

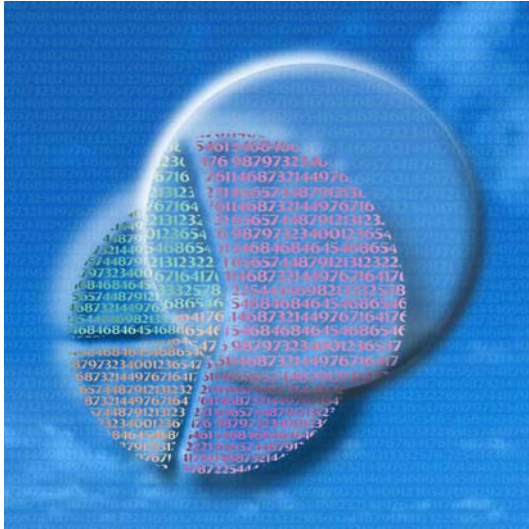
# Sensitivity auditing. Gauging model quality in relation to model use

Andrea Saltelli, European  
Commission, Joint Research  
Centre

Sec Gen advanced seminar  
series, June 10<sup>th</sup>, 2011

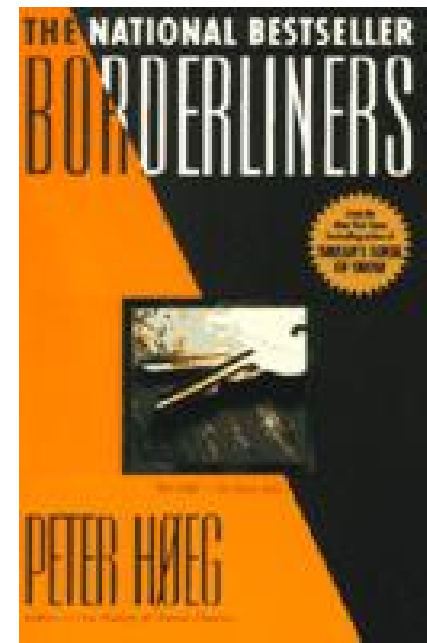
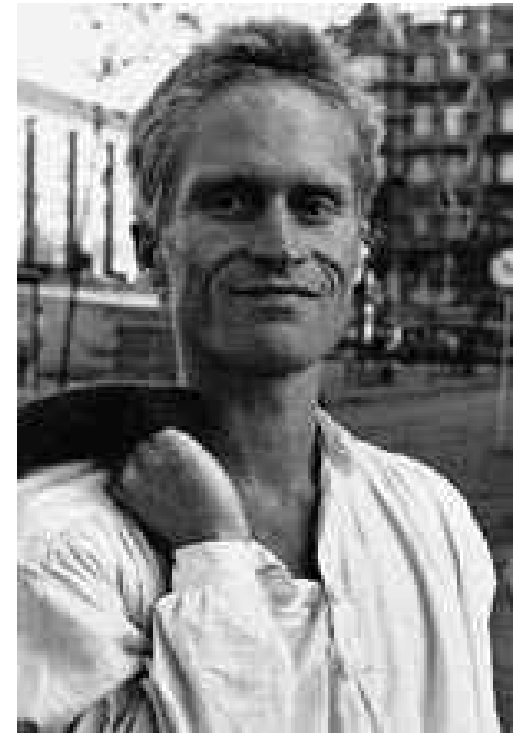
Brussels – Berlaymont Building

[andrea.saltelli@jrc.ec.europa.eu](mailto:andrea.saltelli@jrc.ec.europa.eu)



# A brief foreword on uncertainty

Peter Høeg, a Danish  
novelist, in *Borderliners*  
(Høeg, 1995)

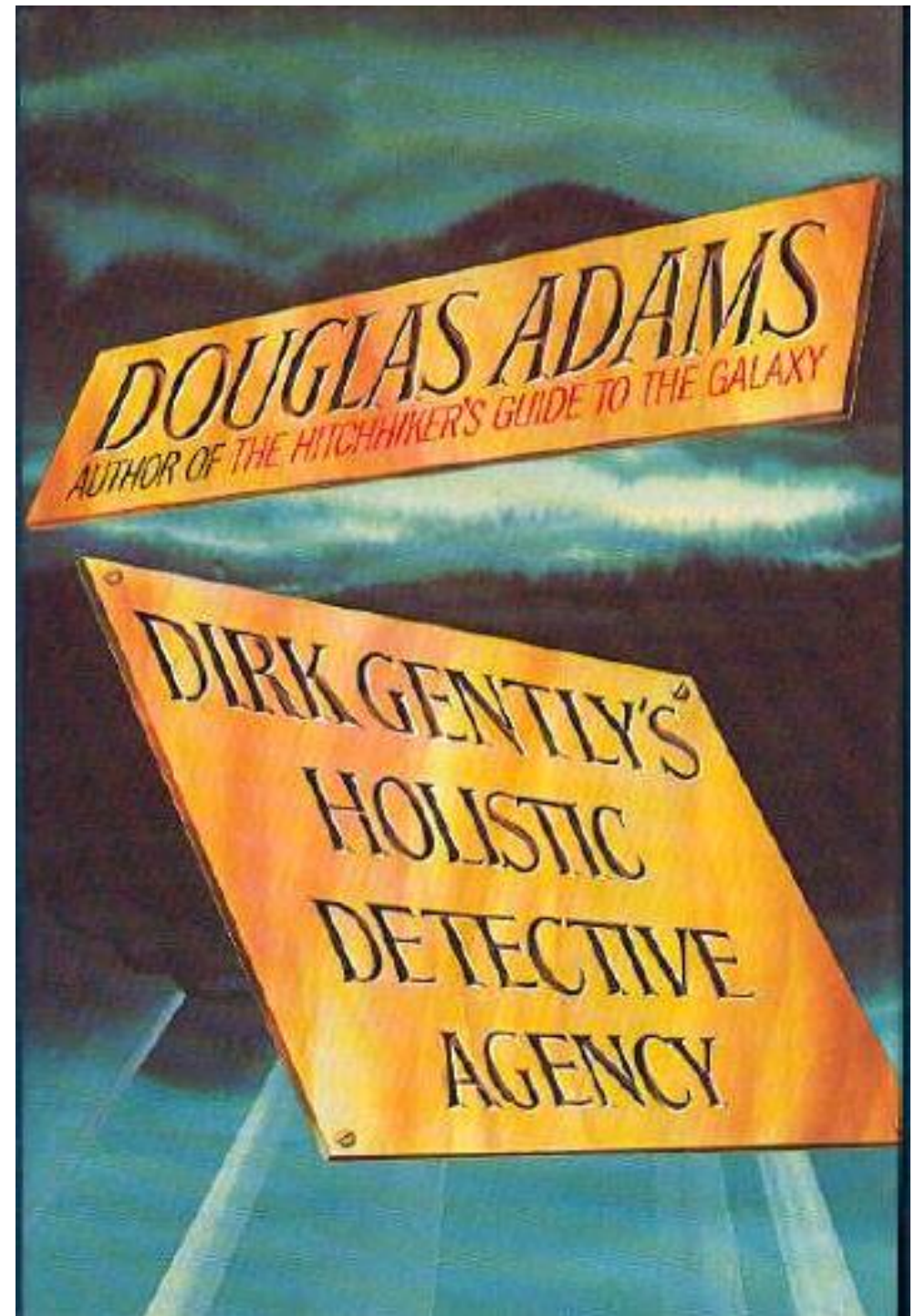


“That is what we meant by science. That both question and answer are tied up with uncertainty, and that they are painful. But that there is no way around them. And that you hide nothing; instead, everything is brought out into the open.”





Pocket Books 1987, p.69



“Well, Gordon’s great insight was to design a program which allowed you to specify in advance what decision you wished it to reach, and only then to give it all the facts. The program’s task, [...], was to construct a plausible series of logical-sounding steps to connect the premises with the conclusion.”

This talk will try to map a middle ground between these two extremes views, between candour and cynicism

- Uncertainty is not the opposite of quality
- Uncertainty cuts both ways
- Find uncertainty before uncertainty finds you





## Model structure uncertainty...

5 consultants, each using a different model were given the same question:  
*“which parts of this particular area are most vulnerable to pollution and need to be protected?”*

*(Refsgaard et al, 2006)*

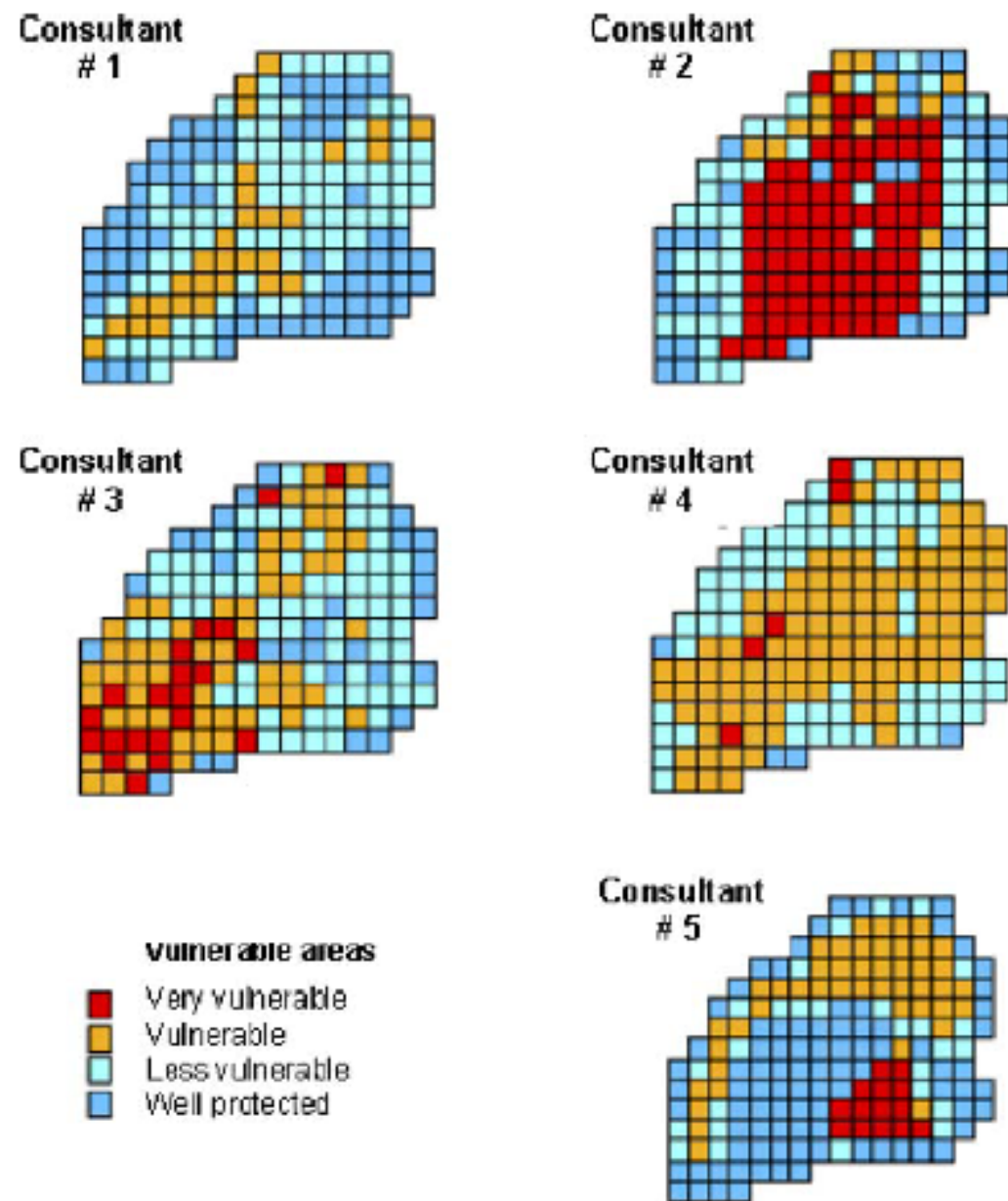
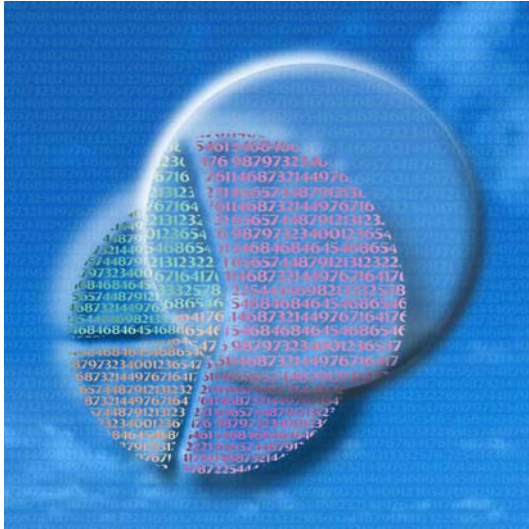


Fig. 1. Model predictions on aquifer vulnerability towards nitrate pollution for a 175 km<sup>2</sup> area west of Copenhagen [11].



Has the '**crunch**'  
something to do  
with  
**mathematical**  
**modeling?**

[« Back to Article](#)

WIRED MAGAZINE: 17.03

# Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon    02.23.09



$$\Pr[T_A < 1, T_B < 1] = \Phi_2(\Phi^{-1}(F_A(1)), \Phi^{-1}(F_B(1)), \gamma)$$

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

Do we need better models?

How about better ways of using  
them?

.. where 'better' has both normative and technical  
sides ...

“Yet we now know that the collective endeavour of these other very nice entrepreneurial scientists [the mathematicians who are employed in the world of finance] has resulted in the creation of a mountain of toxic fake securities. A sobering thought.”

**Jerome Ravetz, Morals and manners in modern science, *Nature* Vol. 457, 5 February 2009.**



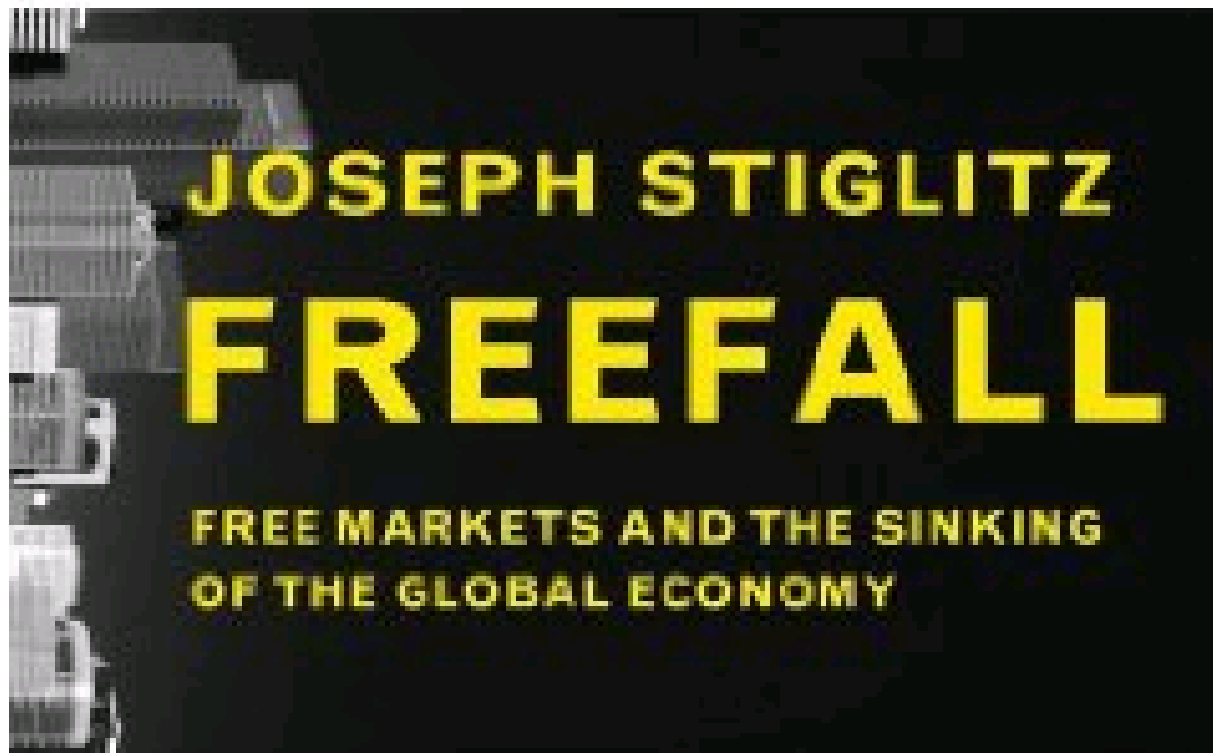


## How did the world end up in such a mess?

Mix greed and stupidity with fraud, corruption and terrible and/or not-enforced regulations and you destroy confidence and trust which are at the heart of the current slump

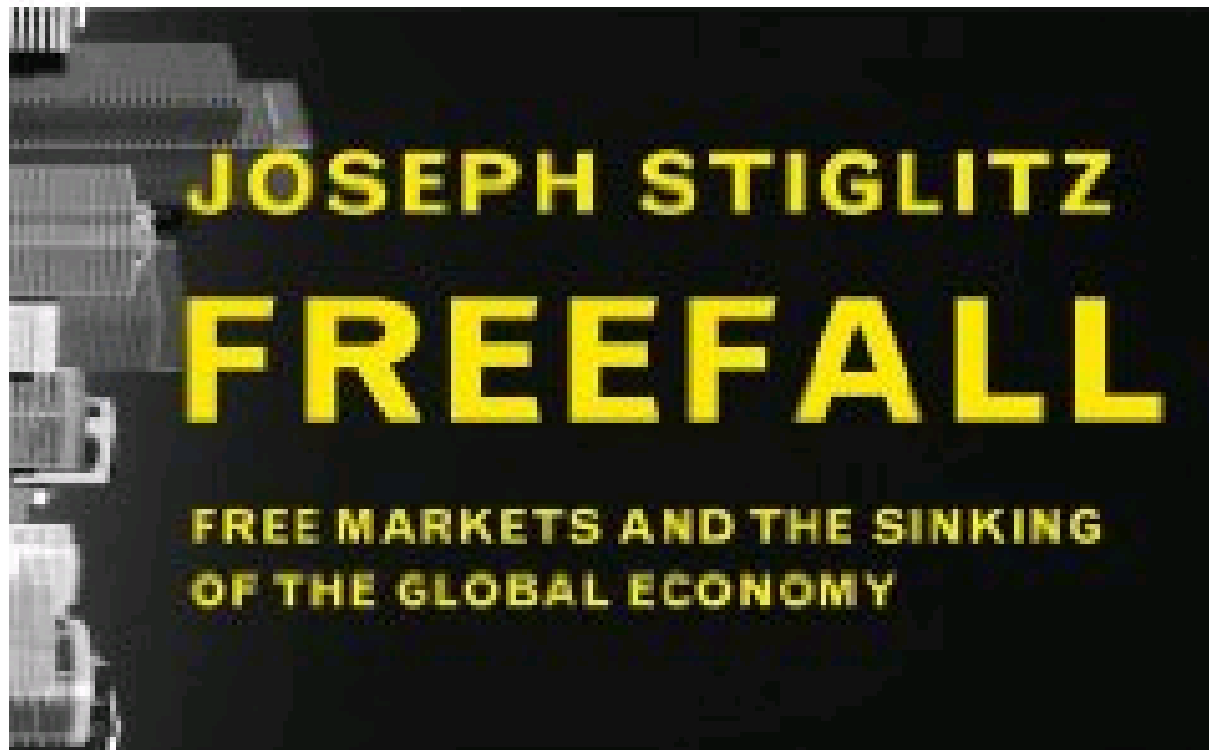
‘Perverse incentives and flawed models – accelerated by a race to the bottom’, p. 92

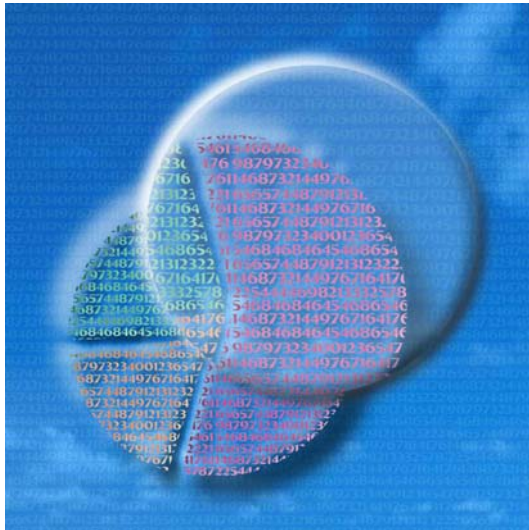
‘[...] the premise of securitization was diversification, but diversification only works if the loans that make up the security are not correlated’, p. 93



“Complexity – going beyond transparency

[...] Part of the agenda of computer models was to maximize the fraction of, say, a lousy sub-prime mortgage that could get an AAA rating, then an AA rating, and so forth,[...] This was called rating at the margin, and the solution was still more complexity”, p. 161





Has modelling  
losing legitimacy?

<<[...] most simulation models will be complex, with many parameters, state-variables and non linear relations. Under the best circumstances, such models have many degrees of freedom and, with judicious fiddling, can be made to produce virtually any desired behaviour, often with both plausible structure and parameter values.>>



George M.  
Hornberger,  
Professor at  
University of Virginia

HORNBERGER and Spear (1981).

*See: Douglas Adams' "Dirk Gently's  
Holistic Detective Agency", 1987*



<<Groundwater models cannot be validated [!]>> Konikov and Brechenovets, 1992.

Reviewed by Oreskes in 1994:

“Verification, Validation and Confirmation of numerical models in the earth sciences”.

Both papers focused on the impossibility of model validation.



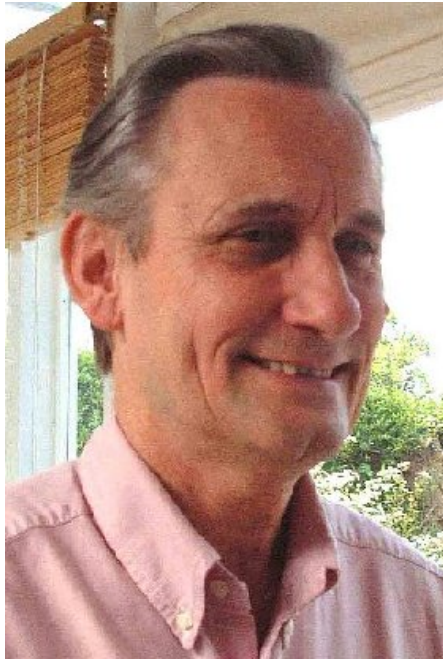
According to Oreskes, natural systems are never closed, and models put forward as description of these are never unique.

Models can never be 'verified' or 'validated', but only 'confirmed' or 'corroborated'.



Naomi Oreskes, Professor at University of California, San Diego.

# The critique of models



Peter Young,  
Professor at  
Lancaster

More than one model is  
compatible with the same  
set of data or evidence.

Equifinality ? →

← Indeterminacy?



Keith Beven,  
Professor at  
Lancaster



The IFPRI had raised about \$460,000 for the modeling, which would have provided insights to help policymakers [...]

When economist Carl Pray heard about plans for the first international assessment of agricultural research, a gold standard sprang to mind: the Intergovernmental Panel on Climate Change (IPCC). But things didn't turn out the way he expected.

IPCC has been pivotal in proving that climate change is real and linking it to human activities. As an agricultural economist at Rutgers University who has worked in many poor countries, Pray is convinced that agricultural research—and genetic modification in

mentally, socially and economically sustainable development through the generation, access to, and use of agricultural knowledge, science and technology?" Critics say this broad mandate made conflict inevitable and stunted the assessment's analytical rigor.

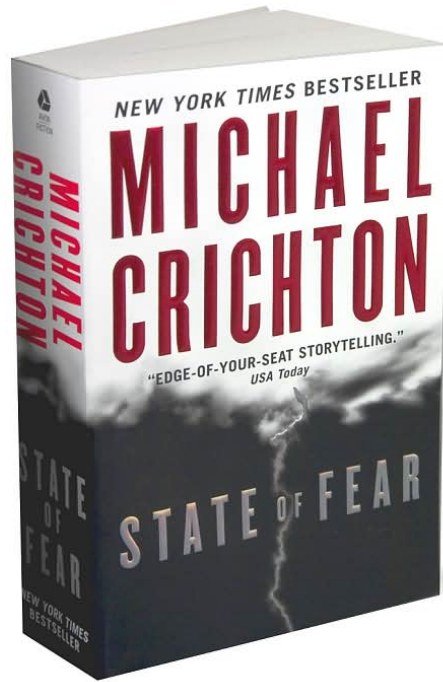
On several key issues, consensus proved elusive. Industry scientists and some academics—mainly agricultural economists and plant biologists—believe the assessment was "hijacked" by participants who oppose genetically modified (GM) crops and other common

the outcome. They note the experience of small-scale farmers, women, have finally been brought to the fore by the assessment. "It really deals with issues of power, influence, and benefits," says Marcia Kishi-Eitman of the Pesticide Action Network North America in San Francisco, California. Toby Kiers, who studies sustainable agriculture at Vrije University in Amsterdam, the Netherlands, agrees. "For technology to be most effective, farmers must be at the center, influencing how it is developed, delivered, and

loaded from www.scient

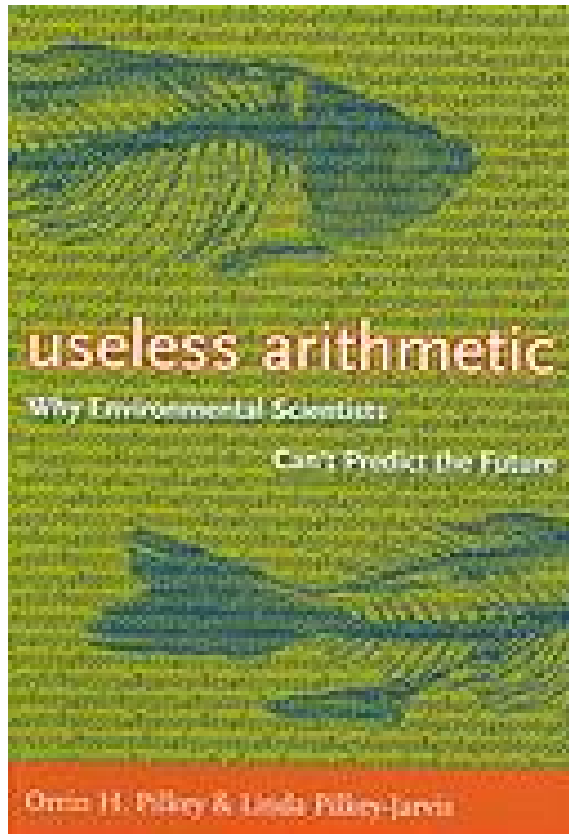
[... ] But Greenpeace [...] objected that the models were not "transparent".

Source: Dueling visions for an hungry world, Erik Stokstad, 14 MARCH 2008, **319** SCIENCE



"They talk as if simulation were real-world data. They 're not. That 's a problem that has to be fixed. I favor a stamp: WARNING: COMPUTER SIMULATION – MAY BE ERRONEOUS and UNVERIFIABLE. Like on cigarettes [...]"

Op. Cit. *p. 556* .



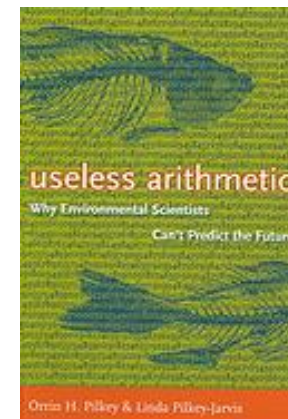
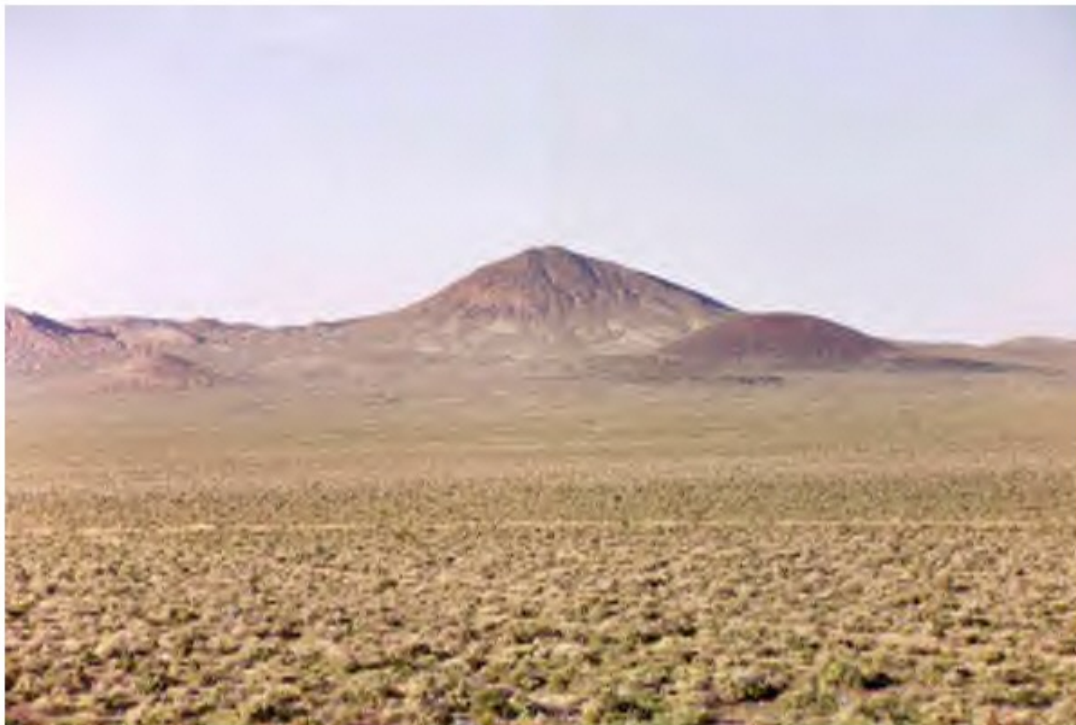
# Useless Arithmetic: Why Environmental Scientists Can't Predict the Future by Orrin H. Pilkey and Linda Pilkey-Jarvis

'Quantitative mathematical  
models used by policy makers  
and government administrators  
to form environmental policies  
are seriously flawed'

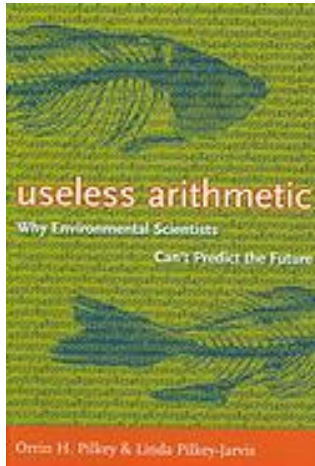


One of the examples discussed concerns the **Yucca Mountain** repository for radioactive waste. TSPA model (for total system performance assessment) for safety analysis.

TSPA is Composed of 286 sub-models.

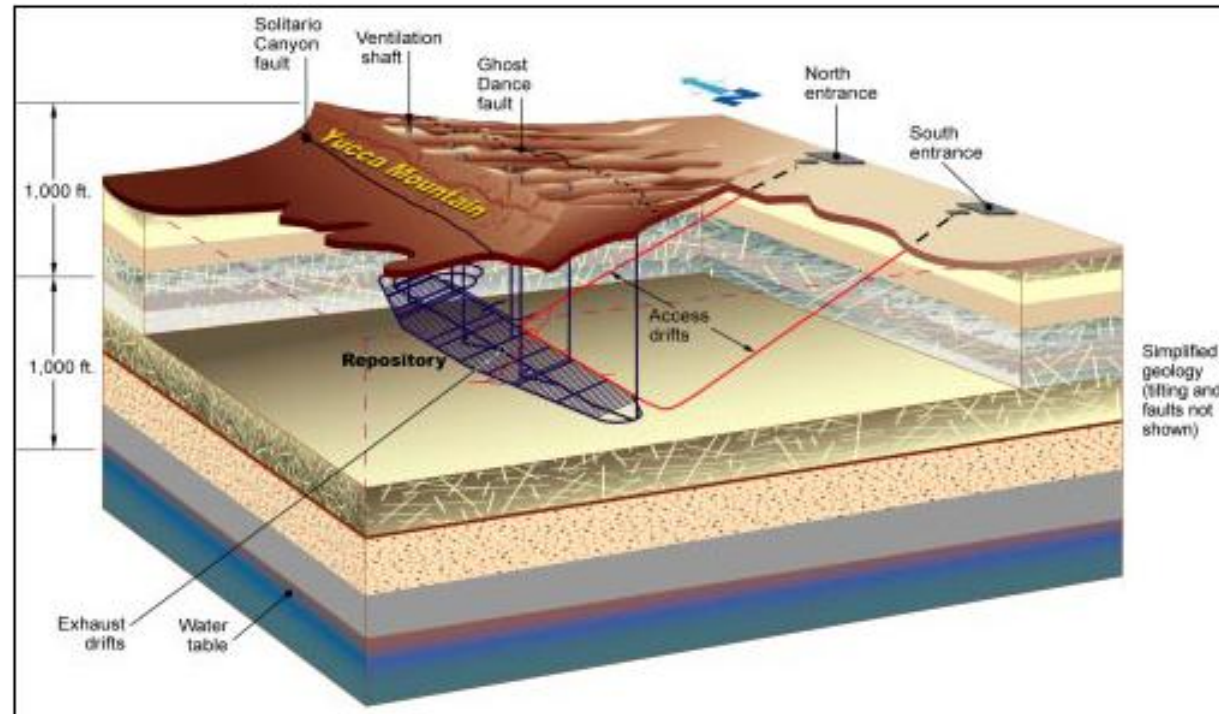
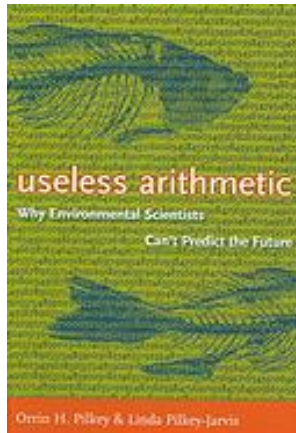






TSPA (like any other model) **relies on assumptions** → one is the low permeability of the geological formation → long time for the water to percolate from surface to disposal.



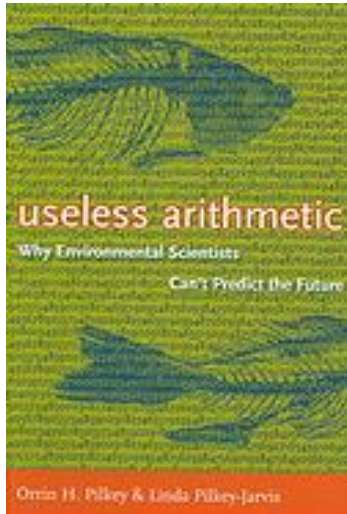


The confidence of the stakeholders in TSPA was not helped when evidence was produced which could lead to an upward revision of 4 orders of magnitude of this parameter (the  $^{36}\text{Cl}$  story)

Type III error in sensitivity: Examples:

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”

Ref. → Robert K. Merton's 'Organized skepticism '

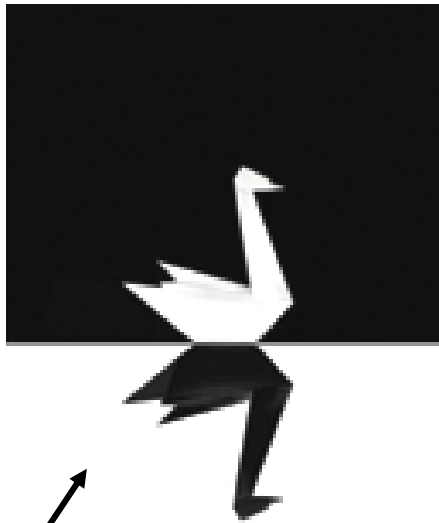
**Communalism** - the common ownership of scientific discoveries, according to which scientists give up intellectual property rights in exchange for recognition and esteem (Merton actually used the term Communism, but had this notion of communalism in mind, not Marxism);

**Universalism** - according to which claims to truth are evaluated in terms of universal or impersonal criteria, and not on the basis of race, class, gender, religion, or nationality;

**Disinterestedness** - according to which scientists are rewarded for acting in ways that outwardly appear to be selfless;

**Organized Skepticism** - all ideas must be tested and are subject to rigorous, structured community scrutiny.

We just can't predict, concludes N. N. Taleb, and we are victims of the ludic fallacy, of delusion of uncertainty, and so on. Modelling is just another attempt to 'Platonify' reality...



Nassim Nicholas Taleb, *The Black Swan*, Penguin, London 2007



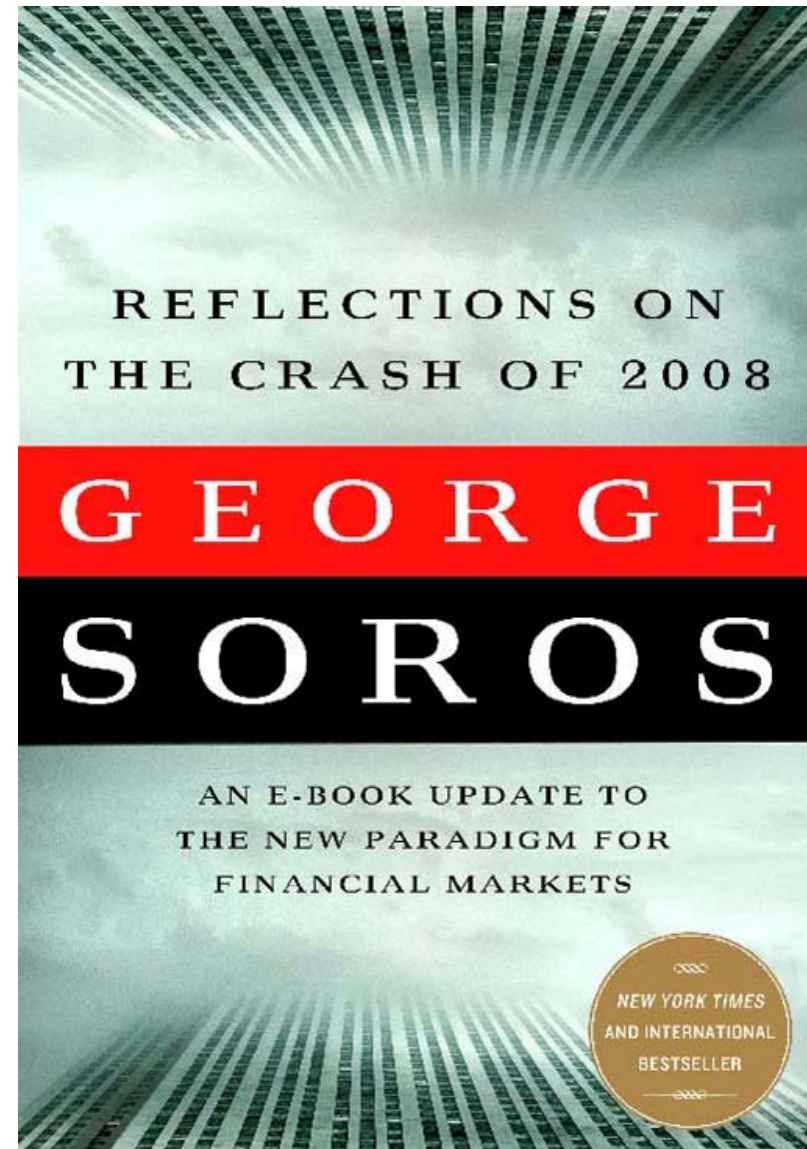
Written before the  
financial crisis



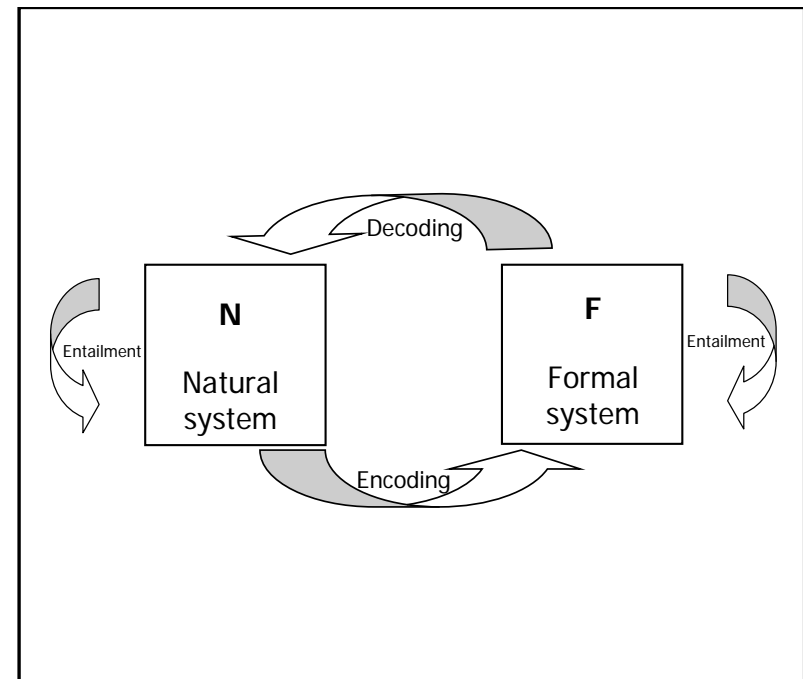
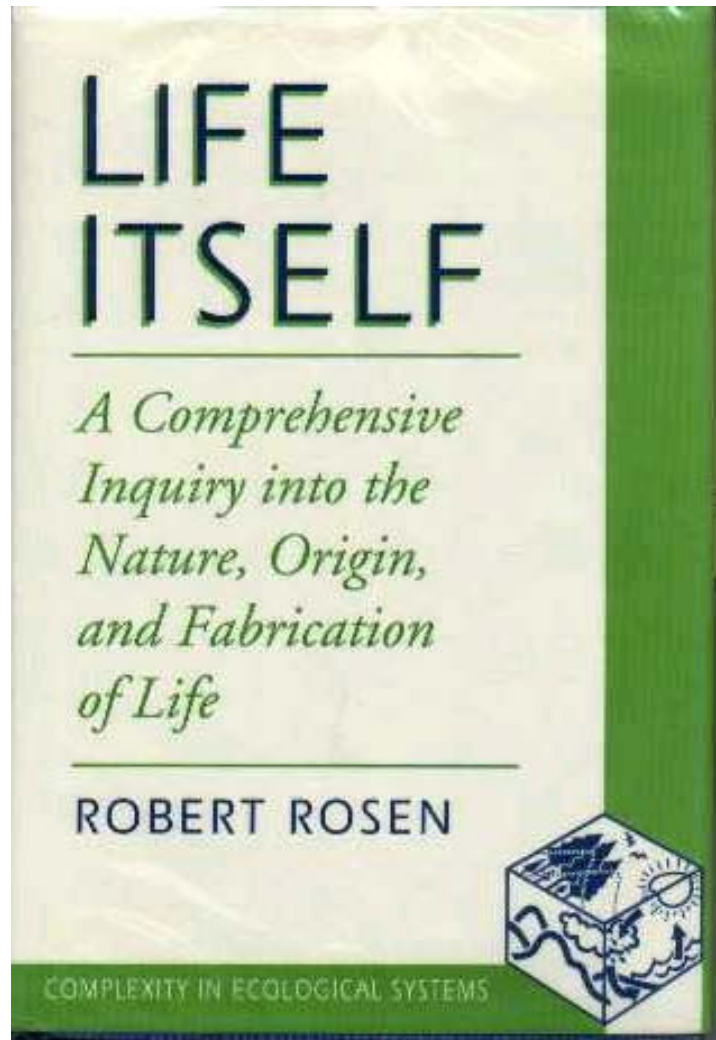
Postulate of 'radical  
fallibility':

"Whenever we acquire  
some useful knowledge,  
we tend to extend it to  
areas where it is no longer  
applicable"

(Taleb's -Platonification')

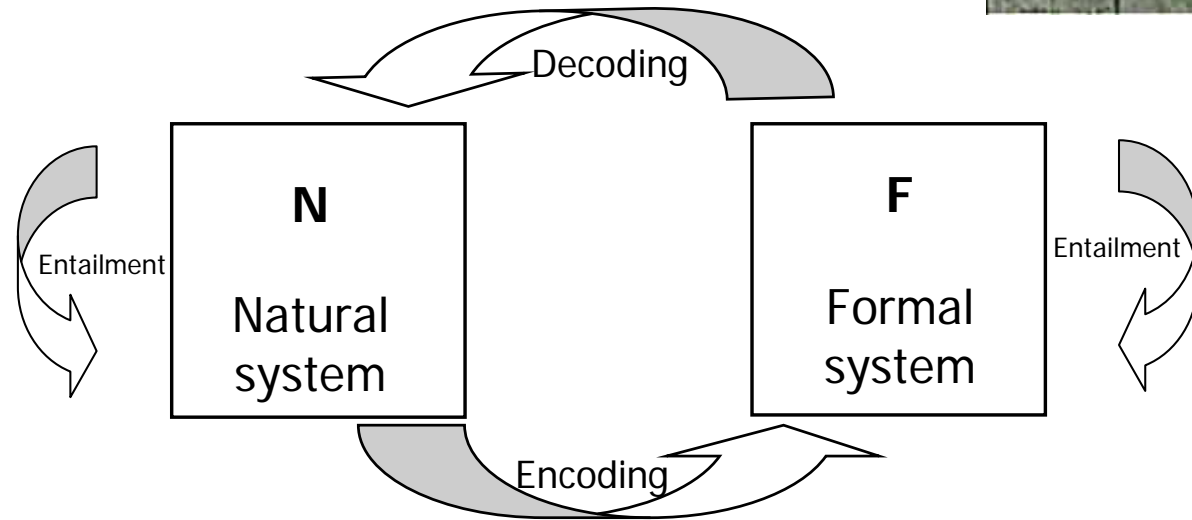






# The critique of models

The  
nature of  
models,  
after  
R. Rosen



## The RIVM media scandal (1999):



### Newspaper headlines:

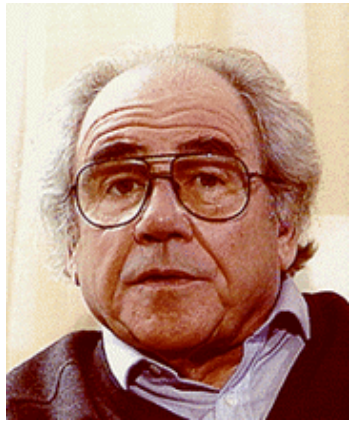
"Environmental institute lies and deceits"

"Fuss in parliament after criticism on environmental numbers"

"The bankruptcy of the environmental numbers"

"Society has a right on fair information, RIVM does not provide it"

Jeroen van der Sluijs, A way out of the credibility crisis around model-use in Integrated Environmental Assessment, *Futures*, **34** (2002) 133-146.

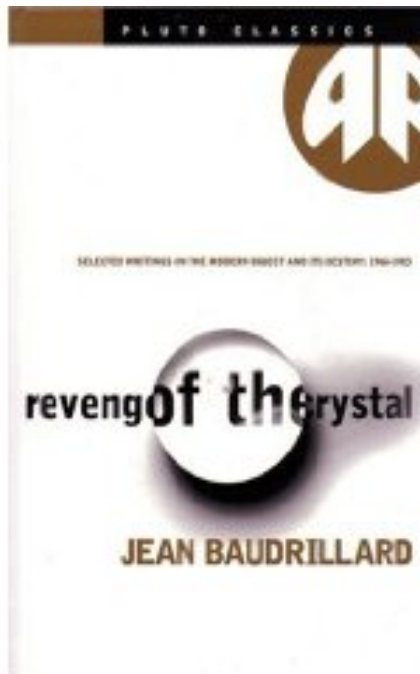


The RIVM media scandal (1999):

*"RIVM over-exact prognoses based on  
virtual reality of computer models"*



Words used by French philosopher Jean Baudrillard in *Revenge of the Crystal*, PLUTO Press 1999, p. 92



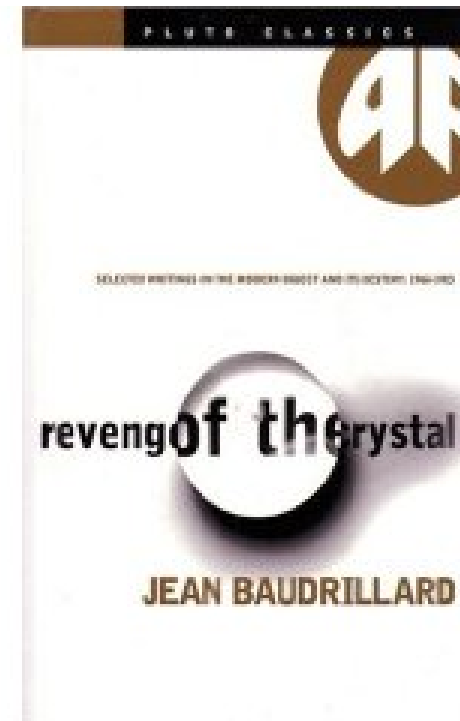
Jeroen van der Sluijs, A way out of the credibility crisis around model-use in Integrated Environmental Assessment, *Futures*, **34** (2002) 133-146.

See also [www.nusap.net](http://www.nusap.net)

He co-authored the RIVM/MNP Guidance on Uncertainty Assessment and Communication (Leidraad).

An immense process of simulation has taken place throughout all of everyday life , in the image of those '**simulation models**' on which operational and computer sciences are based. One 'fabricates' a model by combining characteristics or elements of the real; and, by making them 'act out' a future event, structure or situation, tactical conclusions can be drawn and applied to reality. It can be used as an **analytic tool under controlled scientific conditions**. In mass communication, **this procedure assumes the force of reality, abolishing and volatilizing the latter** in favour of that neo-reality of a model materialized by the medium itself.

Jean Baudrillard, *Revenge of the Crystal*, PLUTO Press 1999, p. 92





## RIVM/MNP Guidance for Uncertainty Assessment and Communication



### MINI-CHECK

### ELABORATION

#### 1. Problem Framing

In our assessment we pay attention to: (i) existing views on the problem other than the client's (including our own view), (ii) the interwovenness with other problems, (iii) possibly relevant aspects of the problem that are not dealt with in the research questions, (iv) the role the study is expected to play in the policy process, and (v) the way the study connects to previous studies on the subject.

*Indicate whether elaboration is or is not required and why (possibly for specific parts). If it is required, then go to Quickscan question 1.*

Wholly

Partly

Insufficiently

#### 2. Involvement of Stakeholders

We have a clear picture of: (i) the relevant stakeholders, (ii) their views and roles with respect to the problem, and (iii) the problem aspects about which they disagree. On the basis of all this, we have decided *if, how* (in formulating research questions, contributing information/data, evaluating findings/results), and when (in the beginning, during, after) we should involve which stakeholders in this assessment.

*Indicate whether elaboration is or is not required and why (possibly for specific parts). If it is required, then go to Quickscan question 2.*

Wholly

Partly

Insufficiently

#### 3. Selection of Indicators

We can provide adequate backing for the selection of indicators and their mutual relationships; we have considered alternative indicators, and in our report we discuss the limitations of the use of these indicators for this problem; we know the level of support among scientists and within society (including decision makers/politicians) for the use of these indicators.

*Indicate whether elaboration is or is not required and why (possibly for specific parts). If it is required, then go to Quickscan question 2.*

Wholly

Partly

Insufficiently

## PNS at NL Environmental Assessment Agency: Tools & checklists for Knowledge Quality Assessment

The position reflects the level of knowledge

Level of knowledge	low	high
<b>NH3 emission</b>		
Modelability		
Empirical basis		
Theoretical understanding		
<b>VOC emission from paint</b>		
Modelability		
Empirical basis		
Theoretical understanding		
<b>PM10 emission</b>		
Modelability		
Empirical basis		
Theoretical understanding		

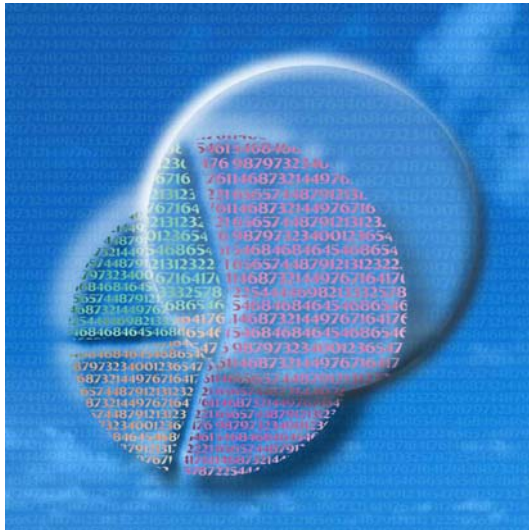
SCIENCE VOL 316 13 APRIL 2007

**"Today, eight years on from the Dutch scandal, no one makes more strenuous efforts than does the Netherlands' RIVM to accommodate and cope with the uncertainties of environmental data and models, hence to achieve the greatest possible quality in generating environmental foresight."**



(Bruce Beck)

Universiteit Utrecht



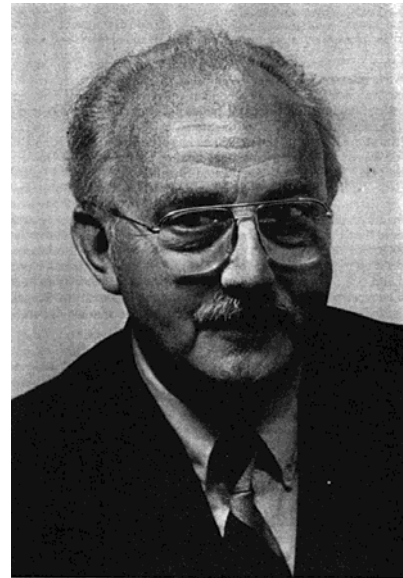
More snippets  
of wisdom



"The uncertainties which are more carefully scrutinised are usually those which are the least relevant" (*lampposting*, Jeroen van der Sluijs).

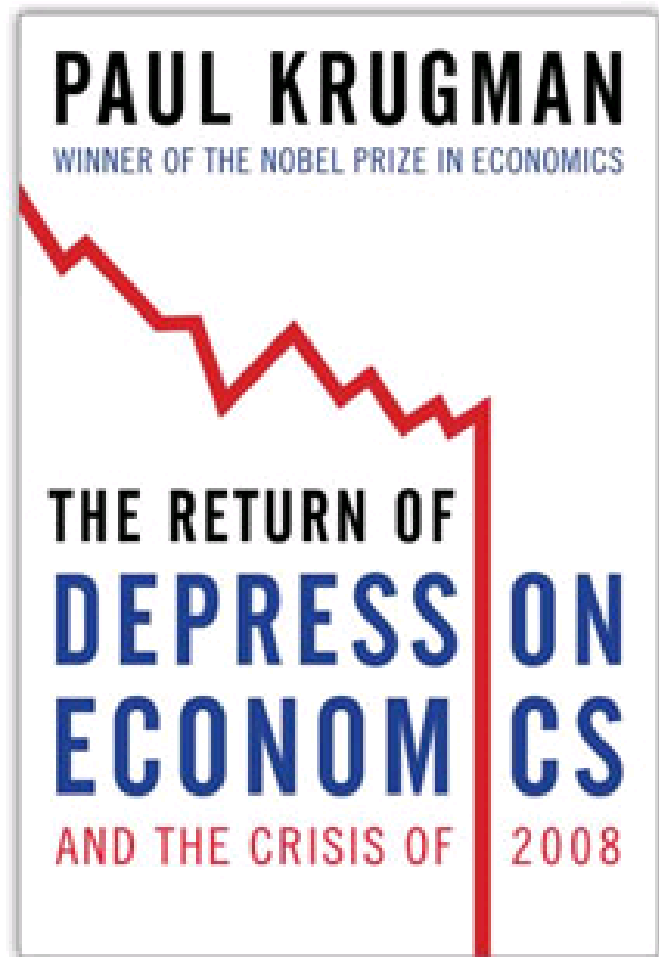
Nassim Nicholas Taleb calls this 'The delusion of uncertainty' .

George Box: 'all models are wrong, some are useful'



*George E. P. Box*

Box, G.E.P., Robustness in the strategy of scientific model building, in Robustness in Statistics, R.L. Launer and G.N. Wilkinson, Editors. 1979, Academic Press: New York.

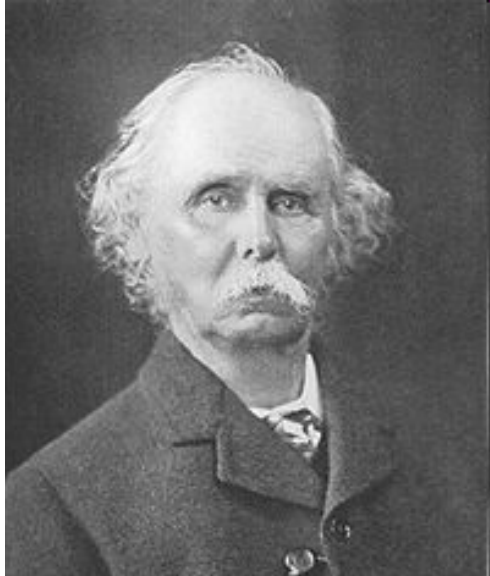


Anyway the equation and diagrams of formal economics are, more often than not, no more than a scaffolding used to help construct an intellectual edifice. Once that edifice has been built to a certain point, the scaffolding can be stripped away, leaving only plain English behind. (Paul Krugman, *The return of Depression Economics*, 2009).

Frank H. Knight, 1921. Risk, Uncertainty, and Profit.

We live in a world of contradiction and paradox, a fact of which perhaps the most fundamental illustration is this: that the existence of a problem of knowledge depends on the future being different from the past, while the possibility of the solution of the problem depends on the future being like the past."





(1) Use mathematics as a shorthand language, rather than as an engine of inquiry.

(2) Keep to them till you have done.

(3) **Translate into English.**

(4) Then illustrate by examples that are important in real life.

(5) **Burn the mathematics.**

(6) If you can't succeed in (4), burn (3). This last I [Marshall] did often.

**Alfred Marshall**, Memorials of Alfred Marshall, ed. A.C. Pigou (London: Macmillan, 1925), 427.

# GIGO (Garbage In, Garbage Out) Science – or pseudo-science – “where uncertainties in inputs must be suppressed least outputs become indeterminate”

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



But: It is possible to disentangle **evidence based policy** from **policy based evidence**? E.g. in impact assessment work?

see Benoît GODIN on Eugenics and the birth of R&D stats: The Culture of Numbers: From Science to Innovation, INRS, Montreal, Canada, Communication presented to the Government-University-Industry Research Roundtable (GUIRR) US National Academy of Sciences, Washington, May 21, 2010.

... but many other data based stories as well: Tobacco & health, capital punishment & crime rate ...

Oreskes, N., Conway E. M., 2010, Merchants of Doubt, Bloomsbury Press

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



Scientists (including statisticians), do make value judgments.

Examples: How low should be a low probability be; arbitrary choices in the level of significance or level of confidence to be selected.

Rudner, R. 1953, The scientist qua scientist makes value judgments, *Philosophy of Science*, **20**(1), 1-6.

## **The Scientist Qua Scientist Makes Value Judgments**

Richard Rudner

*Philosophy of Science*, Vol. 20, No. 1 (Jan., 1953), pp. 1-6

...clearly the scientist as scientist does make value judgments. For, since no scientific hypothesis is ever completely verified, in accepting a hypothesis the scientist must make the decision that the evidence is sufficiently strong or that the probability is sufficiently high to warrant the acceptance of the hypothesis. Obviously our decision regarding the evidence and respecting how strong is "strong enough", is going to be a function of the importance, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis.

## Statistics for policy: three models

A **rational-positivist** model for the use of indicators and policy (good quality statistics underpin good policies)

**Discursive-interpretive** model (statistics contribute to a process of framing of and focusing on an issue among the many competing for public's attention)

**Strategic** model (statistics is used by parties competing for a given constituency).

see Boulanger, P-M., Political uses of social indicators: overview and application to sustainable development indicators. *International Journal of Sustainable Development*, **10** (1,2):14-32, 2007.

## **Impact Assessment Scene setter**

[there are] credibility problems that are structural  
[to impact assessment], testing our own bread,  
and [there is a] need for some kind of  
external/objective validation.

IAB scrutiny already provides some validation of  
this kind, but especially for heavily model-based  
work stronger medicine is needed.

[undisclosed house source]

Reproducibility

is a necessary  
condition for



Transparency

is a necessary  
condition for



Legitimacy

is a necessary  
condition for



Epistemic authority

## Implications

Ensure **external validation** of one's own funding

Use econometric modelling (lots of data, few equations) for inference (**proof of causality**) rather than closed form models

If using closed form models use **simple** ones

## Implications (ctd)

If using closed form complex model

- distinguish **analytic** from **evidential** use
- be ready to negotiate on the basis of ad **hoc simplified representations** of the models;
- have a **pedigree**.

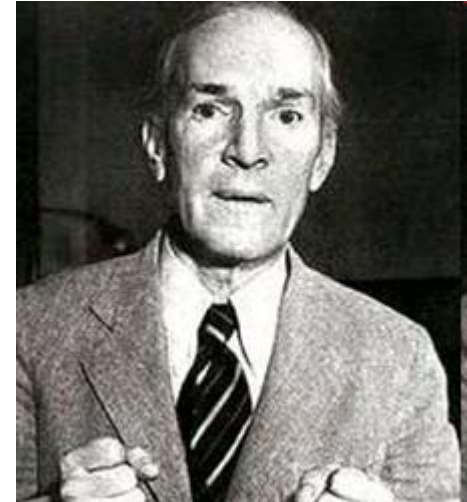


Reasons for doing the above:

- Practitioners' and guidelines' recommendation
- Loss of trust in institutions / science / mathematical models
- Stakeholders' scepticism

Upton Sinclair

“It is difficult to get a man to understand something when his salary depends upon his not understanding it”

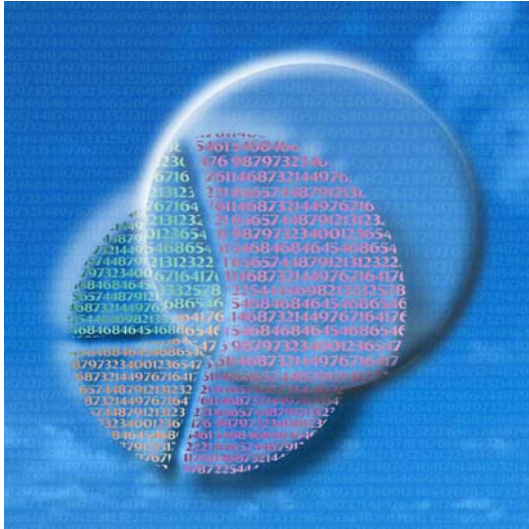


Thus far:

- a) The making of a model is not scientifically prescribed
- b) Models are particularly vulnerable to instrumental or otherwise unethical use
  - i.e. they are more vulnerable than laboratory based experiments

c) Stakeholders will tend to expect or suspect instrumental use of models. They will “believe everything was possible and that nothing was true” (H. Arendt)

Could sensitivity analysis help?



Econometricians'  
take on the matter

'Sensitivity analysis would Help'



Edward E. Leamer, UCLA

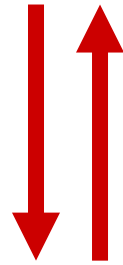
<<I have proposed a form of **organised sensitivity analysis** that I call "**global sensitivity analysis**" in which a neighborhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighborhood of assumptions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

Edward E. Leamer, 1990, Let's  
Take the Con Out of Econometrics,  
*American Economics Review*, **73**  
(March 1983), 31-43.





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



**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’.

With the ashes of the mathematical models used to rate mortgage-backed securities still smoldering on Wall Street, now is an ideal time to revisit the sensitivity issues.

Tantalus on the Road to Asymptopia

Edward E. Leamer, 2010 *Journal of Economic Perspectives*, **24**, (2), 31–46.



"... my observation of economists at work who routinely pass their data through the filters of many models and then choose a few results for reporting purposes." *Ibidem*



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

*Ibidem*

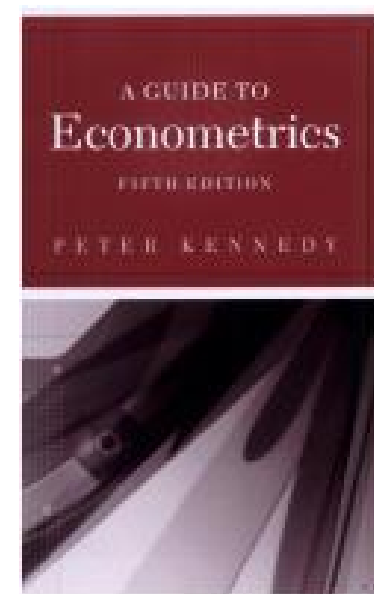
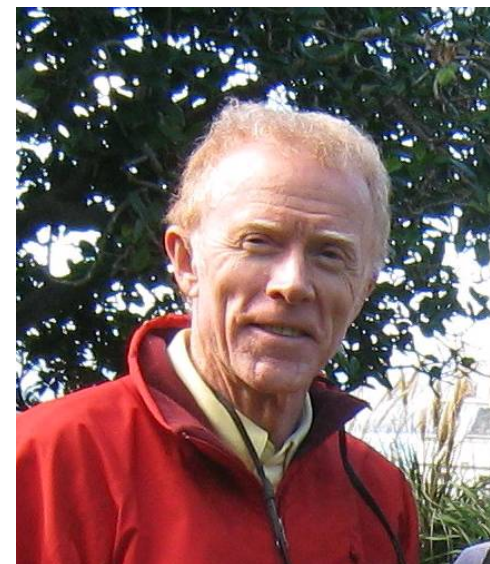
"or, to put it another way, a fanatical commitment to fanciful formal models is often needed to create the appearance of progress." *Ibidem*

# The critique of models <-> Uncertainty

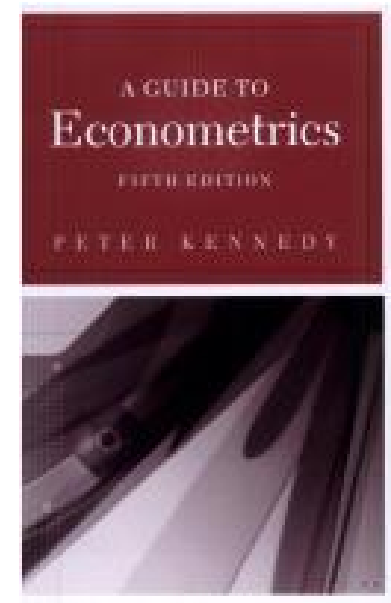
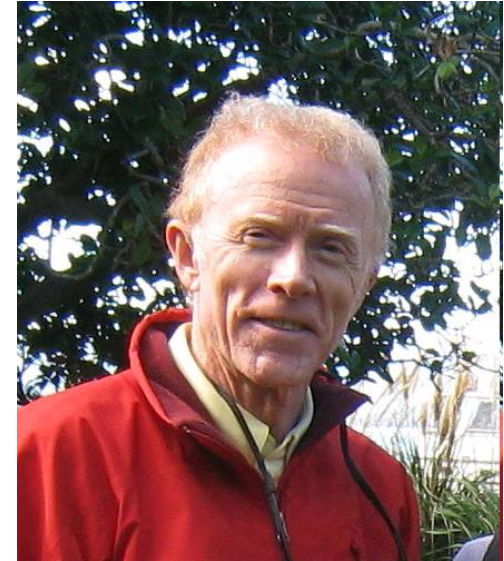
Peter Kennedy, A Guide to Econometrics.

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 >>



<<When reporting a sensitivity analysis, researchers should explain fully their specification search so that the readers can judge for themselves how the results may have been affected. This is basically an 'honesty is the best policy' approach, advocated by Leamer'.>>



Definition of uncertainty and sensitivity analysis.

**Sensitivity analysis:** The study of how **uncertainty** in the output of a model (numerical or otherwise) can be **apportioned** to different **sources** of uncertainty in the model input.

**Uncertainty analysis:** Focuses on just **quantifying the uncertainty** in model output.

In sensitivity analysis:

Type I error: assessing as important a non important factor

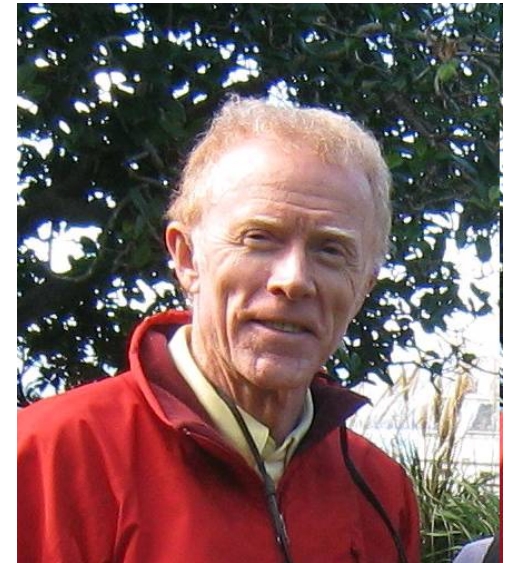
Type II: assessing as non important an important factor

Type III: analysing the wrong problem



The spectre of type III errors:  
= answering the wrong question

Peter Kennedy's II commandment of  
applied econometrics: 'Thou shall  
answer the right question', Kennedy  
2007



The spectre of type III errors:

Donald Rumsfeld version: "Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know we know. We also know there are **known unknowns**; that is to say we know there are some things we do not know. But there are also **unknown unknowns** -- the ones we don't know we don't know."





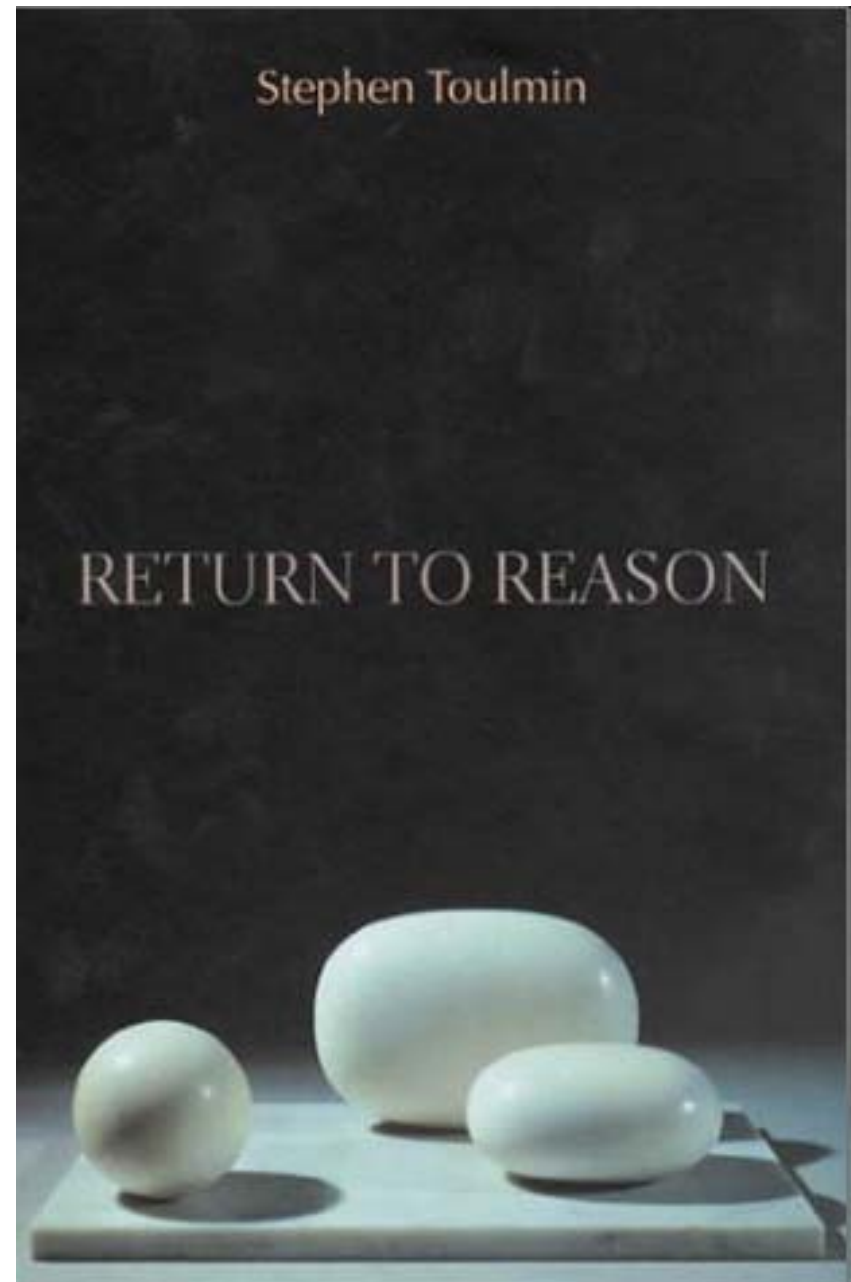
Do the sum right

Versus

Do the right sums

*(Stephen Toulmin)*

A plea for reasonableness  
versus rationality

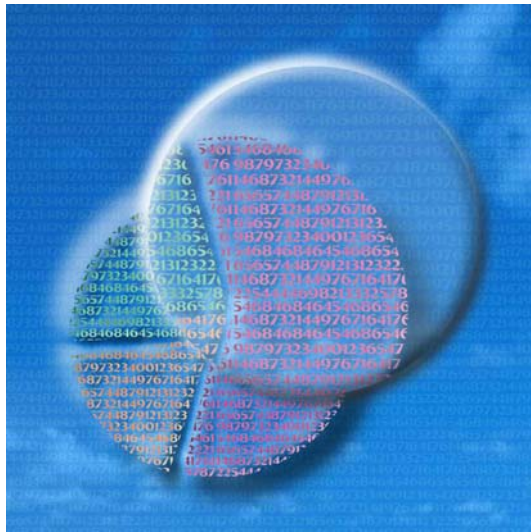


SA → implications for model quality

What constitutes an input for the analysis depends upon how the analysis is set up ...

... and will instruct the modeller about those factors which have been included... thus

→ the modeller will remain ignorant of the importance of those variables which have been kept fixed.



# The spiritual comfort of guidelines



The US OFFICE OF  
MANAGEMENT AND BUDGET  
(OMB)

in its **controversial** 'Proposed  
Risk Assessment Bulletin' also  
prescribes how to do a  
sensitivity analysis.



#### 4. Standard for Characterizing Uncertainty

Influential risk assessments should characterize uncertainty with a **sensitivity analysis** and, where feasible, through use of a **numeric distribution**

factors →  
prioritization

[...] **Sensitivity analysis** is particularly useful in pinpointing which assumptions are **appropriate candidates for additional data collection** to narrow the degree of uncertainty in the results. **Sensitivity analysis is generally considered a minimum, necessary component of a quality risk assessment report.**

**OFFICE OF MANAGEMENT AND BUDGET**

**Proposed Risk Assessment Bulletin (January 9, 2006)**

<http://www.whitehouse.gov/omb/infoereg/>





"John Graham has led the White House mission to change agencies' approach to risk " *ibidem* in ***Nature***

Proposed Risk Assessment Bulletin (January 9, 2006)  
<http://www.whitehouse.gov/omb/inforeg/>

## Why controversial?

"The aim is to bog the process down, in the name of transparency" (Robert Shull). [...] the proposed bulletin resembles several earlier efforts, including rules on 'information quality' and requirements for cost-benefit analyses, that make use of the OMB's extensive powers to weaken all forms of regulation.

Colin Macilwain, Safe and sound? ***Nature***, 19 July 2006.



---

## Part IX

# Office of Management and Budget

---

Guidelines for Ensuring and Maximizing  
the Quality, Objectivity, Utility, and  
Integrity of Information Disseminated by  
Federal Agencies; Notice; Republication



The OMB about  
transparency

<http://www.whitehouse.gov/omb/inforeg/>

[models should be made available to a third party so that it can ] **use the same data, computer model or statistical methods to replicate the analytic results reported in the original study.**

[...] The more important benefit of transparency is that **the public will be able to assess how much an agency's analytic result hinges on the specific analytic choices** made by the agency.

Friday, February 22, 2002

Graphic - Federal Register, Part IX

**Office of Management and Budget**

Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication

<http://www.whitehouse.gov/omb/inforeg/>

Concreteness about analytic choices allows, for example, the implications of alternative technical choices to be readily assessed. **This type of sensitivity analysis is widely regarded as an essential feature of high-quality analysis**, yet sensitivity analysis cannot be undertaken by outside parties unless a high degree of transparency is achieved. The OMB guidelines do not compel such sensitivity analysis as a necessary dimension of quality, but the transparency achieved by reproducibility will allow the public to undertake sensitivity studies of interest.

Friday, February 22, 2002

Federal Register, Part IX

**Office of Management and Budget**

Guidelines for Ensuring and Maximizing the Quality,

Objectivity, Utility, and Integrity of Information

Disseminated by Federal Agencies

<http://www.whitehouse.gov/omb/inforeg/>

See also ibidem

Open-Source Policy

Modelling, by Max Henrion,

15 June 2006

## Implications

Ensure external validation of one's own funding

Use econometric modelling (lots of data, few equations) for inference (proof of causality) rather than closed form models

If using closed form models use simple ones

## Implications (ctd)

If using closed form complex model

- distinguish **analytic** from **evidential** use
- be ready to negotiate on the basis of ad **hoc simplified representations** of the models;
- have a **pedigree**.

Reasons for doing the above:

- Practitioners' and guidelines' recommendation
- Loss of trust in institutions / science / mathematical models
- Stakeholders' scepticism

Reproducibility

is a necessary  
condition for

Transparency

is a necessary  
condition for

Legitimacy

is a necessary  
condition for

Epistemic authority



One more remark on simplification: it can be a good thing to have a lumped version of a model.

*“As the complexity of a system increases, our ability to make precise and yet significant statements about its behaviour diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.”*

Zadeh's incompatibility principle (1965).



Lofti Zadeh

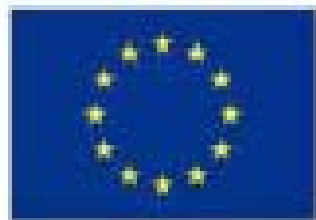


## Guidance on the Development, Evaluation, and Application of Environmental Models



" [SA] methods should preferably be able to deal with a model regardless of assumptions about a model's linearity and additivity, consider interaction effects among input uncertainties, [...], and evaluate the effect of an input while all other inputs are allowed to vary as well."

European Commission



## IMPACT ASSESSMENT GUIDELINES

15 January 2009

SEC(2009) 92

“... Sensitivity analysis can be used to explore how the impacts of the options you are analysing would change in response to variations in key parameters and how they interact.”

## Other prescriptions

Intergovernmental Panel on Climate Change  
(IPCC, 1999, 2000)

The IPCC mentions the existence of  
"...sophisticated computational techniques  
for determining the sensitivity of a model  
output to input quantities...", while in fact  
recommending merely local (derivative  
based) methods.

Although the IPCC background papers advise the reader that [... the sensitivity is a local approach and is not valid for large deviations in non-linear functions...], they do not provide any prescription for non-linear models.

# Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties

**IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties**

**Jasper Ridge, CA, USA**

**6-7 July 2010**

*[...] These notes define a common approach and calibrated language that can be used broadly for developing expert judgments and for evaluating and communicating the degree of certainty in findings of the assessment process*

## Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties

IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties  
Jasper Ridge, CA, USA  
6-7 July 2010

The AR5 will rely on two metrics for communicating the degree of certainty in key findings:

- Confidence in the validity of a finding, based on the type, amount, quality, and consistency of evidence (e.g., mechanistic understanding, theory, data, models, expert judgment) and the degree of agreement. Confidence is expressed qualitatively.
- Quantified measures of uncertainty in a finding expressed probabilistically (based on statistical analysis of observations or model results, or expert judgment).

## ➔ Why SA

- communicate limits to knowledge
- ascertain relative merit of theories
- achieve simplicity/parsimony and hence transparency ...
- ...

## My own prescriptions:

- 1) Choose carefully one target variable
- 2) Explore carefully the input factors space; be quantitative; spot interactions among factors → make the analysis in one shot, and not piecewise.
- 3) Look at uncertainties before going public with findings (with E.E. Leamer and P. Kennedy)



## **Do I need uncertainty and sensitivity analysis in impact assessment?**

- ➔ I want to be sure that my model-based inference can stand in court
- ➔ I want to look into a stakeholder's black box

## Examples of troubles to be anticipated :

- ➔ <<You treated X as a constant when we know it is uncertain by at least 30%>>
- ➔ <<Beware: It would be sufficient for a 5% error in X to make your statement about Z fragile>>
- ➔ <<Your model is but one of the plausible models – you neglected model uncertainty>>

- ➔ << You have maximized instrumentally your level of confidence in the results >>
- ➔ << You have artificially inflated uncertainty >>
- ➔ << Your framing is not socially robust >>

All of the above can be used to defend an assessment as well as to invalidate one.

Uncertainty can be instrumentally amplified

*Industry groups are fighting  
government regulation by  
fomenting scientific uncertainty*

# DOUBT

By David Michaels  
Photographs by Mindy Jones

## Is Their Product

Few scientific challenges are more complex than understanding the health risks of a chemical or drug. Investigators cannot feed toxic compounds to people to see what doses cause cancer. Instead, laboratory researchers rely on animal tests, and

vinyl chloride, chromium, benzene, benzidine, nickel, and a long list of other toxic chemicals and medications. What is more, Congress and the administration of President George W. Bush have encouraged such tactics by making it easier for private citizens to challenge government-funded research. At

Scientific American, Jun2005, Vol. 292, Issue 6

*Industry groups are fighting  
government regulation by  
fomenting scientific uncertainty*

# DOUBT

By David Michaels  
Photographs by Mindy Jones

## Is Their Product

**F**ew scientific challenges are more complex than understanding the health risks of a chemical or drug. Investigators cannot feed toxic compounds to people to see what doses cause cancer. Instead laboratory researchers rely on animal tests, and vinyl chloride, chromium, benzene, benzidine, nickel, and a long list of other toxic chemicals and medications. What is more, Congress and the administration of President George W. Bush have encouraged such tactics by making it easier for critics to sue to challenge government-funded research. At

In the US, intense 'exchange' between EPA's investigators and OMB's administrators.

*Industry groups are fighting  
government regulation by  
fomenting scientific uncertainty*

# DOUBT

By David Michaels  
Photographs by Mindy Jones

## Is Their Product

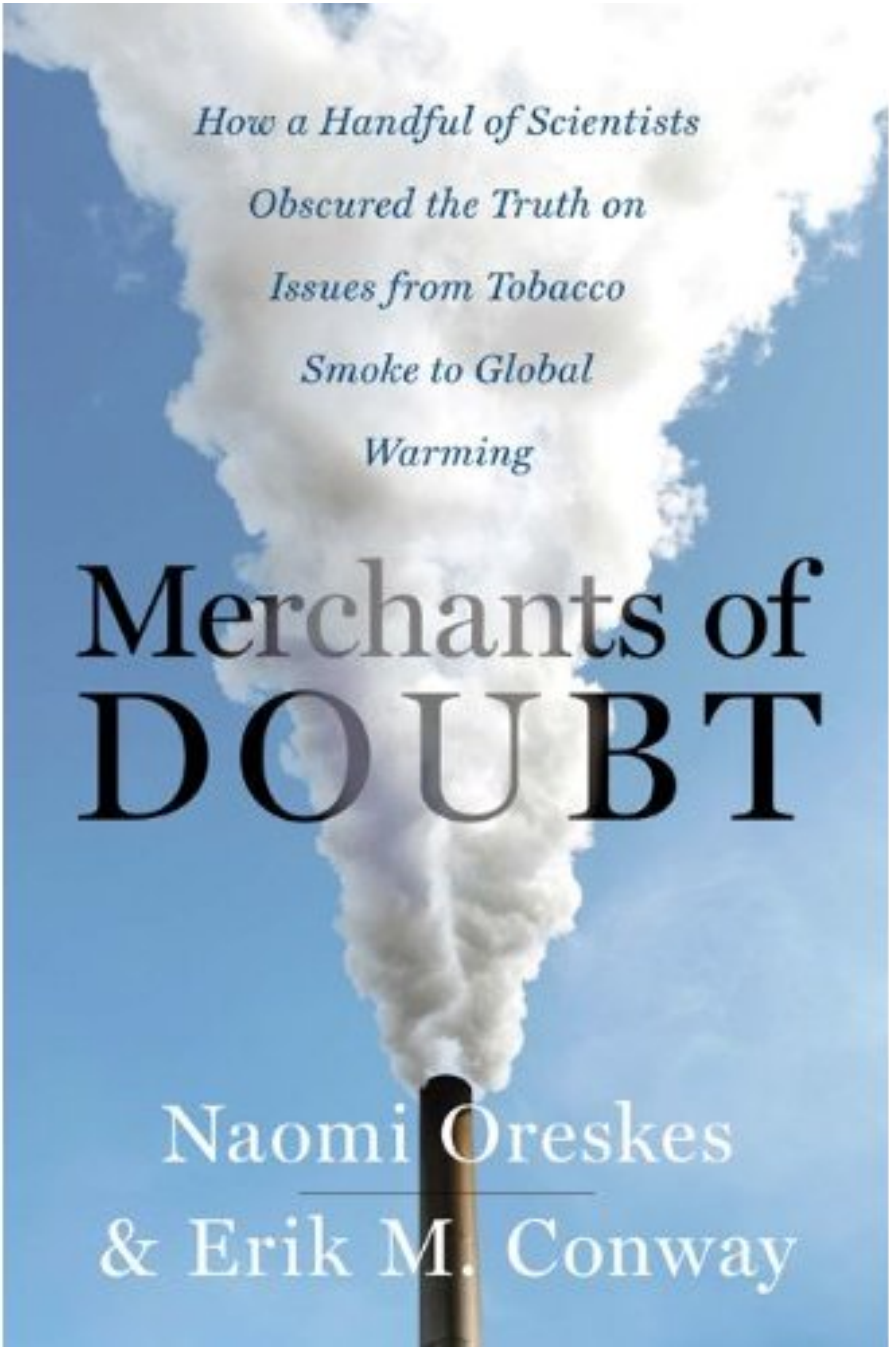
Few scientific challenges are more complex than understanding the health risks of a chemical or drug. Investigators cannot feed toxic compounds to people to see what doses cause cancer. Instead, laboratory researchers rely on animal tests, and vinyl chloride, chromium, benzene, benzidine, nickel, and a long list of other toxic chemicals and medications. What is more, Congress and the administration of President George W. Bush have encouraged such tactics by making it easier for industry groups to challenge government-funded research. At

- Fabrication (and politicisation) of uncertainty

The example of the US Data quality act and of the OMB "Peer Review and Information Quality" which

"seemed designed to maximize the ability of corporate interests to manufacture and magnify scientific uncertainty".





*How a Handful of Scientists  
Obscured the Truth on  
Issues from Tobacco  
Smoke to Global  
Warming*

# Merchants of DOUBT

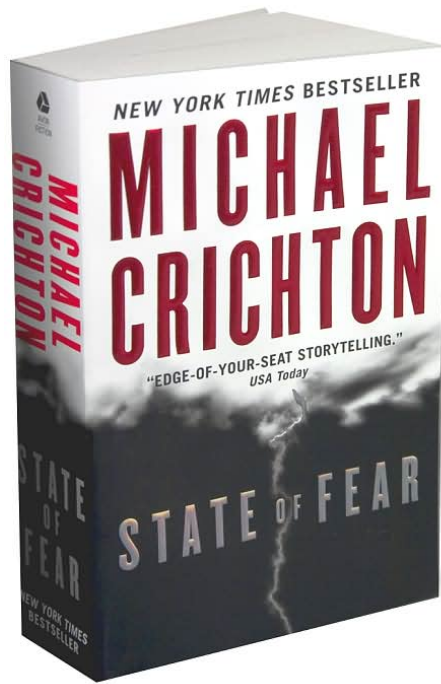
Naomi Oreskes  
& Erik M. Conway



Uncertainty can be instrumentally downplayed



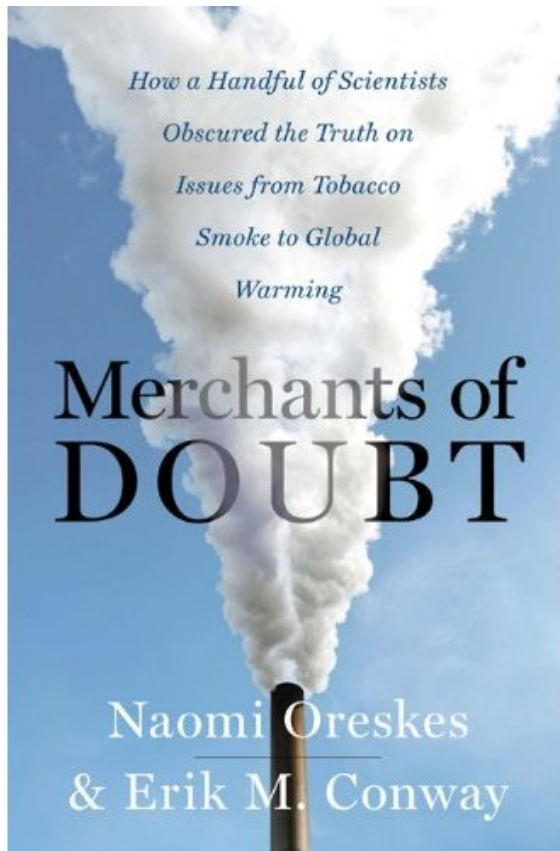
According to  
Leamer it was so in  
relation to whether  
capital punishment  
detects murder (in  
the seventies)



According to Crichton  
it was so on global  
change ... as it was  
done on Eugenics at  
the beginning of the  
XX century.

See also Richard S. Lindzen,  
Science and Politics: **Global  
Warming and Eugenics**, from  
Risks, Costs, and Lives  
Saved, R.W. Hahn, editor,  
Oxford University Press, New  
York, 1996.

So? Is global change's story like  
Eugenics' or like tobacco's?



For these authors it is  
like for tobacco

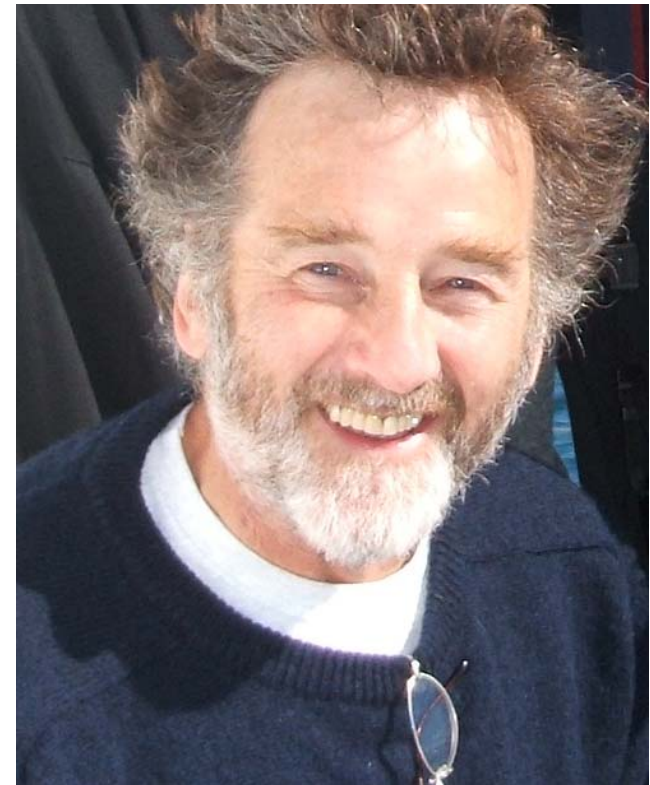
## BOOKS & ARTS

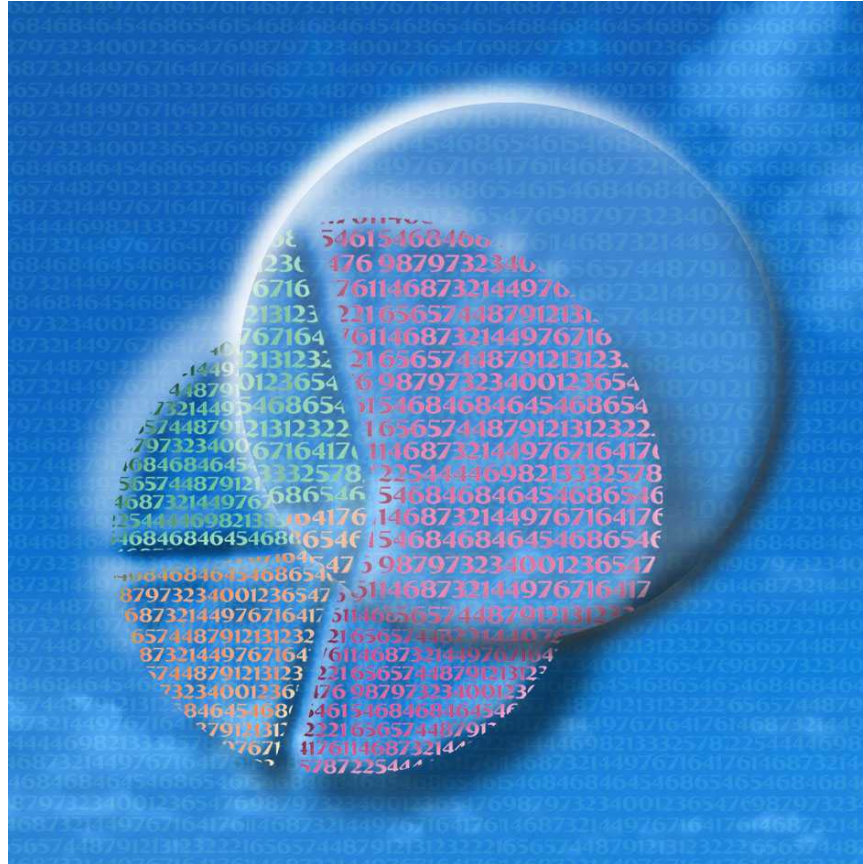
---

# When doubt becomes a weapon

**Brian Wynne** wishes that a book on the vulnerability of scientific evidence to attack by ideologists had grappled more with the larger question of why science is such an easy target.

“Oreskes and Conway could have gone further in asking how scientific uncertainty should be interpreted in policy, and how science can be led to overreach itself in arbitrating public facts, meanings and norms.”





Perfunctory  
sensitivity  
analysis?

**From:** Saltelli, A., D'Hombres, 2010, Sensitivity analysis didn't help. A practitioner's critique of the Stern review, *GLOBAL ENVIRONMENTAL CHANGE*, **20**, 298-302.



# The case of Stern's Review – Technical Annex to postscript



William Nordhaus,  
University of Yale



Nicholas Stern, London  
School of Economics

Stern, N., Stern Review on the Economics of Climate Change. UK Government Economic Service, London, [www.sternreview.org.uk](http://www.sternreview.org.uk).

Nordhaus W., Critical Assumptions in the Stern Review on Climate Change, SCIENCE, 317, 201-202, (2007).

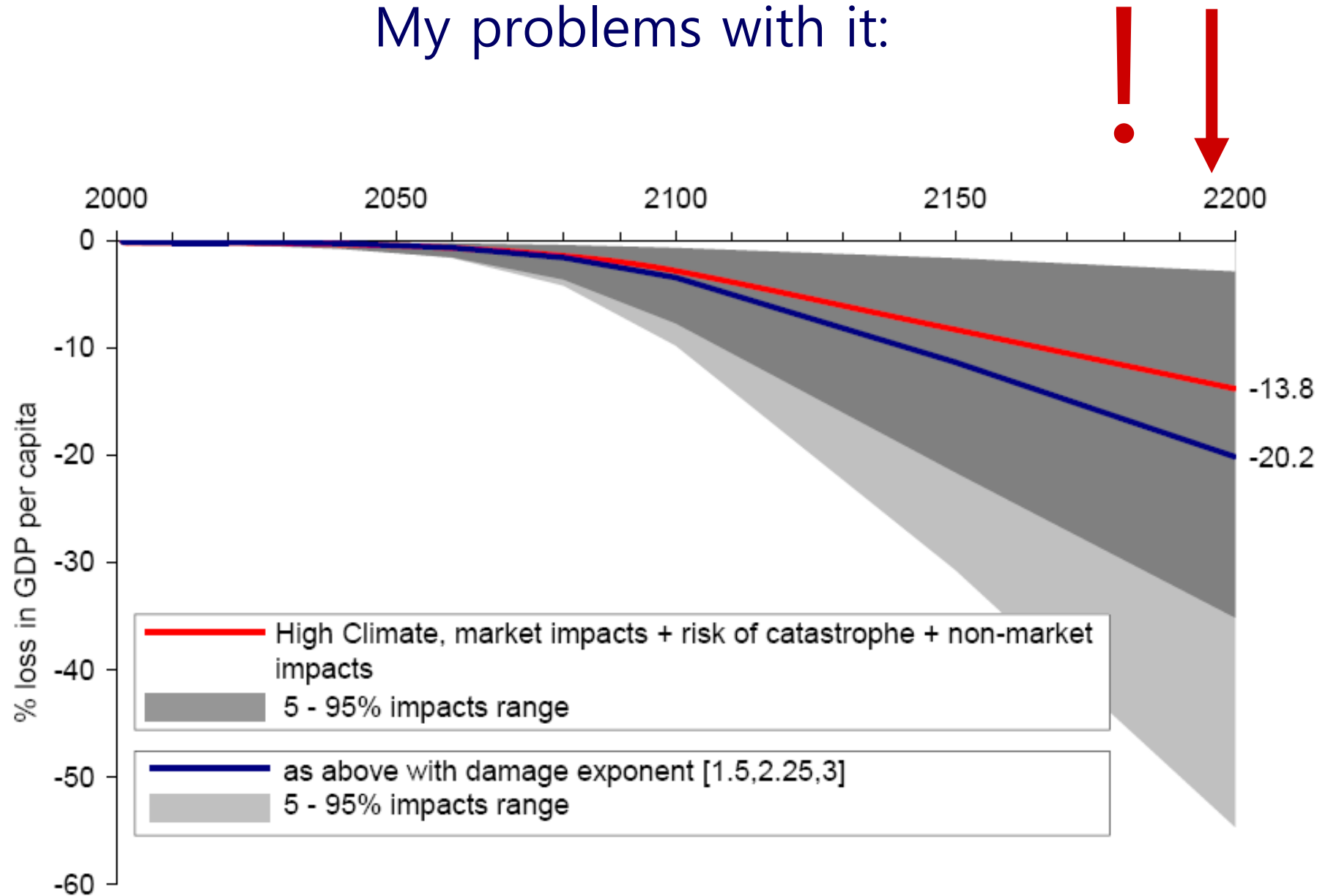
Stern's Review – Technical Annex to postscript (a sensitivity analysis of a cost benefit analysis)

The Stern - Nordhaus exchange on *SCIENCE*

Nordhaus → falsifies Stern based on 'wrong' range of discount rate (~ you GIGOing)

Stern → 'My analysis shows robustness'

My problems with it:

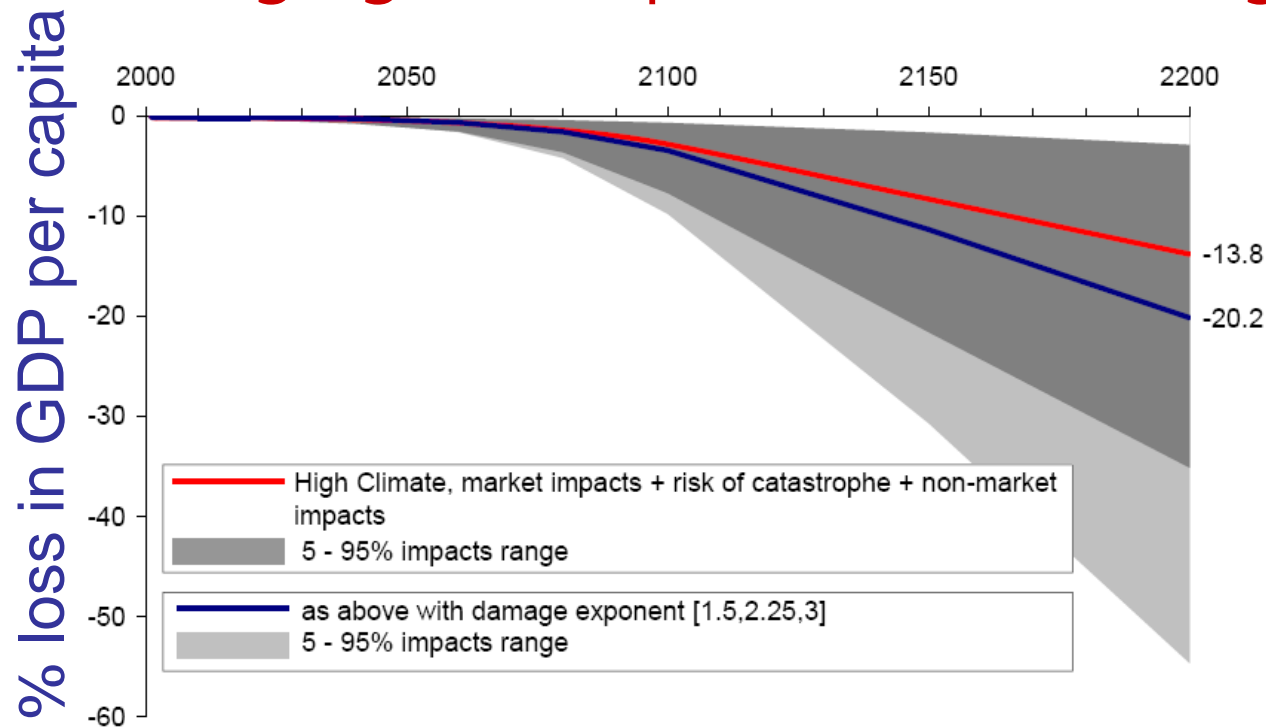


... but foremost Stern says:

changing assumptions → important effect

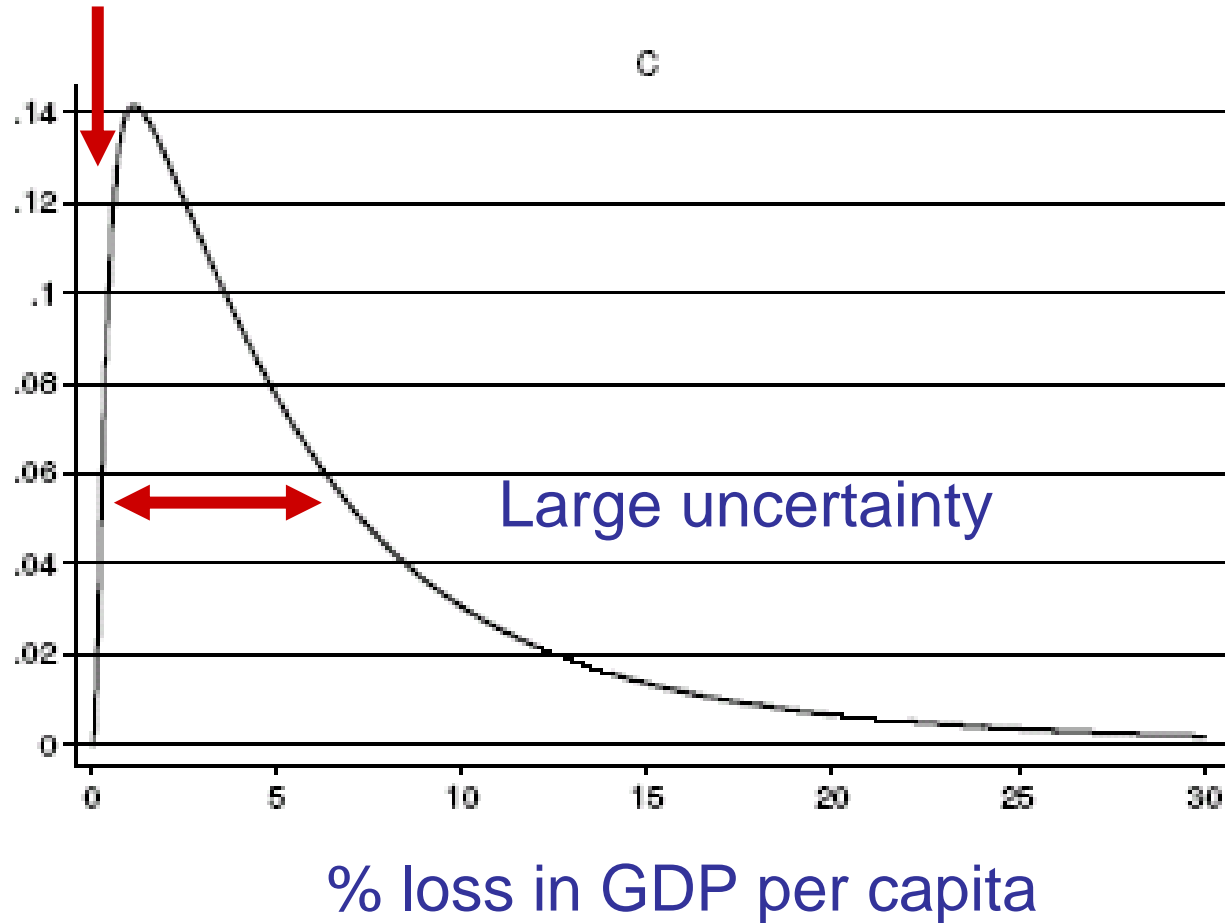
when instead he should admit that:

changing assumptions → all changes a lot



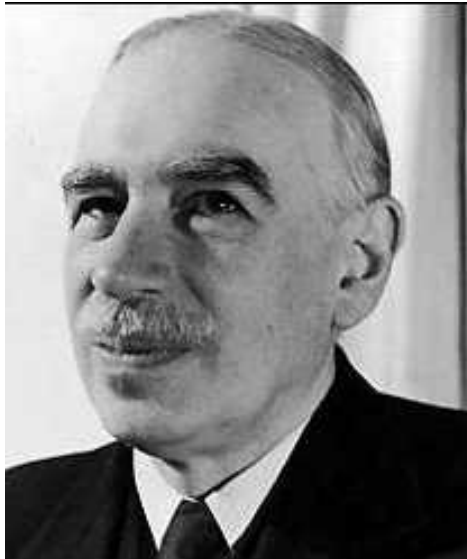
## How was it done? A reverse engineering of the analysis

Missing points



Same criticism applies to Nordhaus – both authors  
frame the debate around numbers which are ...

... precisely wrong



# Comments on the Stern Review's Economics of Climate Change\*

Sir Partha Dasgupta, FBA FRS  
Frank Ramsey Professor of Economics  
University of Cambridge

November 11, 2006  
(Revised: December 12, 2006)

About discount factors. A critique of the way  
delta (intergenerational) and eta (aversion to  
inequality) are set in the review;

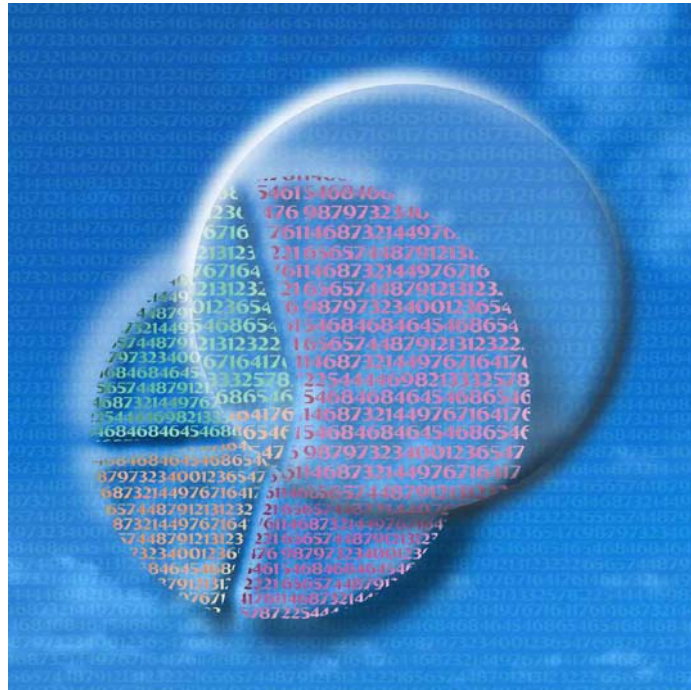
*“[...] to suppose that eta is 1 is also to  
suppose that starvation isn't all that painful!”*

## Comments on the Stern Review's Economics of Climate Change\*

Sir Partha Dasgupta, FBA FRS  
Frank Ramsey Professor of Economics  
University of Cambridge

*But the conclusion I have reached is that the strong, immediate action on climate change advocated by the authors is an implication of their views on intergenerational equity; it isn't driven so much by the new climatic facts the authors have stressed.*





Are statistical  
practices for SA  
taken up?

**From:** Saltelli, A., Annoni P., 2010  
How to avoid a perfunctory sensitivity analysis,  
*Environmental Modeling and Software*, **25**, 1508-  
1517.

What do these have in common?

J. Campbell, *et al.*, *Science* **322**, 1085 (2008).

R. Bailis, M. Ezzati, D. Kammen, *Science* **308**, 98 (2005).

E. Stites, P. Trampont, Z. Ma, K. Ravichandran, *Science* **318**, 463 (2007).

J. Murphy, *et al.*, *Nature* **430**, 768-772 (2004).

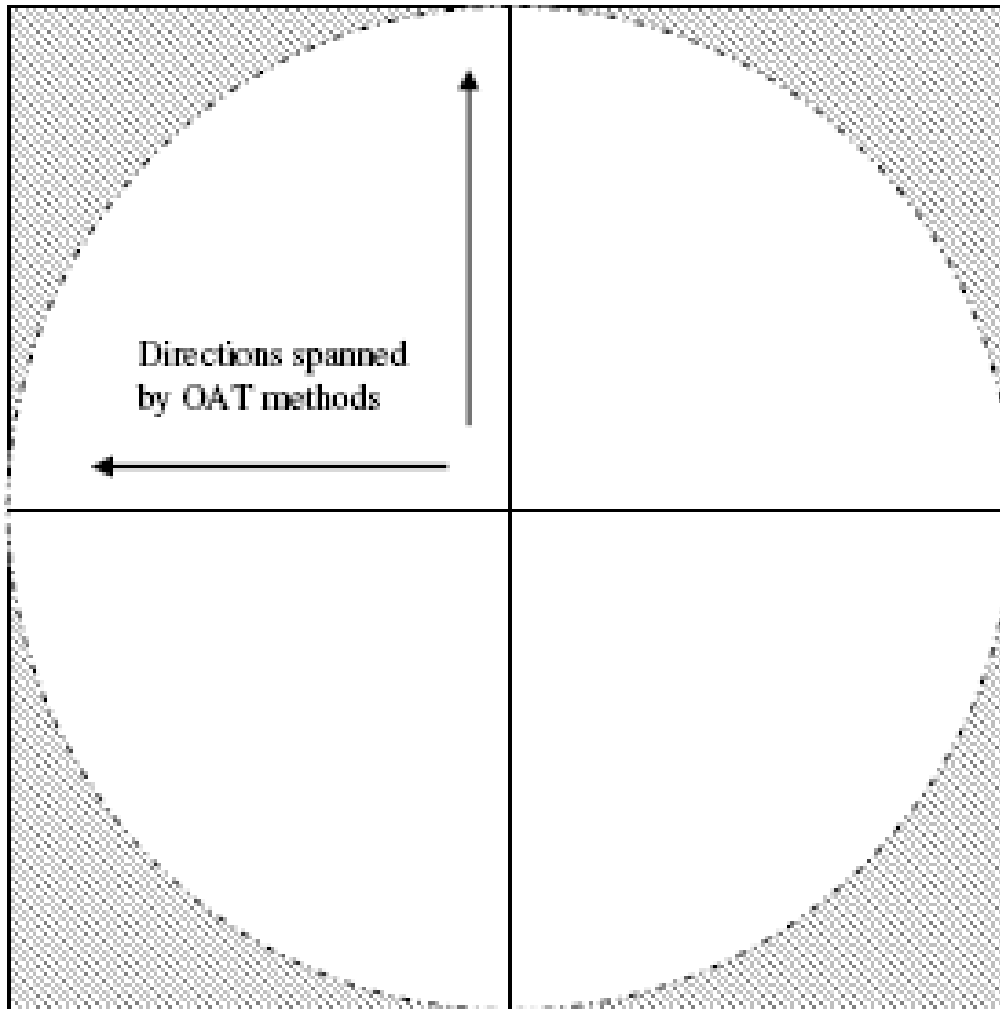
J. Coggan, *et al.*, *Science* **309**, 446 (2005).

OAT

# Why not just changing one factor at a time (OAT)?

•**OMB A4:** Use a numerical sensitivity analysis to examine how the results of your analysis vary with plausible changes in assumptions, choices of input data, and alternative analytical approaches. Sensitivity analysis is especially valuable when the information is lacking to carry out a formal probabilistic simulation. Sensitivity analysis can be used to find 'switch points' -- critical parameter values at which estimated net benefits change sign or the low cost alternative switches. Sensitivity analysis usually proceeds by changing one variable or assumption at a time, but it can also be done by varying a combination of variables simultaneously to learn more about the robustness of your results to widespread changes. Again, however, major rules above the \$1 billion annual threshold require a formal treatment.

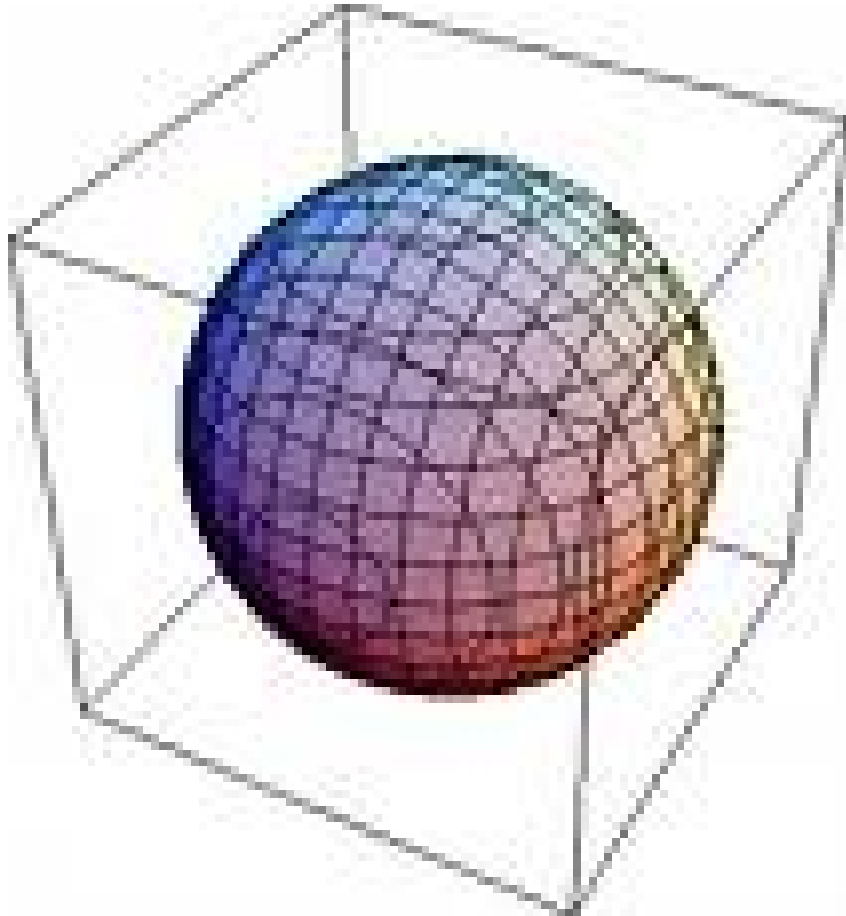
# OAT in 2 dimensions



Area circle / area  
square = ?

~ 3/4

# OAT in 3 dimensions



Volume sphere /  
volume cube =?

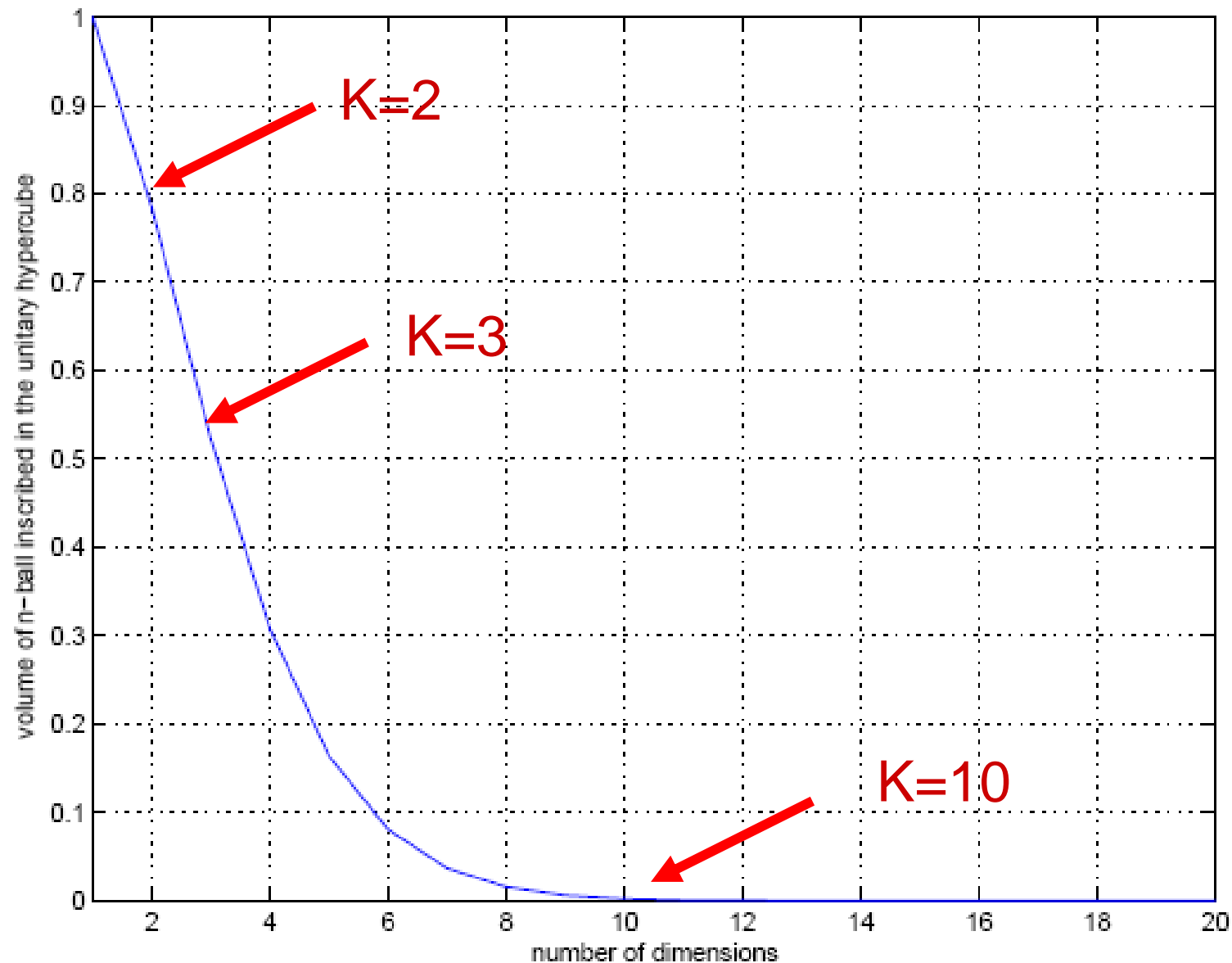
$\sim 1/2$

# OAT in 10 dimensions

Volume hypersphere / volume  
ten dimensional hypercube  $\sim 0.0025$



# OAT in $k$ dimensions





# An environmental case study

Describe a chain of species mutating one into another without backward reactions



## An environmental case study

The Bateman equations describe the concentrations  $N_i$  of  $k$  species in linear chain governed by rate constants  $\lambda_i$ :

$$\begin{aligned}\frac{dN_1}{dt} &= -\lambda_1 N_1 \\ \frac{dN_i}{dt} &= \lambda_{i-1} N_{i-1} - \lambda_i N_i \quad (i = 2, \dots, k)\end{aligned}$$

if  $N_1 \neq 0$  and  $N_i = 0 \forall i > 1$

We want to simulate  $N_k(t)$  with different  $k$

$$\begin{aligned}N_k(t) &= \frac{N_1(0)}{\lambda_k} \sum_{i=1}^k \lambda_i \alpha_i \exp(-\lambda_i t) \\ \alpha_i &= \prod_{j=i+1}^k \frac{\lambda_j}{(\lambda_j - \lambda_i)}\end{aligned}$$

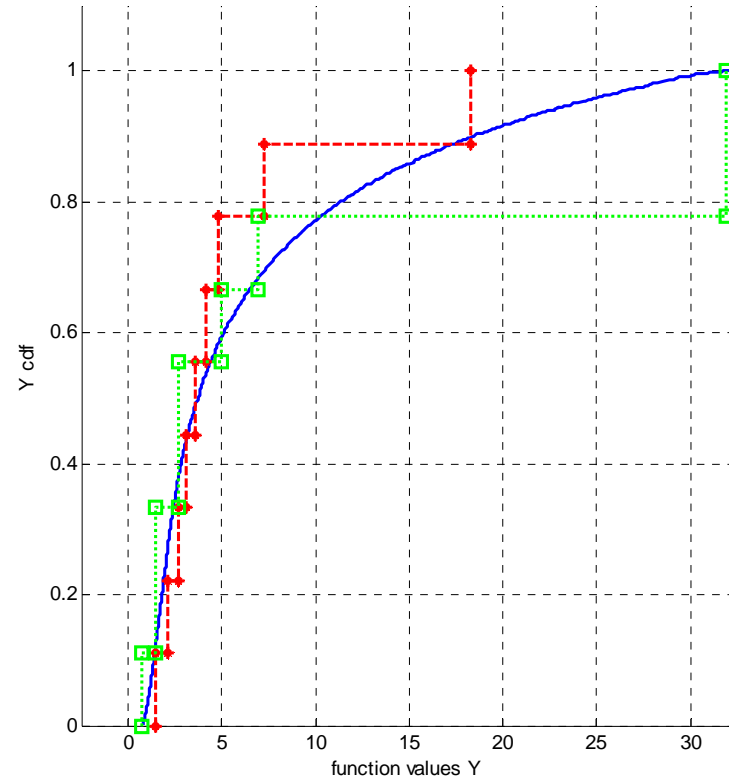
## An environmental case study

### Our settings:

- Six experiments with increasing number of species  $k$  involved (that is number of parameters)
- Fixed time instant  $t$
- $\lambda_i$  randomly sampled from an uniform  $U[1,100]$
- Concentration of initial species  $N_1(0)=100$
- Comparison between OAT and Elementary Effect method with (roughly) the same number of runs

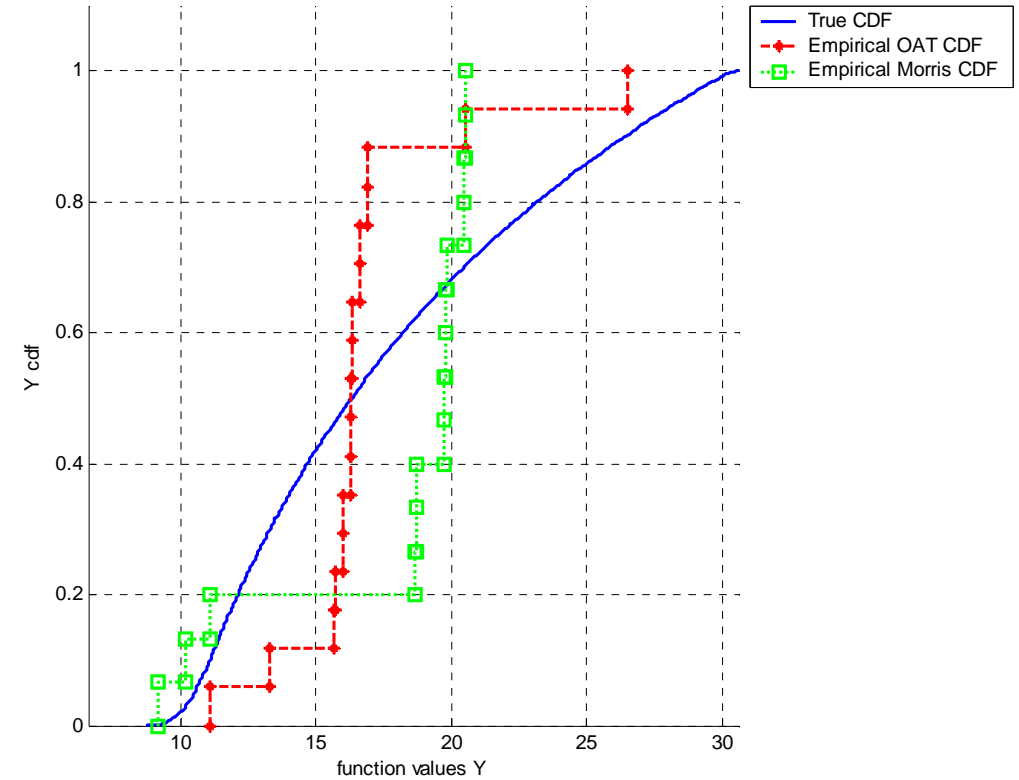
## 2 species

Experiment with 2 factors - # of OAT runs: 9 # of Morris runs: 9 Bateman equati



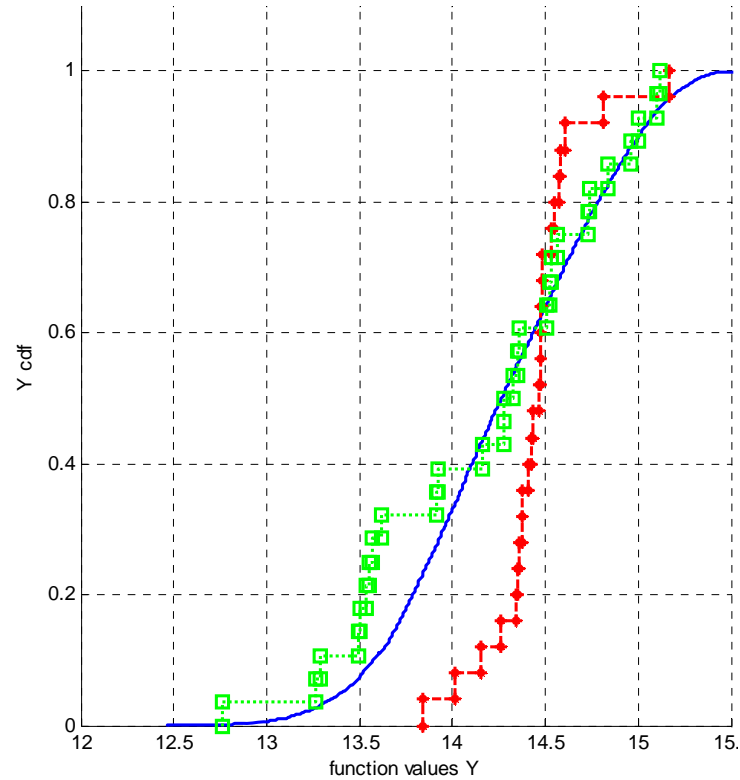
## 4 species

Experiment with 4 factors - # of OAT runs: 17 # of Morris runs: 15 Bateman equations



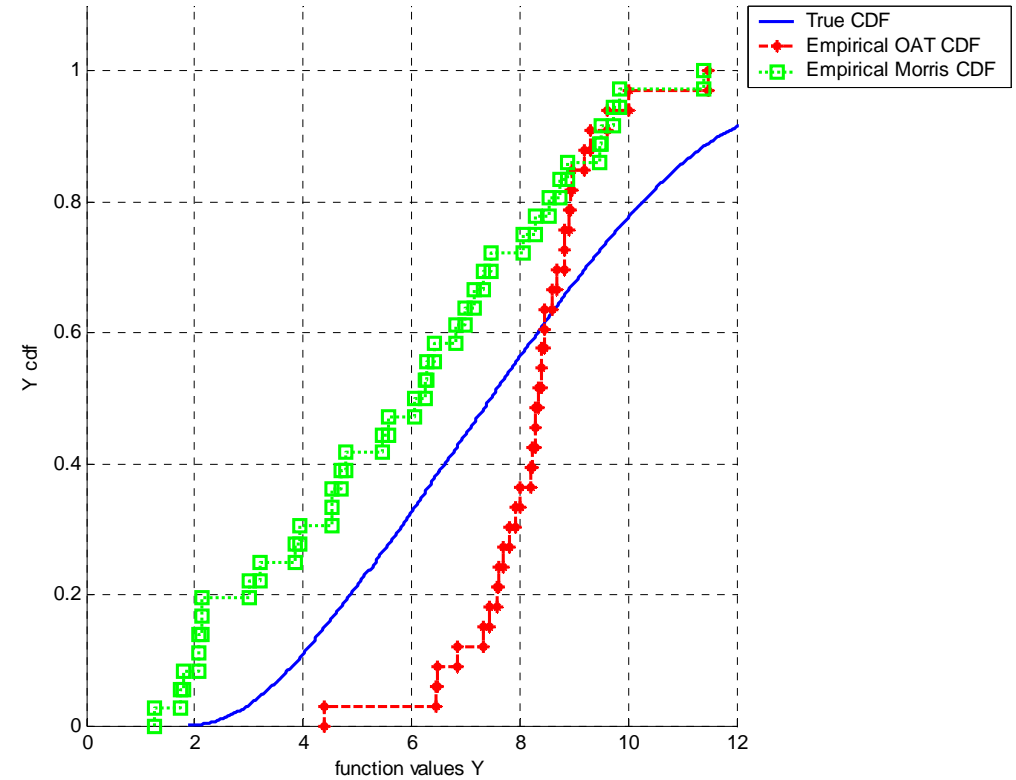
## 6 species

Experiment with 6 factors - # of OAT runs: 25 # of Morris runs: 28 Bateman eq



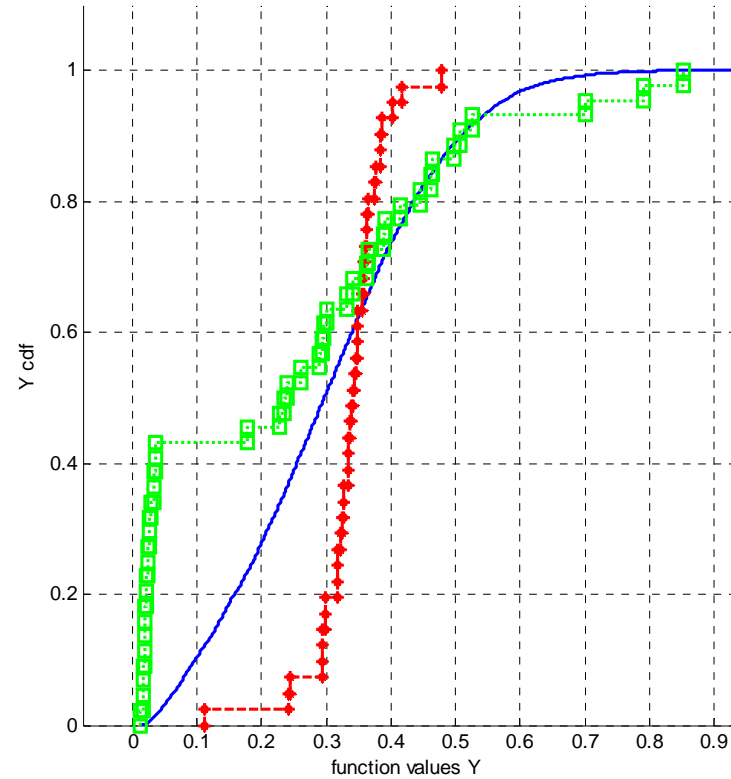
## 8 species

Experiment with 8 factors - # of OAT runs: 33 # of Morris runs: 36 Bateman equations



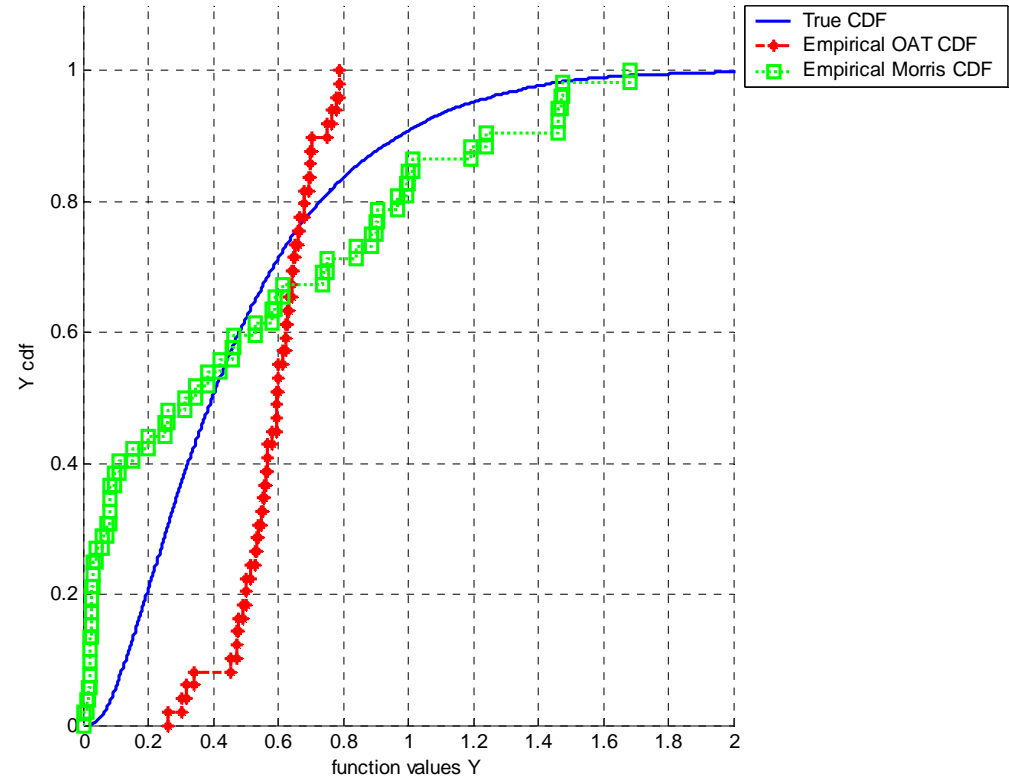
## 10 species

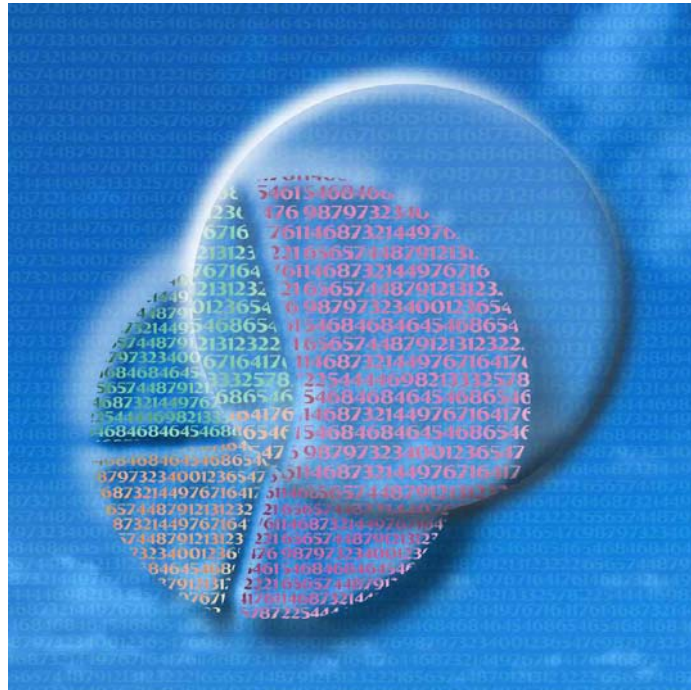
Experiment with 10 factors - # of OAT runs: 41 # of Morris runs: 44 Bateman  $\epsilon$



## 12 species

Experiment with 12 factors - # of OAT runs: 49 # of Morris runs: 52 Bateman equations





Can something  
be done to  
ease adoption?

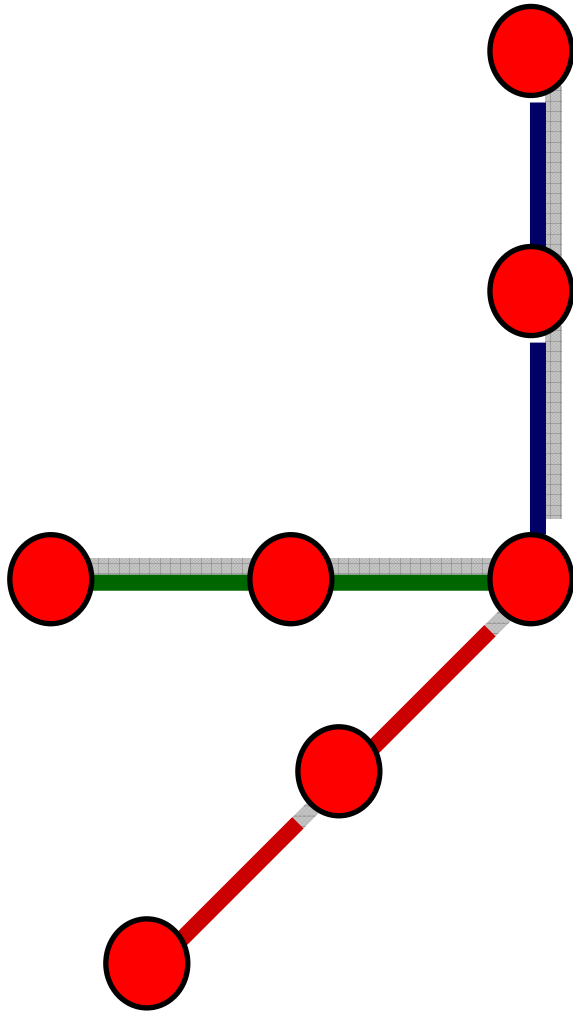
**From:**

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications*, **181**, 259-270.

**and**

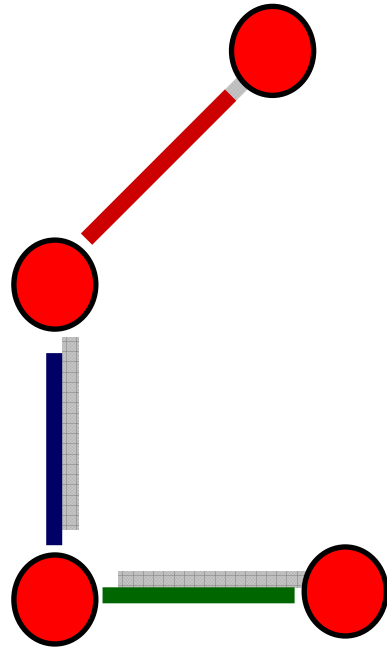
Campolongo F, Saltelli A, Cariboni, J, 2010, From screening to quantitative sensitivity analysis. A unified approach, Submitted to *Computer Physics Communication*.





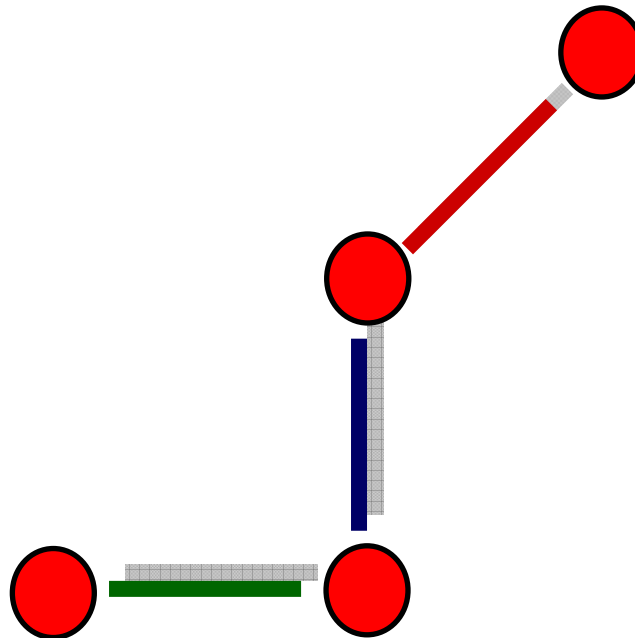
In 3 dimensions, OAT,  
7 points

This is what is done

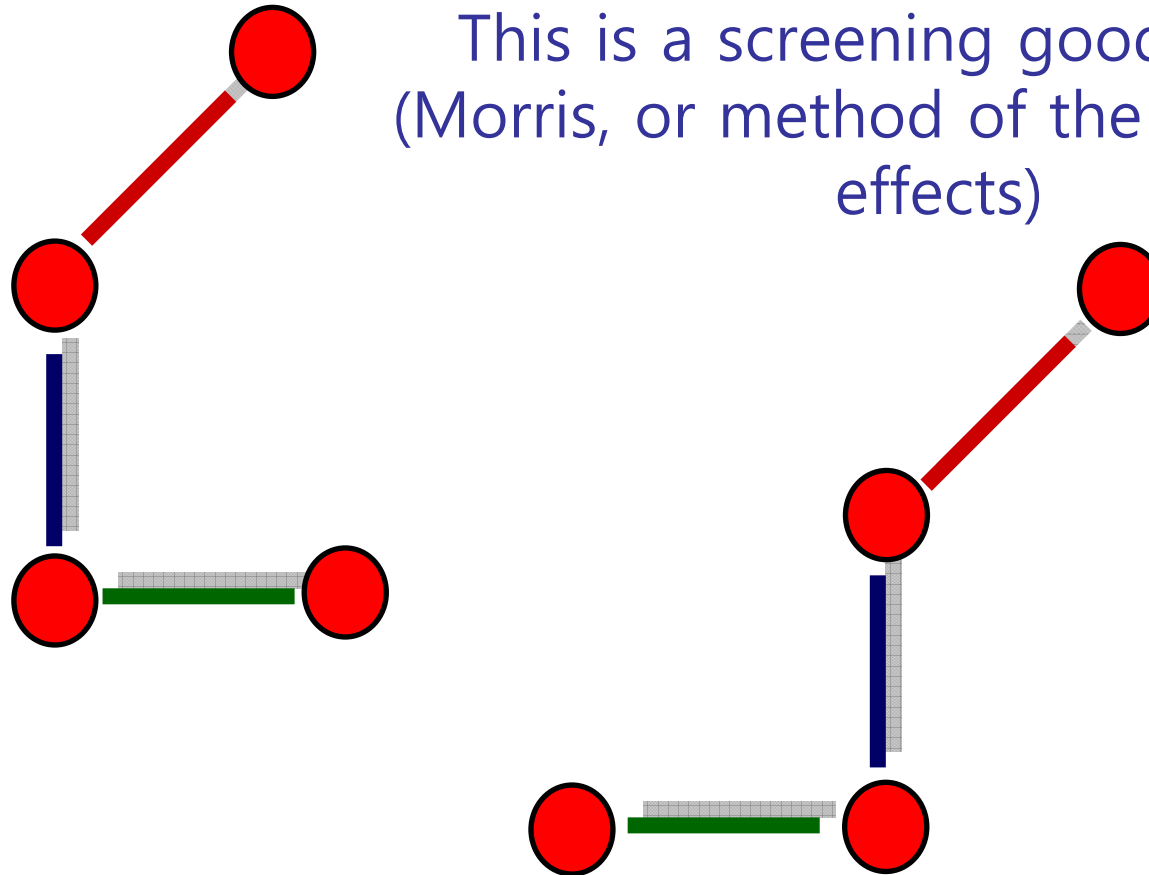


In 3 dimension, 8 screening  
points in a trajectory  
arrangement

This is what could be done

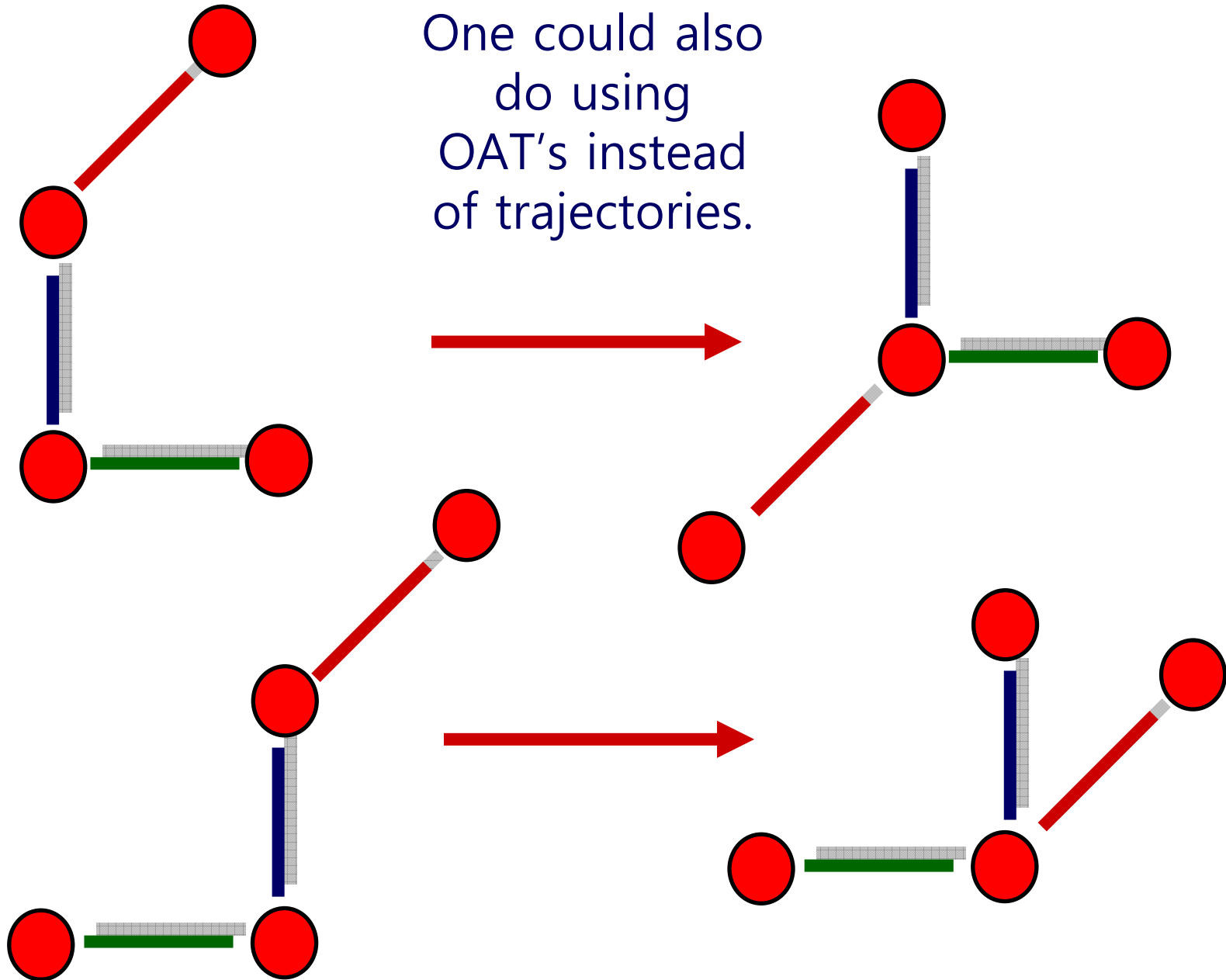


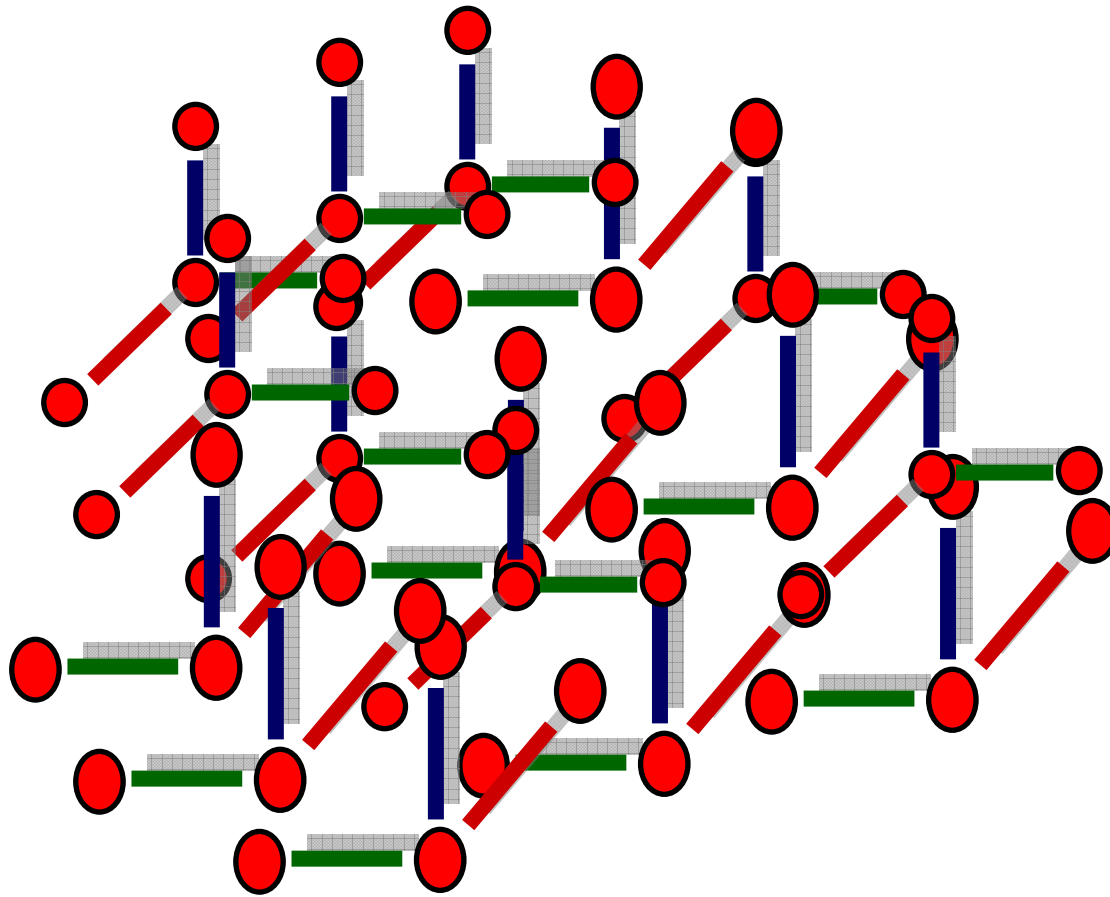
This is a screening good practice  
(Morris, or method of the elementary  
effects)



**See:** Campolongo, F., Cariboni, J., and Saltelli, A., 2007, An effective screening design for sensitivity analysis of large models, *Environmental Modelling and Software*, **22**,1509-1518.

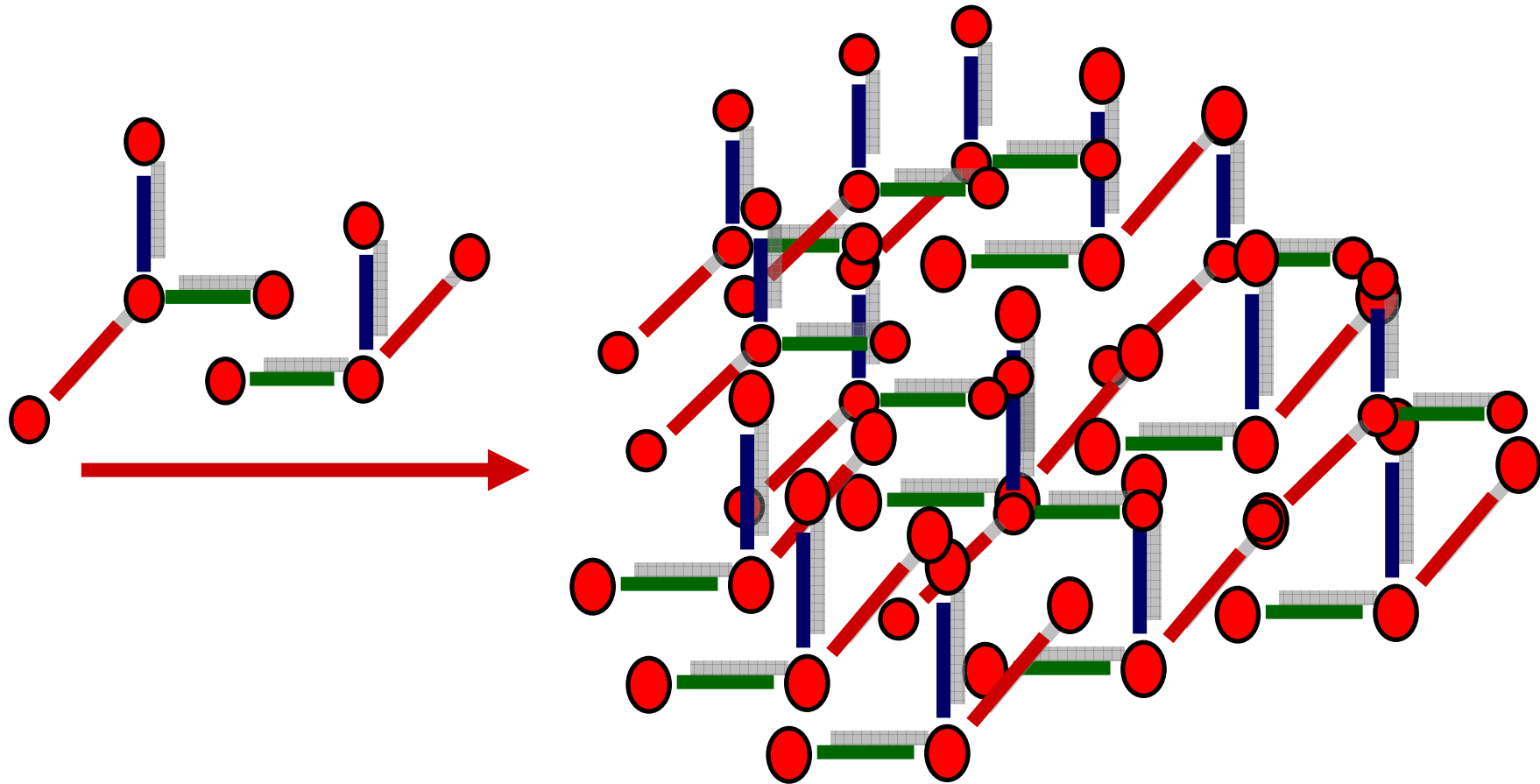
One could also  
do using  
OAT's instead  
of trajectories.





Increasing the  
number of OAT's the  
test becomes  
quantitative...

...because this design  
is the same used for  
the total sensitivity  
index ST (see next!)



Thus one can start EE-wise (few points) and continue variance-based, without discarding points, by just changing the estimator (from that for EE to that for ST).

How does JRC deploy it?  
E.g. in impact assessment

# Sensitivity analysis of QUEST III

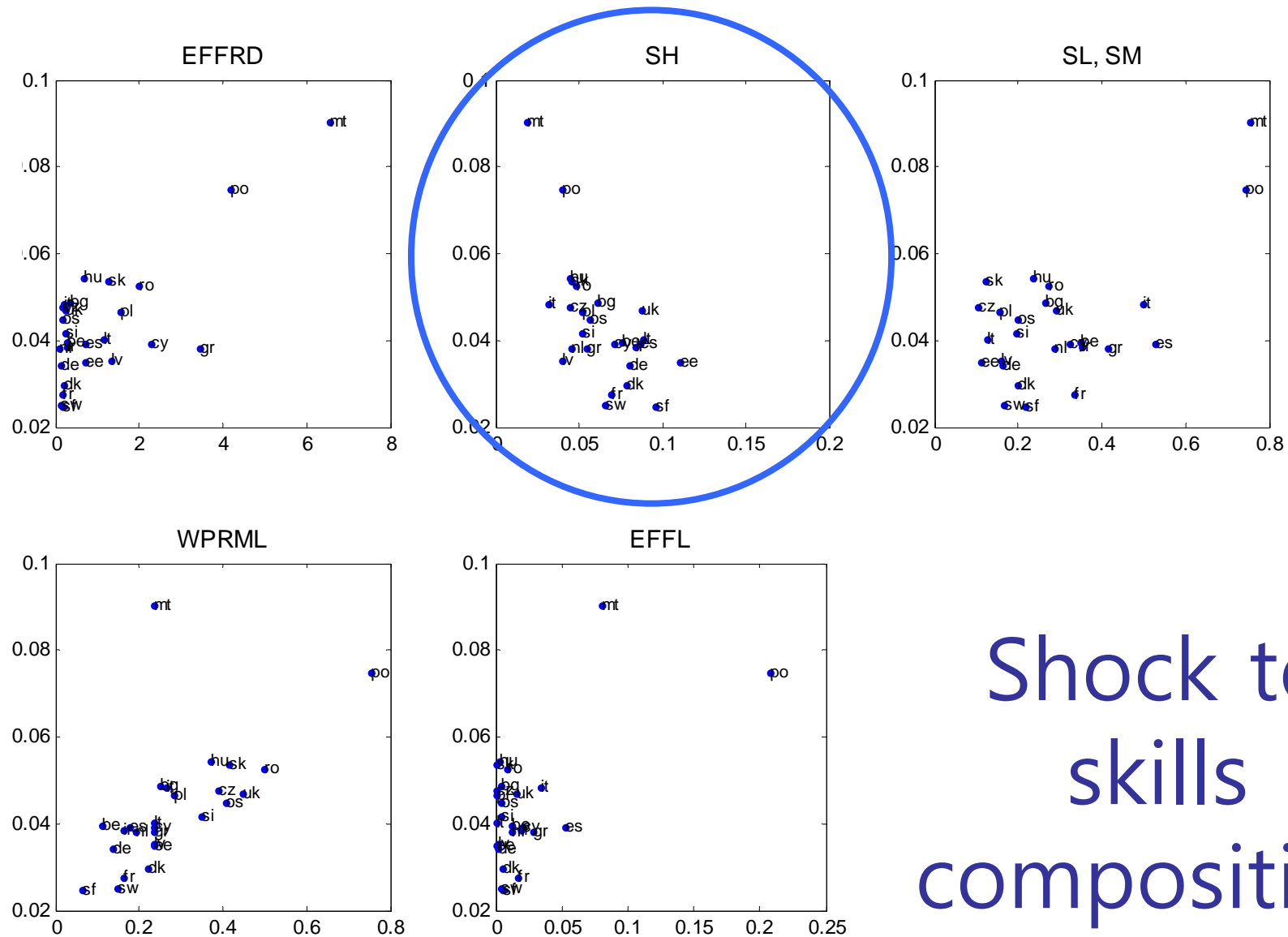
Structural reforms EU 2020

Simulation of QUEST III under the 27 different parameterizations (the EU countries)



# With ECFIN: Model behaviour across EU countries

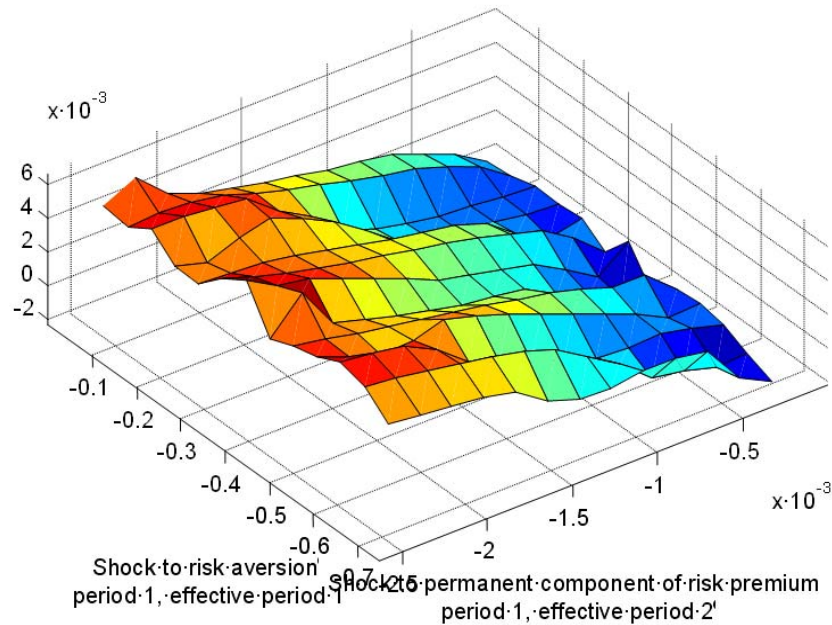
GDP



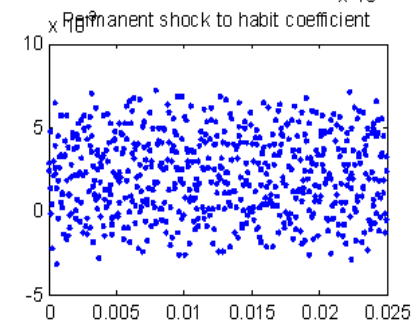
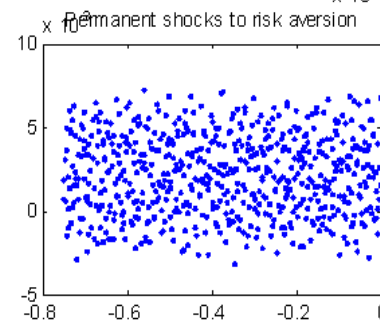
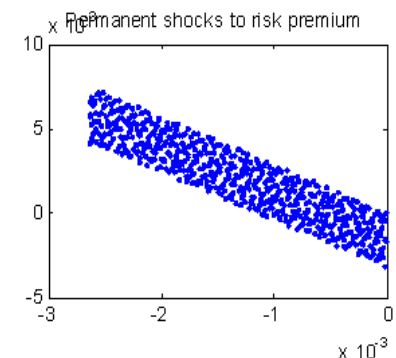
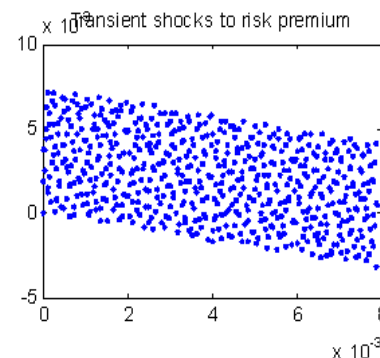
Shock to  
skills  
composition

# With MARKT: Analysis of Solvency II impact assessment

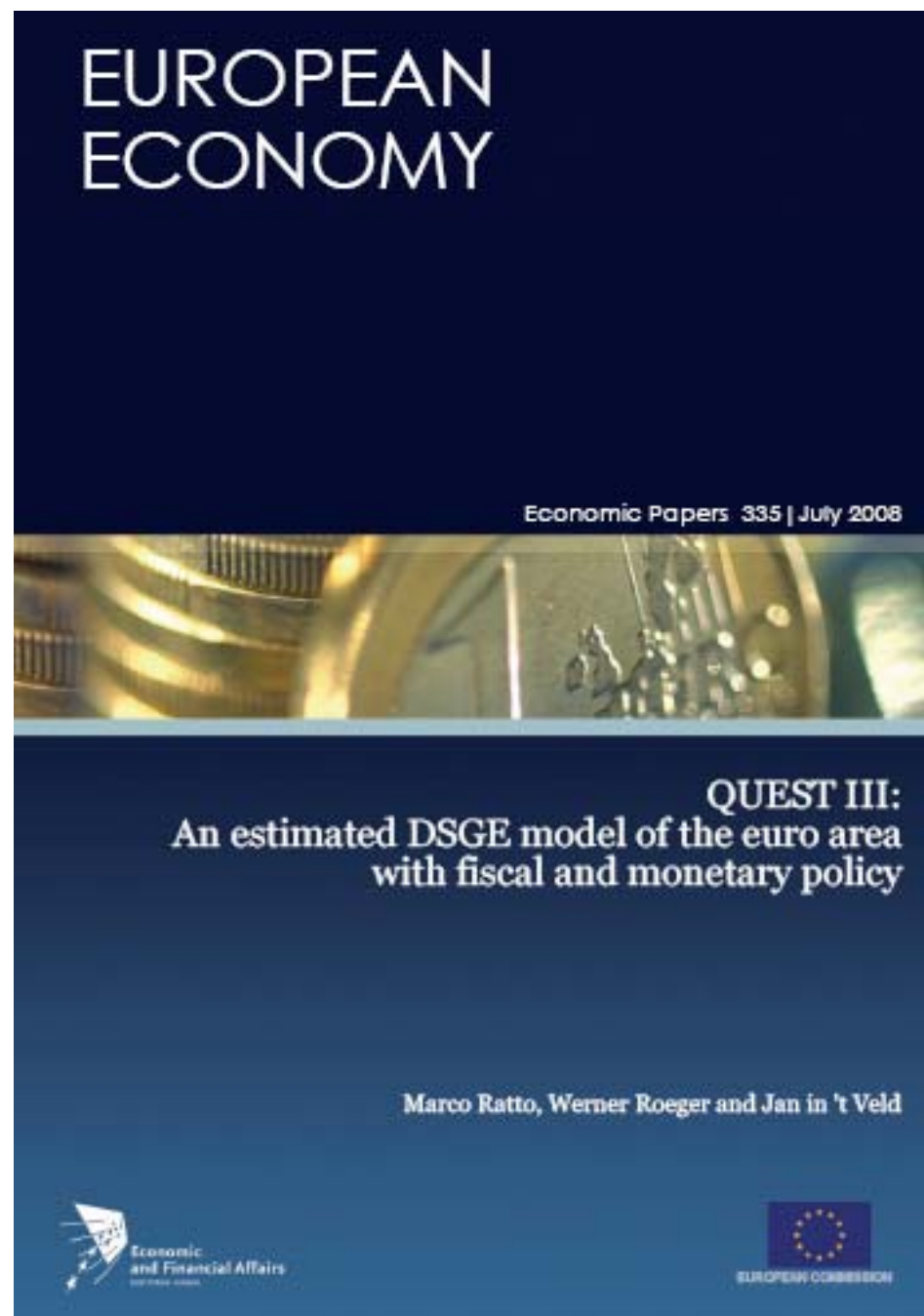
## Response to individual shocks



Distributions of response to combinations of shocks

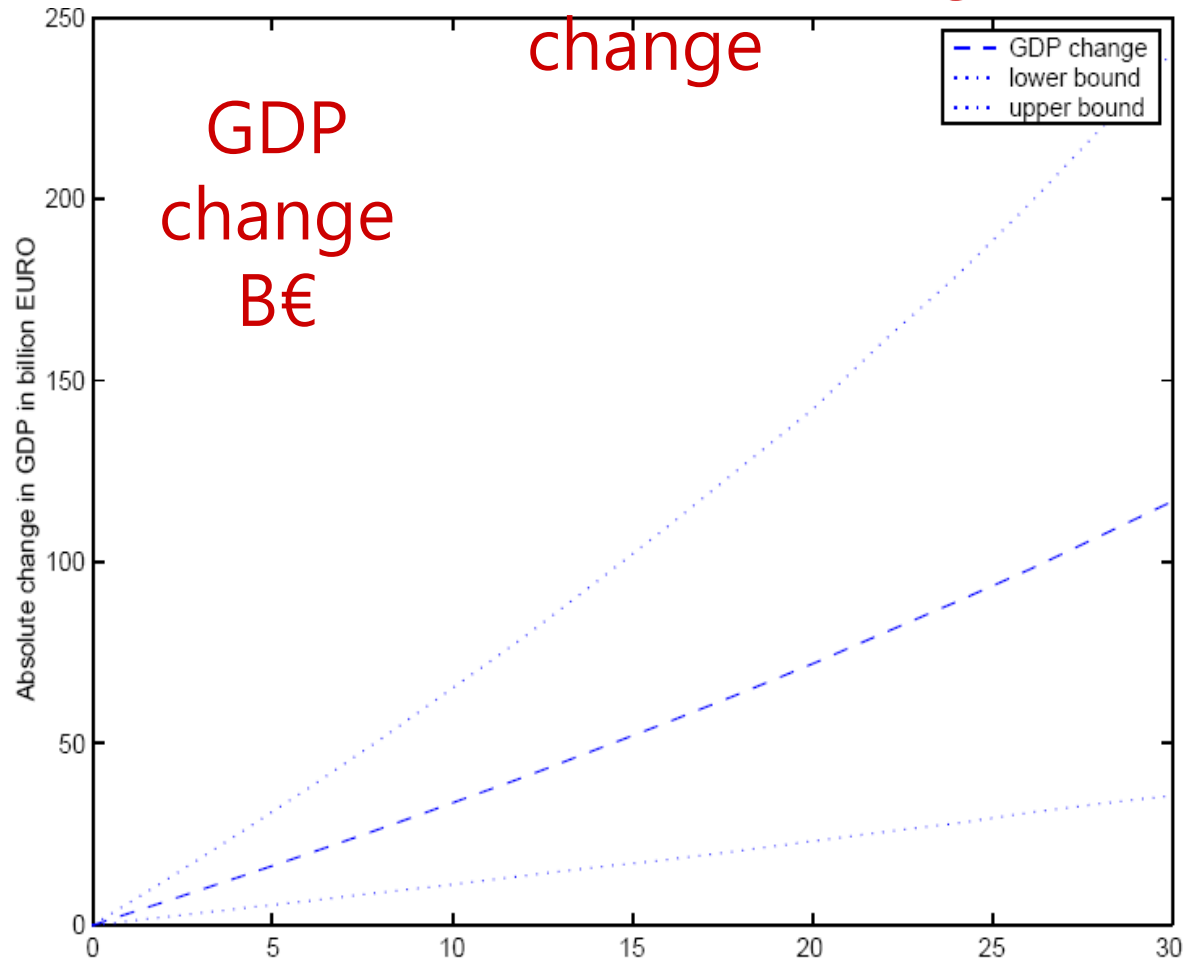


More about this from  
Marco Ratto (JRC) and  
Jan in 't Veld (ECFIN),  
modelling seminar



# Clearing & Settlement impact assessment

## Uncertain relation between trading cost and GDP



Percentage reduction in trading cost

# The Global Competitiveness Report 2010–2011



Klaus Schwab, World Economic Forum

## Appendix B: The Joint Research Centre assessment of the Global Competitiveness Index

MICHELA NARDO, European Commission Joint Research Centre  
PAOLA ANNONI, European Commission Joint Research Centre

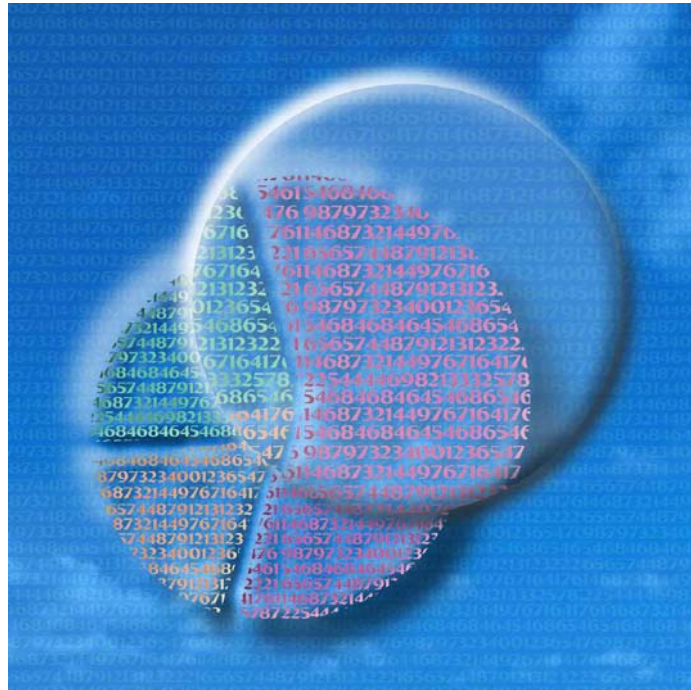
Attempting to summarize complex concepts such as competitiveness in a single metric or index raises a number of empirical challenges. These include data quality, indicator selection, indicator importance, correlation assessment and so on. It does not take

## Robustness assessment (=sensitivity analysis) of indices

### Page 10:

#### “Adjustments to the GCI

Over the past year, the Global Competitiveness Index has been put through a rigorous analysis by the Joint Research Centre of the European Commission (JRC).”

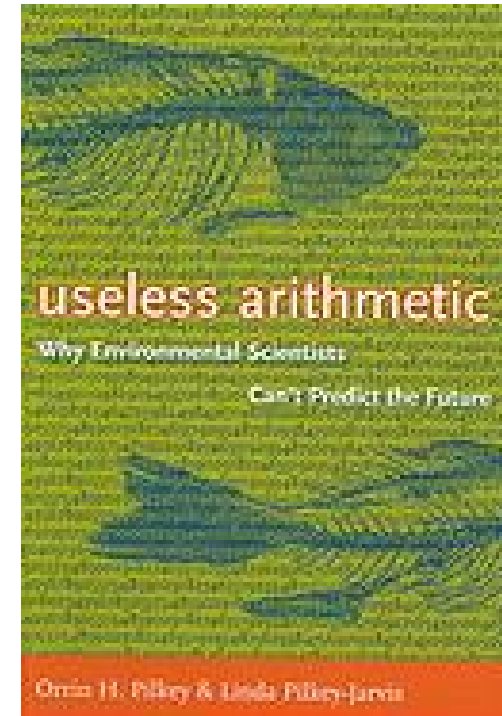


Beware the  
delusion of  
technical fixes!

Good quality in the tools are a  
necessary but not a sufficient  
condition for a good quality  
assessment

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.

[...] 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."





# REFERENCES

**EPA guidelines:**

<http://www.epa.gov/crem/library/cred%20guidance%200309.pdf>,  
p.69-76, accessed August 06, 2009.

**OMB:**

Office of Management and Budget, Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Federal Register / Vol. 67, No. 36 / Friday, February 22, 2002 / Notices, p. 8456,  
<http://www.whitehouse.gov/omb/fedreg/reproducible2.pdf>, accessed August 5, 2009.

**OMB:**

Office of Management and Budget's (OMB) Office of Information and Regulatory Affairs (OIRA), January 9, 2006, Proposed Risk Assessment Bulletin,  
<http://www.whitehouse.gov/omb/inforeg/proposed%20risk%20assessment%20bulletin%20010906.pdf>, p. 16-17, accessed August 5, 2009.

## REFERENCES

Baudrillard, J. (1999) *Revenge of the Crystal: Selected Writings on the Modern Object and Its Destiny, 1968-83*, Pluto Classics, p.92.

Campolongo, F., J. Cariboni, and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software* 22, 1509-1518.

Funtowicz, S.O., and Ravetz, J.R. (1990) *Uncertainty and Quality in Science for Policy*, Springer, Dordrecht, p.54.

Funtowicz, S.O., and Ravetz, J.R. Science for the Post Normal age, 1993, *Futures*, **25**, 739–755.

Homma, T. and A. Saltelli (1996), Importance measures in global sensitivity analysis of model output, *Reliability Engineering and System Safety*, 52(1), 1-17.

Kennedy, P. (2007) *A guide to econometrics*, Fifth edition, Blackwell Publishing, p.396.

Leamer, E., Let's take the con out of econometrics, and Sensitivity analysis would help. In C. Granger (ed.) (1990) *Modelling Economic Series*, Clarendon Press, Oxford, or:  
Leamer, E., 1990, Let's Take the Con Out of Econometrics, *American Economics Review*, **73** (March 1983), 31-43.

## REFERENCES

- Mac Lane, S. (1988) IIASA's Credibility, SCIENCE, Letters, 241, 1144.
- Mac Lane, S. (1988) IIASA and Modeling, SCIENCE, Letters, 242, 1623-1624.
- Michaels, D., Doubt is their product, Scientific American, Jun2005, 292, Issue 6
- Morris, M. (1991), Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2), 161-174.
- Nordhaus W., (2007) Critical Assumptions in the Stern Review on Climate Change, SCIENCE, 317, 201-202.
- Pilkey, O. H. and L. Pilkey-Jarvis, (2007) Useless Arithmetic. Why Environmental Scientists Can't Predict the Future, Columbia University Press, New York.

## REFERENCES

Rosen, R. (1991) Life Itself - a Comprehensive Inquiry into Nature, Origin, and Fabrication of Life, Columbia University Press, p. 49-55.

Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn, Design and Analysis of Computer Experiments, Statistical Science Volume 4, Number 4 (1989), 409-423.

Saisana M., Saltelli A., Tarantola S. (2005) Uncertainty and Sensitivity analysis techniques as tools for the quality assessment of composite indicators, Journal of the Royal Statistical Society-A, 168 (2), 307-323.

Saltelli, A., Tarantola, S. Campolongo, F., 2000, Sensitivity analysis as an ingredient of modelling, Statistical Science, 15 (4), 377-395.

Saltelli A. Tarantola S., 2002, On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal, Journal of American Statistical Association, 97 (459), 702-709.

## REFERENCES

- Saltelli, A. (2002), Making best use of model valuations to compute sensitivity indices. *Computer Physics Communications*, 145, 280-297.
- Saltelli, A., M. Ratto, S. Tarantola and F. Campolongo (2005) Sensitivity Analysis for Chemical Models, *Chemical Reviews*, 105 (7), 2811 - 2828.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications*, 181, 259-270.
- Saltelli, A., D'Hombres, B., 2010, Sensitivity analysis didn't help. A practitioner's critique of the Stern review, *Journal of Global Environmental Change*, 25, 1508-1517.
- Saltelli, A., Annoni Paola, 2010, How to avoid a perfunctory sensitivity analysis, *Environmental Modeling and Software*, 25 1508-1517.
- Sarewitz, D., Liberating Science from Politics, 2006, *American Scientist*, 94 (2), 104-105.
- Stern, N., (2006) Stern Review on the Economics of Climate Change. UK Government Economic Service, London, [www.sternreview.org.uk](http://www.sternreview.org.uk). Technical Annex to the Postscript, available at [http://www.hm-treasury.gov.uk/media/1/8/Technical\\_annex\\_to\\_the\\_postscript\\_P1-6.pdf](http://www.hm-treasury.gov.uk/media/1/8/Technical_annex_to_the_postscript_P1-6.pdf)

## REFERENCES

Stern, N., and Taylor C. (2007) Climate Change: Risk, Ethics, and the Stern Review, *SCIENCE*, **317**, 203-204.

Stokstad, E. (2008) Dueling Visions for a Hungry World, *SCIENCE*, **319**, 1474-1476.

Taleb, N. N. (2007) The Black Swan: The Impact of the Highly Improbable, Random House.

van der Sluijs, J.P. (2002) A way out of the credibility crisis around model-use in Integrated Environmental Assessment, *Futures*, **34**, 133-146.

Young, P.C., Parkinson, S. and Lees, M.J., (1996), Simplicity out of complexity: Occam's razor revisited, *Journal of Applied Statistics*, 23, 165-210.

Zidek, J., 2006, Editorial: (post-normal) statistical science, *Journal of the Royal Statistical Society*, 69(1), 1-4.