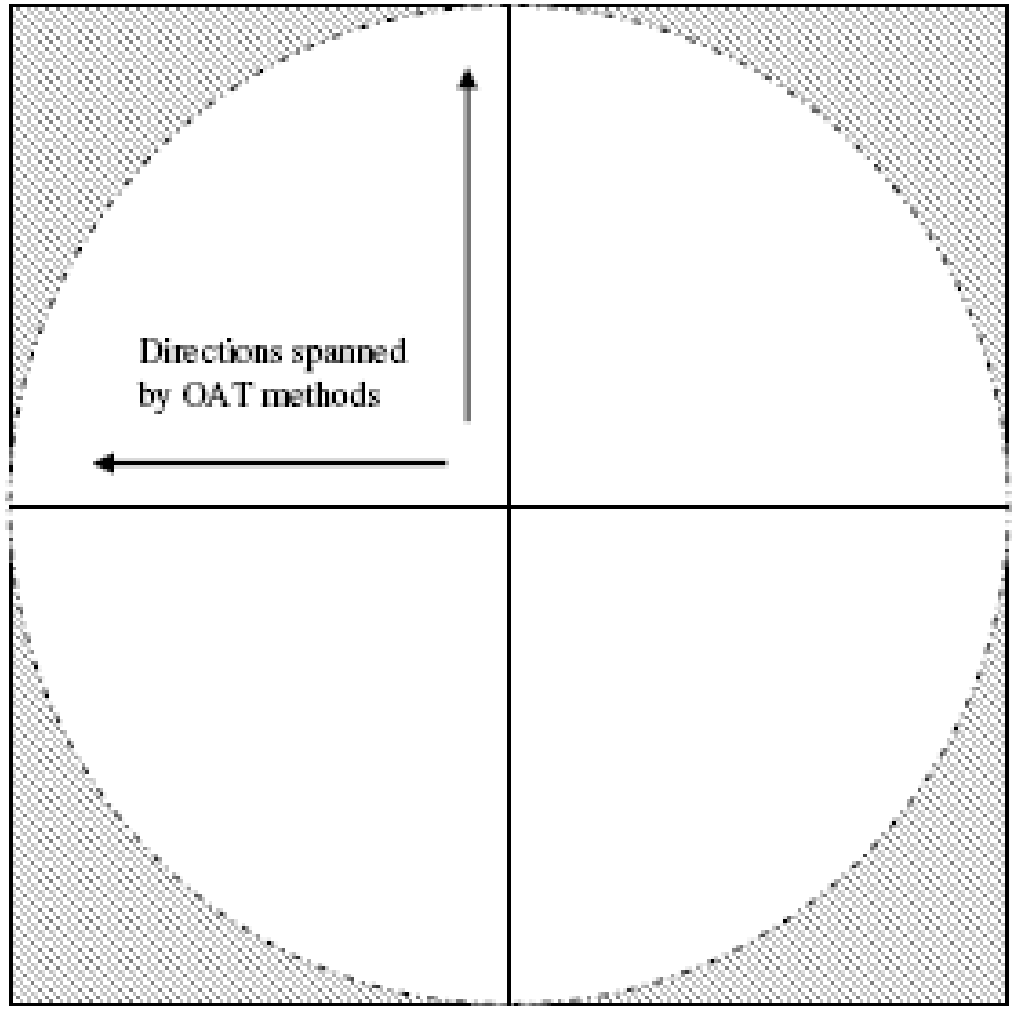
journal homepage: www.elsevier.com/locate/envsoft

How to avoid a perfunctory sensitivity analysis

Andrea Saltelli*, Paola Annoni

Joint Research Center, Institute for the Protection and Security of the Citizen, via E.Fermi, 2749, Ispra VA 21027, Italy

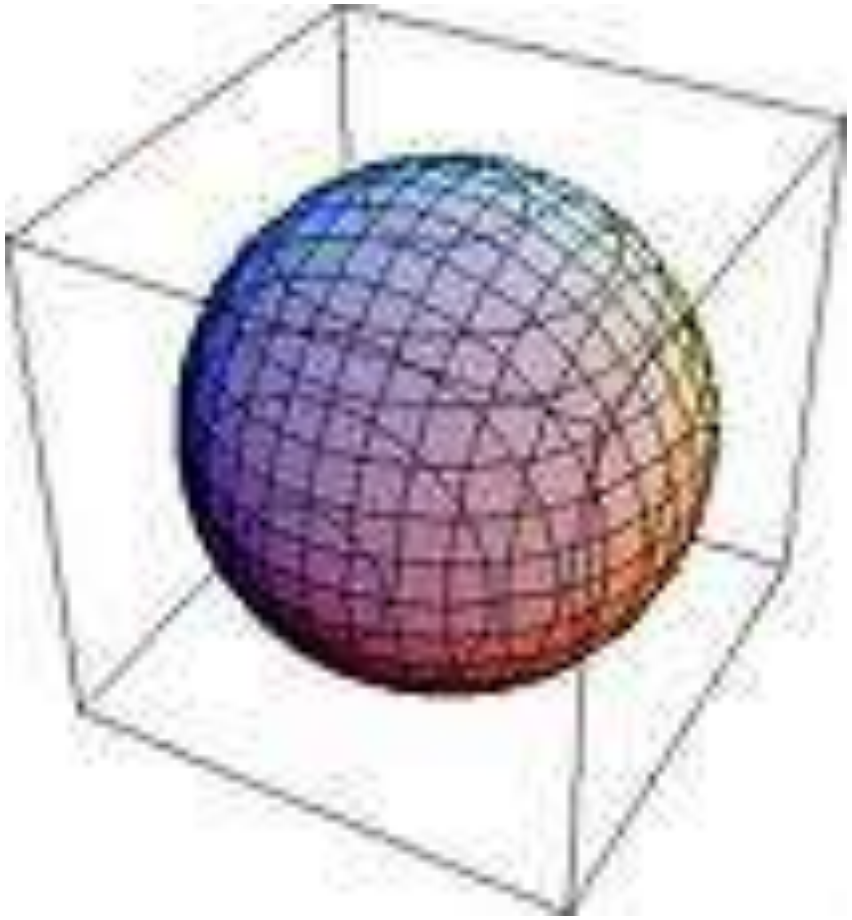
OAT in 2 dimensions



Area circle / area square =?

~ 3/4

OAT in 3 dimensions

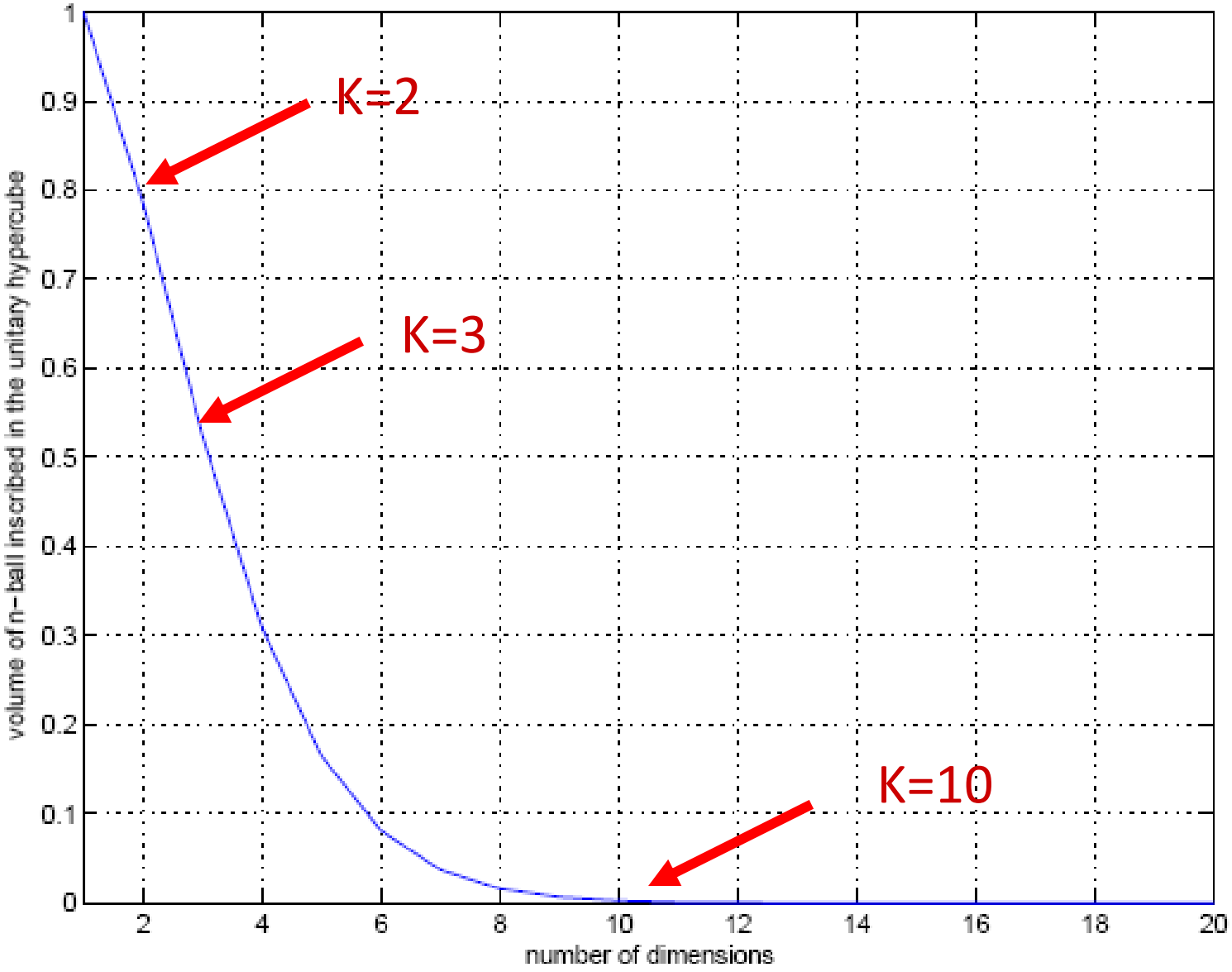


Volume sphere / volume cube =?

~ 1/2

OAT in 10 dimensions; Volume hypersphere / volume ten dimensional hypercube =? ~ 0.0025

OAT in k dimensions

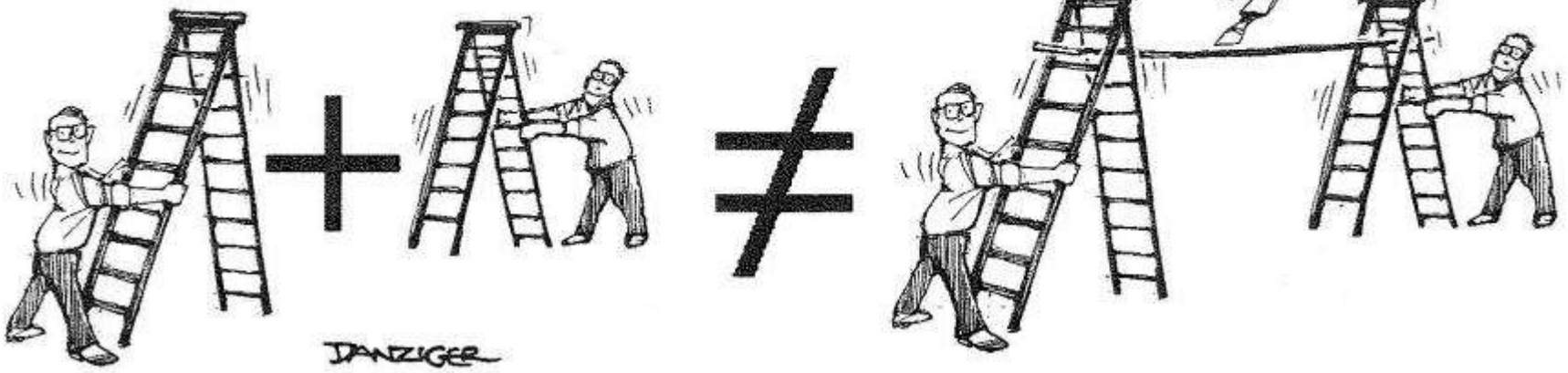


OAT does not capture interactions

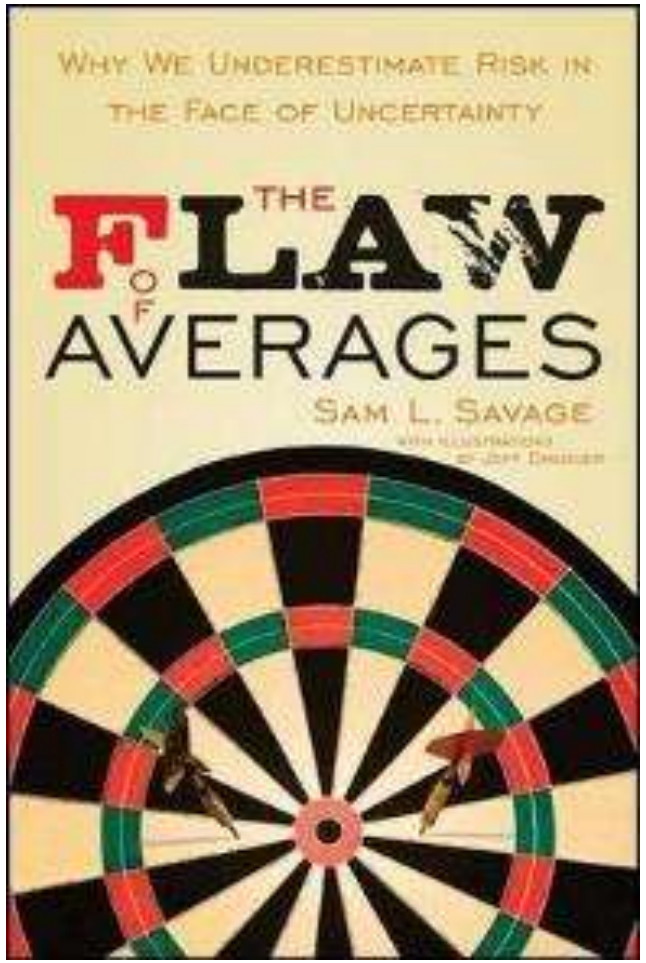
➔ The resulting analysis is non conservative

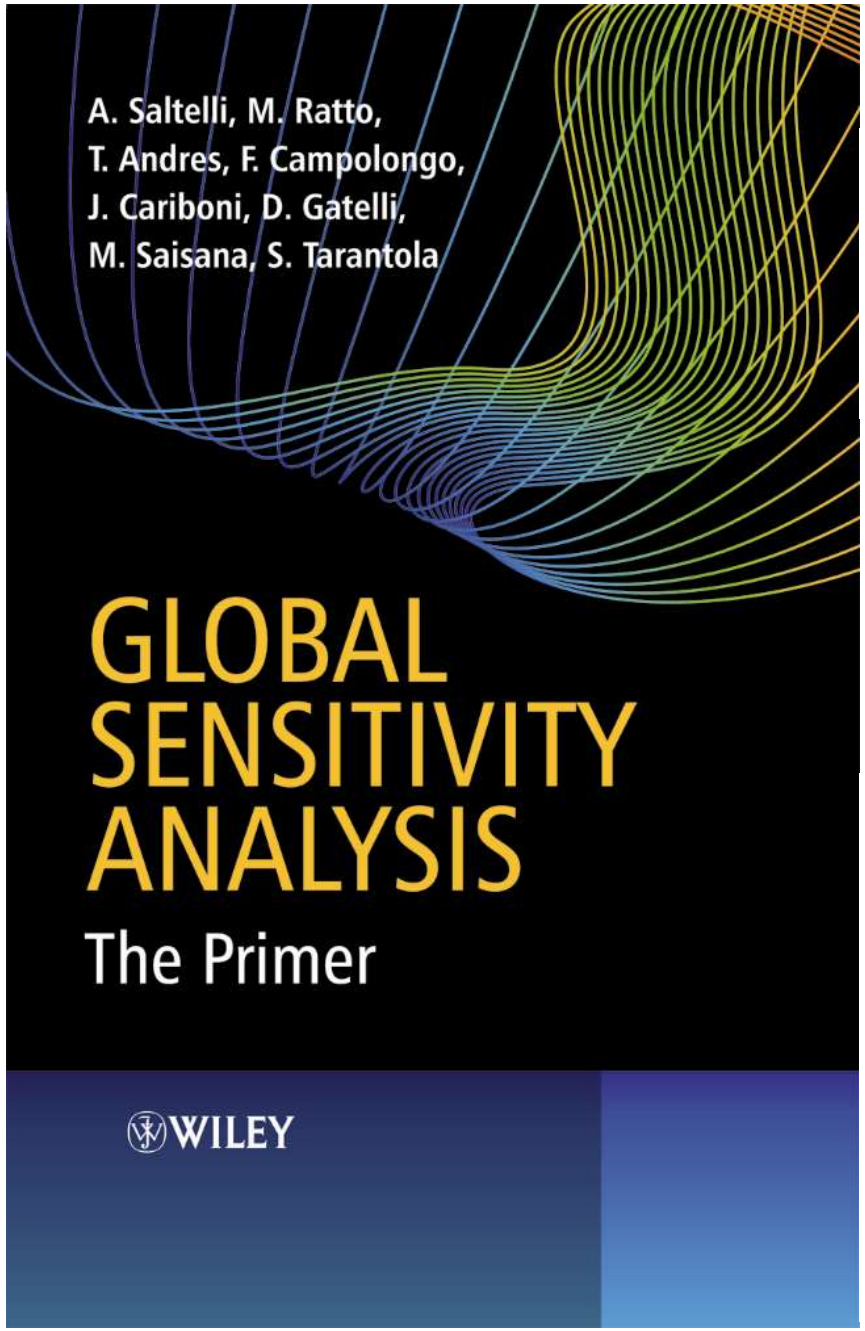
How would you test the scaffolding?

How coupled ladders are shaken in most of available literature



How to shake coupled ladders





All about variance-based sensitivity analysis and Sobol indices

(available online)

1.6 EXERCISES

1. Prove that

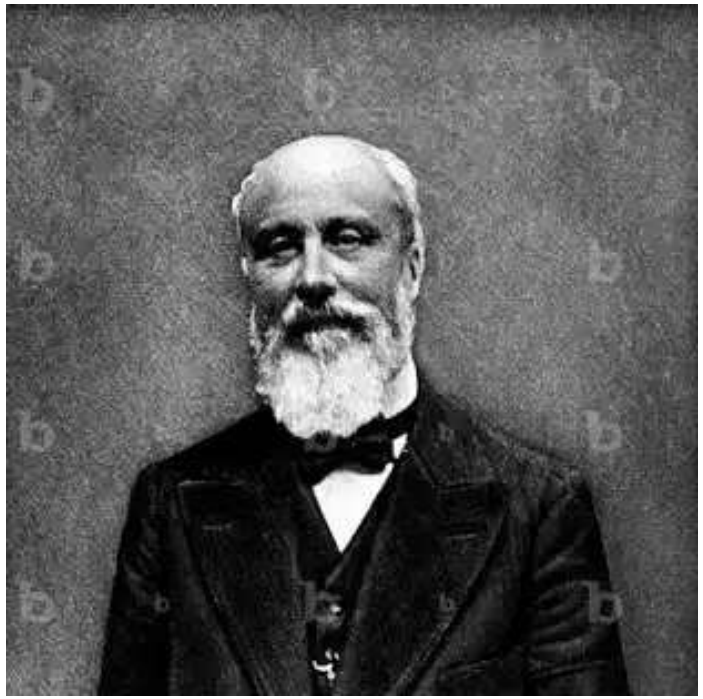
$$V(Y) = E(Y^2) - E^2(Y).$$

2. Prove that for an additive model of two independent variables X_1 and X_2, fixing one variable can only decrease the variance of the model.

Part II - Is GSA important? According to whom?

"… inferences from mathematical models of phenomena to real physical applications must also be demonstrated to be approximately correct when the assumptions of the model are only approximately true" [1]

Pierre Duhem (1861–1916) and his 'Principle of stability' ~1906

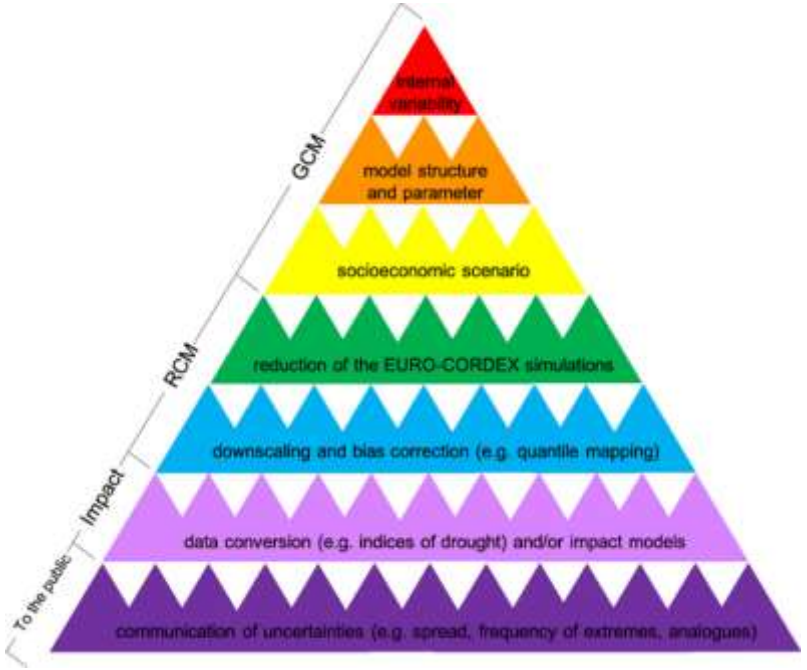


[1] See: S. C. Fletcher, 'The Principle of Stability', Philosopher's Imprint, vol. 20, no. 3, 2020, Accessed: Sep. 08, 2024. [Online].

Models are fragile

Duhem's principle of stability [1], links to the occurrence of either Butterfly [2] or Hawkmoth effects [3] … as well as to the accumulation of parametric error in a model, the so-called uncertainty cascade [4], that is the subject of global sensitivity analysis studies

[1] S. C. Fletcher, 'The Principle of Stability', Philosopher's Imprint, vol. 20, no. 3, 2020, Accessed: Sep. 08, 2024. [Online].
[2] H. G. Schuster, Deterministic Chaos: An Introduction, 2nd Rev edition. Weinheim: Vch Pub, 1998.
[3] E. Winsberg, 'Appendix: Structural Stability and the "Hawkmoth Effect"', in Philosophy and Climate Science, Cambridge: Cambridge University Press, 2018, pp. 232–246. doi: 10.1017/9781108164290.016.
[4] M. Christie, A. Cliffe, P. Dawid, and S. S. Senn, Simplicity, Complexity and Modelling. Wiley, 2011.



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

Edward E. Leamer, Let's Take the Con Out of Econometrics, *American Economics Review*, **73** (March **1983**), 31-43.

Edward E. Leamer

"One reason these methods are rarely used is their honesty seems destructive."

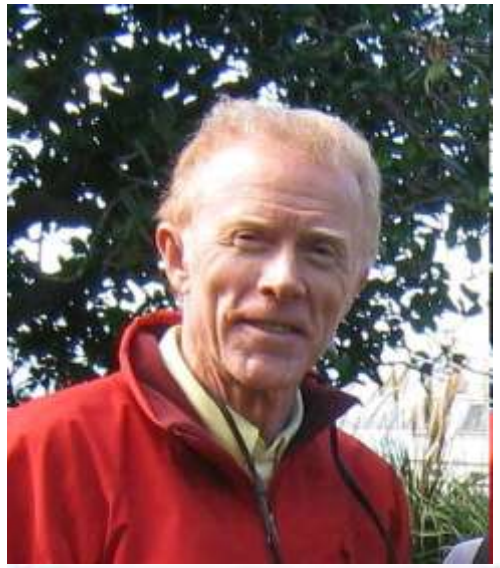
Tantalus on the Road to Asymptopia, Edward E. Leamer, 2010 *Journal of Economic Perspectives*, 24, (2), 31-46.

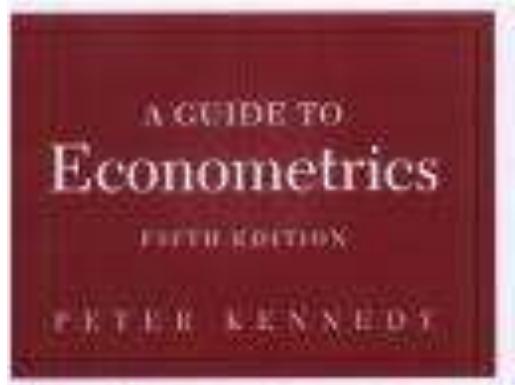
27 years

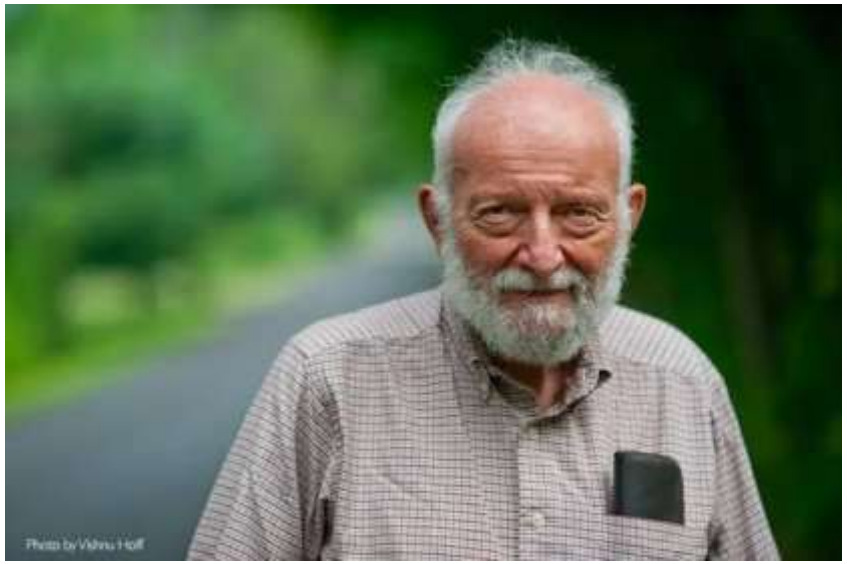
Edward E. Leamer, Let's Take the Con Out of Econometrics, *American Economics Review*, 73 (March 1983), 31-43.

Peter Kennedy, A Guide to Econometrics. Anticipating criticism by applying sensitivity analysis. This is one of the ten commandments of applied econometrics according to Peter Kennedy:

<<Thou shall confess in the presence of sensitivity.
Corollary: Thou shall anticipate criticism >>







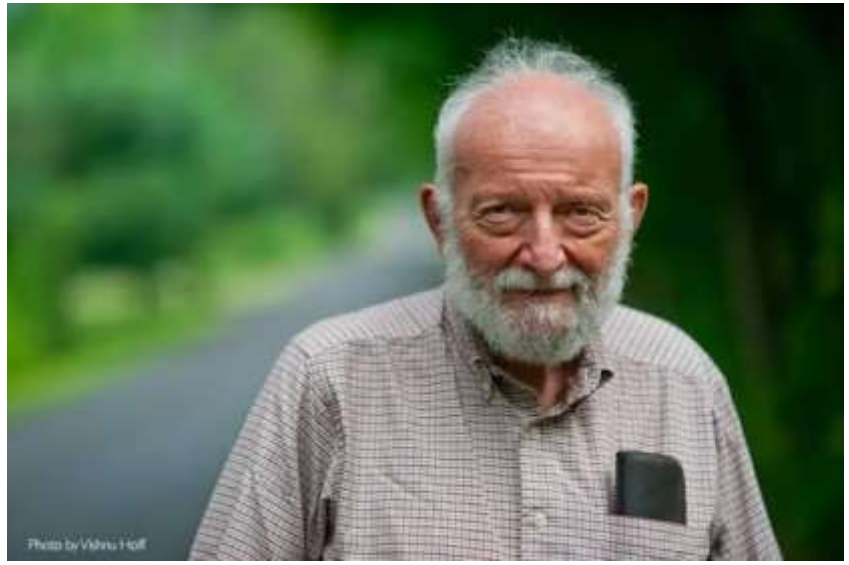
Richard Levins
1930-2016

Richard Levins' robustness analysis:

"Our truth is the intersection of independent lies"

➔ Construct multiple models with different assumptions and treat a result as credible only if it persists across those independent formulations

➔ Identify conclusions that are invariant to idealizations rather than artifacts of any single model.



A fate similar to that of Leamer's global sensitivity analysis befell Richard Levins' robustness analysis

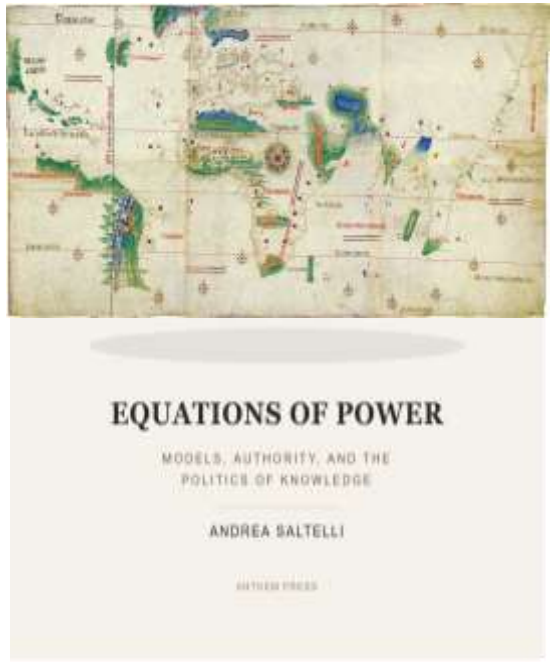
Richard Levins. The Strategy of Model Building in Population Biology. American Scientist, 54(4):421–431, 1966.

40 years

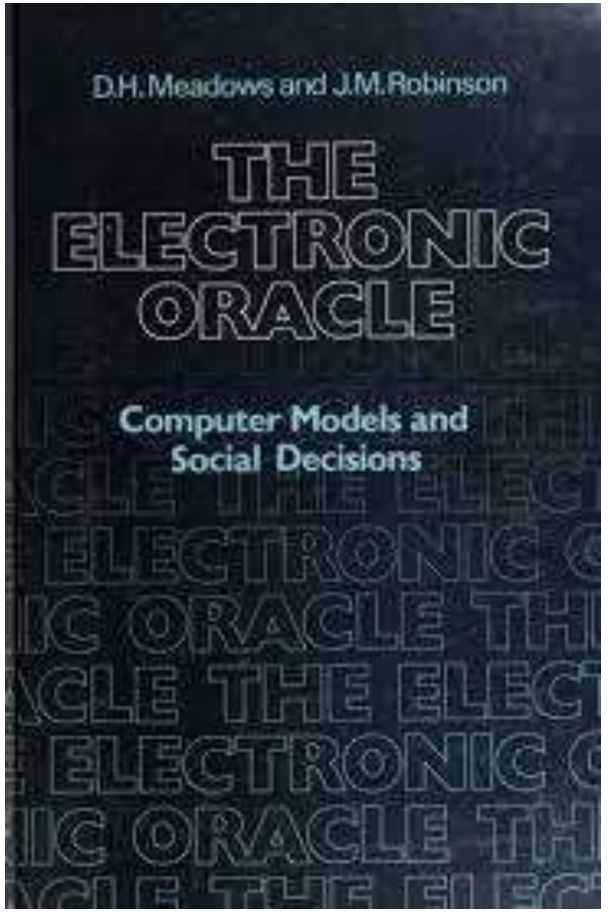
Michael Weisberg. Robustness Analysis. Philosophy of Science, 73(5):730–742, 2006.

Validation methods have historically lacked traction, and Dana Meadows lamented this ‘going on for decades’ in 1985 … Why?

A discussion here

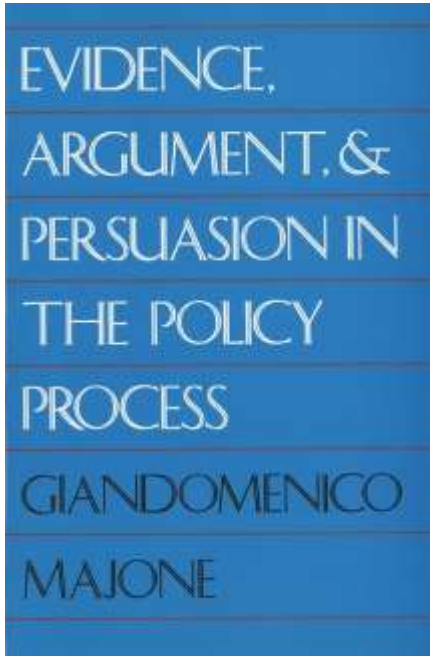


To appear



Published in 1985

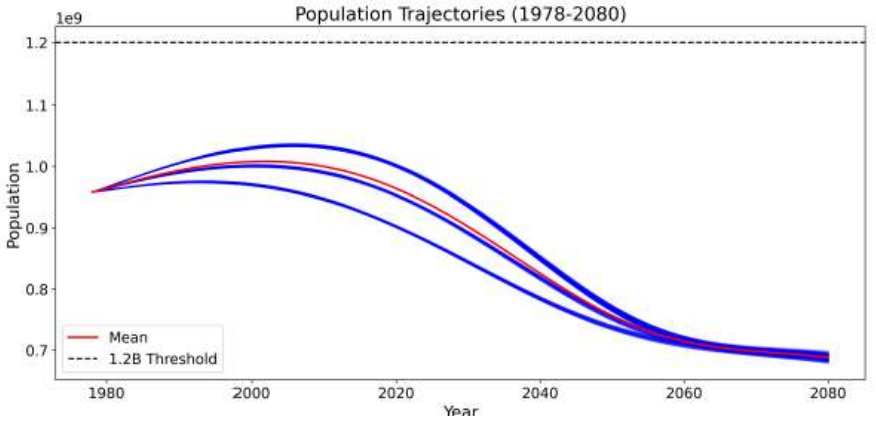
Giandomenico Majone



"Are the results from a particular model more sensitive to changes in the model and the methods used to estimate its parameters, or to changes in the data?"

p. 62, Giandomenico Majone. Evidence, argument, and persuasion in the policy process. Yale University Press, 1989.

The Song Jian / one-child policy case. We find near-insensitivity to inputs …



Download This Paper | Open PDF in Browser

SSRN

What Demographic Prediction Can and Cannot Achieve

14 Pages • Posted: 19 May 2026

Samuele Lo Piano
Gdańsk University of Technology - Faculty of Management and Economics

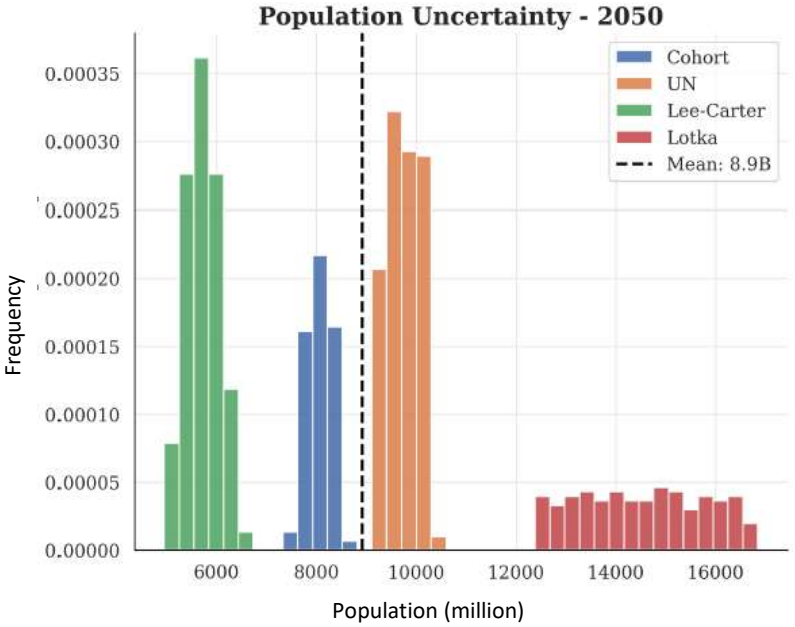
Marta Kuc Czarnecka
Gdańsk University of Technology

Roger Pielke
University of Colorado at Boulder

Andrea Saltelli
UPF Barcelona School of Management

Date Written: May 15, 2026

… and projections **today in use** are dramatically model-dependent



But the real strength of the models, in my mind at least, were in sensitivity analysis (where one could examine the response of the model to parameters or structures that were not known with precision (i.e., sensitivity analysis), and in the examination of the behavior of the model components relative to that of the real system in question (i.e., validation)

… By undertaking sensitivity analysis and validation, a great deal can be learned about the real system, including what you do not know. (Hall, 2020)

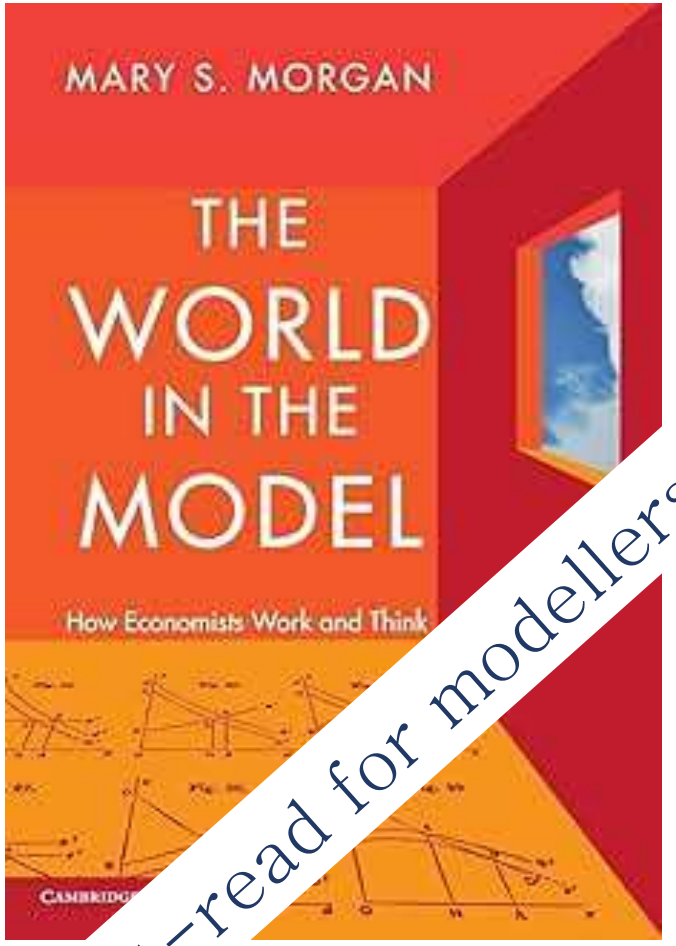
Charles A.S. Hall

Hall, C. A. S. (2020). Systems Ecology and Limits to Growth: History, Models, and Present Status. In G. S. Metcalf, K. Kijima, & H. Deguchi, eds., *Handbook of Systems Sciences*, Singapore: Springer, , pp. 1–38.

Mary S. Morgan

Hall's words resonate with Mary Morgan's 'the world in the model'

Playing with a model's moving parts to see what this gives, before seeing if it chimes with 'the model in the world'

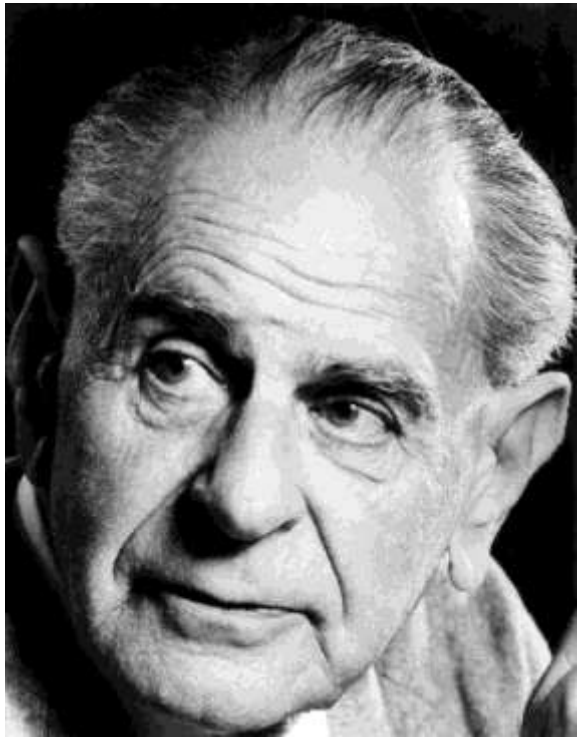


A must-read for modellers

Models that take no risks

The empirical content of a theory is measured by the risks it takes — the observations that could prove it wrong (severe tests)

Climate models projecting to 2300. Nuclear waste safety models covering a million years. Digital twins simulating ecosystem dynamics with 50-year horizons



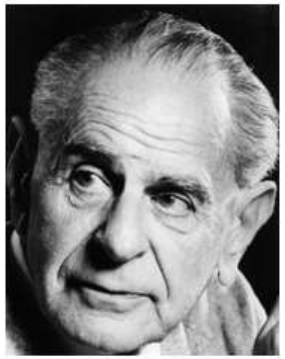
Karl Popper
(1902–1994)

These models take no risks. No severe tests here

Models that take no risks

The empirical content of a theory is measured by the risks it takes — the observations that could prove it wrong (severe tests)

Climate models projecting to 2300. Nuclear waste safety models covering a million years. Digital twins simulating ecosystem dynamics with 50-year horizons



Karl Popper
(1902-1994)

These models take no risks. No severe tests here

Global sensitivity analysis as models' risky tests

Part III – Modelling cultures

Leo Breiman
1928-2005

Leo Breiman's 'Two cultures' work

"The goal is not interpretability, but accurate information"

➔ Embrace model-agnostic algorithmic culture that prioritizes empirical predictive accuracy

Leo Breiman. Statistical Modeling: The Two Cultures Statistical Science, 16(3):199–231, August **2001**

Two cultures

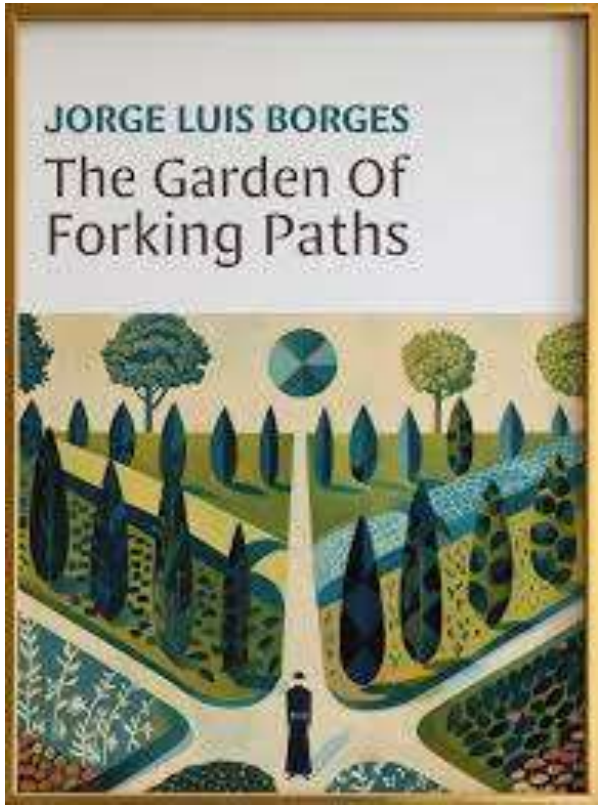
The Data Modelling Culture
Data are generated by a specific stochastic model; estimate its parameters.

The Algorithmic Modelling Culture
The data-generating mechanism is unknown; focus on predictive accuracy using algorithms.

Leo Breiman. Statistical Modeling: The Two Cultures Statistical Science, 16(3):199–231, August 2001

The Algorithmic Modelling Culture
Provided an epistemological justification for machine learning and, in time, AI

The Data Modelling Culture
Discovered the ‘Garden of Forking Paths; opened the season of multi-analyst experiments and multiverse analysis



Application to Machine Learning

Example use SA to ascertain that an algorithm does not make implicit use of protected attributes

More reading about explainable ML

Lu, Xuefei, and Emanuele Borgonovo. 2026. “How Do I Explain That Decision? A New Approach Combining Counterfactuals and Sensitivity Analyses.” European Journal of Operational Research 334 (1): 200–215. https://doi.org/10.1016/j.ejor.2026.04.035.

Giulia Vannucci, Roberta Siciliano, Andrea Saltelli, 2026, Global sensitivity analysis in random forests: unveiling generative variable importance. Statistical Methods and Applications, 34(5), https://doi.org/10.1007/s10260-026-00839-y.

Bénesse, C., Gamboa, F., Loubes, J.-M., & Boissin, T. (2022). Fairness seen as Global Sensitivity Analysis. *Machine Learning*. https://doi.org/10.1007/s10994-022-06202-y

More reading about explainable ML

Bénesse, C., Gamboa, F., Loubes, J.-M., & Boissin, T. (2022). Fairness seen as Global Sensitivity Analysis. *Machine Learning*. https://doi.org/10.1007/s10994-022-06202-y

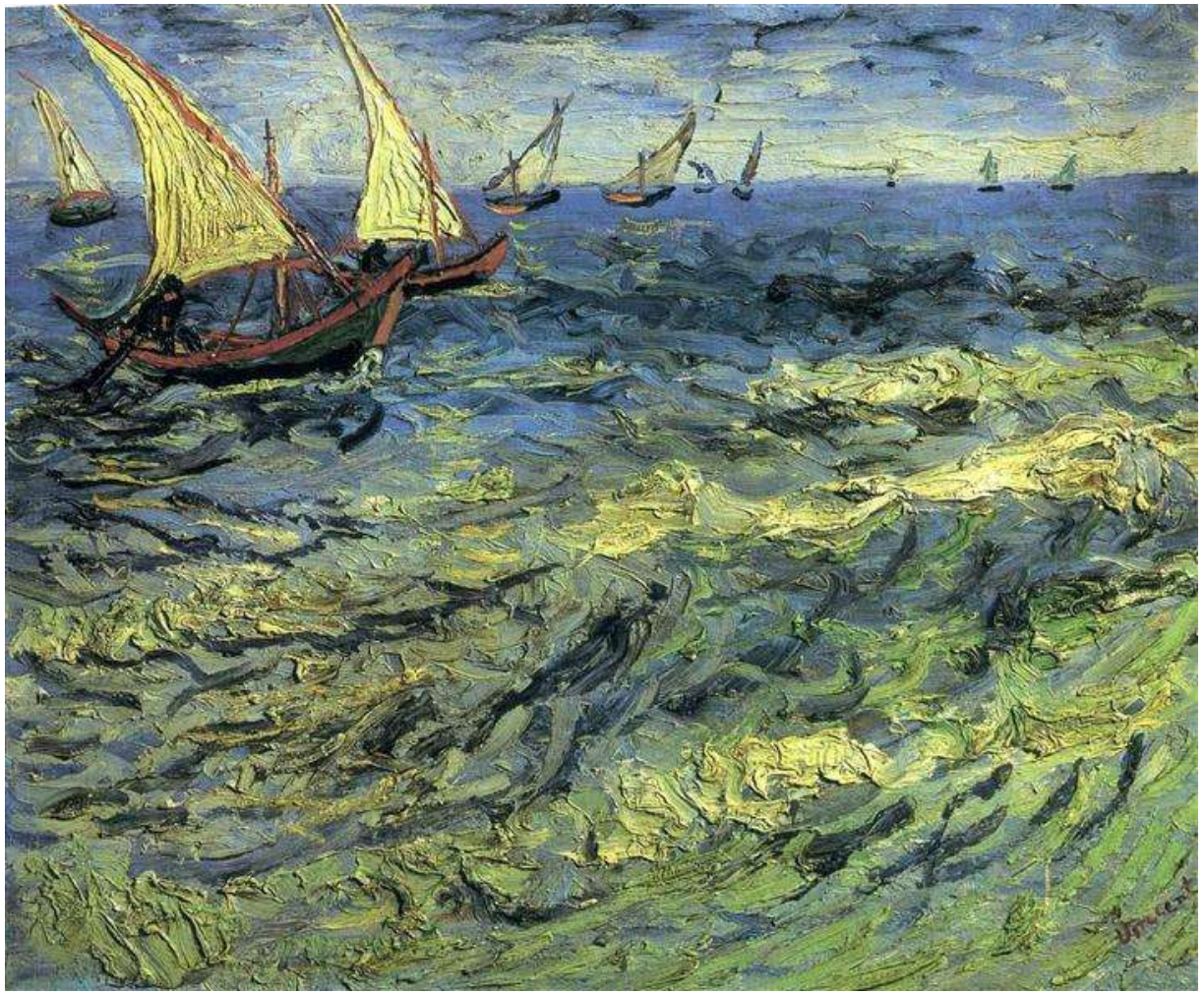
"fairness can be defined using a special framework based on Global Sensitivity Analysis …"

More reading about explainable ML

Lu, Xuefei, and Emanuele Borgonovo. 2026. “How Do I Explain That Decision? A New Approach Combining Counterfactuals and Sensitivity Analyses.” European Journal of Operational Research 334 (1): 200–215. https://doi.org/10.1016/j.ejor.2026.04.035.

“In machine learning, [the idea is]to treat features as players and model predictions as rewards”

Fishing expeditions? Forking paths?



Vincent van Gogh, Fishing Boats at Sea, source: https://www.wikiart.org/

Jorge Luis Borges
(1899-1986)

Taking different narratives within the same novel like Ts'ui Pên

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*

Andrew Gelman† and Eric Loken‡

14 Nov 2013

Many multi-analyst studies ongoing, e.g.:

"Are soccer players with dark skin tone more likely to receive red cards from referees?", Silberzahn et al. 2018

"Is there an impact of compulsory schooling on teenage pregnancy rates?", Huntington-Klein et al. (2021)

Source: BBC https://www.bbc.com/sport/football/68435064

HUNTINGTON-KLEIN, N., ARENAS, A., BEAM, E., BERTONI, M., BLOEM, J. R., BURLI, P., CHEN, N., GRIECO, P., EKPE, G., PUGATCH, T. et al. (2021). The influence of hidden researcher decisions in applied microeconomics. Economic Inquiry 59 944–960.
SILBERZAHN, R., UHLMANN, E. L., MARTIN, D. P., ANSELMI, P., AUST, F., AWTREY, E., BAHNÍK, Š., BAI, F., BANNARD, C., BONNIER, E. et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science 1, 337–356.

Multi-analyst studies ➔ multiverse analysis, though the two are different:

	Multi-Analyst Study	Multiverse Analysis
Unit of variation	Different analysts	Different analytic specifications
Source of variability	Human judgment	Enumerated decision space
Structure	Decentralized	Systematic grid
Scalability	Limited by analyst pool	Limited by comput. Complexity
Epistemic target	Researcher heterogeneity	Specification robustness

➔ **Multi-analyst studies** examine social epistemology — how scientific communities generate variability.

➔ **Multiverse analyses** examine statistical robustness — how conclusions depend on modelling choices.

Defining the multiverse (Simonsohn et al. 2020)

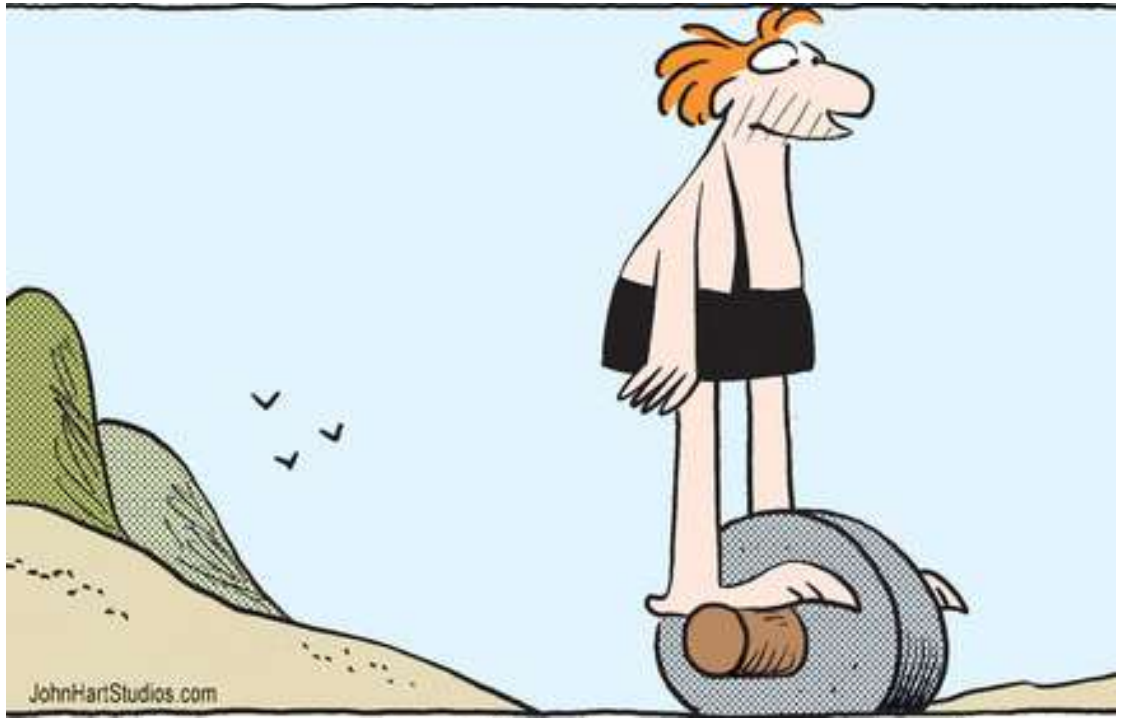
A systematic approach to handling:

- Researchers' degrees of freedom
- Analytic flexibility
- Model specification choice

https://www.istockphoto.com/

Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson. 2020. "Specification Curve Analysis." Nature Human Behaviour 4 (11): 1208–14.

Are different words being used for the same thing?



BC, comic strip by cartoonist Johnny Hart

The Tower of Babel by Pieter Bruegel the Elder (1563). Source: Wikipedia Commons

Wheel rediscovered?

Back to the garden

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*

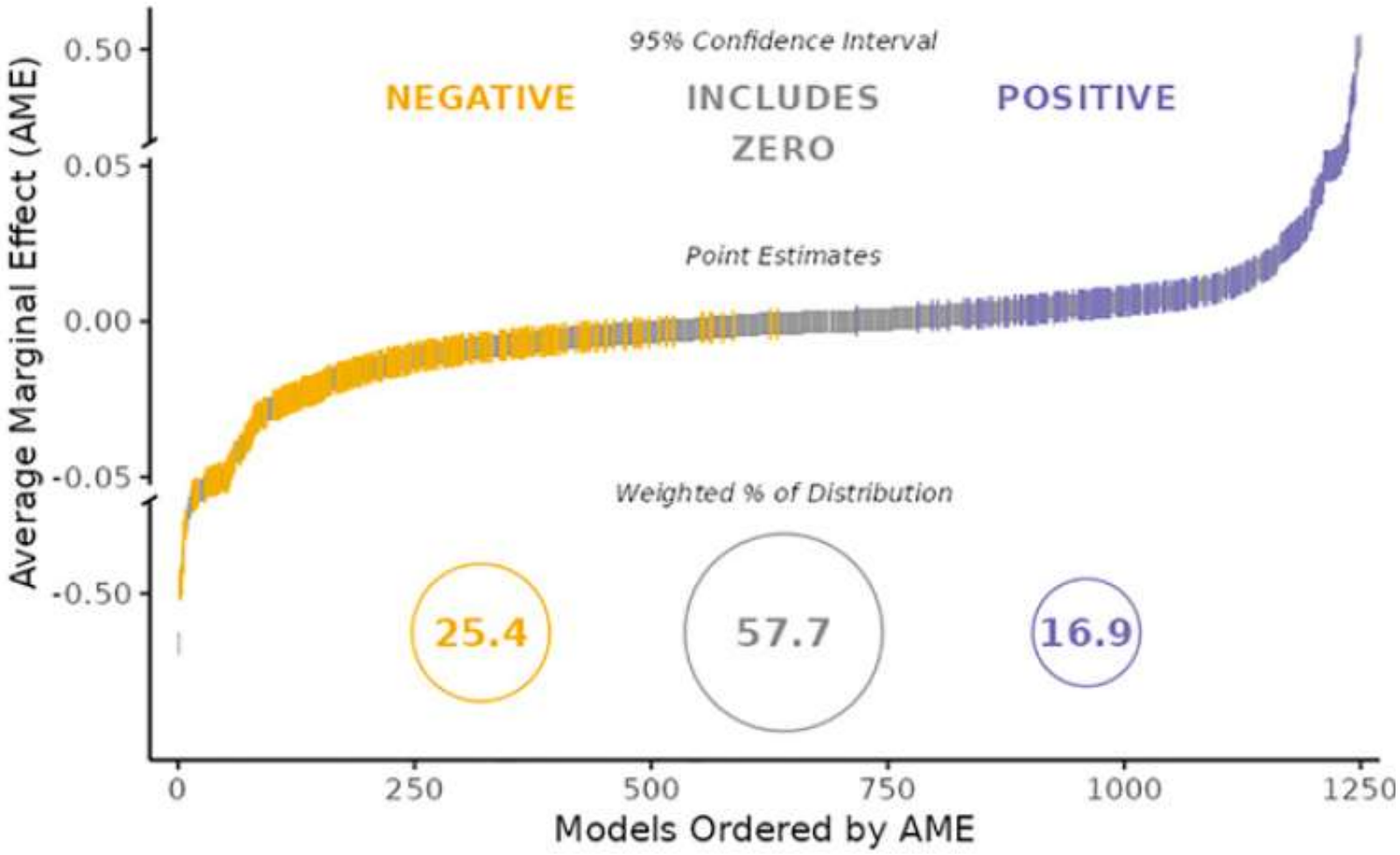
Andrew Gelman† and Eric Loken‡

14 Nov 2013

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OPEN

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022



"Will different researchers [73 teams] converge on similar findings when analysing the same data?

…teams' results varied greatly, ranging from large negative to large positive effects"

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OPEN

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OPEN

Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty

Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022

In the PNAS multi-analyst study the normative (political) orientation of the analysts was found **not** to explain much of the variance

The modelling variables themselves were found to explain at most a few percent of the variance of the results … **where does the variance come from?**

https://www.istockphoto.com/

Simulate different paths using triggers

k_i	Variable	Description
1	*age*	Inclusion or non-inclusion of the covariates age and square age of participants
2	*all − ethnic*	Inclusion or non-inclusion of a covariate that takes into account all ethnic categories
3	*bootstrap*	Inclusion or non-inclusion of the bootstrap procedure
4	*cont − europe*	Inclusion or non-inclusion of continental Europe countries in the analysis
9	*income*	Inclusion or non-inclusion of the covariate related to the income
10	*pop − wb*	Inclusion or non-inclusion of the covariate related to the information about country-specific numerosity of population
11	*rich − dems*	Inclusion or non-inclusion of rich democracies in the analysis
12	*SE*	Inclusion or non-inclusion of clustered standard error
13	*socx − oecd*	Inclusion or non-inclusion of the covariate related to social expenditure
14	*X*1	Inclusion or non-inclusion of the covariates related to the migration indexes
15	*Y*	Choice of policy dependent variable

3 *bootstrap* Inclusion or non-inclusion of the bootstrap procedure
4 *cont − europe* Inclusion or non-inclusion of continental Europe countries in the analysis

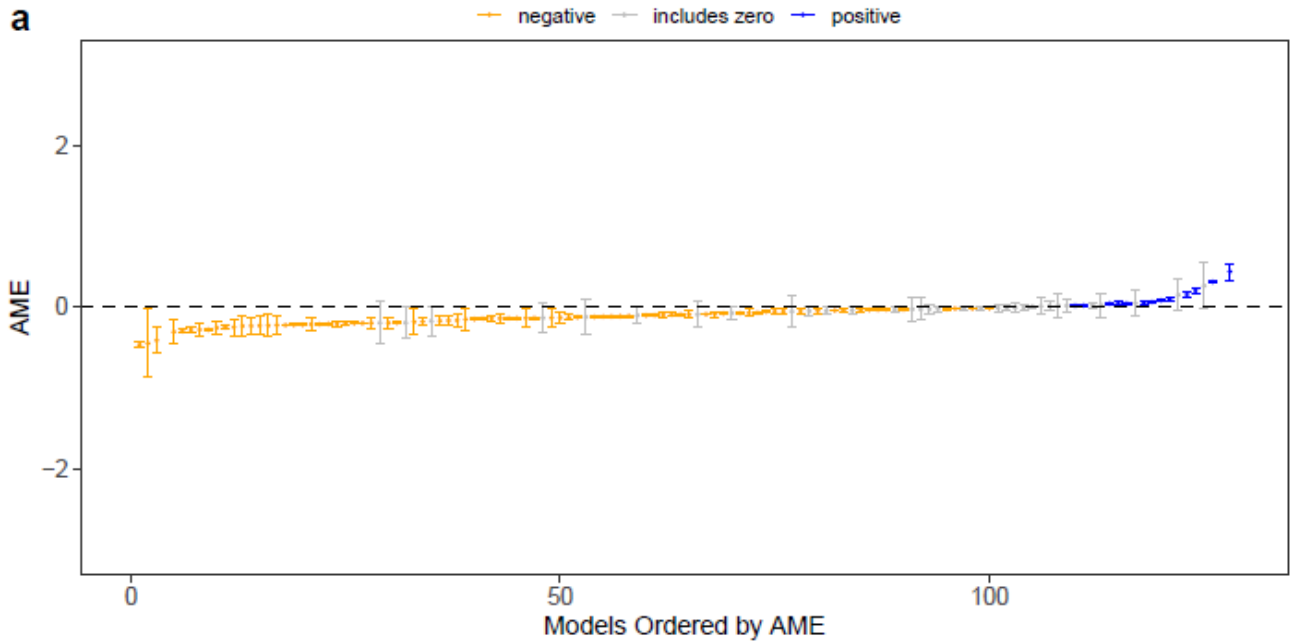
Table 1. Descriptions of variables included in the analysis.

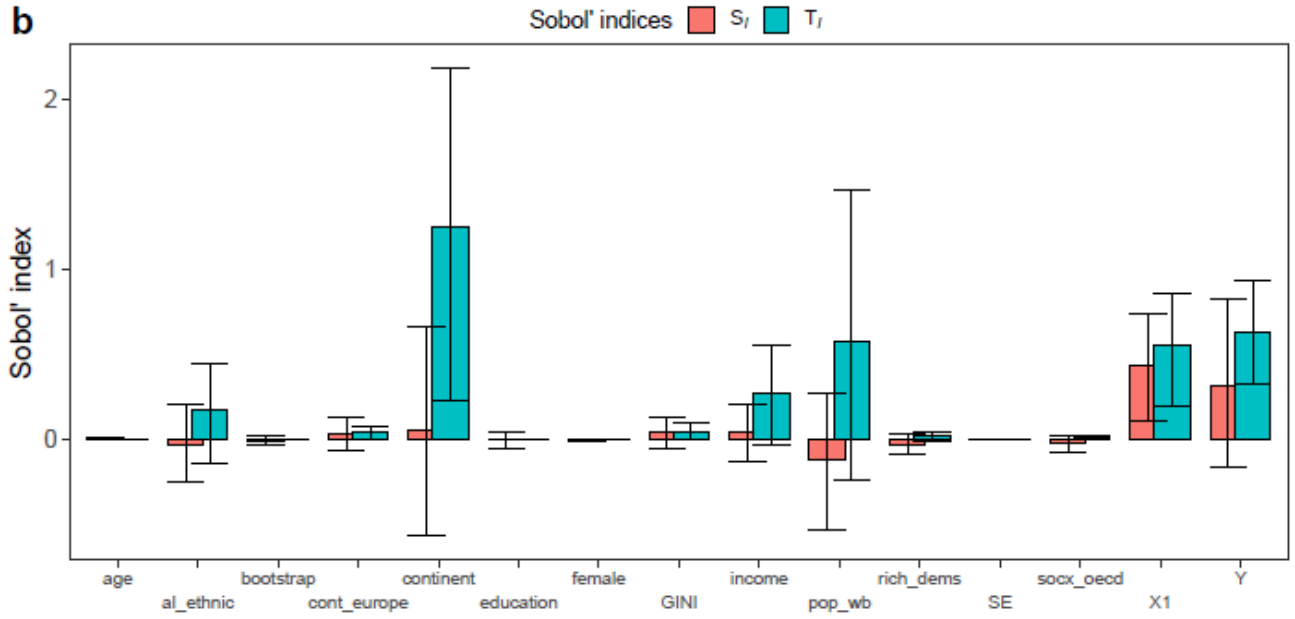
Leamer Reloaded: A Long-Neglected Suggestion Is Rediscovered To Apply Global Sensitivity Analysis To A Quantification Before Publishing It

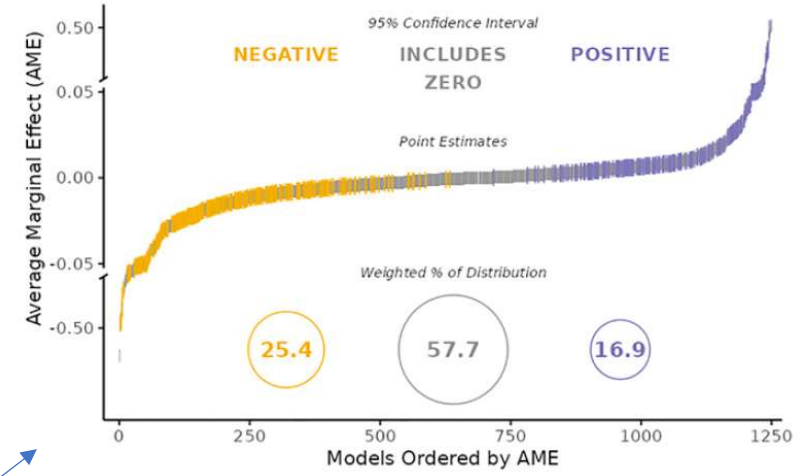
Andrea Saltelli*[1,2], Alessio Lachi[3,4], Arnald Puy[5], and Nate Breznau[6]

OSF

https://osf.io/preprints/metaarxiv/sq34n_v1



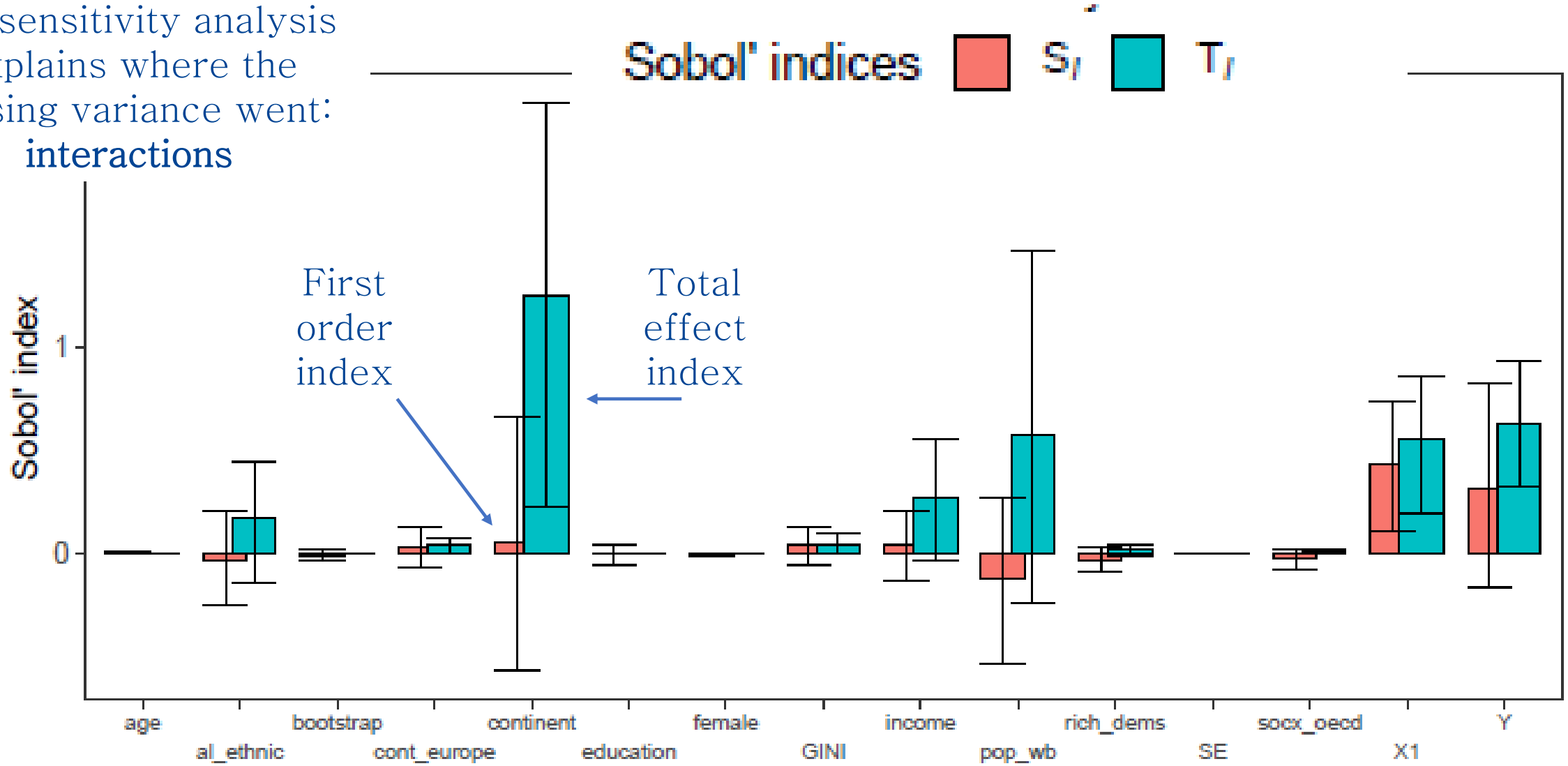




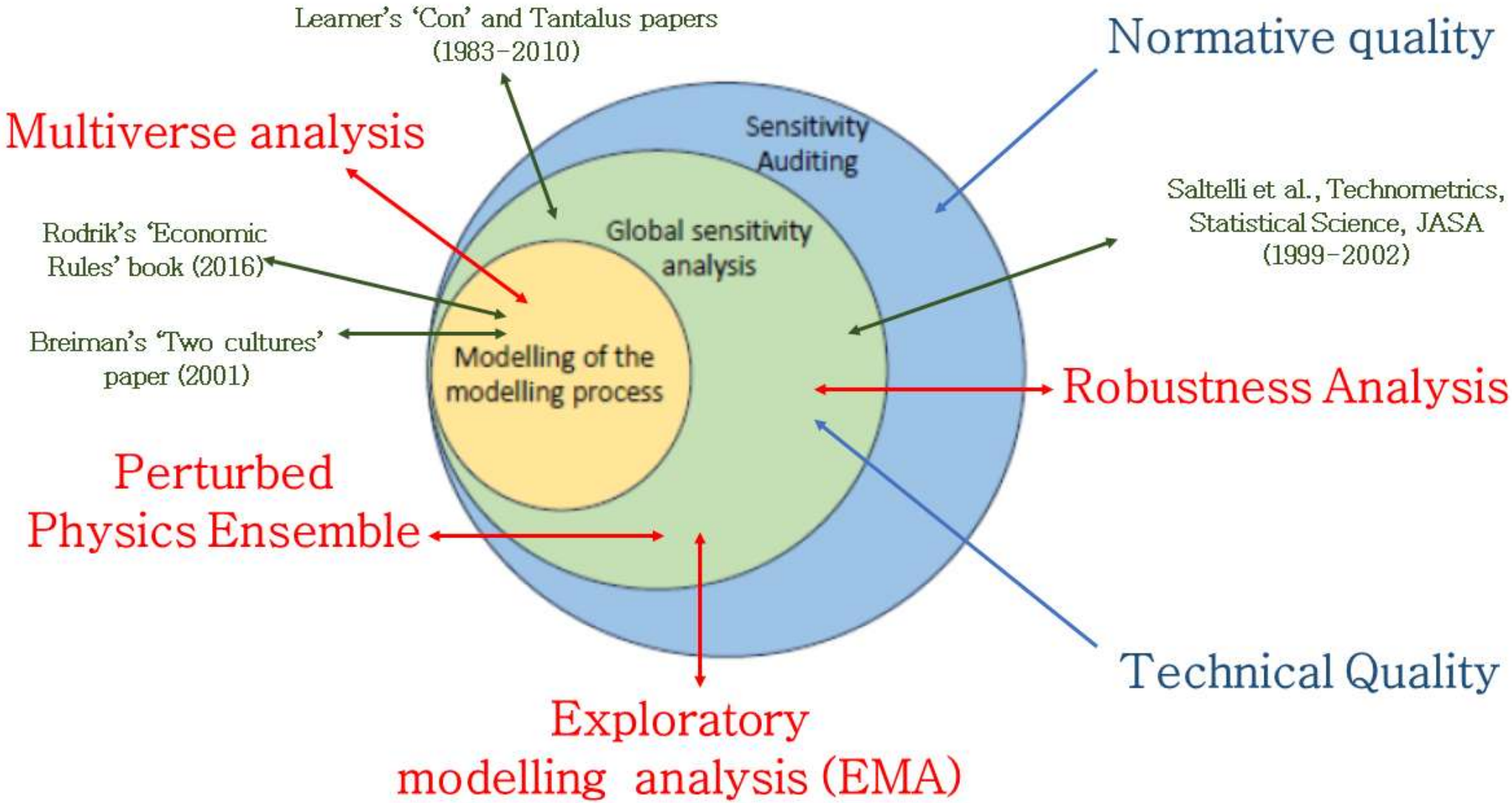
Uncertainty analysis broadly reproduces the results of the multi-analyst experiment

Our sensitivity analysis explains where the missing variance went: **interactions**

Our sensitivity analysis explains where the missing variance went: **interactions**



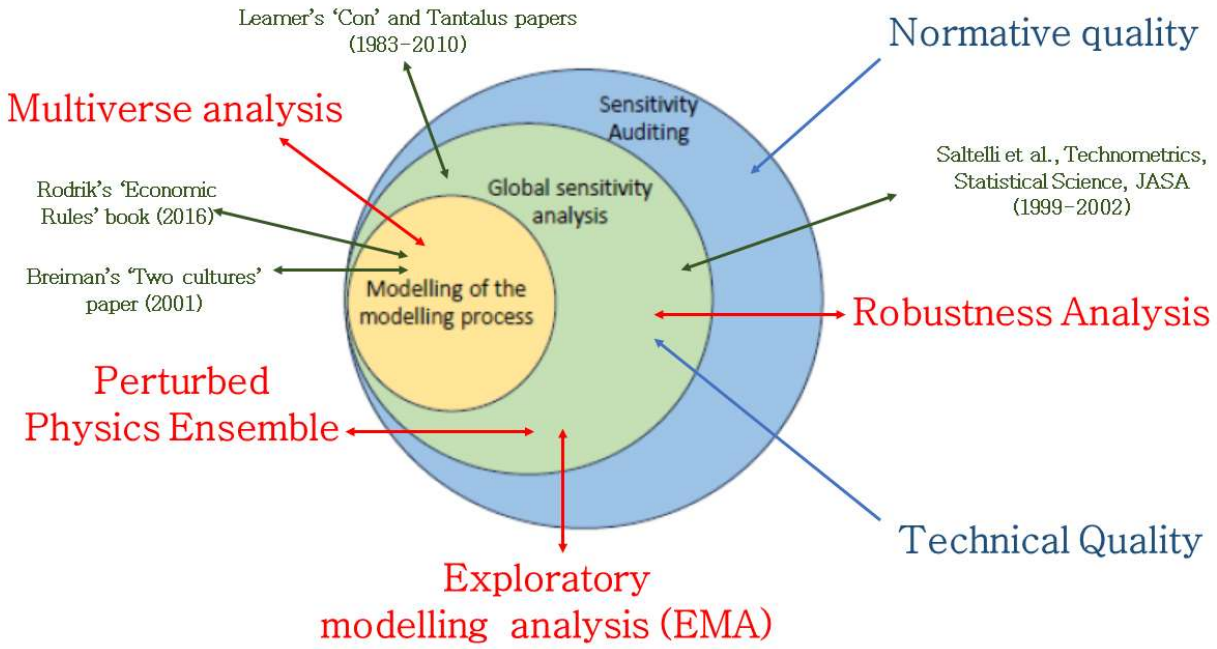
Part IV – GSA and its neighbourhoods


Learner's 'Con' and Tantalus papers (1983-2010)
Normative quality
Multiverse analysis
Sensitivity Auditing
Saltelli et al., Technometrics, Statistical Science, JASA (1999-2002)
Rodrik's 'Economic Rules' book (2016)
Global sensitivity analysis
Breiman's 'Two cultures' paper (2001)
Modelling of the modelling process
Robustness Analysis
Perturbed Physics Ensemble
Technical Quality
Exploratory modelling analysis (EMA)

Exploratory modelling analysis (RAND corporation)

Explore scenarios and simulate

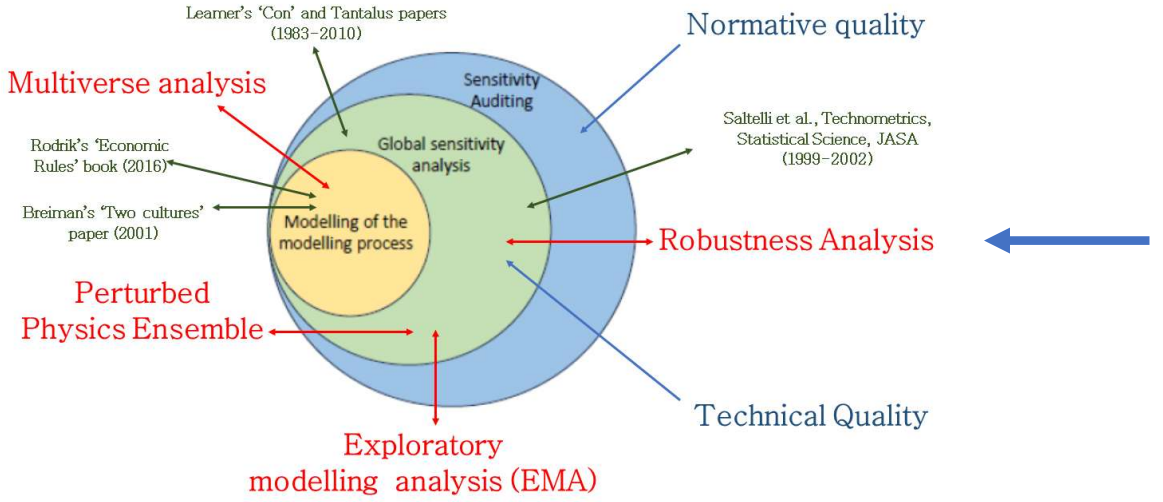
(Too) liberally



Jan H. Kwakkel and Erik Pruyt. Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty. Technological Forecasting and Social Change, 80(3):419–431, March 2013.

Robustness analysis (Houkes et al., 2024)

parameter RA ➔ dependence on model parameters
structural RA ➔ dependence on structural assumption
representational RA ➔ model philosophy, e.g., moving from a math / differential equations to agent-based modelling

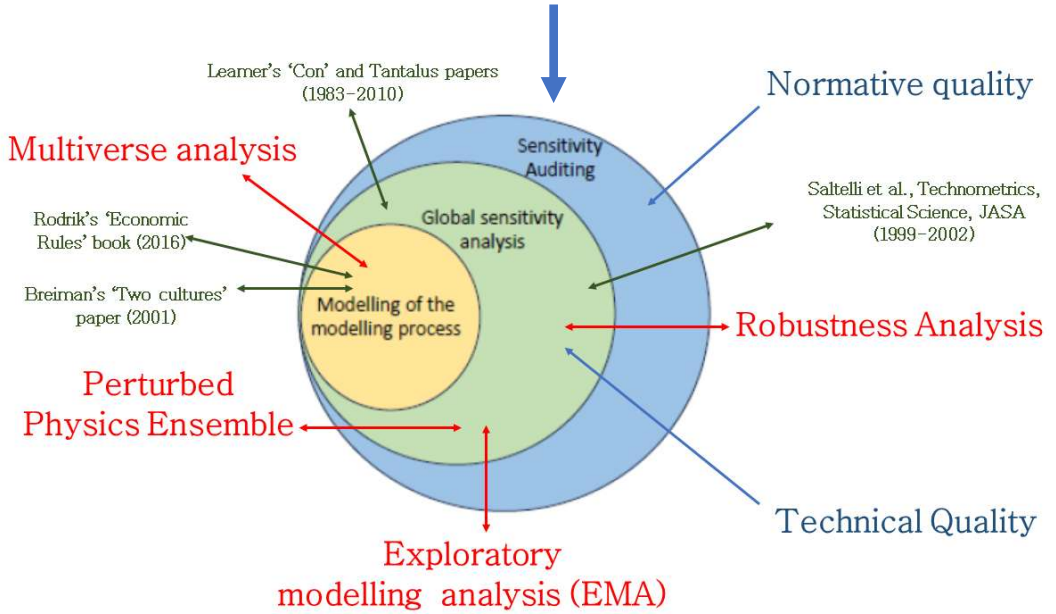


Wybo Houkes, Dunja Šešelja, and Krist Vaesen. Robustness Analysis. In Tarja Knuuttila, Natalia Carrillo, and Rami Koskinen, editors, The Routledge Handbook of Philosophy of Scientific Modeling, pages 195–207. Routledge, Abingdon, Oxon ; New York, NY, 2024.

Modelling of the modelling process and multiverse analysis tells us how sensitive conclusions are to the paths taken …

… But was the right garden entered in the first place?

Sensitivity auditing: "Sensitivity auditing is a wider consideration of the effect of all types of uncertainty, including structural assumptions embedded in the model, and subjective decisions taken in the framing of the problem" (European Commission, 2021).



European Commission. November 2021. "Better Regulation: Guidelines and Toolbox." https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how/better-regulation-guidelines-and-toolbox_en

Int. J. Foresight and Innovation Policy, Vol. 9, Nos. 2/3/4, 2013 213

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

ELSEVIER

Energy Research & Social Science
Volume 127, September 2025, 104296

Original research article

Energy policy-making in the European Union between past and present ☆

Samuele Lo Piano [a b], Andrea Saltelli [b c]

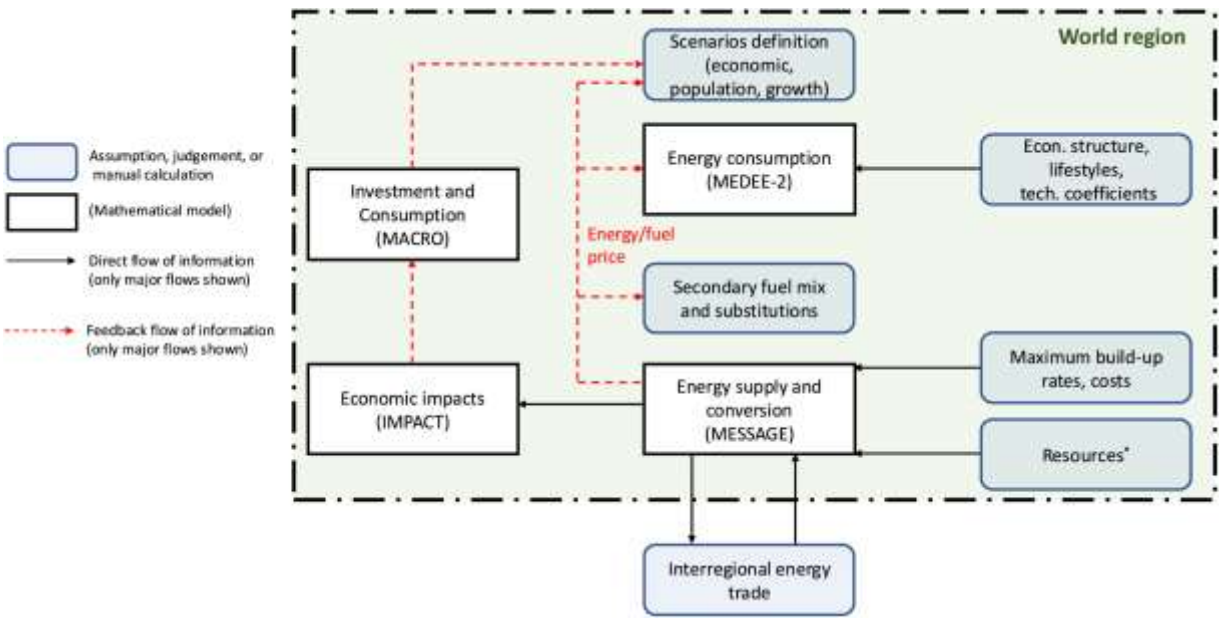
Show more

A recent application to show that old pathologies in modelling persist

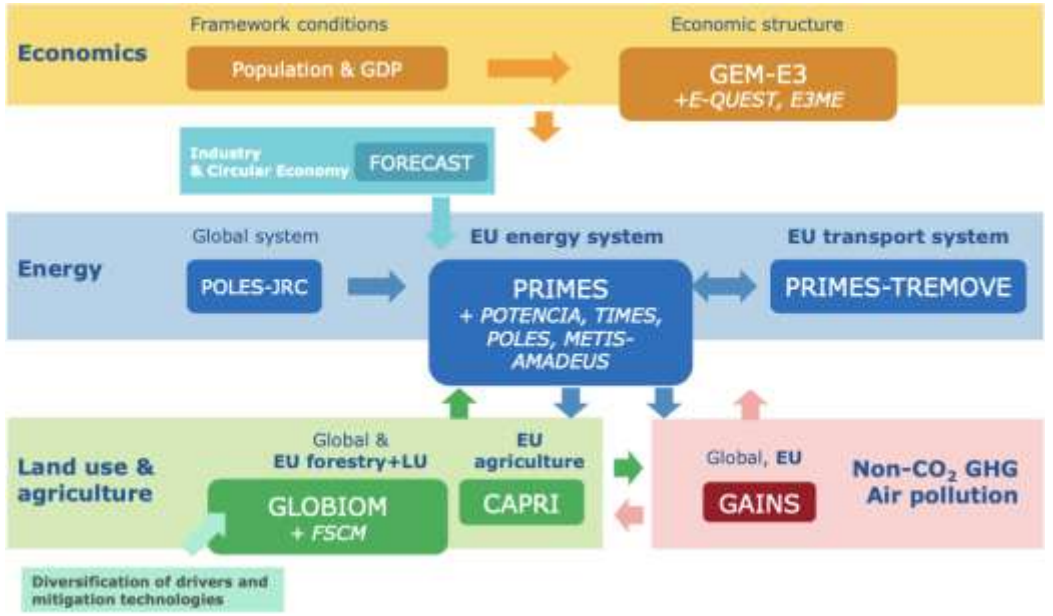
upf. BARCELONA SCHOOL OF MANAGEMENT

- We contrast a 1980s energy study and a 2024 EU impact assessment on emissions.
- Auditing reveals persistent flaws when models are used to inform policy.

IIASA 1980's



EC 2024



The rhetoric of the concatenated suite of models

Home › Policy Sciences › Article

The institutional context of science, models, and policy: The IIASA energy study

Published: September 1984

Volume 17, pages 277–320, (1984) Cite this article

The same way the Digital Twins of the planet are ‘scientifically prescribing’ today …

Brian Wynne (and others such as William Keepin) debunked in the early 80’s the off-the-mark model-based energy future, declared as ‘scientifically prescribed’ by analysts at IIASA …

A fast breeder reactor in the Netherlands, today an amusement park

See a summary here

ELSEVIER

Energy Research & Social Science

Volume 127, September 2025, 104296

Original research article

Energy policy-making in the European Union between past and present ☆

Samuele Lo Piano [a b], Andrea Saltelli [b c]

Show more

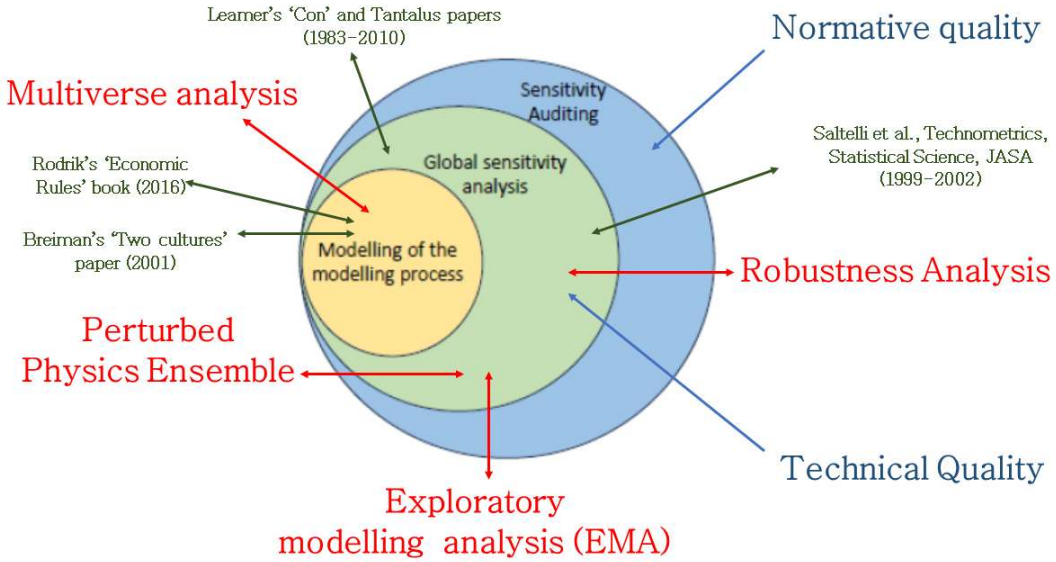
Applying sensitivity auditing to different cases:

- Nutrition and Public Health
- Sustainability, Ecological Footprint (EF)
- Global Hydrological Models
- Sociohydrology, Coupled Human and Water Systems (CHAWS)
- Food Security
- Education, Programme for International Student Assessment (PISA) test

The same pattern appears:

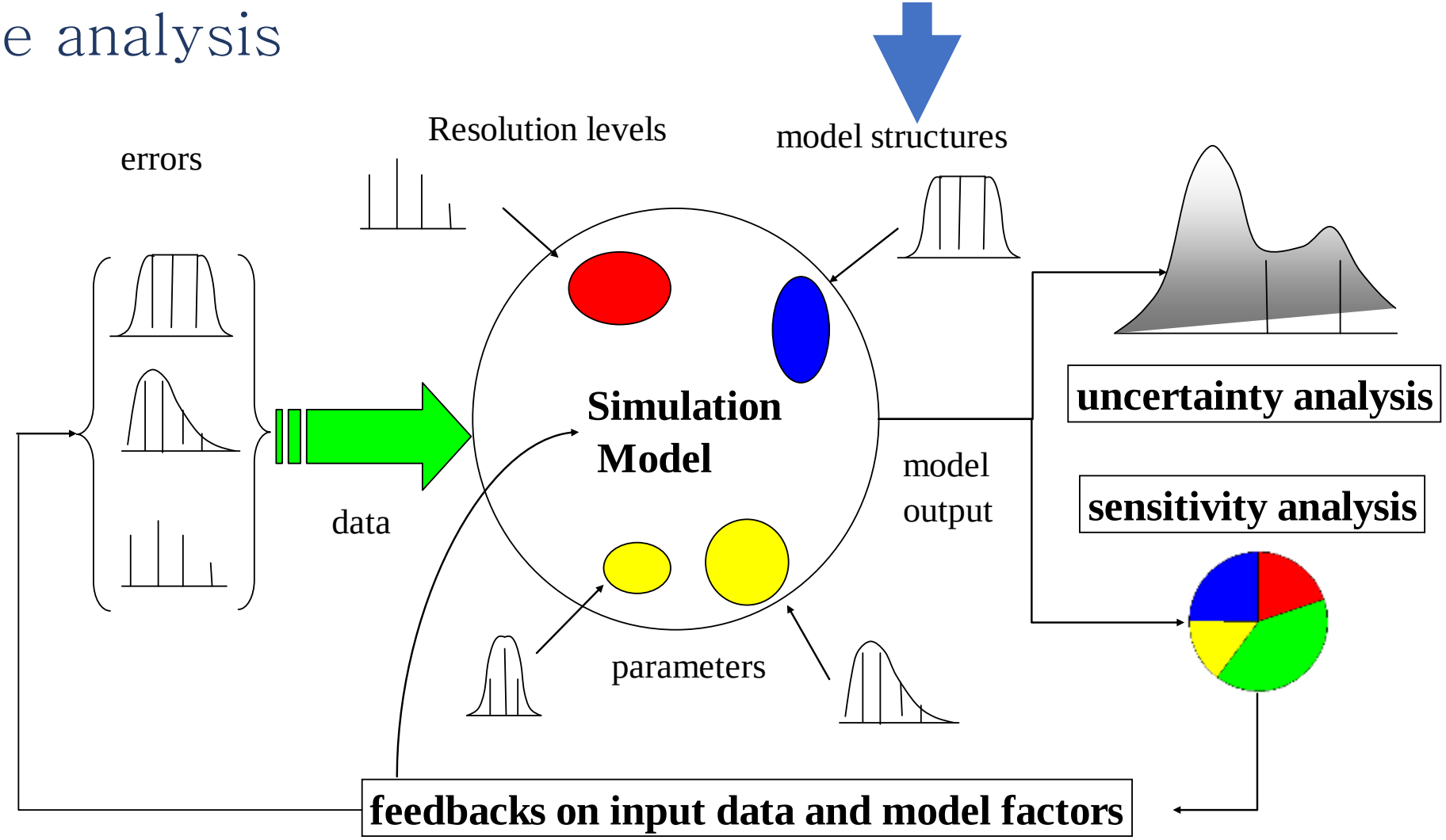
- the most consequential choices are often not numerical;
- they concern problem framing, indicator selection, omitted variables, boundary choices, and implicit values

Perturbed parameter (or physics) ensemble: same as uncertainty quantification and sensitivity analysis when done by physicists in climate modelling



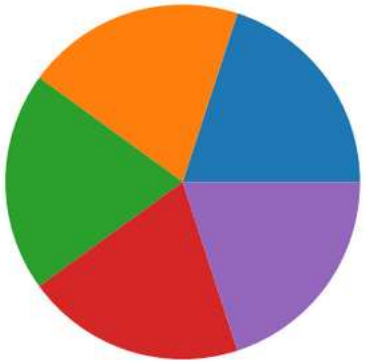
John W. Rostron, David M. H. Sexton, Carol F. McSweeney, Kuniko Yamazaki, Timothy Andrews, Kalli Furtado, Mark A. Ringer, and Yoko Tsushima. The impact of performance filtering on climate feedbacks in a perturbed parameter ensemble. Climate Dynamics, 55(3):521–551, August 2020. doi:10.1007/s00382-020-05281-8

When sampling modelling choices via 'triggers', global sensitivity analysis becomes a modelling of the modelling process = multiverse analysis



Example of using triggers: an integer in [1-5] is used to select one out of 5 possible evapotranspiration formulae at each model run

Trigger=an integer variable, whose value is determined via 'Russian roulette' in Monte Carlo jargon



nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature communications › comment › article

Comment | Open access | Published: 08 June 2022

The delusive accuracy of global irrigation water withdrawal estimates

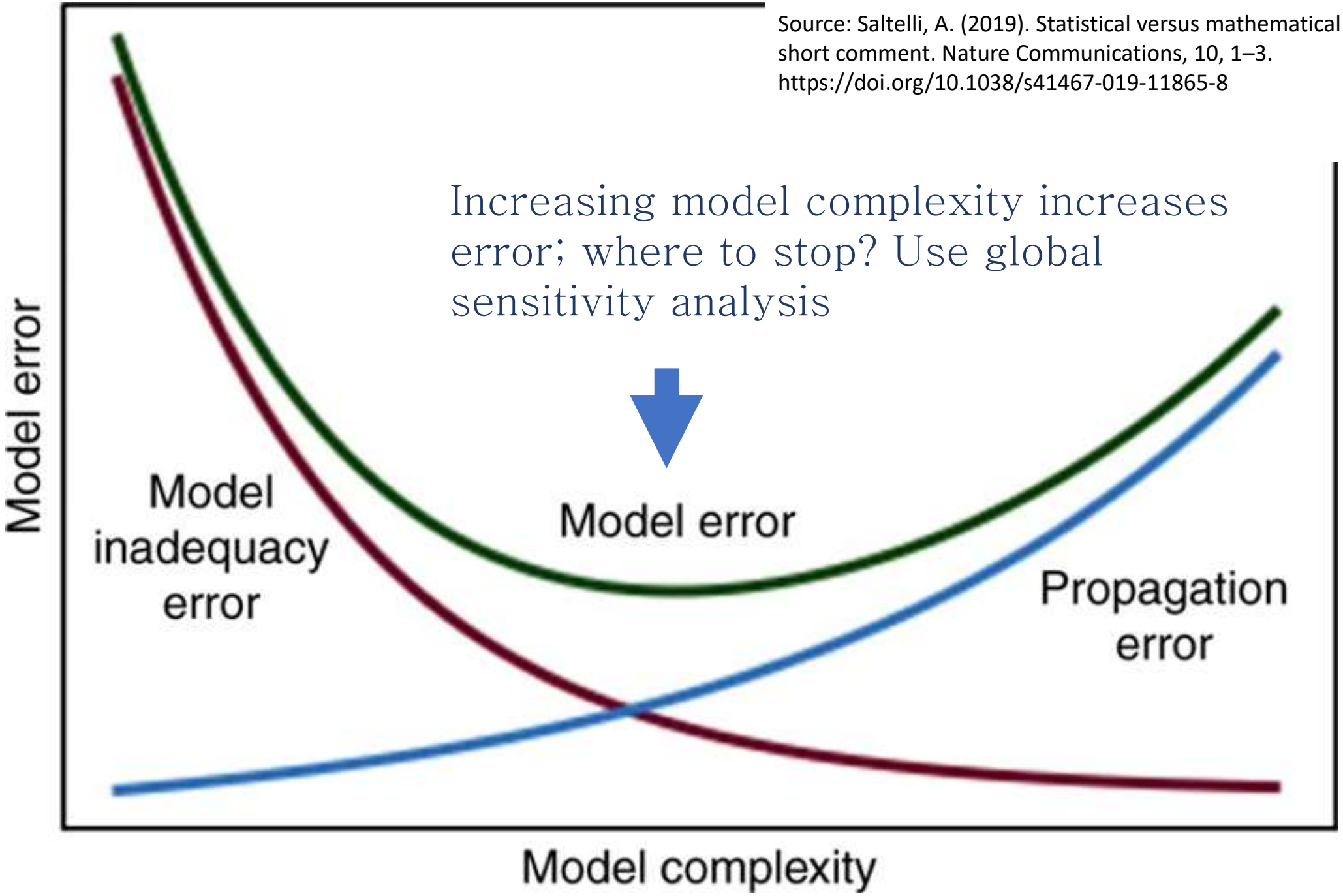
Arnald Puy ✉, Razi Sheikholeslami, Hoshin V. Gupta, Jim W. Hall, Bruce Lankford, Samuele Lo Piano, Jonas Meier, Florian Pappenberger, Amilcare Porporato, Giulia Vico & Andrea Saltelli

Nature Communications **13**, Article number: 3183 (2022) | Cite this article

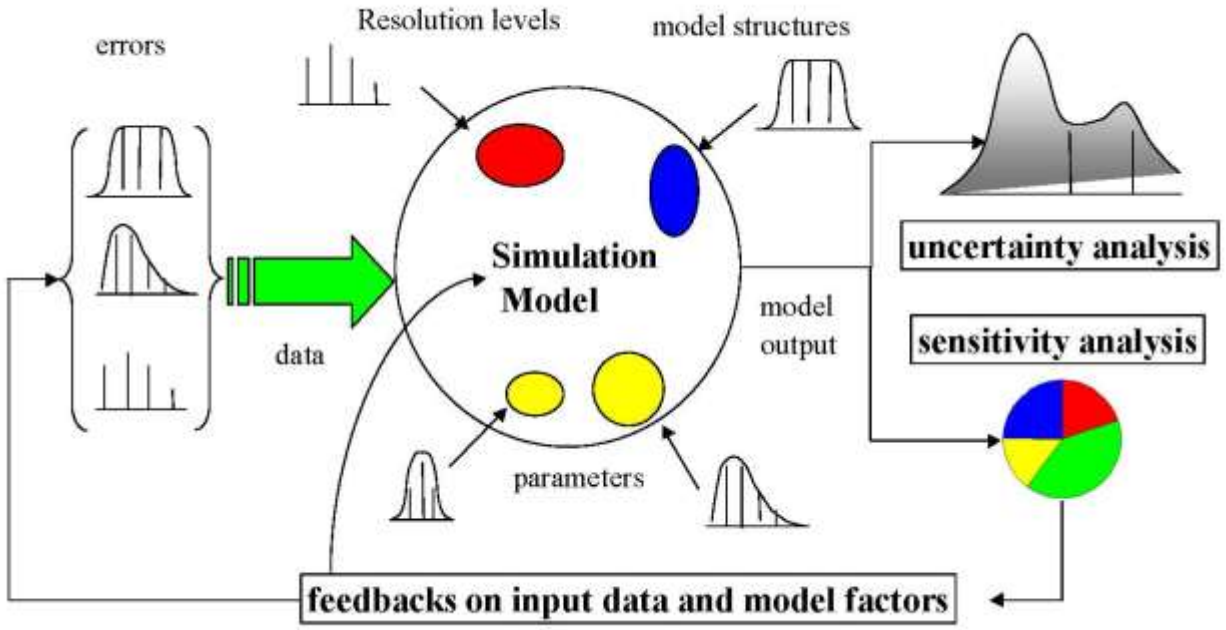
10k Accesses | **79** Citations | **93** Altmetric | Metrics

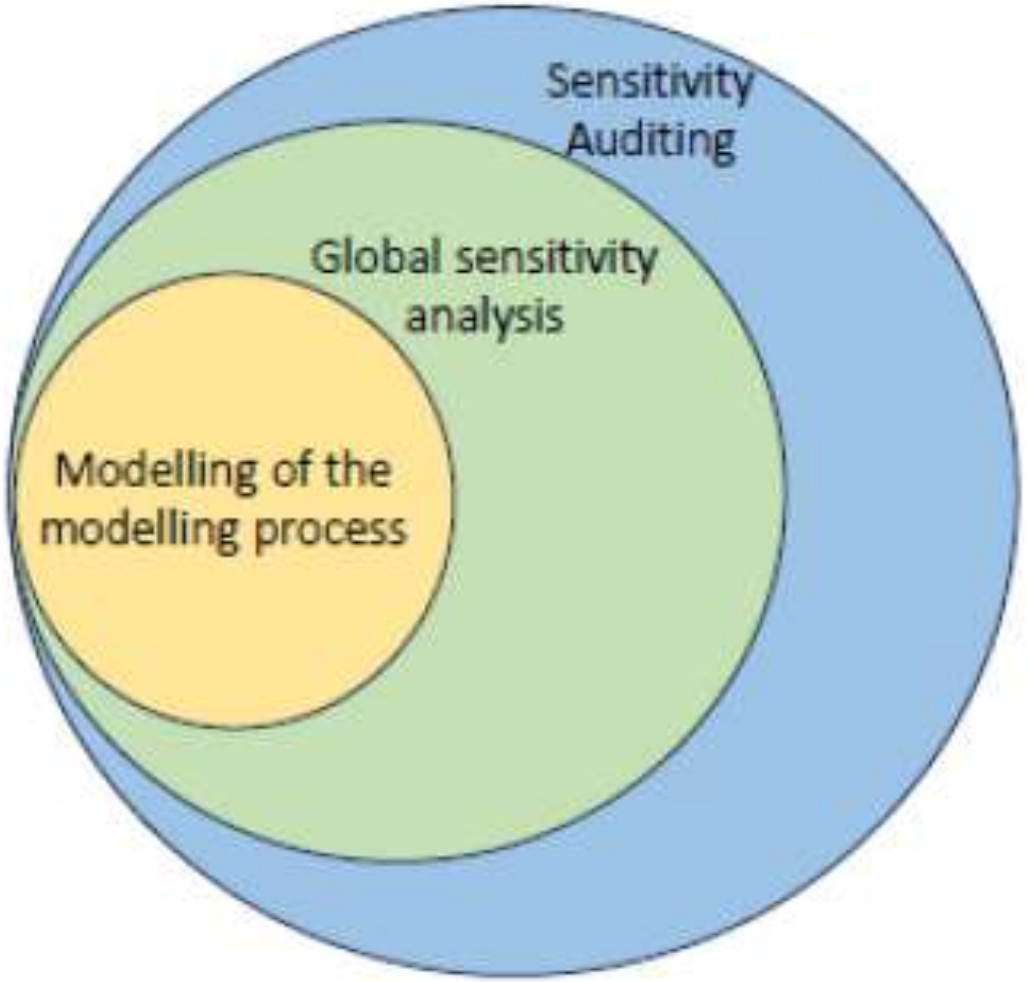
Russian roulette as depicted in the 1925 film The Night Club. Source: Wikipedia Commons

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



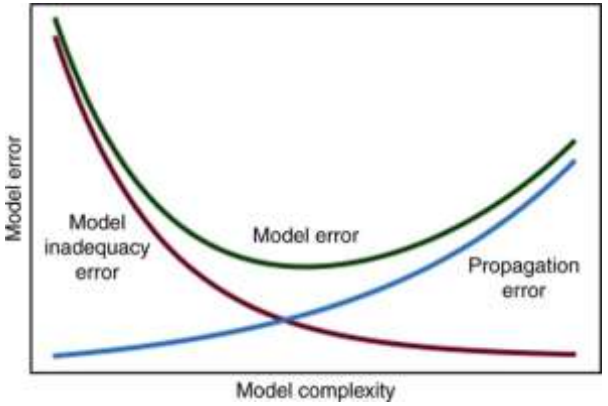
The End


errors
Resolution levels
model structures
Simulation Model
data
model output
uncertainty analysis
sensitivity analysis
parameters
feedbacks on input data and model factors


Sensitivity Auditing
Global sensitivity analysis
Modelling of the modelling process

Sensitivity analysis and auditing can assist sociology of quantification in activities of de- and re-construction (e.g. for statactivists, critical friends, counter-modellers), to

➔ retrace what was assumed
➔ check the level of complexity…



Christoph Becker. Insolvent: How to Reorient Computing for Just Sustainability. The MIT Press, Cambridge, Massachusetts London, England, June 2023.
Ehsan Nabavi. Modelling as intervention technology: Science, politics, and water conflicts. Water Alternatives, 18(2):A18–2–10, 2025

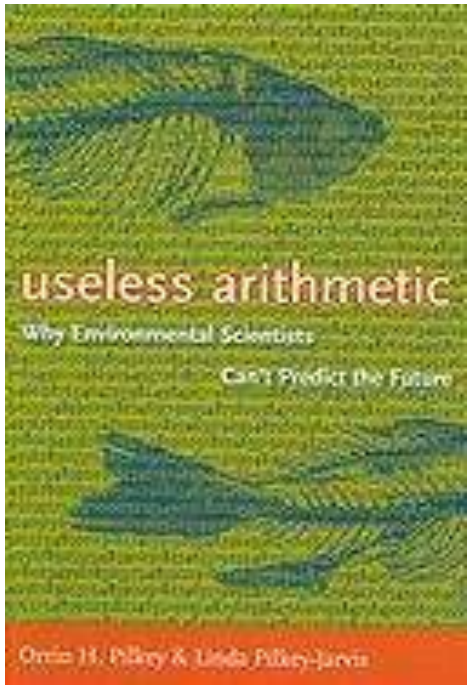
Part V GSA’s dos and don'ts

Don’t confuse the map with the territory

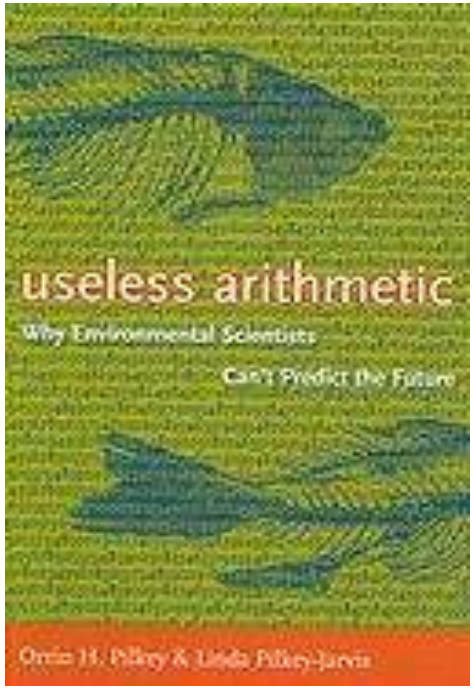
If you do, sensitivity analysis will not save you

<<It is important, however, to recognize that the sensitivity of the parameter in the equation is what is being determined, not the sensitivity of the parameter in nature>>

Orrin H. Pilkey



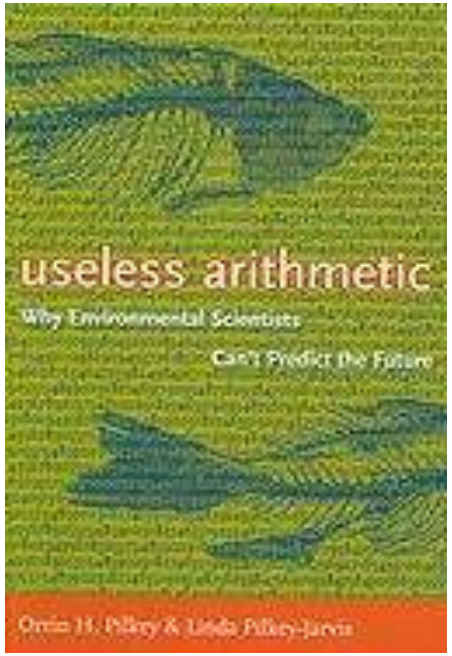
Useless Arithmetic: Why Environmental Scientists Can't Predict the Future
by Orrin H. Pilkey and Linda Pilkey-Jarvis, Columbia University Press, 2009.

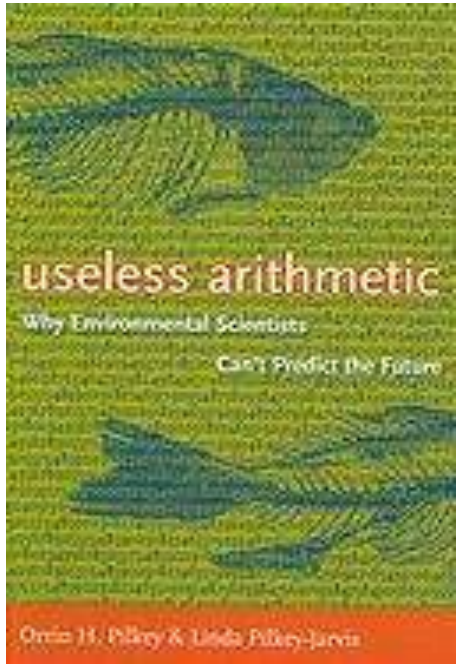


<<…If the model is wrong or if it is a poor representation of reality, determining the sensitivity of an individual parameter in the model is a meaningless pursuit>>

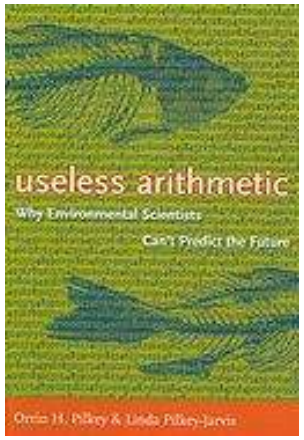
One of the examples discussed concerns the Yucca Mountain repository for radioactive waste. TSPA model (for total system performance assessment) for safety analysis.

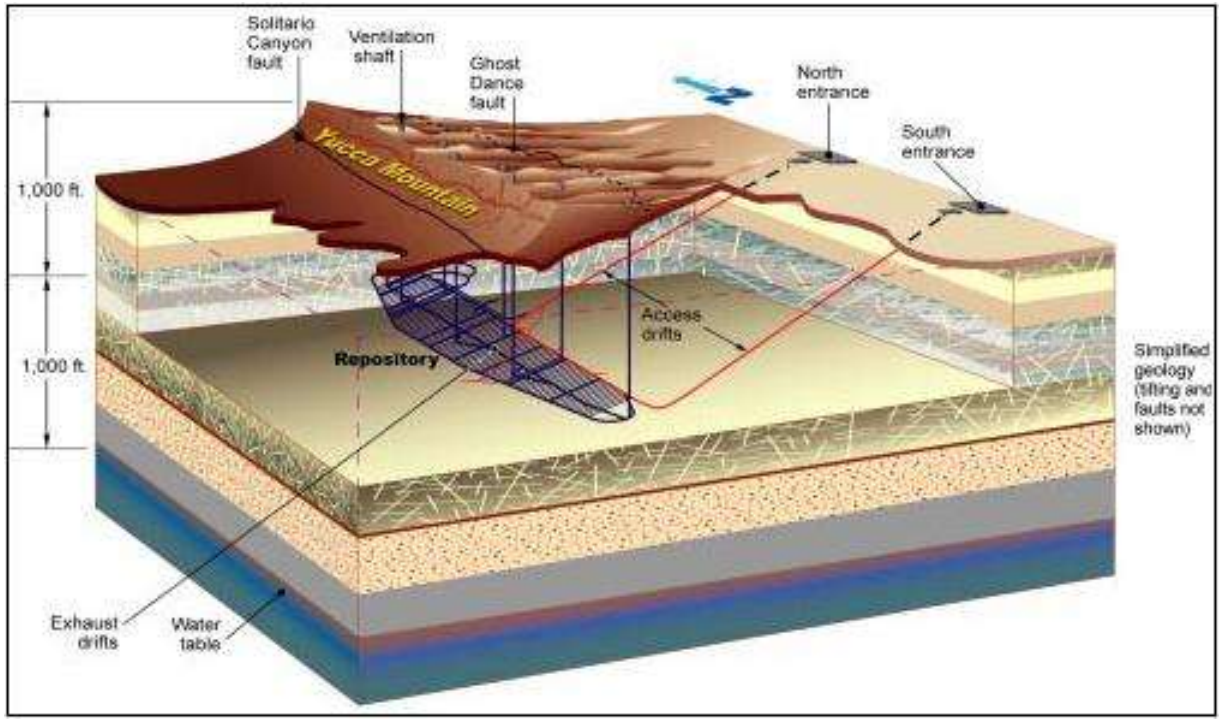
TSPA is Composed of 286 sub-models.





TSPA (like any other model) relies on assumptions → one is the low permeability of the geological formation → long time for the water to percolate from surface to disposal

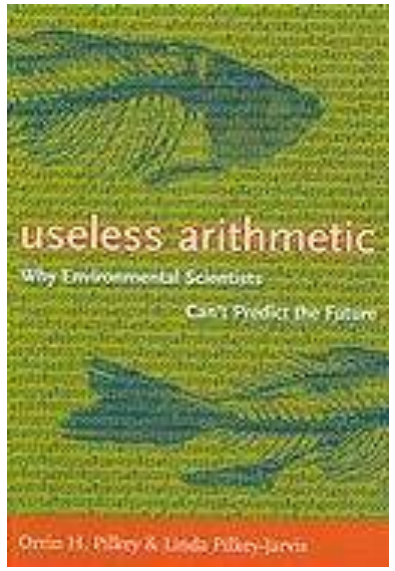




The confidence of the stakeholders in TSPA was not helped when evidence was produced which could lead to an upward revision of 4 orders of magnitude of this parameter
(the ^{36}Cl story)

In the case of TSPA (Yucca mountain) a range of 0.02 to 1 millimetre per year was used for percolation of flux rate.

→… SA useless if it is instead ~ 3,000 millimetres per year.



"Scientific mathematical modelling should involve constant efforts to falsify the model"

➔ Organized skepticism (as per CUDOS)

Communalism, Universalism, Disinterestedness, Organized Skepticism, from sociology of science, Robert K. Merton.

Beware the size of your model

Mind the conjecture of O'Neil

Sensitivity analysis can:

- surprise the analyst,
- uncover technical errors in the model,
- identify critical regions in the space of the inputs,
- avoid type II errors …

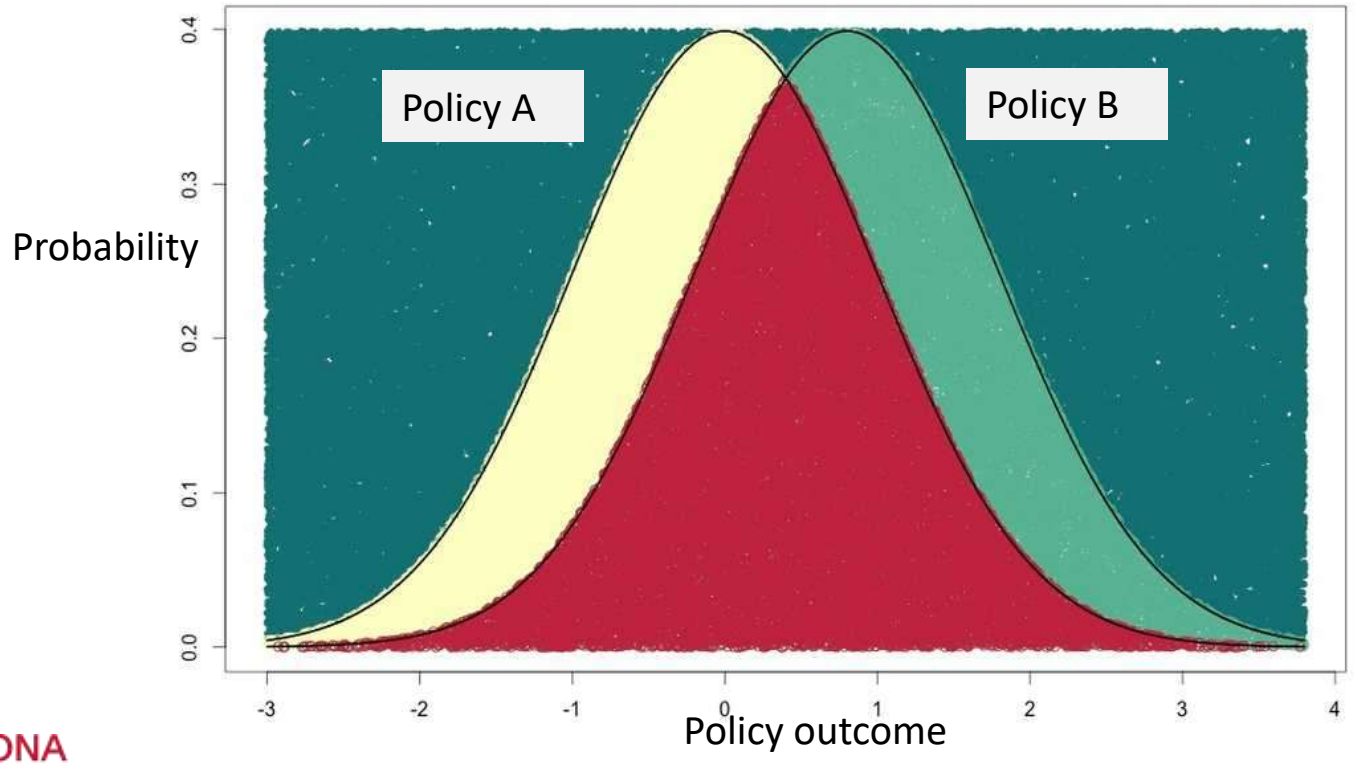
Source: The Simpson, 20th Television Animation (The Walt Disney Company)

Sensitivity analysis can :

surprise the analyst,
uncover technical errors in the model,
identify critical regions in the space of the inputs,
avoid type II errors
establish priorities for research,
simplify models
falsify models (show that a model is false or irrelevant)
defend against your own model being falsified

Sensitivity analysis can:

verify whether policy options (or marketing strategies) can be distinguished from one another given the uncertainties in the system, …



BARCELONA SCHOOL OF MANAGEMENT

What method would one choose to perform sensitivity analysis?

Source: iStock by Getty images

What method would one choose to perform sensitivity analysis?

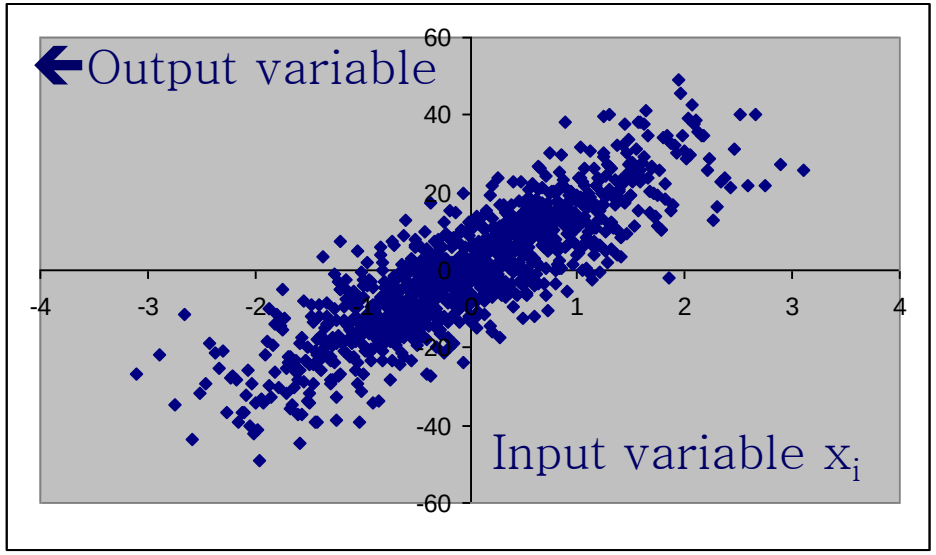
Many of the sensitivity analysis seen in the literature are local or otherwise OAT (One factor At a Time)

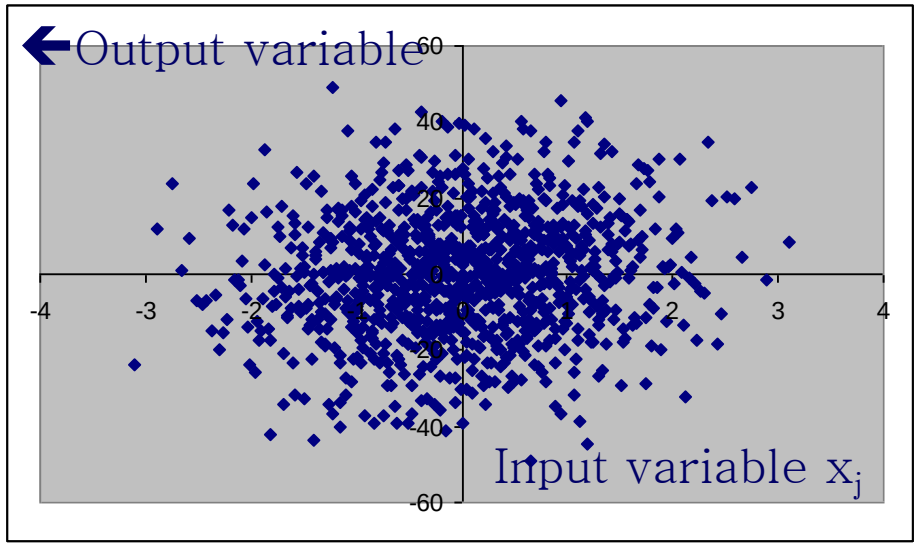
$$y = f(x_1, x_2, \dots x_k)$$

$$\left.\frac{\partial y}{\partial x_i}\right|_{x_i = x_i^0} \longleftarrow \text{Local}$$

Beyond ‘one factor at a time’

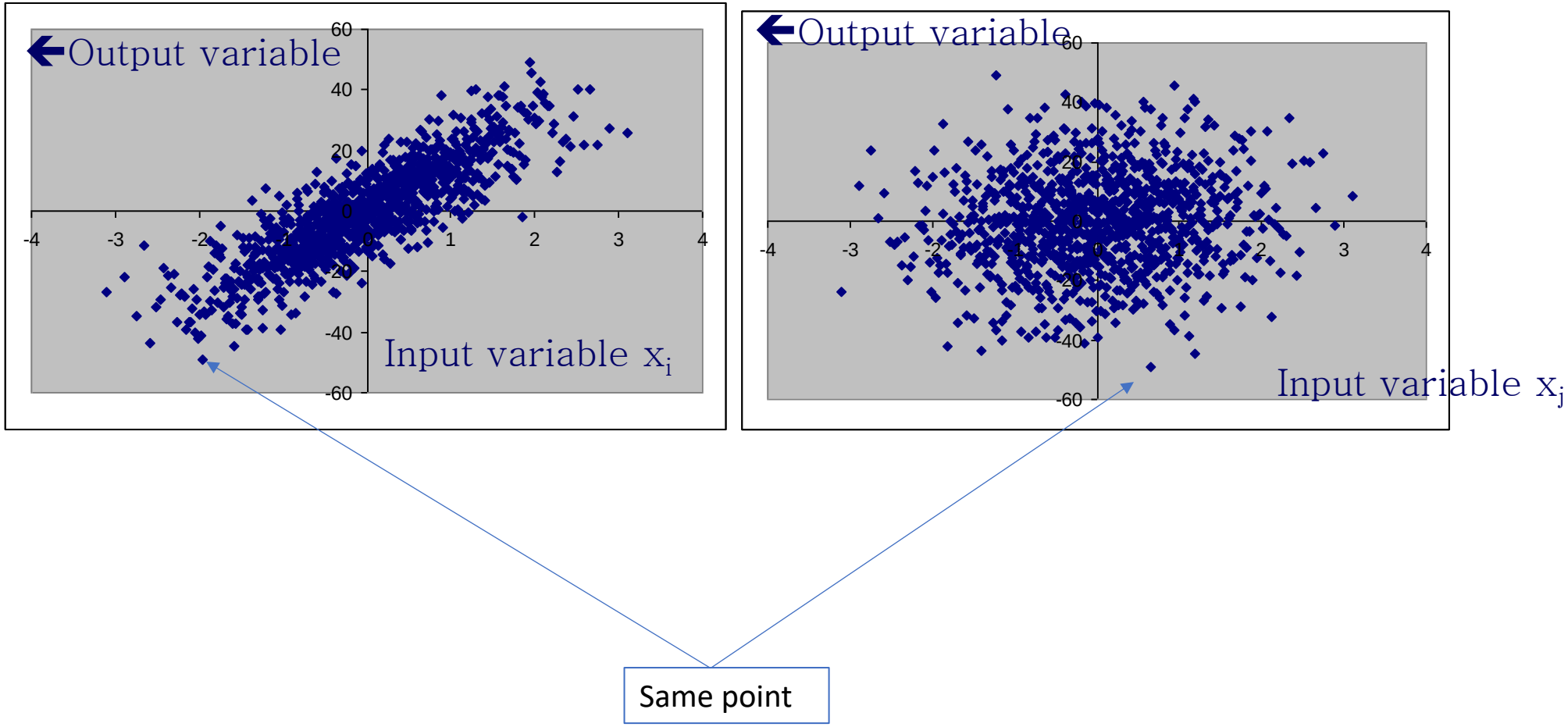
An introduction to variance based methods



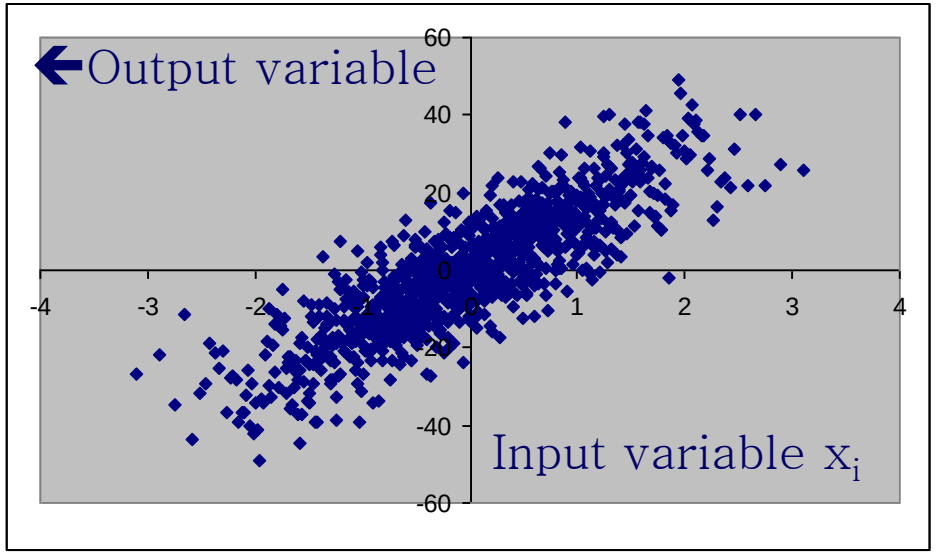


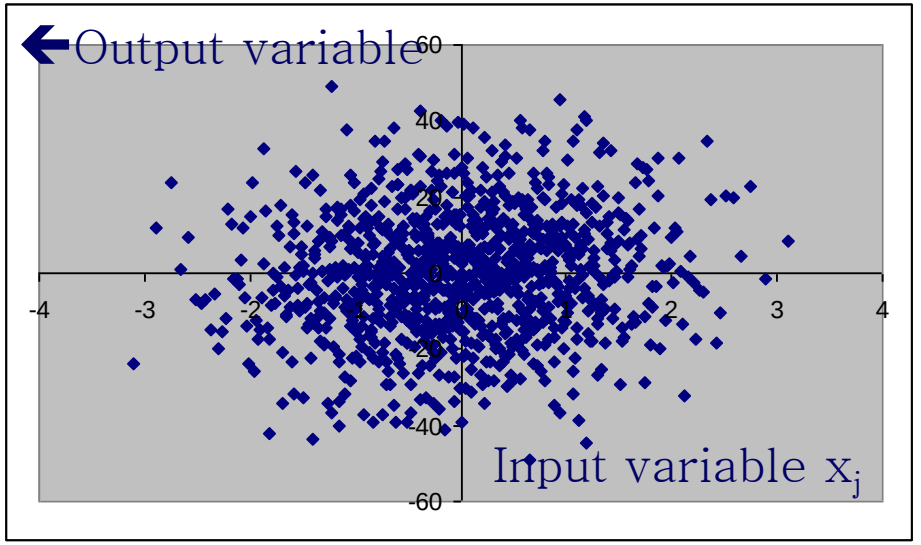
Plotting the output as a function of two different input factors:

The model has been run 1,000 times sampling all uncertain factors; two of them are shown here, x_i (left) and x_j (right). The output variable has first been sorted by increasing x_i then by increasing x_j.



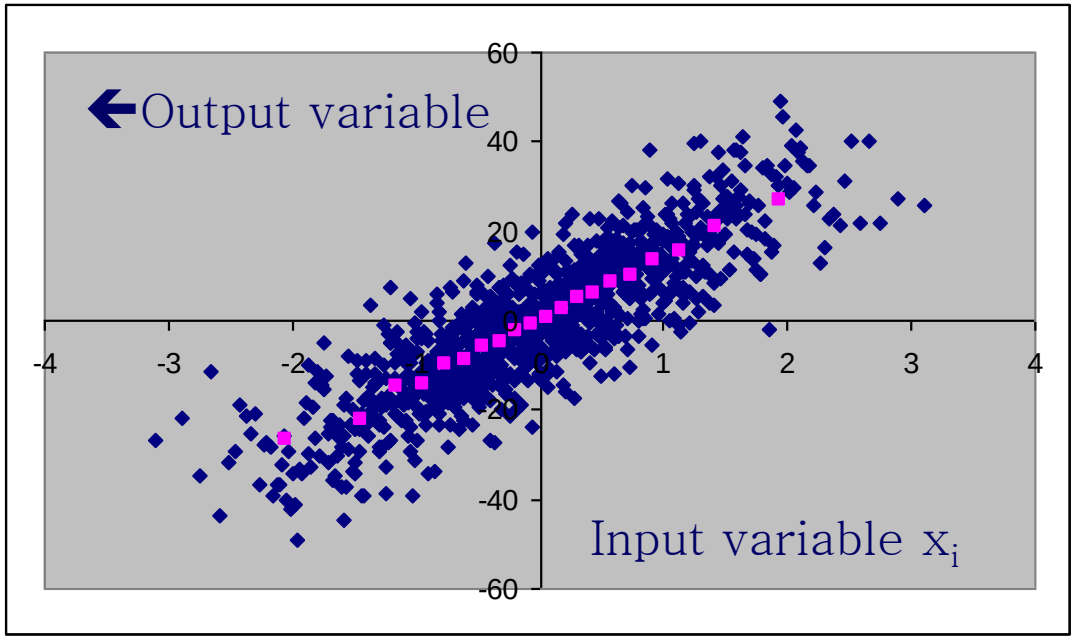
The points are always the same, only sorted differently

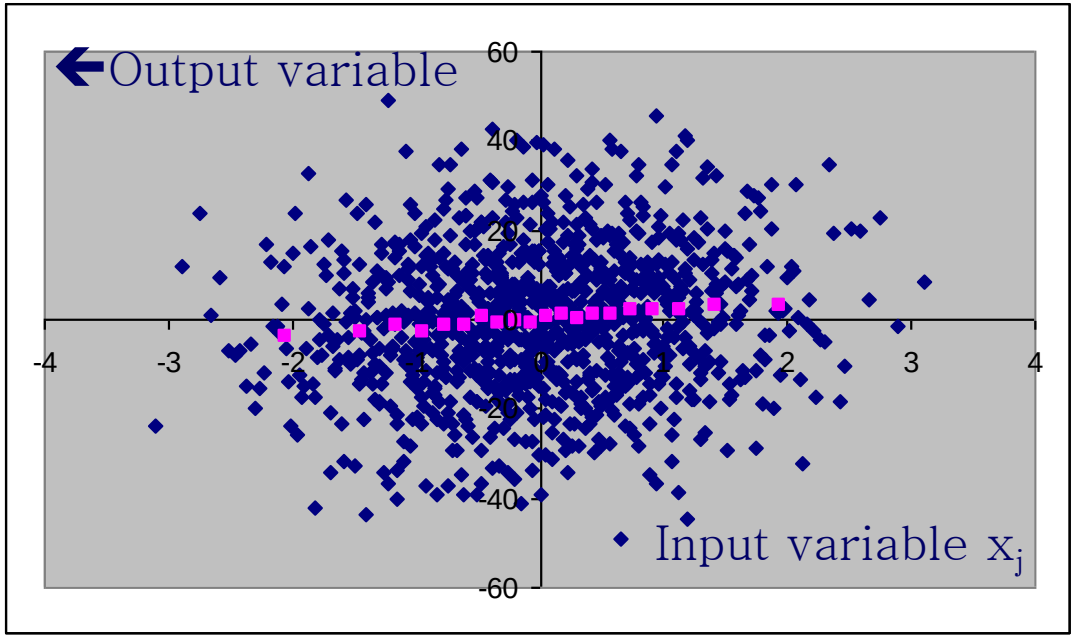




Plotting the output as a function of two different input factors

Which factor is more important?

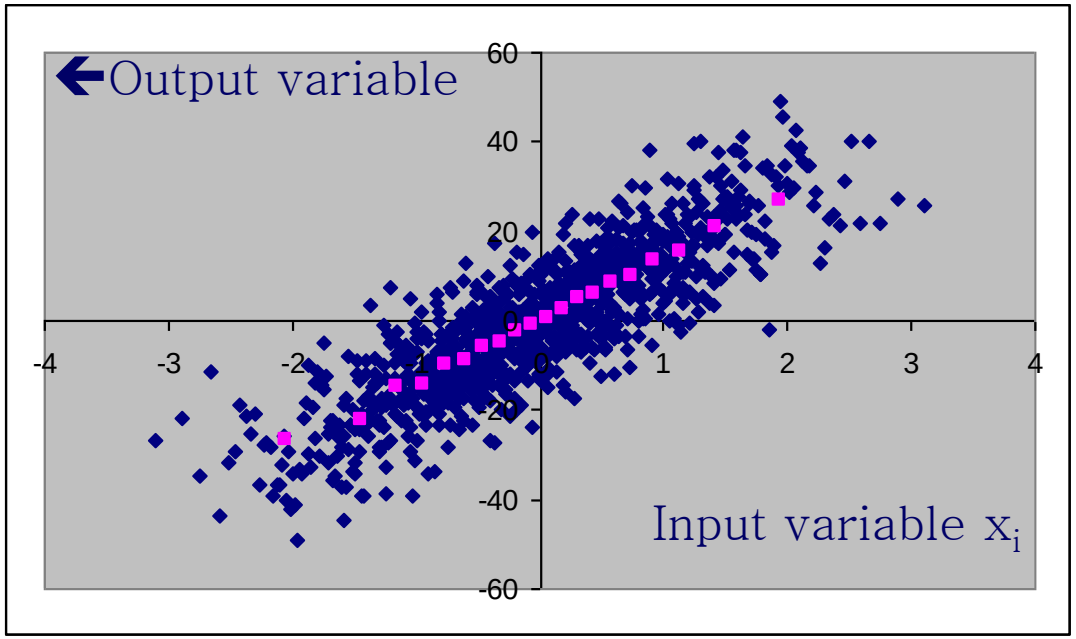




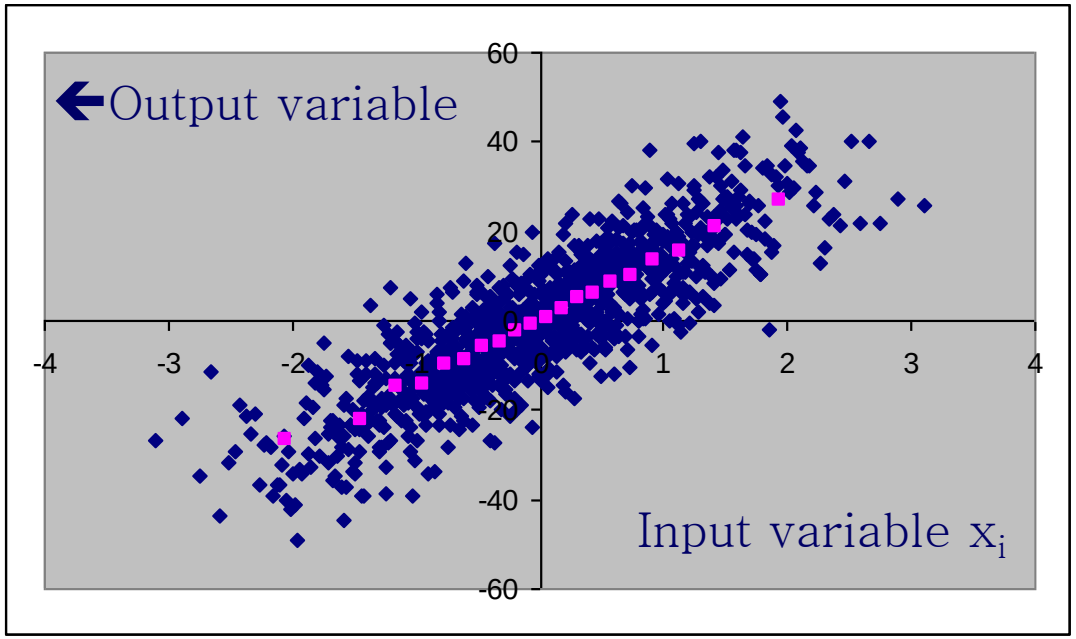
~1,000 blue points

Divide them in 20 bins of ~ 50 points

Compute the bin's average (pink dots)

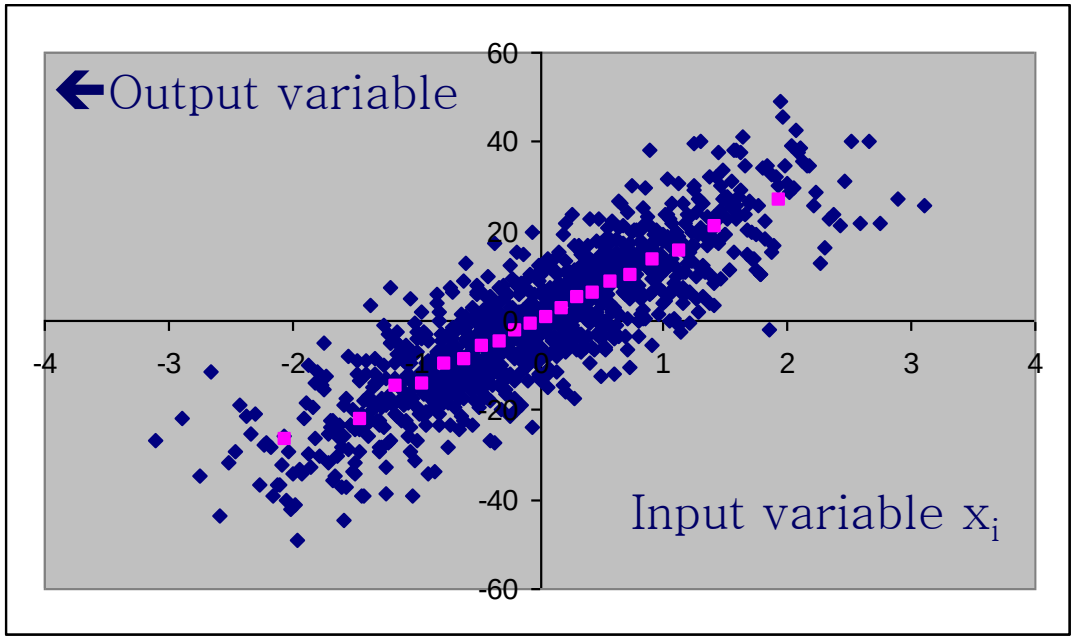


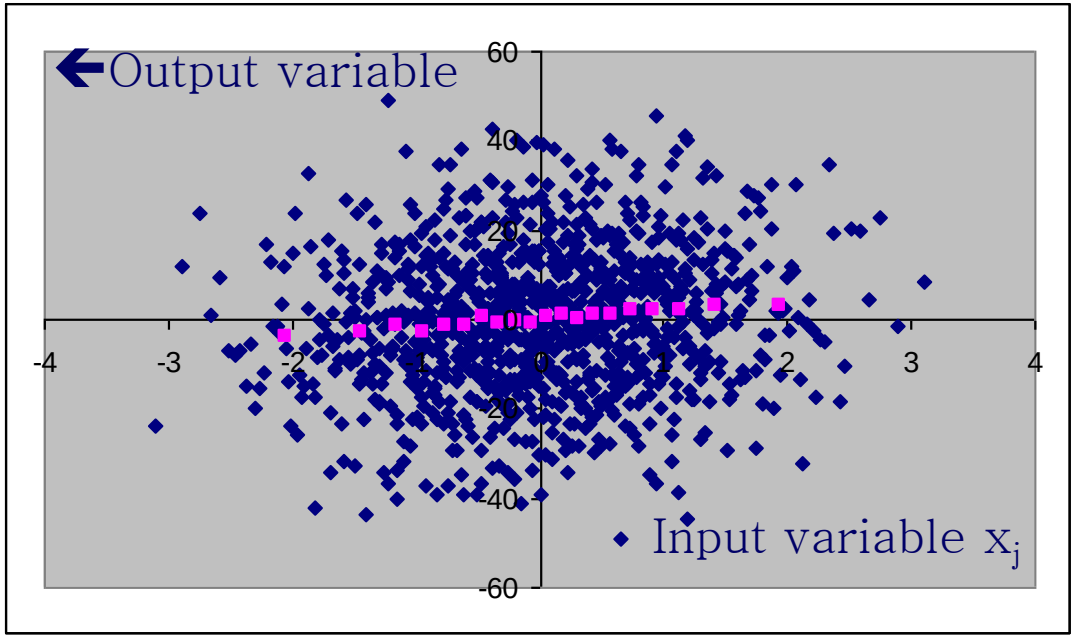
Each pink point is ~ $E_{\mathbf{X}_{\sim i}}\left(Y \middle| X_i\right)$



Taking the variance of the pink points one obtains a sensitivity measure

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)$$





Which factor has the highest

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y\middle|X_i\right)\right)?$$

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

The partial variance divided by the total variance is the so-called sensitivity index of the first order

Don’t run the model just once

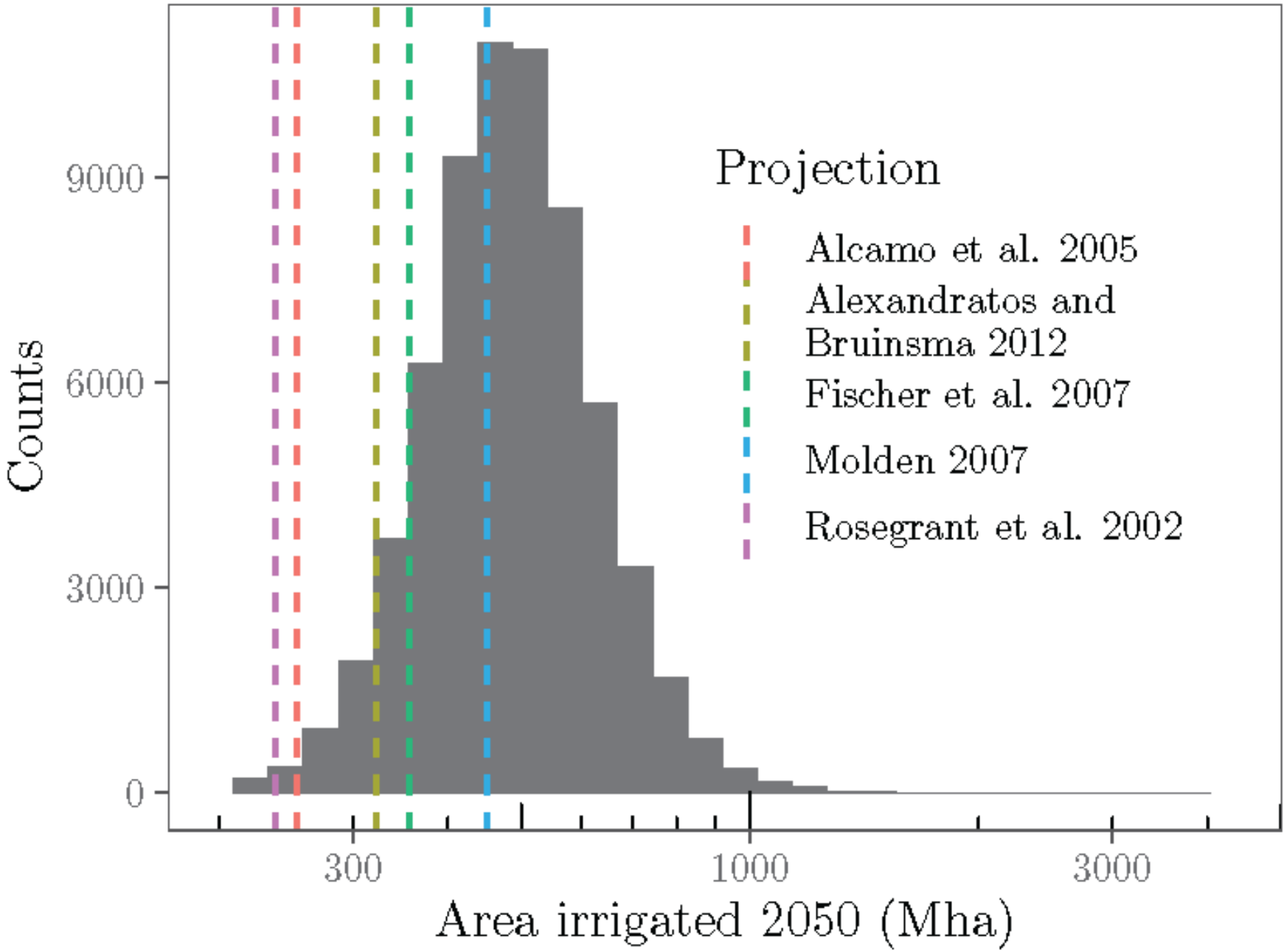
There is much to learn by running the model a few times, especially during model building

Lubarsky's Law of Cybernetic Entomology: there is always one more bug!

Model routinely used to produce point estimates may becomes non conservative when the uncertainty is plugged in

Current Models Underestimate Future Irrigated Areas

- How much land will need to be irrigated by the year 2050?
- Here the dashed lines represent deterministic model predictions from different models and datasets (from FAO & others organizations);
- An uncertainty analysis (grey histogram) reveals that the models are non-conservative: the need might be much larger



Citation:
Puy, A., Lo Piano, S., & Saltelli, A. (2020). Current models underestimate future irrigated areas. *Geophysical Research Letters*, 47, e2020GL087360. https://doi.org/10.1029/2020GL087360

Don’t sample just parameters and boundary conditions

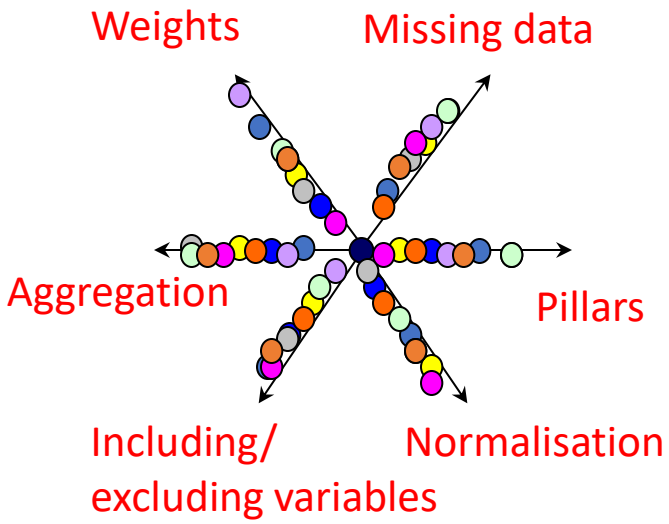
Explore thoroughly the space of the assumptions

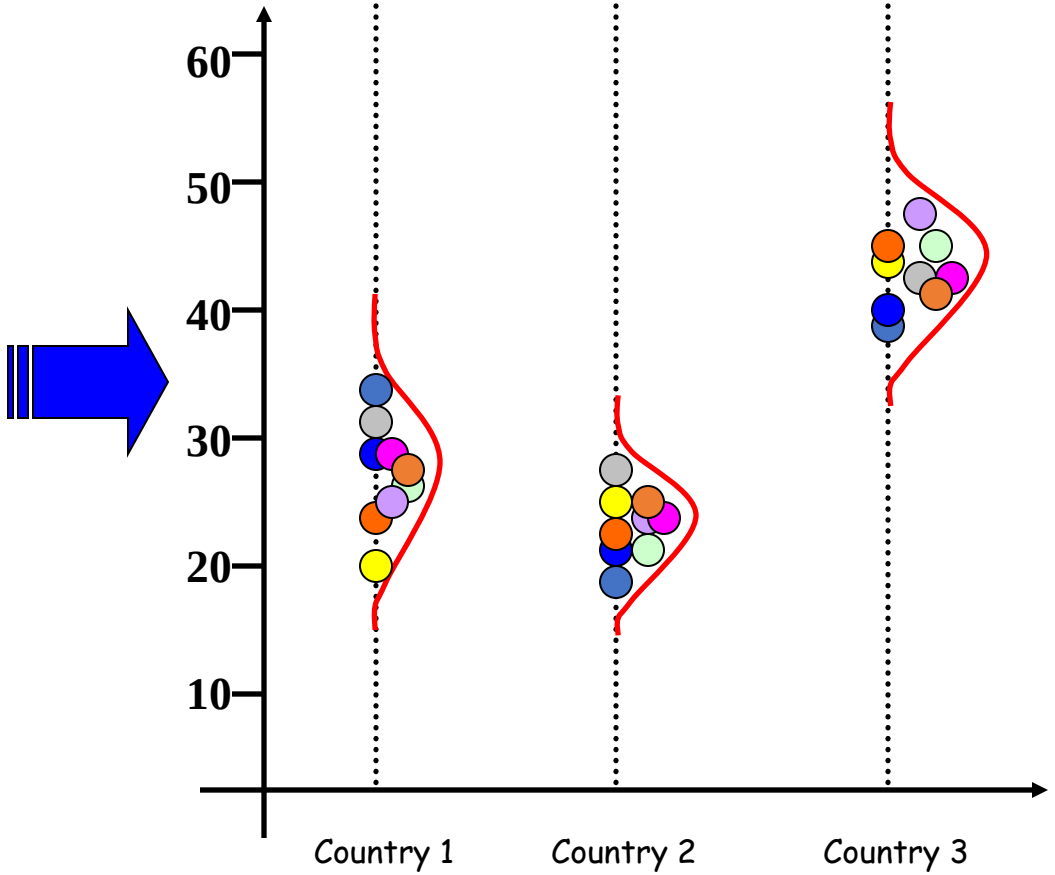
One can sample more than just factors:

- modelling assumptions,
- alternative data sets,
- resolution levels,
- scenarios …

Assumption	Alternatives
Number of indicators	▪ all six indicators included or one-at-time excluded (6 options)
Weighting method	▪ original set of weights, ▪ factor analysis, ▪ equal weighting, ▪ data envelopment analysis
Aggregation rule	▪ additive, ▪ multiplicative, ▪ Borda multi-criterion

Space of alternatives





NEVER vary all factors of the same amount

Be it 5%, 10%, or 20%

New WHO estimates: Up to 190 000 people could die of COVID-19 in Africa if not controlled

07 May 2020

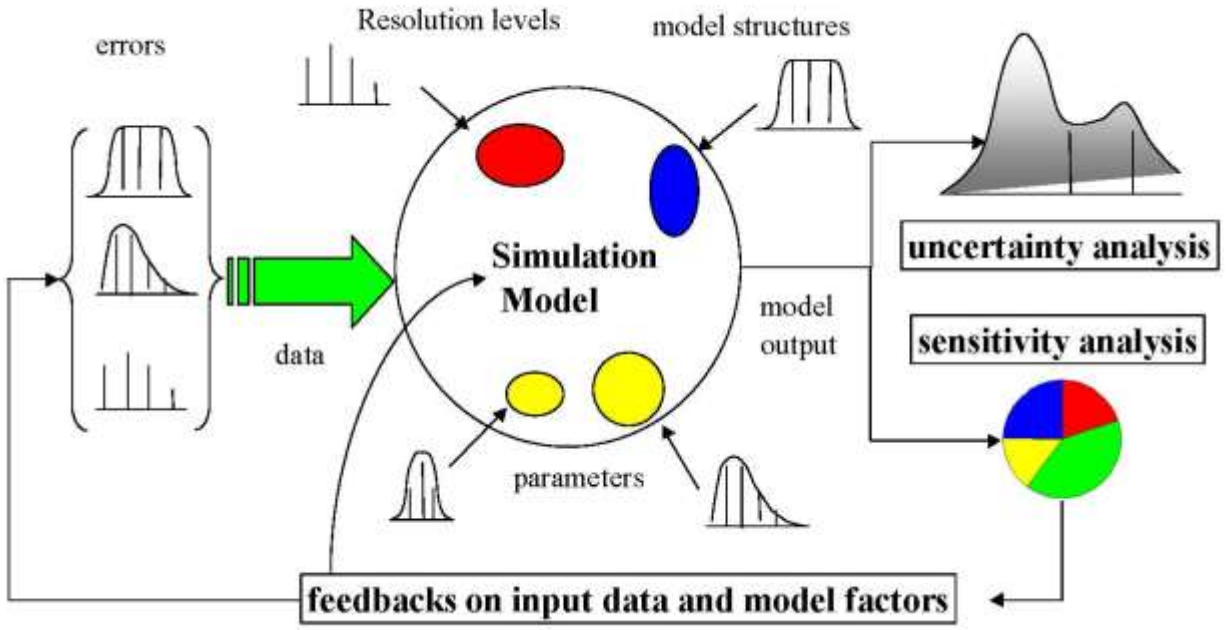
Brazzaville – Eighty-three thousand to 190 000 people in Africa could die of COVID-19 and 29 million to 44 million could get infected in the first year of the pandemic if containment measures fail, a new study by the World Health Organization (WHO) Regional Office for Africa finds. The research, which is based on prediction modelling, looks at 47 countries in the

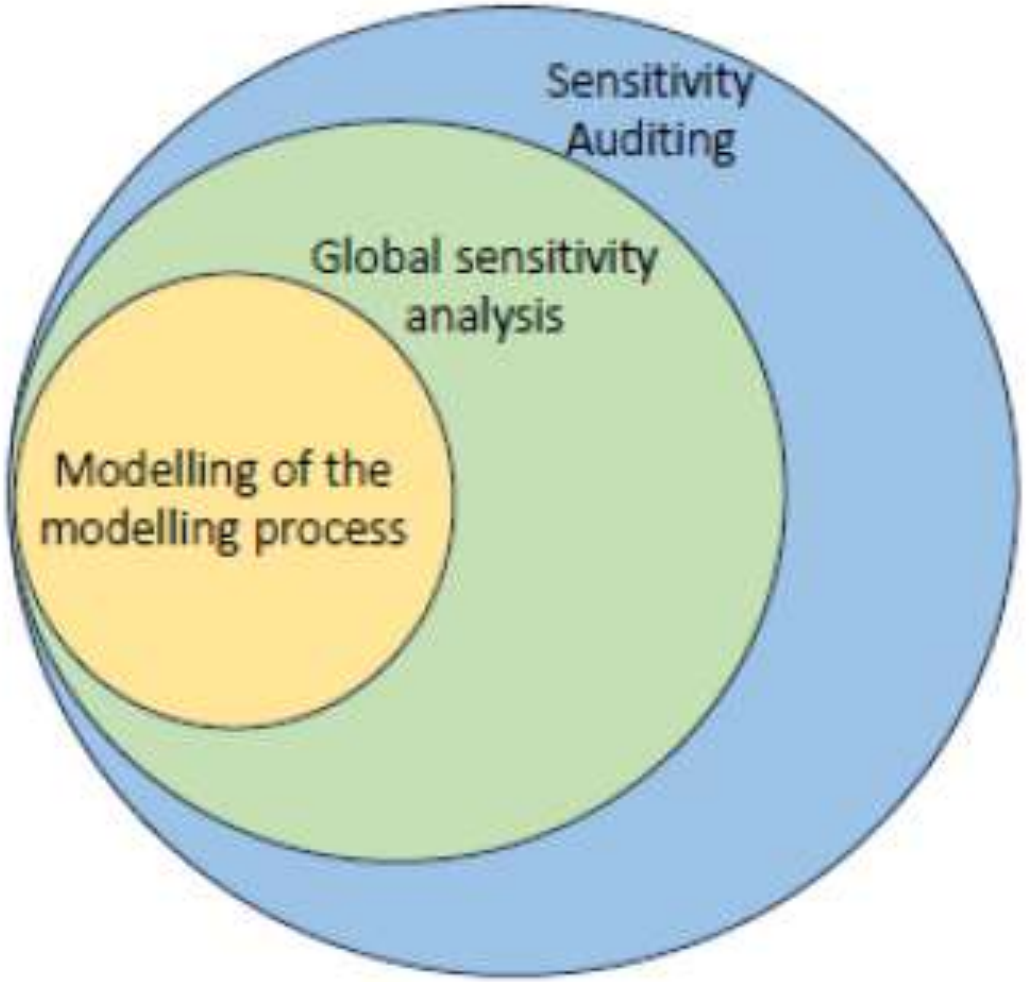
Speculative scenario in which ten uncertain input probabilities are increased by an arbitrary 10% — as if they were truly equally uncertain — with no theoretical or empirical basis for such a choice

World Health Organization
REGIONAL OFFICE FOR Africa

In a numerical experiment relating to a real-life application, the range of uncertainty of each input is a crucial input to the analysis, and often the most expensive to get

The End


errors
Resolution levels
model structures
Simulation Model
data
model output
uncertainty analysis
sensitivity analysis
parameters
feedbacks on input data and model factors


Sensitivity Auditing
Global sensitivity analysis
Modelling of the modelling process

There are terms which capture two-way, three way, … interactions among variables

All these terms are linked by a formula

Variance decomposition (ANOVA)

$$V(Y) =$$

$$\sum_{i} V_i + \sum_{i,\, j>i} V_{ij} + \ldots + V_{123\ldots k}$$

Variance decomposition (ANOVA)

The total variance can be decomposed into main effects and interaction effects up to the order k, the dimensionality of the problem (only for independent factors)

If fact interactions terms are awkward to handle: **just** the **second** order terms for a model with k factors are as many as k(k-1)/2 …

(10 factors=45 second order terms)

How about a single ‘importance’ terms for all effects?

In fact such terms exist and can be computed easily, without knowledge of the individual interaction terms

Thus given a 3-factor model $f(X_1, X_2, X_3)$

Where the variance decomposition would read $1 = S_1 + S_2 + S_3 + S_{12} + S_{13} + S_{23} + S_{123}$

We compute

$$T_1 = S_1 + S_{12} + S_{13} + S_{123}$$

$$T_2 = S_2 + S_{12} + S_{23} + S_{123}$$

$$T_3 = S_3 + S_{13} + S_{23} + S_{123}$$

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

Known as:

- Pearson's correlation ratio (Pearson)
- First order effect (Sobol' & disciples)
- Top marginal variance (Jansen)

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)$$

Top marginal variance=
= the expected reduction in variance than would be achieved if factor Xi could be fixed.
Why?

Because:

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)+$$
$$+E_{X_i}\left(V_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)=V(Y)$$

One of the exercise in the book: prove this using $V(\bullet)=E(\bullet)^2-E^2(\bullet)$

Because:

= reduction in variance if X_i could be fixed.

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)+$$

$$+E_{X_i}\left(V_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)=V(Y)$$

What variance would be left (on average) if X_i could be fixed

For additive systems one can decompose the total variance as a sum of first order effects

$$\sum_i V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)\approx V(Y)$$

How to get from first order to total order

From

$$V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y|X_i\right)\right)$$ Main effect of factor X_i

replacing X_i with $\boldsymbol{X}_{\sim i}$

Gives the main effect of non-X_i

$$V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y|\mathbf{X}_{\sim i}\right)\right)$$

But again:

$$V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y|\mathbf{X}_{\sim i}\right)\right)+$$
$$E_{\mathbf{X}_{\sim i}}\left(V_{X_i}\left(Y|\mathbf{X}_{\sim i}\right)\right)=V\left(Y\right)$$

Main effect on non$-X_i$

$$E_{\mathbf{X}_{\sim i}}\left(V_{X_i}\left(Y|\mathbf{X}_{\sim i}\right)\right)$$

$$V_{\mathbf{X}_{\sim i}}\left(E_{X_i}\left(Y|\mathbf{X}_{\sim i}\right)\right)$$

… all remaining variance must be due to X_i and its interactions; this divided the total variance gives the total effect T_i

How to compute S_i and T_i ?

If k is the number of factors we need $n(2+k)$ simulation where n is the size of a ‘base’ matrix.

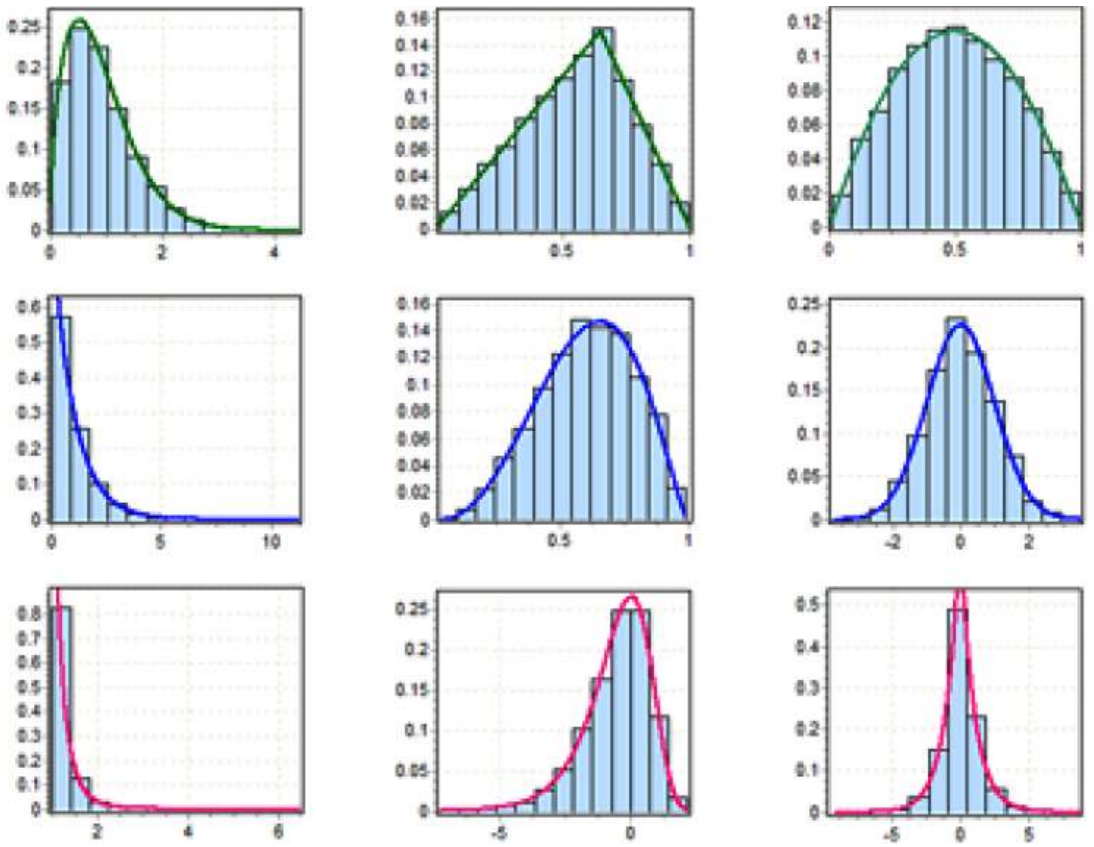
What is needed is a design (as in statistical DOE theory)

Building a Monte Carlo analysis

$$\begin{matrix} x_{11} & x_{12}\ldots & x_{1k} \\ x_{21} & x_{22}\ldots & x_{2k} \\ \ldots & \ldots & \ldots \\ x_{N1} & x_{N2} & x_{Nk1} \end{matrix}$$

Input matrix: each column is a sample of size N from the distribution of a factor

Each row is a sample trial of size k to generate a value of y



Examples of distributions of input factors

$$\mathbf{M}_1 = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & & & \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

Base matrix M_1

$$\mathbf{M}_2 = \begin{pmatrix} x'_{11} & x'_{12} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x'_{2k} \\ \dots & & & \\ x'_{n1} & x'_{n2} & \dots & x'_{nk} \end{pmatrix}$$

Base matrix M_2

$$\mathbf{N}_j = \begin{pmatrix} x'_{11} & x'_{12} & \dots & x_{1j} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x_{2j} & \dots & x'_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'_{n1} & x'_{n2} & \dots & x_{nj} & \dots & x'_{nk} \end{pmatrix}$$

k of these are needed

Matrix quasi-M_2 for factor j

'quasi' because all columns are from M_2 but column j from M_1

$$\mathbf{M}_1 = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & & & \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

Base matrix M_1

$$\mathbf{M}_2 = \begin{pmatrix} x'_{11} & x'_{12} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x'_{2k} \\ \dots & & & \\ x'_{n1} & x'_{n2} & \dots & x'_{nk} \end{pmatrix}$$

Base matrix M_2

S_j

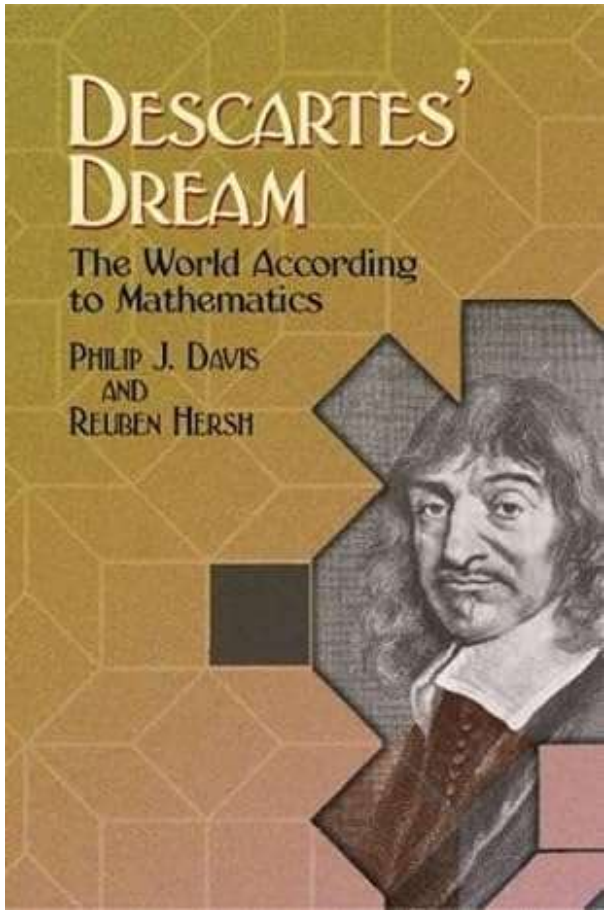
T_j

$$\mathbf{N}_j = \begin{pmatrix} x'_{11} & x'_{12} & \dots & x_{1j} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x_{2j} & \dots & x'_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x'_{n1} & x'_{n2} & \dots & x_{nj} & \dots & x'_{nk} \end{pmatrix}$$

k of these are needed

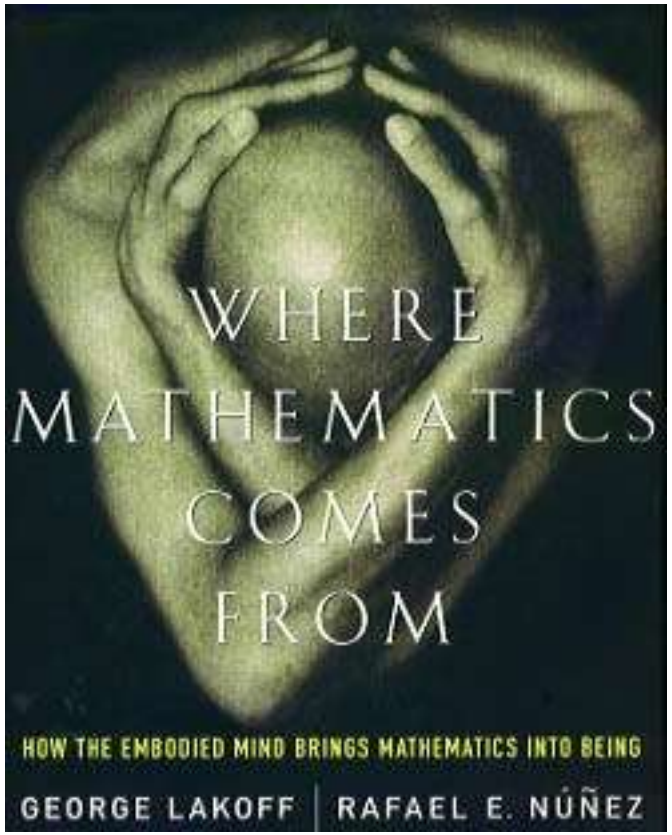
Matrix quasi-M_2 for factor j

Why models live in a state of exception



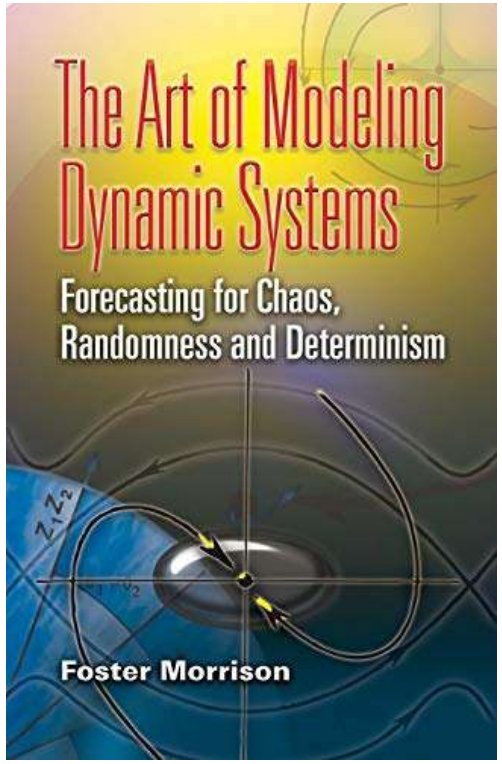
Unparalleled palette of methods / epistemic authority

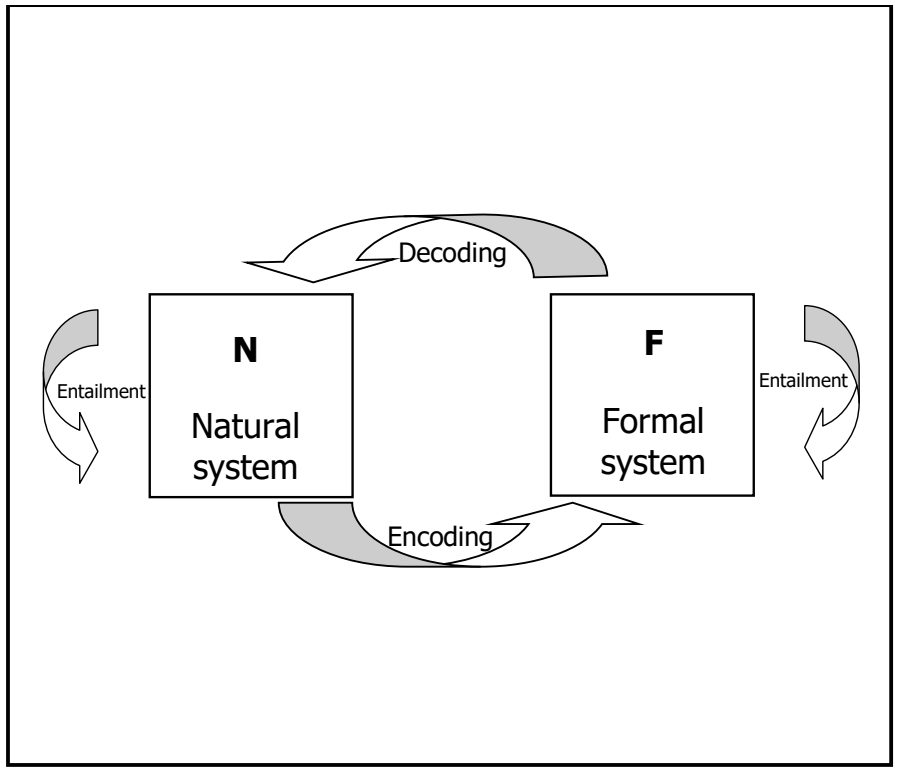
Models dispose of a unique repertoire of methods. Are endowed with unparallel epistemic authority that originates from mathematics, the highest ranked among scientific disciplines (Davies & Hersh, 1986), considered by the fathers of the scientific revolution the language of God himself, up to the point that reconnecting it to human experience is up today an unfinished project (Lakoff & Núñez, 2001).

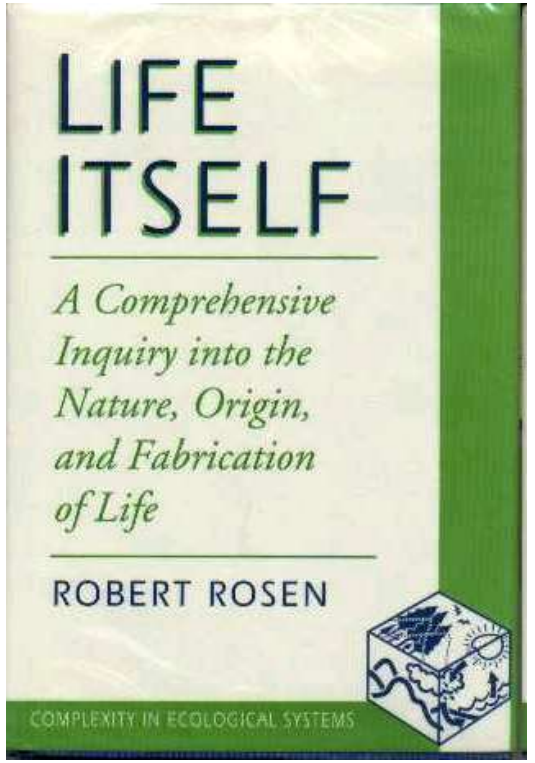


Why models live in a state of exception

Lack of agreed standards. Modelling as art/craft (Rosen).

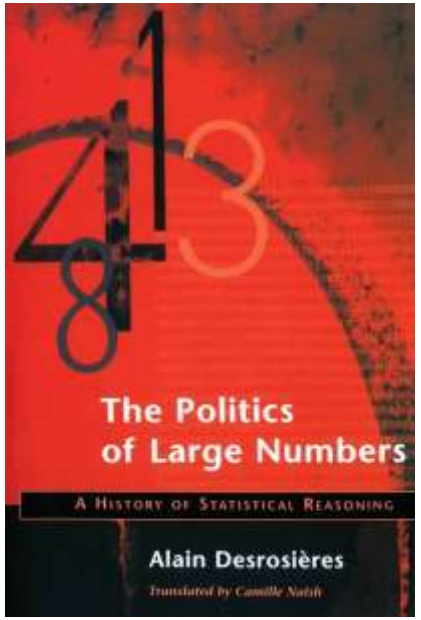






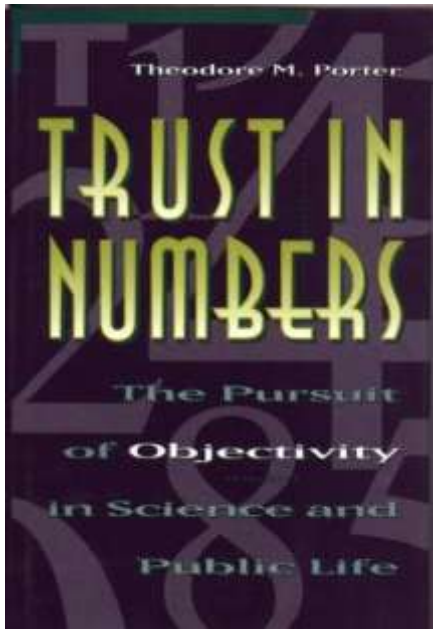
Louie, A.H. 2010. “Robert Rosen’s Anticipatory Systems.” Foresight 12 (3): 18–29.
Padilla, J. J., Diallo, S. Y., Lynch, C. J., & Gore, R. (2018). Observations on the practice and profession of modeling and simulation: A survey approach. SIMULATION, 94(6), 493–506.

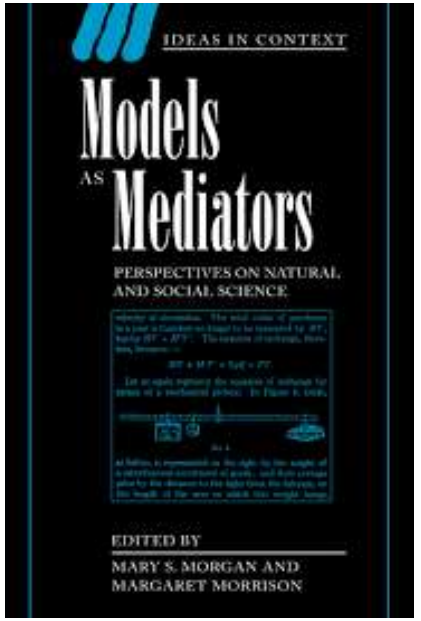
Why models live in a state of exception

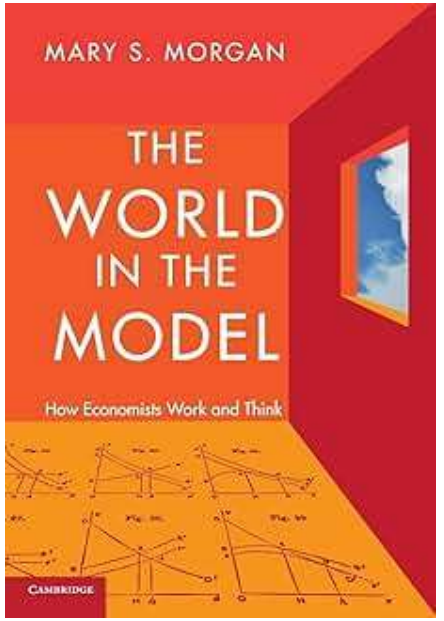


Mathematical models escape sociology of quantification

Statistics has a much deeper connection to sociology, and to sociology of quantification in particular (Desrosières, 1998; Mennicken & Espeland, 2019; Mennicken & Salais, 2022) than mathematical modelling. Sociology of quantification treats impact assessment tools such as cost benefit analysis (Porter, 1995). Little on modelling, see exceptions: (Morgan, 2012, Morgan & Morrison, 1999).







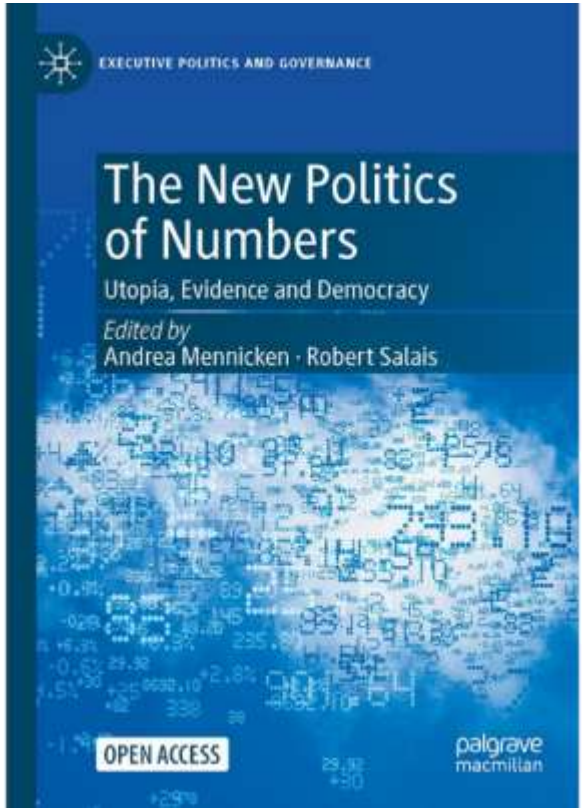
Model have a better pretense to neutrality than other instances of quantification

A technical proof of quality is illusory without a parallel investigation of normative quality; the example of indicators of employment

Technical Quality

Normative quality

Justice: Means versus Freedoms
Author(s): Amartya Sen
Source: *Philosophy & Public Affairs*, Vol. 19, No. 2 (Spring, 1990), pp. 111-121
Published by: Wiley
Stable URL: http://www.jstor.org/stable/2265406
Accessed: 28/10/2014 14:48



Salais, R. (2022). "La donnée n'est pas un donné": Statistics, Quantification and Democratic Choice. In *The New Politics of Numbers: Utopia, Evidence and Democracy*, Andrea Mennicken and Robert Salais, Palgrave Macmillan, pp. 379–415.

Why models live in a state of exception

Mathematical models are extremely malleable

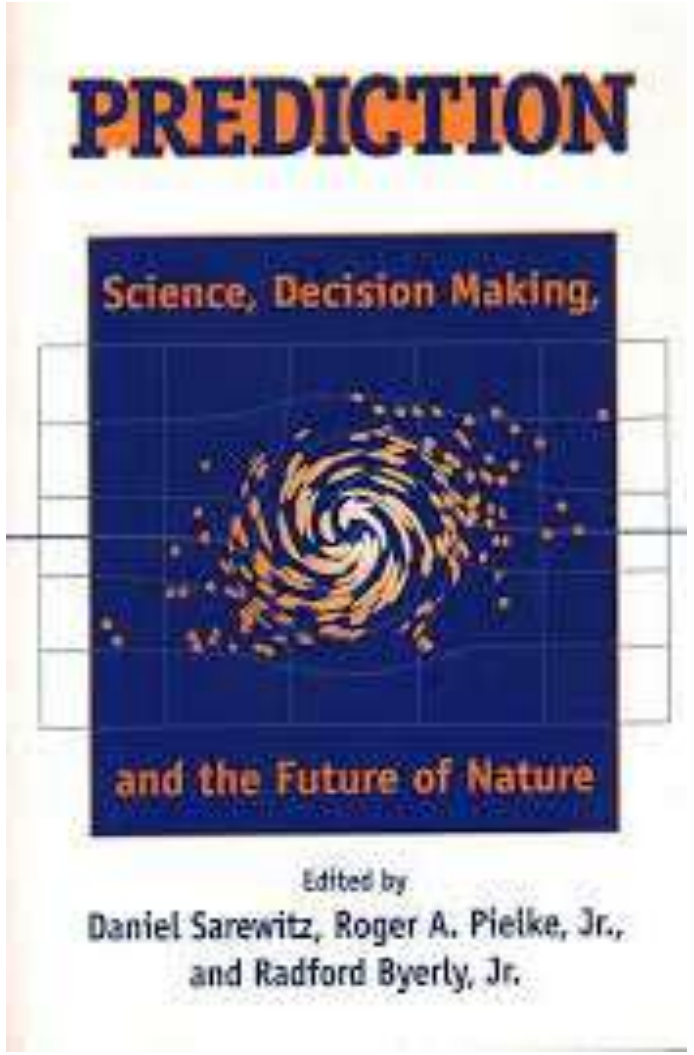
Models lend themselves very naturally to evidence based policy. In statistics you have to reverse the statistical pyramid to achieve the same result – this goes much faster with models

Evidence based policy	Statistics (creating things that hold together for the solution of practical problems)
Policy based evidence	Governance driven quantification (a reversal of the statistical pyramid)

Why models live in a state of exception

Models cannot be falsified

Models do not meet classic (Popperian) criteria of scientificity. Oreskes (2000) has observed that model-based predictions tend to be treated like logical inferences in a classic hypothetic-deductive model. The relation between models and data is often more symbiotic than adversarial. In climate studies this relation has been defined as 'incestuous', exactly to make the point that in modelling studies using data to prove a model wrong may not be straightforward (Edwards, 1999).

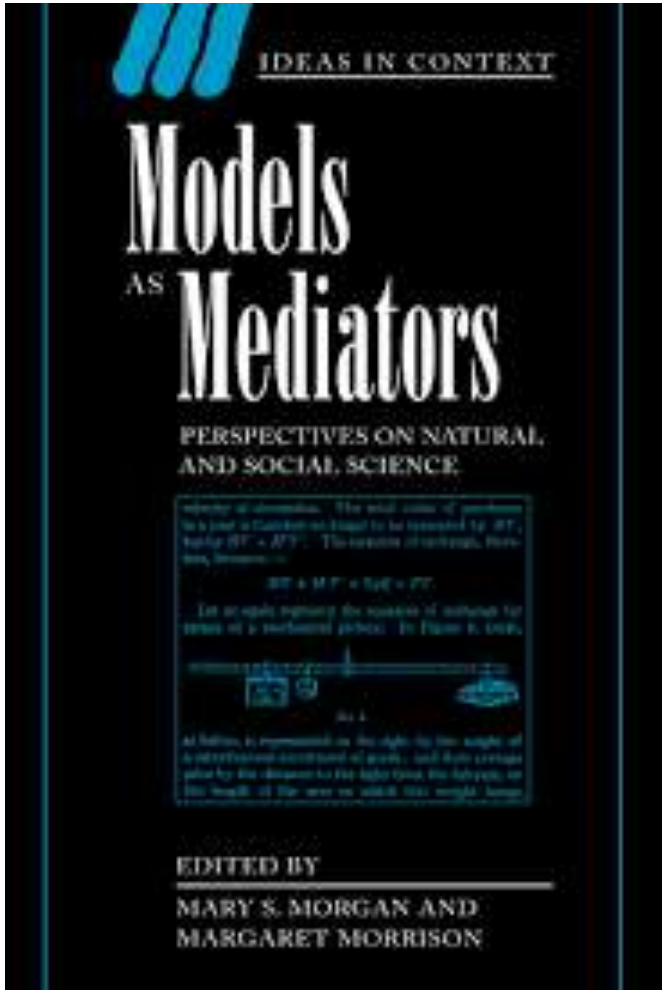


Edwards, P.N. 1999. 'Global Climate Science, Uncertainty and Politics: Data-laden Models, Model-filtered Data'. Science as Culture 8 (4): 437–72.

Why models live in a state of exception

Models as the most effective mediators between theory and reality

Due to their independence from both theory and the world, models act as “mediators”, instruments that advance understanding thanks to the tacit craftsmanship of scientists (Morgan & Morrison 1999).



Why models live in a state of exception

Models as the most effective mediators between theory and reality

Models are metaphors that express “in an indirect form our presuppositions about the problem and its possible solutions”, and can thus assist in an **extended community of peers** to deliberate about social or ecological problems (Ravetz 2023).

Extended peer community

Add languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

The concept of **Extended peer community** belongs to the field of Sociology of science, and in particular the use of science in the solution of social, political or ecological problems. It was first introduced by in the 1990s by Silvio Funtowicz and Jerome R. Ravetz.[1] in the context of what would become Post-normal science. An **Extended peer community** is intended by these authors as a space where both credentialed experts from different disciplines and lay stakeholders can discuss and deliberate.

Consequences descending from state of exception

Gross asymmetry developers/ users

Models operate in a context of asymmetry of knowledge between developers and users (Jakeman *et al.*, 2006). There are ‘black boxes’ also in other families of quantification, typically algorithms or statistics. Yet this asymmetry may be larger for mathematical models.

Jakeman, A.J., R.A. Letcher, and J.P. Norton. 2006. ‘Ten Iterative Steps in Development and Evaluation of Environmental Models’. Environmental Modelling & Software 21 (5): 602–14.

Consequences descending from state of exception

Ritual use

An analogy between statistical and mathematical modelling is in the 'ritual' use of methods. Rituals in statistics are described in Gigerenzer (Gigerenzer, 2018; Gigerenzer & Marewski, 2015). For models here an anecdote by Kenneth Arrow: producing one month-ahead weather forecasts

"... The commanding general is well aware that the forecasts are no good. However, he needs them for planning purposes"

See also Niklas Luhmann 'deparadoxification' (Moeller, 2006); See also politicians' claim: 'We follow the science' during COVID-19

Gigerenzer, Gerd. 2018. 'Statistical Rituals: The Replication Delusion and How We Got There'. Advances in Methods and Practices in Psychological Science 1 (2): 198–218.

Mathematical models: a state of exception?

nature

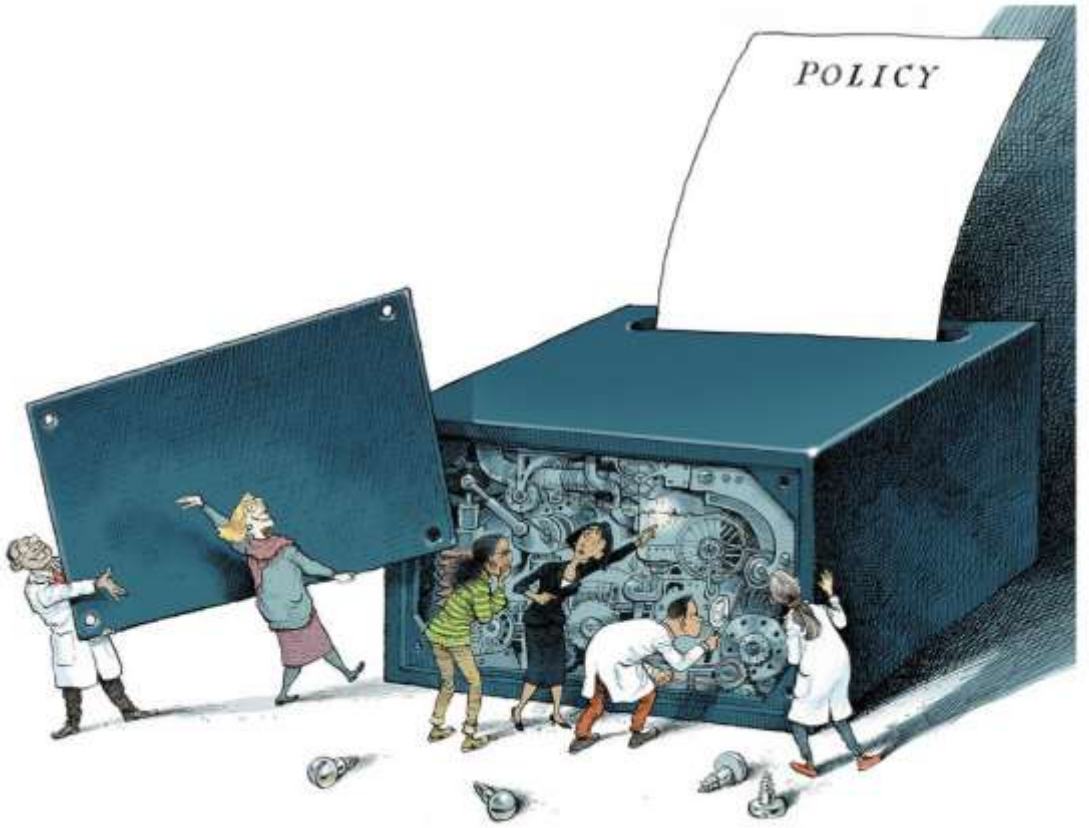
Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › comment › article

COMMENT | 24 June 2020

Five ways to ensure that models serve society: a manifesto

Pandemic politics highlight how predictions need to be transparent and humble to invite insight, not blame.

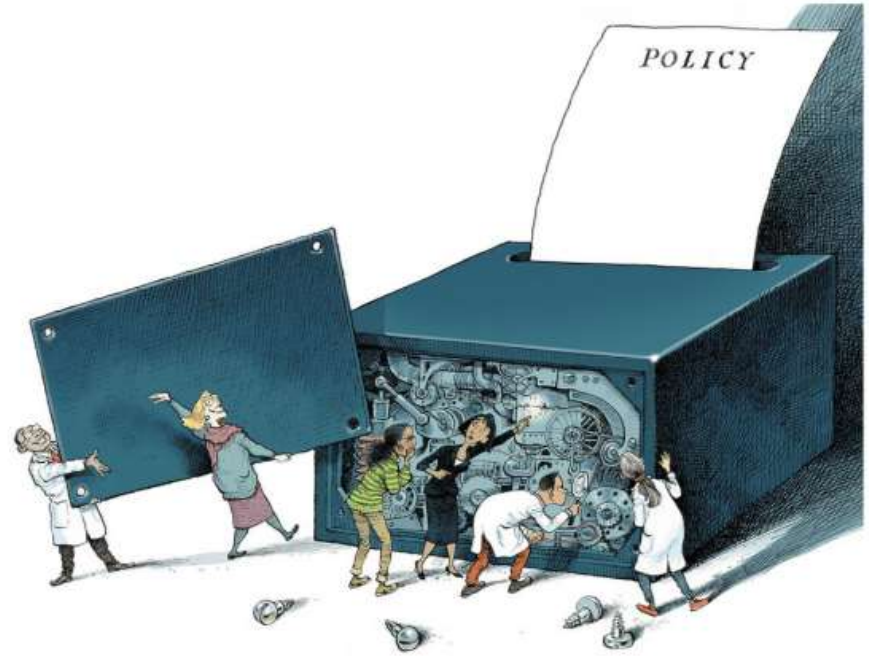


Mind the assumptions

Assess uncertainty and sensitivity

Mind the hubris

Complexity can be the enemy of relevance



Mind the framing

Match purpose and context

Mind the consequences

Quantification can backfire

Mind the unknowns

Acknowledge ignorance

Mind the assumptions

Assess uncertainty and sensitivity

Mind the hubris

Complexity can be the enemy of relevance

Mind the framing

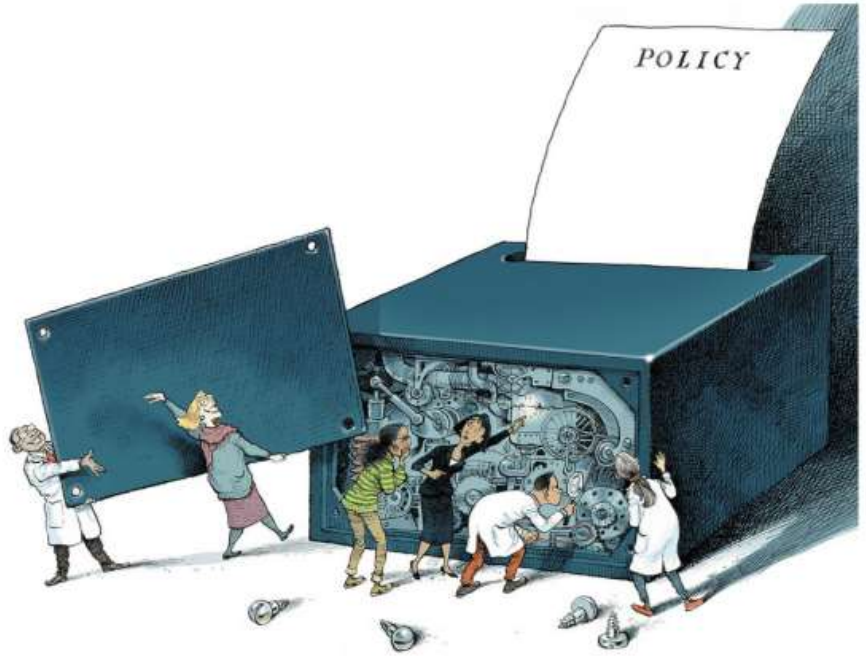
Match purpose and context

Mind the consequences

Quantification can backfire

Mind the unknowns

Acknowledge ignorance

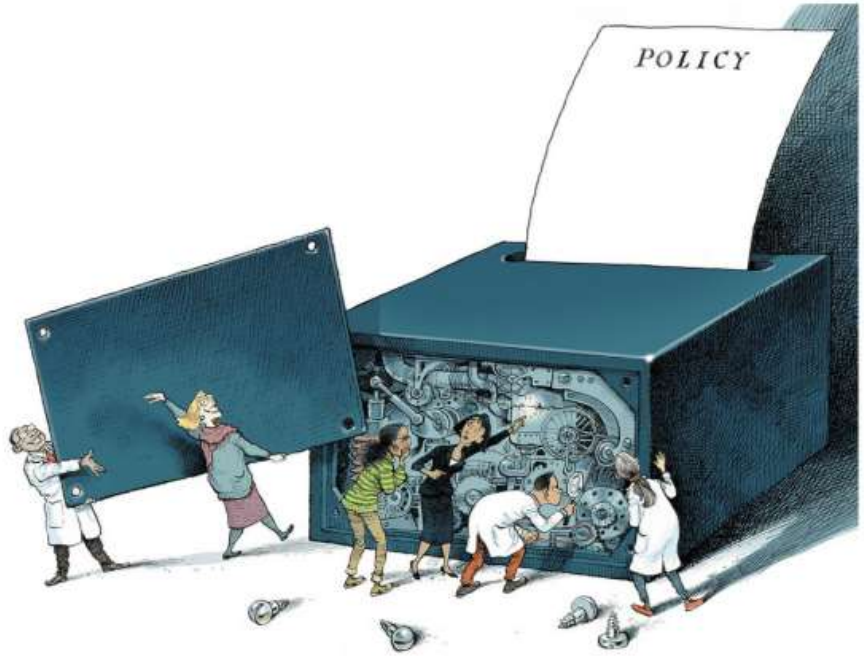


Source: A. Saltelli, G. Bammer, I. Bruno, et al., Five ways to ensure that models serve society: a manifesto, Nature 582 (2020) 482–484.

Mind the assumptions

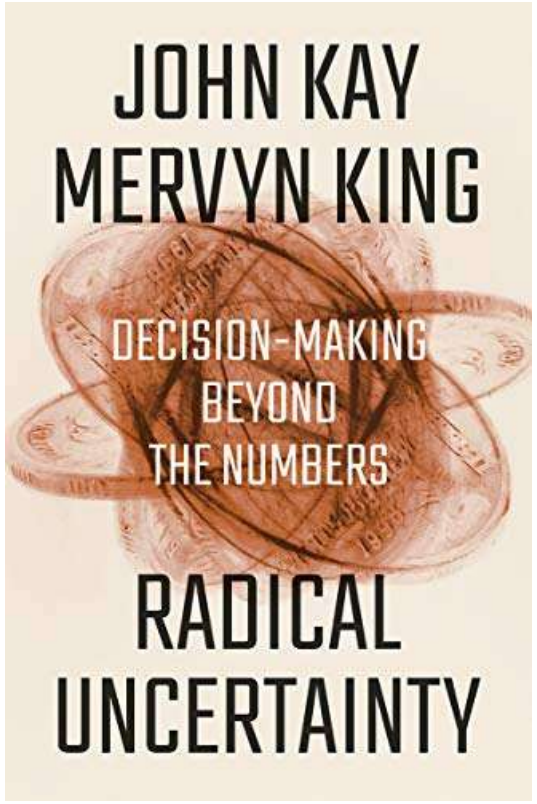
Assess uncertainty and sensitivity

… models require input values for which there is no reliable information…



Source: A. Saltelli, G. Bammer, I. Bruno, et al., Five ways to ensure that models serve society: a manifesto, Nature 582 (2020) 482–484.

Models ask as input information which we don't have – The case of WEBTAG



John Kay

WebTAG: Annual Percentage Change in Car Occupancy (% pa) up to 2036

Journey Purpose	Weekday					Weekend	All Week
	7am-10am	10am-4pm	4pm-7pm	7pm-7am	Weekday Average		
Work	-0.48	-0.4	-0.62	-0.5	-0.44	-0.48	-0.45
Non - Work (commuting and other)	-0.67	-0.65	-0.53	-0.47	-0.59	-0.52	-0.56

Source: J. A. Kay, "Knowing when we don't know," 2012, https://www.ifs.org.uk/docs/john_kay_feb2012.pdf

Mind the assumptions

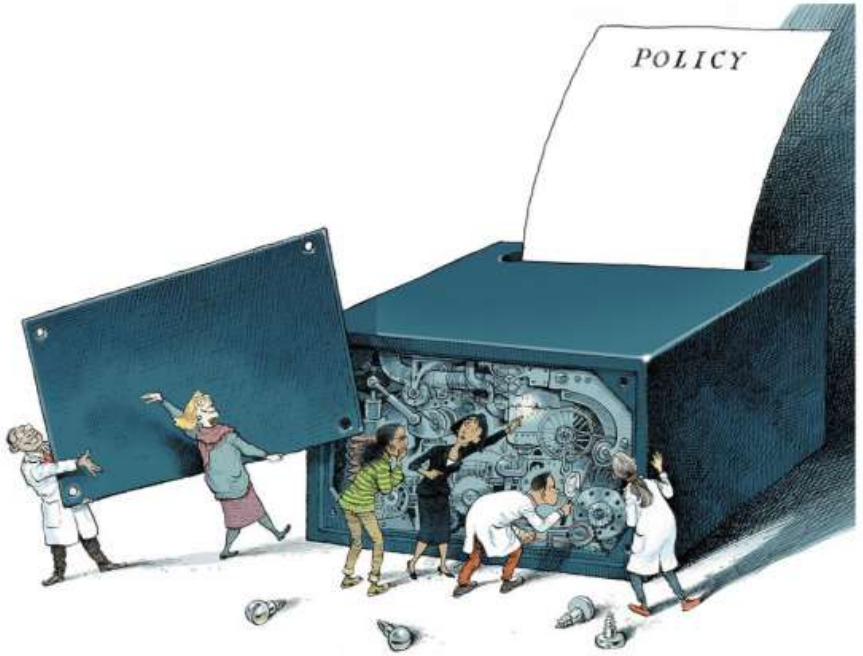
Assess uncertainty and sensitivity

Mind the hubris

Complexity can be the enemy of relevance

Mind the framing

Match purpose and context



Source: A. Saltelli, G. Bammer, I. Bruno, et al., Five ways to ensure that models serve society: a manifesto, Nature 582 (2020) 482–484.

Mind the consequences

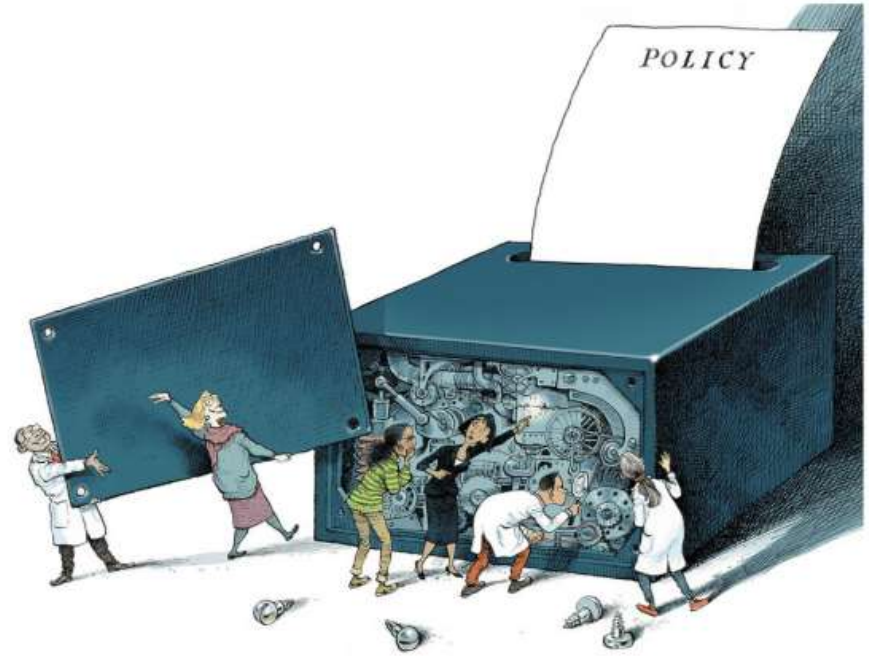
Quantification can backfire

Mind the unknowns

Acknowledge ignorance

Mind the consequences

Quantification can backfire.



« Back to Article

WIRED MAGAZINE: 17.03

Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon 02.23.09

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of Wired.*

Here is what killed your 401(k)…

Li's Gaussian copula function …

Nassim Nicholas Taleb, hedge fund manager and author of *The Black Swan*, is particularly harsh when it comes to the copula. "People got very excited about the Gaussian copula because of its mathematical elegance, but the thing never worked," he says. "Co-association between securities is not measurable using correlation," because past history can never prepare you for that one day when everything goes south. "Anything that relies on correlation is charlatanism."

Felix Salmon, Wired, February 2009

WIRED

Source: https://www.wired.com/2009/02/wp-quant/

Mind the assumptions

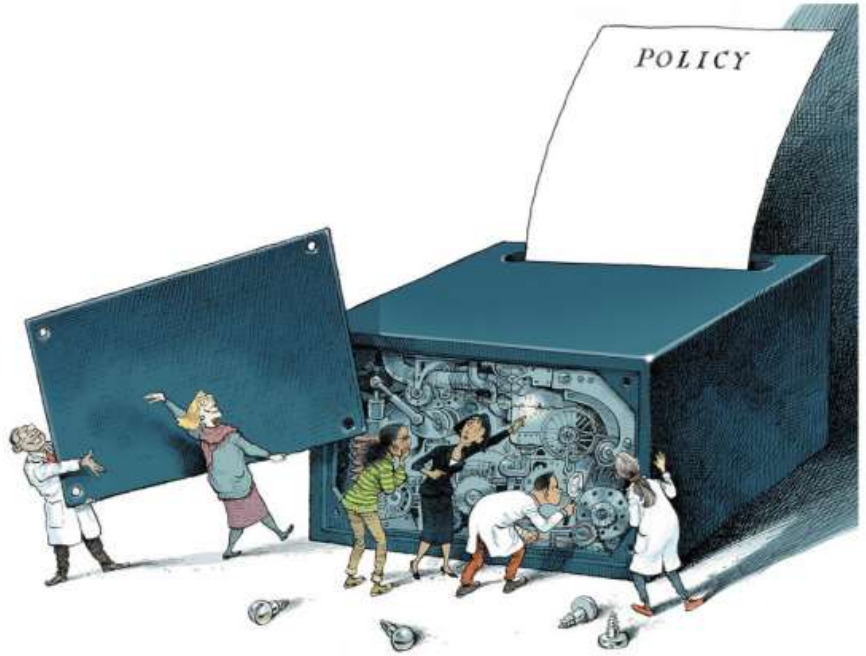
Assess uncertainty and sensitivity

Mind the hubris

Complexity can be the enemy of relevance

Mind the framing

Match purpose and context



Source: A. Saltelli, G. Bammer, I. Bruno, et al., Five ways to ensure that models serve society: a manifesto, Nature 582 (2020) 482–484.

Mind the consequences

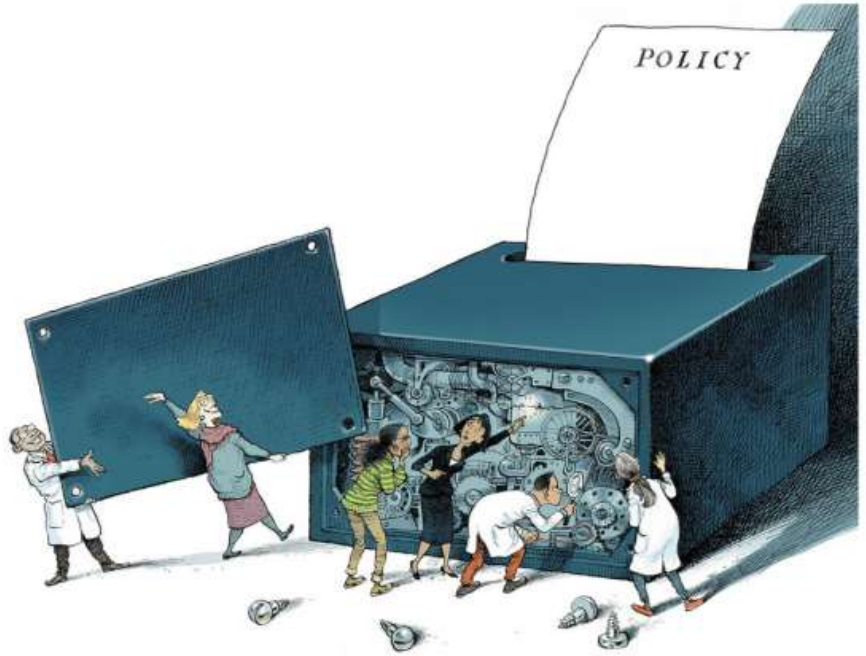
Quantification can backfire

Mind the unknowns

Acknowledge ignorance

Mind the unknowns

Acknowledge ignorance



Source: A. Saltelli, G. Bammer, I. Bruno, et al., Five ways to ensure that models serve society: a manifesto, Nature 582 (2020) 482–484.

From Socrates's "knowing of not knowing" to Nicolaus Cusanus' Docta Ignorantia, ignorance was a virtue until Descartes

Consequences descending from state of exception

Models and trans-science

Models lend themselves to trans-science (Weinberg, 1972).

- How many people will sit in autonomous cars by 2050
- How will the spread of malaria change if global temperature increases by 1.5ºC
- What will be the cost of CO2 averaged over the next three centuries

Weinberg, AM. 1972. ‘Science and Trans-Science’. *Minerva* 10:209–22.

Journal of Statistical Software

sensobol: An R Package to Compute Variance-Based Sensitivity Indices

Arnald Puy, **Samuele Lo Piano**, **Andrea Saltelli**, **Simon A. Levin**

The measures and their ‘settings’ = when to use them

Journal of the American Statistical Association ›
Volume 97, 2002 - Issue 459

Enter keywords, authors, DOI, C

Submit an article | Journal homepage

979
Views
286
CrossRef citations to date
6
Altmetric

Primary Article

On the Relative Importance of Input Factors in Mathematical Models

Safety Assessment for Nuclear Waste Disposal

Andrea Saltelli & Stefano Tarantola

Pages 702-709 | Published online: 31 Dec 2011

Download citation https://doi.org/10.1198/016214502388618447

The measures and their ‘settings’ = when to use them

First order effect	Factor prioritization (orienting research)
Total effect	Factor fixing (model simplification)

ELSEVIER

Computer Physics Communications

Volume 145, Issue 2, 15 May 2002, Pages 280-297

Making best use of model evaluations to compute sensitivity indices

Andrea Saltelli ✉ ⊕

Computing the indices efficiently

Higher order Sobol' indices

Get access >

Art B. Owen ✉, Josef Dick, Su Chen

Information and Inference: A Journal of the IMA, Volume 3, Issue 1, March 2014, Pages 59–81, https://doi.org/10.1093/imaiai/iau001

Published: 01 March 2014 **Article history** ▾

Plenty of code available in R, MATLAB, and Phyton

https://cran.r-project.org/web/packages/sensitivity/sensitivity.pdf

https://cran.rstudio.com/web/packages/sensobol/index.html

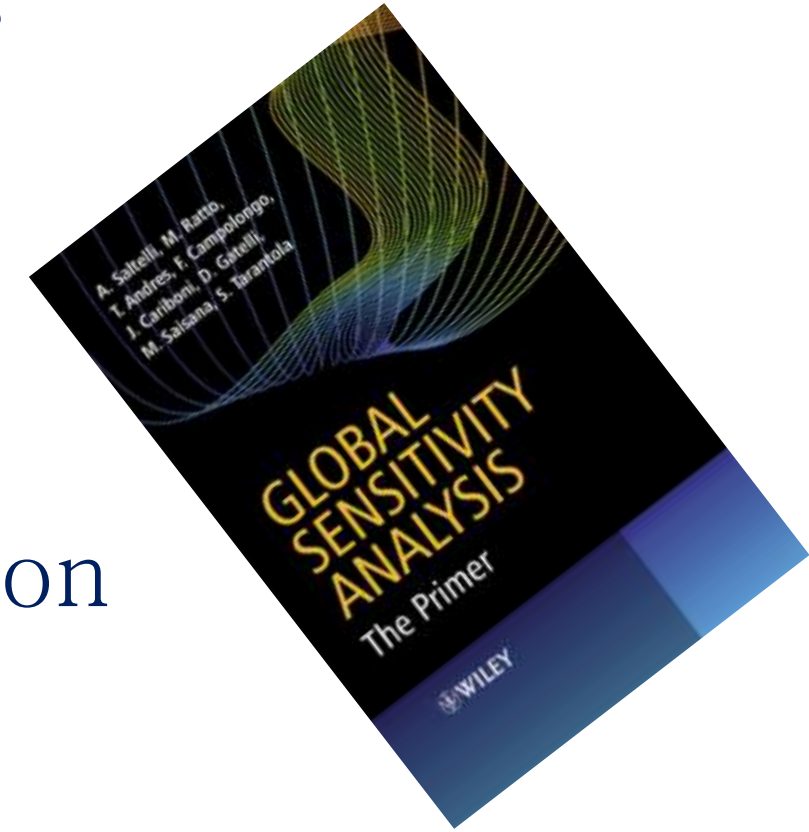
https://www.uqlab.com/ (in MatLab, by Bruno Sudret and his team)

SALib https://salib.readthedocs.io/en/latest/

…but there is more, in R, Phython, Giulia …

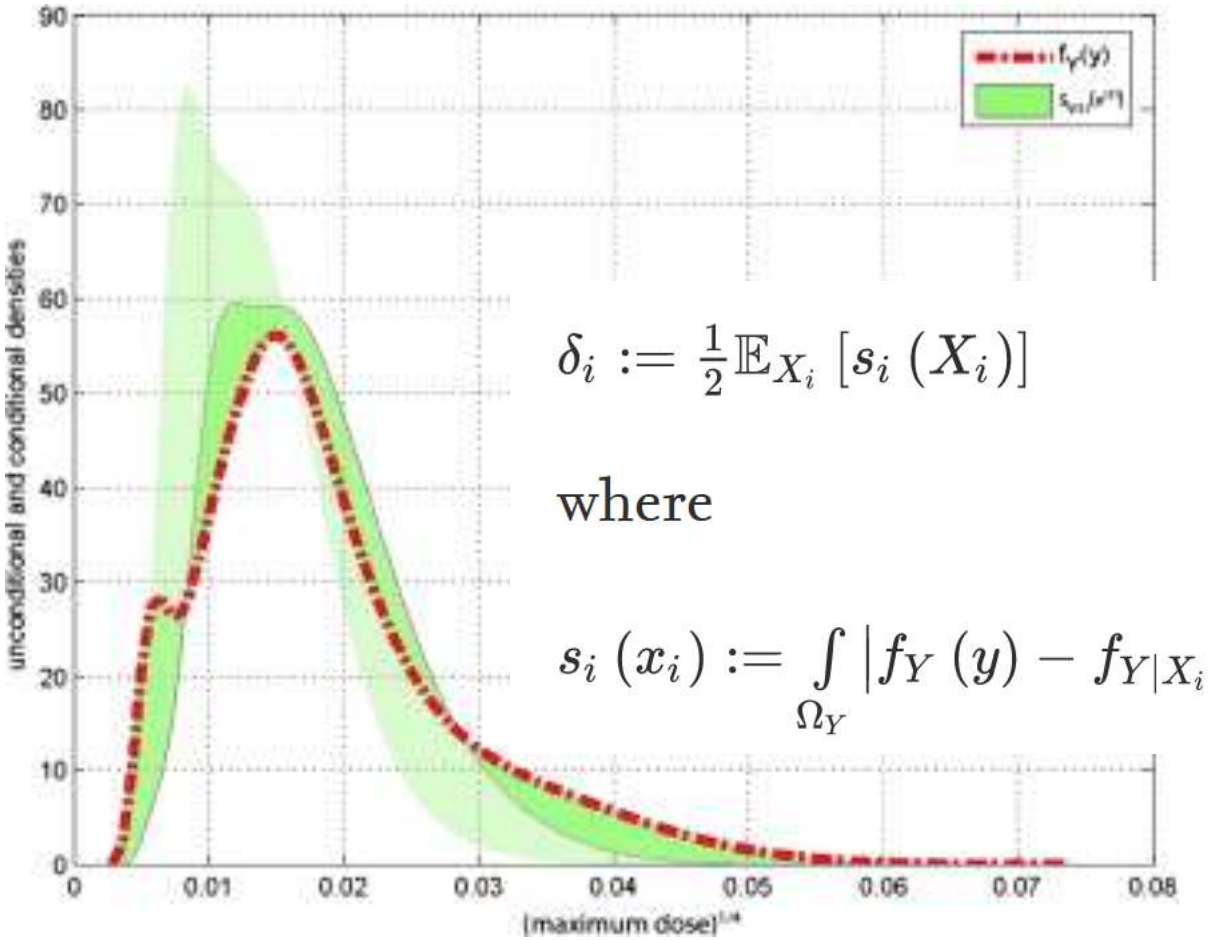
Advantages with variance based methods:

- graphic interpretation scatterplots
- statistical interpretation (ANOVA)
- expressed plain English
- working with sets
- relation to settings such as factor fixing and factor prioritization
- give the effective dimension



Chapter 1 and its exercises

… but there are other methods that can be used for different settings, e.g. moment independents methods, Shapley coefficients, reduced spaces, VARS …



$$\delta_i := \frac{1}{2}\mathbb{E}_{X_i}\left[s_i\left(X_i\right)\right]$$

where

$$s_i\left(x_i\right) := \int_{\Omega_Y} \left| f_Y\left(y\right) - f_{Y|X_i=x_i}\left(y\right) \right| dy$$

ELSEVIER

Environmental Modelling & Software

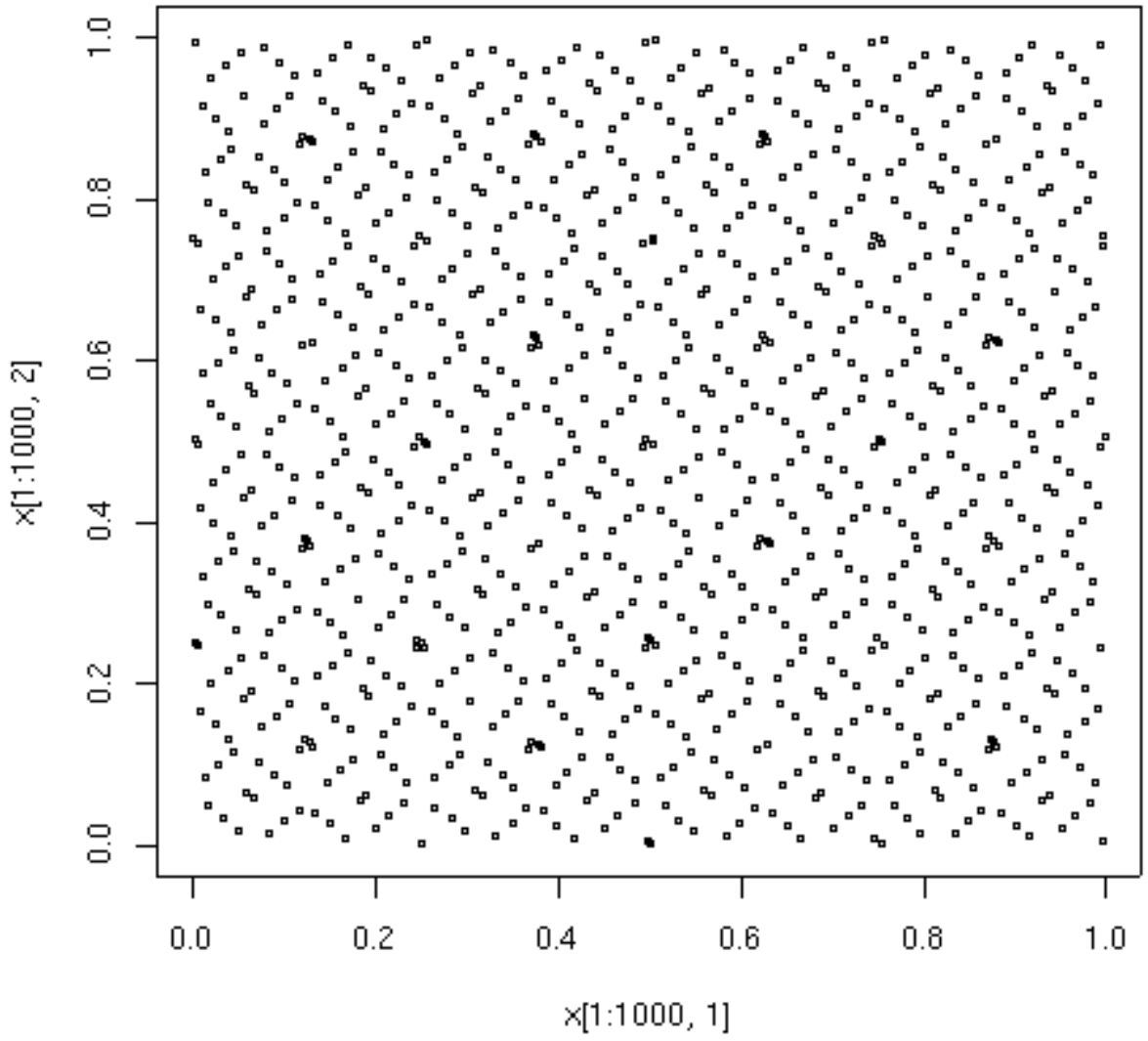
Volume 34, June 2012, Pages 105-115

Model emulation and moment-independent sensitivity analysis: An application to environmental modelling

E. Borgonovo [a], W. Castaings [b, c], S. Tarantola [d]

Quasi random sequences

Ilya M. Sobol’



Cornell University

the

arXiv.org > stat > arXiv:1505.02350

Search...

Help | Adva

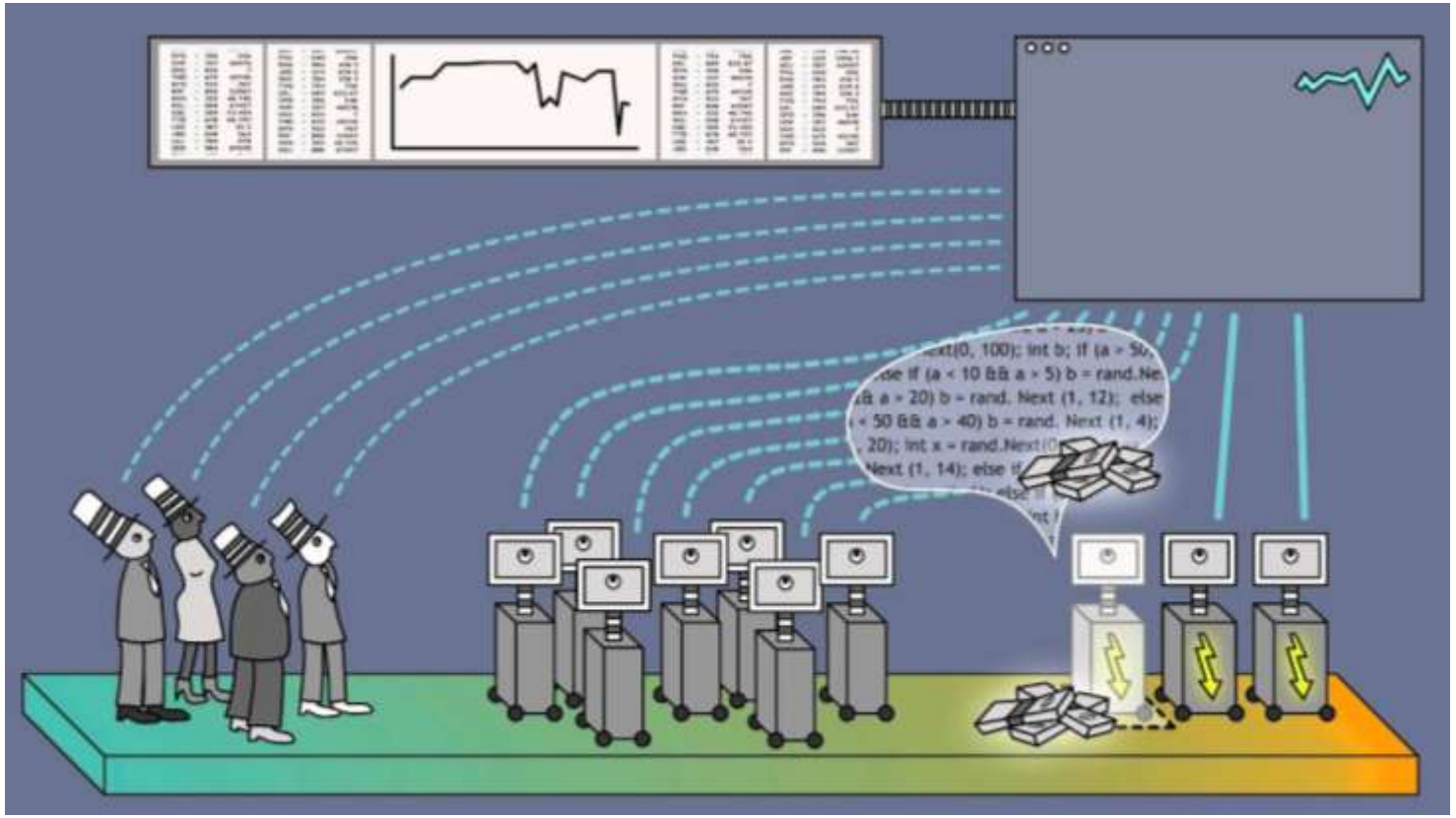
Statistics > Applications

[Submitted on 10 May 2015]

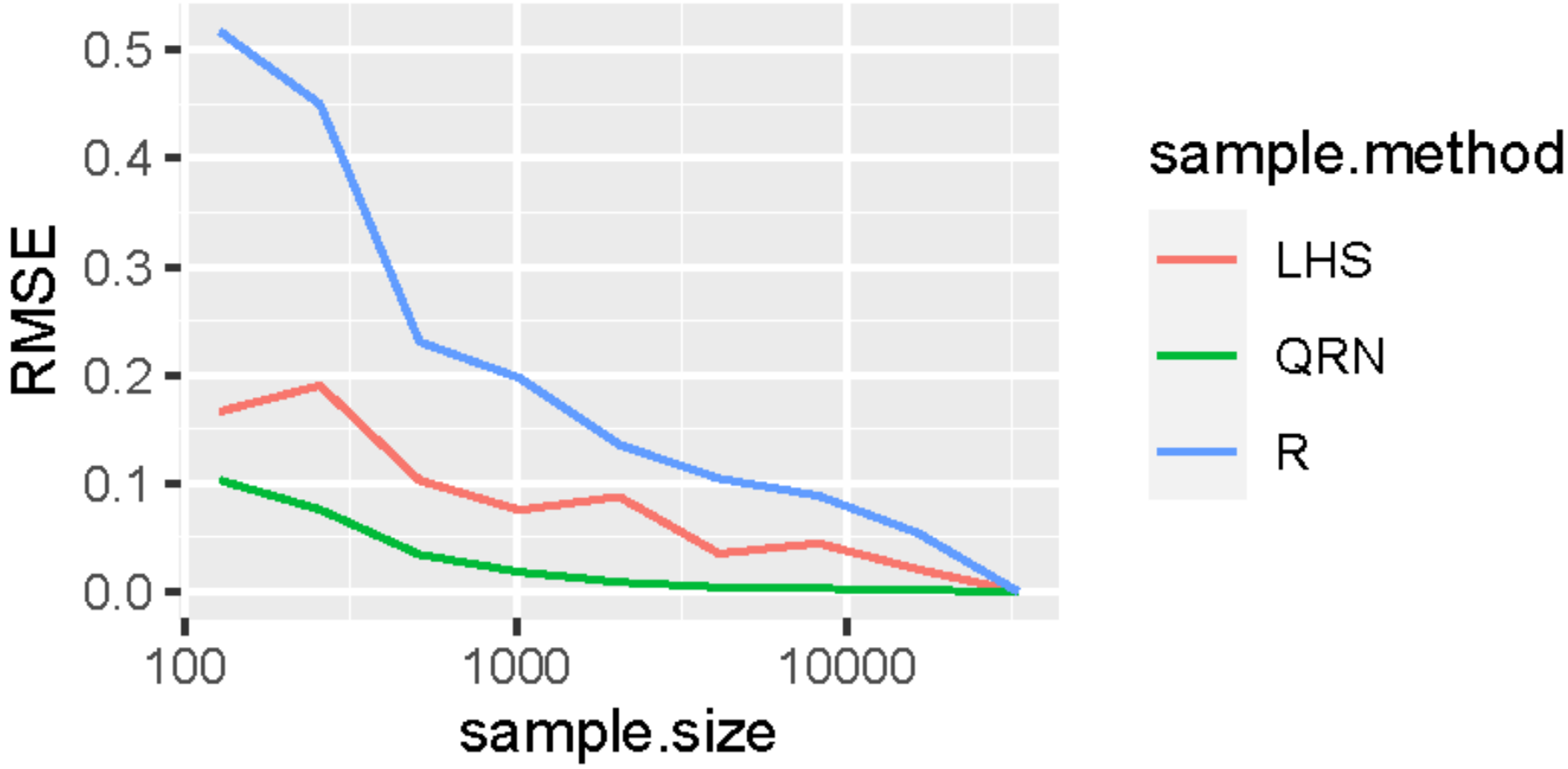
Exploring multi-dimensional spaces: a Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques

Sergei Kucherenko, Daniel Albrecht, Andrea Saltelli

Sobol' LP-TAU are used in high frequency trading



Source: https://www.youtube.com/watch?v=z4nCTdQlH8w



Root mean square error with different designs

What method would one choose to perform sensitivity analysis?

Most of the sensitivity analysis found in the literature are local or otherwise OAT (One factor At a Time)

$$y = f(x_1, x_2, \dots x_k)$$

$$\frac{x_i^0}{y^0}\frac{\partial y}{\partial x_i}\bigg|_{x_i = x_i^0} \longleftarrow \text{Local}$$

What method would one choose to perform sensitivity analysis?

Most of the sensitivity analysis found in the literature are local or otherwise OAT (One factor At a Time)

$$y = f(x_1, x_2, \dots x_k)$$

$$\left.\frac{std(x_i)}{std(y)}\frac{\partial y}{\partial x_i}\right|_{x_i = x_i^0} \longleftarrow \text{Hybrid}$$

$$\left.\frac{\partial y}{\partial x_i}\right|_{x_i=x_i^0}$$ ← Effect on y of perturbing x_i around its nominal value

$$\left.\frac{x_i^0}{y^0}\frac{\partial y}{\partial x_i}\right|_{x_i=x_i^0}$$ ← Effect on y of perturbing x_i both 'normalized' (divided by their nominal value)

$$\left.\frac{std(x_i)}{std(y)}\frac{\partial y}{\partial x_i}\right|_{x_i=x_i^0}$$ ← Effect on y of perturbing x_i both 'standardized' (divided by their standard deviation)