ICA-RUS/CCRP-PJ2 International Workshop 2013
Now and Future of Global Climate Risk Management, Tokyo 5 December 2013

Dealing with uncertainty in Integrated Assessment:

The NUSAP approach

Dr. Jeroen P. van der Sluijs

Copernicus Institute of Sustainable Development
Utrecht University

Daily practice of dealing with uncertain science in policy making

Two dominant strategies: uncertainties are either

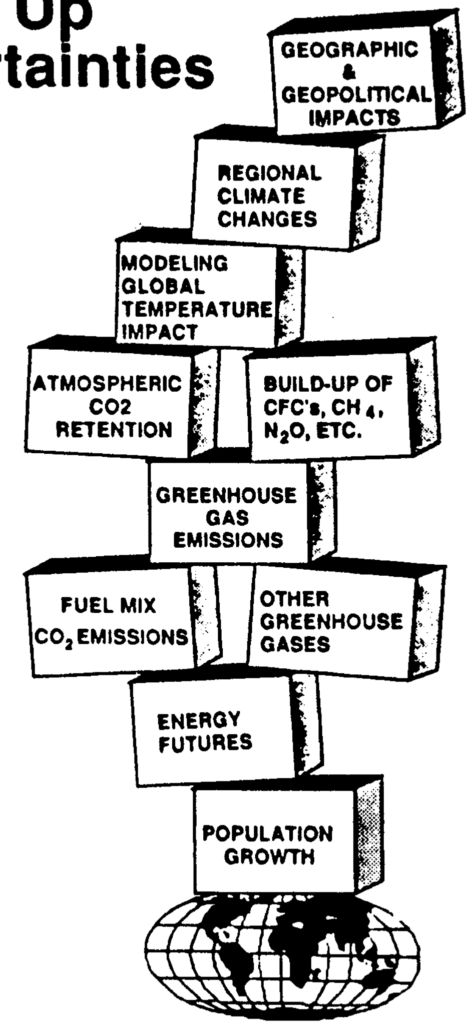
- **downplayed** to promote political decisions (enforced consensus), or
- **overemphasised** to prevent political action

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

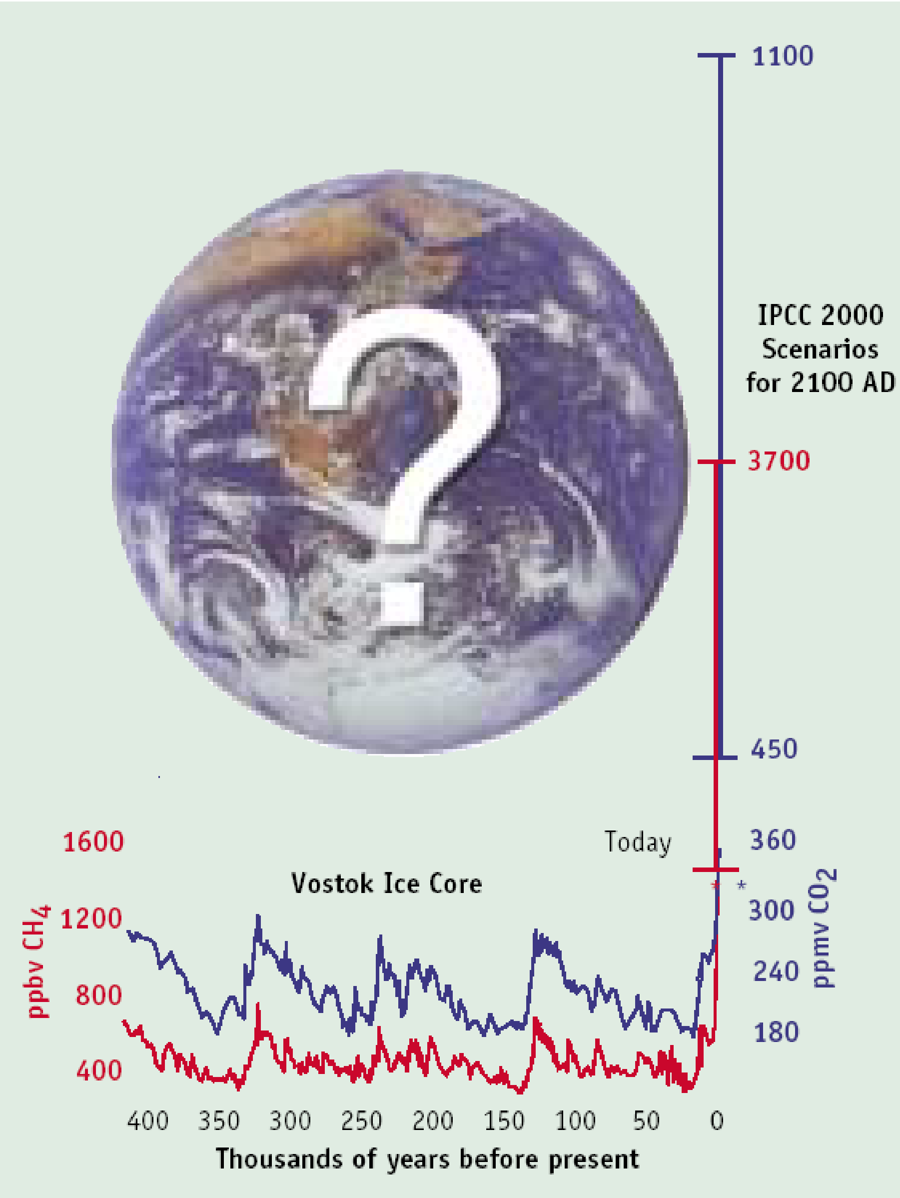
- We need new ways to deal with **uncertainty**, scientific **dissent** & **plurality** in sustainability science.

GLOBAL CLIMATE CHANGE

Piling Up Uncertainties



Sailing into terra incognita?



Complex - *uncertain* - risks

Typical characteristics (Funtowicz & Ravetz):

- Decisions needed **before conclusive scientific evidence** is available;
- Potential **impacts of ‘wrong’ decisions** can be huge
- **Values in dispute**
- Knowledge base characterized by large (partly irreducible, largely **unquantifiable**) uncertainties, multi-causality, knowledge gaps, and **imperfect understanding**
- More research ≠ less uncertainty; unforeseen complexities!
- Assessment dominated by **models**, scenarios, assumptions, extrapolations
- Many (hidden) **value loadings** reside in problem frames, indicators chosen, assumptions made

A practical problem:

Protecting a strategic fresh-water resource

5 scientific consultants 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)

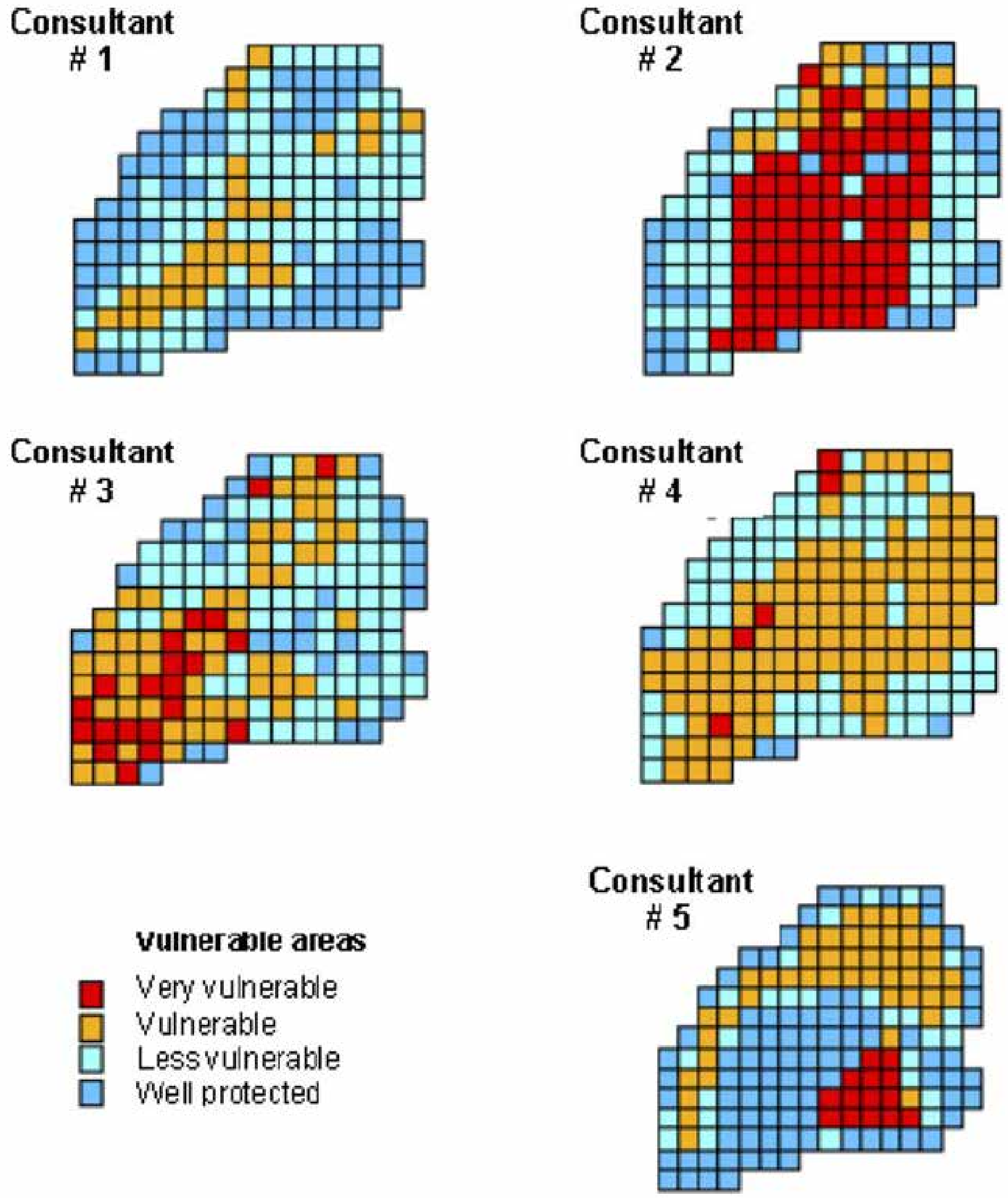


Fig. 1. Model predictions on aquifer vulnerability towards nitrate pollution for a 175 km^2 area west of Copenhagen [11].

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
- Uncertainty can be result of production of knowledge
- Acknowledge that not all uncertainties can be quantified
- Openly deal with deeper dimensions of uncertainty (problem framing indeterminacy, ignorance, assumptions, value loadings, institutional dimensions)
- *Tools:* Knowledge Quality Assessment
  - ***Working deliberatively within imperfections***

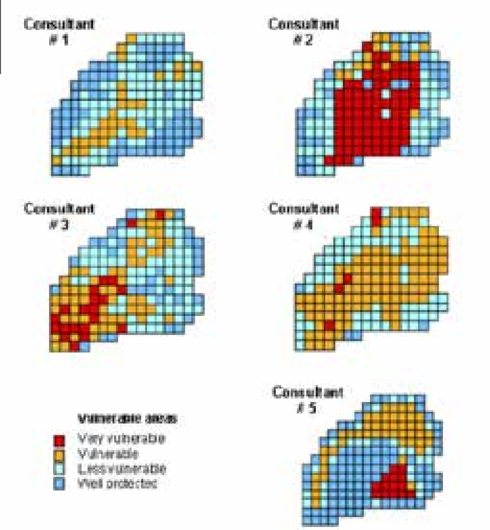


Fig. 1. Model predictions on aquifer vulnerability towards nitrate pollution for a 175 km² area west of Copenhagen [11].

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**

Former chairman IPCC on objective to reduce climate uncertainties:

- "*We cannot be certain that this can be achieved easily and we do know it will take time. Since a fundamentally chaotic climate system is predictable only to a certain degree, our research achievements will always remain uncertain.* ***Exploring the significance and characteristics of this uncertainty is a fundamental challenge to the scientific community.***" (Bolin, 1994)

[Prof. Bert Bolin, 15 March 1925 – 30 December 2007]

Golden rules for sensitivity auditing of models

1. **Check against rhetoric** use of mathematical modelling;
2. Adopt an **'assumption hunting' attitude**;
3. Detect Garbage In Garbage Out (**GIGO**), extended definition Funtowicz and Ravetz (1990) *"… where uncertainties in inputs must be suppressed lest outputs become indeterminate."*;
4. **Find sensitive assumptions** before these find you;
5. Aim for **transparency**;
6. **Do the right sums**;
7. **Focus** the analysis **on** the **key question** answered by the model, **exploring** holistically the **entire space of the assumptions**.

(Saltelli, Pereira, Van der Sluijs, Funtowicz, in press)

http://arxiv.org/abs/1211.2668

Uncertainty is more than a number

Dimensions of uncertainty:

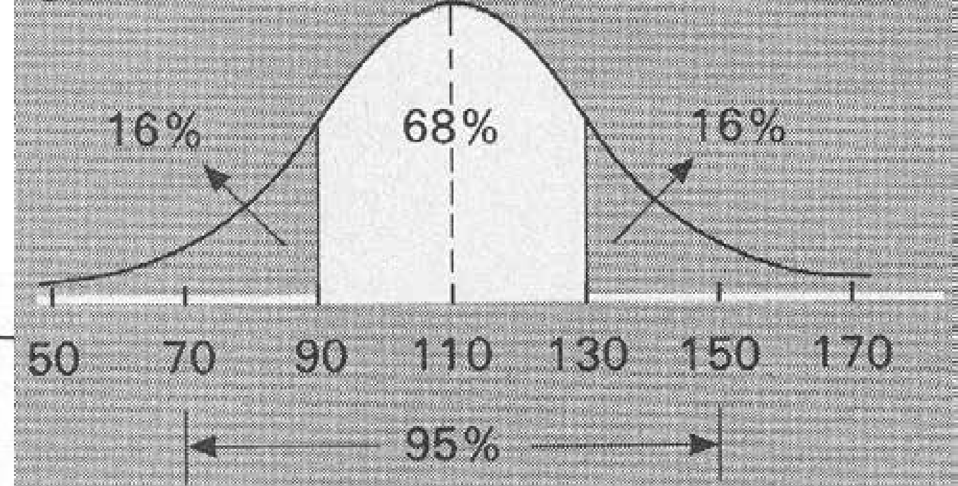
- Technical (inexactness)
- Methodological (unreliability)
- Epistemological (ignorance)
- Societal (limited social robustness)

Reliability intervals normal distributions

$\pm s$ = 68 %

$\pm 2s$ = 95 %

$\pm 3s$ = 99.7 %



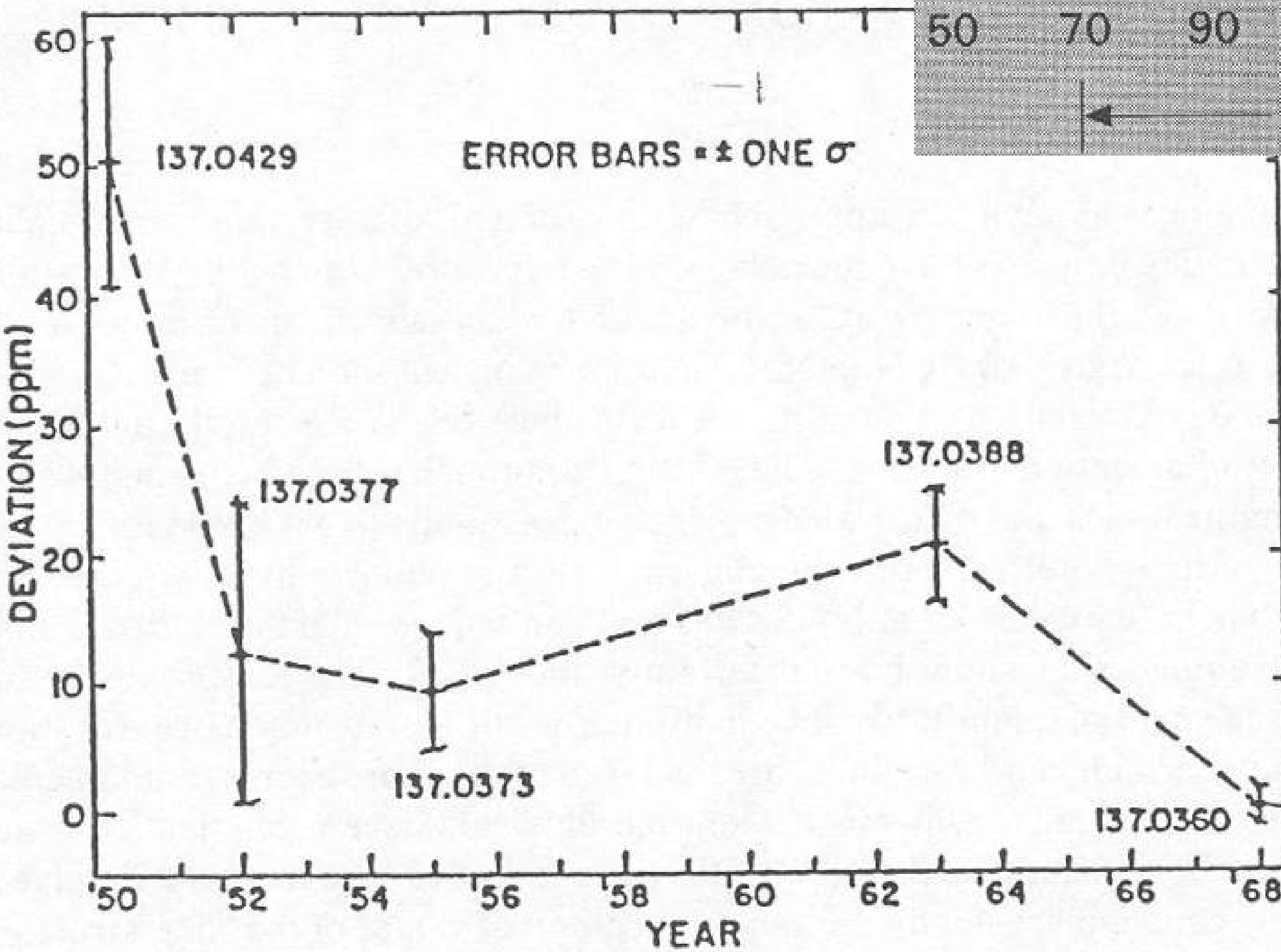
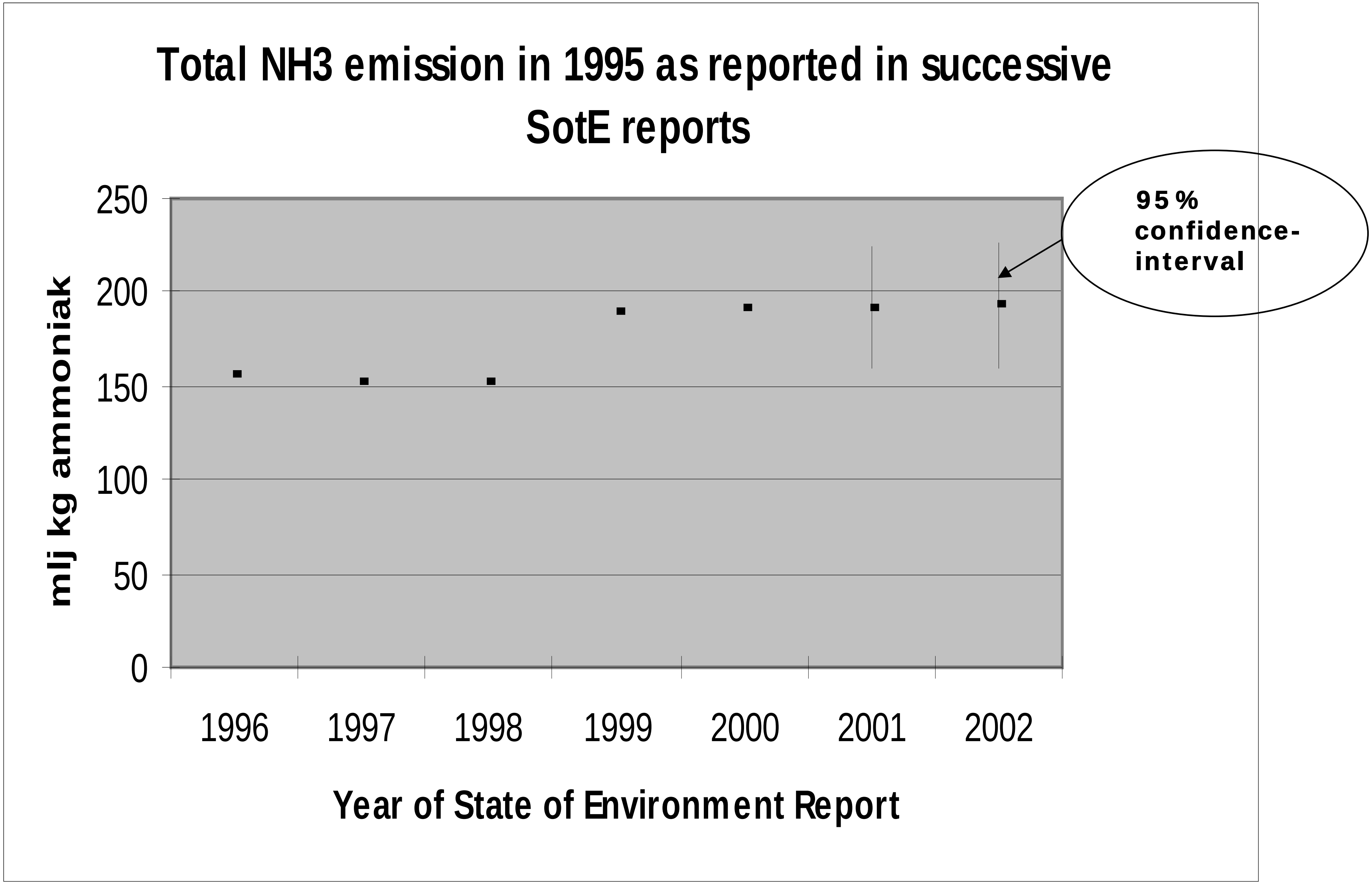


Fig. 1. Successive recommended values of the fine-structure constand α^{-1} (B. N. Taylor *et al.*, 1969, 7)


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

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

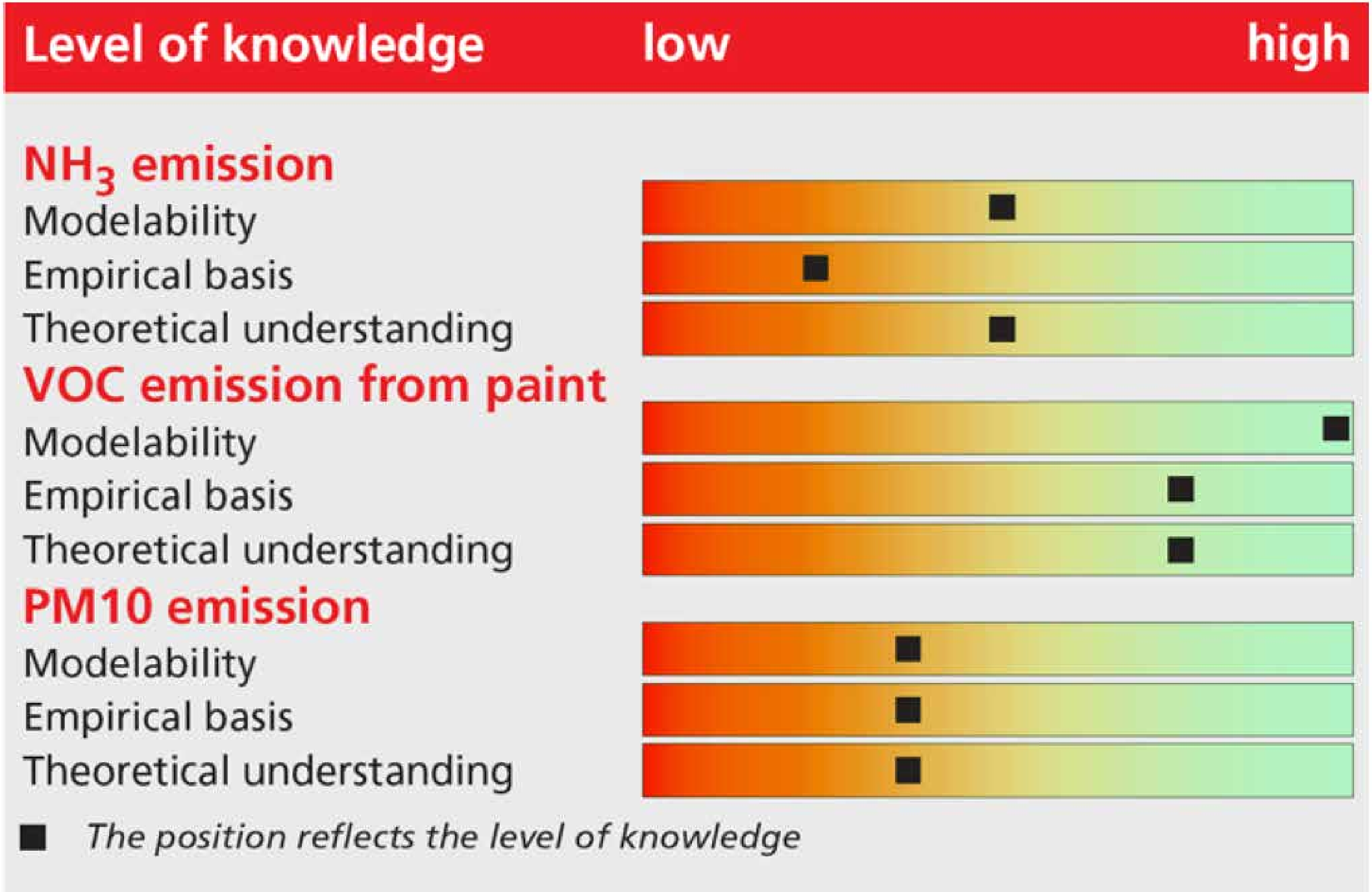
"Pedigree" expresses multi-criteria evaluation of the strength of a number by looking at:

- 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: Air Quality



http://dx.doi.org/10.1088/1748-9326/3/2/024008

Case 1

The IMAGE/TIMER B1 scenario

http://www.nusap.net/workshop/report/finalrep.pdf

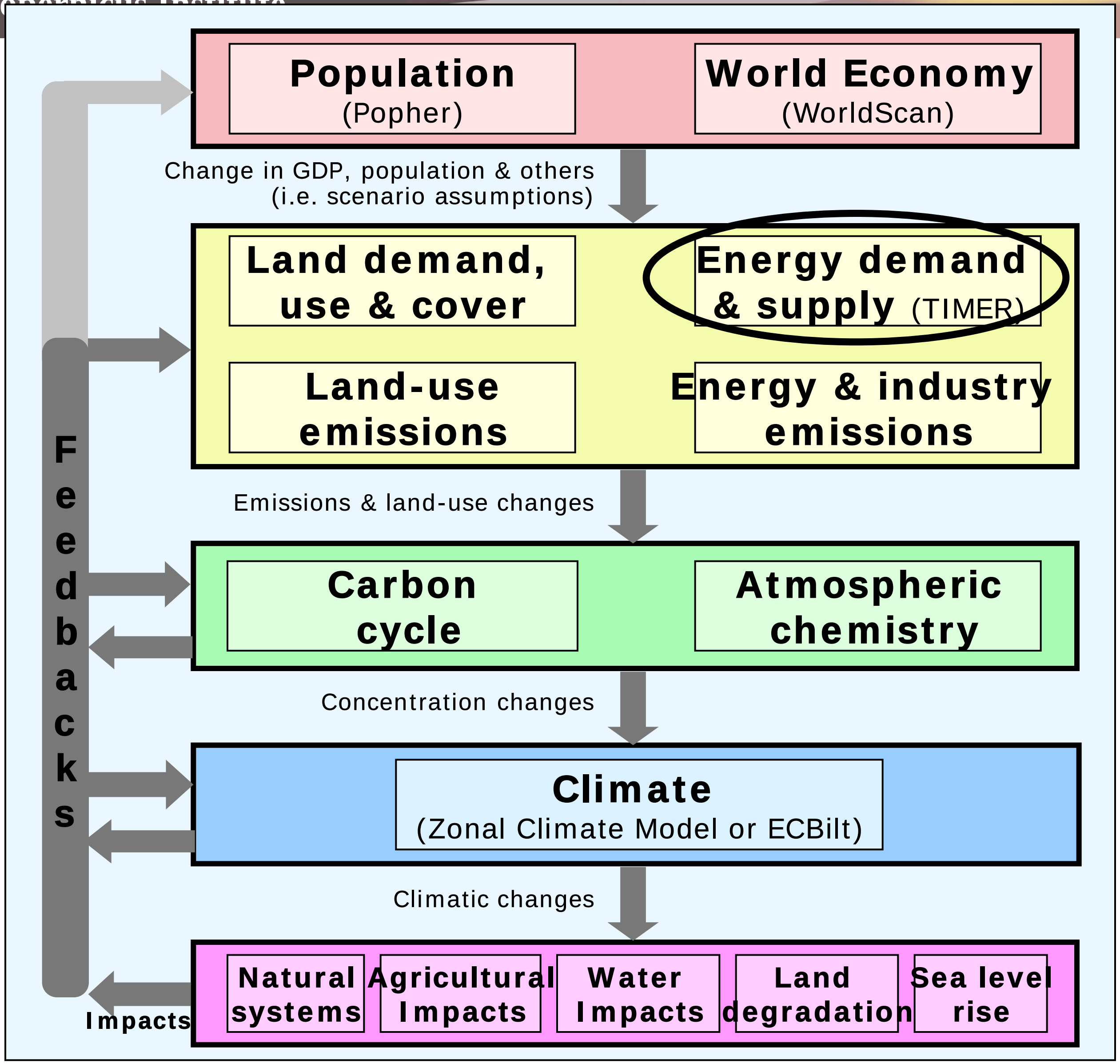
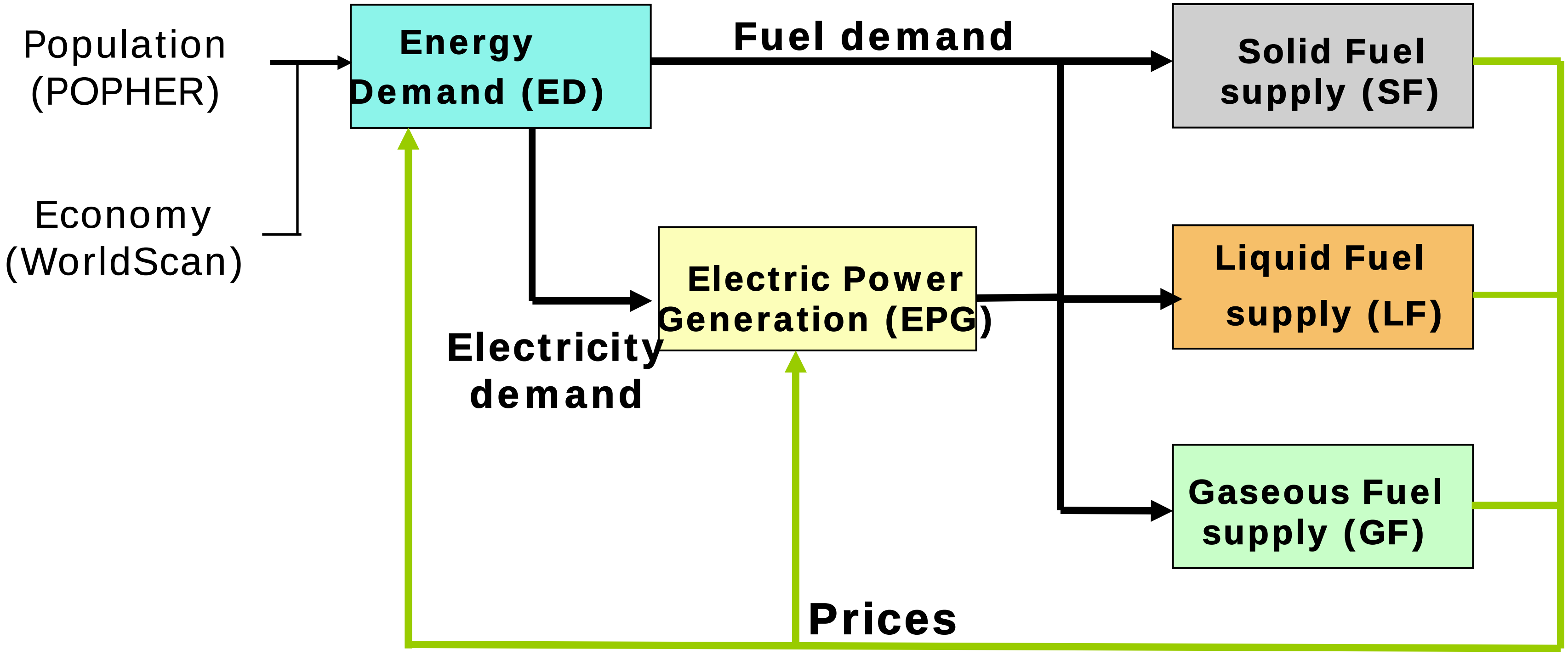


IMAGE 2: Framework of models and Linkages

TIMER Model : five submodels



Inputs: Population, GDP capita^{-1}, activity in energy sectors, assumptions regarding technological development, depletion and others.

Outputs: End-use energy consumption, primary energy consumption.

Applying NUSAP to a complex model was quite a challenge

TIMER model:

- 300 variables
- 19 world regions
- 5 economic sectors
- 5 types of energy carriers
- 2 forms of energy
- some are time series
  - Þ about 160,000 numbers

Morris (1991)

- facilitates global sensitivity analysis in minimum number of model runs
- covers entire range of possible values for each variable
- parameters varied one step at a time in such a way that if sensitivity of one parameter is contingent on the values that other parameters may take, Morris captures such dependencies

Most sensitive model components:

- Population levels and economic activity
- Intra-sectoral structural change
- Progress ratios for technological improvements
- Size and cost supply curves of fossil fuels resources
- Autonomous and price-induced energy efficiency improvement
- Initial costs and depletion of renewables

Parameter Pedigree

- Proxy
- Empirical basis
- Theoretical understanding
- Methodological rigour
- Validation

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

Elicitation workshop

- Focussed on 40 key uncertain parameters grouped in 18 clusters
- 18 experts (in 3 parallel groups of 6) discussed parameters, one by one, using information & scoring cards
- ***Individual*** expert judgements, informed by group discussion

Structural change / Growth elasticity	**Sub module:** dem

Definition: These parameters describe the structural change curve. When an economy grows it is assumed to go through successive stages of development. In TIMER, based on historic analysis, that this is also reflected in terms of the demand for energy services in different energy end-use sectors. For instance, in early stages of development the industry sector is dominated by light industry; in a next stage heavy industry dominates and finally industry with high-value added. Consequently the energy intensity of a economy is assumed to go through a maximum with increasing GDP per capita (at PPP). In TIMER, the structural change formulation can be characterised by two important parameters:

Position maximum: Position of the maximum in the GDP per capita (at PPP) vs energy intensity curve

Saturation level: This parameter represents a theoretical minumum in energy intensity, associated with a saturation in energy demand per capita as a function of GDP per capita (at PPP). Note that this saturation point is assumed to be strongly scenario dependent. In a A-storyline the saturation is not met before 2100, in a B storyline it is.

Background Information:

	B1 range:	**Range over which sensitivity was tested:**
Position maximum:	1189.22, 1.0E+05 1995US$	100.00, 1.0E+05
Saturation level:	0, 3.5E-03 GJ/1995 US$	0, B1 value +50%

Rank in Morris Sensitivity Analysis (maximums are listed from this group of parameters)

Grouped by	Rank	μ(μ)	σ(μ(μ))	μ(σ)
Type:	1	873%	587%	2008%
Module	1	423%	278%	1051%

Dimension	17 Regions	5 Sectors	heat/electricity	5 energy carriers	Other
Variable	x	x	x		

Likely Uncertainty Range: *Maximum:* ± ______ % *Saturation:* ± ______ %

Characterization of variable

		0	1	2	3	4		Elaboration/justification
Value-ladenness	**Negligible**						**High**	

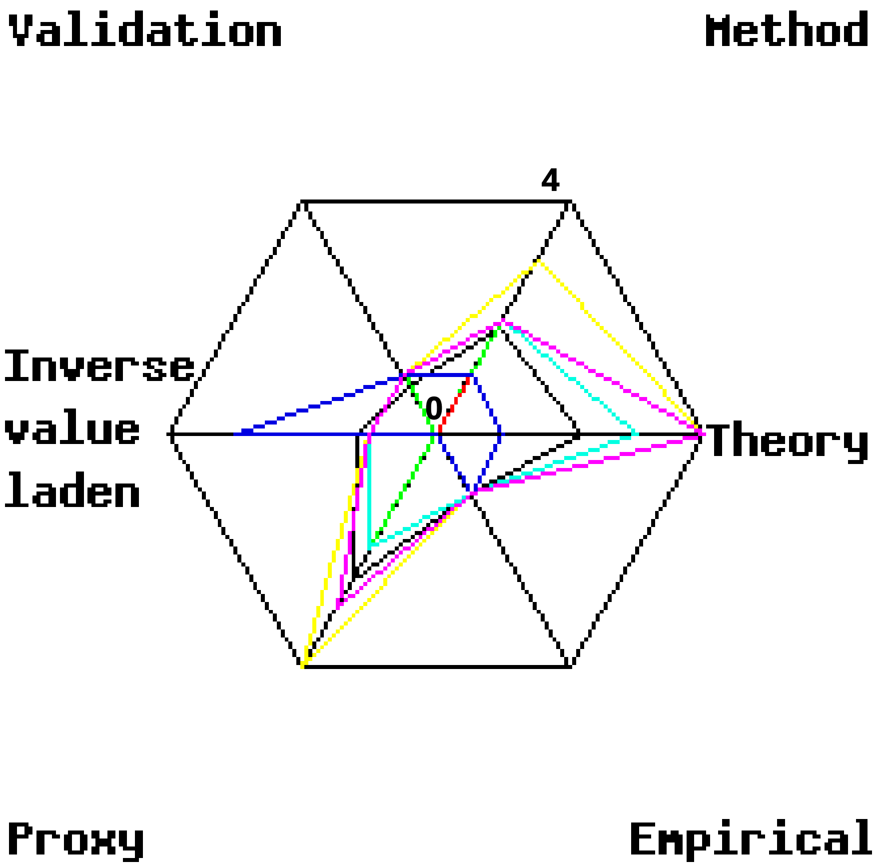
Pedigree

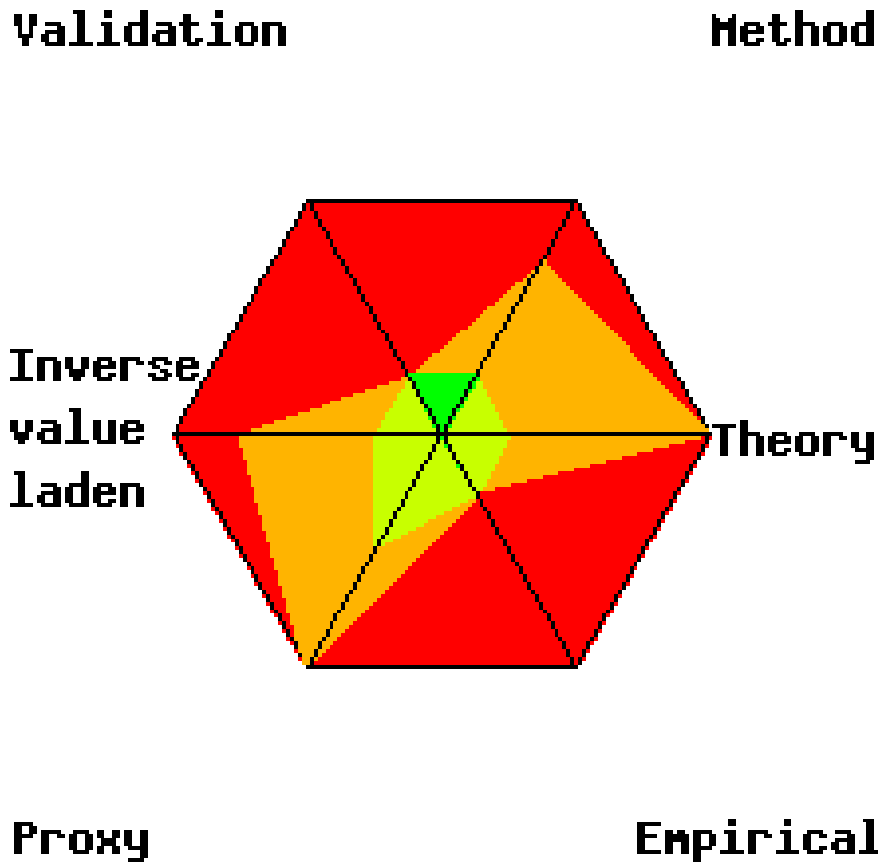
		0	1	2	3	4		Elaboration/justification
Proxy	**Not Related**						**Exact Measure**	
Empirical basis	**Weak**						**Strong**	
Theoretical understanding	**Weak**						**Strong**	
Methodological rigour	**Low**						**High**	
Validation	**No**						**Complete**	

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 card at a time
- If you feel you cannot judge the pedigree scores for a given parameter, leave it blank

Example result *gas depletion multiplier*





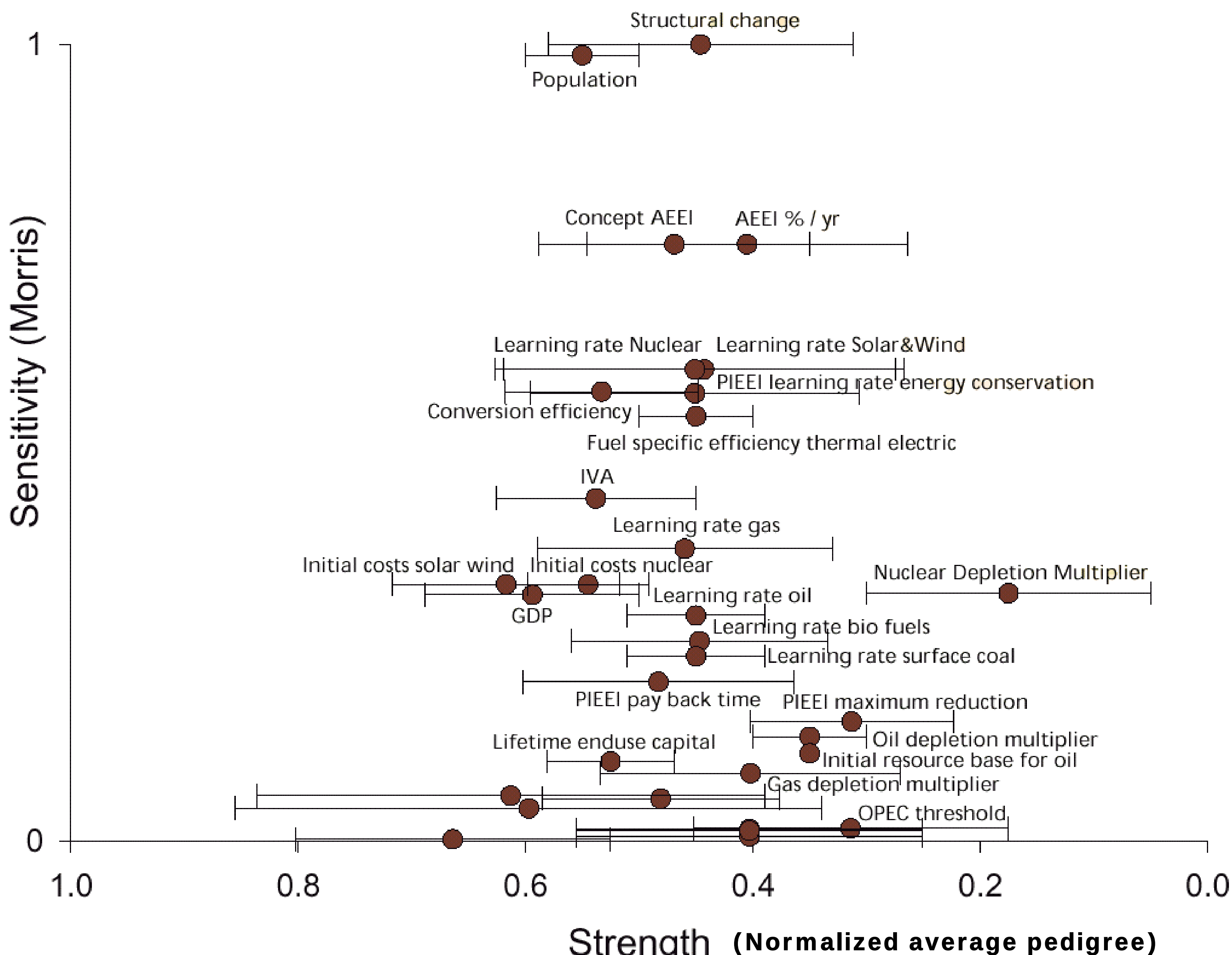
Radar diagram:
Each coloured line represents scores given by one expert

Same data represented as kite diagram:
Green = min. scores, Amber= max scores,
Light green = min. scores if outliers omitted
(Traffic light analogy)

Average scores (0-4)

• proxy	2½	± ½
• empirical	2	± ½
• theory	2	± ½
• method	2	± ½
• validation	1	± ½
• valueladeness	2½	± 1
• competence	2	± ½

Diagnostic Diagram



Case 2
Chains of models

Baakse Beek

local climate impacts for adaptation policy making

Assumptions in model chains

Analysis

1. Identify explicit and implicit assumptions in the calculation chain.

2. Identify and prioritize key-assumptions in the chain.

3. Assess the pedigree of key-assumptions.

4. Identify 'weak' links in the calculation chain.

5. Further analyze the potential value-ladenness of key assumptions.

Revision

6. Revise/extend assessment:

– sensitivity analysis of key-assumptions;

– diversification of assumptions;

– different choices in chain.

Communication

7. Communication:

– key-assumptions;

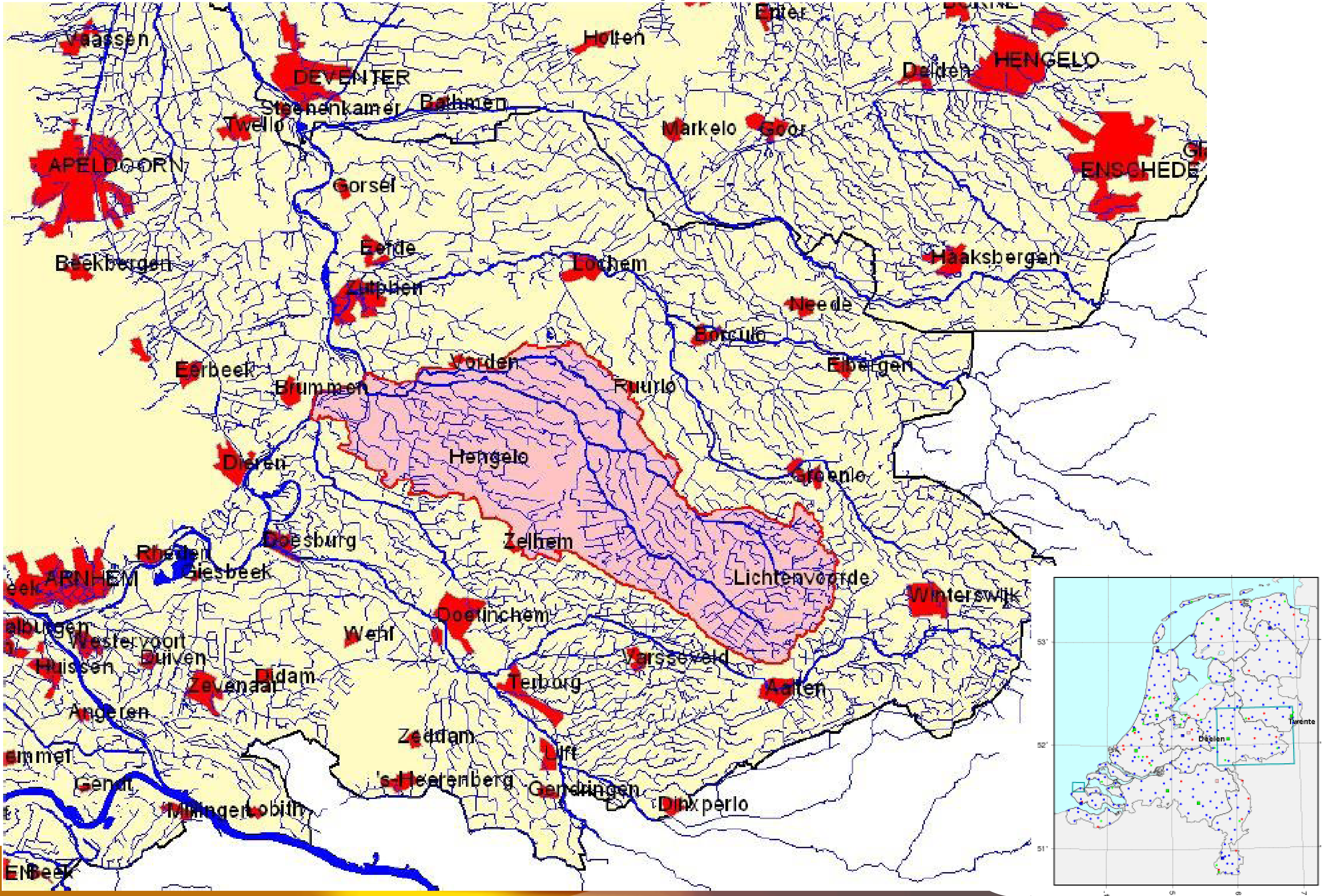
– alternatives and underpinning of choices regarding assumptions made;

– influence of key-assumptions on results;

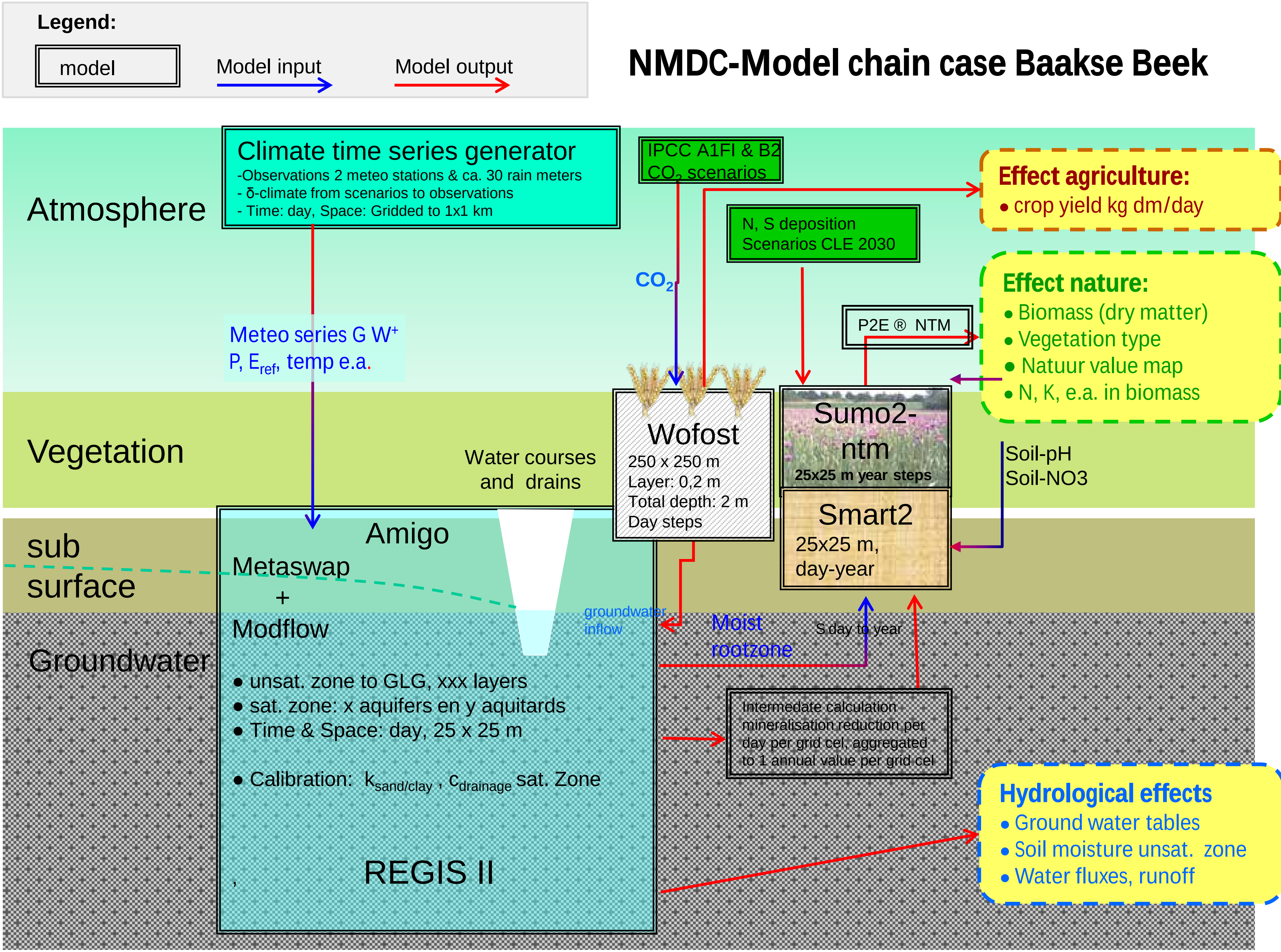
– implications in terms of robustness of results

Kloprogge et al., 2011 http://dx.doi.org/10.1016/j.envsoft.2009.06.009

Baakse Beek area, The Netherlands



NMDC-Model chain case Baakse Beek



Identifying assumptions: think of...

- (over-) Simplifications of reality;
- Up / down scaling in the coupling of models;
- Variables kept constant (in time and space) in the model that vary in reality;
- Feedbacks excluded in the analysis;
- Processes kept outside the system boundary;
- Major sources of uncertainty.

Prioritization & critical appraisal

- 52 assumptions identified: “Gross list”
- 16 respondents each selected a top 10
- Aggregated into a “group top 10”

- Pedigree analysis (“strengths and weaknesses in the underpinning”) of each assumption in “group top 10”

Table 1
Pedigree scheme used to assess assumptions during the workshop.

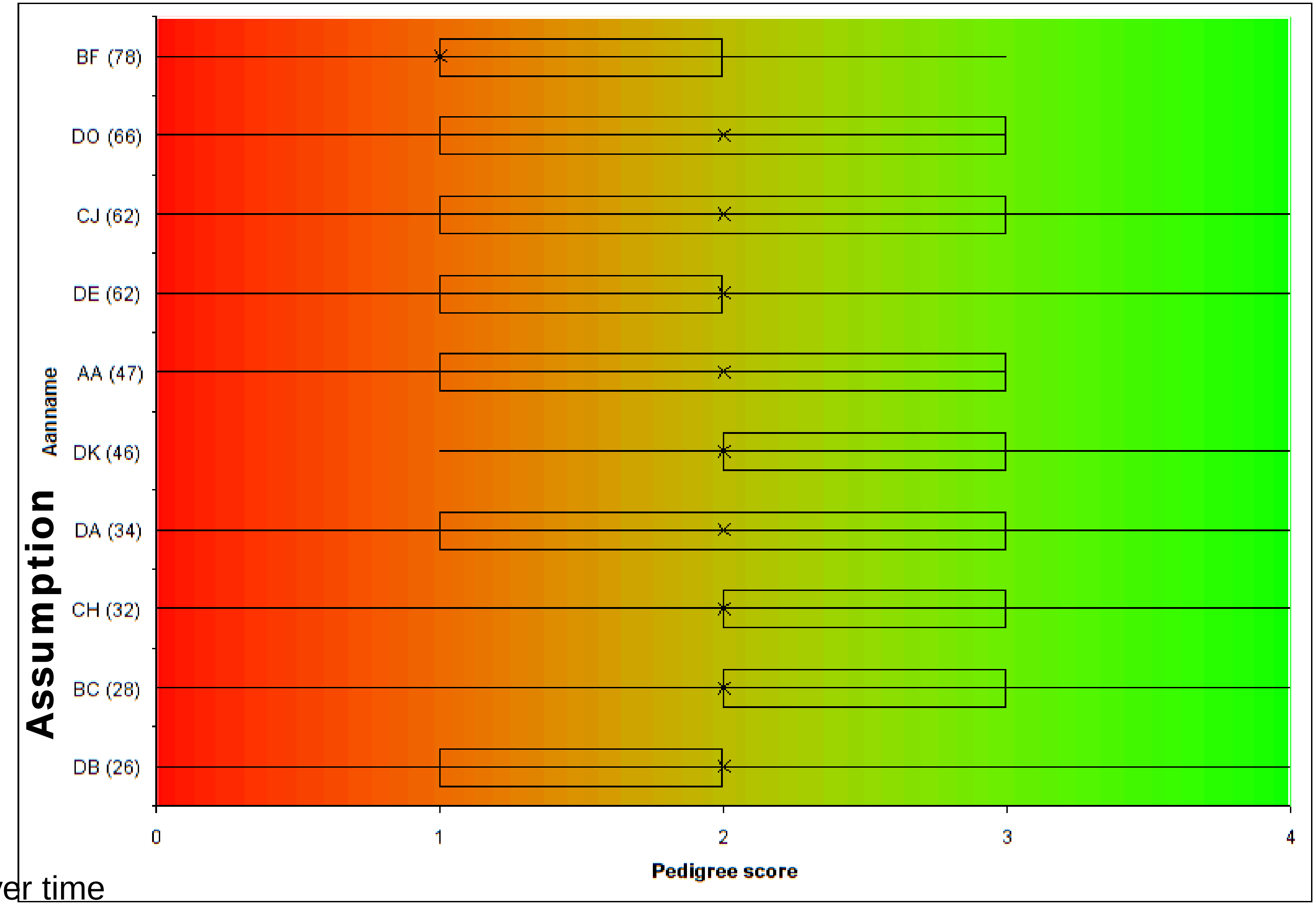
Score	Influence of situational limitations	Plausibility	Choice space	Agreement among peers	Agreement among stakeholders	Sensitivity to views of analyst	Influence on results
4	No such limitations	Very plausible	No alternatives available	Complete agreement	Complete agreement	Not sensitive	Little or no influence
3	Hardly influenced	Plausible	Very limited number of alternatives	High degree of agreement	High degree of agreement	Hardly sensitive	Local impact in the calculations
2	Moderately influenced	Acceptable	Small number of alternatives	Competing perspectives	Competing perspectives	Moderately sensitive	Important impact in a major step in the calculation
1	Importantly influenced	Hardly plausible	Average number of alternatives	Low degree of agreement	Low degree of agreement	Highly sensitive	Moderate impact on end result
0	Completely influenced	Fictive or speculative	Very ample choice of alternatives	Controversial	Controversial	Extremely sensitive	Important impact on end result

Example result pedigree scores

for one of the assumptions

Criteria		Number of votes for pedigree score						Median
		4	3	2	1	0		
a. Influence situational limitations	No such limitations			2	3	1	Completely influenced	1
b. Plausibility	Very plausible		1	4	1		Fictive or speculative	2
c. Choice space	No alternatives		1	4	1		Very ample choice of alternatives	2
d. Agreement among peers	Complete agreement					6	Low degree (controversial)	0
e. Agreement stake-holders	Complete agreement			6			Controversial	2
f. Sensitivity views and interests analyst	Not sensitive					6	Very highly sensitive	0
Total median pedigree score								1.5

g. Influence on results	Little or no influence					6	Important impact on end result	0



BF: land use constant over time
DO: drought stress within one year does not impact nature
CJ: feedbacks via market effects excluded
DE: Model coupling AMIGO-SMART2/SUMO2 around root zone
AA: Completeness of range of climate scenario's
DK: Coupling vegetation and hydrology
DA: Feedbacks via pests, weeds and plant diseases
CH: Developments in crop growth technologies not accounted for
BC: Conductivity of sub surface too homogeneous in the model
DB: Aggregation of daily values Amigo-hydrology to annual number for mineralisation reduction in SMART2/SUMO2

Open Access E-learning course uncertainty assessment

https://proxy.reeds.uvsq.fr/broceliande/KQA?q=node/1584/1584/pathway

Home Page ›› List of Modules ›› KQA ›› Discovering Uncertainty and Knowledge Quality Assessment ›› Expert Elicitation for Uncertainty Quantification

Contact: Brocéliande team | Login | Search

Pathways available

Discovering Uncertainty and Knowledge Quality Assessment

Discovering Uncertainty and Knowledge Quality Assessment

- Science, Policy & Post Normal Science
- Illustrative Examples
- Uncertainty concepts
- Quality
- **Uncertainty assessment tools**
  - *Sensitivity Analysis*
  - *Error propagation equation*
  - *Monte Carlo Analysis*
  - ***Expert Elicitation for Uncertainty Quantification***
  - *NUSAP: Numeral Unit Spread Assessment Pedigree*
  - *Extended Peer Review (review by stakeholders)*
  - *The Uncertainty Guidance: a checklist based approach to uncertainty assessment*
  - *Checklist for model quality assistance*
  - *A method for critical review of assumptions in model-based assessments*
  - *Scenario Analysis*

Expert Elicitation for Uncertainty Quantification

Expert elicitation is a structured process to elicit subjective judgemen
quantitative risk analysis to quantify uncertainties in cases where the
data available to infer on uncertainty. Usually the subjective judgeme
probability density function (PDF) reflecting the expert's degree of be
Expert elicitation in the context of uncertainty quantification aims at a
specifying probabilistic information regarding uncertainty, in a structu
is applied in situations where there is scarce or insufficient empirical
uncertainty, and where it is relevant to obtain inscrutable and defens
Several elicitation protocols have been developed amongst which the much-used Stanford/SRI Protocol is the first (Spetzler and von Holstein, 1975; see also Morgan and Henrion, 1990; chapter 6 and 7). Expert elicitation typically involves the following steps:

(1) Identify and select experts;

(2) Explain to the expert the nature of the problem and the elicitation procedure. Create awareness of biases in subjective judgements and explore these.

(3) Clearly define the quantity to be assessed and chose a scale and unit familiar to the expert.

(4) Discuss the state of knowledge on the quantity at hand (strengths and weaknesses in available data, knowledge gaps, qualitative uncertainties).

(5) Elicit extremes of the distribution.

(6) Assess these extremes: could the range be broader than stated?

(7) Further elicit and specify the distribution (shape and percentiles or characterising parameters).

(8) Verify with the expert that the distribution that you constructed from the expert's responses correctly represents the expert's beliefs.

(9) Decide whether or not to aggregate the distributions elicited from different experts (this only makes sense if the experts had the same mental models of the quantity for which a distribution was elicited).

List of Grains in relation cross language

Illustration

Expert elicitation of health risks of Ultra Fine Particles (English)

Additional information