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

- 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)
<table>
<thead>
<tr>
<th>Sector of application</th>
<th>Expts In %</th>
<th>Variables In %</th>
<th>Elicitations In %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear appl.</td>
<td>17</td>
<td>53</td>
<td>30</td>
</tr>
<tr>
<td>Chemical appl.</td>
<td>13</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Natural disasters</td>
<td>45</td>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td>(water &amp; volcanoes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space &amp; Rockets</td>
<td>9</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Animal health</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Occupational risks</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Banking</td>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>
Applications of the PC method

Separate assessments
(pairwise comparisons)

293 experts
202 variables
14,826 elicitations
<table>
<thead>
<tr>
<th>Sector of application</th>
<th>Expts In %</th>
<th>Variables In %</th>
<th>Elicitations In %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety Management applications</td>
<td>39</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>Chemical Process applications</td>
<td>5</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Reliability of landfill technologies</td>
<td>28</td>
<td>17</td>
<td>41</td>
</tr>
<tr>
<td>Water pollution applications</td>
<td>28</td>
<td>32</td>
<td>16</td>
</tr>
</tbody>
</table>
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 ≠ 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

<table>
<thead>
<tr>
<th>EJ Study</th>
<th>Variables of interest</th>
<th>Seed variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dispersion</td>
<td>Plume dispersion coefficients</td>
<td>Near-field tracer experiments (domain)</td>
</tr>
<tr>
<td>Environmental transport</td>
<td>Transfer coefficients</td>
<td>Cumulative concentrations (adjacent)</td>
</tr>
<tr>
<td>Dose-response models</td>
<td>Human dose response</td>
<td>Animal dose response (adjacent)</td>
</tr>
<tr>
<td>Option pricing</td>
<td>Quarterly rates</td>
<td>Weekly rates (domain)</td>
</tr>
</tbody>
</table>
Seed variables model human dose response (lethal toxicity due to large releases of chemicals)

<table>
<thead>
<tr>
<th>Chemical</th>
<th># of seeds</th>
<th>Kinetics</th>
<th>Mechanisms</th>
<th>Target organs</th>
<th>Functional disturbance</th>
<th>Health effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylonitrile</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>e</td>
<td>2</td>
</tr>
<tr>
<td>Ammonia</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Hydrogen fluoride</td>
<td>9</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Sulphur trioxide</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Azinphos-methyl</td>
<td>10</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
Results of scoring experts

Bayesian Updates: no      Weights: global     DM Optimisation: yes
Significance Level:    0.00169    Calibration Power:          1

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

(c) 1999 TU Delft
Dry Deposition Range Graphs: itemwise

- Item no.: 61 Item name: DD-E-1 1.6 mu Scale: LOG
- Experts
- 1
  [-------------------*---------]
- 2
  [-------------------------------------*-------------------]
- 3
  [------------------*-------------- ----]
- 4
  [--------*--------------------]
- 5
  [---- ---*--------]
- 6
  [---*------]
- 7
  [-------*------------------]
- 8
  [--------------------------*---------------]
- prf wgt
  [=======================================*=================================]
- eq wgt
  [=======================================*=================================]
- Real:
  [=======================================*=================================]
  0.38
  0.002
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