Science, Policy and Epidemiology Workshop
Florence, November 21-23, 2011

Dealing with uncertainty II

concepts, tools, pitfalls

Dr. Jeroen P. van der Sluijs

Copernicus Institute for Sustainable Development and Innovation
Utrecht University

&

Research in Ecological Economics, Eco-Innovation & Tool Development for Sustainability
Université de Versailles Saint-Quentin-en-Yvelines, France

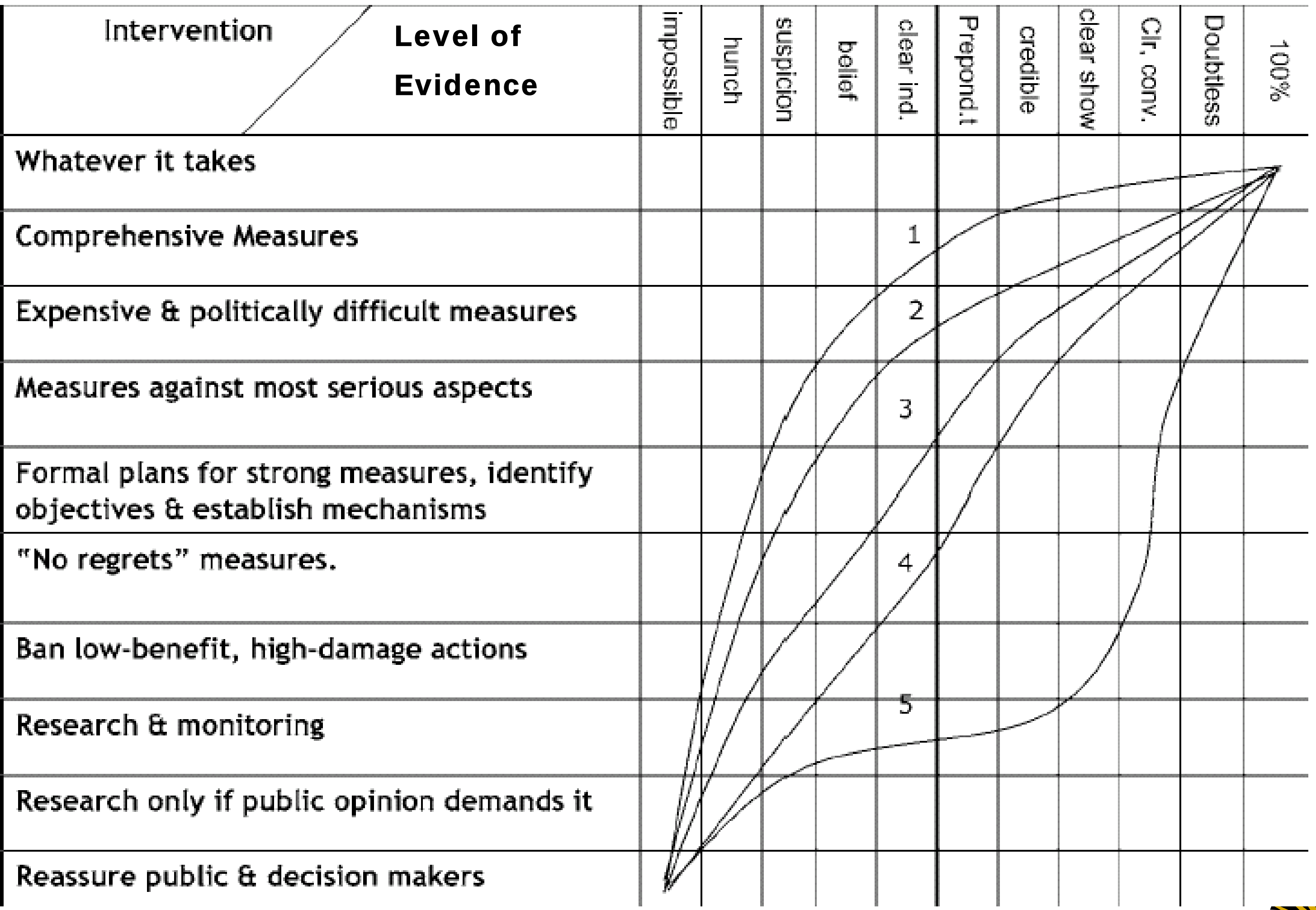
Insights on uncertainty

- 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)
- Add focus needed from reducing uncertainty towards reflective methods to explicitly cope with uncertainty and quality

Weiss 2003/2006 evidence scale

10. Virtually certain
9. Beyond a reasonable doubt
8. Clear and Convincing Evidence
7. Clear Showing
6. Substantial and credible evidence
5. Preponderance of the Evidence
4. Clear indication
3. Probable cause: reasonable grounds for belief
2. Reasonable, articulable grounds for suspicion
1. No reasonable grounds for suspicion
0. Insufficient even to support a hunch or conjecture

Even where there is agreement on “level of evidence”, there usually is substantial societal disagreement on what level of intervention is justified.



Attitudes according to Weiss 2003:

1. Environmental absolutist

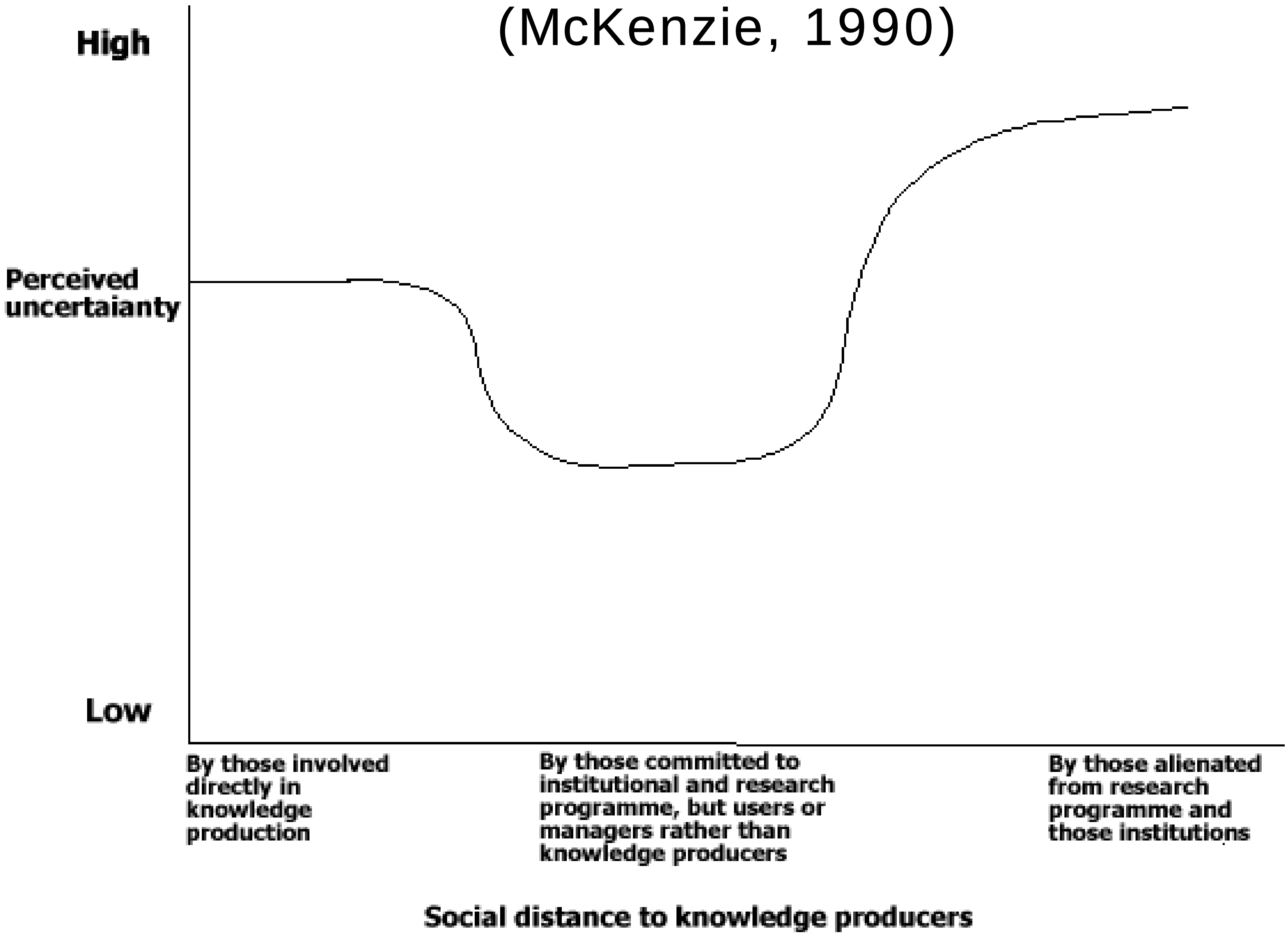
2. Cautious environmentalist

3. Environmental centrist

4. Technological optimist

5. Scientific absolutist

The certainty trough (McKenzie, 1990)



Consensus approach IPCC problematic

- Undue certainty (high error costs!)
- promotes anchoring towards previously established consensus positions
- Hides diversity of perspectives
- Constrains decision-makers options
- Underexposes dissent
  - hampers both scientific debates and policy debates

http://dx.doi.org/10.1016/j.cosust.2010.10.003
http://www.nature.com/news/2011/111005/full/478007a.html

Published online 5 October 2011 | *Nature* **478**, 7 (2011) | doi:10.1038/478007a

Column: World View

The voice of science: let's agree to disagree

Consensus reports are the bedrock of science-based policy-making. But disagreement and arguments are more useful, says Daniel Sarewitz.

Daniel Sarewitz

In case of complex problems, the "Speaking truth to power" model fails because:

- Truth cannot be known and is thus not a substantial aspect of the issue
- *"… good scientific work has a product, which should … correspond to Nature as closely as possible… But* ***the working judgements on the product are of its quality, and not of its logical truth****."*

(Funtowicz and Ravetz 1990, p. 30)

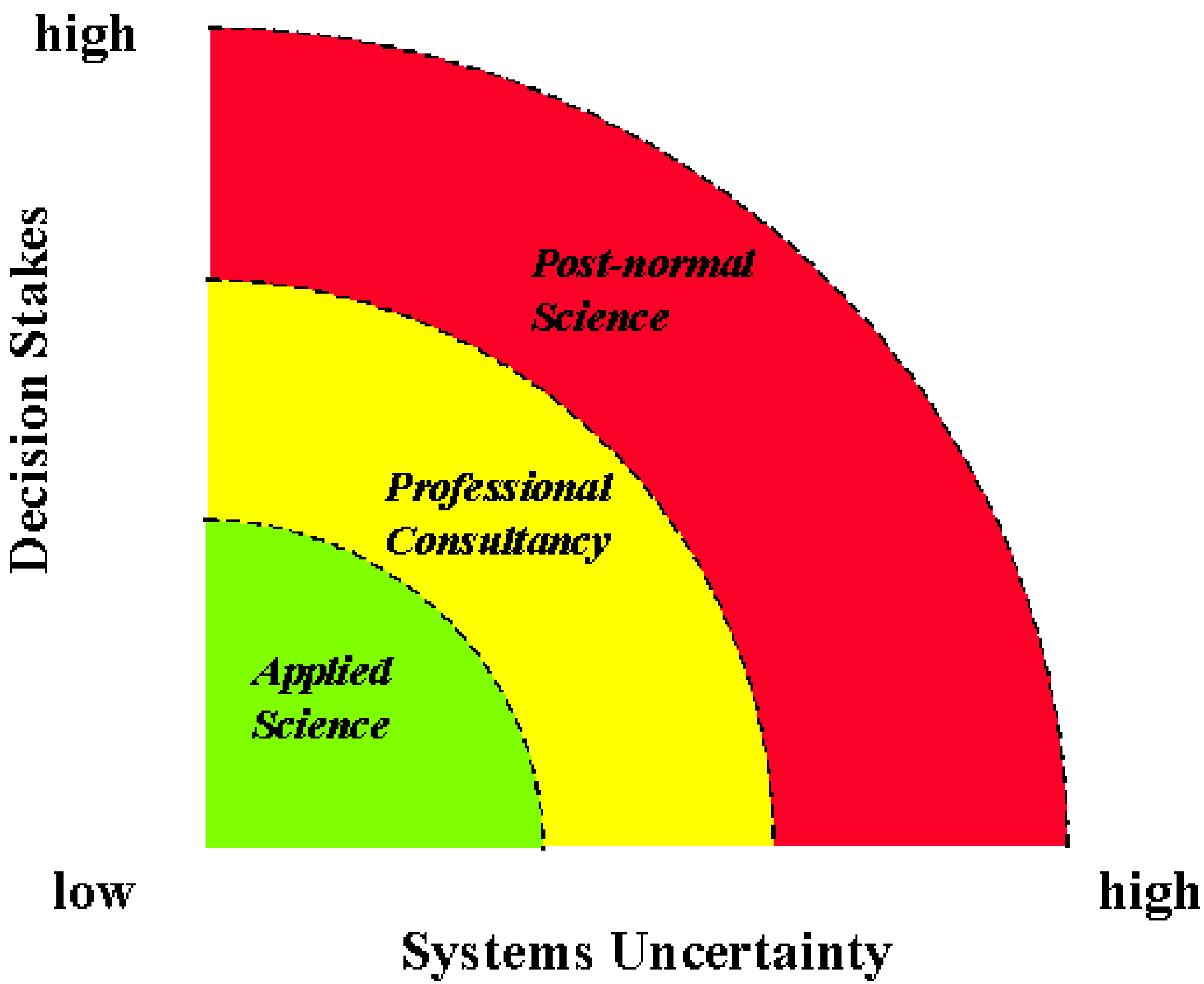
The alternative model: PNS

Extended participation:

working deliberatively within imperfections

- Science is only one part of relevant **evidence**
- Critical dialogue on strength and relevance of evidence
- Interpretation of evidence and attribution of policy meaning to knowledge is democratized
- **Tools for Knowledge Quality Assessment empower all stakeholders to engage in this deliberative process**

(Funtowicz, 2006; Funtowicz & Strand, 2007)



Funtowicz and Ravetz, Science for the Post Normal age, *Futures*, 1993

Elements of Post Normal Science

- Appropriate management of uncertainty quality and value-ladenness
- Plurality of commitments and perspectives
- Internal extension of peer community *(involvement of other disciplines)*
- External extension of peer community *(involvement of stakeholders in environmental assessment & quality control)*

Plurality and uncertainty in risk assessment: lessons learned

- **Diversity of the knowledge base**:
  - It must be based on the full spectrum of available scientific knowledge;
- **Robustness of the knowledge claims**
  - Include uncertainty, dissent and criticism in the analysis, synthesis and assessments;
- Make thorough **Knowledge Quality Assessment the key task in the science policy interface** and develop a joint language to communicate limitations to our knowledge and understanding clearly and transparently
  - Bayesian likelihood terminology is misleading, it unduly suggests certainty;
- Make use of **information of non-scientific sources** (local knowledge)
  - But scrutinize this information and be clear on its status;
- **Clarify values, stakes and vested interests** that play a role in research and in the political and socioeconomic context within which the research is embedded.

(Maxim and van der Sluijs, 2007)

RIVM / De Kwaadsteniet (1999)

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

Newspaper headlines:

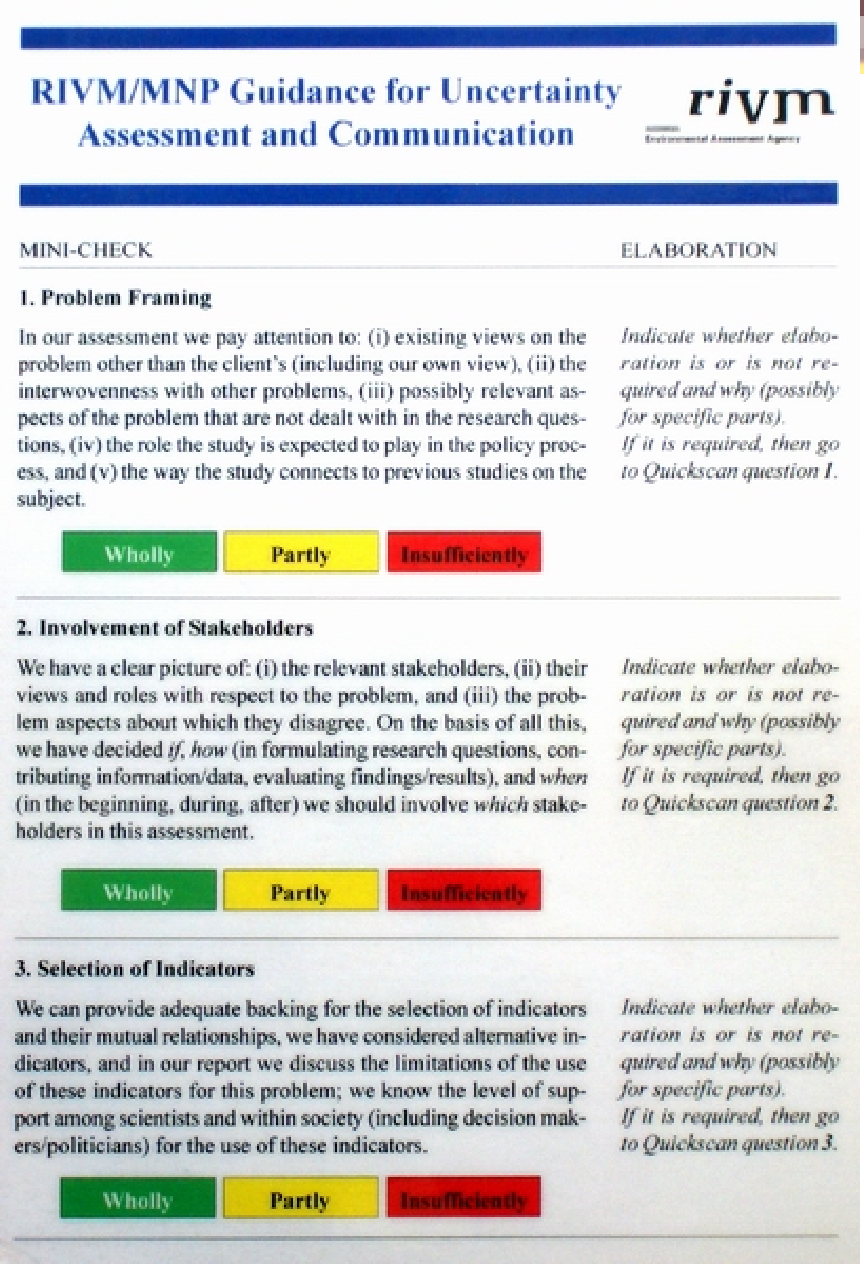
- Environmental institute lies and deceits
- Fuss in parliament after criticism on environmental numbers
- The bankruptcy of the environmental numbers
- Society has a right on fair information, RIVM does not provide it

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

PNS in 2008:

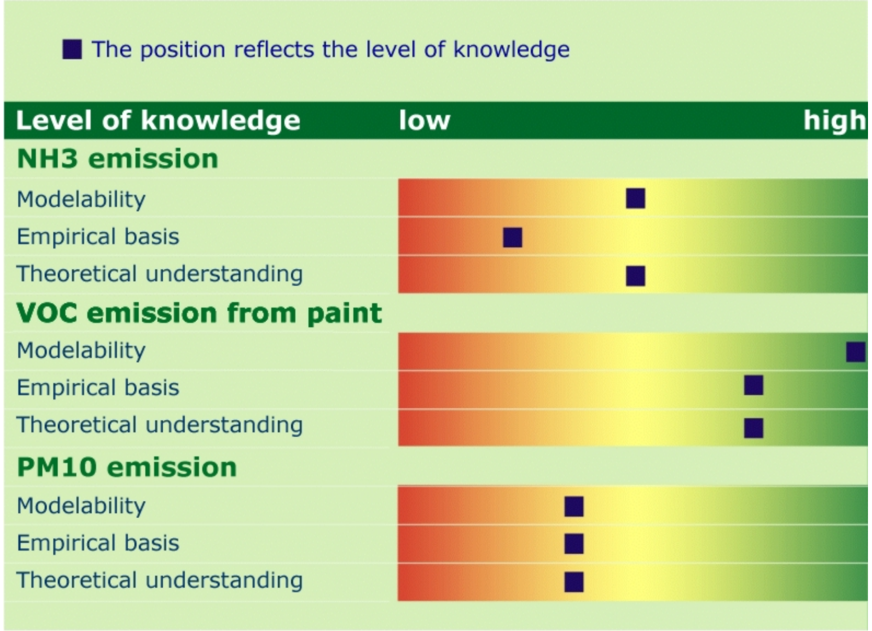
Tools & checklists for Knowledge Quality Assessment



RIVM/MNP Guidance for Uncertainty Assessment and Communication

rivm

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. Wholly / Partly / Insufficiently	*Indicate whether elaboration is or is not required and why (possibly for specific parts). If it is required, then go to Quickscan question 1.*
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. Wholly / Partly / Insufficiently	*Indicate whether elaboration is or is not required and why (possibly for specific parts). If it is required, then go to Quickscan question 2.*
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. Wholly / Partly / Insufficiently	*Indicate whether elaboration is or is not required and why (possibly for specific parts). If it is required, then go to Quickscan question 3.*



***SCIENCE* VOL 316 13 APRIL 2007**

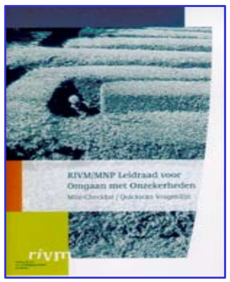
"Today, eight years on from the Dutch scandal, no one makes more strenuous efforts than does the Netherlands' RIVM to accommodate and cope with the uncertainties of environmental data and models, hence to achieve the greatest possible quality in generating environmental foresight."

(Bruce Beck)

RIVM-MNP Uncertainty Guidance

Reminder list
Invokes Reflection
Portal to QS

Mini-Checklist

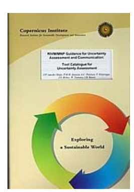


Quickscan Questionnaire

Further Guidance
Advice
Hints & Implications

Quickscan Hints & Actions List

Tool Catalogue for Uncertainty Assessment



Detailed Guidance

Advice on Quantitative + Qualitative tools for UA

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).

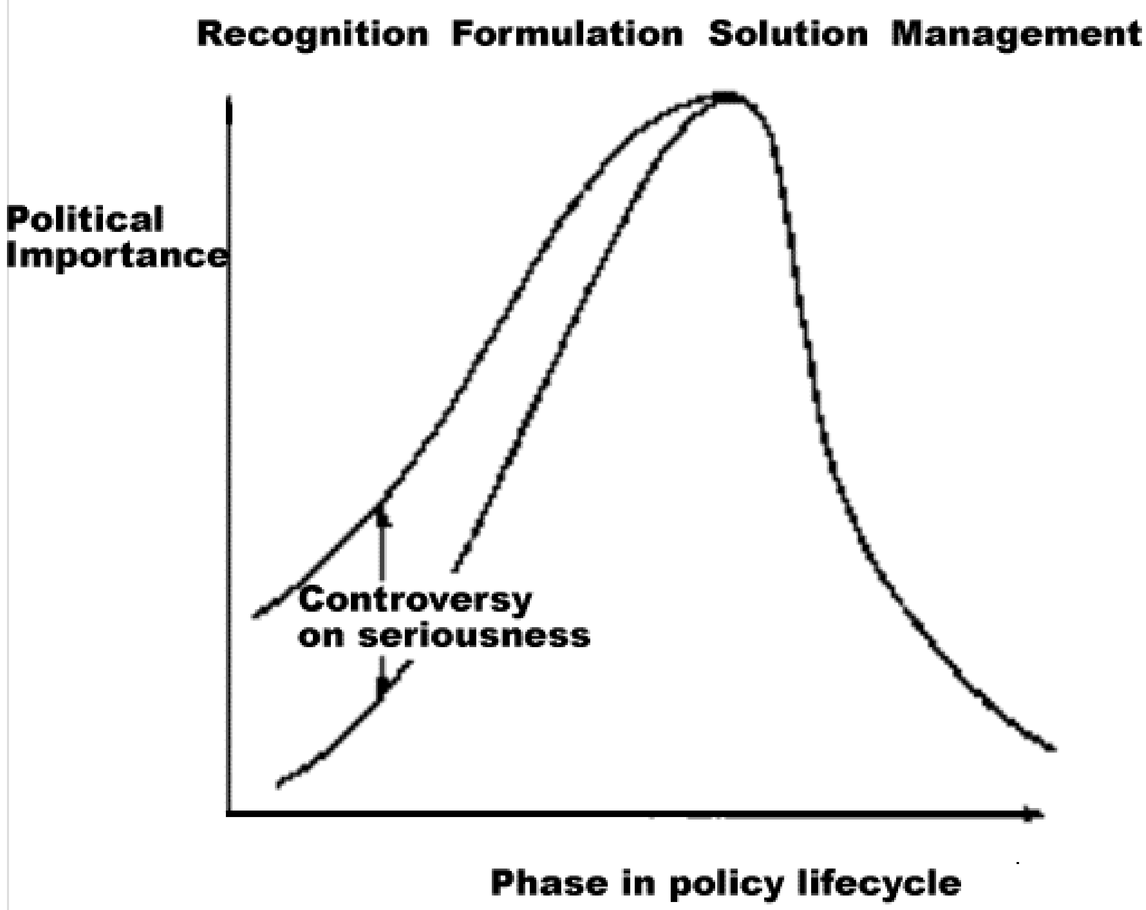
Context validation (Dunn, 1999).
The validity of inferences that we have estimated the proximal range of rival hypotheses.

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.

What is the role of the assessment in the policy process?

- ad hoc policy advice
- to evaluate existing policy
- to evaluate proposed policy
- to foster recognition of new problems
- to identify and/or evaluate possible solutions
- to provide counter-expertise
- other

In different phases of problem lifecycle, different uncertainties are salient



Different problem-types need different uncertainty management strategies

		Consensus on relevant norms and values: *No*	Consensus on relevant norms and values: *Yes*
Certainty about Relevant knowledge	*No*	UNSTRUCTURED PROBLEM Science as problem recognizer Policy as learning	MODERATELY STRUCTURED PROBLEM (ENDS) Science as advocate Policy as negotiation
	Yes	MODERATELY STRUCTURED PROBLEM (MEANS) Science as mediator Policy as accommodation	STRUCTURED PROBLEM Science as problem solver Policy as rule

		Consensus about values	
		No	Yes
Consensus about knowledge	No	*Unstructured* • Ignorance • Value-ladenness • Problem framing • Scenario uncertainty • Public debate • Conflict management • Reflexive science.	*Moderately structured (ends)* • Unreliability • Scenario uncertainty • Ignorance • Stakeholder involvement • Extended peer review
	Yes	*Moderately structured (means)* • Value ladenness • Strategic knowledge use • Accomodation • Reflexive science.	*Structured* • Statistical uncertainty • Normal scientific procedures • Statistical approaches

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
- Source of knowledge
- Quality control of the science (for instance: review of assumptions)

Extended Peer Involvement in practice

MNP Guidance on Stakeholder Particpation

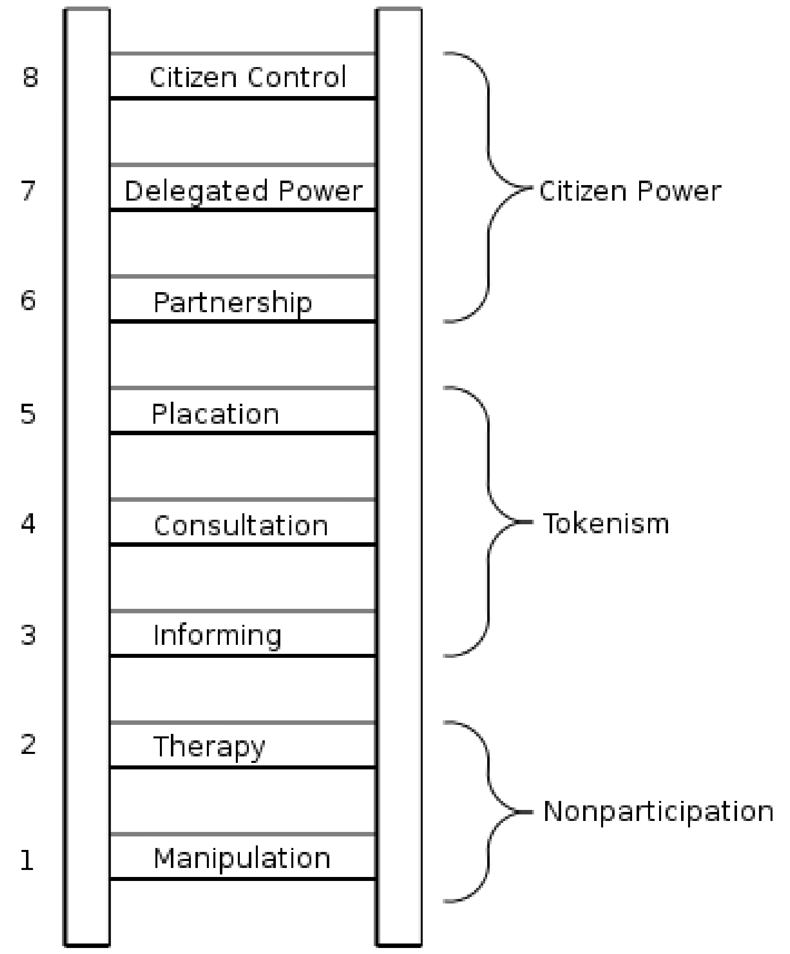
- **Why** participation?
- **What** should participation be about?
- **Who** to involve?
- **How much** participation?
- **What form**?

Incentives for participatory risk assessment

- Instrumental
  - decrease conflict/increase acceptance of or trust in the science
- Normative
  - process should be legitimate/ democracy
- Substantive
  - relevant wisdom is not limited to scientific specialists and public officials
  - Bounded rationality
  - Increase quality

(Stern & Fineberg, ***Understanding Risk, Informing Decisions in a Democratic Society***, 1996)

Sherry Arnstein 1969
Ladder of Citizen Participation
Implemented at MNP 2007



	Level of ambition	Direction of communication
Interactief	Co-decide	MNP & SH*
	Co-produce	MNP ↔ SH
	Take advice / Consult	MNP ← SH
Niet interactief	Listen	MNP ← SH
	Study	MNP SH
	Inform	MNP → SH
	No participation	MNP SH

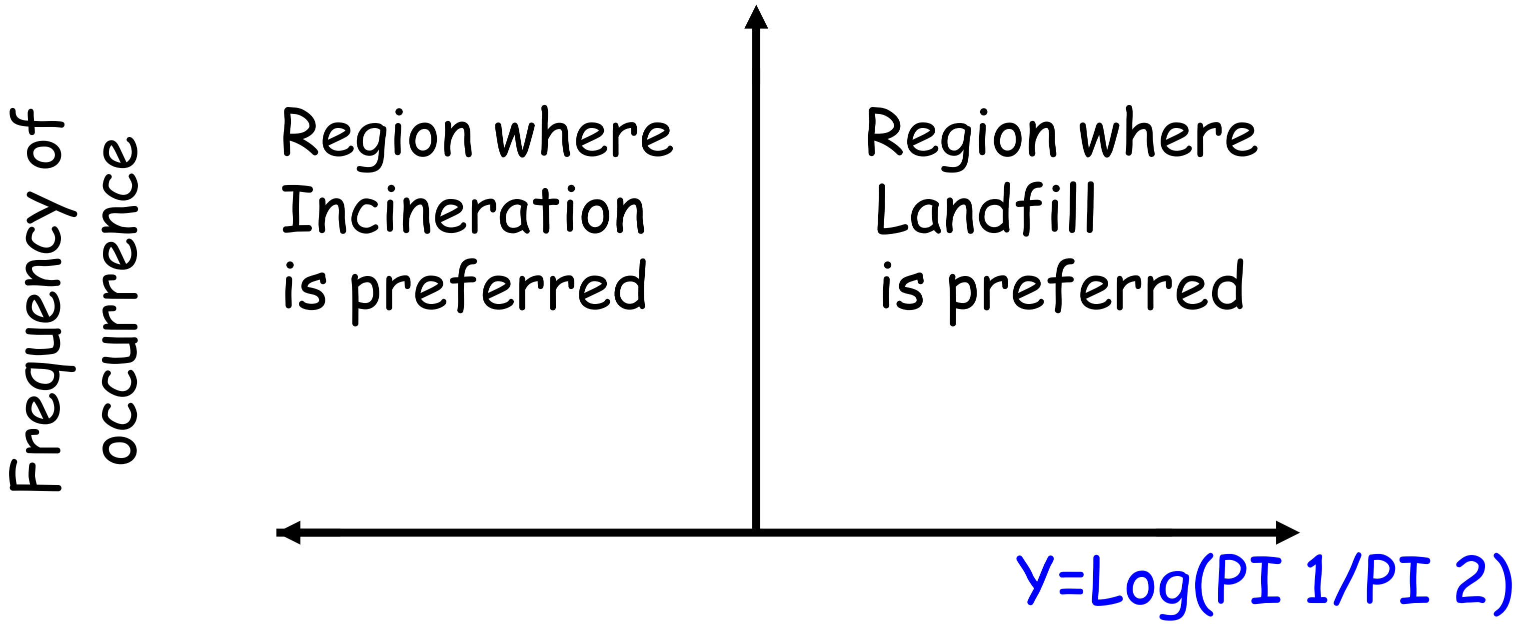
*SH = stakeholders

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?

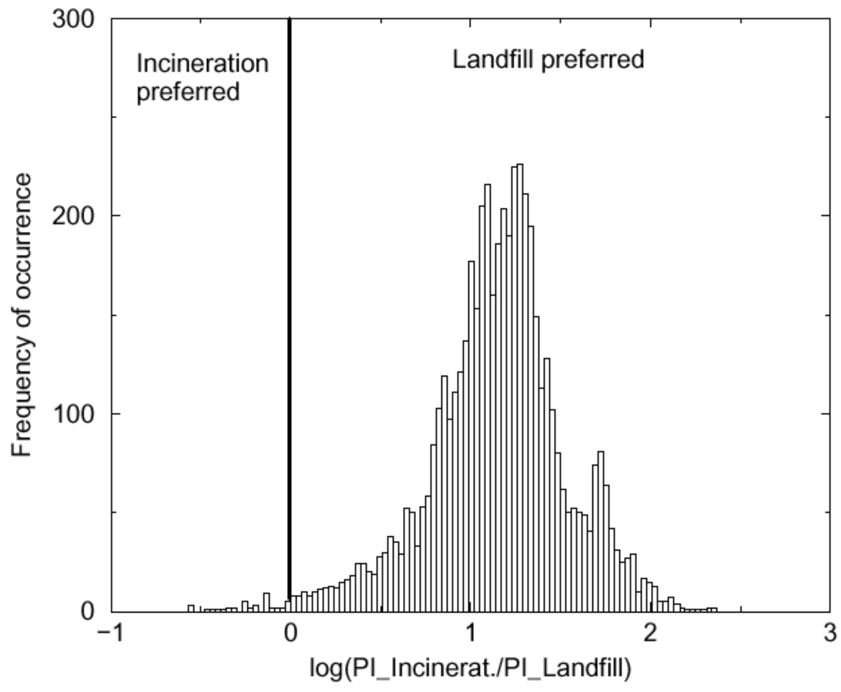
High uncertainty is not the same as low quality

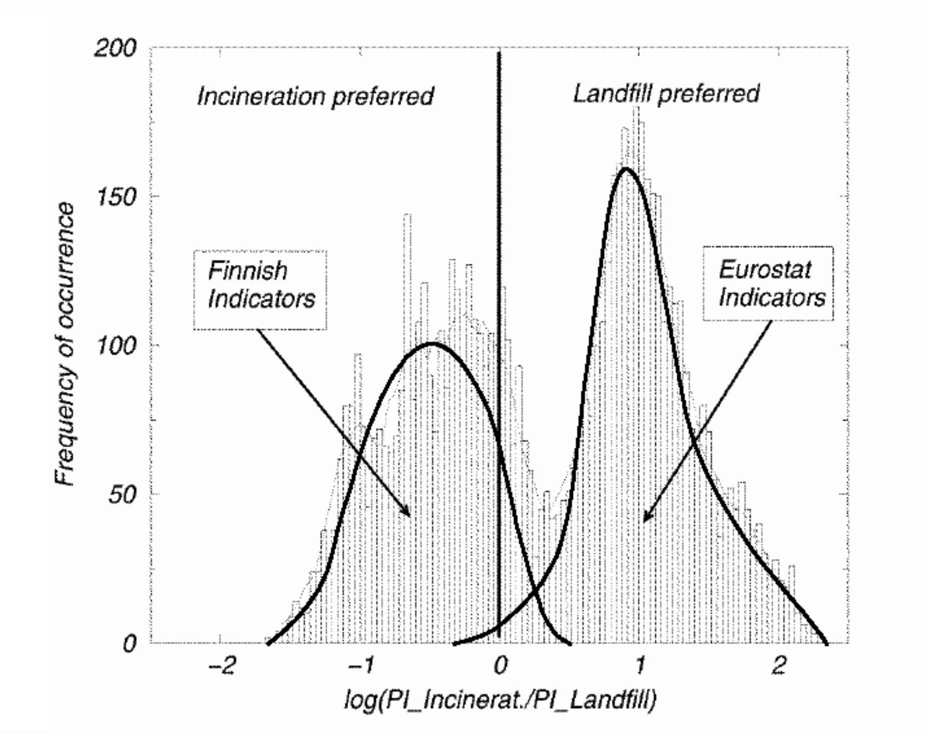
Example: imagine the inference is Y = the logarithm of the ratio between the two pressure-on-decision indices PI1 and PI2



High uncertainty is not the same as low quality,

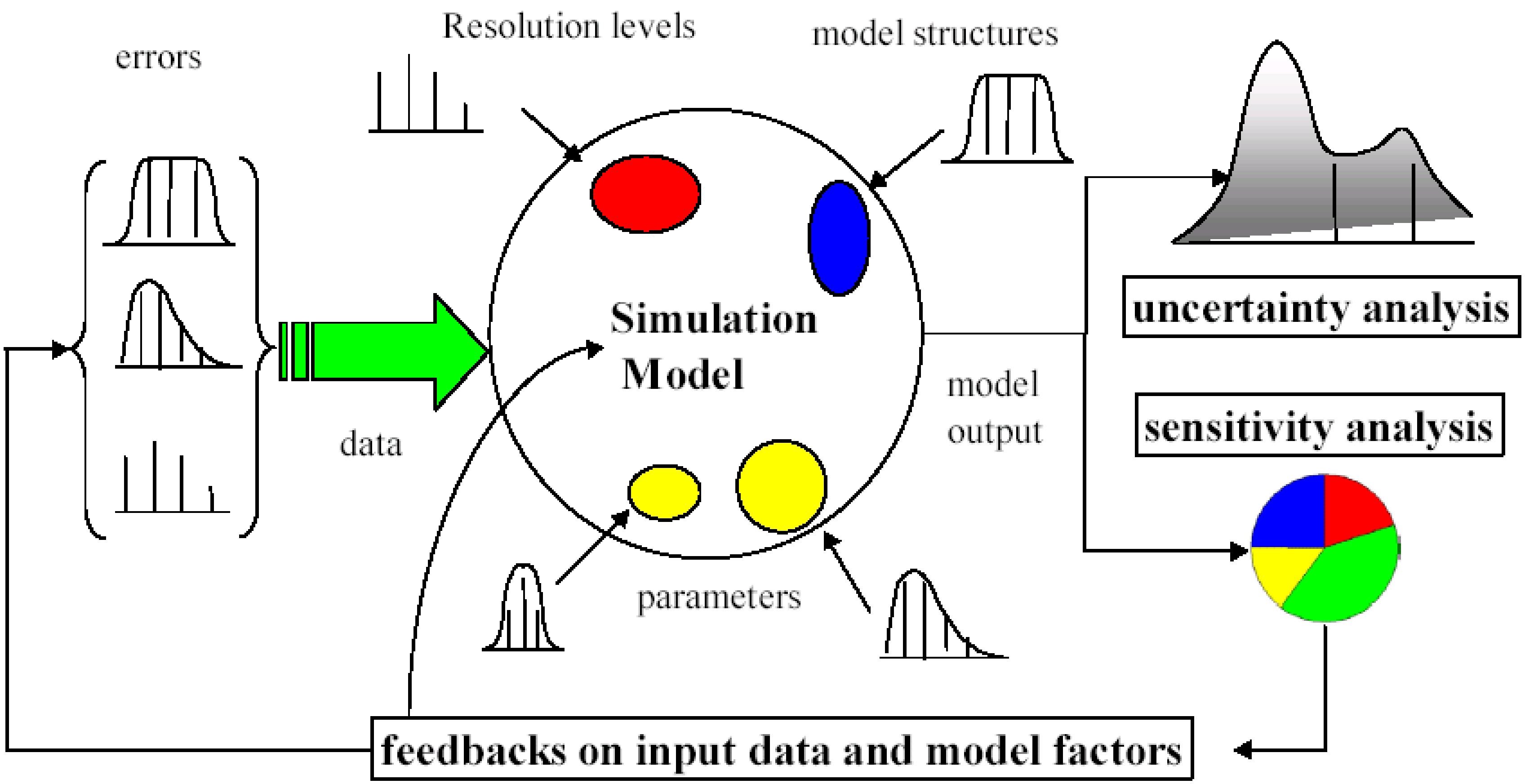
but..... methodological uncertainty can de dominant

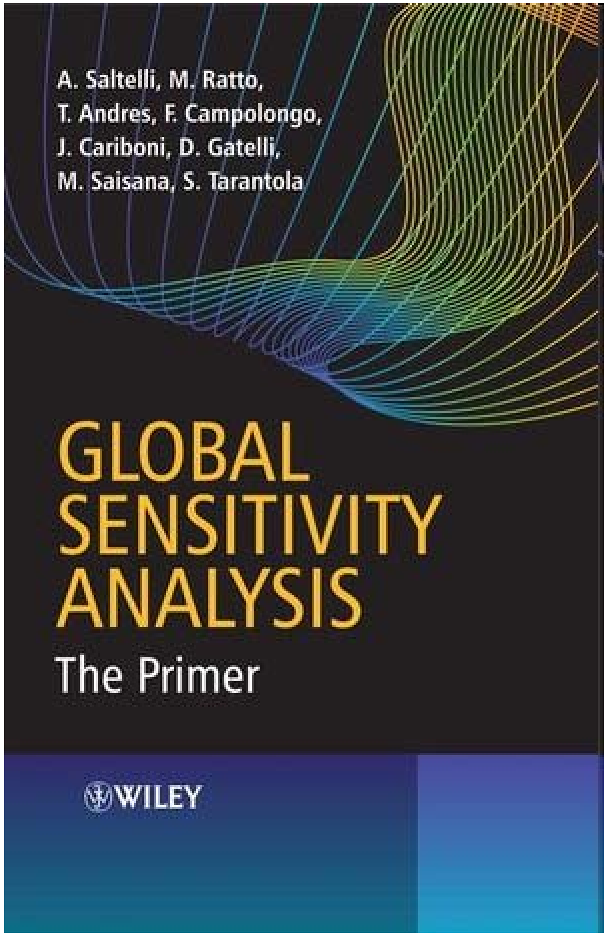


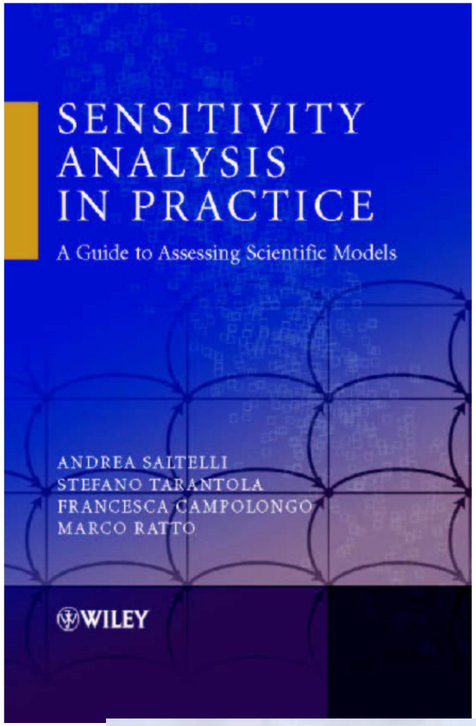


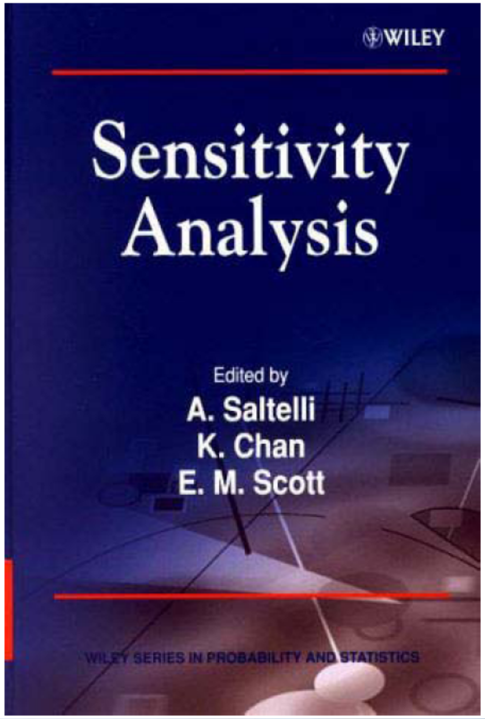
(slide borrowed from Andrea Saltelli)

Uncertainty analysis = Mapping assumptions onto inferences
Sensitivity analysis = The reverse process









Andrea Saltelli

Applied Statistics group at EU Joint Research Centre

Do we know enough to quantify?

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

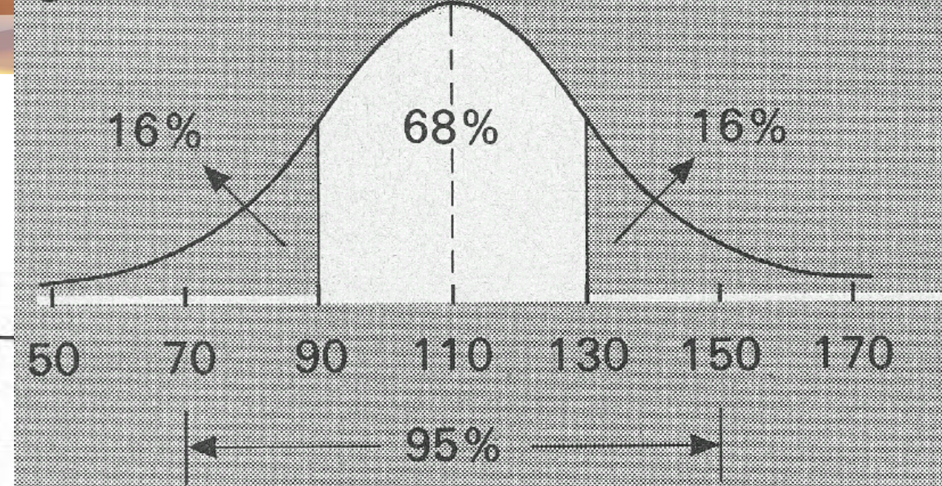
- Full probability density function
  - Robust, well defended distribution
- Bounds
  - Well defended percentile bounds
- First order estimates
  - Order of magnitude assessment
- Expected sign or trend
  - Well defended trend expectation
- Ambiguous sign or trend
  - Equally plausible contrary trend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

Reliability intervals normal distributions

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

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

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



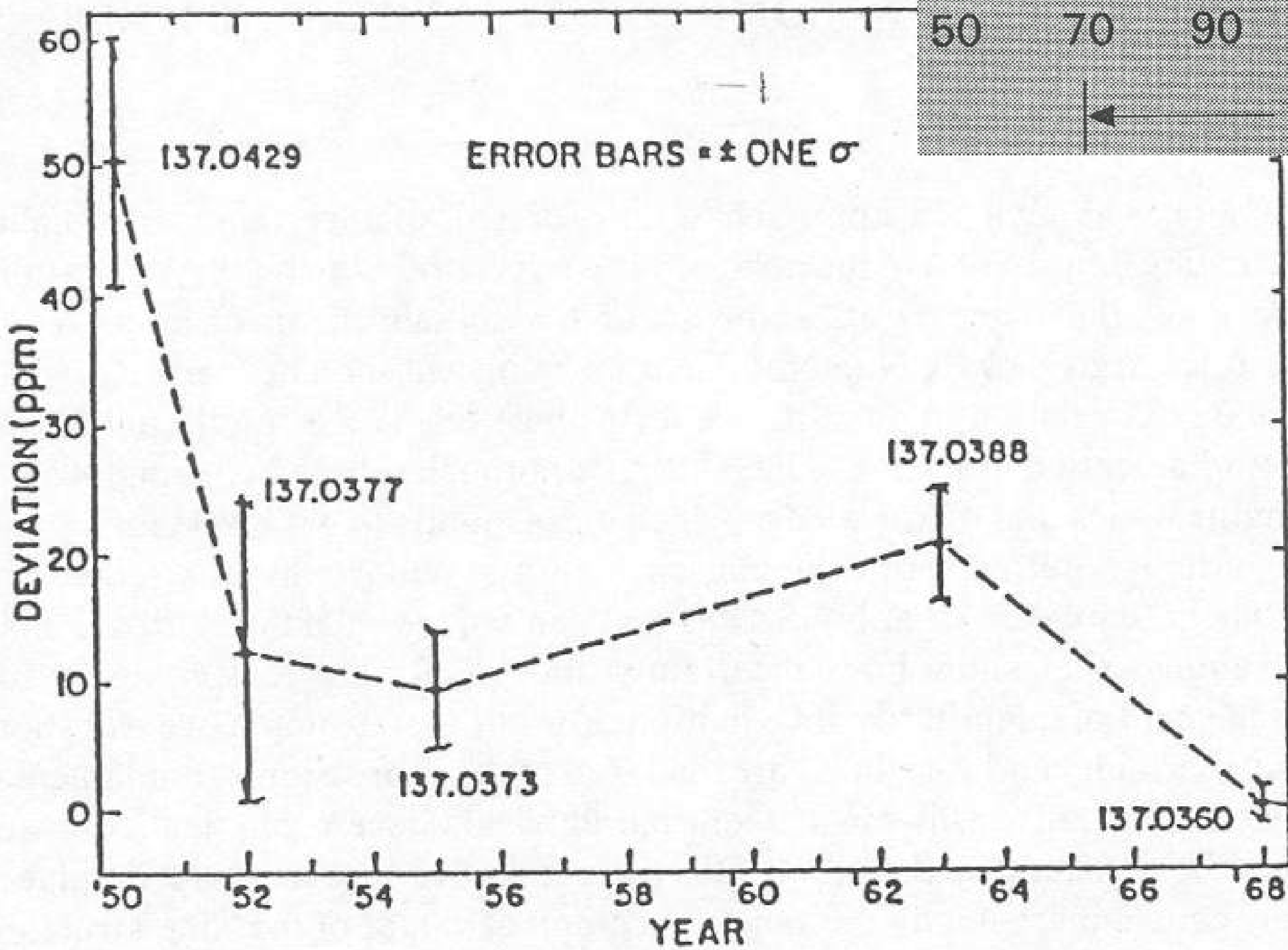
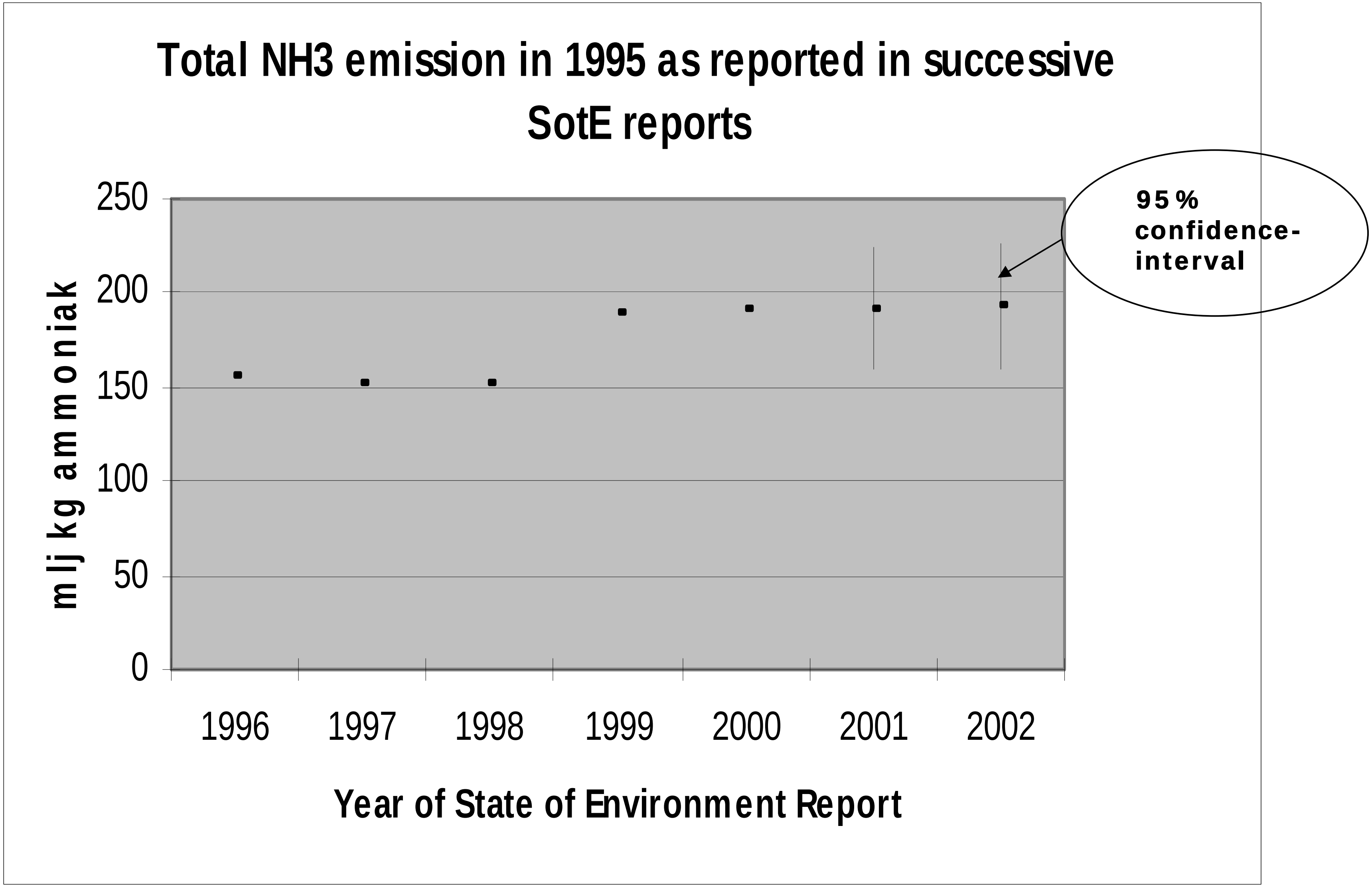


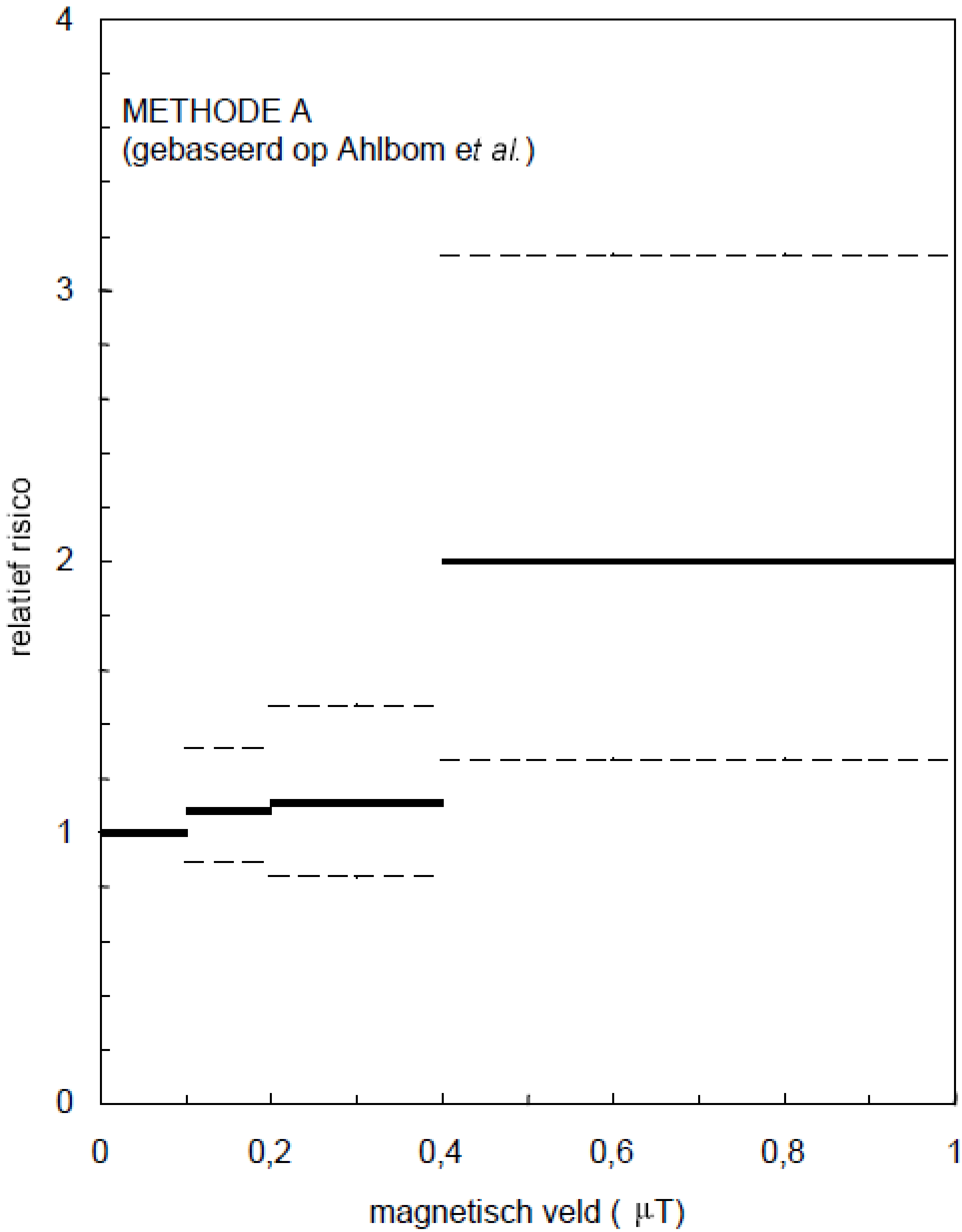
Fig. 1. Successive recommended values of the fine-structure constand α^{-1} (B. N. Taylor *et al.*, 1969, 7)


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

Uncertainty is more than a number

Dimensions of uncertainty:

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



Uncertainty in Relative Risk of child leukemia as a function of magnetic field strength of overhead power lines

Technical:
95% - interval

Methodological:
- Indirect exposure metric
- small sample size
  > 0.4 µT, n=106
- selection bias

Epistemological
- Causality unknown

Societal
- Distrust in outcomes that point at low risks measure

NUSAP

Qualified Quantities

- Numeral
- Unit
- Spread
- Assessment
- Pedigree

(Funtowicz and Ravetz, 1990)

*NUSA**P**:* Pedigree

Evaluates the strength of the number by looking at:

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

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

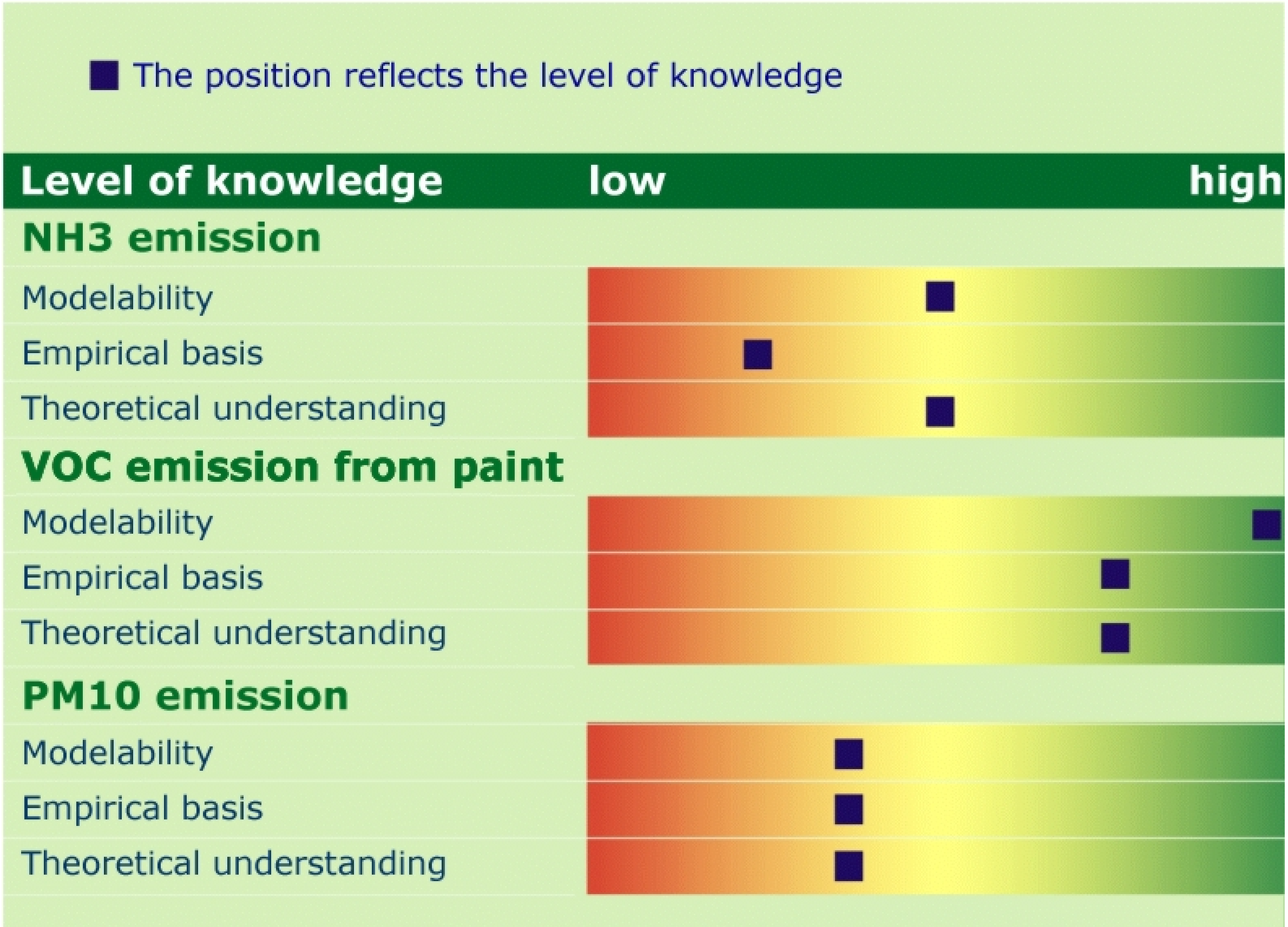
Trafic-light analogy <1.4 red; 1.4-2.6 amber; >2.6 green

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.

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

Example: Air Quality



Mapping and prioritization of relevant uncertainties

- Highlight uncertainties in typology relevant to this problem
- Set priorities for uncertainty assessment
- Select uncertainty assessment tools from the tool catalogue

Typology of uncertainties

- Location
- Level of uncertainty
  - statistical uncertainty, scenario uncertainty, recognised ignorance
- Nature of uncertainty
  - knowledge-related uncertainty, variability-related uncertainty
- Qualification of knowledge base (Pedigree)
  - weak, fair, strong
- Value-ladenness of choices
  - small, medium, large

Locations of uncertainties:

- Context

  ecological, technological, economic, social and political

  representation
- Expert judgement

  narratives, storylines, advices
- Model

  model structure, technical model, model parameters, model inputs
- Data

  measurements, monitoring data, survey data
- Outputs

  indicators, statements

UNCERTAINTY MATRIX		Level of uncertainty (from determinism, through probability and possibility, to ignorance)			Nature of uncertainty		Qualification of knowledge base (backing)			Value-ladenness of choices		
Location ↓		Statistical uncertainty (range+ chance)	Scenario uncertainty (range as 'what-if' option)	Recognized ignorance	Knowledge-related uncertainty	Variability-related uncertainty	Weak –	Fair 0	Strong +	Small –	Medium 0	Large +
Context	Ecological, technological, economic, social and political representation											
Expert judgement	Narratives; storylines; advices											
Model: Model structure	Relations											
Model: Technical model	Software & hardware implementation											
Model: Model parameters												
Model: Model inputs	Input data; driving forces; input scenarios											
Data (in general sense)	Measurements; monitoring data; survey data											
Outputs	Indicators; statements											

Table 1: Uncertainty typology overhead power lines and child leukaemia

Uncertainty characterizations	Nature Epistemic / Ontic	Range Statistical/ Scenario	Recognized ignorance	Methodo-logical Unreliability	Value diversity among analysts
Context uncertainty					
Share overhead power lines in total exposure ELF EMF	E	Sc	++	+	+
The nature of the effect (cancer) is stochastic	O	St			
Model structure uncertainty					
Shape possible exposure-response relationship	E	Sc	+		+
Responsible damaging characteristic(s) ELF EMF	E	Sc	++	++	+
Parameter uncertainty					
Relative Risk as function of magnetic field strength	E/O	St			
Calibration: poor and/or inaccurate measurements; uncertainty about number of exposed children	E	St		+	
Input data uncertainty					
Exposure assessment indirect	E/O	St/Sc	++	++	+
Selection bias	E	St		++	+

Tool catalogue

For each tool:

- Brief description
- Goals and use
- What sorts and locations of uncertainty does this tool address?
- What resources are required to use it?
- Strengths and limitations
- guidance on application & complementarity
- Typical pitfalls of each tool
- References to handbooks, example case studies, web-sites, experts etc.

Tool catalogue

- Sensitivity Analysis
- Error propagation equations
- Monte Carlo analysis
- Expert Elicitation
- Scenario analysis
- NUSAP
- PRIMA
- Checklist model quality assistance
- Assumption analysis
- ……

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

Uncertainty communication criteria

- Meet requirements of good scientific practice
- Audiences should have access to the uncertainty information
- Essential uncertainty information should be located in sections of the report that are most likely to be read by the audiences

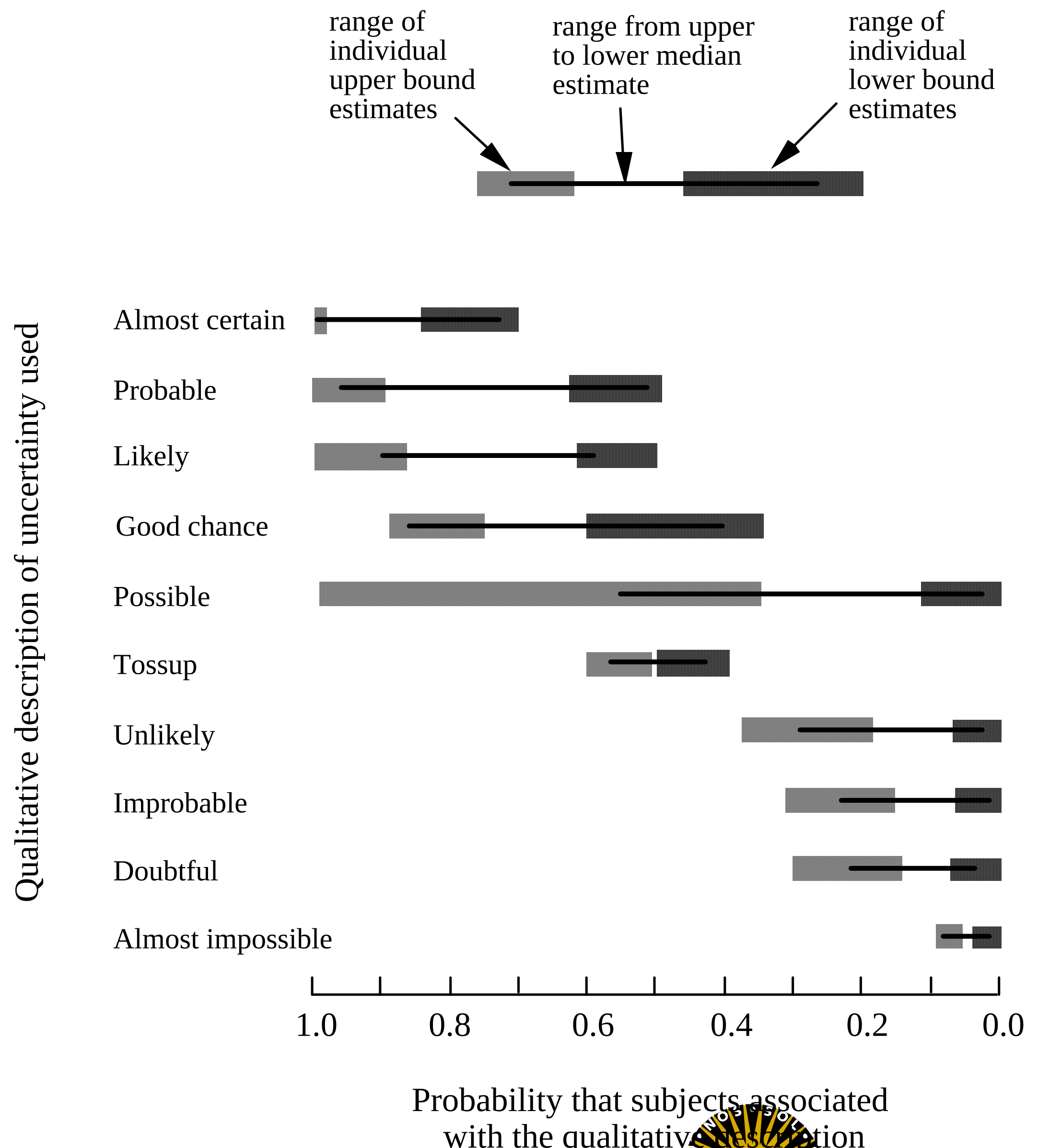
Uncertainty communication criteria
- continued -

- Clear
  - avoid misinterpretation
  - avoid bias
  - avoid differences in interpretation between individuals
- Easy to process and understand
- Meet information needs of the target audiences
- Useful
- Credible

Verbal expression	Chance (per cent)	Chance (fraction)
Virtually certain	More than 99% chance that the result is true	≥ 99 out of 100
Very likely	90–99% chance that the result is true	≥ 9 out of 10 and ≤ 99 out of 100
Likely	66–90% chance that the result is true	≥ 2 out of 3 and ≤ 9 out of 10
Medium likelihood	33–66% chance that the result is true	Between 1 and 2 out of 3
Unlikely	10–33% chance that the result is true	≤ 1 out of 3 and ≥ 1 out of 10
Very unlikely	1–10% chance that the result is true	≤ 1 out of 10 and ≥ 1 out of 100
Exceptionally unlikely	Less than 1% chance the result is true ditto	≤ 1 out of 100

IPCC WGI Proposal for Interpretation and Use of Probabilistic Terms

Words have different meanings for different people…



Read-experiments policy reports Vaessen and Ellis (2000)

- time spent on reading the report is often limited
- most reading time is spent on the primary layer (summary, introduction, conclusions and recommendations, and chapters containing essential answers to the report's topics)
- the index, summary and introduction are often used to select chapters and sections that the reader will read
- selecting sections that the reader will read is also done by browsing chapter, section and paragraph titles
- readers with a negative attitude towards the conclusion of the report read less parts of the text that readers with a positive attitude
- compared to the background layer, more sentences of the primary layer are read
- the information eventually read is quite limited; also important uncertainty information is often not read

Progressive Disclosure of Information

“PDI entails implementation of several layers of information to be progressively disclosed from non-technical information through more specialised information, according to the needs of the user”

(Pereira and Corral, 2002)

	Outer PDI layers	Inner PDI layers
Contents	Uncertainties can be integrated in the message (implicit in the wording used, such as using the word "may" or "might")	Uncertainties mentioned separately and explicitly
	Uncertainties as essential contextual information on the assessment results	Uncertainties as part of scientific accounting on the approach used in the study and on the assessment results
	Uncertainties translated to the political and societal context	Account of the 'bare' uncertainties from a scientific point of view
	Emphasis on policy relevance of uncertainties	Balanced account of uncertainties in all parts of the assessment
	Emphasis on implications of uncertainties	Emphasis on nature, extent and sources of uncertainties
	Implications of uncertainties for the assessment results and the policy advice given	Implications of uncertainties for representativeness of a study, value of the results, and further research
Style	Scientific information translated into 'common language'	Scientific information with a high technical sophistication
	Use of jargon to be avoided	Use of jargon allowed
Degree of detail	Details only if considered policy relevant	Highly detailed (each layer offers more detailed information than the previous PDI layer)

Triggers that increase policy relevance of uncertainty

- High influence on policy advice given
- Indicator outcomes close to a policy goal, threshold or norm
- Indicator outcomes point at serious effects or catastrophic events
- Being wrong in one direction is very different than being wrong in the other when it comes to policy advice
- Controversies among stakeholders are involved
- Value-laden choices and assumptions are in conflict with stakeholder views and interests
- Fright factors/media triggers are involved
- Persistant misunderstandings among audiences
- If audiences are expected to distrust outcomes that point at low risks