Fourth Subjective Bayesian Meeting
Department of Statistics, University of Warwick
Friday 2nd September 2016

EXPERT ELICITATION

Willy Aspinall
Bristol Earth Sciences
and
Aspinall & Associates
University of BRISTOL | Cabot Institute

Expert elicitation techniques

- Delphi method
- Team building, decision conferencing, etc
- Structured Expert Judgment with performance scoring

Key question: How do we measure performance?

Credibility comes from performance .....

The Classical Model is the only procedure with empirical control on performance measures

EJ for RATIONAL CONSENSUS:

Goal: comply with scientific method principles and combine experts' judgments to get a **Good Probability Assessor**

Parties pre-commit to a method which satisfies necessary conditions for scientific method:

Traceability/accountability

Neutrality (don't encourage untruthfulness)

Fairness (ab initio, all experts equal)

Empirical control (performance meas't)

Withdrawing from SEJ *post hoc* incurs burden of proof on Expert

Cooke's "Classical Model" for SEJ

** Cooke, R.M. (1991) Experts in Uncertainty. Oxford University Press, 321pp.*

What is a GOOD subjective probability/uncertainty assessor?

- **Calibration, statistical likelihood**
  - Are the expert's probability or uncertainty statements statistically accurate?

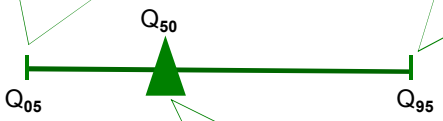
    ⇨ p-value of statistical test
- **Informativeness**
  - Is probability mass concentrated in a small region, relative to uniform background measure?

    ⇨ information measure

Three-quantile elicitation question format for an unknown parameter or probability item…

1. To reflect the uncertainty distribution associated with this item, what do you think is a plausible value for the 5th percentile?

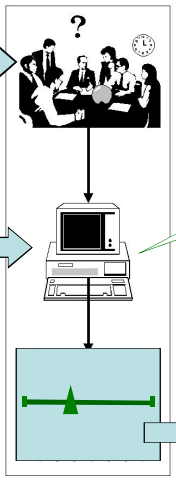
2. To reflect the uncertainty distribution associated with this item, what do you think is a plausible value for the 95th percentile?



3. To locate the uncertainty distribution central tendency, what do you think is a plausible value for the 50th percentile (median)?
[n.b. this value need not be symmetrical between outer quantiles]

Pooling experts' target item judgments with Classical Model performance weights

Target item = Prob [event | data]?
Experts give uncertainty percentiles: Q_{05}, Q_{50}, Q_{95}

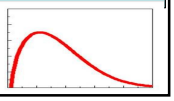


Classical Model algorithm

Experts' performance-based weights from scores on seed item set calibration:
$W_j = C_j * I_j$

$DM_i = \Sigma W_{ij} * Q_i$

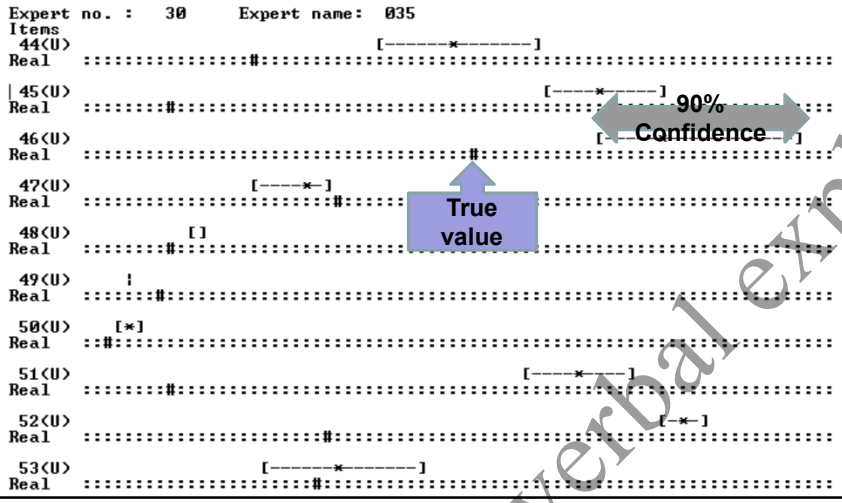
Weighted combination of pooled group responses creates "Decision-Maker" DM_i uncertainty quantiles for Item i; usually fit to suitable distribution:



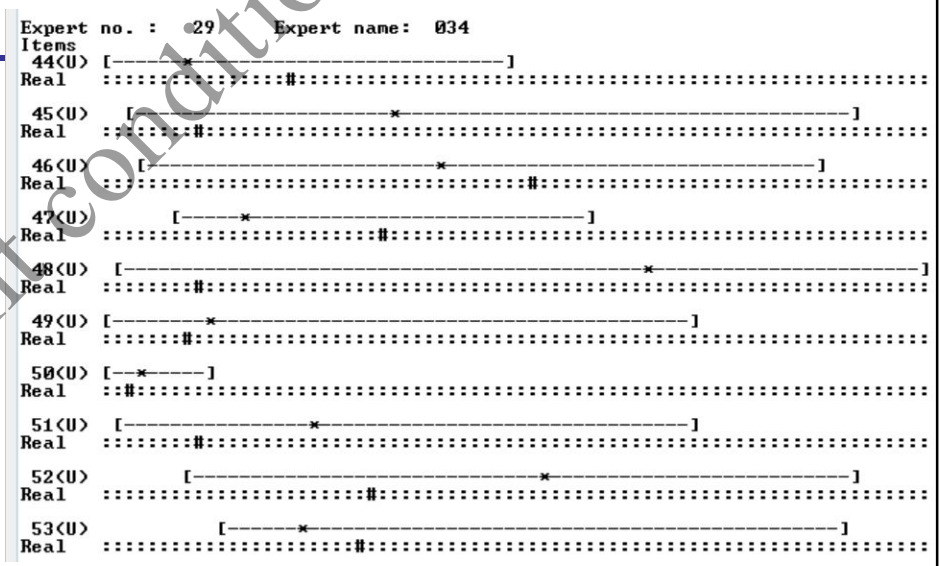
Individual expert **j**: weight W_j is product of his/her statistical accuracy C and informativeness I, averaged over the set of seed items

Expert Performance CAN be objectively measured

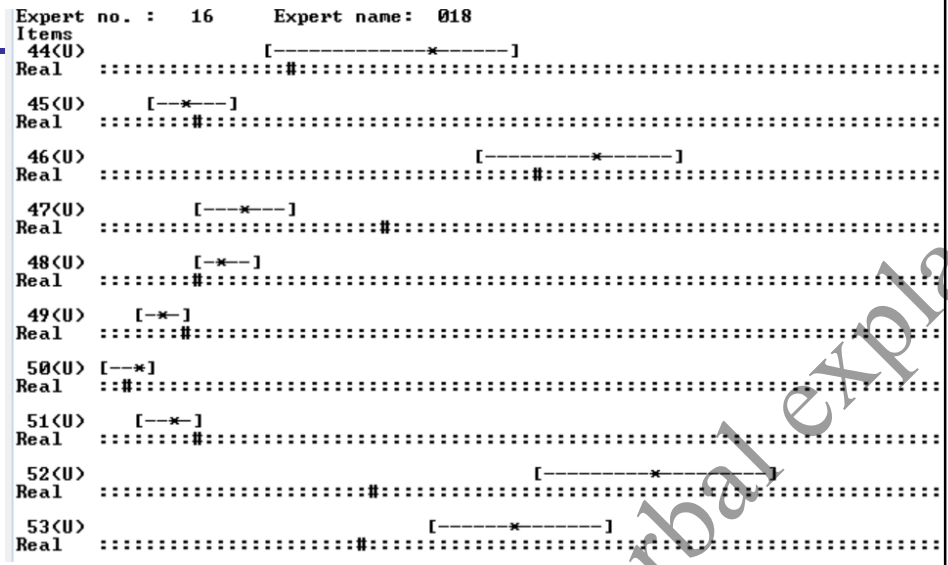
Very High Information, Very Poor Statistical Accuracy



Low Information, Good Statistical Accuracy



High Information, Decent Statistical Accuracy



Scoring for expert e:

- **Calibration score**: Cal(e) = probability of FALSELY rejecting the statistical hypothesis that e's probability statements are accurate based on realizations
- **Information score:** Inf (e) = (Shannon relative information wrt background) = ability to concentrate high probability in small intervals.

Calibration

For each variable X_i, i = 1..n

Assess: $\_\_a_i\_\_\_$ $\_\_b_i\_\_\_$ $\_\_c_i\_\_\_$

5% 50% 95%

Expert believes

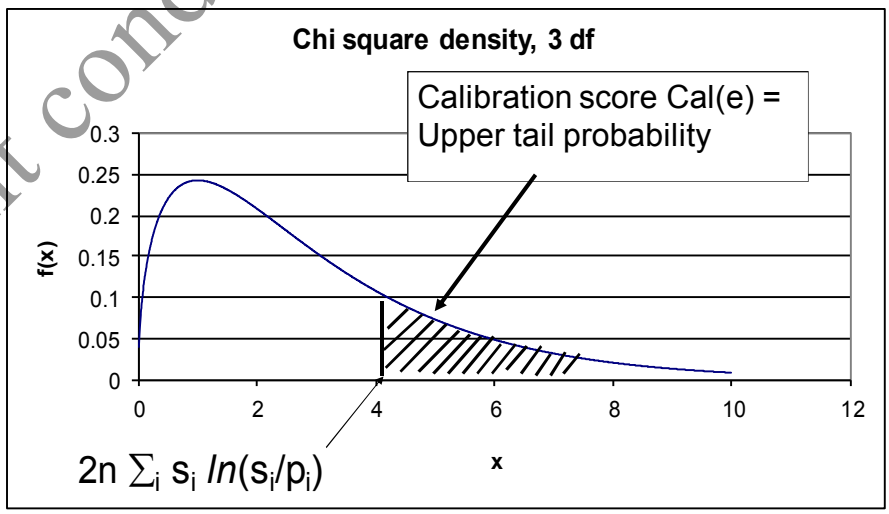
$p_1 = \text{Prob}(X_i \le a_i) = 0.05$, $p_2 = \text{Prob}(a_i < X_i \le b_i = 0.45)$, etc

Let $x_1 \ldots x_n$ be realizations of $X_1 \ldots X_n$

$s_1 = \#\{i \mid x_i \le a_i\} / n$, $s_2 = \#\{i \mid a_i < X_i \le b_i\} / n$, etc

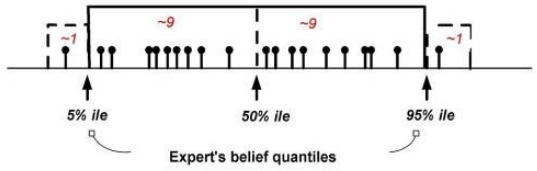
Then $2n \sum_{i=1..4} s_i \, ln \, (s_i / p_i) \sim$ Chi square, 3df.

Calibration Score

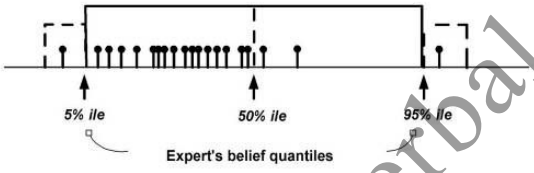


Good and "poor" expert statistical accuracy performance

Expected spread of 20 statistically independent realization draws from a 'well-calibrated' expert's distribution:



This expert is penalized for lop-sided support:



Information score

Compute Shannon relative information wrt background:

$$I(e,i) = I(f_{e,i}(x) | \mu(x)) = \ldots \Sigma_{j=1..4}\, p_j \ln(p_j / m_j)$$

m_j is background measure of inter-quantile interval j, for item i.

Inf score(e) = average information: $(1/\#\text{items})\ \Sigma_i\, I(e,i)$

Optimization

- significance level α is chosen to optimize combined score of DM:

$f_\alpha(i) = \Sigma_{e=1\ldots E} f_{e,i} \times \text{Cal}(e) \times \text{Inf}(e) \times 1_\alpha\{\text{Cal}(e)\}$

For each α, compute calibration $\times$ information; choose α for which this is maximum.

Combining Experts' judgements

$f_{e,i}$ = expert e's density for item/variable i.

Performance Based Combinations

Global weight decision maker

- proportional to expert's combined score, (with optimization).

$f_{gw}(i) = \Sigma_{e=1..E} w_e f_{e,i}$; $\Sigma_{e=1..E} w_e = 1$.

Weight depends on Expert performance over all Items

Item weight decision maker

- product of calibration and information for each item (with optimization).

$f_{iw}(i) = \Sigma_{e=1..E} w_{e,i} f_{e,i}$; $\Sigma_{e=1..E} w_{e,i} = 1$.

Weight depends on Expert and specific Item

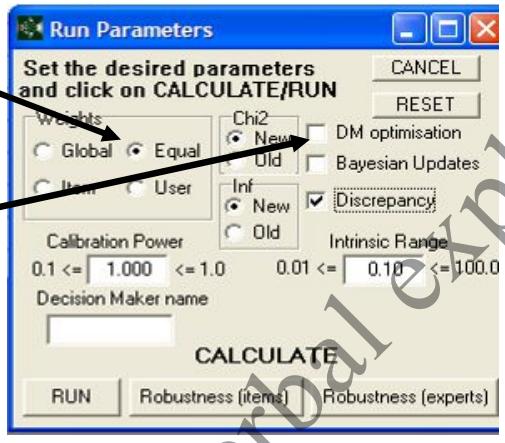
Equal weight decision maker

$f_{eq}(i) = (1/E) \Sigma_{e=1..E} f_{e,i}$

EXCALIBUR checks - Discrepancy

Run EXCALIBUR with Equal Weights and Discrepancy:

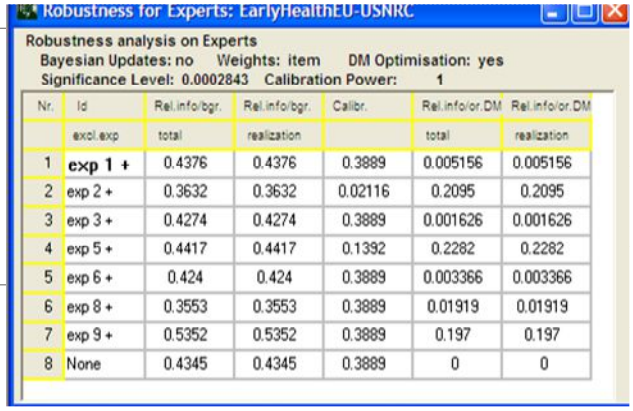
Shows how much the experts differ from the "average expert"



Robustness

Run Robustness (Items or Experts) to see how omitting item or expert would affect results:

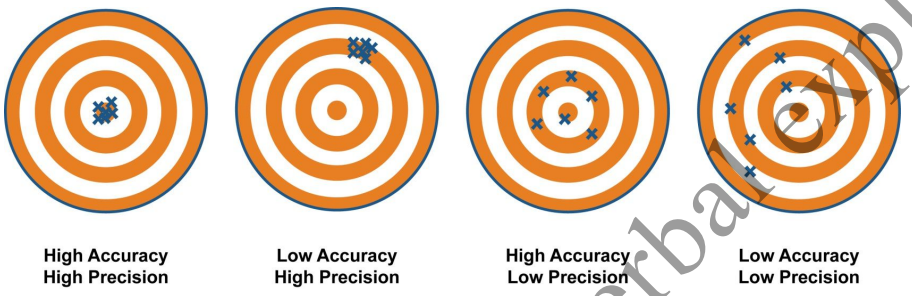
is the mean difference wrt original DM smaller than the differences between experts themselves?



Nr.	Id	Rel.info/bgr.	Rel.info/bgr.	Calibr.	Rel.info/or.DM	Rel.info/or.DM
	excl.exp	total	realization		total	realization
1	**exp 1 +**	0.4376	0.4376	0.3889	0.005156	0.005156
2	exp 2 +	0.3632	0.3632	0.02116	0.2095	0.2095
3	exp 3 +	0.4274	0.4274	0.3889	0.001626	0.001626
4	exp 5 +	0.4417	0.4417	0.1392	0.2282	0.2282
5	exp 6 +	0.424	0.424	0.3889	0.003366	0.003366
6	exp 8 +	0.3553	0.3553	0.3889	0.01919	0.01919
7	exp 9 +	0.5352	0.5352	0.3889	0.197	0.197
8	None	0.4345	0.4345	0.3889	0	0

Characterizing experts' judgement traits

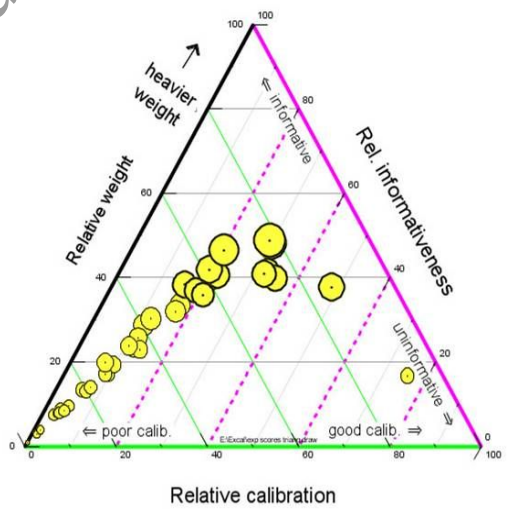
- Calibration questions (seed items) should resemble the target questions as closely as possible…..
- Used to measure the expert's performance
  - Accuracy – good calibration
  - Precision – high information



High Accuracy High Precision | Low Accuracy High Precision | High Accuracy Low Precision | Low Accuracy Low Precision

Expert weights from the Classical Model

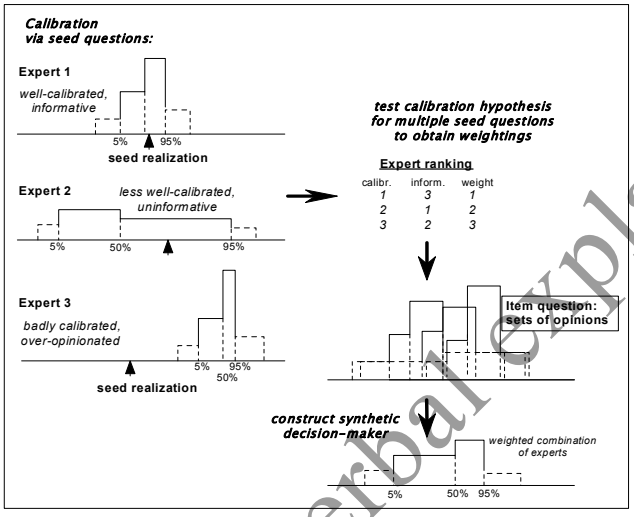
Typical profile: experts relative weights from informativeness and calibration scores



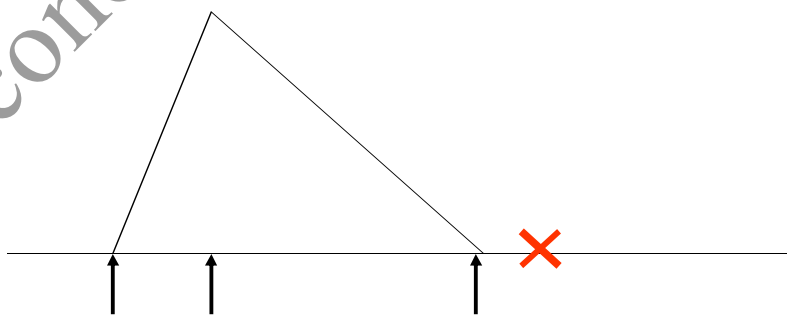
Cartoon

An optimal decision on any question of interest can then be obtained from the weighted sum of the opinions of a group of experts:

$$DM_i = \Sigma W_j * Q_i$$



Why not Triangular?



An observation outside range is infinitely surprising: disables statistical treatment

Identify seed variables

1. *Ask for values of observable or potentially observable quantities.*
2. *Formulate questions in a manner consistent with the way in which an expert represents the relevant information in his knowledge base.*

Practical issues

1. *The seed variables should sufficiently cover the case structures for elicitation. Particularly, when one expert panel should tackle different sub fields, seed variables must be provided for all sub fields.*
2. *For each panel at least 10 seed variables are needed, preferably more.*
3. *Seed variables may be identified as such in the elicitation - but maybe not …….*
4. *If possible, the analyst should be unaware of the values of the seed variables during the elicitation.*

FAQ's(1)

- **From an expert:** *I don't know that*
  - **Response**: *No one knows, if someone knew we would not need to do an expert judgment exercise. We are tying to capture your uncertainty about this variable. If you are very uncertain then you should choose very wide confidence bounds.*
- **From an expert:** *I can't assess that unless you give me more information.*
  - **Response**: *The information given corresponds with the assumptions of the study. We are trying to get your uncertainty* conditional *on the assumptions of the study. If you prefer to think of uncertainty conditional on other factors, then you must try to unconditionalize and fold the uncertainty over these other factors into your assessment.*
- **From an expert:** *I am not the best expert for that.*
  - **Response**: *We don't know who are the best experts. Sometimes the people with the most detailed knowledge are not the best at quantifying their uncertainty.*
- **From an expert:** *Does that answer look OK?*
  - **Response**: *You are the expert, not me.*
- **From the problem owner:** *So you are going to score these experts like school children?*
  - **Response**: *If this is not a serious matter for you, then forget it. If it is serious, then we must take the quantification of uncertainty seriously. Without scoring we can never validate our experts or the combination of their assessments.*

FAQ's(2)

- **From the problem owner:** *The experts will never stand for it.*
  - **Response** *We've done it many times, the experts actually like it.*
- **From the problem owner:** *Expert number 4 gave crazy assessments, who was that guy?*
  - **Response**: *You are paying for the study, you own the data, and if you really want to know I will tell you. But you don't need to know, and knowing will not make things easier for you. Reflect first whether you really want to know this.*
- **From the problem owner:** *How can you give an expert weight zero?*
  - **Response:** *Zero weight does not mean zero value. It simply means that this expert's knowledge was already contributed by other experts and adding this expert would only add a bit of noise. The value of unweighted experts is seen in the robustness of our answers against loss of experts. Everyone understands this when it is properly explained.*
- **From the problem owner:** *I prefer to use the equal weight combination.*
  - **Response**: *So long as the calibration of the equal weight combination is acceptable, there is no scientific objection to doing this. The analyst's job is to indicate the best combination, according to the performance criteria, and to say what other combinations are scientifically acceptable.*

EXCALIBUR applications

- Early applications:
- Space
- (propulsion system reliability)
- Space
- (space debris impact)
- Space
- (strength of composites)
- Industrial
- (flange connection failures)
- Industrial
- (fuelling crane failure)
- Hydrology
- (groundwater contamination; reservoir erosion modelling)

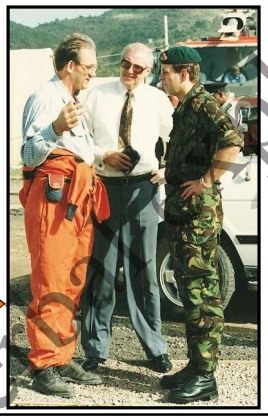
- More recent applications:
- Climate change
- (radwaste storage)
- Seismology
- (earthquake hazards)
- Bioterror
- (malacious agents……)
- Medical
- (SARS; vCJD risk models, XMRV, CWD … etc)
- Volcanology
- (eruption risks…..)

Montserrat volcano July 1995 – November 2015 twenty years of expert elicitations

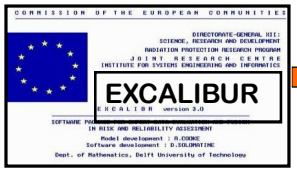
Highlighting some issues that have emerged from practical elicitations for various hazard and risk assessments ….. I'm to blame for the content

Prompted by the Guadeloupe 1976 experience….

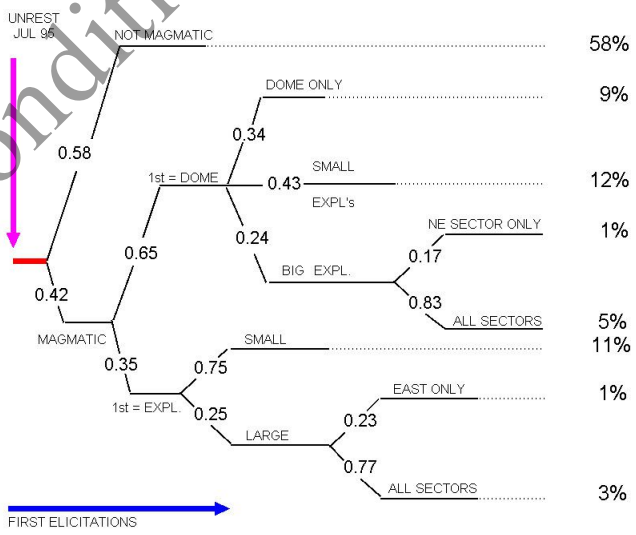
….in Montserrat, we put in place a formalised procedure for providing scientific advice to the authorities



…..using the EXCALIBUR procedure, developed originally for the European Space Agency

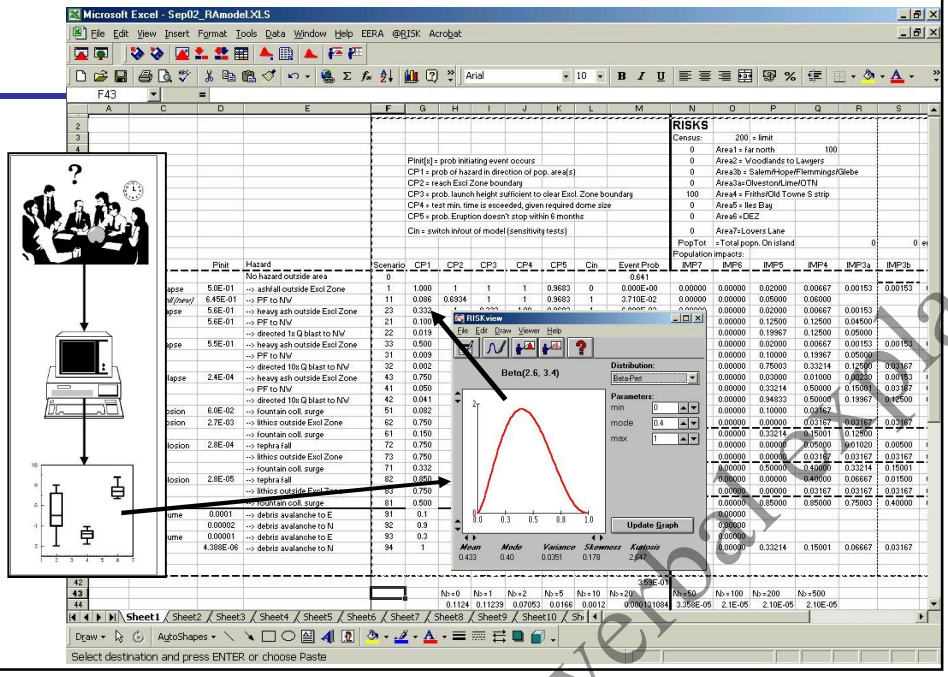


Montserrat volcano: event tree

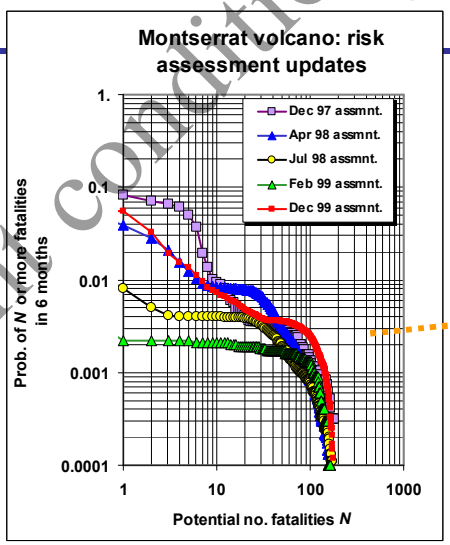


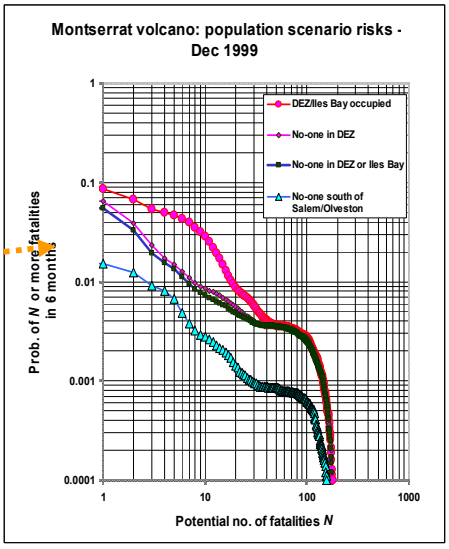
VOLCANO CRISIS: INITIAL PROBABILITY TREE

Constructing the joint hazard – risk Monte Carlo simulation model

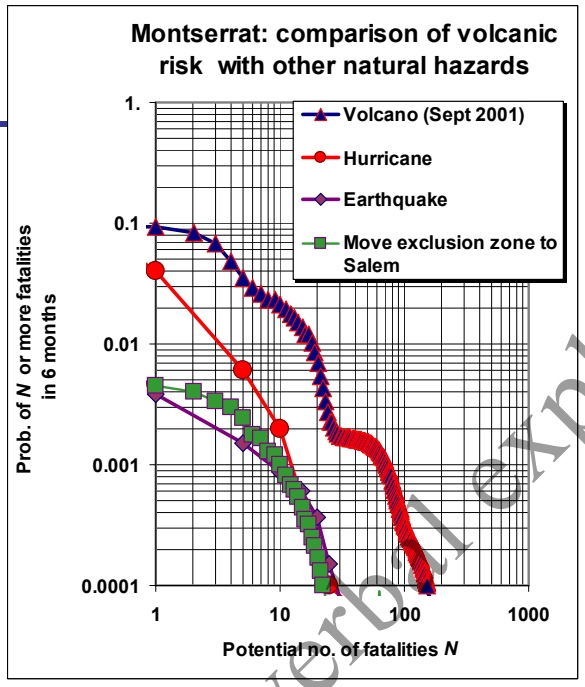


Population risk curves: regular updates,…….. and mitigation by staged evacuation





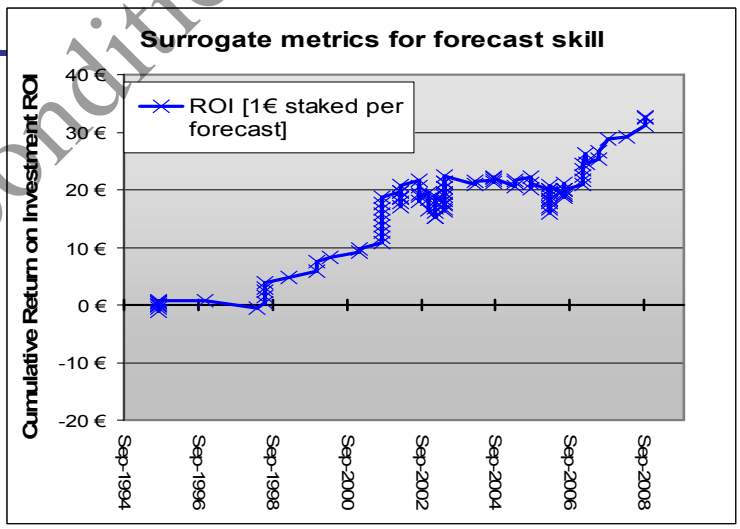
"Acceptable" risk levels?



….comparison to other natural hazards??

Pandemic flu

Communicating forecast skill



[Hagedorn, R., Smith, L.A. (2008) Communicating the value of probabilistic forecasts with weather roulette. Meteorol. Appl. Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/met.9.]

Vesuvius, and the future threat to Naples

3G WORLD NEWS

New computer study shows widespread devastation in final phase of eruption

5mins after it begins
Column of red-hot ash, gas and rock collapses on to mountain, blasting towards populated areas

15mins
Pyroclastic flow sweeps through homes built illegally on slopes

20mins
Flows burst through town from shins to Pompeii

Magma erupts into the sky at 400mph

AD79 eruption: plaster casts of victim who were burnt to death are now a tourist attraction

VESUVIUS

Illicit homes sprawl down slopes

Magma travels at 60 mph

Up to 300,000 could be killed

"Today people aren't prepared for an evacuation. It's wrong to bet on carrying out the entire evacuation in just three to five days. If there was an eruption tomorrow it's quite possible we'd see huge traffic jams, car crashes and people using guns to make their escape," he said. "We need a plan to start the evacuation of vulnerable people, like pensioners, first. People need to know what to do and they have to understand they won't be able to get out by car. You can walk 10 miles in three hours, that's quite a good safety margin."

Some Neapolitans are hoping Vesuvius will stir soon. Guides who work near its summit believe a volcano that splutters, grows and occasionally spits fire and lava is a better crowd-puller than a dull, impo- vision only to descript the ... hear th... waiting ... men."

THE SUNDAY TIMES MARCH 4, 2007

Vesuvius blast could kill 300,000

John Follain

THE next eruption of Vesuvius could kill at least 300,000 people, nearly 20 times as many as the AD79 disaster that buried the ancient city of Pompeii, according to Italian government research.

More than half a million people live in the so-called "red zone" of 18 towns in a four-mile radius of the volcano and most would die if an evacuation could not be completed in time, the research says.

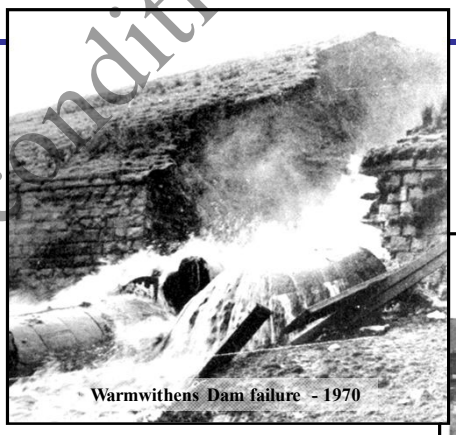
The findings are from a study by some of Europe's leading volcanologists and public health experts, including Dr Peter Baxter of Cambridge University's Department of Public Health.

The destruction of Pompeii, the worst affected city, has inspired many books and films, including Robert Harris's 2003 bestseller, which features Pliny the Elder and which is to be adapted by the director Roman Polanski in a £100m movie.

Some 2.5m tourists visited Pompeii last year, where the people by bus from each of the 18 towns.

Professor Giuseppe Luongo of the University of Naples, former director of the Vesuvius Observatory which monitors the volcano, believes plans are inadequate and local people are ill-informed about them.

From fire to water……



Warmwithens Dam failure - 1970

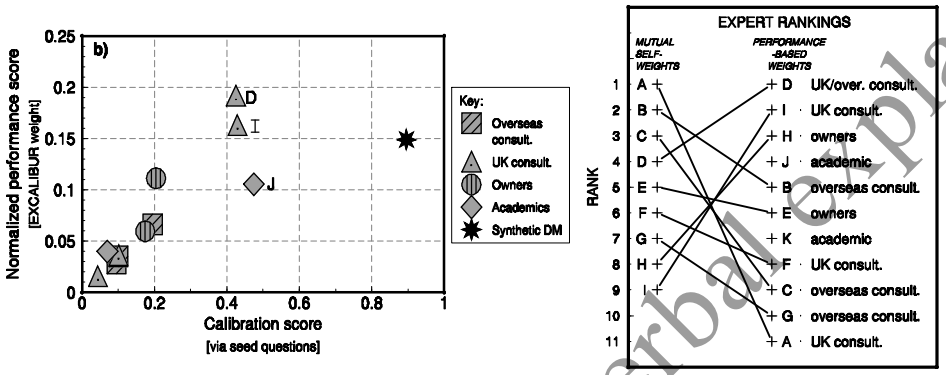
Cowlyd Reservoir inspection party - 1917

..risk assessment and reservoir safety in the UK

36

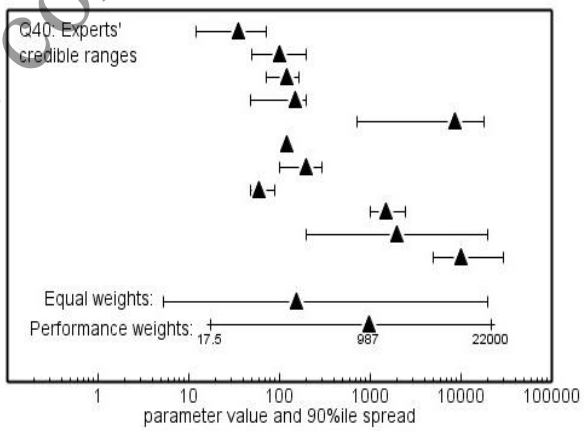
The reservoir engineers: performance-based scores, and mutual weighting rankings

Objective: to develop a generic quantitative model for accelerated internal erosion in Britain's population of 2,500 ageing dams, using elicited quantities for key variables



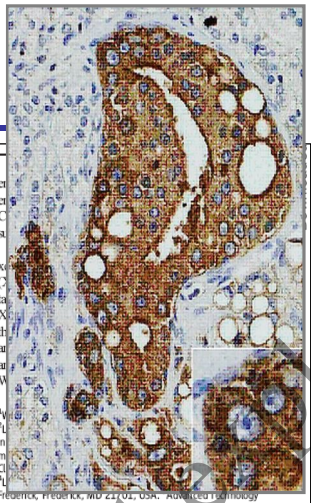
Experts' spreads for one parameter, and alternative ways of pooling weighted opinions

Experts' opinions on the time-to-failure (in days from first detection) of the 10%ile slowest cases



Note the "two schools of thought" effect…and the strong 'opinionation' of many experts

Big news!



Detection of an Infectious Retrovirus, XMRV, in Blood Cells of Patients with Chronic Fatigue Syndrome

Vincent C. Lombardi,[1]* Francis W. Ruscetti,[2]* Jaydip Das Gupta,[3] Max A. Pfost,[1] Kathryn S. Hagen,[1] Daniel L. Peterson,[1] Sandra K. Ruscetti,[4] Rachel K. Bagni,[5] Cari Petrow-Sadowski,[6] Bert Gold,[2] Michael Dean,[2] Robert H. Silverman,[3] Judy A. Mikovits[1]†

Chronic fatigue syndrome (CFS) is a debilitating disease of unknown etiology that is estimated to affect 17 million people worldwide. Studying peripheral blood mononuclear cells (PBMCs) from CFS patients, we identified DNA from a human gammaretrovirus, xenotropic murine leukemia virus–related virus (XMRV), in 68 of 101 patients (67%) as compared to 8 of 218 (3.7%) healthy controls. Cell culture experiments revealed that patient-derived XMRV is infectious and that both cell-associated and cell-free transmission of the virus are possible. Secondary viral infections were established in uninfected primary lymphocytes and indicator cell lines after their exposure to activated PBMCs, B cells, T cells, or plasma derived from CFS patients. These findings raise the possibility that XMRV may be a contributing factor in the pathogenesis of CFS.

Chronic fatigue syndrome (CFS) is a disorder of unknown etiology that affects multiple organ systems in the body. Patients with CFS display abnormalities in immune system function, often including chronic activation of the innate immune system and a deficiency in natural killer cell activity (*1*, *2*). A number of viruses, including ubiquitous herpesviruses and

Frederick, Frederick, MD 21702, USA. [5]Advanced Technology Program, National Cancer Institute–Frederick, Frederick, MD 21701, USA. [6]Basic Research Program, Scientific Applications International Corporation, National Cancer Institute–Frederick, Frederick, MD 21701, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: judym@wpinstitute.org

www.sciencemag.org SCIENCE VOL 326 23 OCTOBER 2009

XMRV Expert Elicitation Workshop

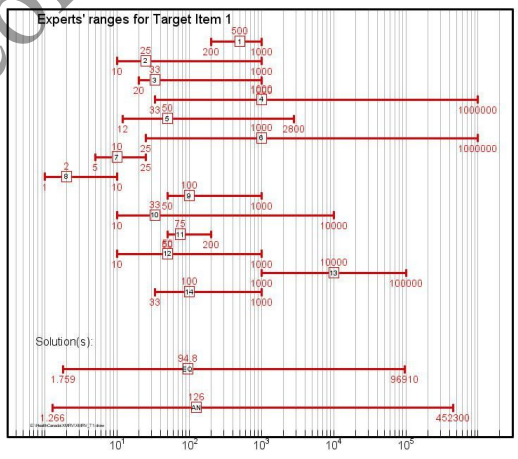
International panel in Ottawa, Canada

Target Question Grouping

Questions	Subject Area
1-7	Prevalence
8-11	Risk Parameters
12-15	Latency
16-22	Routes of Transmission
23-25	Risk Mitigation
26-30	Disease Relationships (causal and non-causal)

Target Questions 1, 3-6

A set of target questions that asked about the current prevalence of XMRV infection in the world (1), Canada (3), USA (4), UK (5) and France (6) in the general adult population? (1 in xxxxx)



Expert Weighted:

- 1 in 126
- Range: 1.2-452,300

Prevalence: Target Questions 1, 3-6

A set of target questions that asked about the current prevalence of XMRV infection in the world (1), Canada (3), USA (4), UK (5) and France (6) in the general adult population? (1 in xxxxx)

Country	Expert Weighted	Expert Range
Canada	1 in 334	1 in 12.1 – 1 in 305,500
USA	1 in 278.8	1 in 12.0 – 1 in 305,500
UK	1 in 450.2	1 in 12.4 – 1 in 305,500
France	1 in 450.2	1 in 12.4 – 1 in 305,500

Table. Published studies October 2009 to June 2011

First author, country	Journal, date	Patients positive for XMRV?
Lombardi, USA	Science, October 2009	Yes (67%)
Erlwein, UK	PLoS One, January 2010 & March 2011 (re-analysis)	No
van Kuppelweld, Netherlands	British Medical Journal, February 2010	No
Groom, UK	Retrovirology, February 2010	No
Swizer, USA	Retrovirology, July 2010	No
Lo, USA	Proc Natl Acad Sci, August 2010	No (but 86.5% MLV)
Hong, China	Virology Journal, September 2010	No
Henrich, USA	J Infect Dis, November 2010	No
Hohn, Germany	PloS One, December 2010	No
Satterfield, USA	Retrovirology, February 2011	No
Furuta, Japan	Retrovirology, March 2011	No
Schutzer, USA	Ann Neurol, April 2011	No
Shin, USA	Journal of Virology, May 2011	No
Knox, USA	Science, May 2011	No

Judgment in the face of scientific uncertainty: the last word in rationality…

46

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

nature climate change
ARTICLES
PUBLISHED ONLINE: 23 DECEMBER 2012 | DOI:10.1038/NCLIMATE1778

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

J. L. Bamber[1]* and W. P. Aspinall[2]

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 AR4 estimates of future sea-level rise from this source may have been underestimated. Various approaches for addressing this problem have been tried, including semi-empirical models and conceptual studies. Here, we report a formalized pooling of expert views on uncertainties in future ice-sheet contributions using a structured elicitation approach. We find that the median estimate of such contributions is 29 cm—substantially larger than in the AR4—while the upper 95th percentile value is 84 cm, implying a conceivable risk of a sea-level rise of greater than a metre by 2100. On the critical question of whether recent ice-sheet behaviour is due to variability in the ice sheet–climate system or reflects a long-term trend, expert opinion is shown to be both very uncertain and undecided.

news & views

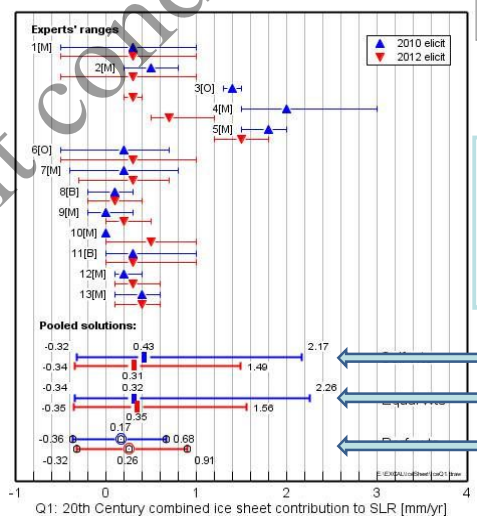
EXPERT JUDGEMENT ASSESSMENT

Quantifying uncertainty on thin ice

The contribution of ice sheets to sea-level rise still has large uncertainties that are yet to be quantified.

R. M. Cooke

Nature Climate Change 3, 311–312 (2013)
doi:10.1038/nclimate1860



2010
2012

[**O**] = **O**bservationalist
[M] = Modeler

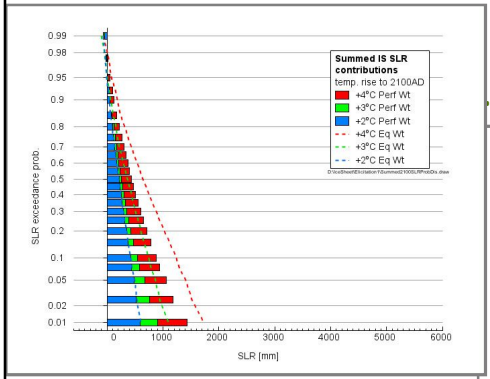
Self – weights

Equal weights

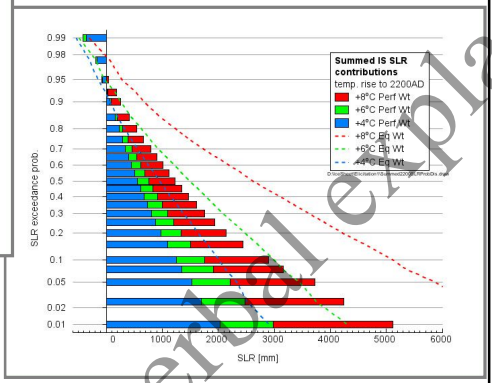
Performance based weights

Pooled expert judgements on combined ice-sheet contributions to sea-level rise:

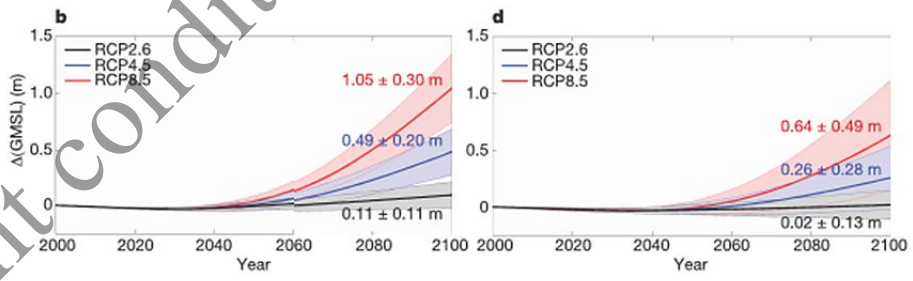
2100CE (+2° C, +3° C, +4° C) - Perf Wts & Equal Wts pooling



, and 2200CE (+4° C, +6° C,



2016 context: possible/plausible future sea-level rise due to (Antarctic) ice sheet melting …..



DeConto & Pollard (2016 Nature doi:10.1038/nature17145):
"Antarctica has the potential to contribute more than a metre of sea-level rise by 2100 and more than 15 metres by 2500, if emissions continue unabated. In this case atmospheric warming will soon become the dominant driver of ice loss, but prolonged ocean warming will delay its recovery for thousands of years."

Variables in Bamber & Aspinall elicitation

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

$$SLR_{IS} = Discharge_{IS} + Runoff_{IS} - Accumulation_{IS}$$

where $_{IS}$ denotes GrIS, WAIS or EAIS.

- Quantities were elicited for +2°C, +3°C and +4°C warming scenarios by 2100CE and for +4°C, +6°C and +8°C warming by 2200CE.

- Contribution units were elicited from experts in terms of mass change, converted to millimeters sea level rise.

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 experts agreed that uncertainties are large and dependences could have an appreciable effect.

- Dependence between variables of interest is often relatively benign - often uncertainties on variables strongly dominate any "dependence effect" – but this is not always the case: use of tail independent Normal copulae has been charged with inducing excessive risk taking on Wall Street

Salmon, F., 2009, "Recipe for Disaster: The Formula That Killed Wall Street" Wired, February 23: http://www.wired.com/techbiz/it/magazine/17-03/wp_quant?current Page=all)

Dependence elicitation

- Dependence between variables X and Y can be 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.

Remarks

- Therefore to allow for the possibility of tail dependence, experts are asked, in addition to the 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?

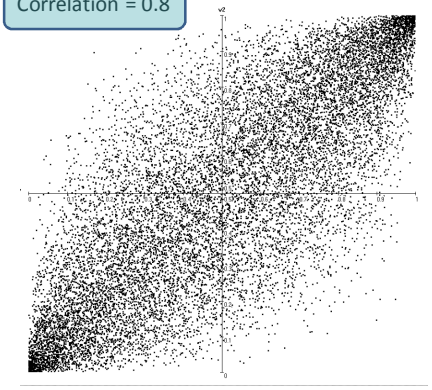
  - If these uncertainties are independent, the elicited probability should 0.05; probabilities greater than 0.05 indicate positive association, less than 0.05 indicate negative association.

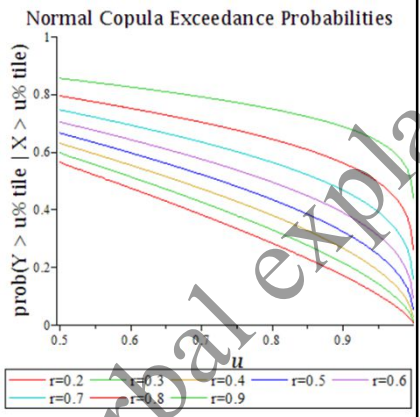
  Observing the relations between these two elicited exceedance probabilities, a choice can be made for analysis, typically, from one-parameter copula families: those with tail independence (e.g. Normal or Frank), and those with tail dependence (e.g. Gumbel or reverse Clayton).

Tail independence: Normal Copula

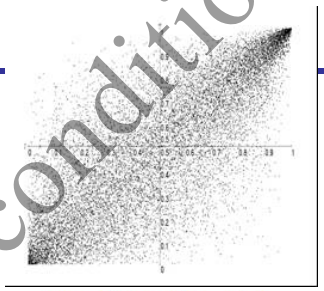
$$\frac{P(Y > \text{u-th quantile} \mid X > \text{u-th quantile})}{1 - u} \rightarrow 1 \text{ as } u \rightarrow 1$$

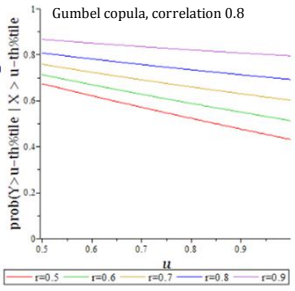
Correlation = 0.8



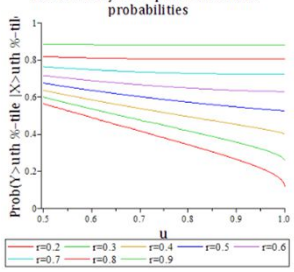


Other copulas are available





Reverse Clayton copula exceedance probabilities

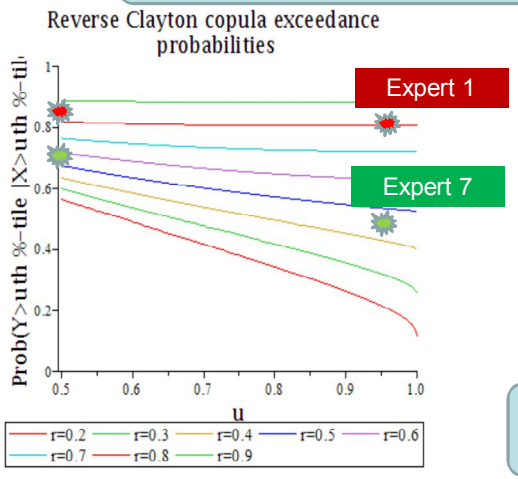


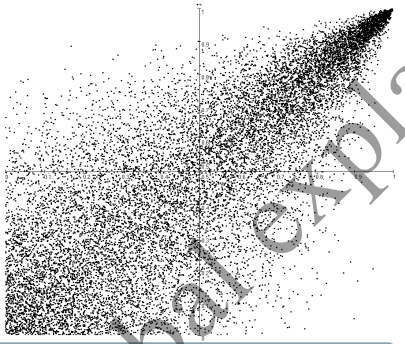
The message from the Gumbel or Reverse Clayton copula is:

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"..

Eliciting upper tail dependence in ice sheet processes

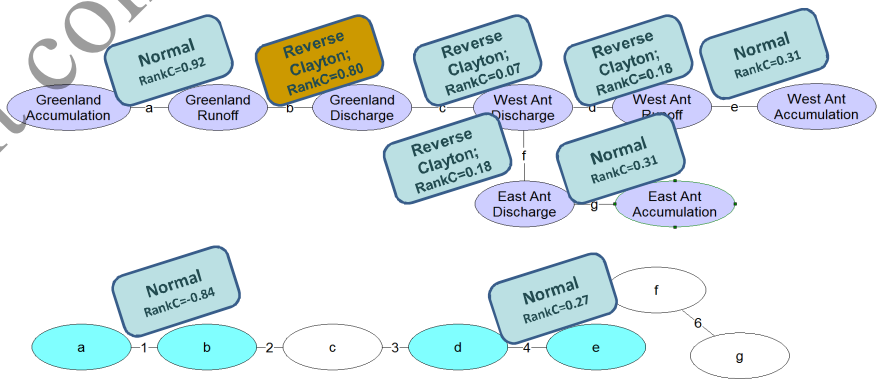
Joint conditional quantile probabilities for Greenland ice sheet
Pr (Discharge > $Disch_{50\%ile}$ | Runoff > $Runoff_{50\%ile}$)
Pr (Discharge > $Disch_{95\%ile}$ | Runoff > $Runoff_{95\%ile}$)





Expert 1 : correlation ~ 0.8
strongly upper tail dependent (above)
Expert 7 : correlation ~ 0.4

Representing Expert 1's elicited dependences with a nested low-order Vine



Bedford, T, Daneshkhah, A & Wilson, KJ 2015, 'Approximate uncertainty modeling in risk analysis with vine copulas' Risk Analysis., 10.1111/risa.12471

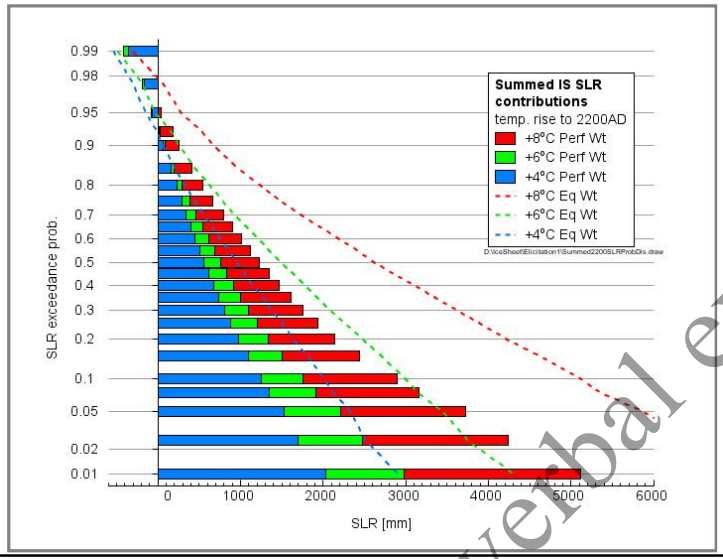
Examining tail effects of experts' dependence elicitations:

Ice sheet contribution to SLR in 2100 @ 3C warming [mm]					
	Mean	StDev	5%-ile	50%-ile	95%-ile
Expert 1 indep	295	196	53	263	691
Expert 1 Tail Dep	295	222	53	249 ↓	752 ↑
Expert 7 indep	407	189	171	377	787
Expert 7 Tail Dep	406	225	150	359 ↓	861 ↑

Uncertainty and dependence effects, and influence of judgment pooling method:

	Ice sheet contribution to SLR in 2100 @ 3C warming [mm]				
Decision maker (expert combination)	Mean	StDev	5%-ile	50%-ile	95%-ile
Performance weights, Independent	335	200	71	307	719
Performance weights, Normal copulae	337	216	64	305	749
Performance weights, Tail Dependent	338	229	71	292	785
Equal weights, Independent	615	270	238	581	1120

Pooled expert judgments on combined ice-sheet contributions to SLR: to be revised in light of tail dependence analysis …..



Summing up:

It is feasible to elicit multivariate dependences AND tail dependences from experts, and apply these to uncertainty modelling.

In the case of ice sheet melting due to global warming, the impact of dependence between ice sheets on tail estimates of sea level rise in 2100CE is not huge, but does make SLR tails marginally heavier.

For other complex geophysical processes, dependence effects could be more significant for extreme event estimation …….

Cross-Validation / research

- **Supplementary Online Material - Cross Validation**

By Roger M. Cooke - Resources for the Future & Delft Univ. of Technology Dept Mathematics [Nov 9, 2015]

- Try: *rogermcooke.net* or email me: *willy@aspinall.associates*

Thank you!

Classical Model elicitation – a few published case histories

- *Bamber, J. and Aspinall, W.P. (2013) An expert judgement assessment of future sea level rise from the ice sheets. Nature Climate Change, 3, doi:10.1038/nclimate1778*
- *Tyshenko, M.G., et al (2012) Expert judgement and re-elicitation for prion disease risk uncertainties. International Journal of Risk Assessment and Management, 16(1-3), 48-77.*
- *Aspinall, W. (2010) Opinion - A route to more tractable expert advice. Nature 463, 294-295.*
- *Cooke, R.M. and Goossens, L.L.H.J. (2008) "TU Delft expert judgment data base." Reliability Engineering & System Safety 93, 657-674.*
- *Aspinall, W.P. (2006) Structured elicitation of expert judgment for probabilistic hazard and risk assessment in volcanic eruptions. In: Mader, H.M., Coles, S.G., Connor, C.B. & Connor, L.J. (eds) Statistics in Volcanology. Special Publications of IAVCEI, 1. Geological Society, London, 15-30.*

Classical Model elicitation – be careful!!

"I know nothing about the subject, but I'm happy to give you my expert opinion."

Thank you