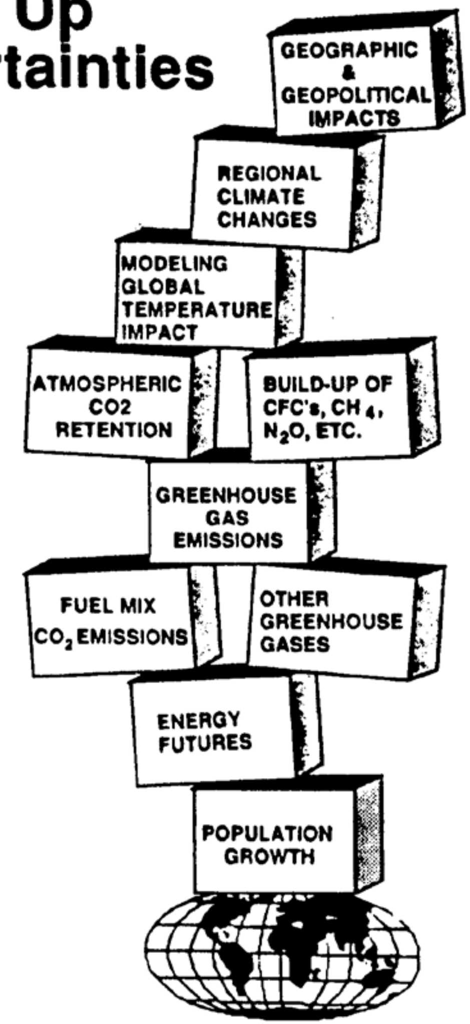
Knowledge Quality Assessment

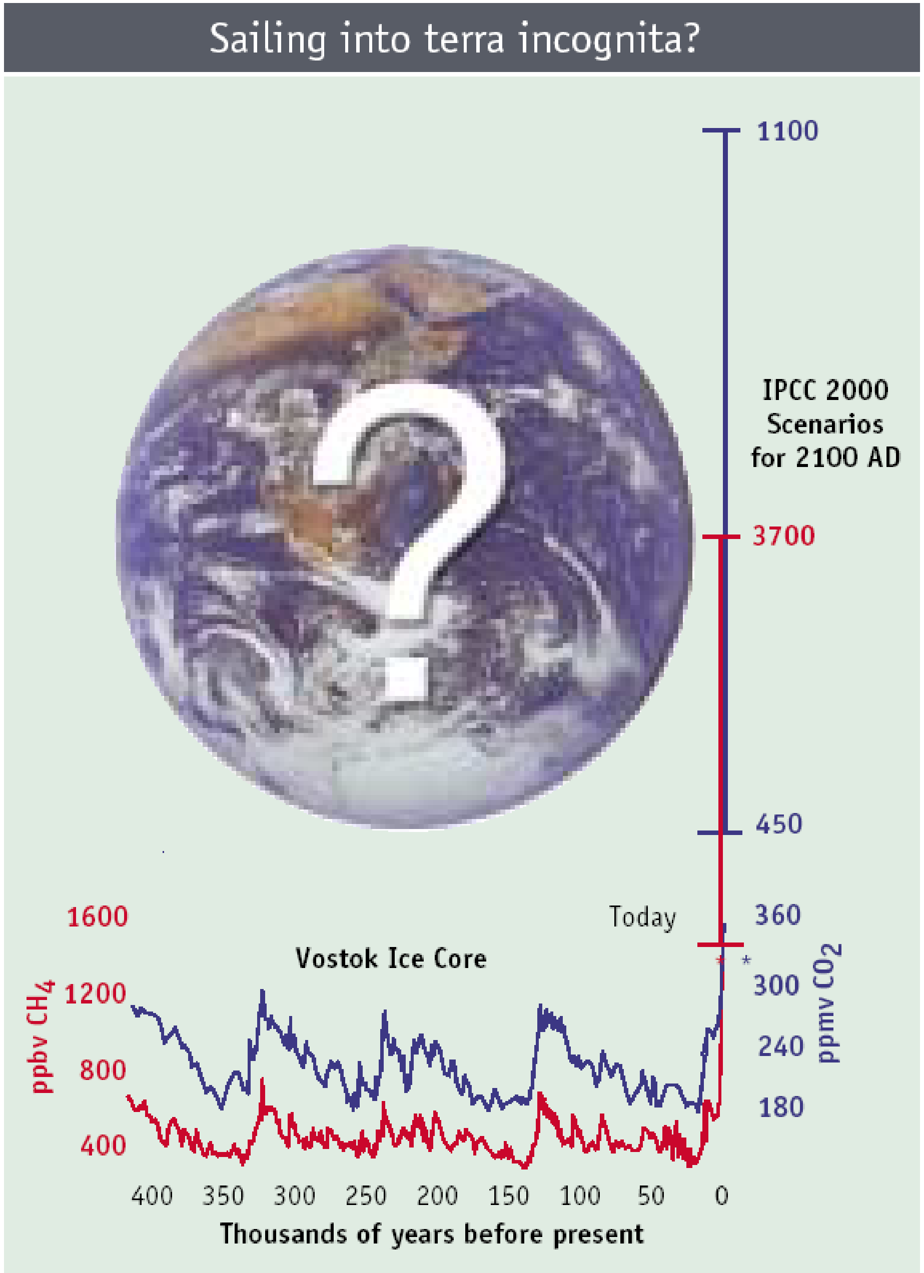
an intro to The Guidance & NUSAP

Jeroen P. van der Sluijs

GLOBAL CLIMATE CHANGE

Piling Up Uncertainties





How does science-policy interface cope with uncertainties

Two strategies dominate:

- **Overselling certainty**
  - to promote political decisions (enforced consensus), or
- **Overemphasising uncertainty**
  - to prevent political action

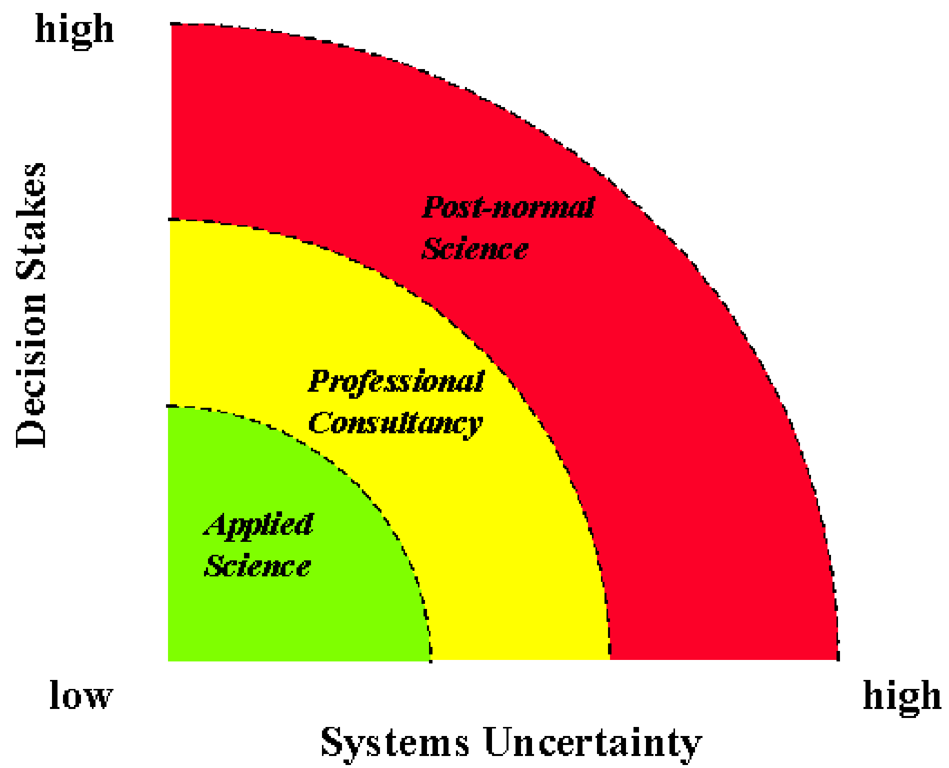
- Both promote decision strategies that are **not fit for meeting the challenges** posed by the uncertainties and complexities faced.

- Need for a third voice next to alarmists and skeptics: coping with uncertainty, scientific dissent & plurality in science for policy.

Complex - *uncertain* - risks

Typical characteristics:

- Decisions urgent
- Stakes high
- Values in dispute
- Irreducible & unquantifiable uncertainty



- Assessment: models, scenarios, assumptions, extrapolations
- (hidden) value loadings in problem frames, indicators chosen, assumptions made

- **Knowledge Quality Assessment!**

(Funtowicz & Ravetz, 1993)
http://www.uu.nl/wetfilos/wetfil10/sprekers/Funtowicz_Ravetz_Futures_1993.pdf

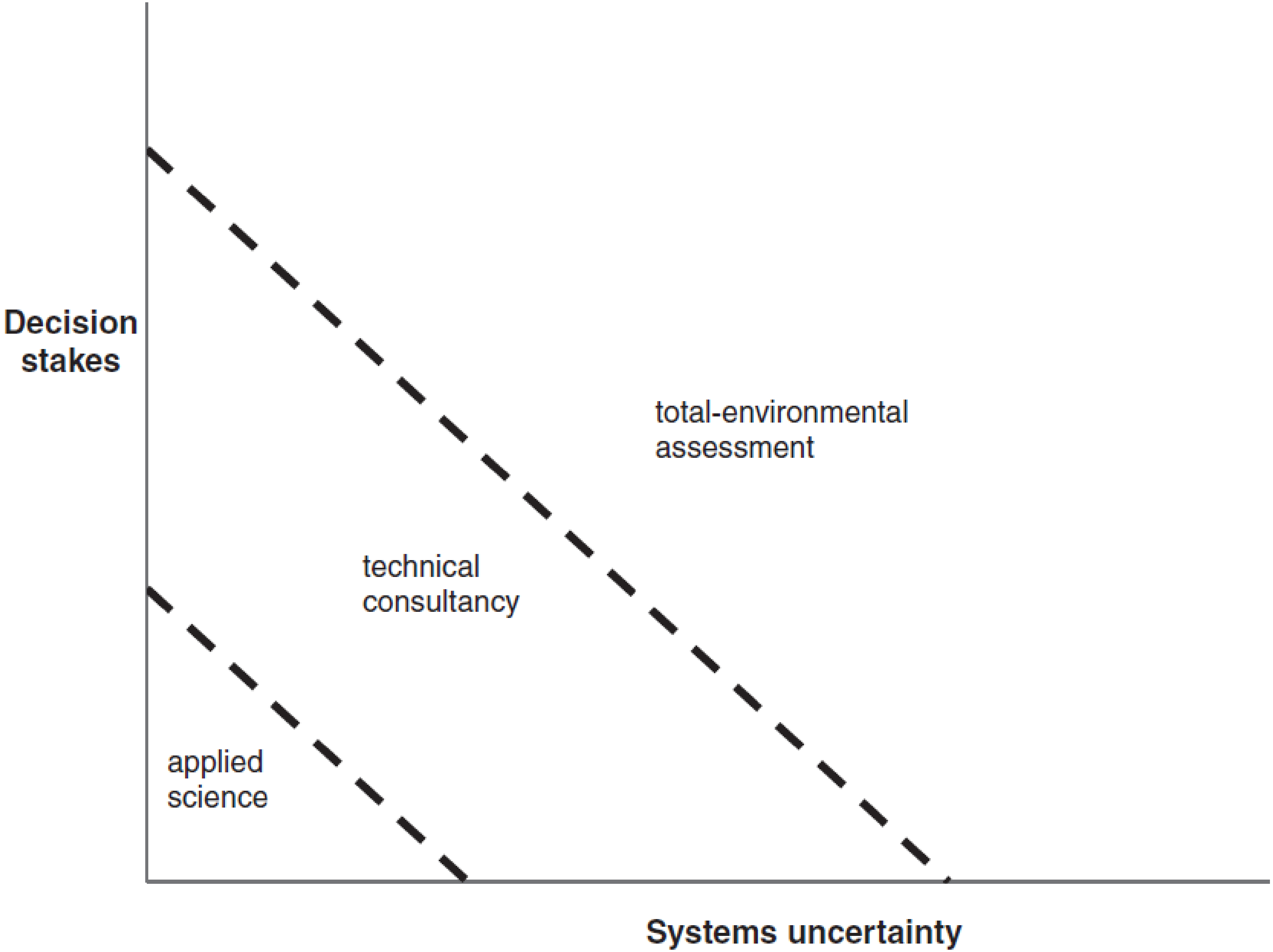


Figure 28.2 The original diagram of three types of risk assessment.
Source: redrawn after Funtowicz & Ravetz (1985).
Note: "Total-environmental assessment" would later be relabelled as "post-normal science".

Illustrative example

Protecting a strategic fresh-water resource under the Water Supply Act Denmark

Case:

- Important aquifer west of Copenhagen
- groundwater abstraction 12 million m^3/year
- Copenhagen County had to prepare an action plan for protection of groundwater against pollution
- Scientist were asked to assess aquifer's vulnerability to pollution in a 175 km^2 area

A practical problem:

Protecting a strategic fresh-water resource

5 scientists addressed same question:

“which parts of this area are most vulnerable to nitrate pollution and need to be protected?”

(Refsgaard, Van der Sluijs et al, 2006)

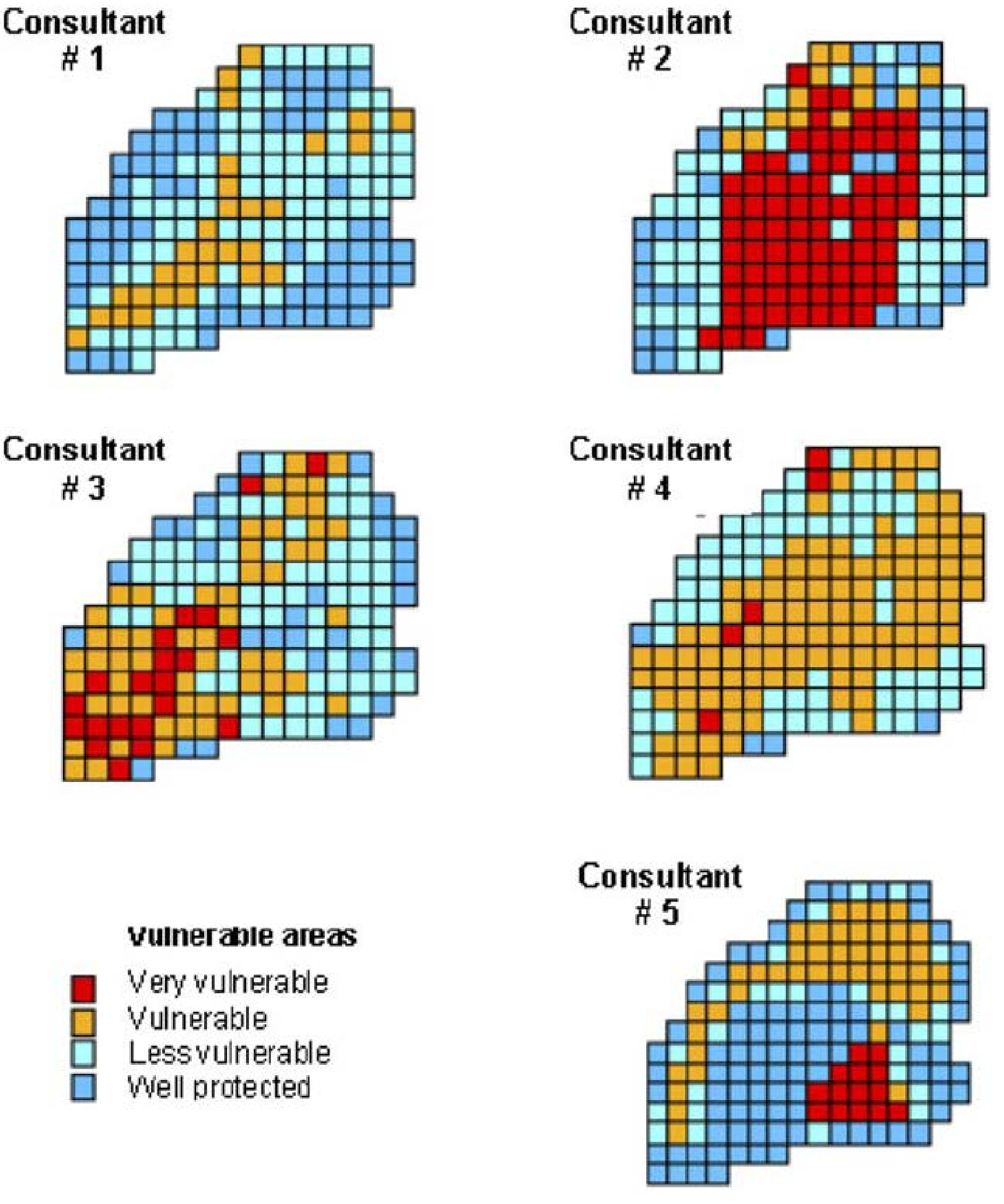


Fig. 1. Model predictions on aquifer vulnerability towards nitrate pollution for a 175 km^2 area west of Copenhagen [11].

3 framings of uncertainty

'deficit view'

- Uncertainty is provisional
- Reduce uncertainty, make ever more complex models
- *Tools:* quantification, Monte Carlo, Bayesian belief networks
    - ***Speaking truth to power***

'evidence evaluation view'

- Comparative evaluations of research results
- *Tools:* Scientific consensus building; multi disciplinary expert panels
- focus on robust findings
    - ***Speaking [consensus] to power***

'complex systems view / *post-normal* view'

- Uncertainty is intrinsic to complex systems
- Openly deal with deeper dimensions of uncertainty
- *Tools:* Knowledge Quality Assessment
    - ***Working deliberatively within imperfections***

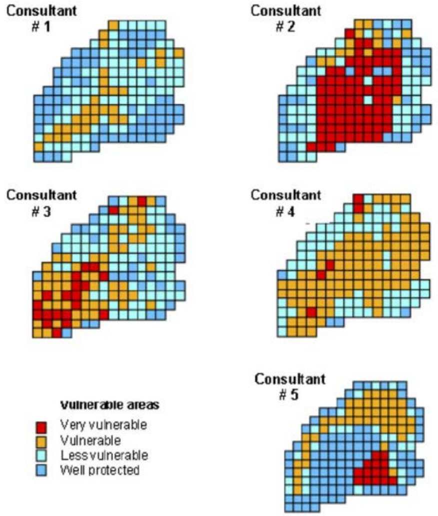


Fig. 1. Model predictions on aquifer vulnerability towards nitrate pollution for a 175 km² area west of Copenhagen [11].

How to act upon such uncertainty?

- **Bayesian** approach: 5 priors. Average and update likelihood of each grid-cell being red with data (but oooops, there is no data and we need decisions now)
- IPCC approach: Lock the 5 consultants up in a room and don't release them before they have **consensus**
- **Nihilist** approach: Dump the science and decide on an other basis
- **Precautionary** robustness approach: protect all grid-cells
- **Academic bureaucrat** approach: Weigh by citation index (or H-index) of consultant.
- Select the consultant that you **trust** most
- Real life approach: Select the consultant that best fits your **policy agenda**
- Post normal: explore the relevance of our ignorance: **working deliberatively within imperfections**

Clark & Majone 1985

Critical Appraisal of Scientific Inquiries with Policy Implications

1. Criticism by whom?

Critical roles

- Scientist
- Peer group
- Program Manager or Sponsor
- Policy maker
- Public interests groups

Clark & Majone 1985

Criticism of what?

Critical modes:

- Input
  - data; methods, people, competence, (im)matureness of field
- Output
  - problem solved? hypothesis tested?
- Process
  - good scientific practice, procedures for review, documenting etc.

Table 1. Critical criteria. **(Clark & Majone, 1985)**

Critical Role	Critical Mode Input	Output	Process
Scientist	Resource and time constraints; available theory; institutional support; assumptions; quality of available data; state of the art.	Validation; sensitivity analyses; technical sophistication; degree of acceptance of conclusions; impact on policy debate; imitation; professional recognition.	Choice of methodology (e.g., estimation procedures); communication; implementation; promotion; degree of formalization of analytic activities within the organization.
Peer Group	Quality of data; model and/or theory used; adequacy of tools; problem formulation. Input variables well chosen? Measure of success specified in advance?	Purpose of the study. Are conclusions supported by evidence? Does model offend common sense? Robustness of conclusions; adequate coverage of issues.	Standards of scientific and professional practice; documentation; review of validation techniques; style; interdisciplinarity.
Program Manager or Sponsor	Cost; institutional support within user organization; quality of analytic team; type of financing (e.g., grant vs. contract).	Rate of use; type of use (general education, program evaluation, decisionmaking, etc.); contribution to methodology and state of the art; prestige. Can results be generalized, applied elsewhere?	Dissemination; collaboration with users. Has study been reviewed?
Policymaker	Quality of analysts; cost of study; technical tools used (hardware and software). Does problem formulation make sense?	Is output familiar and intelligible? Did study generate new ideas? Are policy indications conclusive? Are they consonant with accepted ethical standards?	Ease of use; documentation. Are analysts helping with implementation? Did they interact with agency personnel? With interest groups?
Public Interest Groups	Competence and intellectual integrity of analysts. Are value systems compatible? Problem formulation acceptable? Normative implications of technical choices (e.g., choices of data).	Nature of conclusions; equity. Is analysis used as rationalization or to postpone decision? All viewpoints taken into consideration? Value issues.	Participation; communication of data and other information; adherence to strict rules of procedure.

Clark & Majone 1985

Meta quality criteria:

- Adequacy
  - reliability, reproducibility, uncertainty analysis etc.
- Value
  - Internal: how well is the study carried out?
  - External: fitness for purpose, fitness for function
  - Personal: subjectivity, preferences, choices, assumptions, bias
- Effectiveness
  - Does it help to solve practical problems
- Legitimacy
  - numinous: natural authority, independence, credibility, competence
  - civil: agreed procedures

Scandal at the Netherlands Environmental Assessment Agency

RIVM / De Kwaadsteniet (1999)

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

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

Science, Technology, & Human Values
36(3) 362-388

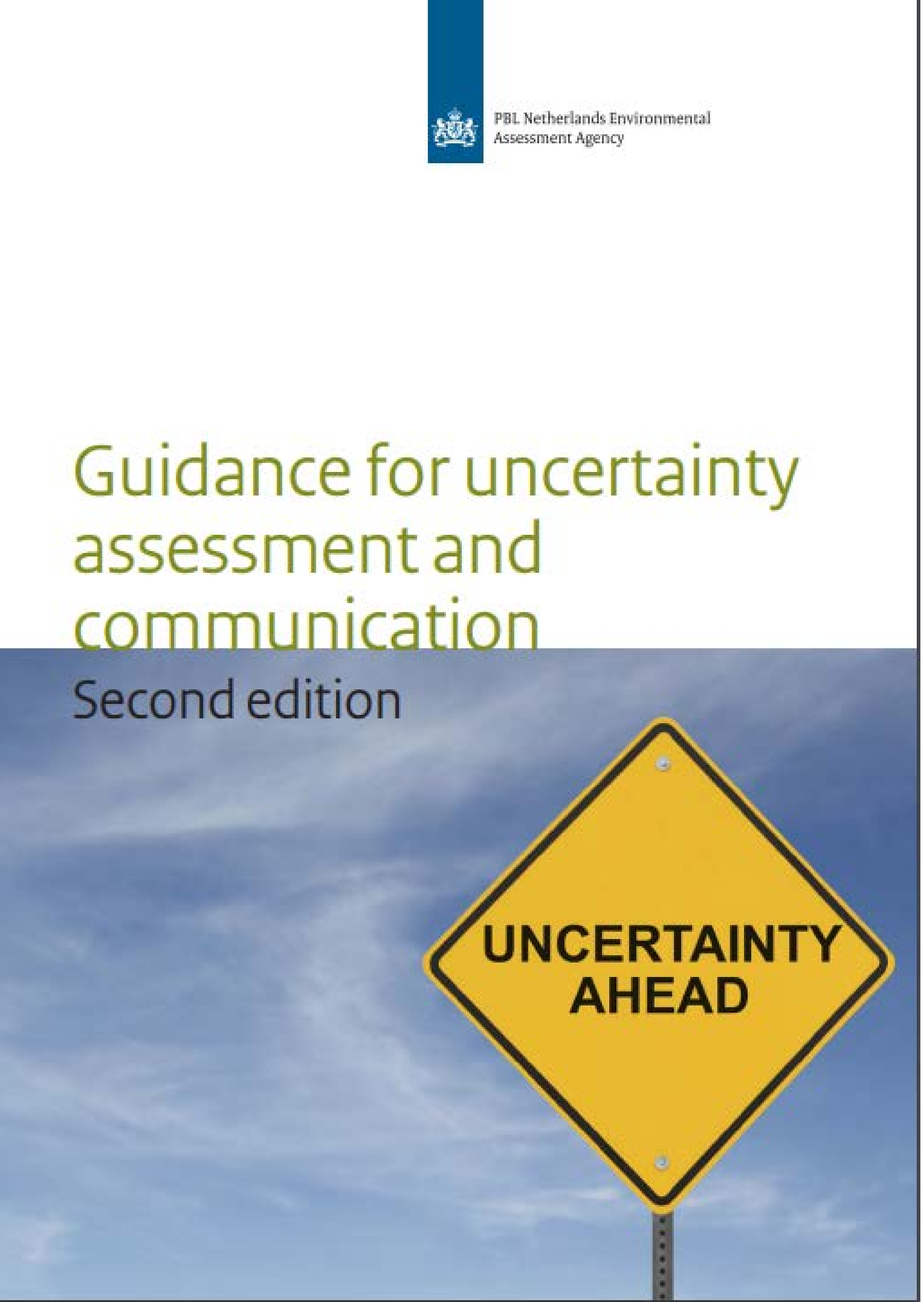
DOI: 10.1177/0162243910385797
http://sthv.sagepub.com

Post-Normal Science in Practice at the Netherlands Environmental Assessment Agency

Arthur C. Petersen,[1] Albert Cath,[2] Maria Hage,[1] Eva Kunseler,[1] and Jeroen P. van der Sluijs[3,4]

Abstract

About a decade ago, the Netherlands Environmental Assessment Agency (PBL) unwittingly embarked on a transition from a technocratic model of science advising to the paradigm of "post-normal science" (PNS). In response to a scandal around uncertainty management in 1999, a Guidance for "Uncertainty Assessment and Communication" was developed with advice from the initiators of the PNS concept and was introduced in 2003. This was followed in 2007 by a "Stakeholder Participation" Guidance. In this article, the authors provide a combined insider/outsider perspective on the transition process. The authors assess the extent to which the PNS paradigm has delivered new approaches in the agency's practice and analyze two projects—on long-term options for Dutch sustainable development policy and for urban development policy—the latter in somewhat more detail. The authors identify several paradoxes PBL encounters when putting the PNS concept into practice. It is concluded that an openness to other



http://www.pbl.nl/en/publications/guidance-for-uncertainty-assessment-and-communication

NL Environmental Assessment Agency (RIVM / MNP) Guidance: Systematic reflection on uncertainty & quality in:

Foci	Key issues
Problem framing	Other problem views; interwovenness with other problems; system boundaries; role of results in policy process; relation to previous assessments
Involvement of stakeholders	Identifying stakeholders; their views and roles; controversies; mode of involvement
Selection of indicators	Adequate backing for selection; alternative indicators; support for selection in science, society, and politics
Appraisal of knowledge base	Quality required; bottlenecks in available knowledge and methods; impact of bottlenecks on quality of results
Mapping and assessing relevant uncertainties	Identification and prioritisation of key uncertainties; choice of methods to assess these; assessing robustness of conclusions
Reporting uncertainty information	Context of reporting; robustness and clarity of main messages; policy implications of uncertainty; balanced and consistent representation in progressive disclosure of uncertainty information; traceability and adequate backing

Problem framing and context

- Explore rival problem frames
- Relevant aspects / system boundary
- Typify problem structure
- Problem lifecycle / maturity
- Role of study in policy process
- Uncertainty in socio-political context

Type-III error:
Assessing the wrong problem by incorrectly accepting the false meta-hypothesis that there is no difference between the boundaries of a problem, as defined by the analyst, and the actual boundaries of the problem (Dunn, 1997).

Context validation (Dunn, 1999).
The validity of inferences that we have estimated the proximal range of rival hypotheses.

Context validation can be performed by a participatory bottom-up process to elicit from scientists and stakeholders rival hypotheses on causal relations underlying a problem and rival problem definitions.

Involvement of stakeholders

- Identify relevant stakeholders.
- Identification of areas of agreement and disagreement among stakeholders on value dimensions of the problem.
- Recommendations on when to involve different stakeholders in the assessment process.

Roles of stakeholders

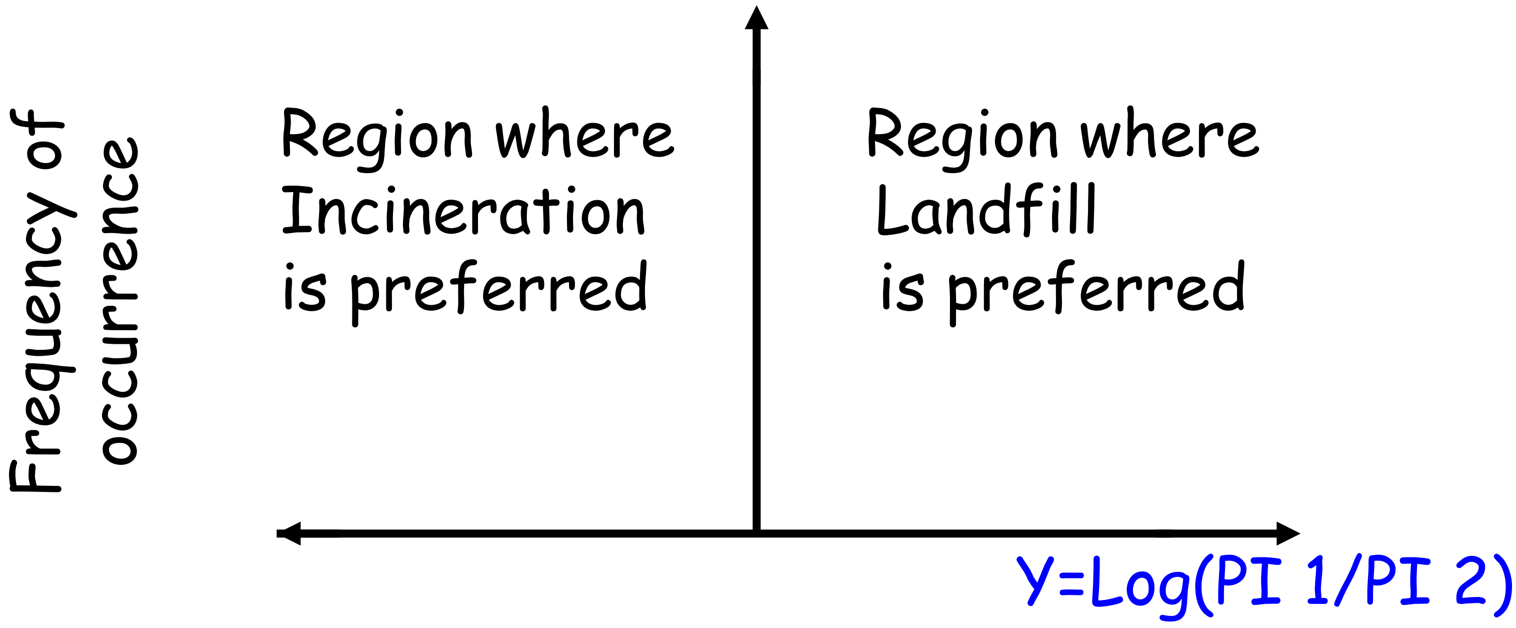
- (Co-) definer of the problems to be addressed
  - What knowledge is relevant?
- Source of knowledge
- Quality control of the science (for instance: review of assumptions)

Indicators

- How well do indicators used address key aspects of the problem?
- Use of proxies
- Alternative indicators?
- Limitations of indicators used?
- Scale and aggregation issues
- Controversies in science and society about these indicators?

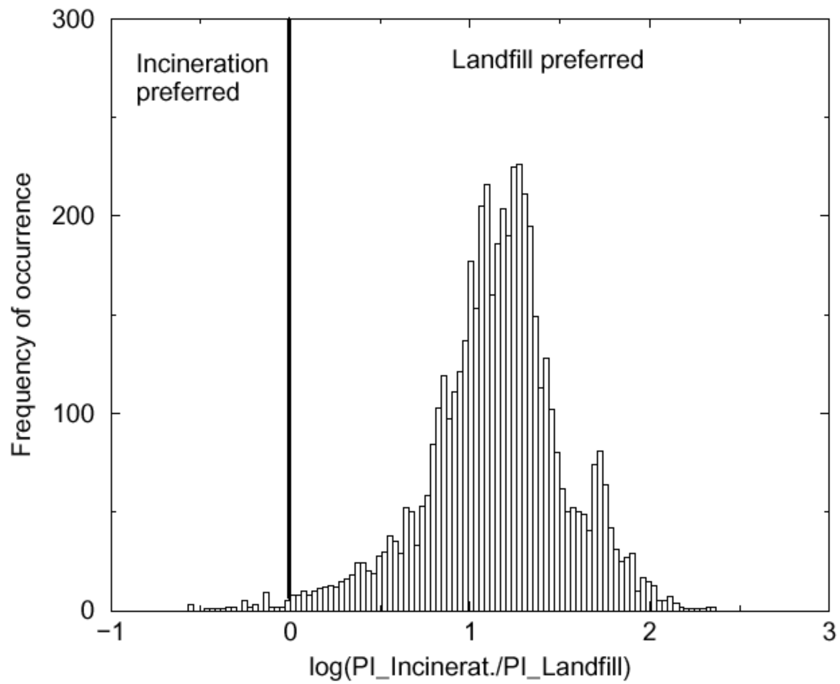
High uncertainty is not the same as low quality

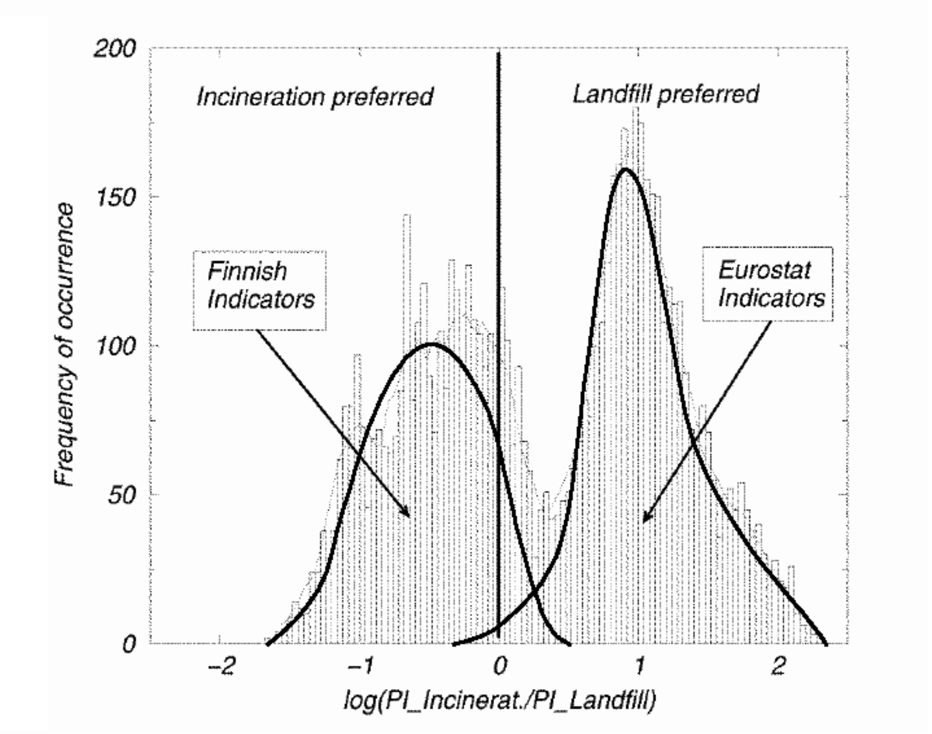
Example: imagine the inference is Y = the logarithm of the ratio between the two pressure-on-decision indices PI1 and PI2



High uncertainty is not the same as low quality,

but..... methodological uncertainty can de dominant





(slide borrowed from Andrea Saltelli)

Do we know enough to quantify?

Risbey & Kandlikar (2007): What format is in accordance with the level of knowledge on the quantity?

- Full probability density function
  - Robust, well defended distribution
- Bounds
  - Well defended percentile bounds
- First order estimates
  - Order of magnitude assessment
- Expected sign or trend
  - Well defended trend expectation
- Ambiguous sign or trend
  - Equally plausible contrary trend expectations
- Effective ignorance
  - Lacking or weakly plausible expectations

Uncertainty is more than a number

Dimensions of uncertainty:

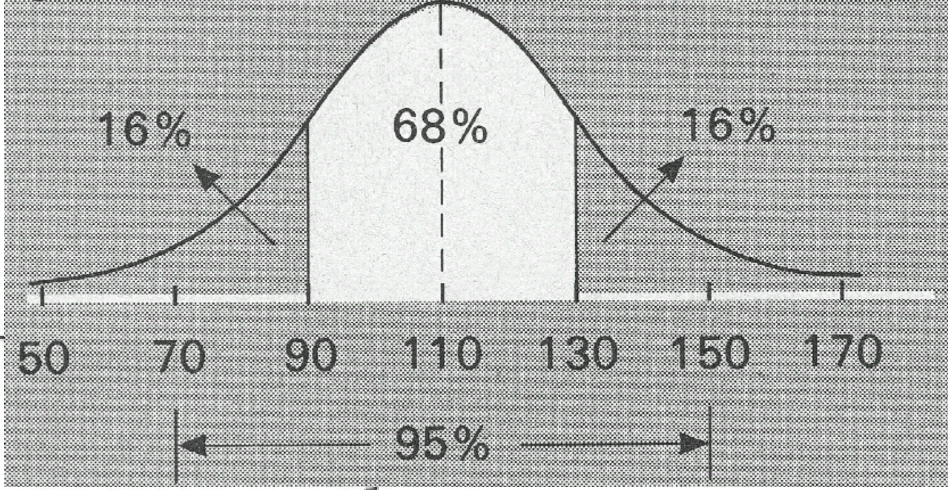
- Technical (inexactness)
- Methodological (unreliability)
- Epistemological (ignorance)
- Societal (limited social robustness)

Reliability intervals normal distributions

$\pm\,\sigma = 68\ \%$

$\pm\,2\sigma = 95\ \%$

$\pm\,3\sigma = 99.7\ \%$



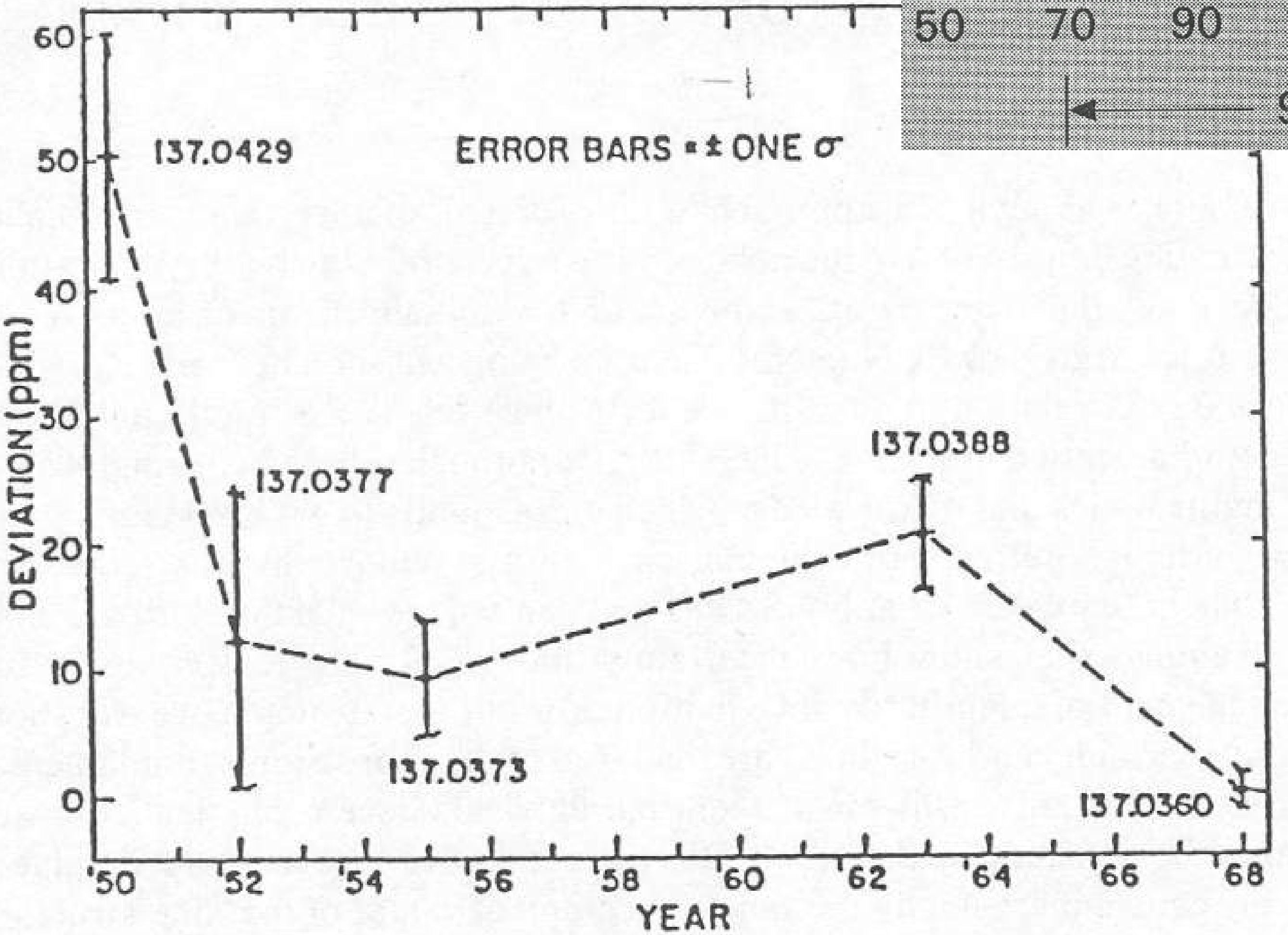
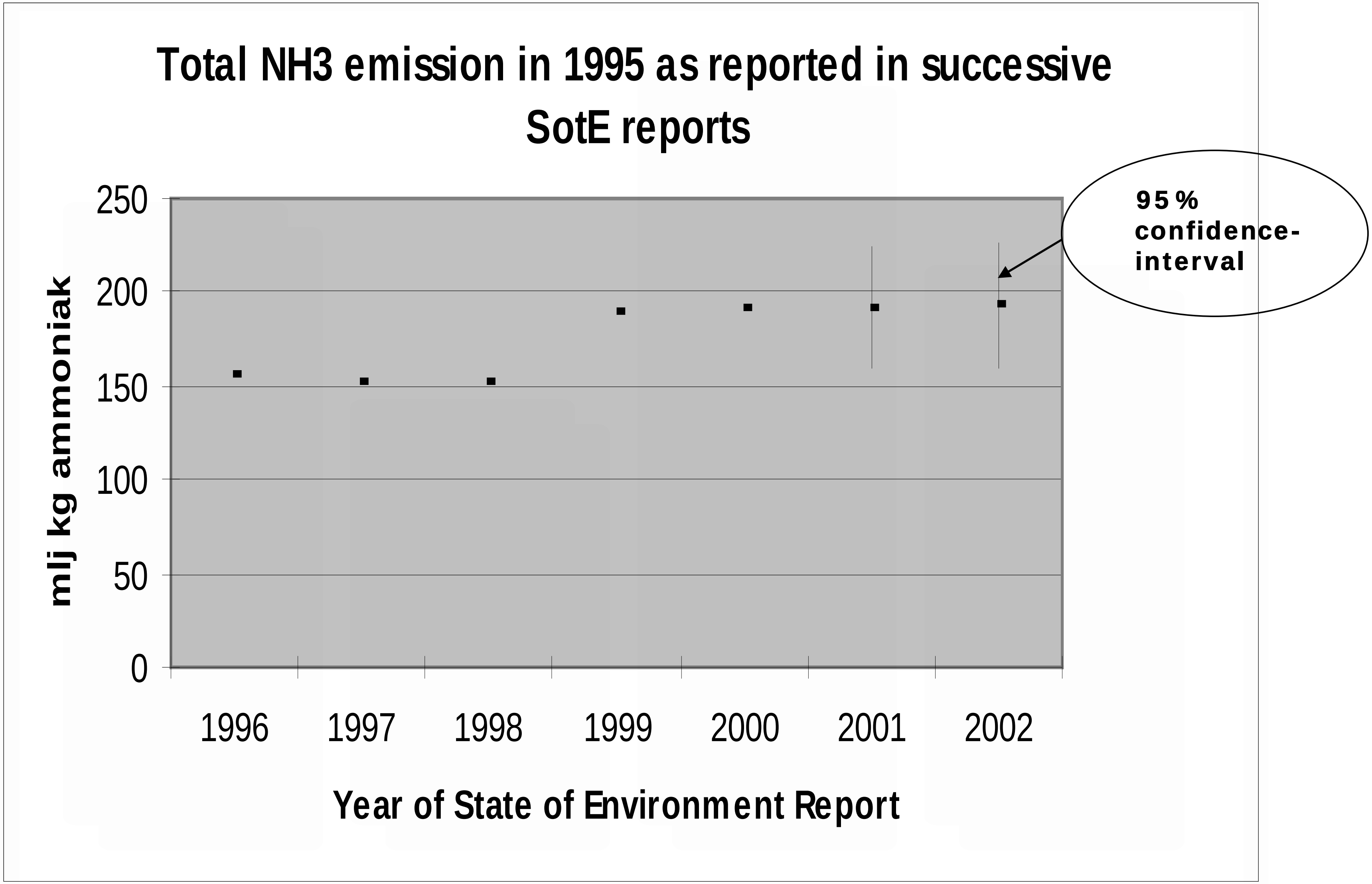


Fig. 1. Successive recommended values of the fine-structure constand α^{-1} (B. N. Taylor *et al.*, 1969, 7)



Total NH3 emission in 1995 as reported in successive SotE reports
95% confidence-interval
mlj kg ammoniak
250
200
150
100
50
0
1996
1997
1998
1999
2000
2001
2002
Year of State of Environment Report

NUSAP: Qualified Quantities

Classic scientific notational system:

- **N**umeral **U**nit **S**pread

For problems in the post-normal domain, add two qualifiers:

- **A**ssessment & **P**edigree

  "Assessment" expresses expert judgement on reliability of numeral + spread

  "Pedigree" expresses multi-criteria evaluation of the strength of a number by looking at:

  - Background history by which the number was produced
  - Underpinning and scientific status of the number

Example Pedigree matrix parameter strength

Code	Proxy	Empirical	Theoretical basis	Method	Validation
4	Exact measure	Large sample direct mmts	Well established theory	Best available practice	Compared with indep. mmts of same variable
3	Good fit or measure	Small sample direct mmts	Accepted theory partial in nature	Reliable method commonly accepted	Compared with indep. mmts of closely related variable
2	Well correlated	Modeled/derived data	Partial theory limited consensus on reliability	Acceptable method limited consensus on reliability	Compared with mmts not independent
1	Weak correlation	Educated guesses / rule of thumb est	Preliminary theory	Preliminary methods unknown reliability	Weak / indirect validation
0	Not clearly related	Crude speculation	Crude speculation	No discernible rigour	No validation

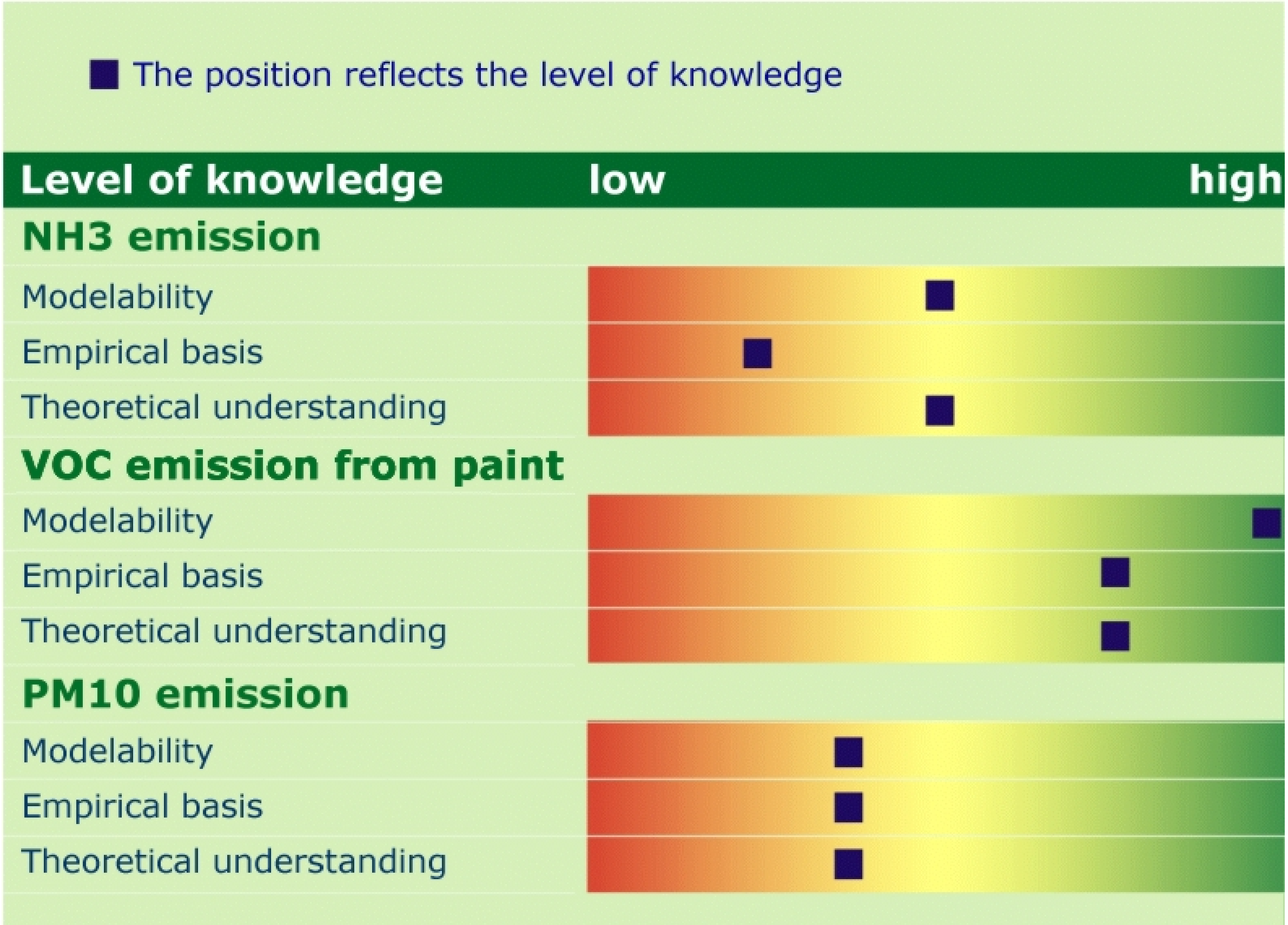
Example Pedigree results

	Proxy	Empirical	Method	Validation	Strength
NS-SHI	3	3.5	4	0	0.66
NS-B&S	3	3.5	4	0	0.66
NS-DIY	2.5	3.5	4	3	0.81
NS-CAR	3	3.5	4	3	0.84
NS-IND	3	3.5	4	0.5	0.69
Th%-SHI	2	1	2	0	0.31
Th%-B&S	2	1	2	0	0.31
Th%-DIY	1	1	2	0	0.25
Th%-CAR	2	1	2	0	0.31
Th%-IND	2	1	2	0	0.31
VOS % import	1	2	1.5	0	0.28
Attribution import	1	1	2	0	0.25

Trafic-light analogy <1.4 red; 1.4-2.6 amber; >2.6 green

This example is the case of VOC emissions from paint in the Netherlands, calculated from national sales statistics (NS) in 5 sectors (Ship, Building & Steel, Do It Yourself, Car refinishing and Industry) and assumptions on additional thinner use (Th%) and a lump sum for imported paint and an assumption for its VOC percentage. See full research report on www.nusap.net for details.

Example: Air Quality



In summary, NUSAP

- Has a strong theoretical foundation in the theory of knowledge and the philosophy of science
- Addresses all three dimensions of uncertainty: technical (inexactness), methodological (unreliability) and epistemological (border with ignorance) in an coherent way
- Provides a systematic framework for synthesising qualitative and quantitative assessments of uncertainty
- Can act as a bridge between the quantitative mathematical disciplines and traditions and the qualitative discursive and participatory disciplines and traditions in the field of uncertainty management.
- Helps to focus research efforts on the potentially most problematic model components
- Pinpoints specific weaknesses in these components
- Provides those who produce, use and are affected by policy-relevant knowledge a tool for a critical self-awareness of their engagement with that knowledge. It thereby fosters extended peer review processes.

Uncertainty matrix

UNCERTAINTY MATRIX			Level of uncertainty (from 'knowing for certain' (deterministic knowledge) to 'not even knowing what you do not know' (total ignorance))			Nature of uncertainty		Qualification of knowledge base (backing)			Value-ladenness of choices		
Location ↓			Statistical uncertainty (range+ chance)	Scenario uncertainty (range indicated as 'what-if' option)	Recognised ignorance	Knowledge-related uncertainty	Variability-related uncertainty	Weak –	Fair 0	Strong +	Small –	Medium 0	Large +
Context	Assumptions on system boundaries and ecological, technological, economic, social and political context												
Expert judgement		Narrative; storyline; advice											
Model	Model structure	Relations											
Model	Technical model	Software and hardware implementation											
Model	Model parameters												
Model	Model inputs	Input data; driving forces; input scenarios											
Data (in a general sense)	Measurements; monitoring; surveys												
Outputs		Indicators; statements											

Reporting

- Make uncertainties explicit
- Assess robustness of results
- Discuss implications of uncertainty findings for different settings of burden of proof
- Relevance of results to the problem
- Progressive disclosure of information -> traceability and backing

Insights on uncertainty

- More research tends to increase uncertainty
  - reveals unforeseen complexities
  - Complex systems exhibit irreducible uncertainty (intrinsic or practically)
- Omitting uncertainty management can lead to scandals, crisis and loss of trust in science and institutions
- In many complex problems unquantifiable uncertainties dominate the quantifiable uncertainty
- High quality $\neq$ low uncertainty
- Quality relates to **fitness for function** (robustness, PP)
- Shift in focus needed from reducing uncertainty towards reflective methods to explicitly cope with uncertainty and quality

AFTERNOON PRACTICUM

Group exercise

Assess the pedigree of the model used in the following study:

Extinction risk from climate change

(Thomas *et al.*, *Nature*, 8 January 2004)

Main message of this paper:

- In 2050, 15-37% of species ‘committed to extinction’ due to climate change for a mid-range climate scenario

Extinction risks from climate change

Species-Area relationship:

- numbers of species that become extinct or threatened by habitat loss from climate change

$$S = c\,A^{z}$$

S = number of species

A = area,

c = constant

$z \approx 0.25$

Ratio of number of species that can live in a habitat of area *A* before (0) and after (t) climate change 'predicts' extinction rate:

$$\frac{S_t}{S_0} = \frac{c\,A_t^{\ z}}{c\,A_0^{\ z}} = (A_t\,/A_0)^z$$

Species committed to extinction

Climate scenario 2050	universal dispersal	no dispersal
> 2.0 °C	21–32%	38–52%
1.8–2.0 °C	15–20%	26–37%
0.8–1.7 °C	9–13%	22–31%

(Thomas et al., 2004)

Rule of thumb

Warming rate 1°C / century corresponds to:

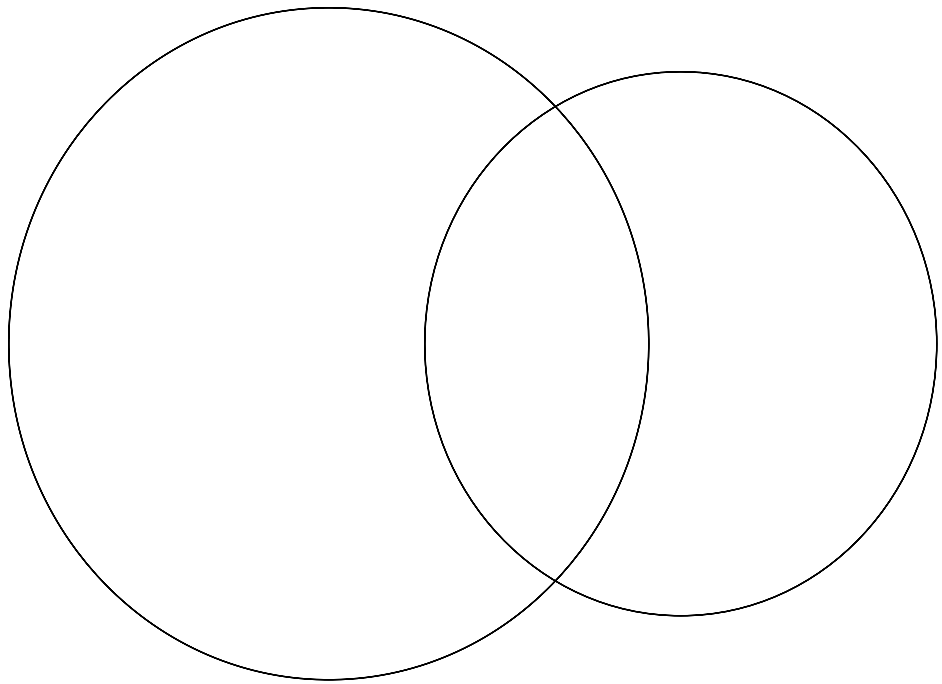
- ± 20 cm sea level rise
- ± 100 km shift of climate zone / century
- ± 150 m upward shift alpine climate zone/century

Climate tolerances of ecosystems

Ecosystem	Climate tolerance (°C/century)
Alpine ecosystem	0
Oak forest	0.12
Mangrove forest	0.50
Coastal wetlands	0.75
Coral reefs equator	1
Coral reefs N/S borders	5

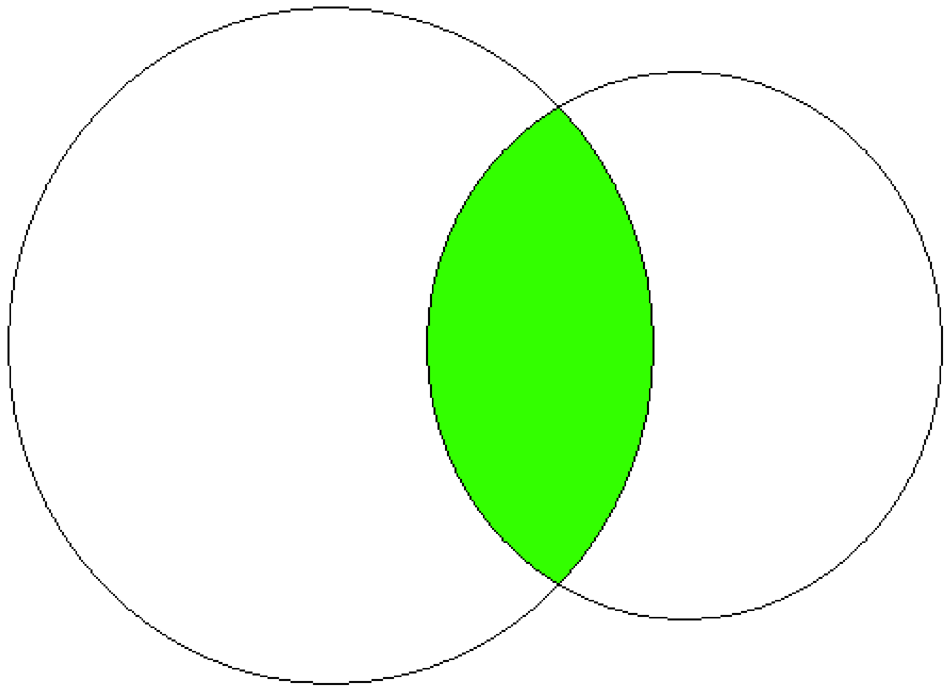
(Hinkley, 1997)

Habitat before Climate Change **Habitat after Climate Change**



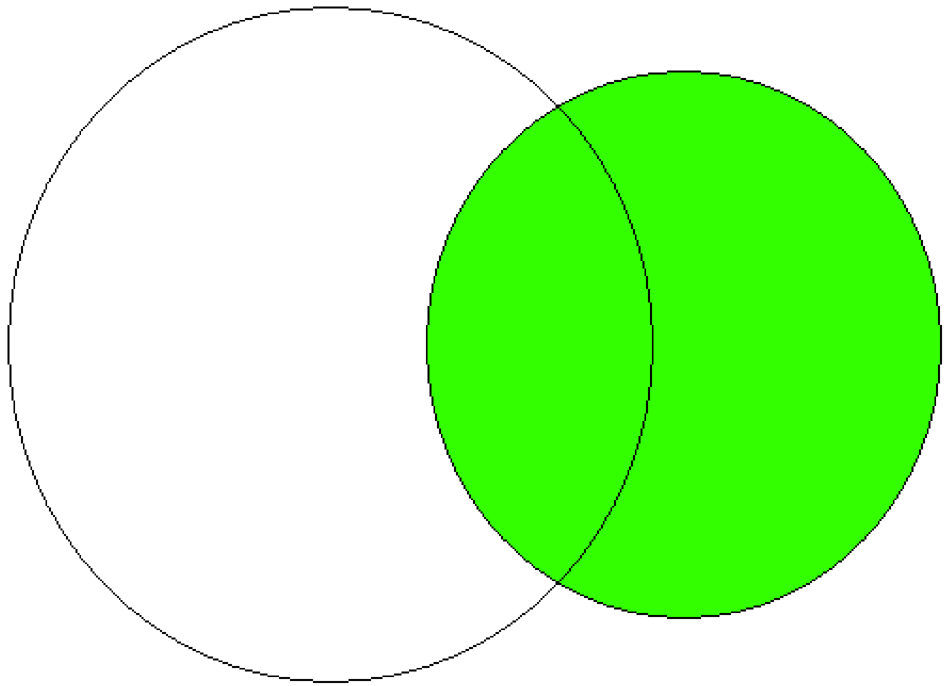
Habitat before Climate Change

Habitat after Climate Change



Assumption: No dispersal

Habitat before Climate Change **Habitat after Climate Change**



Assumption: Full dispersal

Pedigree matrix for evaluating models

Score	Supporting empirical evidence		Theoretical understanding	Representa-tion of understood underlying mechanisms	Plausibility	Colleague consensus
	Proxy	Quality and quantity				
4	Exact measures of the modelled quantities	Controlled experiments and large sample direct measurements	Well established theory	Model equations reflect high mechanistic process detail	Highly plausible	All but cranks
3	Good fits or measures of the modelled quantities	Historical/field data uncontrolled experiments small sample direct measurements	Accepted theory with partial nature (in view of the phenomenon it describes)	Model equations reflect acceptable mechanistic process detail	Reasonably plausible	All but rebels
2	Well correlated but not measuring the same thing	Modelled/derived data Indirect measurements	Accepted theory with partial nature and limited consensus on reliability	Aggregated parameterized meta model	Somewhat plausible	Competing schools
1	Weak correlation but commonalities in measure	Educated guesses indirect approx. rule of thumb estimate	Preliminary theory	Grey box model	Not very plausible	Embrionic field
0	Not correlated and not clearly related	Crude speculation	Crude speculation	Black box model	Not at all plausable	No opinion

Instructions

- Do the Pedigree assessment as an **individual** expert judgement, we do not want a group judgement
- Main function of group discussion is clarification of concepts
- Group works on one pedigree criterion at a time
- If you feel you cannot judge on of the pedigree criteria, leave it blank