20 Years of Expert Judgement at TUDelft

Louis Goossens & Roger Cooke

PSAM9, Hong Kong
18-23 May 2008

The Delft Method of Expert Judgement Elicitation

- Developed by Roger Cooke – early 90-ies
- Support from
  - Ministry of Housing, Physical Planning and the Environment (the Netherlands)
  - European Commission
- Expert Judgement Procedures Guide (*EUR 18820, 2000*)

- Main goal: RATIONAL CONSENSUS in Decision-Making

Applications of the Delft Method

- In total (we elicited mostly the 5, 50, 95 percentile assessments of unknown variables)
  - 587 experts
  - 4,137 variables (“the unknowns and knowns”)
  - 67,759 elicitations (total number of questions)

Sector of application	*Expts In %*	*Variables In %*	*Elicitations In %*
Nuclear appl.	17	53	30
Chemical appl.	13	11	7
Natural disasters (water & volcanoes)	45	20	48
Space & Rockets	9	4	2
Animal health	8	6	4
Occupational risks	2	2	1
Banking	4	3	6

Applications of the PC method

Separate assessments

(pairwise comparisons)

293 experts

202 variables

14,826 elicitations

Sector of application	*Expts In %*	*Variables In %*	*Elicitations In %*
Safety Management applications	39	41	40
Chemical Process applications	5	10	6
Reliability of landfill techniologies	28	17	41
Water pollution applications	28	32	16

Goal of EJ study

- Census : data
- Consensus:
  - Political: one expert/stakeholder – one vote
  - Rational: pre-commit to method...*post hoc* withdrawal incurs burden of proof

Rational Consensus

Necessary but not sufficient:

- Scrutability/accountability
- Neutrality
- Fairness
- Empirical control

Variables

- Query variables: questions to the experts
- Target variables: unknown parameters
- Seed variables: unknown to the expert, but known to the analyst
- Each question is an *experiment*
- If query = target, the target variables must be *observable*
- If query $\neq$ target, the query variables are *post-processed* with probabilistic inversion techniques

Performance measures

- Calibration (statistical likelihood)
- Information (wrt background measure)

- Range graphs expertwise

Other issues

- Choosing experts
- Training
- Biases
- Expert names
- .....

Expert involvement

- Training meeting (*panel of experts*)
  - Discuss questionnaires
  - Discuss scope of analysis
  - Train in providing subjective assessments
- Working period for preparing assessments (*individual experts)*
- Elicitation session (*individual expert and 2 analysts: substantive and normative*)
  - Assessments of variables (3 quantile points)
  - Documented rationale (models, assumptions, uncertainties)
  - List of dependencies (between variables: if var. A > 50%, which % of var. B > 50%?)

Expert selection criteria

- Reputation in the field of interest
- Experimental experience in field of interest
- Publications in field of interest
- Awards
- Familiarity with uncertainty concepts
- Diversity in background
- Balance of views
- Interest in the project
- Availability for the project

Rational consensus
Performance based weights

Need ‘calibration’ or ‘seed’ variables to

- Evaluate expert performance
- Construct performance based DM
- Verify DM’s performance: *Empirical control*

DM = Decision Maker

Seed variables: examples

EJ Study	Variables of interest	Seed variables
Dispersion	Plume dispersion coefficients	Near-field tracer experiments **(domain**)
Environmental transport	Transfer coefficients	Cumulative concentrations **(adjacent**)
Dose-response models	Human dose response	Animal dose response **(adjacent**)
Option pricing	Quarterly rates	Weekly rates **(domain**)

Seed variables model human dose response (lethal toxicity due to large releases of chemicals)

Chemical	*# of seeds*	*Kinetics*	*Mechan-isms*	*Target organs*	*Functio-nal dis-turbance*	*Health effects*
Acrylonitrile	10	8	4	2	-	2
Ammonia	10	3	1	3	3	3
Hydrogen fluoride	9	6	-	-	-	3
Sulphur trioxide	10	2	1	3	1	6
Azinphos-methyl	10	6	2	1	1	5

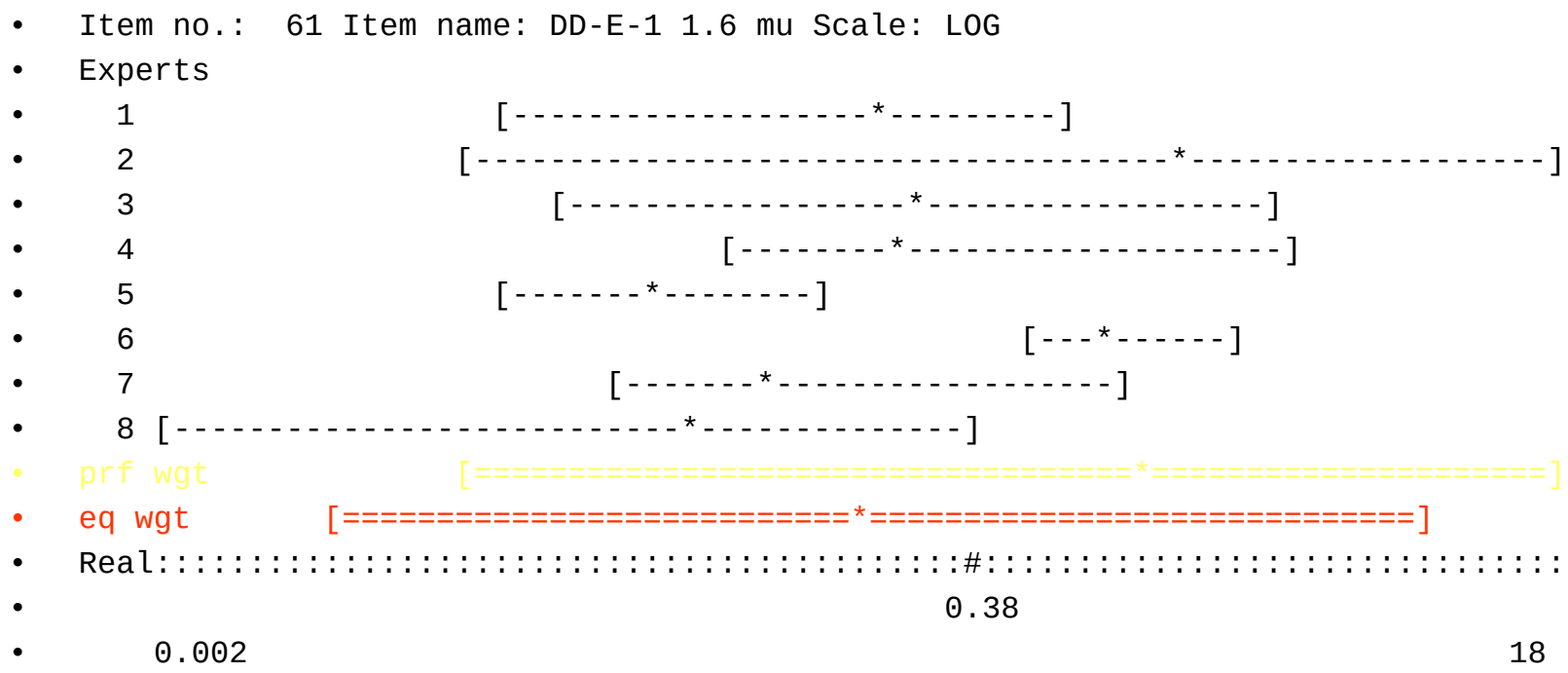
EU-USNRC Dry Deposition

03/07/2003

Results of scoring experts
Bayesian Updates: no Weights: global DM Optimisation: yes
Significance Level: 0.00169 Calibration Power: 1

Nr.	Id	Calibr.	Mean relat total	Mean relat realizatii	Numb real	UnNormaliz weight	Normaliz.w without DM	Normaliz.w with DM
1	Expert1	3.064E-5	0.9411	0.7044	14	0	0	0
2	Expert2	0.5271	0.3593	0.1661	14	0.08754	0.9339	0.4675
3	Expert3	0.00169	0.679	0.41	14	0.000693	0.007393	0.003701
4	Expert4	0.00169	0.7177	0.7231	14	0.001222	0.01304	0.006527
5	Expert5	2.054E-8	0.789	0.7201	14	0	0	0
6	Expert6	0.002203	1.188	1.341	14	0.002955	0.03152	0.01578
7	Expert7	0.00169	0.6474	0.7826	14	0.001323	0.01411	0.007064
8	Expert8	0.0008759	0.9759	0.5431	14	0	0	0
9	perf wgt	0.6587	0.2429	0.142	14	0.09351		0.4994
10	eq wgt	0.00169	0.1524	0.1677	14	0.0002834		0.002998

Dry Deposition Range Graphs: itemwise



Conclusions (1)

1. Expert judgment is Scientific data
2. Experts' performance as subjective probability assessors is highly variable
3. Experts like performance measurement
4. Valid measures of performance exist
5. Performance-based combinations of expert judgements outperform the equal weight combinations, and the best expert

Conclusions (2)

Uncertainty is that which disappears when we become certain.

Less uncertainty -> better decisions