Expert Judgment research coda:
(a) Dependence elicitation
(b) Expected Relative Frequency (ERF) Decision-maker
(c) Paired Comparison analysis of stakeholder preferences, with Probabilistic Inversion ranking

Willy Aspinall

---

Ice sheet melting – projected contributions to future sea-level rise

An expert judgement assessment of future sea level rise from the ice sheets

J. L. Rambo and W. P. Aspinall (2013)

A major gap in predictive capability concerning the future evolution of the ice sheets was identified in the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change. As a consequence, it has been suggested that the AR5 would utilize a “broadened” approach from that used in the AR4 report. The commitment of the AR5 to ensure that the characteristics of the AR4 experts are maintained is reflected in this report.

A major gap in predictive capability concerning the future evolution of the ice sheets was identified in the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change. As a consequence, it has been suggested that the AR5 would utilize a “broadened” approach from that used in the AR4 report. The commitment of the AR5 to ensure that the characteristics of the AR4 experts are maintained is reflected in this report.
Aspinall on Dependence; ERF, & PCPI
"We have very high confidence (90 percent) that global mean sea level will rise at least 0.2 meters (8 inches) and no more than 2.0 meters (6.6 feet) by 2100."

(Global Sea Level Rise Scenarios for the United States National Climate Assessment, 2012)
Monte Carlo simulation results using experts' quantiles:
SLR rate at 2100AD, three temperature scenarios: 2°C, 3°C, 4°C

Pooled expert judgements on combined ice-sheet contributions to sea-level rise: 2100AD; 2200AD
Variable dependences

- For each of the three ice sheets, Greenland (GrIS), West Antarctica (WAIS), and East Antarctica (EAIS), the contribution to SLR per unit time is modeled as:

\[ SLR_{is} = Discharge_{is} + Runoff_{is} - Accumulation_{is} \]

where \( IS \) denotes GrIS, WAIS or EAIS.

- The units are mass converted to millimeters sea level rise.

- These quantities were elicited for 2°C, 3°C and 4°C warming scenarios by 2100 and 4°C, 6°C and 8°C warming by 2200.

- The uncertainties are large and dependences could have an appreciable effect.

Dependence elicitation

Dependence between variables of interest is often relatively benign in the sense that the uncertainties on the variables strongly dominates any "dependence effect". This is not always the case, however.

- Dependence tout court between variables X and Y was captured by asking experts:

  *Suppose X is observed and its value is above your median, what is then your probability that Y is also above your median?*

- IF X and Y are independent, then the answer should be 0.5; larger values indicate positive dependence and lower values indicate negative dependence. Experts quickly buy into this format.
• The use of tail independent copulae is not conservative and has been charged with inducing excessive risk taking on Wall Street (Solamon 2009). Therefore to allow for the possibility of tail dependence, experts are asked, in addition to previous, median-related question:

Suppose X is observed and its value is above your 95 percentile, what is then your probability that Y is also above your 95 percentile?

• if these uncertainties are independent, the elicited probability should 0.05; probabilities greater than 0.05 indicate positive association, less than 0.05 indicate negative association.

By observing the relations between these two exceedance probabilities, a choice can be made, typically, between one-parameter copula families, namely two with tail independence (normal and Frank) and two with tail dependence (Gumbel and reverse Clayton).

---

Vines - a way of collecting (conditional) bivariate correlation constraints that can be specified independently of each other

![Vines Diagram](image)

**Vines and Continuous Non-parametric Bayesian Belief Nets with Emphasis on Model Learning**

Dorota Kurowicka; Roger M. Cooke

In: RETHINKING RISK MEASUREMENT AND REPORTING

RiskBooks: edited By Klaus Böcker
(Tail?) Dependence captured in a low order Regular Vine

A "vine" is a generalization of a rank correlation dependence tree, allowing arbitrary correlation matrices to be modelled. \[\text{[Bedford & Cooke, 2001]}\]

Tail independence: Normal Copula

\[
\frac{\Pr(Y > \text{u-th quantile} \mid X > \text{u-th quantile})}{1 - u} \rightarrow 1 \text{ as } u \rightarrow 1
\]

Correlation = 0.8
Various other copulas are available

The message from the Gumbel or Reverse Clayton copula is:

*If something bad happens to X, there is good reason to fear something bad will happen to Y, and the reasons get stronger as the coupling between X and Y increases, and as "bad" gets "worse".*
Representing Expert 1’s elicited dependences

Examining two highest weighted experts dependence elicitations:

<table>
<thead>
<tr>
<th>Ice sheet contribution to SLR in 2100 @ 3C warming [mm]</th>
<th>Mean</th>
<th>StDev</th>
<th>5%-ile</th>
<th>50%-ile</th>
<th>95%-ile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1 indep</td>
<td>295</td>
<td>196</td>
<td>53</td>
<td>263</td>
<td>691</td>
</tr>
<tr>
<td>Expert 1 Tail Dep</td>
<td>295</td>
<td>222</td>
<td>53</td>
<td>249</td>
<td>752</td>
</tr>
<tr>
<td>Expert 7 indep</td>
<td>407</td>
<td>189</td>
<td>171</td>
<td>377</td>
<td>787</td>
</tr>
<tr>
<td>Expert 7 Tail Dep</td>
<td>406</td>
<td>225</td>
<td>150</td>
<td>359</td>
<td>861</td>
</tr>
</tbody>
</table>
Pooled expert judgements on combined ice-sheet contributions to sea-level rise: 2100AD; 2200AD

- Point-wise value estimation using expert judgment
Motivation for new model:
an occasional result with the Classical Model

These two experts get equal statistical accuracy scores:

- an expert's averaged relative number of “good” answers over several trial items is proportional to the arithmetic mean of probabilities that his/her answers fall in defined “good intervals”

- this measure provides an index for expert’s ability to give answers close to true value

- leading to a set of differential weights for judgments of an expert group, and an ERF-based pooling DM for target items (central value counterpart to the Cooke Classical Model statistical accuracy DM).

Define a “closeness” score for each of \( N \) test questions. Average these scores as a definitive reward, \( L \).

If \( F^e_i \) denotes the cumulative distribution function of expert \( e \) for the \( i \)-th question, the formula for the reward \( L \) of expert \( e \) is

\[
L(e) = \frac{1}{N} \sum_{i=1}^{N} \left( F^e_i (x_i + \% x_i) - F^e_i (x_i - \% x_i) \right)
\]

[percentage \( \% x_i \) decided by analyst a priori]

**Basis of ERF model**

**Influence of Informativeness on the L score**

Red Expert: true value within distribution’s support - gets non-zero weight, but has wide distribution.

Green Expert provides more concentrated distribution, with the same mode as Red Expert.

Green Expert awarded higher \( L \) score because this range is more probable under his/her distribution.
Cooke C*I and ERF (central) accuracy scores

ERF DM gets a higher weight for central accuracy than other DMs (right frame) [but outscored in by the Best Expert 9]

Calibration & Informativeness comparison of DMs
Two-dimensionality of the four rewards: one associated with Expected Accuracy and Relative Error, the other with Calibration and Informativeness, the latter two anti-correlated.

Intuitively we associate these two dimensions with estimation “accuracy” and “uncertainty”.

Based on cross-validation simulations
Aspinall on Dependence; ERF, & PCPI

Cross-validation tests of performance-based weights

Probability of picking an individual expert better than the DM, according to mean Cooke C’I score or Expected Relative Frequency accuracy rewards.

<table>
<thead>
<tr>
<th></th>
<th>C’I reward</th>
<th>ERF accuracy reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vesuvius</td>
<td>0%</td>
<td>15%</td>
</tr>
<tr>
<td>Gas95</td>
<td>9%</td>
<td>45%</td>
</tr>
<tr>
<td>Pbearl</td>
<td>14%</td>
<td>43%</td>
</tr>
<tr>
<td>Disper</td>
<td>0%</td>
<td>27%</td>
</tr>
<tr>
<td>Depos</td>
<td>0%</td>
<td>50%</td>
</tr>
</tbody>
</table>

None of these probabilities is greater than 50%!

Picking a single “best expert” *a priori* is hazardous, as the optimum DM for either type of reward is generally better than all individual experts.

*See also Cooke, ElSaadany & Huang (2008 RESS Issue on Expert Judgement); and Burgman et al (2011)*

• Paired comparison with probabilistic inversion
Paired comparisons

Person A prefers a Mercedes 300 to a BMW

\[ > \]

and he prefers BMW > Corsa

\[ > \]

what about his choice between Merc and Corsa?

Paired Comparison Example Case: Energy Policies

1. **Tax@pump**: 1$ per gallon gasoline surcharge, to be used for research in renewables

2. **Tax Break**: (a) No sales tax on purchase of new hybrid or electric car; (b) First car owners can deduct purchase cost from their income tax; (c) No sales tax on bio-diesel or ethanol; (d) Tax credits for energy efficiency home improvements (insulation, double glass windows, solar panels)

3. **Vehicle Tax**: Annual road tax 1$ per lb on all light duty vehicles, no tax rebate for driving to work or parking, to be used for research in fuel efficient vehicles and bio fuels.

4. **CO$_2$ cap**: CO$_2$ emissions cap on electricity generation.

5. **Subsidies for clean coal**: Give subsidies for clean coal with carbon sequestration to make coal competitive with natural gas.

6. **Do Nothing**
Enter Data in UNIBALANCE

Check each Stakeholders’ consistency

p-value for rejecting the hypothesis that pairwise preferences are at random, based on the nr of circular triads.
Check that group is statistically different from ‘random stakeholders’

- Variance in stakeholder-averaged ranks
- How much stakeholders pairwise agree

Paired Comparison Data Analysis

- Is each stakeholder’s preference non-random?
- Is the agreement between stakeholders non-random?
  - Coefficient of agreement
  - Coefficient of concordance
Petrology Workshop findings

Hazard Assessment

Experts 3, 5, 7, 9 and 11 provided choices that are rationally coherent;
the high p-values for the others suggest random responses (see next slide)

Hazard Assessment

Note very low coeff. of agreement and high p-value for group responses, and warning!!
Hazard Assessment

Relative ranking scores and variances

197 random draws from the response distributions inferred by Probabilistic Inversion from the group sample
Research

Experts 2, 3, 5, 6, 7, 11, 12 provide non-random choices

Expert 4’s responses are almost totally random

Note: experts are not listed in the same order as for hazard assessment responses

Research

Although agreement coeff is quite low, the group’s ranking choices do not appear random

P = 0.035
Research

Relative ranking scores and variances

### Probabilistic inversion

<table>
<thead>
<tr>
<th>Item name</th>
<th>Score</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gas</td>
<td>0.9414</td>
<td>0.2720</td>
</tr>
<tr>
<td>2. Timescales</td>
<td>0.6544</td>
<td>0.2720</td>
</tr>
<tr>
<td>3. Petrogenesis</td>
<td>0.6585</td>
<td>0.2810</td>
</tr>
<tr>
<td>4. Geophys</td>
<td>0.5288</td>
<td>0.2624</td>
</tr>
<tr>
<td>5. Observatories</td>
<td>0.3119</td>
<td>0.2400</td>
</tr>
<tr>
<td>6. Fluid</td>
<td>0.4666</td>
<td>0.2666</td>
</tr>
<tr>
<td>7. Communication</td>
<td>0.4127</td>
<td>0.2972</td>
</tr>
</tbody>
</table>

Entropy: 1.0071

---

Research

200 random draws from the response distributions inferred by Probabilistic Inversion from the group sample
Aspinall on Dependence; ERF, & PCPI
Results from paired comparison PI of twelve vCJD transmission risks

<table>
<thead>
<tr>
<th>Transmission Route</th>
<th>Score</th>
<th>St. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Platelet transfusion</td>
<td>0.5599</td>
<td>0.2318</td>
</tr>
<tr>
<td>2  FFP plasma transfusion</td>
<td>0.6266</td>
<td>0.2501</td>
</tr>
<tr>
<td>3  Whole blood transfusion</td>
<td>0.7384</td>
<td>0.2075</td>
</tr>
<tr>
<td>4  Dura Mater transplant</td>
<td>0.9520</td>
<td>0.0370</td>
</tr>
<tr>
<td>5  Packed red blood cells</td>
<td>0.6002</td>
<td>0.2742</td>
</tr>
<tr>
<td>6  Dental tissue graft</td>
<td>0.2756</td>
<td>0.2484</td>
</tr>
<tr>
<td>7  Corneal transplant</td>
<td>0.6953</td>
<td>0.2263</td>
</tr>
<tr>
<td>8  Hematopoietic stem cell transplant</td>
<td>0.3197</td>
<td>0.1752</td>
</tr>
<tr>
<td>9  Human derived urine fertility products</td>
<td>0.3160</td>
<td>0.2034</td>
</tr>
<tr>
<td>10 Bone marrow transplant</td>
<td>0.4966</td>
<td>0.2195</td>
</tr>
<tr>
<td>11 pdFVIII</td>
<td>0.4353</td>
<td>0.2524</td>
</tr>
<tr>
<td>12 pdFXI</td>
<td>0.3992</td>
<td>0.2340</td>
</tr>
</tbody>
</table>

- [http://ewi.tudelft.nl/over-de-faculteit/afdelingen/toegepaste-wiskunde/risico-analyse/software/unibalance/unibalance/](http://ewi.tudelft.nl/over-de-faculteit/afdelingen/toegepaste-wiskunde/risico-analyse/software/unibalance/unibalance/)

or Google UNIBALANCE TUDelft
Thank you!
Back to a climate change image for the last word .......

GLOBAL WARMING