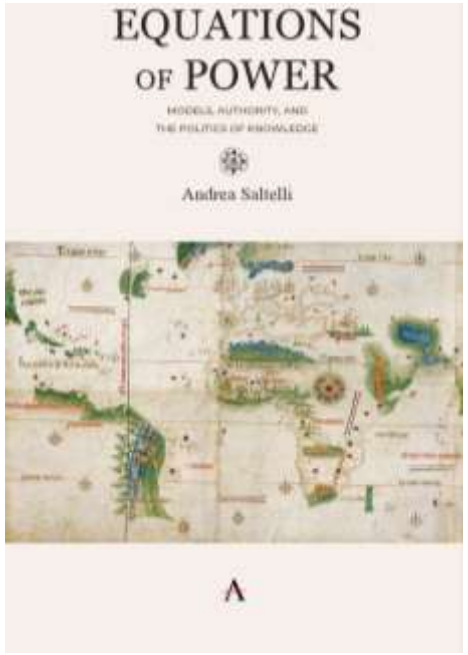
Methodology for comparative model assessment and validation

By Andrea Saltelli, at the Faculty of Management and Economics of the Gdańsk University of Technology (Politechnika Gdańska), October 28 2026 – 28 października 2026, godz. 13:15–15:00, sala 103/104 WZiE

Based on Chapter 5 of Equations of Power: Models, Authority, and the Politics of Knowledge (Anthem Press, 2027)



“Ce qui est simple est toujours faux. Ce qui ne l’est pas est inutilisable.”

(“If it’s simple, it’s always false. If it’s not, it’s unusable.”)

— Paul Valéry, Œuvres II (1942)

Paul Valéry, Œuvres II (1942), Bibliothèque de la Pléiade, Gallimard, Paris, 1960, p. 864. Epigraph of Chapter 5, Equations of Power.

Sensitivity analysis and validation

"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." (*Italics* by Hall)

Charles A. S. Hall, ecologist

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, pp. 1–38. Springer, Singapore. p. 13.

The map for today: technical versus normative quality

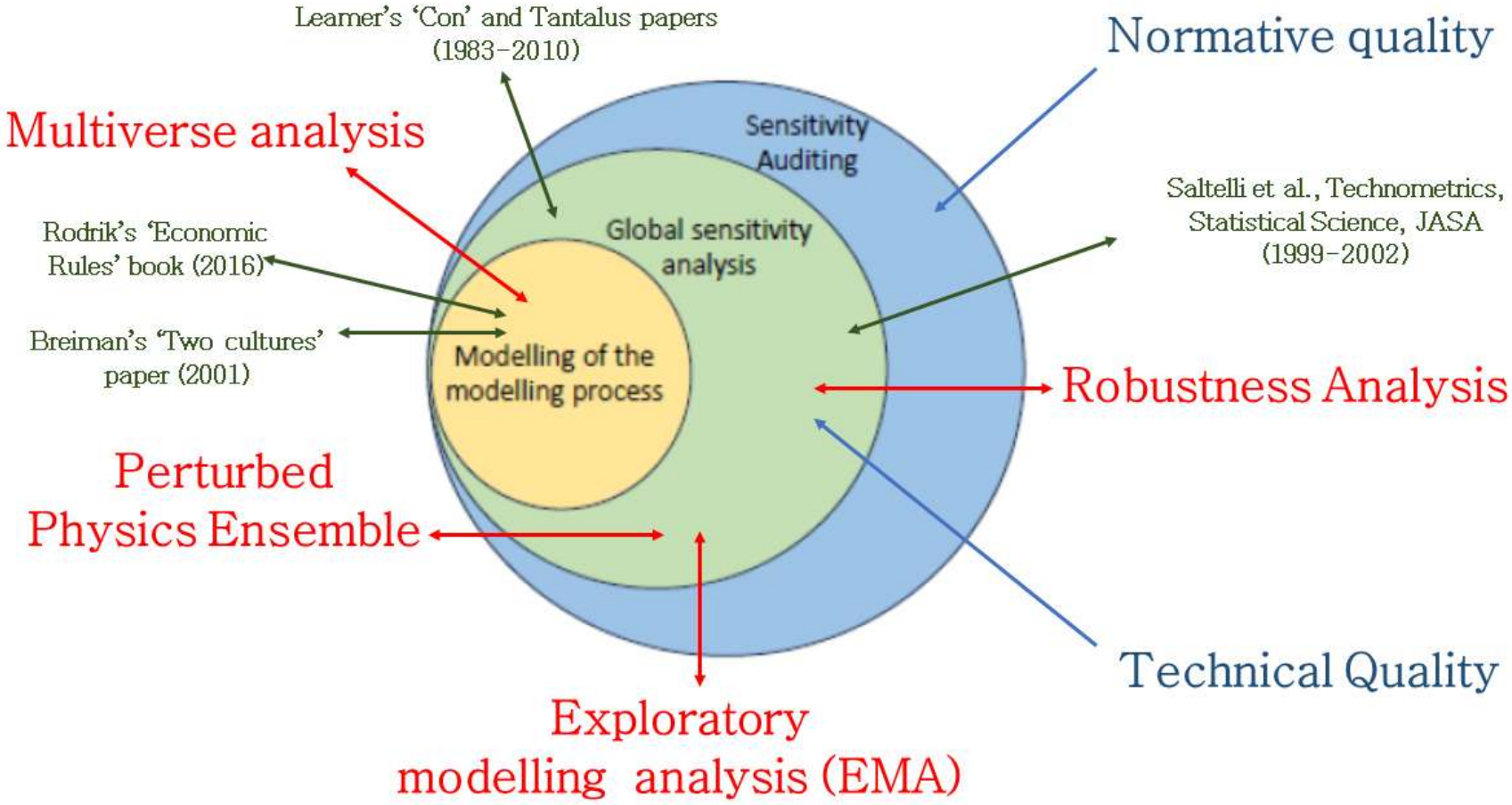
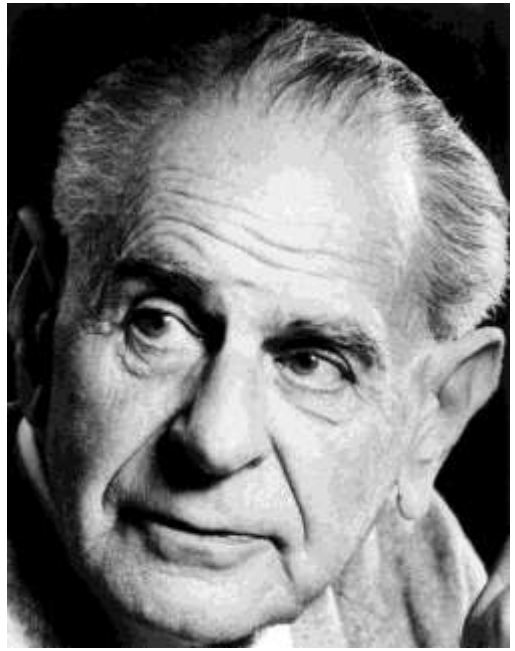


Figure 5.1, Equations of Power: Models, Authority, and the Politics of Knowledge (Anthem Press, 2027).

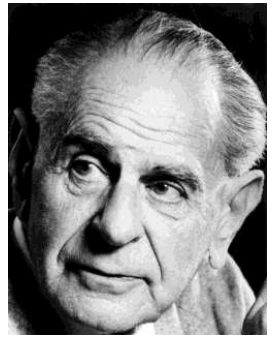
Part I – Global sensitivity analysis as a severe test



Karl Popper
(1902-1994)

Models that take no risks

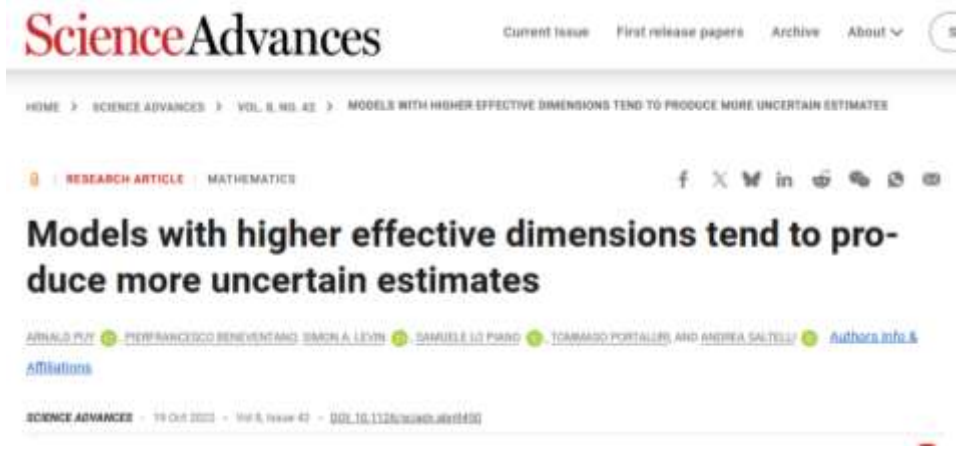
Popper: The empirical content of a theory is measured by the risks it takes — the observations that could prove it wrong (severe tests)

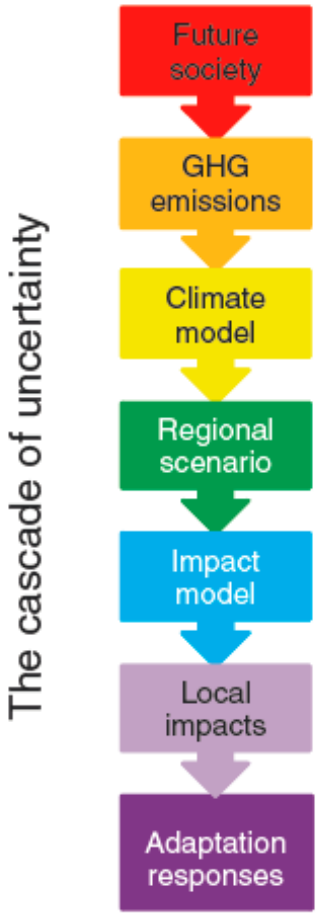


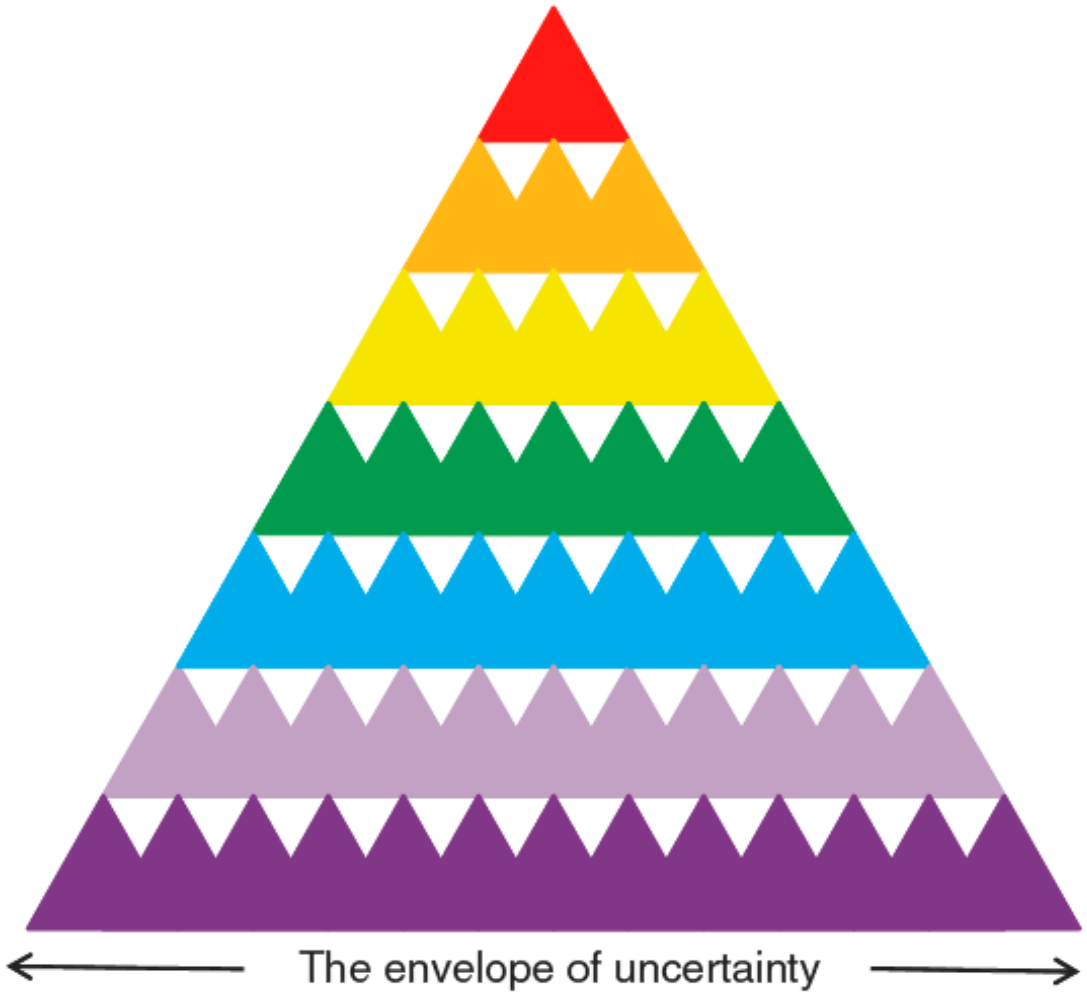
- Climate models projecting the cost of CO2 to 2300.
- Nuclear waste safety models covering a million years.
- Digital twins simulating ecosystem dynamics with 50-year horizons

These models take no risks. No severe tests here

Uncertainty analysis: the study of the uncertainty in model output

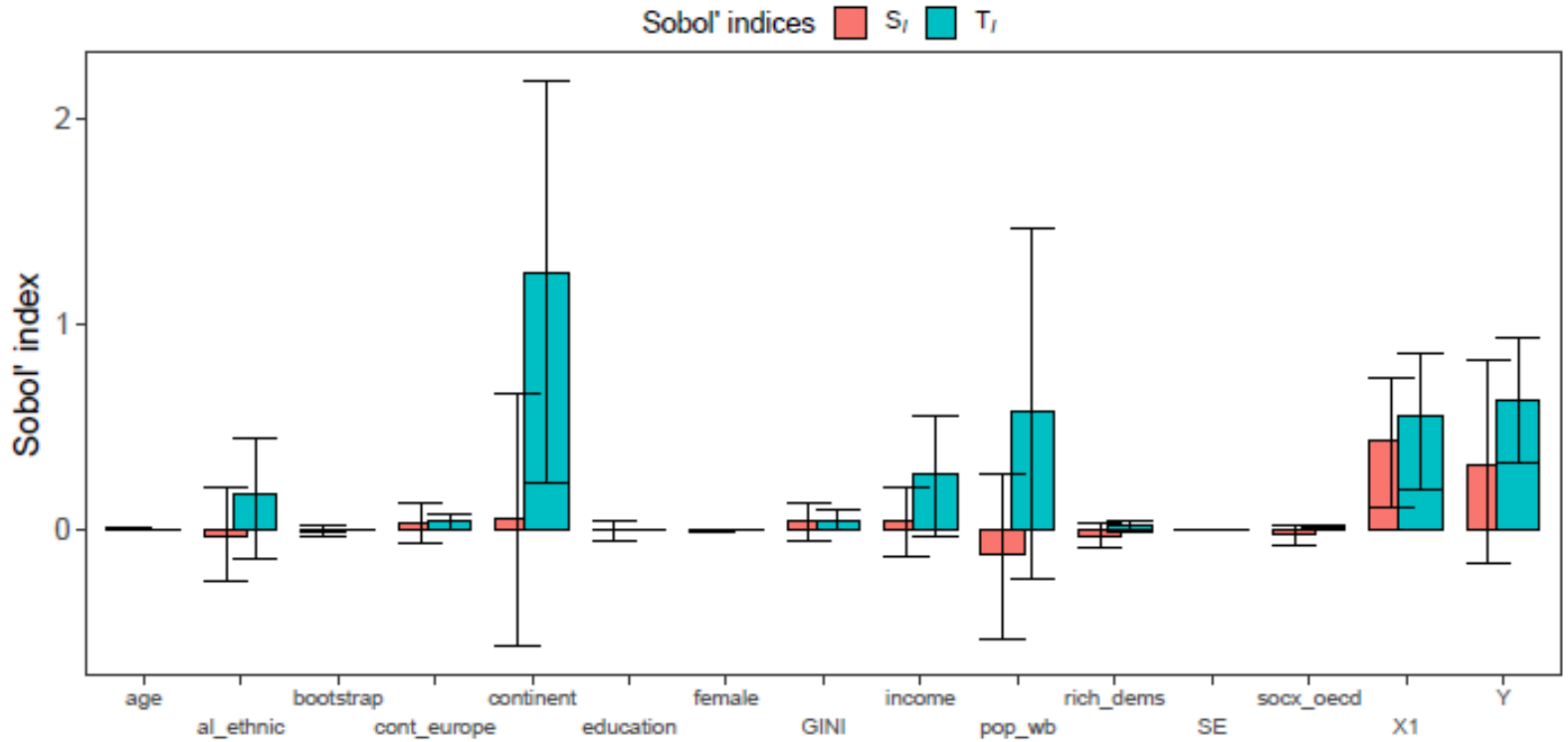






Source: https://www.climate-lab-book.ac.uk/2014/cascade-of-uncertainty/

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



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

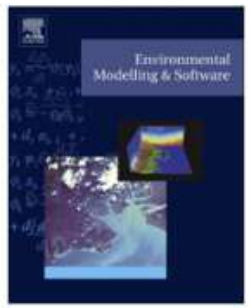
A geometric proof

Environmental Modelling & Software 25 (2010) 1508–1517

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

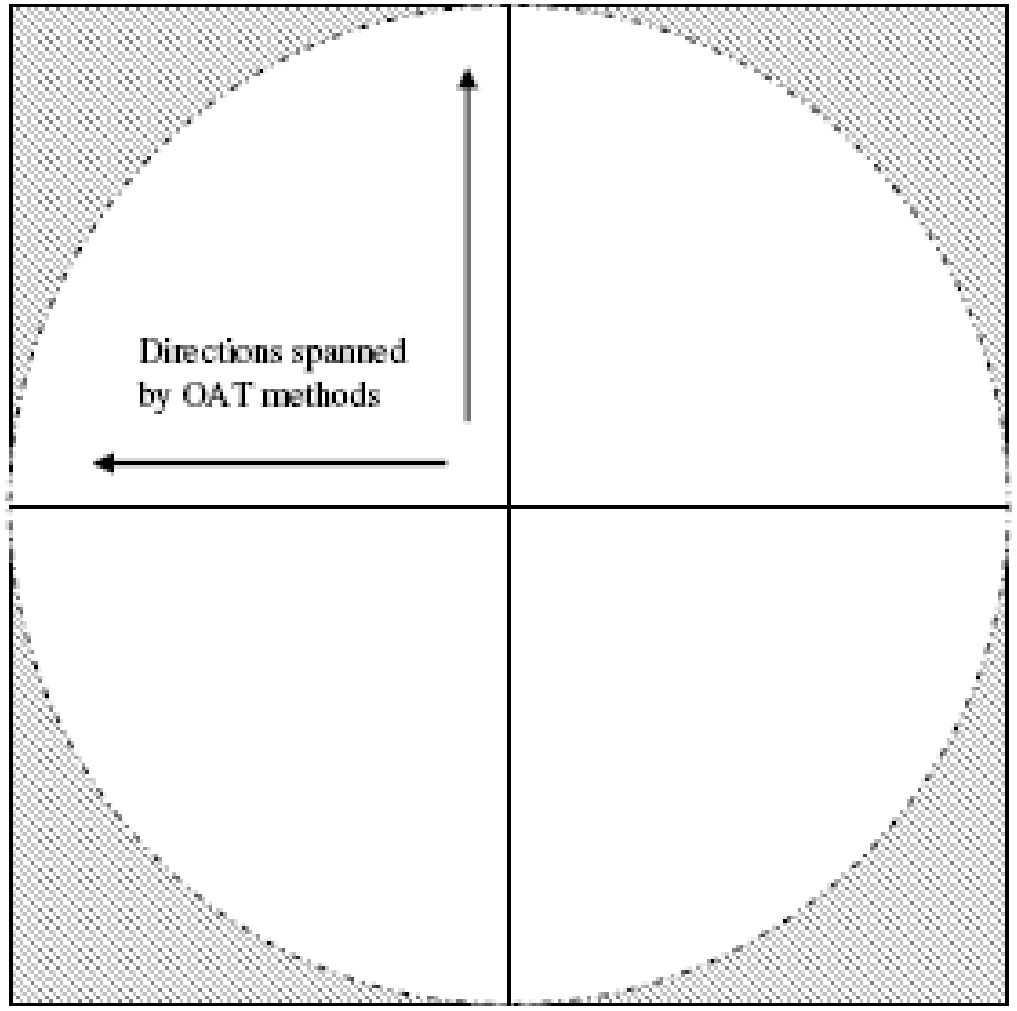


How to avoid a perfunctory sensitivity analysis

Andrea Saltelli*, Paola Annoni

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

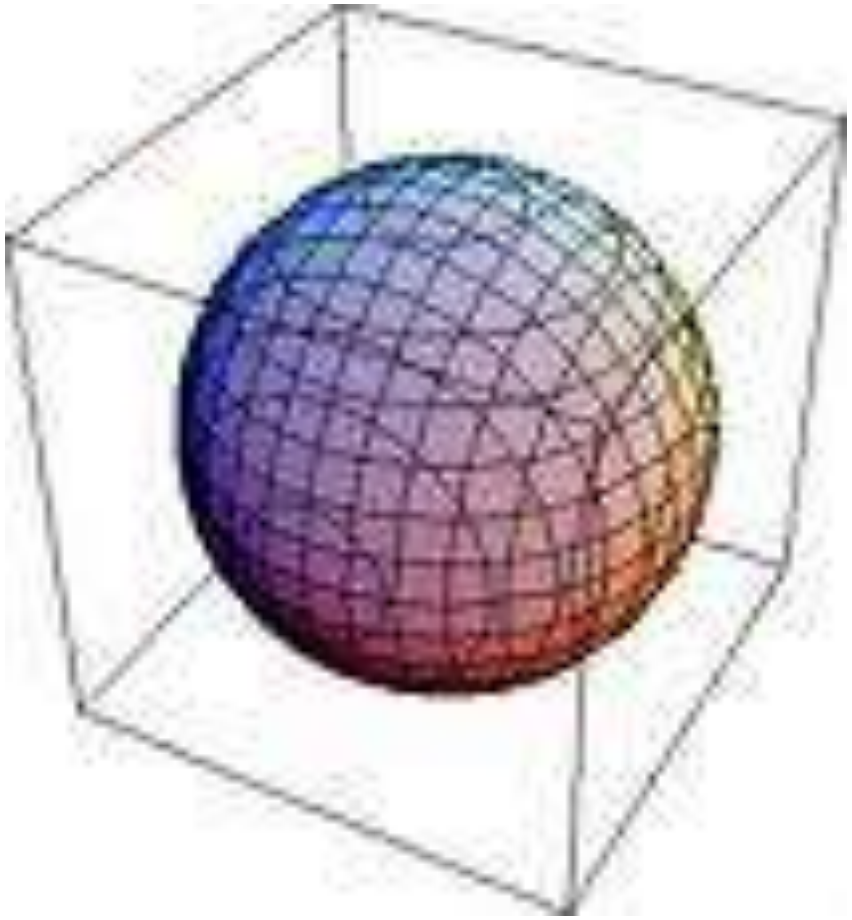
OAT in 2 dimensions



Area circle / area square =?

~ 3/4

OAT in 3 dimensions



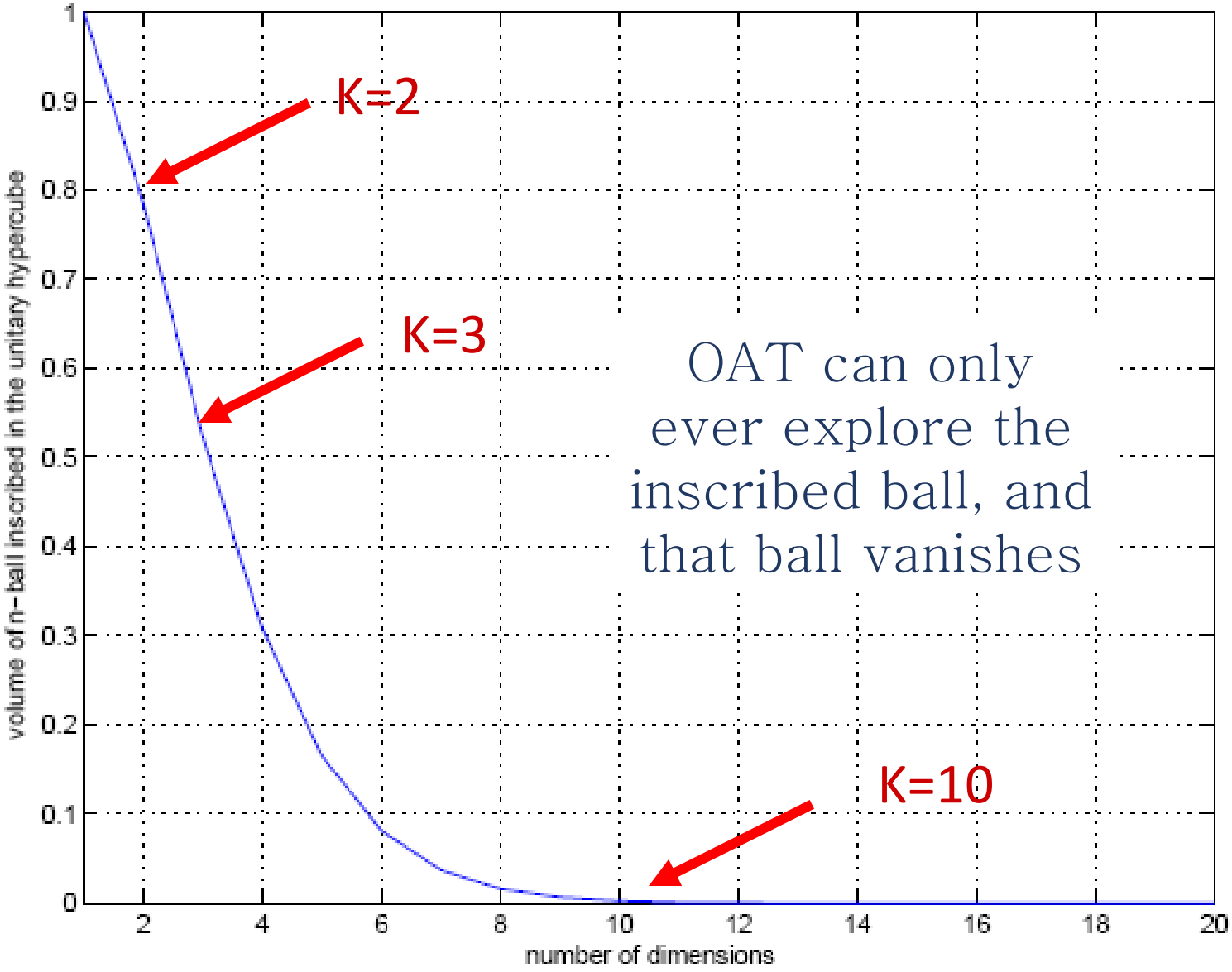
Volume sphere / volume cube =?

~ 1/2

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

~ 0.0025

OAT in k dimensions

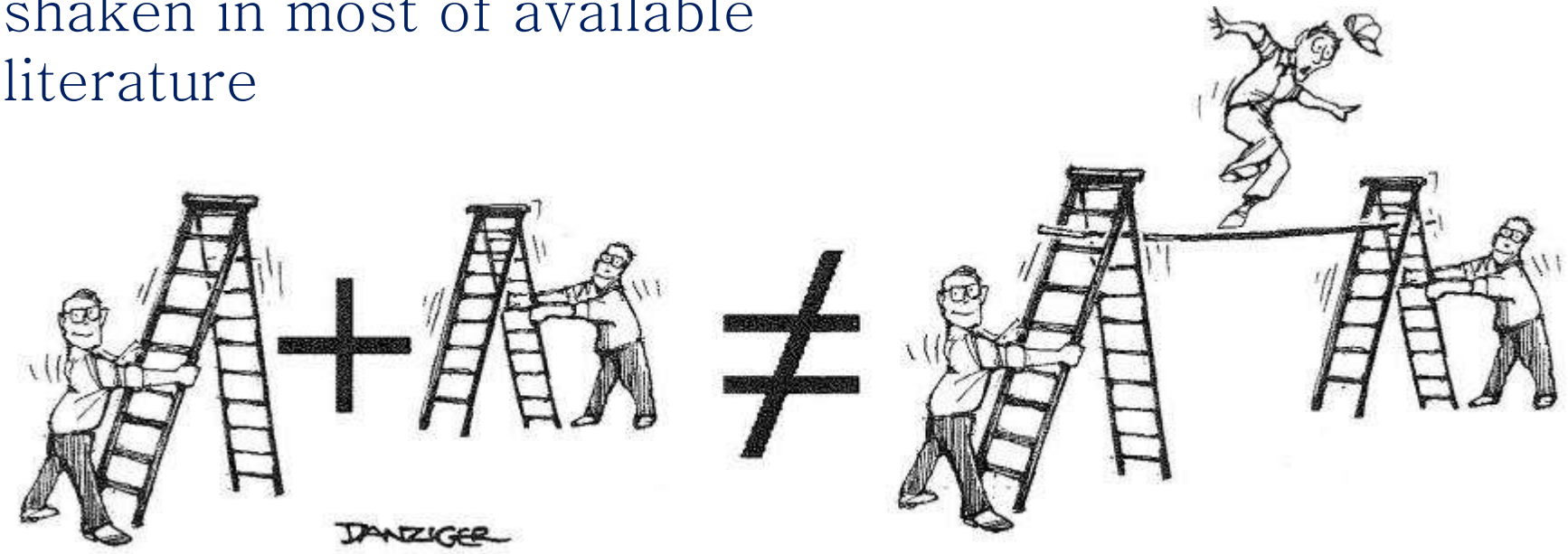


OAT does not capture interactions

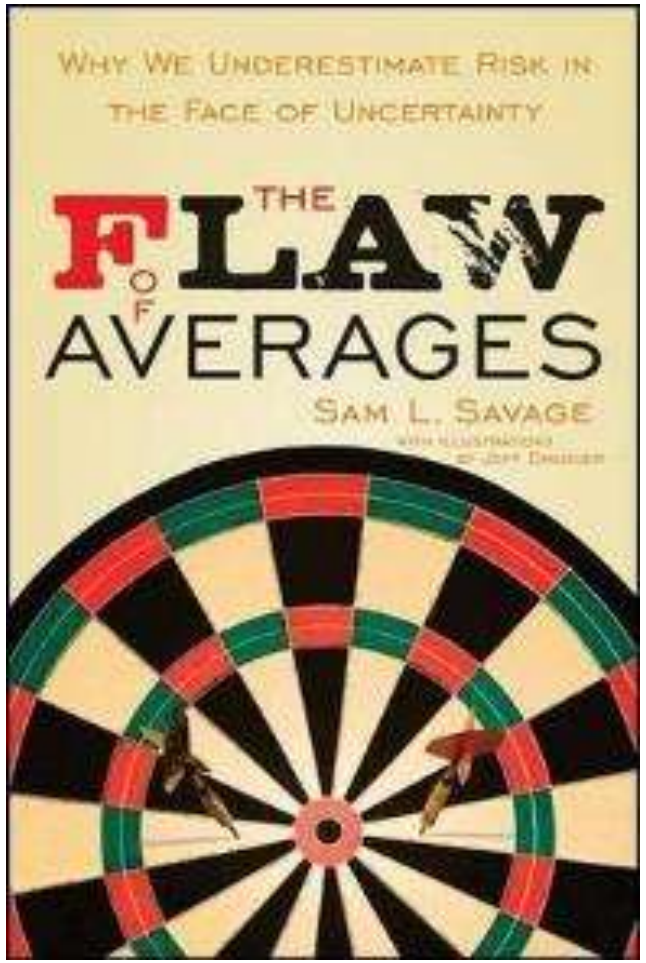
➔ The resulting analysis is non conservative

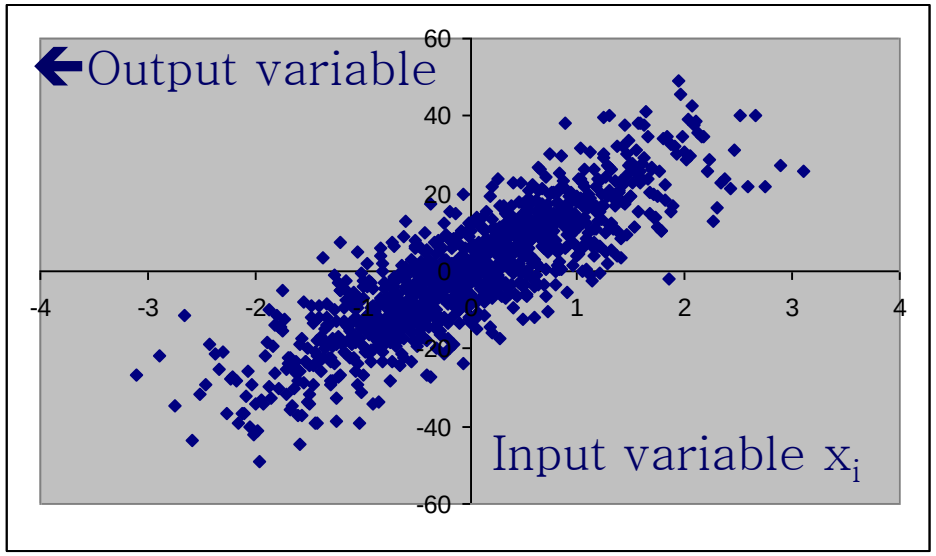
How would you test the scaffolding?

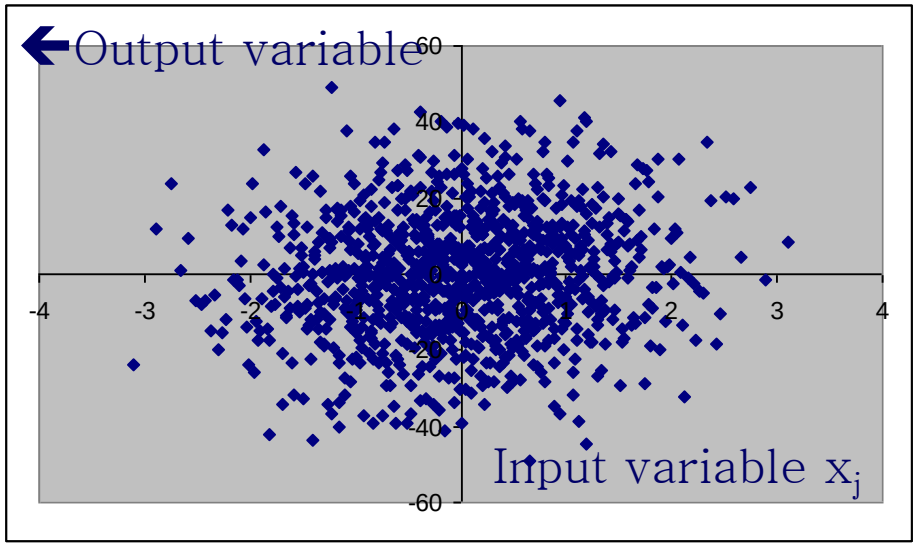
How coupled ladders are shaken in most of available literature



How to shake coupled ladders

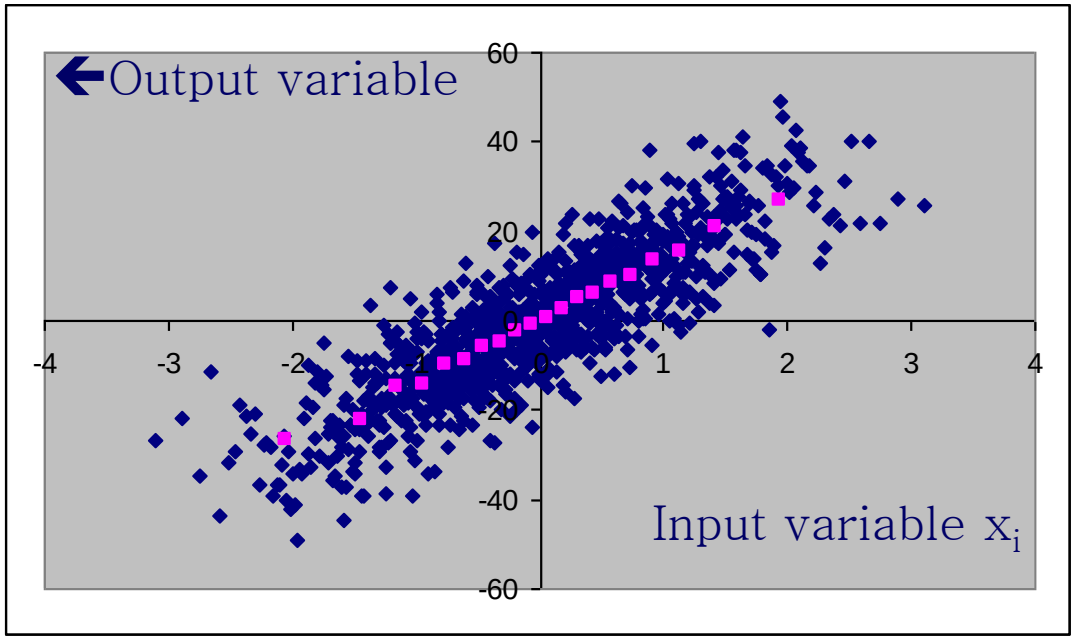




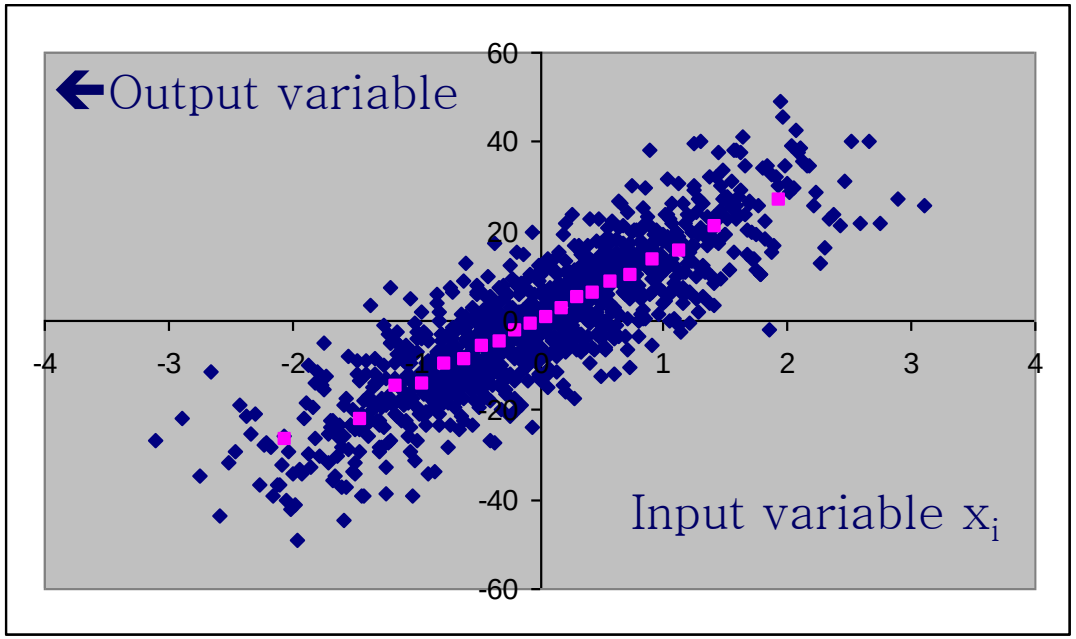


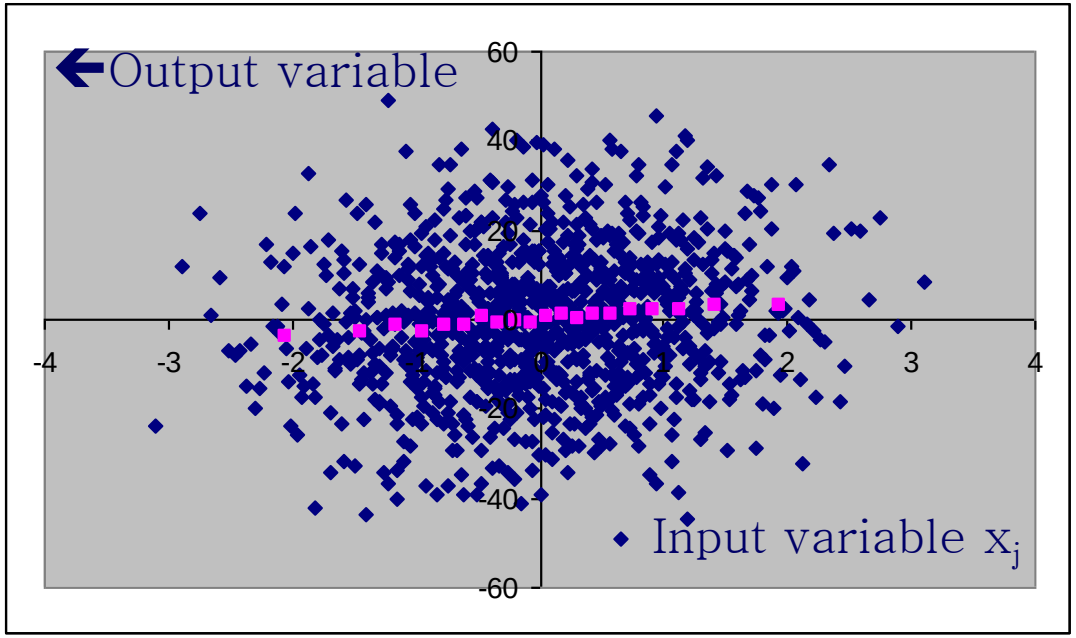
Plotting the output as a function of two different input factors

Which factor is more important?



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





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

Total effect term

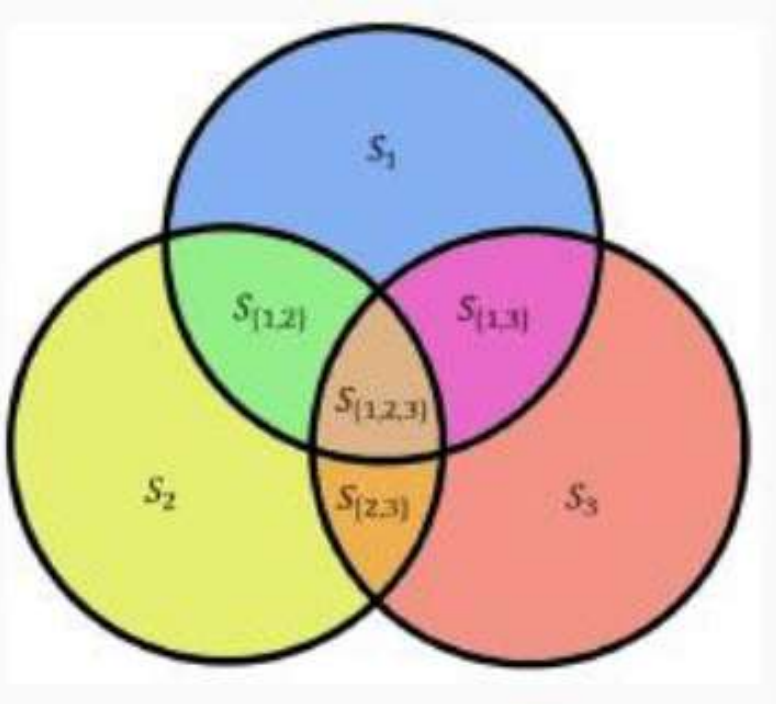
$$T_i = \frac{V_{X_i}\left(E_{\mathbf{X}_{\sim i}}\left(Y \middle| X_i\right)\right)}{V(Y)}$$

Homma, Toshimitsu, and Andrea Saltelli. 1996. ‘Importance Measures in Global Sensitivity Analysis of Nonlinear Models’. *Reliability Engineering & System Safety* 52 (1): 1–17. https://doi.org/10.1016/0951-8320(96)00002-6.

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



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

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

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

“Sensitivity analyses would help” (Leamer, 1985)

JOURNAL ARTICLE

Sensitivity Analyses Would Help

Edward E. Leamer

The American Economic Review
Vol. 75, No. 3 (Jun., 1985), pp. 308-313 (6 pages)
Published By: American Economic Association
https://www.jstor.org/stable/1814801

GSA as a severe test

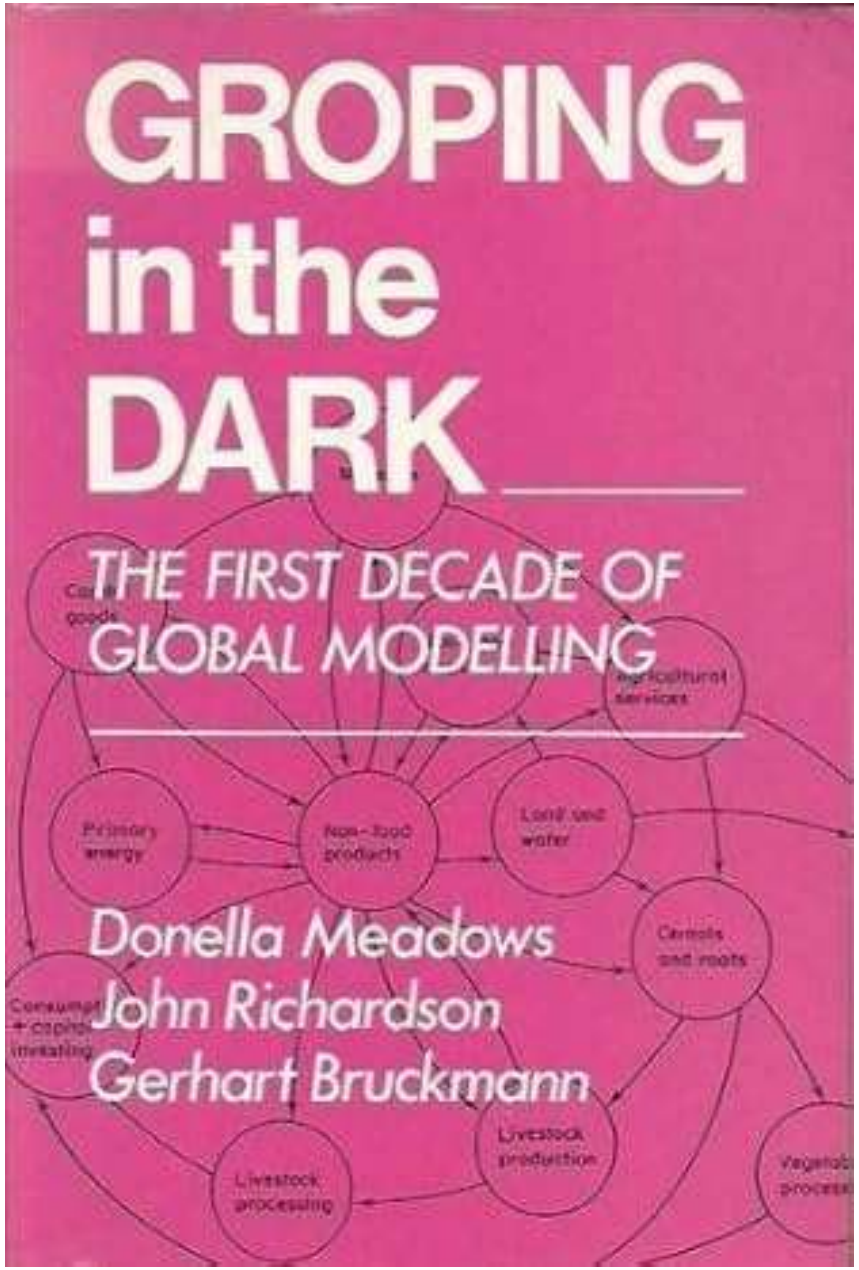
Does the conclusion survive a test designed to refute it?

When the inference dissolves under the exploration of its own input space, the model's lack of empirical content is exposed.

Edward E. Leamer, 2010

"One reason these methods are rarely used is their honesty seems destructive; or, to put it another way, a fanatical commitment to fanciful formal models is often needed to create the appearance of progress."

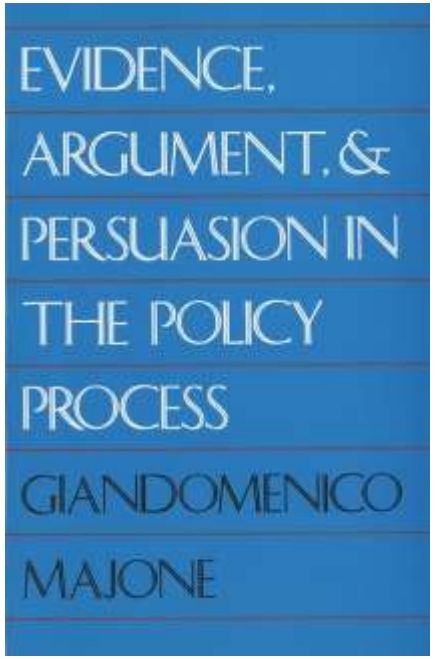
Leamer, E. E. (2010). Tantalus on the road to asymptopia. Journal of Economic Perspectives, 24(2), 31–46.



1982

"We call a parameter sensitive **not when a change in it changes the value of an output number but when it changes the policy conclusions** we would draw from the model" (p. 132)

Giandomenico Majone



- Data
- Analyst
- Model
- Objective

"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.

Which cause should drive the result?

Material cause = the data

The result should depend on it

Final cause = the objective

The result should depend on it

Formal cause = the model

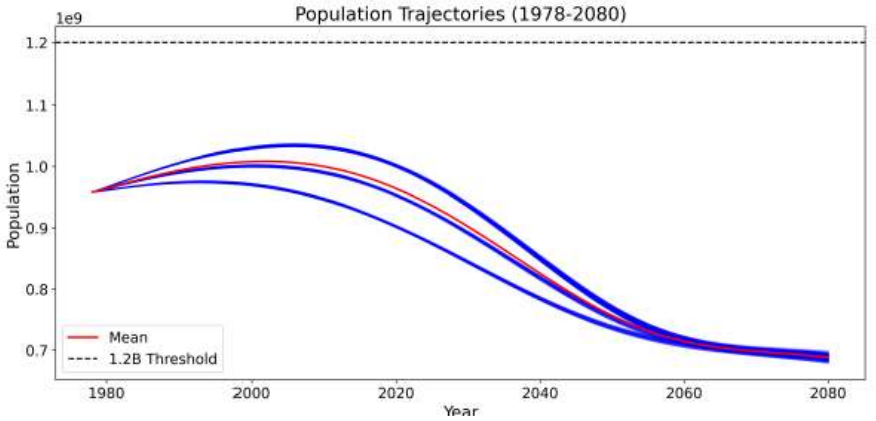
The result should depend on it least

Efficient cause = the analyst

The result should depend on it least

Models should be sensitive to the data and the objective, and insensitive to the model and the analyst using it.

The Song Jian / one-child policy 1980's model was found to be near-insensitive 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

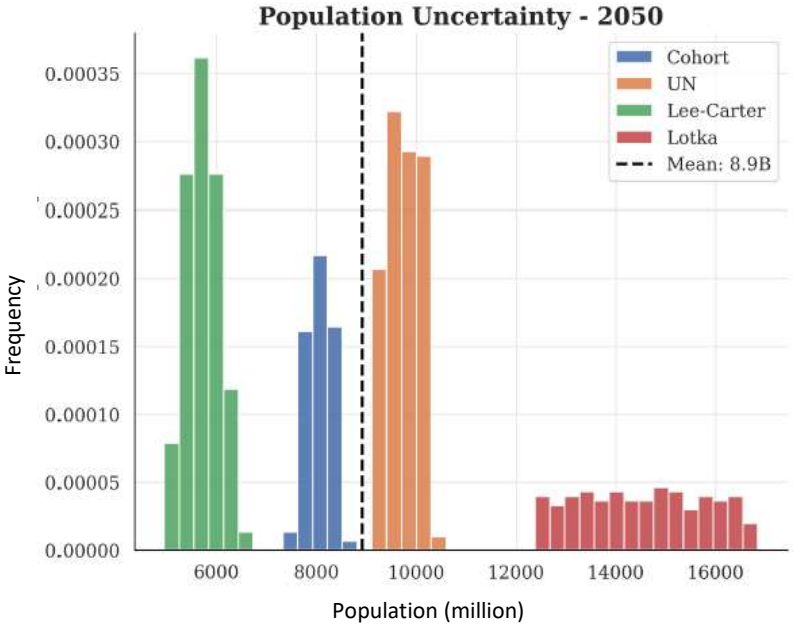
Marta Kuc Czarnecka
Gdańsk University of Technology

Roger Pielke
University of Colorado at Boulder

Andrea Saltelli
UPF Barcelona School of Management

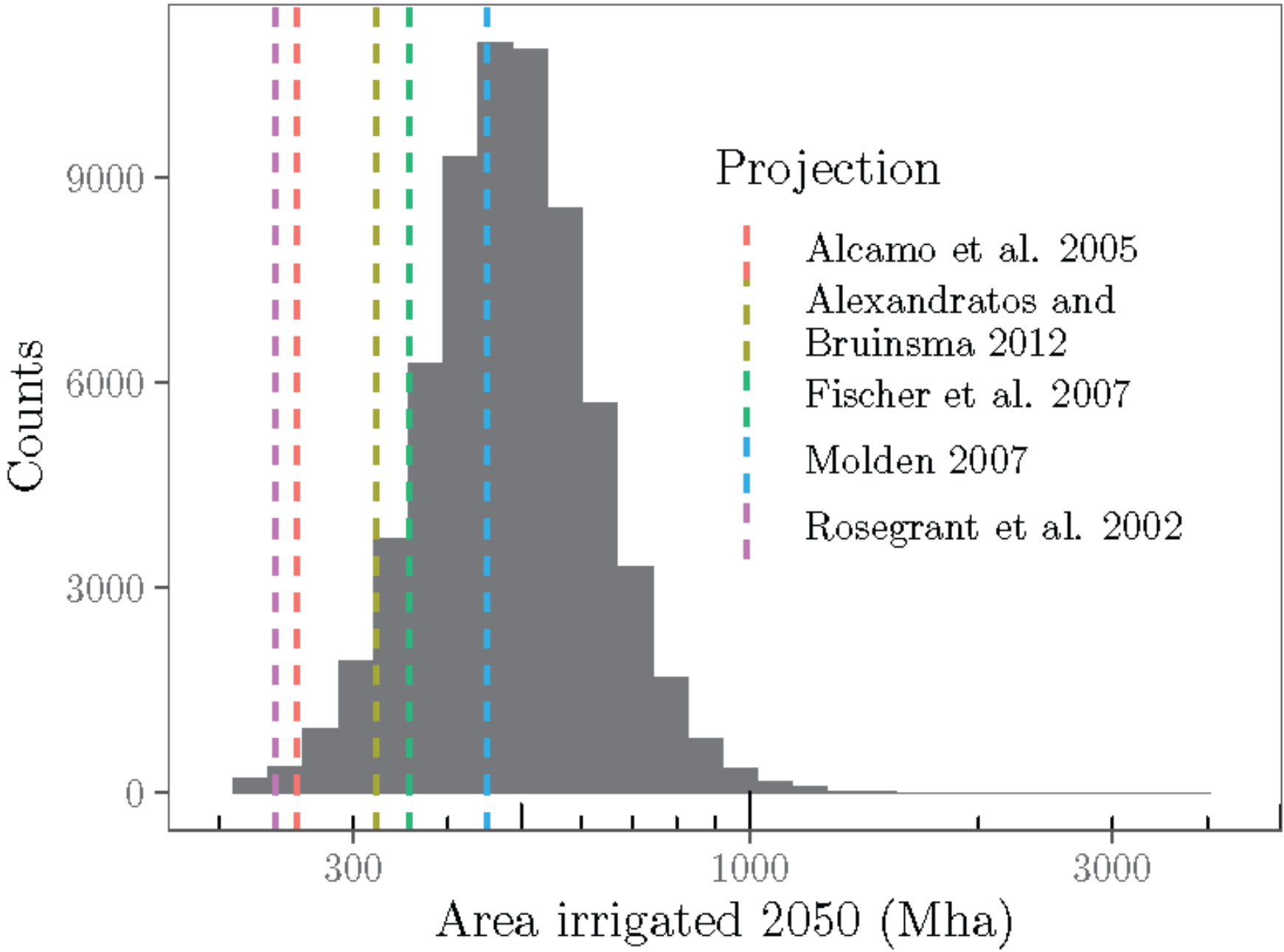
Date Written: May 15, 2026

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



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

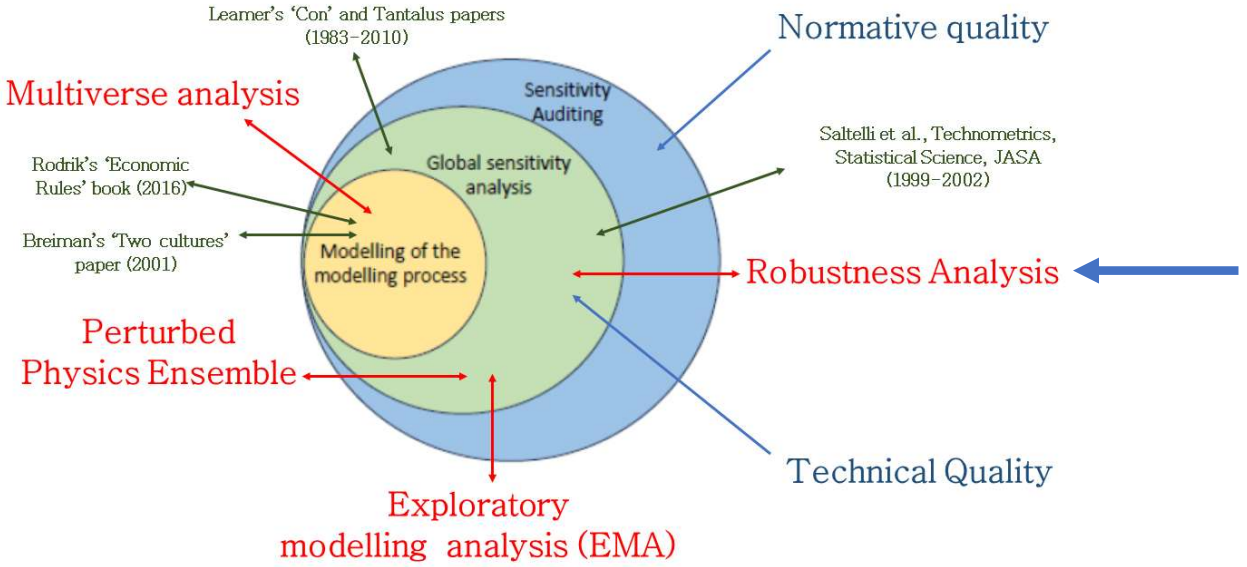


Citation:
Puy, A., Lo Piano, S., & Saltelli, A. (2020). Current models underestimate future irrigated areas. *Geophysical Research Letters*, 47, e2020GL087360. https://doi.org/10.1029/2020GL087360

Part II – From one model to many

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.

Robust theorems: “the intersection of independent lies”

“Therefore, we attempt to treat the same problem with several alternative models each with different simplifications but with a common biological assumption. Then, if these models, despite their different assumptions, lead to similar results we have what we can call a robust theorem which is relatively free of the details of the model. Hence our truth is the intersection of independent lies.” (Levins, 1966, p. 423)

Orzack & Sober (1993)

Critical: Robustness is a non-empirical form of confirmation

Weisberg (2006)

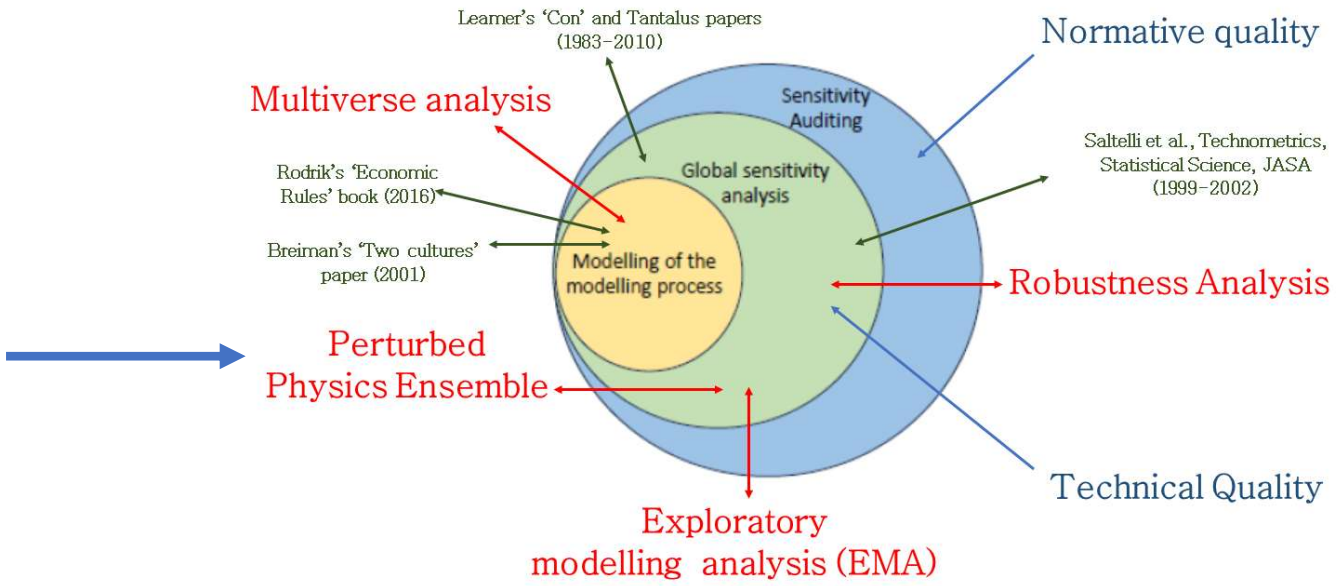
Supportive: A four-step procedure linking common structures across models to robust properties

What GSA adds

The common structure must also be shown to be influential

Levins, R. (1966). The strategy of model building in population biology. American Scientist, 54(4), 421–431. Orzack, S. H., Sober, E. (1993). Quarterly Review of Biology, 68(4), 533–546. Weisberg, M. (2006). Robustness analysis. Philosophy of Science, 73(5), 730–742.

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

Butterflies and hawkmoths

The butterfly effect

Sensitive dependence on initial conditions (SDIC)

Can be explored by propagating uncertainty

The hawkmoth effect

Structural model error (SME)

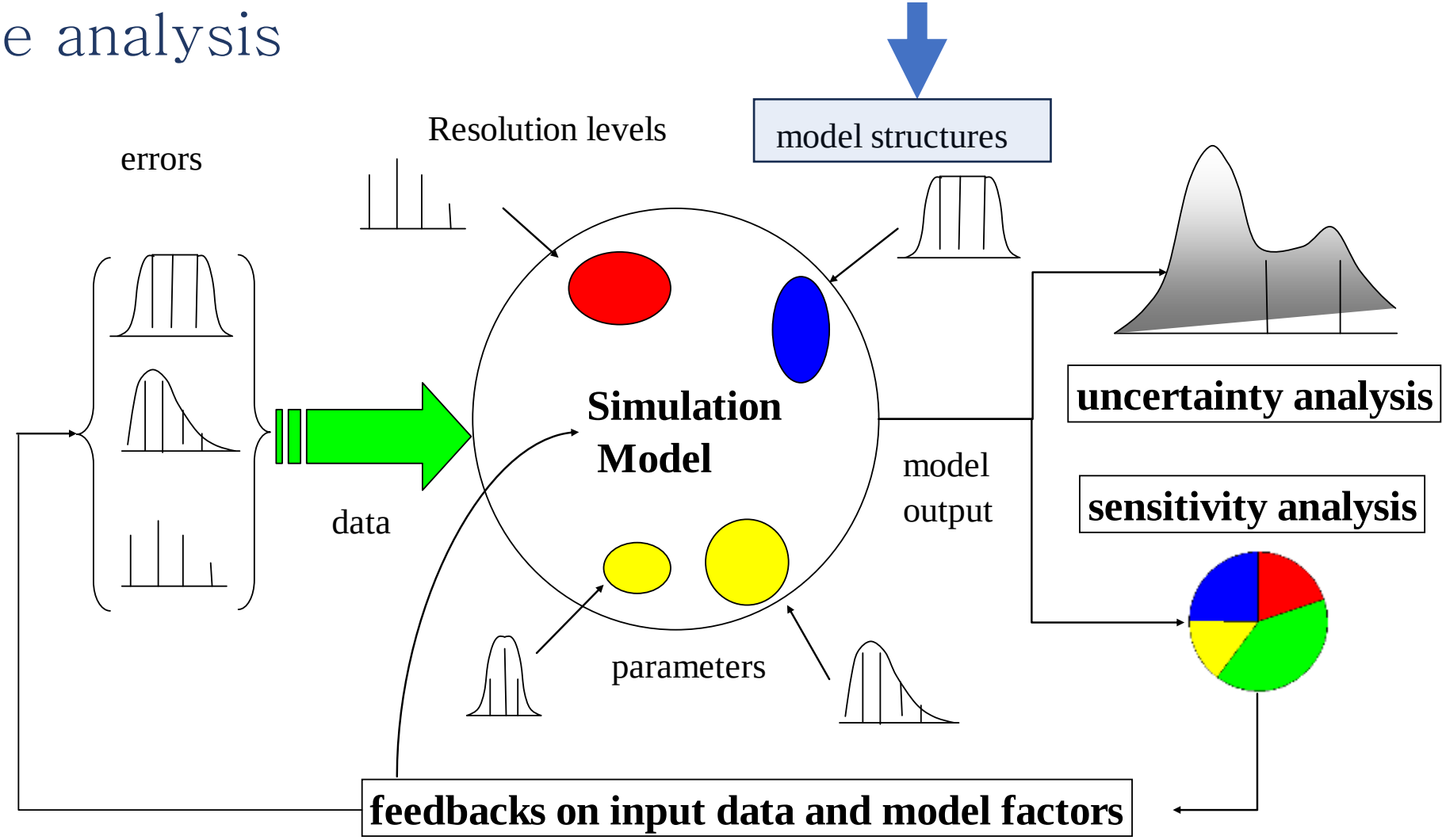
Imposes often terminal limits on predictive power (Frigg et al., 2014)

Yet structural uncertainty **can itself be propagated** — e.g. by sampling among alternative formulations of the same physical law (evapotranspiration).

Via negativa (N.N. Taleb's books): such analyses establish that a model cannot support the inference asked of it — and may cure the modeller of hubris.

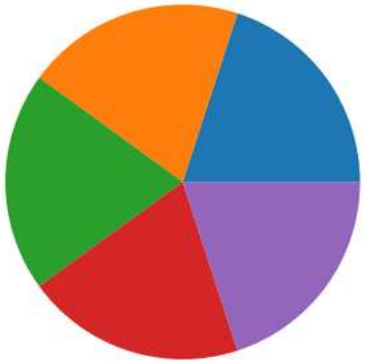
Frigg, R., Bradley, S., Du, H., Smith, L. A. (2014). Laplace's demon and the adventures of his apprentices. Philosophy of Science, 81(1), 31–59.

Sample modelling choices via ‘triggers’. Global sensitivity analysis becomes thus 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

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

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OPEN

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

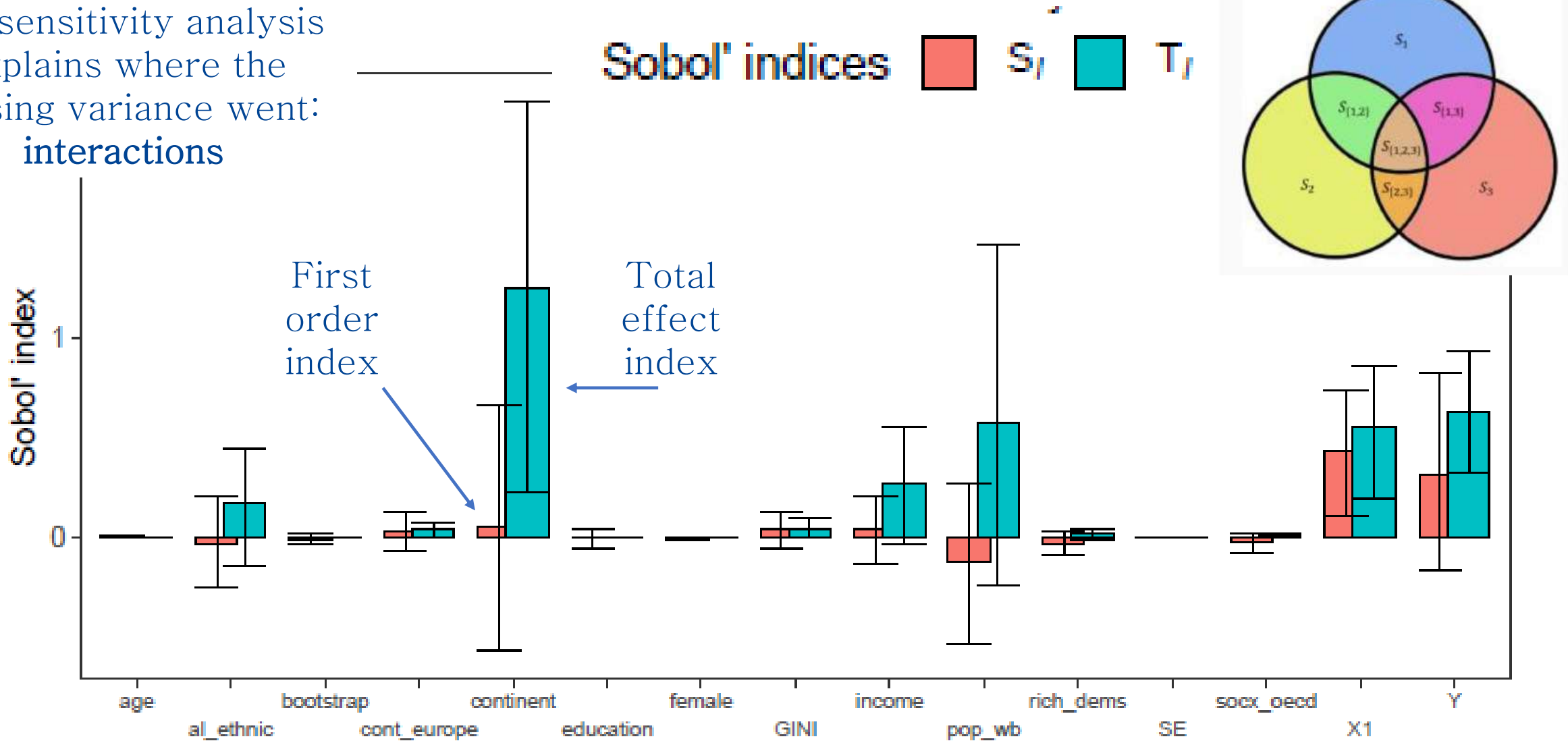
Edited by Douglas Massey, Princeton University, Princeton, NJ; received March 6, 2022; accepted August 22, 2022

The 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 per cent of the variance of the results … Where **does the variance come from?**

https://www.istockphoto.com/

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



Many models fit the same data … ➔ and tell different stories

Equifinality
Hydrology (Beven, 1993)

The Rashomon effect
Machine learning (Breiman, 2001)

The importance of the same variable may vary across models

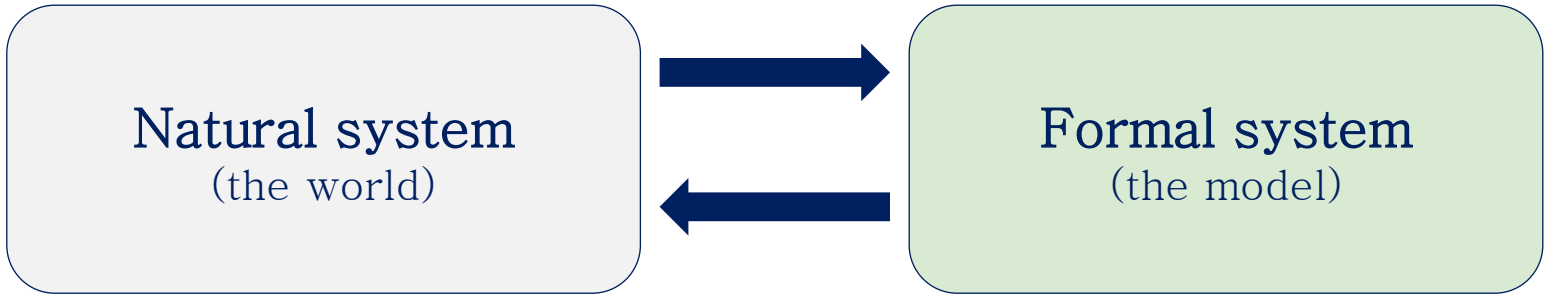
→ a sensitivity analysis that keeps the model fixed may be incomplete

→ hence: modelling of the modelling process

Beven, K. (1993). Advances in Water Resources, 16(1), 41–51. Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science, 16(3), 199–231. Donnelly, J., Katta, S., Borgonovo, E., Rudin, C. (2025). Doctor Rashomon and the UNIVERSE of Madness. arXiv:2510.12734.

Part III – Validation of what? The limits of GSA

The paradox of global sensitivity analysis



GSA operates here: it tests the coherence of the encoding, not whether the decoding informs about the world

"The first act of selection – a sort of proto-sensitivity analysis – occurs in the mind of the modeller, who chooses which aspects of a system are 'relevant'."

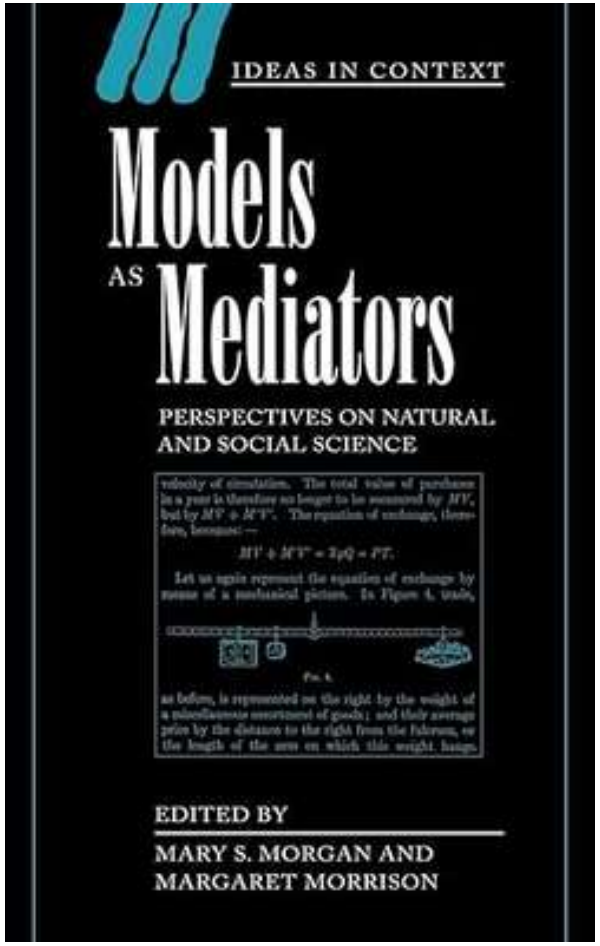
GSA then chases which of those aspects drive the results — but the second selection is subordinate to the first.

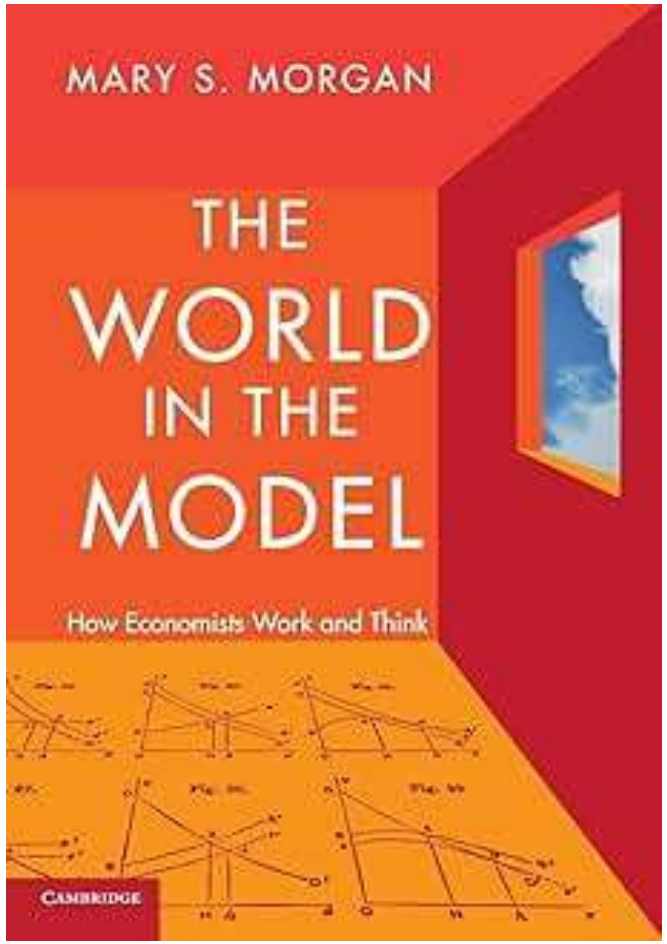
It tests the fabric of the model, rather than the model against the world it claims to represent.

Modelling relation after Rosen, R. (1991). Life Itself. Columbia University Press. Equations of Power, §5.2.

Mary S. Morgan

Margaret Morrison



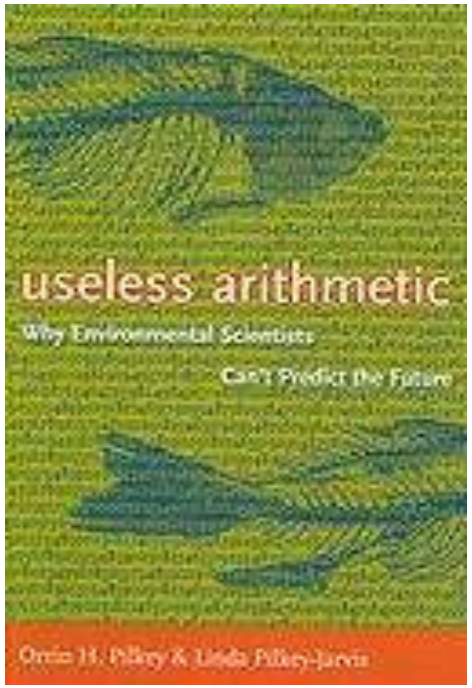


Models as mediators between theories and the world.

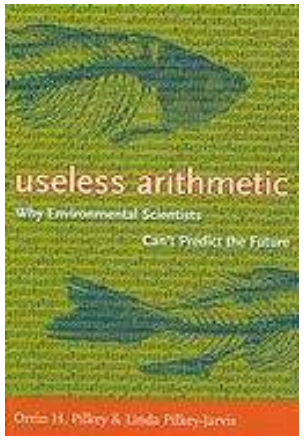
World in the model, model in the world

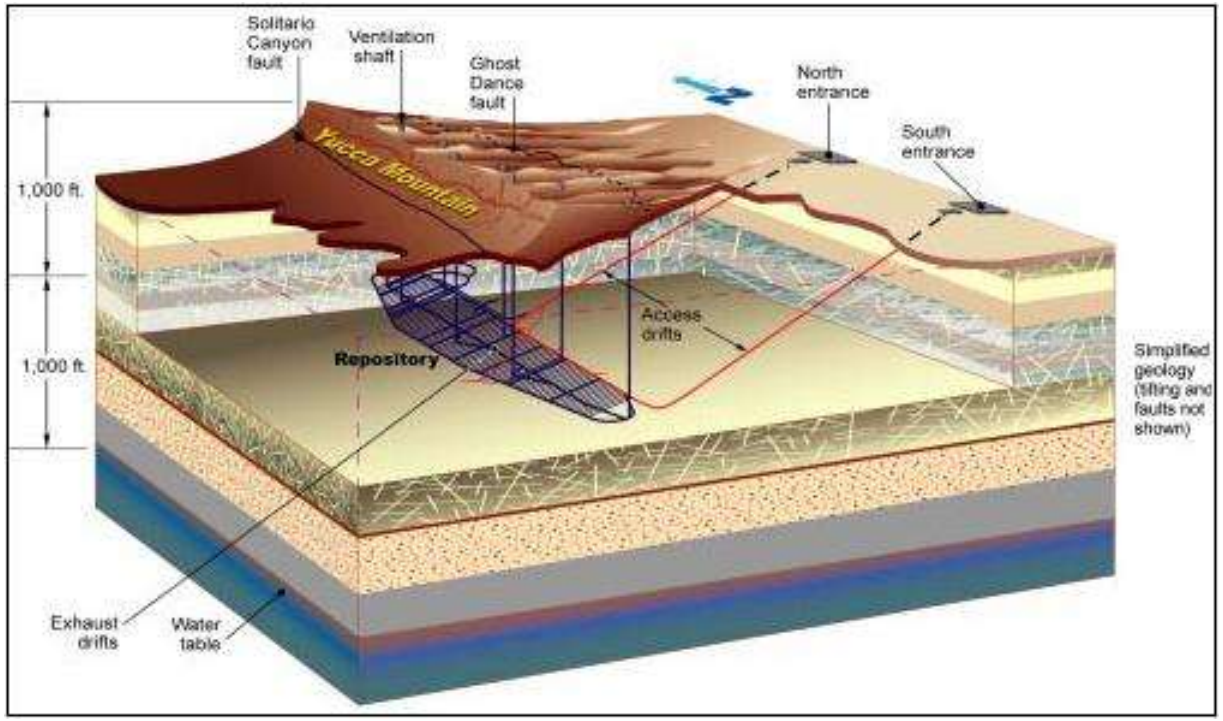
<<It is important, however, to recognize that the sensitivity of the parameter in the equation is what is being determined, not the sensitivity of the parameter in nature>>

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.





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.

Where does the rabbit come from?

A critique: “Monte Carlo doesn’t tell you anything that wasn’t already baked into the simulation. The distribution of the output comes from assumptions in the input (modulo bugs): a probability model for the parameters in the simulation. It comes from what you program the computer to do. Monte Carlo reveals the consequences of your assumptions about randomness. The rabbit goes into the hat when you build the probability model and write the software. The rabbit does not come out of the hat without having gone into the hat first.” (Philip B. Stark)

A question: How liberal is liberal enough?

Give a factor the widest plausible range and find it non-influential: you have learned something of value. Drop plausibility and let uncertainty explode: what did you learn?

Stark, P. B. (2022). Pay no attention to the model behind the curtain. Pure and Applied Geophysics, 179(11), 4121–4145.

The quality of an analysis ⟷ Its input

From: Uncertainty and Quality in Science for Policy by Silvio Funtowicz and Jerry Ravetz, Springer 1990.

Funtowicz & Ravetz's GIGO (Garbage In, Garbage Out) Science – or pseudo-science – "where uncertainties in inputs must be suppressed least outputs become indeterminate"

Leamer, Edward E. 1985. 'Sensitivity Analyses Would Help'. The American Economic Review 75 (3): 308–13.

Leamer's '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'.

NEVER vary all factors by 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

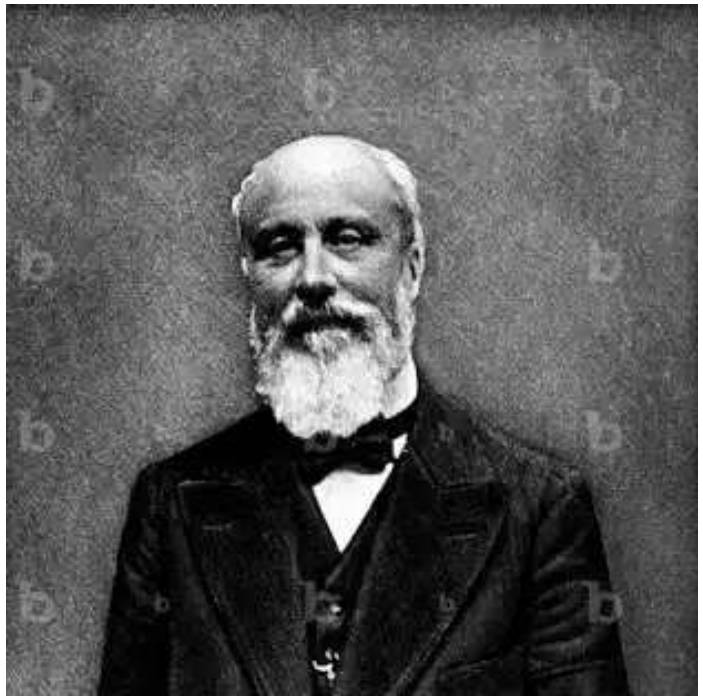
Brazzaville – Eighty-three thousand to 190 000 people in Africa could die of COVID-19 and 29 million to 44 million could get infected in the first year of the pandemic if containment measures fail, a new study by the World Health Organization (WHO) Regional Office for Africa finds. The research, which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

Speculative scenario in which ten uncertain input probabilities are varied 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

"… 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].

Part IV – Beyond the formal system: sensitivity auditing and models in competition

From sensitivity analysis to sensitivity auditing

Uncertainty and sensitivity are also, and perhaps eminently, political.

Sensitivity auditing espouses post-normal science and its NUSAP notation: a quantity together with an account of its quality.

N	U	S	A	P
Numeral	Unit	Spread	Assessment	Pedigree

Familiar: the number, its units, its range

Judgement of confidence; scored evaluation of the process

What NUSAP adds is not precision but candour.

Saltelli, A., Guimarães Pereira, Â., van der Sluijs, J. P., Funtowicz, S. (2013). What do I make of your latinorum? Sensitivity auditing of mathematical modelling. Int. J. Foresight and Innovation Policy, 9(2/3/4), 213–234. Funtowicz, S., Ravetz, J. R. (1990). Uncertainty and Quality in Science for Policy. Kluwer.

1. Check against rhetorical use of mathematical modelling
2. Adopt an ‘assumption hunting’ attitude
3. Detect garbage in, garbage out (GIGO)
4. Find sensitive assumptions before they find you
5. Aim for transparency
6. Do the right sums
7. Focus the analysis on the key question answered by the model, exploring the entire space of the assumptions holistically

Verbatim from

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

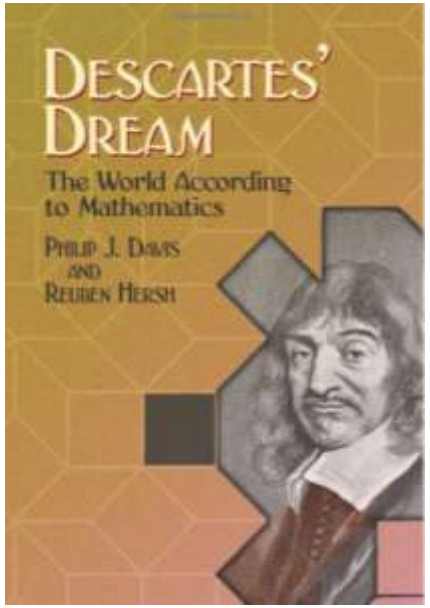
What do I make of your latinorum? Sensitivity auditing of mathematical modelling

Genuine or bogus? A validation checklist

"... [there are] certain criteria by which the mathematician judges whether an "application" of mathematics is genuine or bogus."

- Does the depth of the real-world problem justify the complexity of the mathematical model?
- Are any genuine mathematical reasonings or non-trivial calculations carried out that require the resources of the mathematical model being proposed?
- Are the coefficients or parameters in the equations capable of being determined in a meaningful and reasonably accurate way?
- Are the conclusions capable of being tested against real-world data?
- Do any non-obvious practical conclusions follow from the analysis?

Eminently sensible — if only one could identify the subject implementing them.



Davis, P. J., Hersh, R. (1986). Descartes' Dream: The World According to Mathematics. Penguin Books, London.

Quantitative story telling: testing the story a model tells

Feasibility

Compatibility with external constraints

e.g. biophysical limits of a resource

Viability

Compatibility with internal constraints

e.g. existing regulatory regimes or laws

Desirability

Compatibility with normative values adopted in the given society

e.g. would citizens willingly submit to the policy?

If the story told by a model is a policy prescription, its robustness is tested against all three — ideally in a participatory setting.

Saltelli, A., Giampietro, M. (2017). What is wrong with evidence based policy, and how can it be improved? Futures, 91, 62–71.

Models in competition: comparative assessment as an institution

Countermodelling

Different modelling teams mobilised by different actors (Zayandeh-Rood basin, Iran)

Adversarial collaboration

Proposed in medicine as a complement to consensus-building

Red teams

Their absence from the UK COVID decision room was lamented by statisticians

Hot tubbing

Concurrent expert evidence in Australian courts: experts in the witness box together

"These instances are instructive precisely because they are exceptional. The default condition remains one where models operate without meaningful challenge."

Nabavi, E. (2025). Water Alternatives, 18(2). Kent, D. M., Ioannidis, J. P. A. (2026). JAMA, doi:10.1001/jama.2025.23911. Wood, S. N. et al. (2025). JRSS A, 189(1), qnaf092. McClellan, P. (2010). Method with experts – concurrent evidence. Journal of Court Innovation, 3(1), 259–268.

None is sufficient on its own

Sensitivity analysis
can reveal what drives a model's output but cannot question the model's purpose.

Sensitivity auditing
can question the purpose but depends on the existence of critical communities willing to do the questioning.

Statactivism and countermodelling
can mobilise such communities but face the asymmetries of power, resources, and access that characterise the present political economy of science.

Equations of Power, §5.9.

Fight a number with a number: statactivism

Statistical judo

Gaming the metrics, using Goodhart's law strategically

Denouncing unfair indicators

e.g. the middle-class bias of the French consumer price index

Brand-new measures

A new index that better defends the purchasing power of the poor

Making the invisible visible

Areas of invisibility or exclusion that escape statistical offices

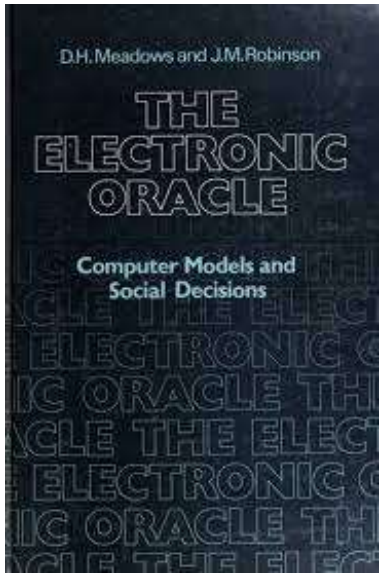
Bruno, I., Didier, E., Prévieux, J. (2014). Statactivisme. Comment lutter avec des nombres. Zones, La Découverte, Paris.

Hall was right: the main reason to build a model may well be to play a sensitivity analysis on it

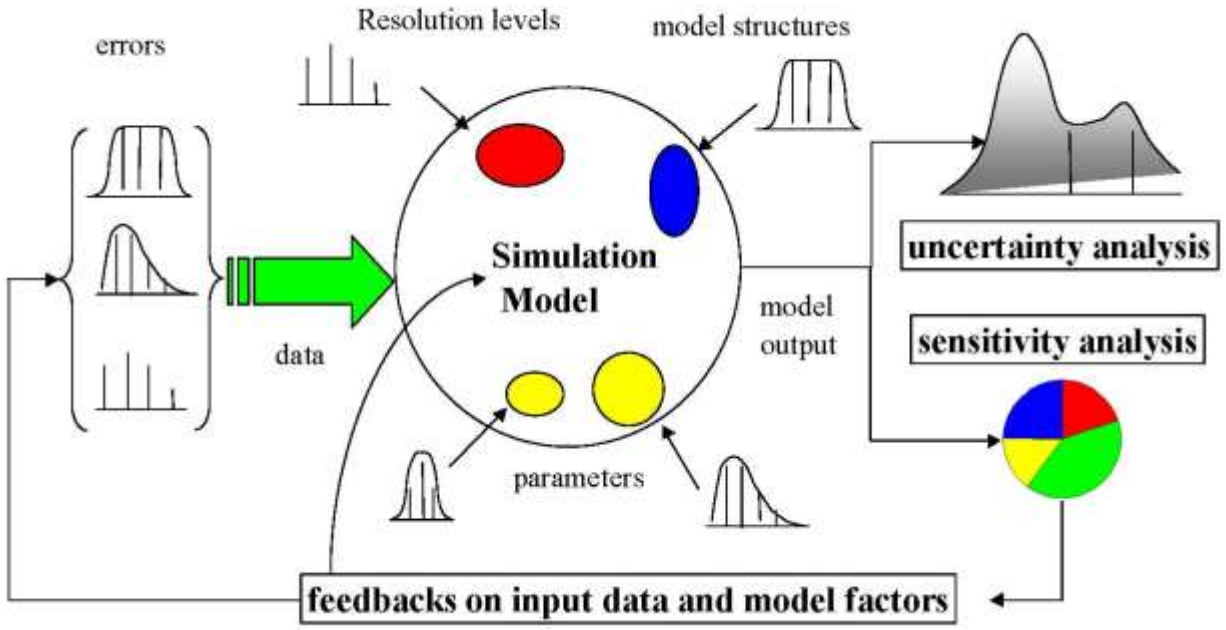
The question is whether this conflicts with the political economy of mathematical modelling

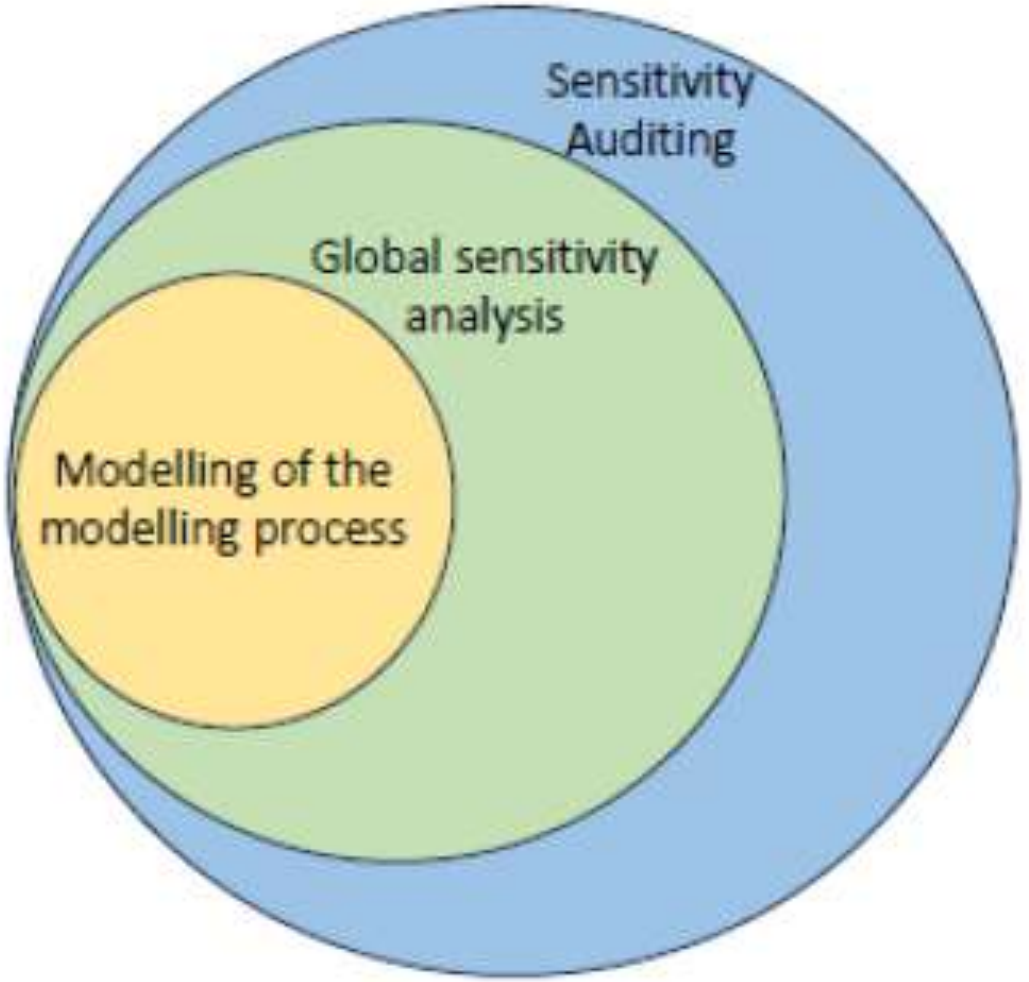
Decades of scarce traction for opening the black box of models: a debt we now pay with AI

Dona Meadows lamented the scarce traction already in 1985



The End


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

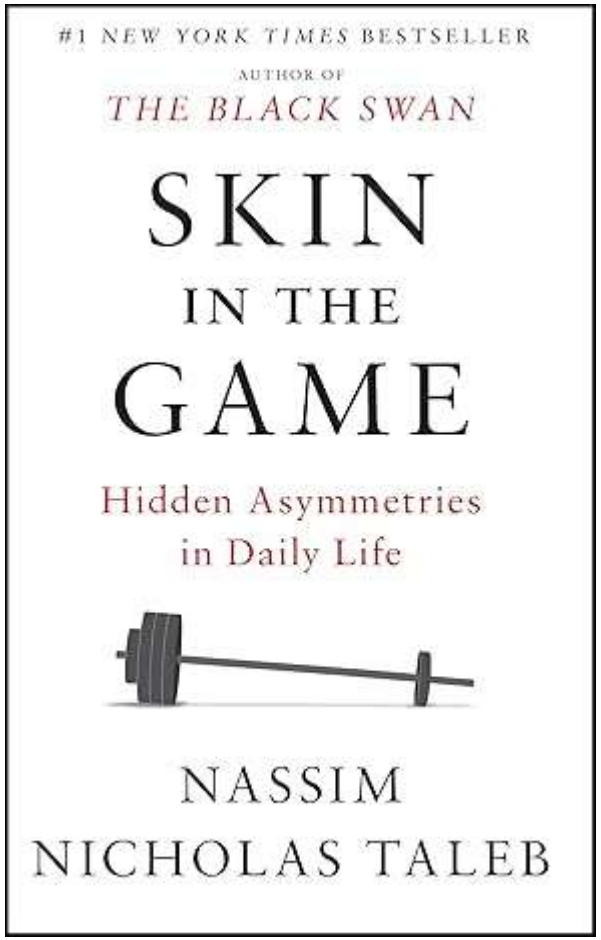

Sensitivity Auditing
Global sensitivity analysis
Modelling of the modelling process

Backup slides

What would help

A model should not be used for risk analysis unless:

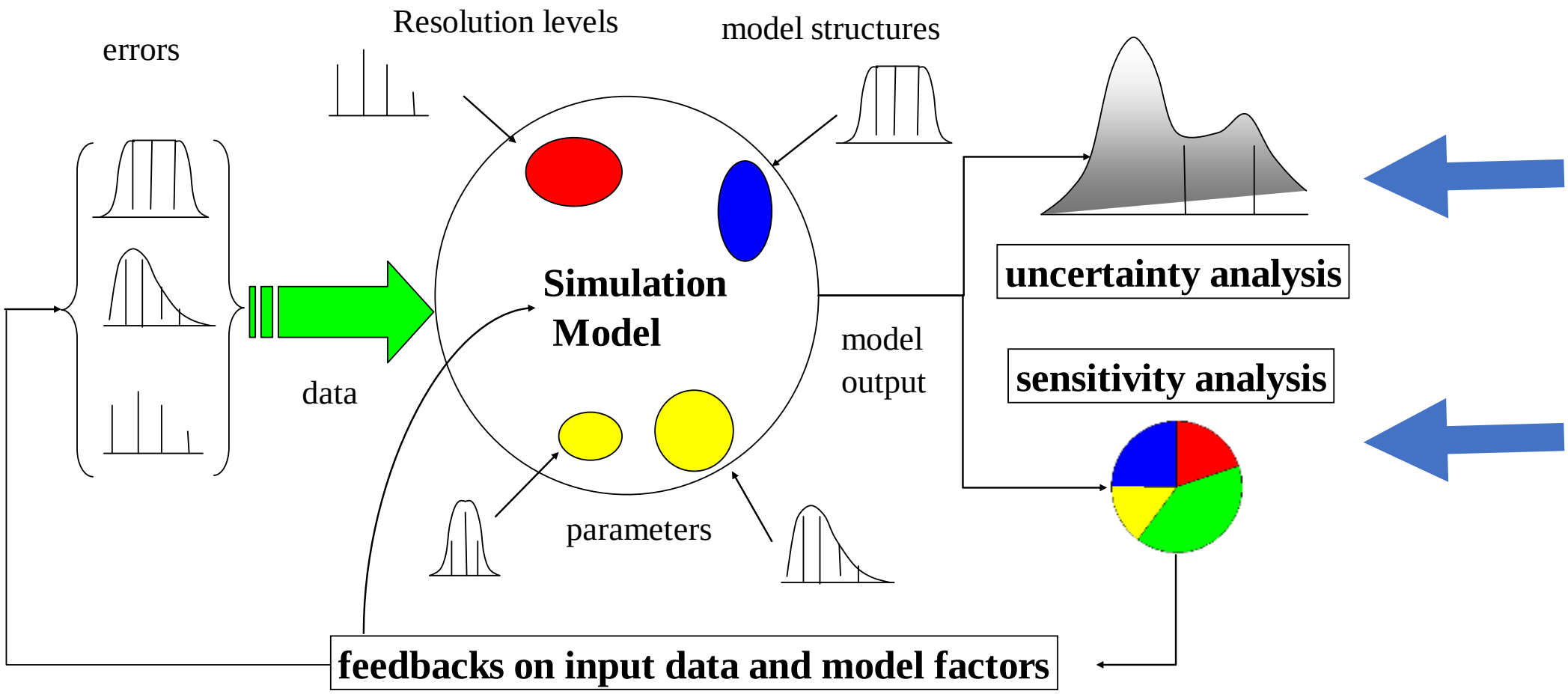
- Its **sensitivity** to key assumptions has been explored — Global sensitivity analysis
- Its structural and **framing** choices have been made explicit and tested — Modelling of the modelling process
- **Affected parties** have had the opportunity to identify which assumptions matter to them — Sensitivity auditing



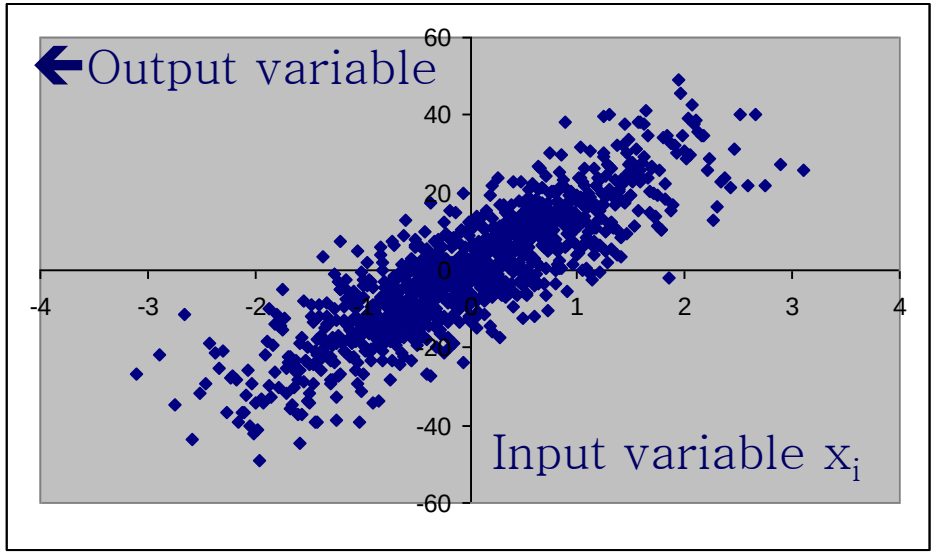
Ignorantia and *via negativa*

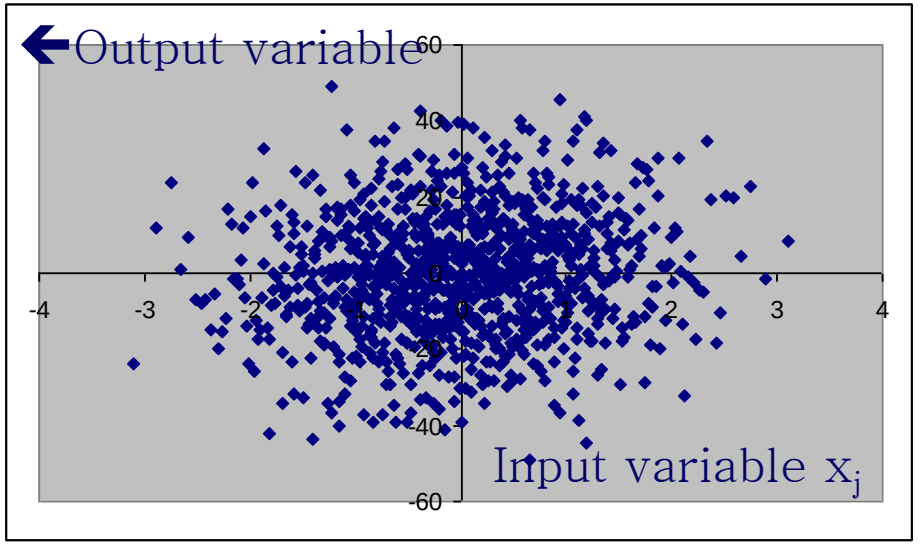
"Via negativa: the principle that we know what is wrong with more clarity than what is right, and that knowledge grows by subtraction. Also, it is easier to know that something is wrong than to find the fix. Actions that remove are more robust than those that add because addition may have unseen, complicated feedback loops"

Stefano Tarantola, circa 2008



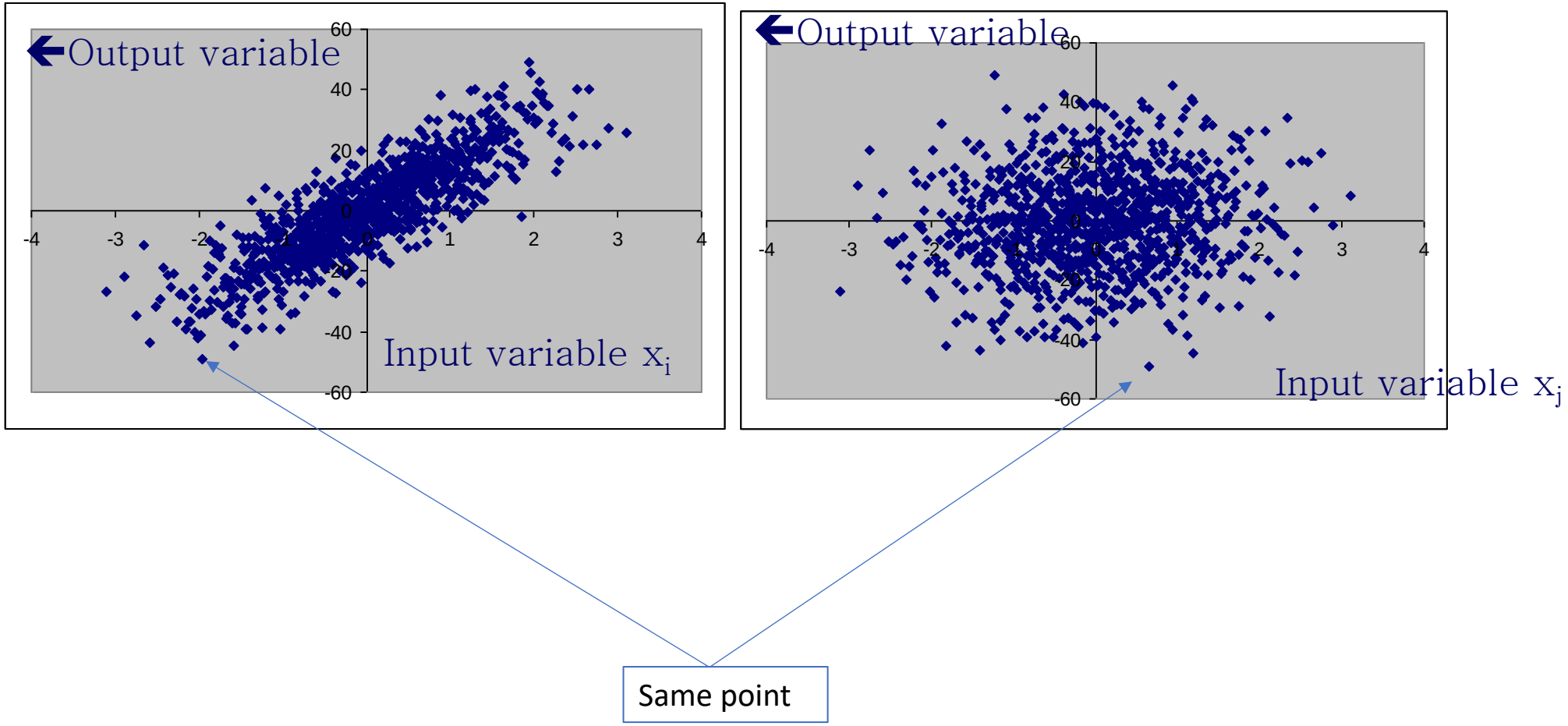
Variance-based methods



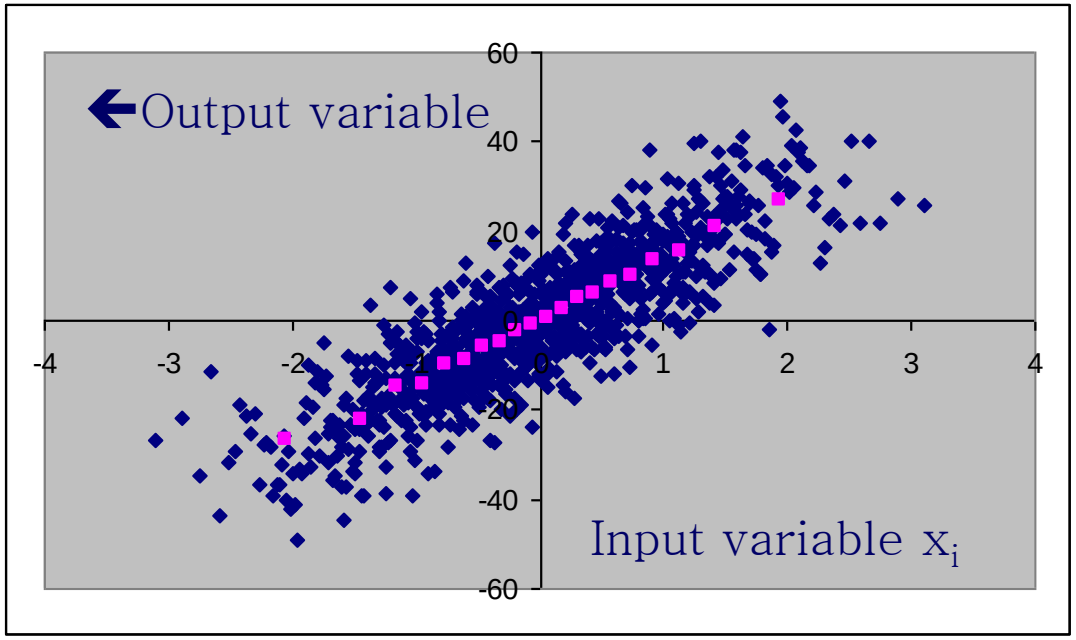


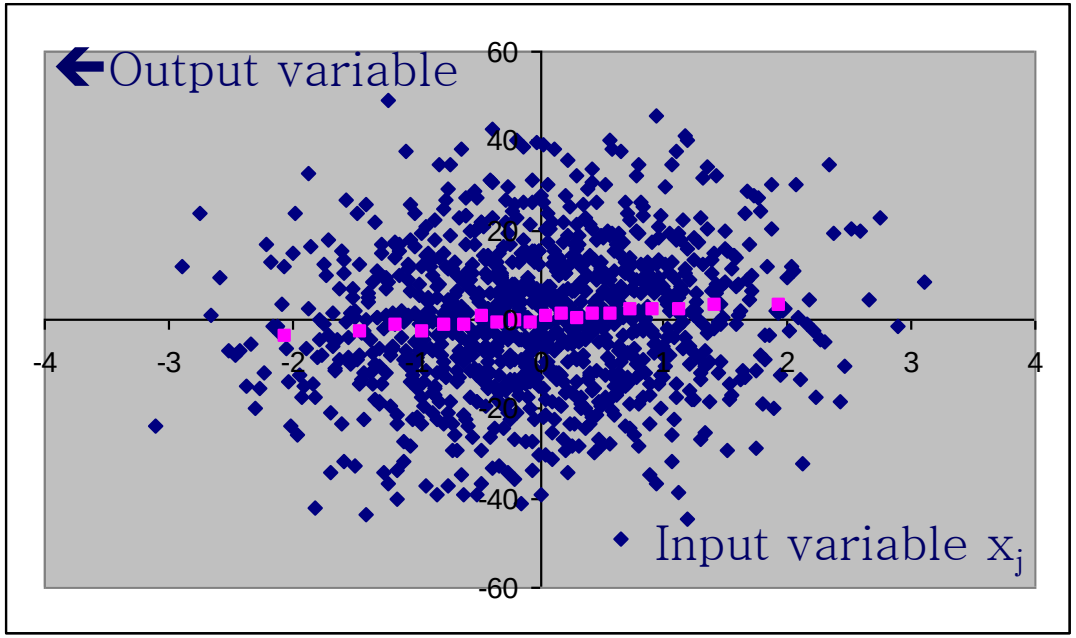
Plotting the output as a function of two different input factors:

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





~1,000 blue points

Divide them in 20 bins of ~ 50 points

Compute the bin's average (pink dots)

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

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

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)

In fact, interaction 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’ term for all effects?

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

Before moving to the total index, let us finish with the first-order one

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

Known as:

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

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

We need to understand what the numerator is in plain English

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

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

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

Because:

= then this is the reduction in variance if X_i could be fixed

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

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

(By inspection) this is 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

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

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:

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

If this = main effect on non$-X_i$

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

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

… then 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.

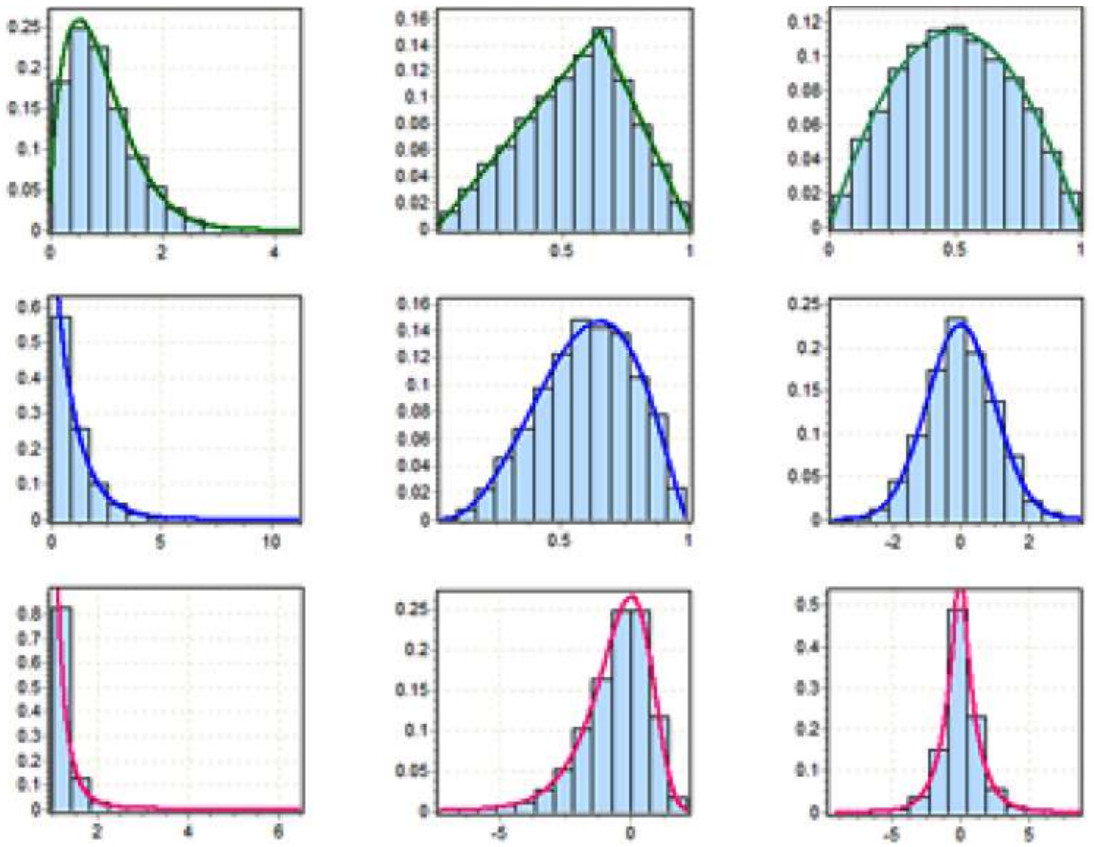
What is needed is a design (as in statistical DOE theory)

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

Saltelli A., 2002, Making best use of model evaluations to compute sensitivity indices, Computer Physics Communications, 145, 280-297.

How about Shapley?

$$S_i \leq \mathrm{Sh}_i \leq T_i,$$

Shapley sits between S_i and T_i for uncorrelated input

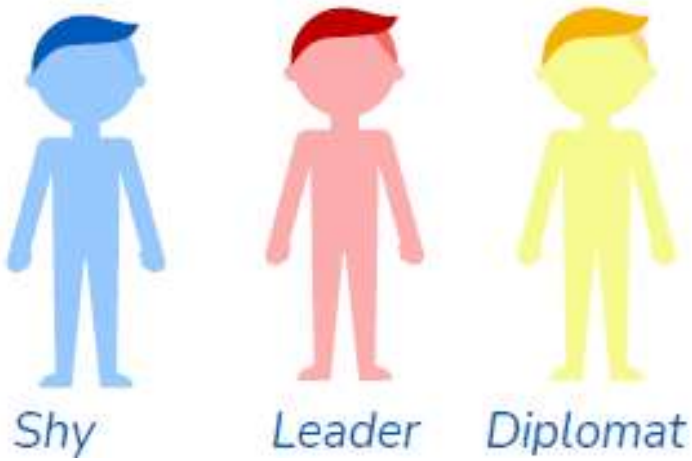
Why?

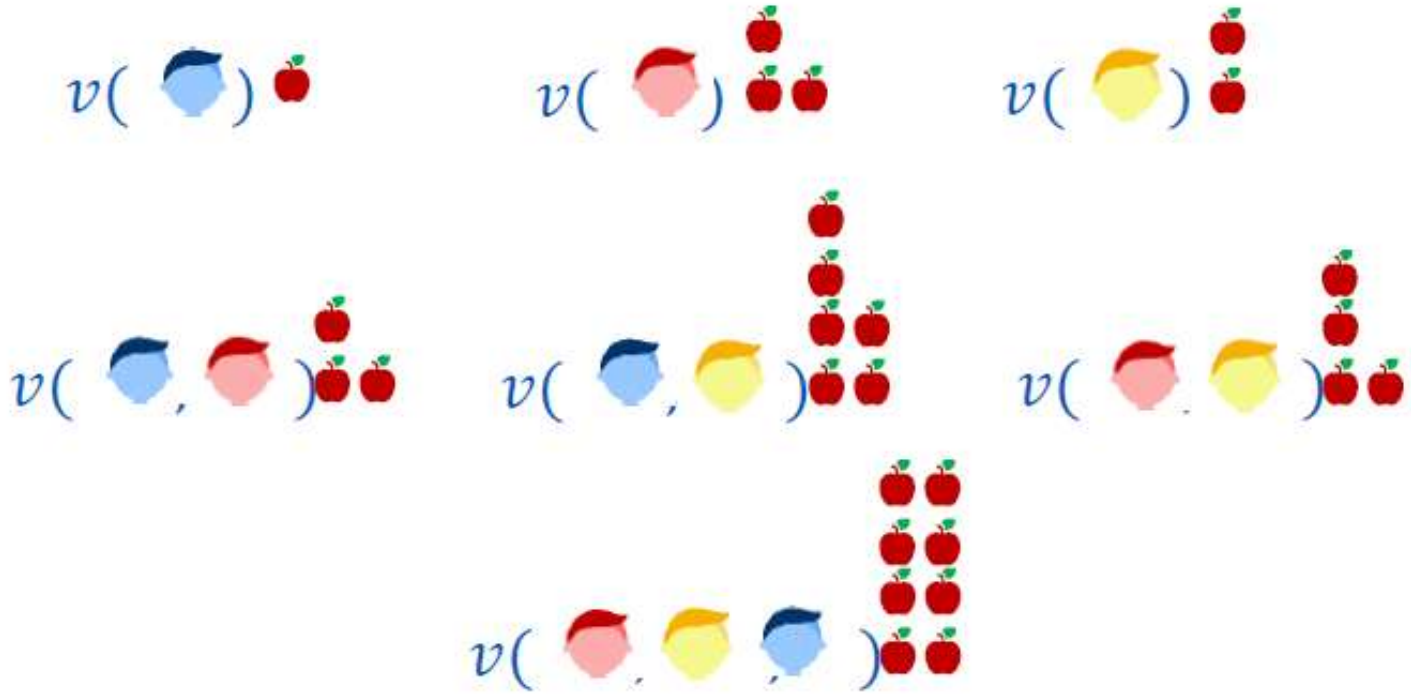
Owen, A. B. (2014). "Sobol' indices and Shapley value."
SIAM/ASA Journal on Uncertainty Quantification, 2(1), 245–251.

$$S_i \leq \mathrm{Sh}_i \leq T_i,$$

- Si gets none of the interactions
- Ti gets all of the interactions belonging to X_i
- Shapley value splits each interaction term *equally* among the variables
- Shapley effects sum exactly to the total variance (or to 1 if normalised)

$T_i - S_i$ is the interaction that Shapley redistributes.





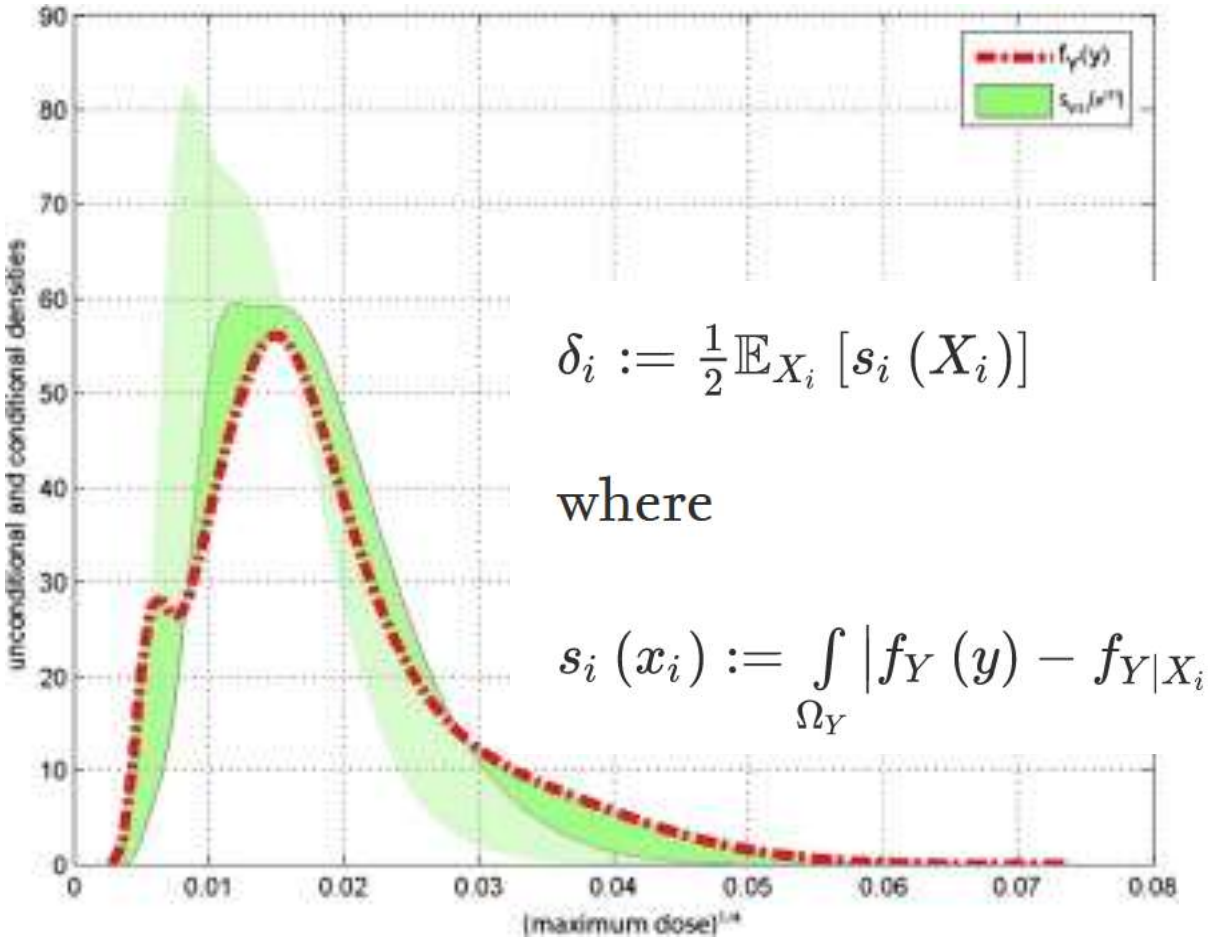
Images courtesy from Bertrand Iooss, Electricité de France (EDF)

A link between $I(X^j)$, a measure of permutation importance in ML, and the total sensitivity index

$$ST_j = \frac{I(X^j)}{2Var(Y)}$$

Antoniadis, Anestis, Sophie Lambert-Lacroix, and Jean-Michel Poggi. 2021. 'Random Forests for Global Sensitivity Analysis: A Selective Review'. Reliability Engineering & System Safety 206 (February): 107312. https://doi.org/10.1016/j.ress.2020.107312.

… but there are other methods that can be used for different settings, e.g. moment-independent:



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

where

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

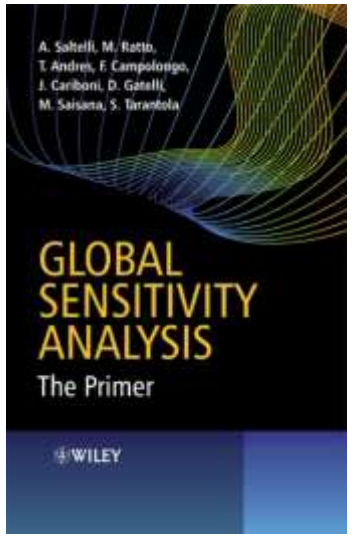
ELSEVIER

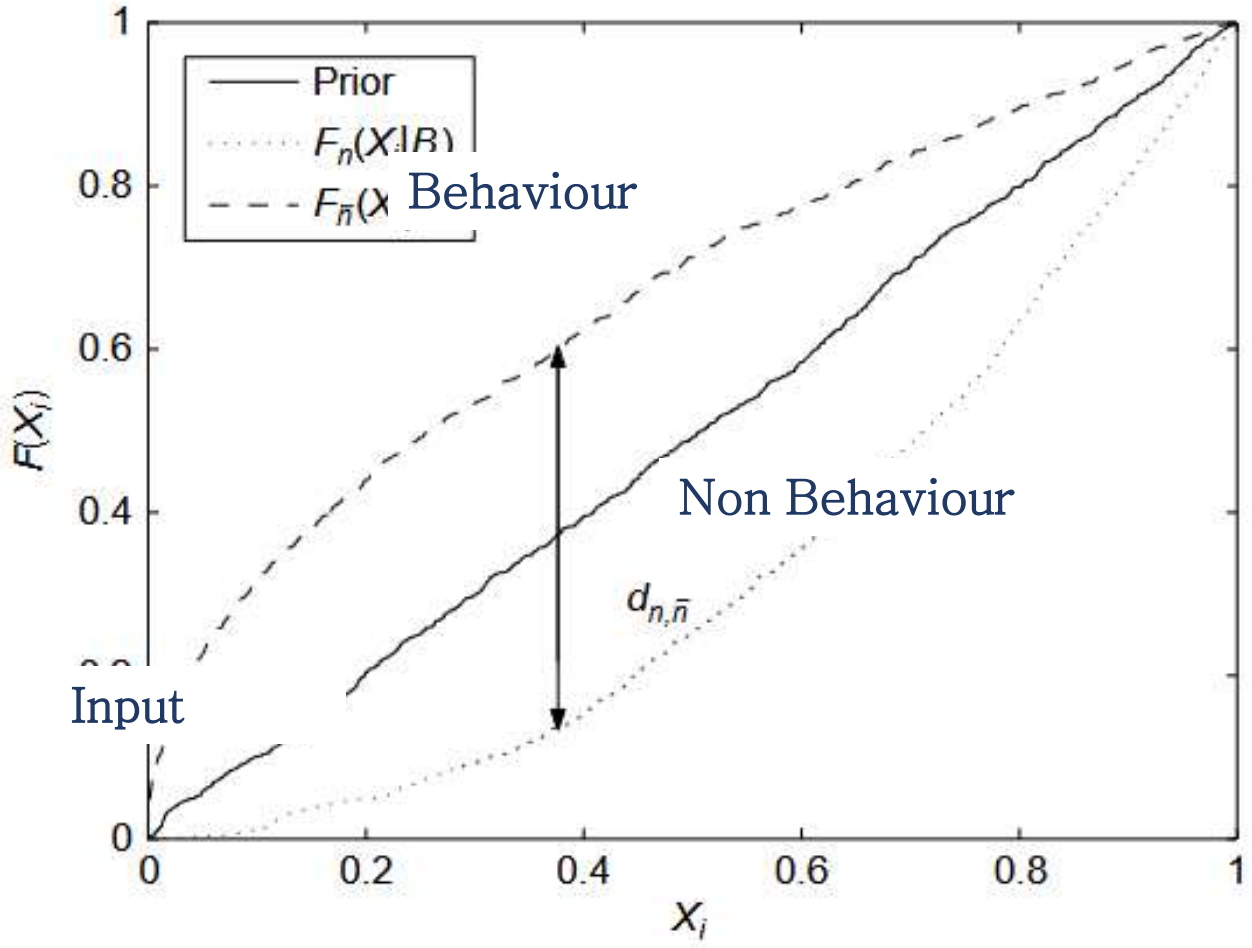
Environmental Modelling & Software

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]

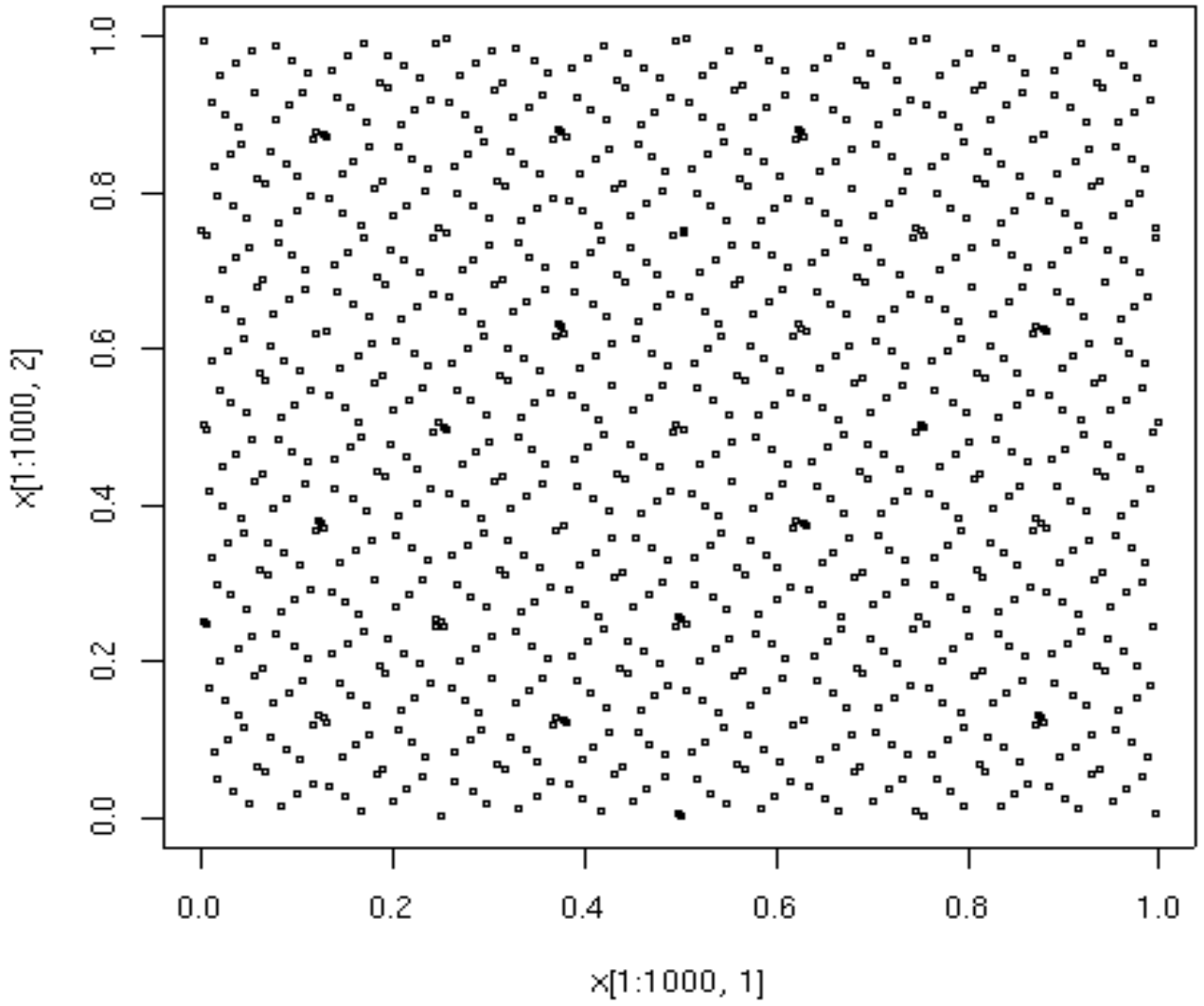




Regionalised sensitivity analysis; the input distribution is split according to Y being **Behaviour** or **Non-Behaviour**

Exploring the multidimensional cube

(Go quasirandom)



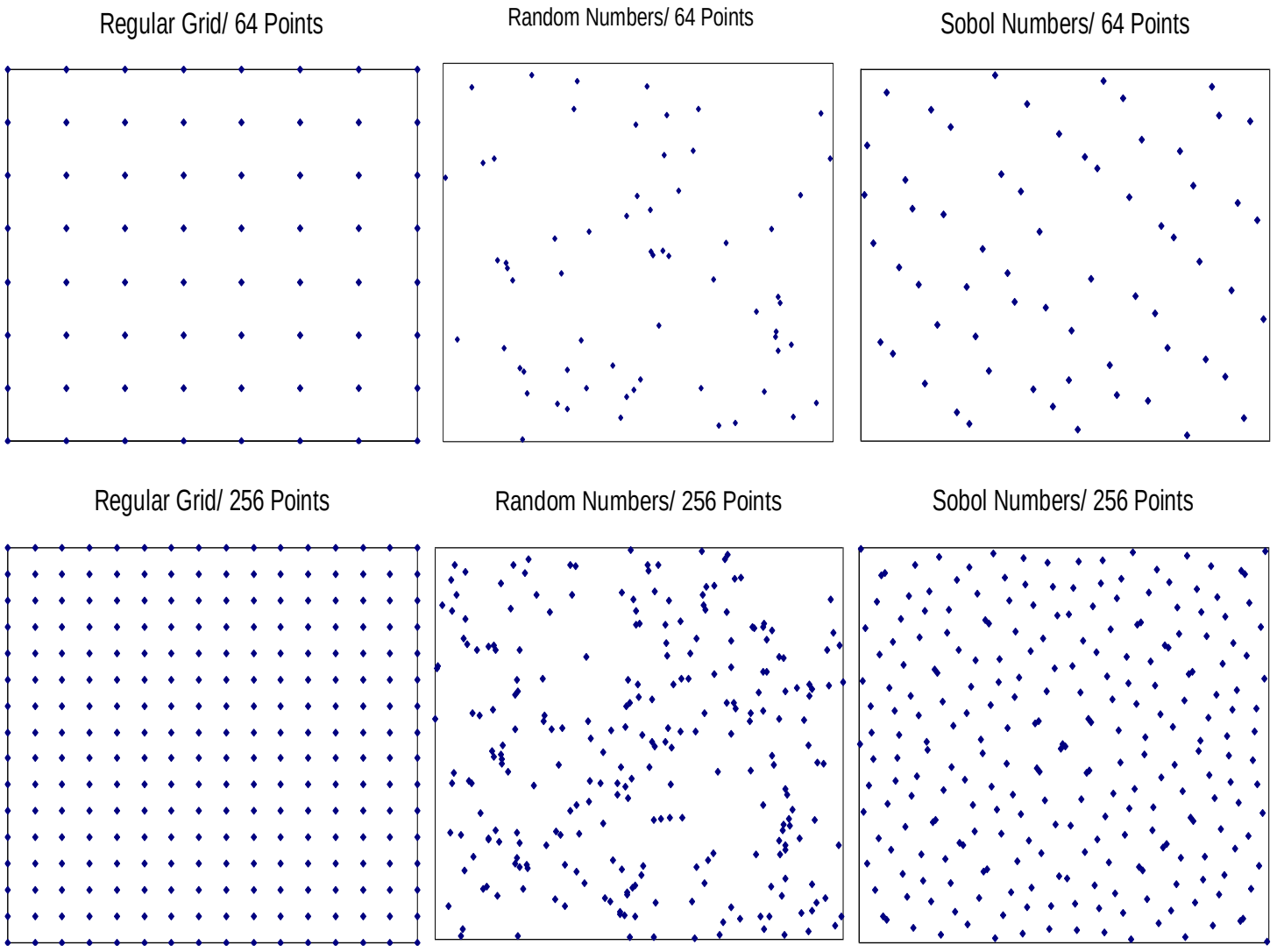
An LP_τ sequence

Ilya M Sobol'
1926-2025

Sensitivity buffs use quasirandom sequences, preferentially Sobol' LP-TAU sequences

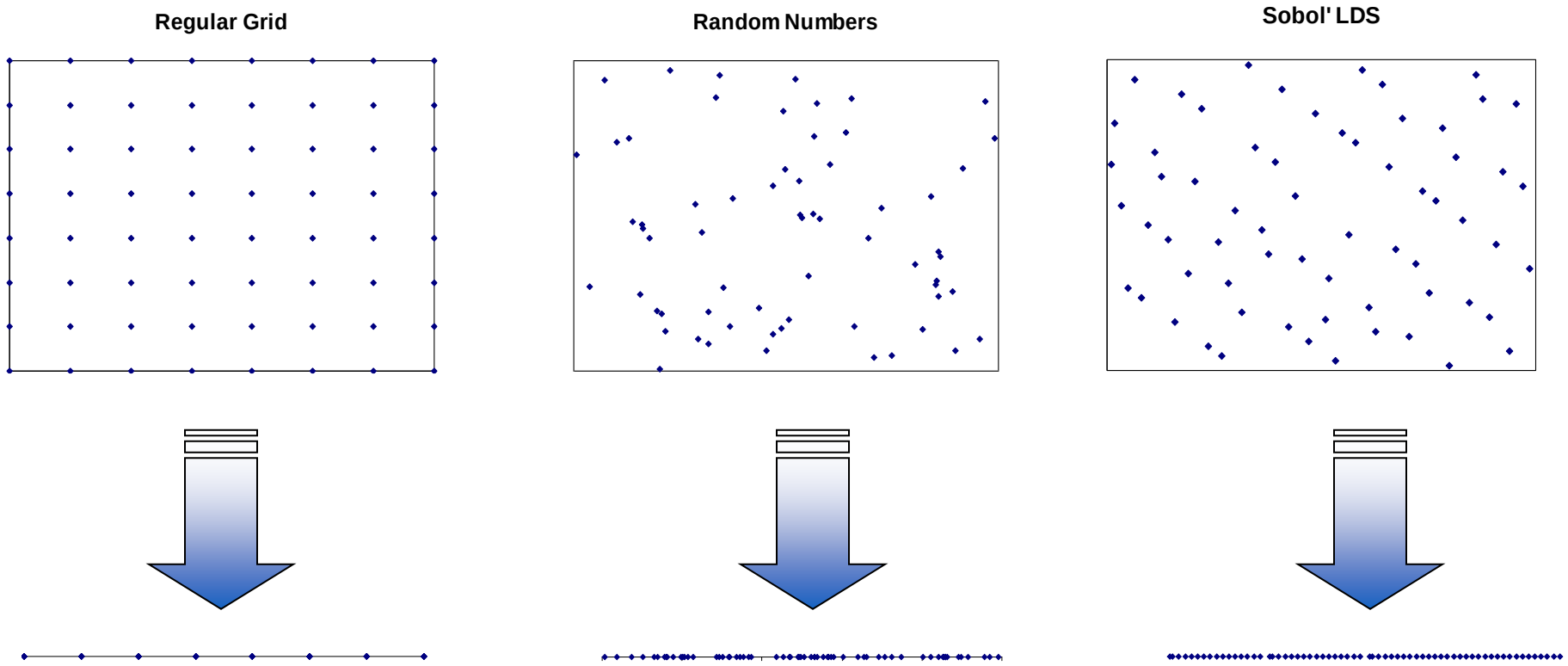
Sobol' Sequences vrs Random numbers and regular grid

Slide courtesy of Sergei Kucherenko, Imperial College London



Slide courtesy of Sergei Kucherenko, Imperial College >London

Projections of Different 2D sequences to 1D

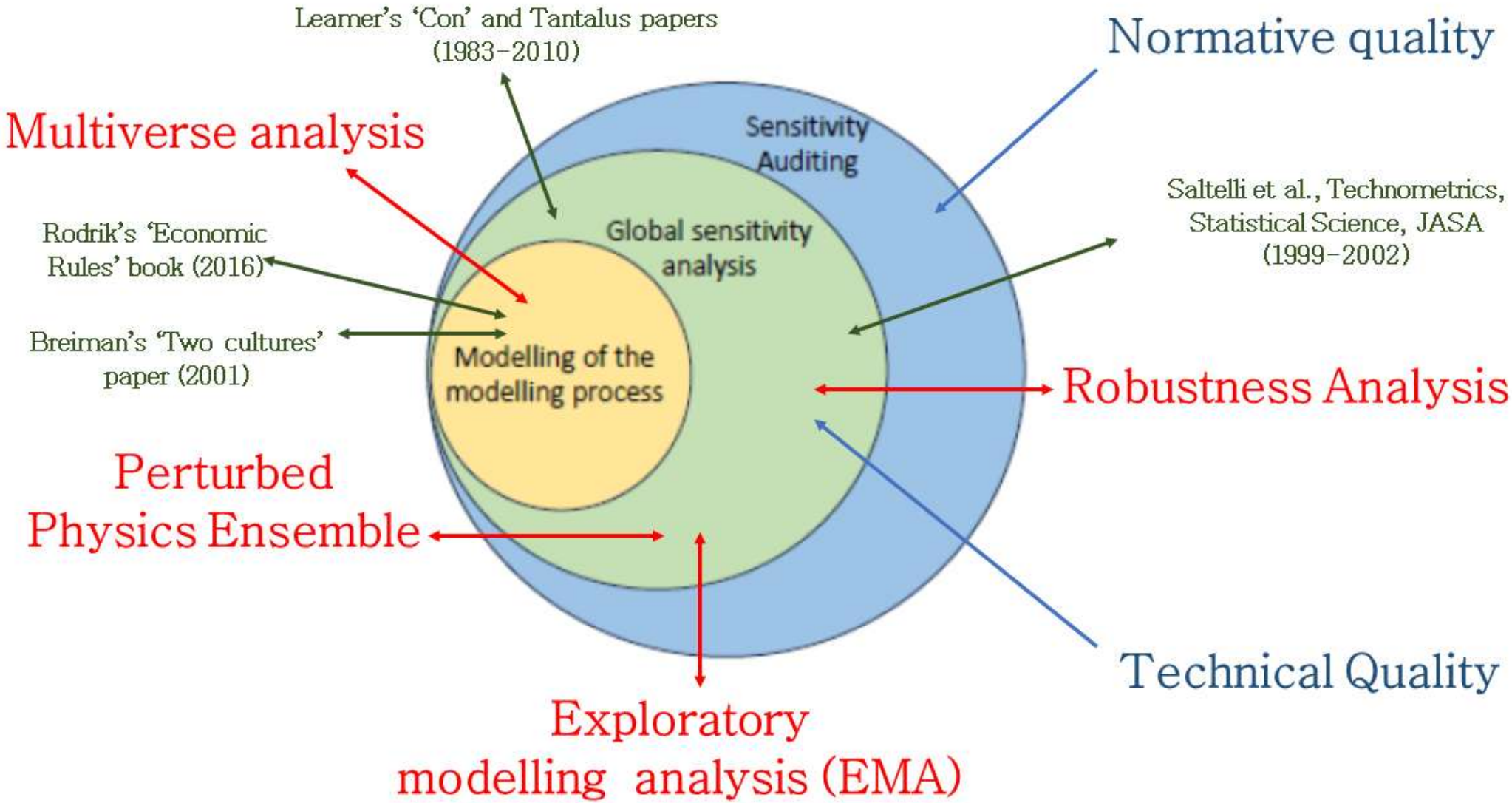


Shadow effect
(64 -> 8 points)

Clustering

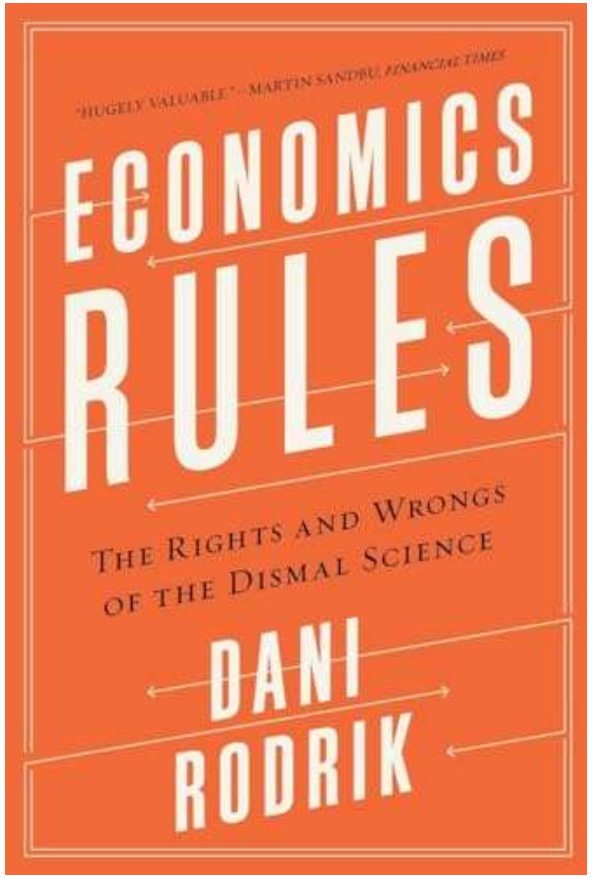
Uniform distribution

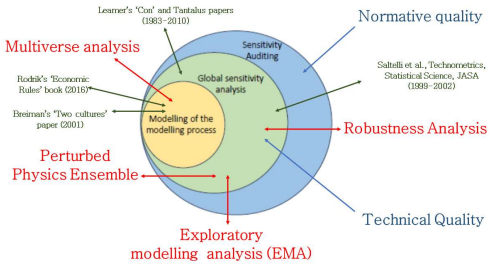
Part II – Who does GSA?


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)

Rodrik's rules about models (and GSA)

- Economics is a collection of models; cherish their diversity.
- It's a model, not the model.
- Make your model simple enough to isolate specific causes and how they work, but not so simple that it leaves out key interactions among causes.
- Unrealistic assumptions are OK; unrealistic critical assumptions are not OK



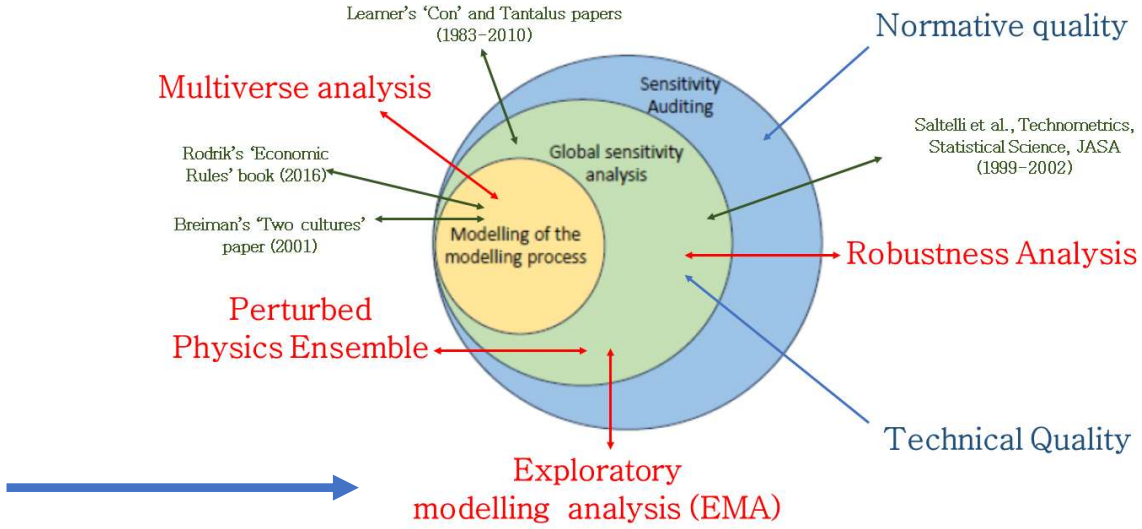


→ When modelling, move horizontally rather than vertically

Exploratory modelling analysis (RAND corporation)

Explore scenarios and simulate

How wide is wide enough?



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.

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

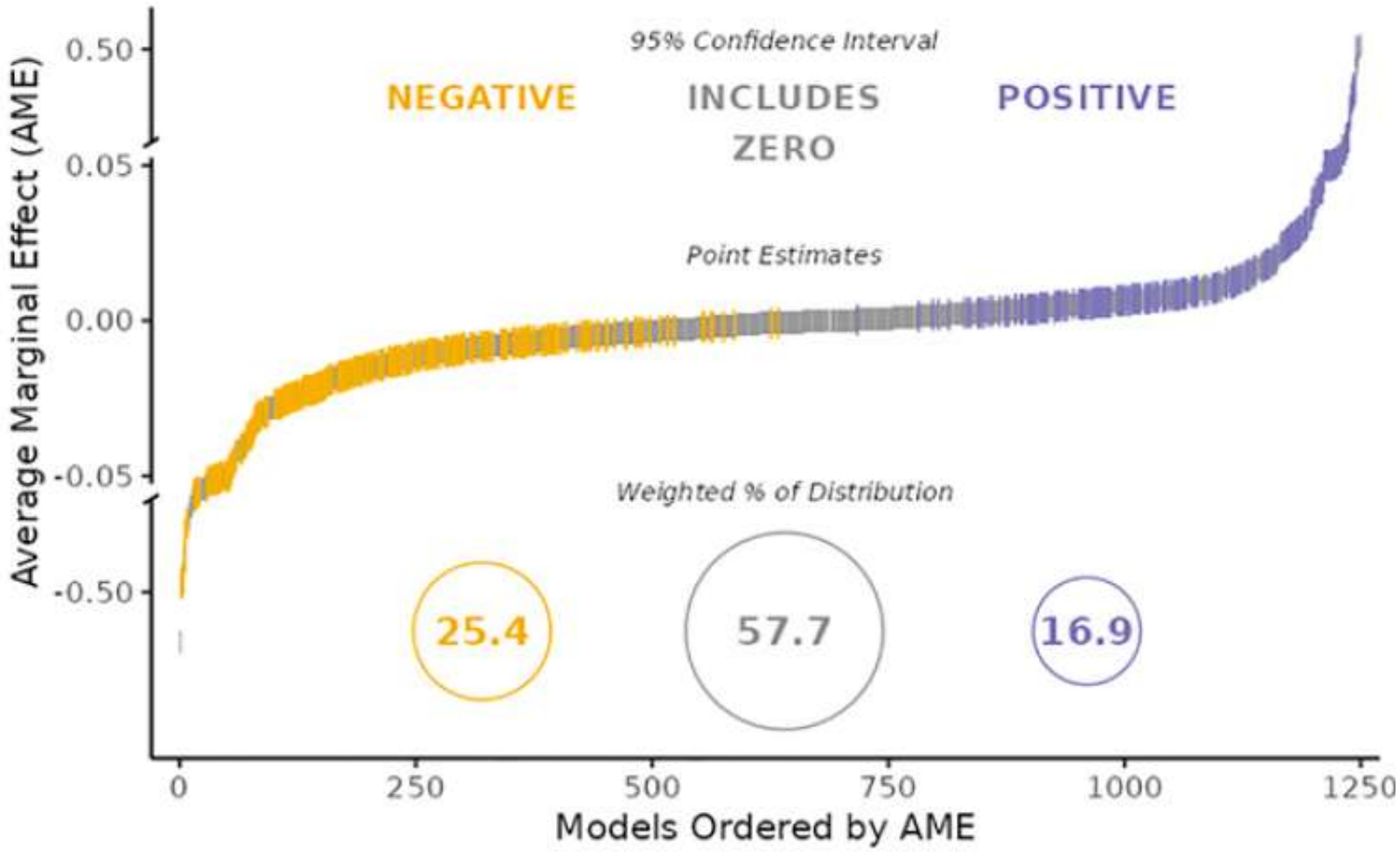
"fairness can be defined using a special framework based on Global Sensitivity Analysis …"

More reading about explainable ML: combining counterfactual and Shapley values

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”

Sensitivity analysis in the garden



"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

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

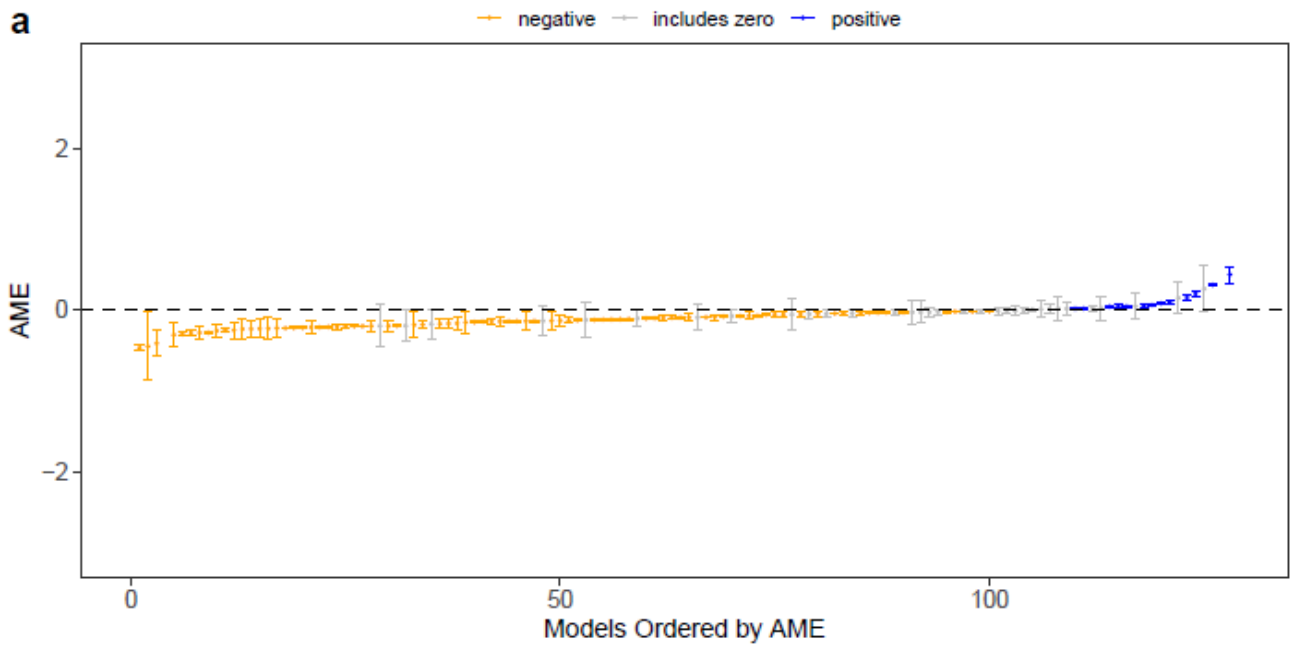
Table 1. Descriptions of variables included in the analysis.

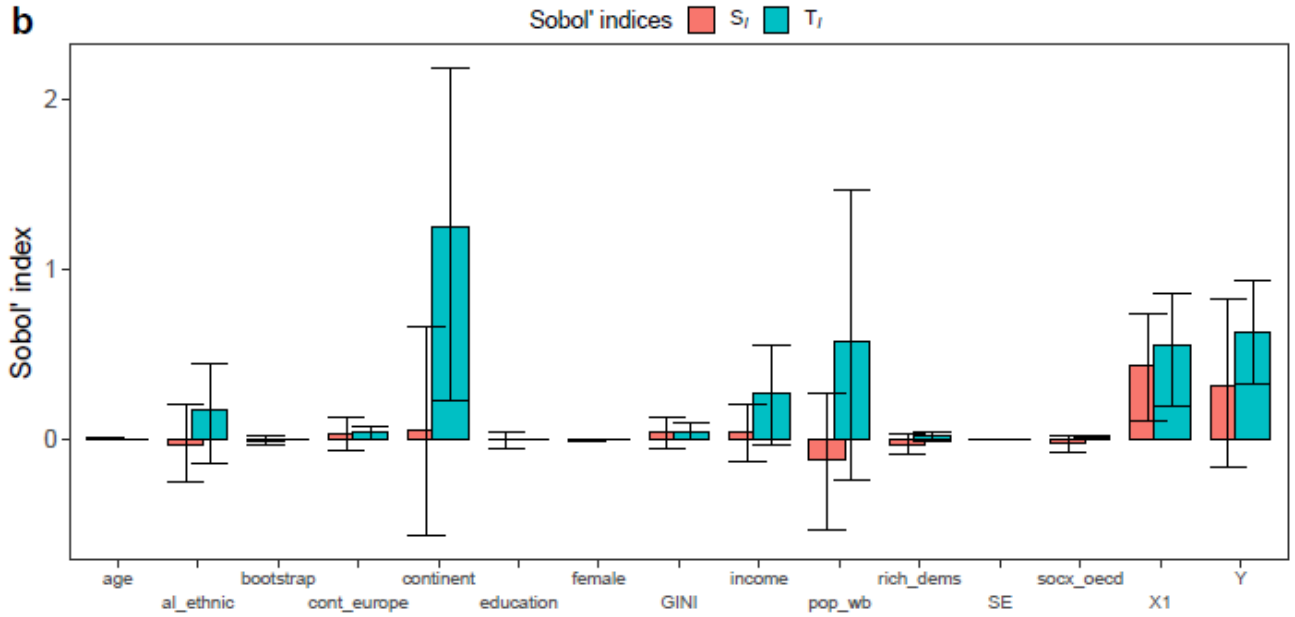
Leamer Reloaded: A Long-Neglected Suggestion Is Rediscovered To Apply Global Sensitivity Analysis To A Quantification Before Publishing It

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

OSF

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





Our uncertainty analysis broadly reproduces the results of the multi-analyst experiment

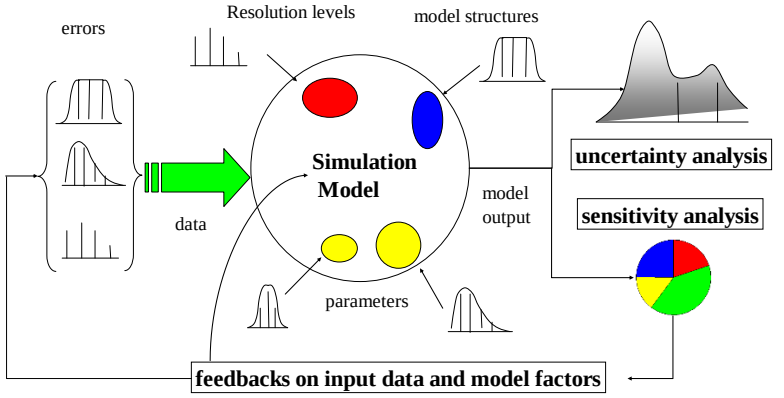
Our sensitivity analysis explains where the missing variance went: **interactions**

Part III GSA’s dos and don'ts

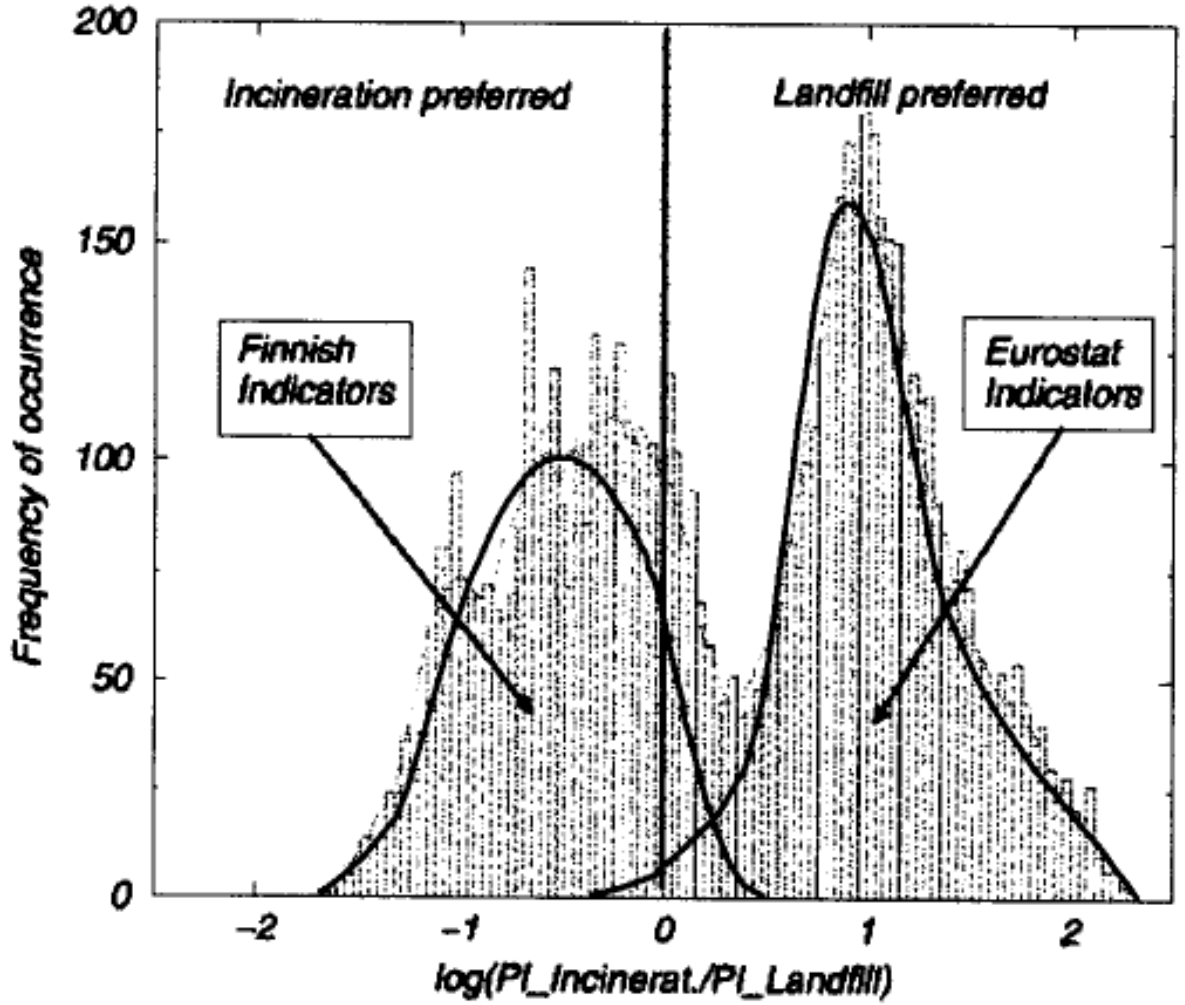
➔ Don't sample just parameters and boundary conditions: explore the space of assumptions thoroughly

One can sample more than just factors:

- modelling assumptions,
- alternative data sets,
- resolution levels,
- policies



Look at policy options versus existing uncertainties: e.g., test whether the choice of the model determines the choice of the policy



November 2000

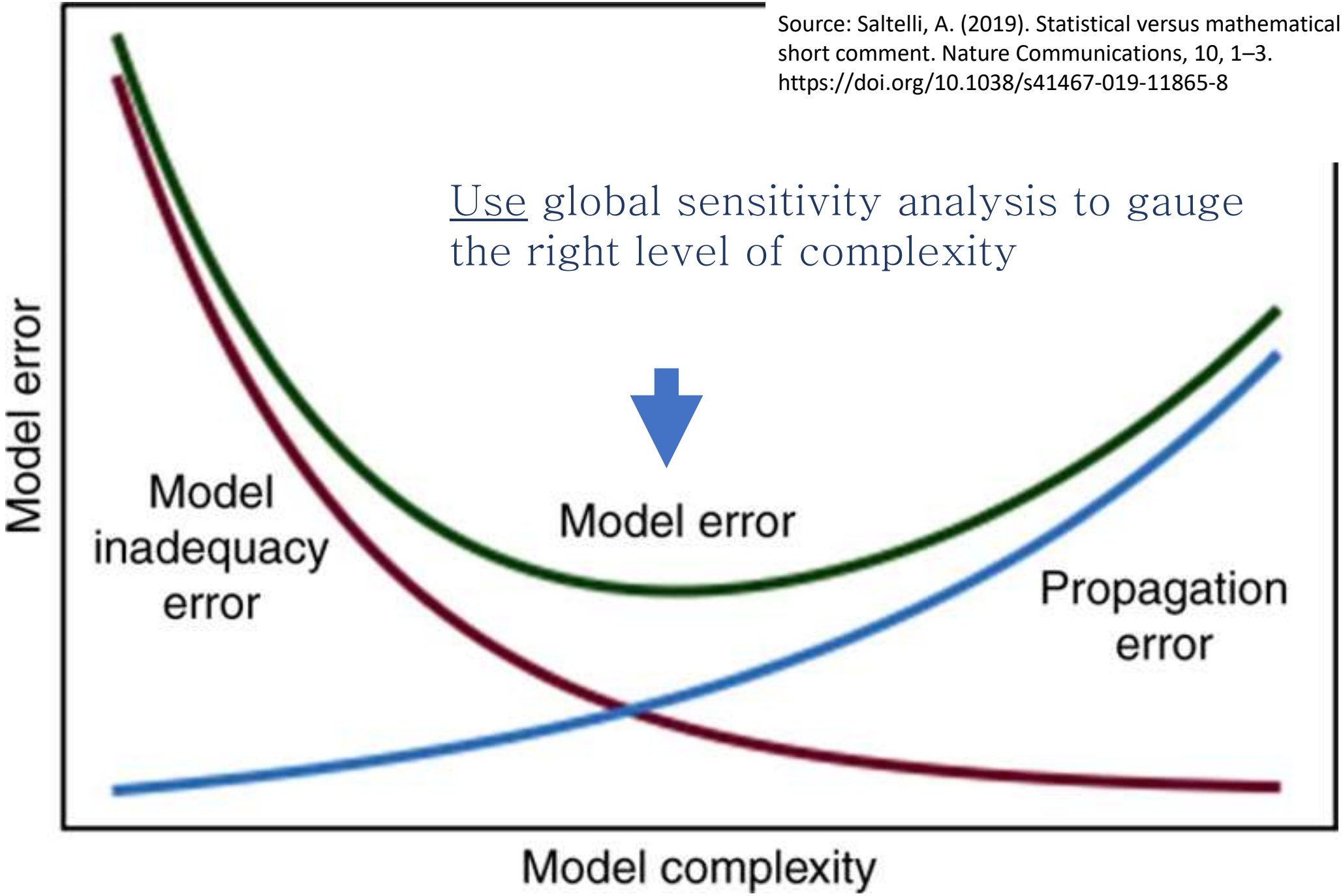
Sensitivity Anaysis as an Ingredient of Modeling

F. Campolongo, A. Saltelli, S. Tarantola

BARCELONA SCHOOL OF MANAGEMENT

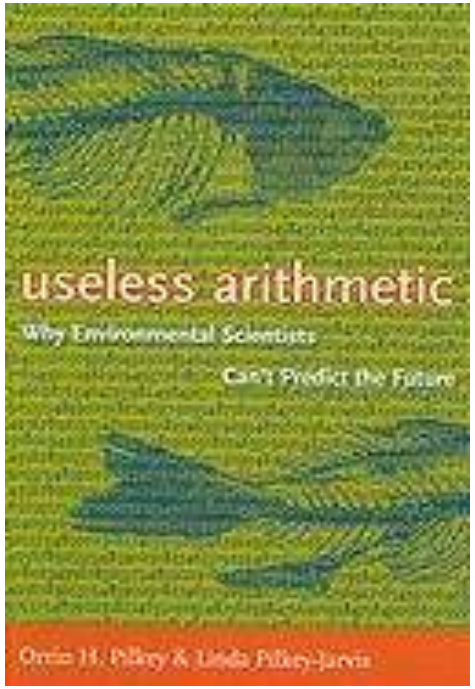
Source: Saltelli, A. (2019). Statistical versus mathematical modelling: A short comment. Nature Communications, 10, 1–3. https://doi.org/10.1038/s41467-019-11865-8

Use global sensitivity analysis to gauge the right level of complexity



Don’t confuse the map with the territory

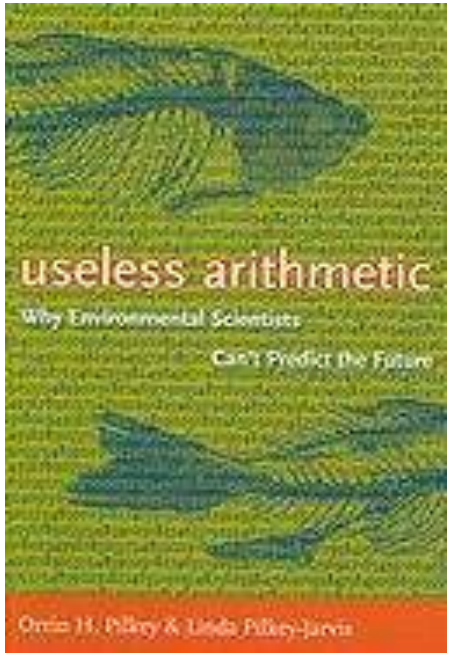
If you do, sensitivity analysis will not save you

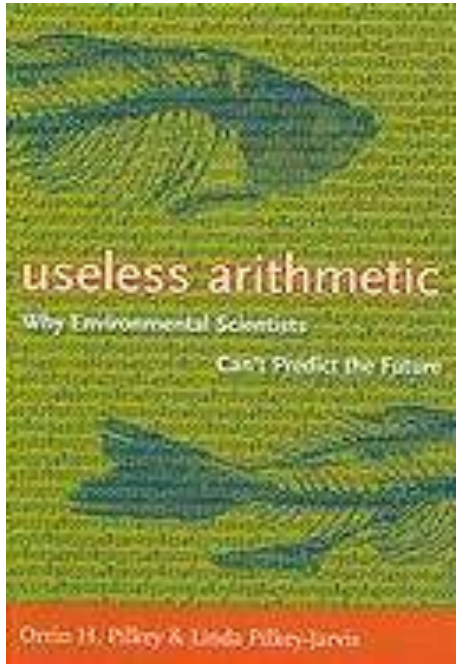


<<…If the model is wrong or if it is a poor representation of reality, determining the sensitivity of an individual parameter in the model is a meaningless pursuit>>

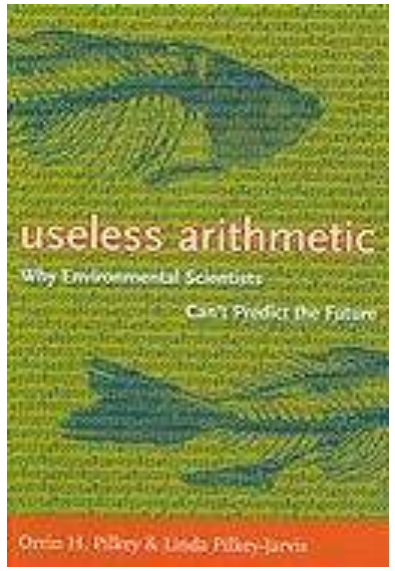
One of the examples discussed concerns the Yucca Mountain repository for radioactive waste. TSPA model (for total system performance assessment) for safety analysis.

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



"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.

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!

A model routinely used to produce point estimates may become non-conservative when the uncertainty is plugged in

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