The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific
and Cultural Organization

International Atomic
Energy Agency

H4.SMR/1645-19

"2nd Workshop on Earthquake Engineering for Nuclear Facilities: Uncertainties in Seismic Hazard"

14 - 25 February 2005

Performance-based expert judgement - a structured elicitation approach & case histories

W. Aspinal

Aspinall & Associates
Bucks, U.K.

IAEA/ICTP Workshop on
Earthquake Engineering for Nuclear Facilities -
Uncertainties in Seismic Hazard Assessment

"Performance-based expert judgement - a structured elicitation approach & case histories"

Trieste, Italy, 14 – 25 February 2005

Unit 18 – Willy Aspinall

Aspinall & Associates (UK)

Willy@aspinall.demon.co.uk

INTRODUCTION

In this morning's session, we showed an example illustrating inter-expert variation in a PSHA for a low seismicity area……

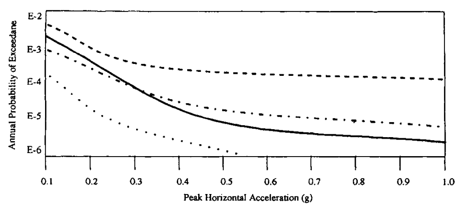


Figure 1: Inter-expert variation in expected seismic hazard curves for a site with sparse regional data

……and a plot of inter-expert variability when providing parameters for a seismic hazard model

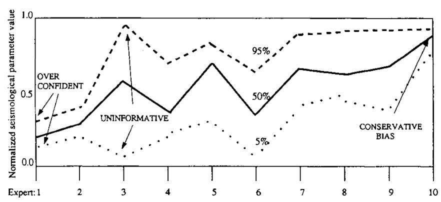


Figure 2. Inter-expert variation in 5%, 50% and 95% estimates of seismological parameters

Recognising that not all subject-matter experts are equal, in this presentation we describe a formal procedure that can be used to provide a performance-based rankings for experts, their judgements and opinions.

We start with a short description of expert judgement elicitation in a volcanic eruption crisis

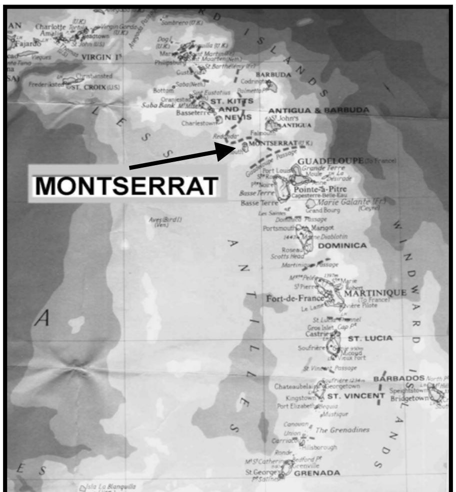
First, I must acknowledge:

Dr. Gordon Woo, Prof. Roger Cooke and Prof. Steve Sparks FRS

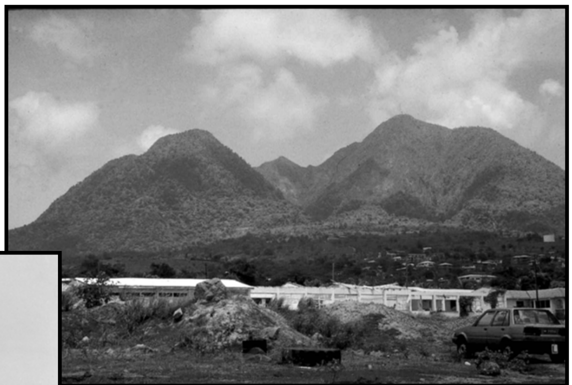
Montserrat Volcano Observatory
British Airways
Kellogg Brown & Root / DEFRA

Institute for Advanced Studies, Bristol University

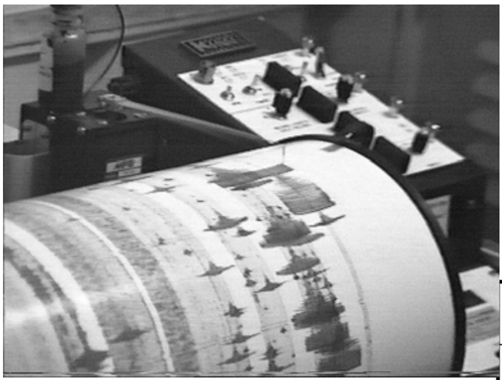
and the Montserrat eruption, 1995…….

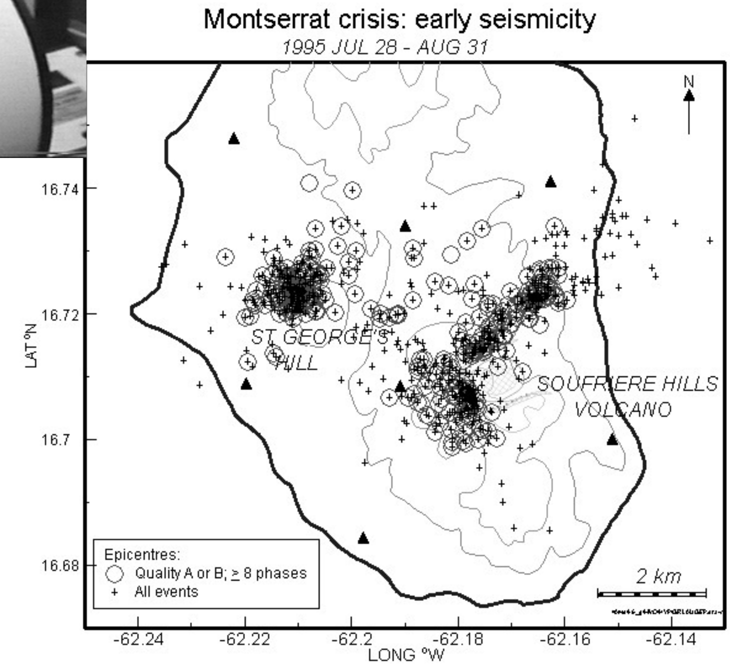


Soufrière Hills, Montserrat, in former times……



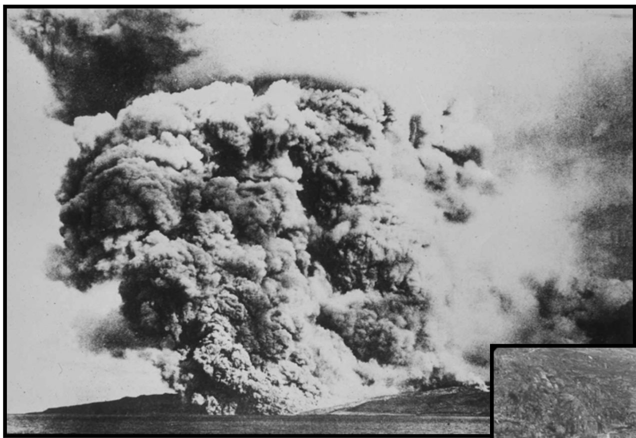
…and in July 1995





Precursory seismic activity………

A regional history of volcanic disasters in the Eastern Caribbean



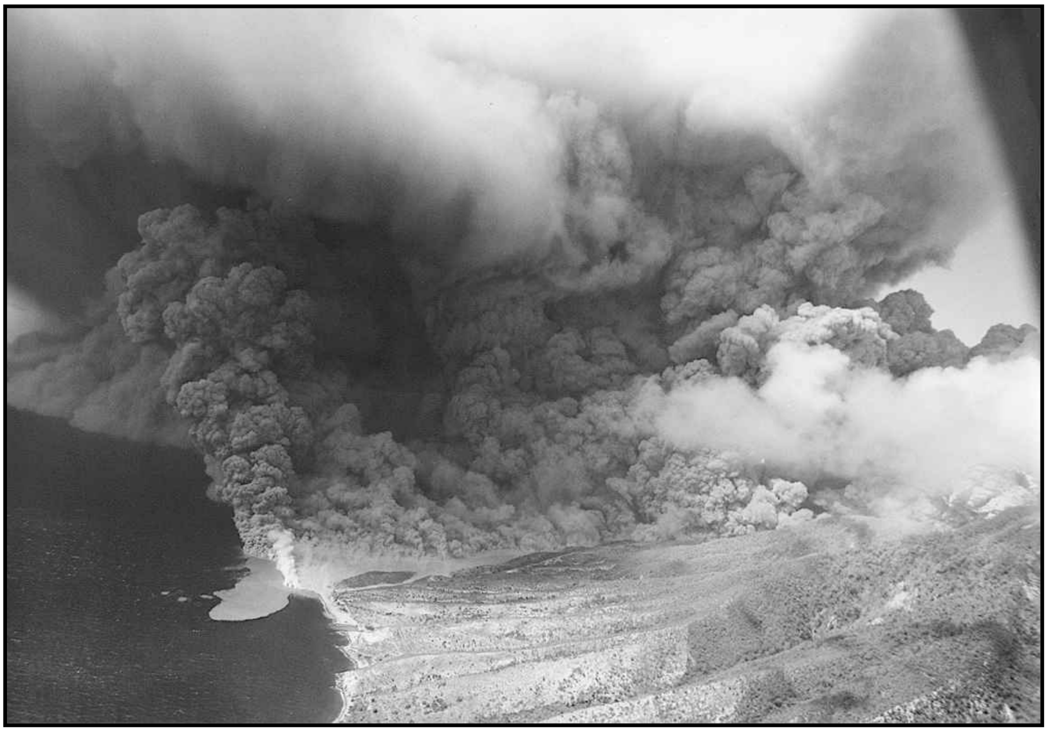
Mt Pelée, Martinique 1902

29,000 people die when political priorities take precedence over public concerns

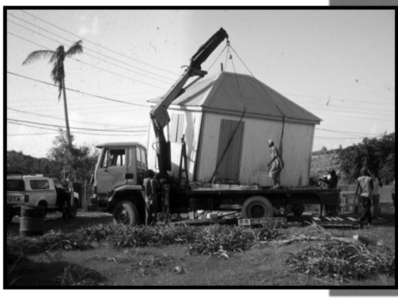
Then, Guadeloupe, 1976……

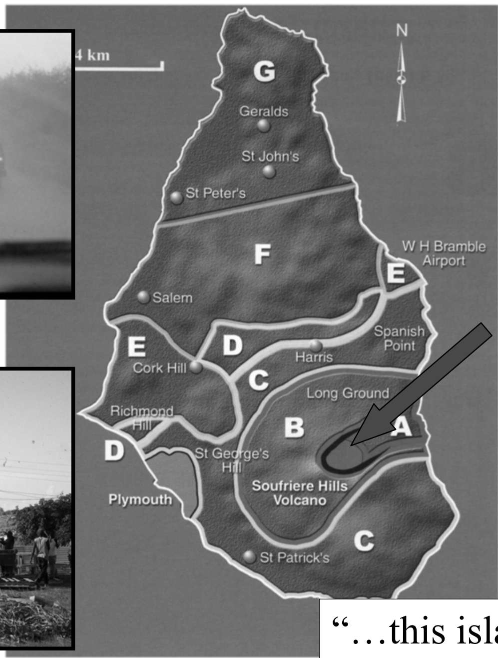
….a volcanic crisis leads to a major evacuation, but the eruption is stillborn; scientists are embroiled in public controversy, severe criticism and recriminations

In Montserrat, a magmatic eruption is confirmed, and escalates progressively in intensity and danger….



Living with an erupting volcano: hazard zones for crisis micro-management

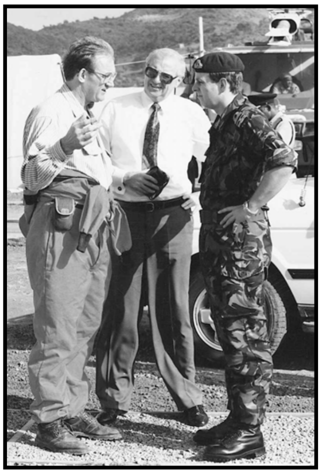




"…this island is exactly the wrong size for an eruption…"

Prompted by the Guadeloupe 1976 experience….

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



using a procedure developed originally for the European Space Agency

Alternative approaches to pooling expert opinions:

simple averaging

committee

decision conferencing (Bonano 1990)

the Delphi method

equal weights (Coppersmith & Youngs 1990)

expert self-weighting (TERA 1980)

group mutual weightings

mathematical theory of scoring rules $\Rightarrow$

Cooke (1991): "Classical" model for pooling opinions

and implementation in the EXCALIBR program

The basis of Cooke's "classical" model

Given a set of known (or knowable) seed items, for each expert

test hypothesis H_0: "This expert is well calibrated", leading to likelihood of acceptance at some defined significance level, and use this likelihood to define his **Calibration score:**

$$C_j = 1 - \chi_R{}^2(2 * M * I(s_j, p) * Power)$$

…where *j* denotes the expert, *R* is no. of quantiles (=degrees of freedom), *M* is the number of seed variables used in calibration, and *I(s,p)* is a measure of information.

Cj **corresponds to the asymptotic probability of seeing a deviation between *s* and *p* at least as great as *I(s,p),* under the hypothesis.**

The basis of Cooke's "classical" model

– **Entropy score**

estimate individual's information score relative to a uniform or loguniform density function from:

$$I_j(s_j, p) = \frac{1}{n}\sum_{i=1}^{n} s_i \ln(\frac{s_i}{p_i})$$

where ***s_i is a sample distribution obtained from the expert on the seed variables, and p_i is a suitable reference density function, depending on the appropriate scaling for the item.***

The basis of Cooke's "classical" model

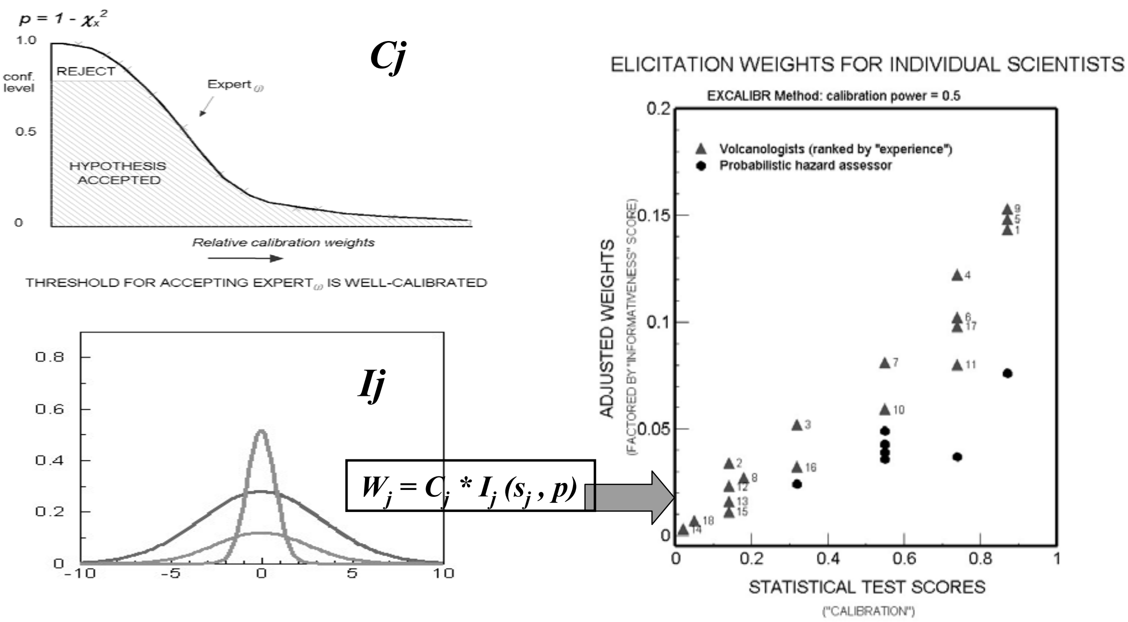
– **Individual's expert weighting**

compute individual's weight from product of his Calibration and Entropy scores (where the latter is now estimated from all variables, seeds and unknowns):

$$W_j = C_j * I_j(s_j, p)$$

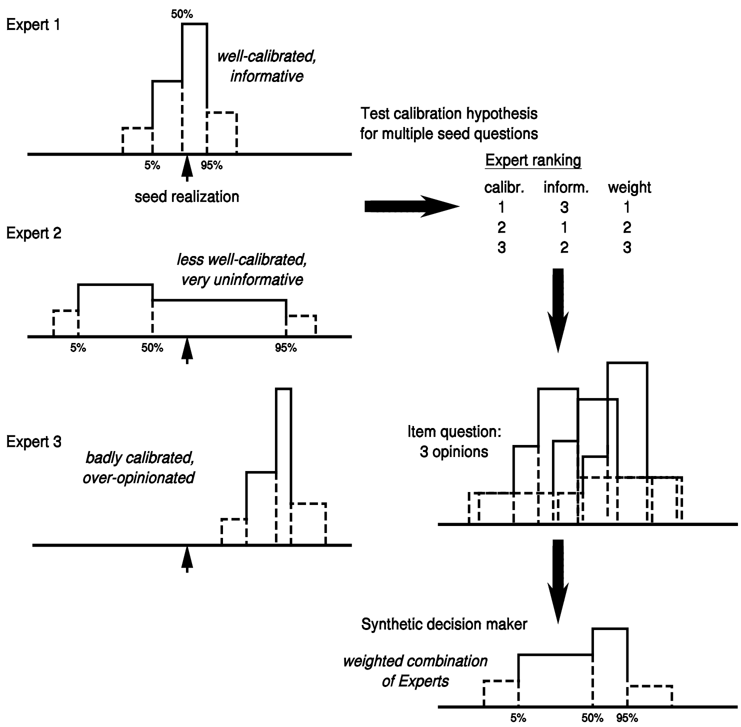
and normalise the W_j across all experts to get relative weights.

EXCALIBR: ranking opinions of individual experts by 'asymptotically proper' scoring rules



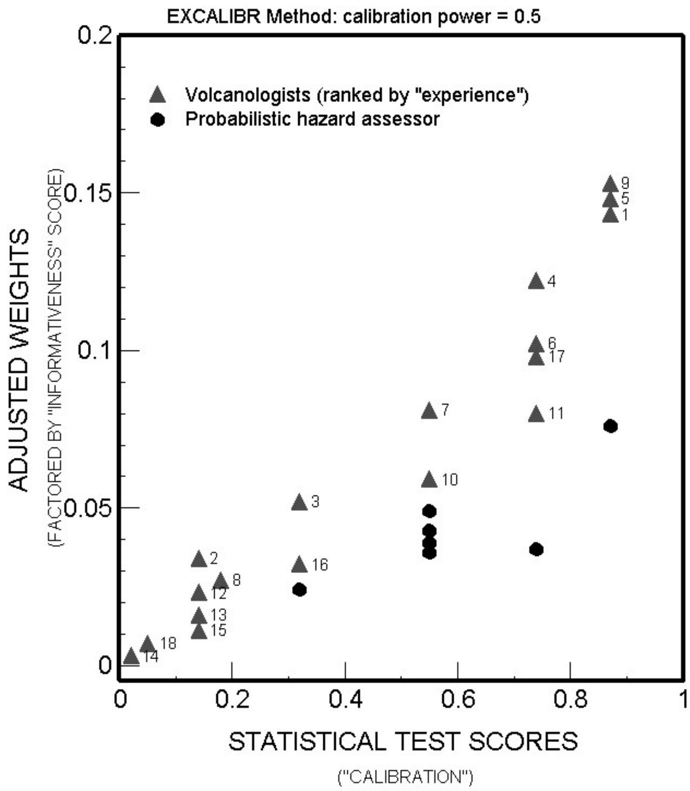
An optimal decision on any question of interest can then be obtained from the weighted sum of the opinions of a group of experts:

$$DM_i = \Sigma W_j * Q_i$$



The art of being a good expert is to get your net just the right size for catching the 'scientific fish'……

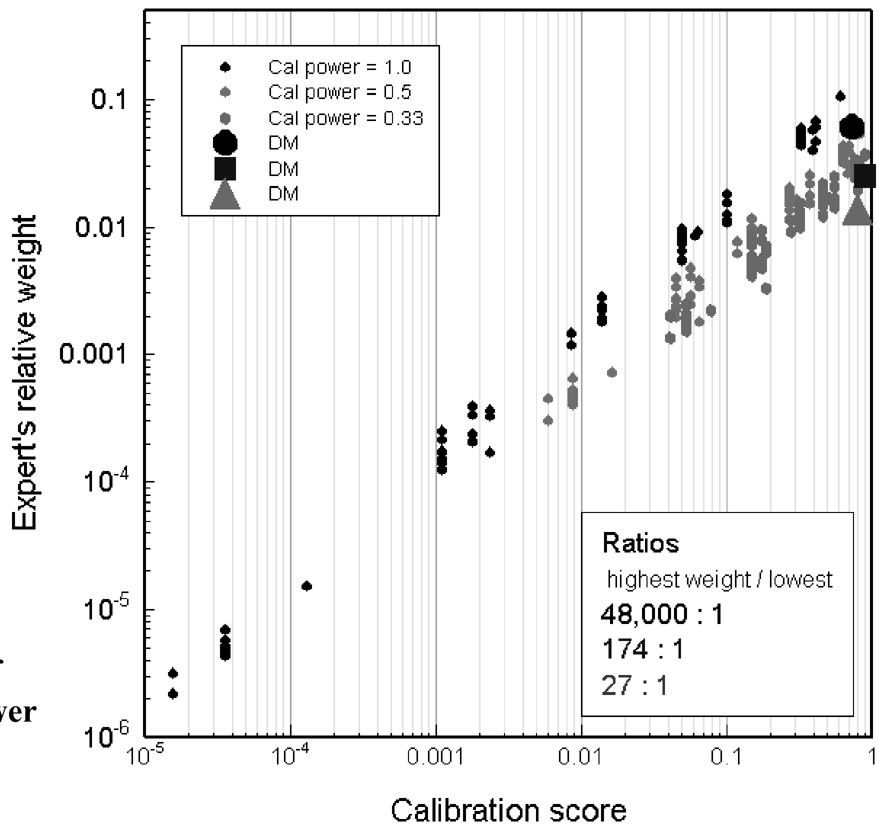
ELICITATION WEIGHTS FOR INDIVIDUAL SCIENTISTS

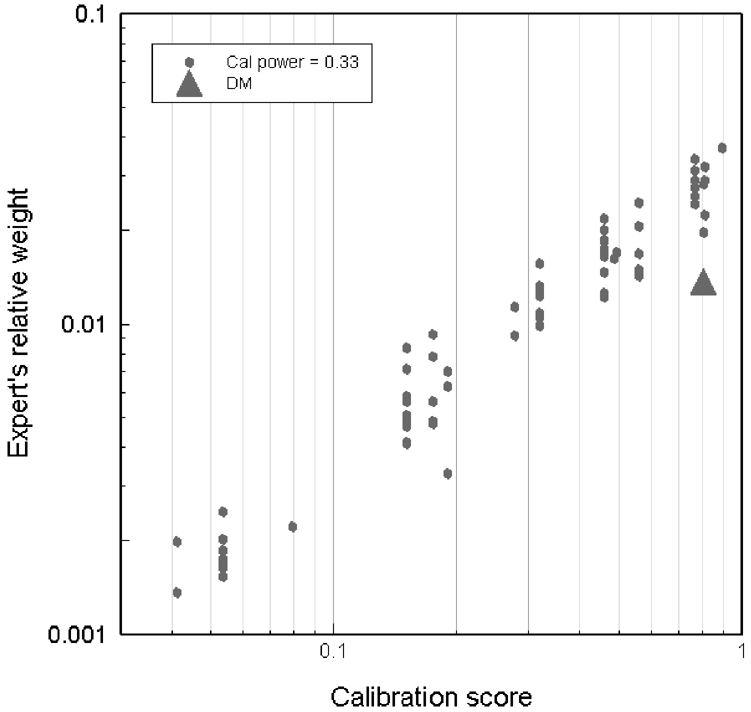


"The method is considered superior to other schemes for pooling judgements, using asymptotically strictly proper scoring rules that reward good calibration and penalise low informativeness"

Calibrations of 75 'experts' at MVO…..

and influence of Calibration Power

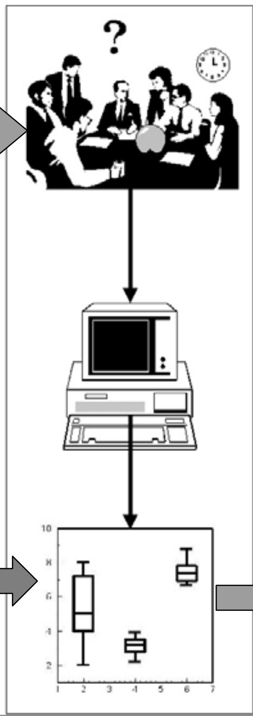




Typically, for the volcano work we have used a low Calibration Power setting, to constrain the range of weights that are applied for decision-making

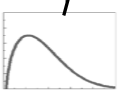
---

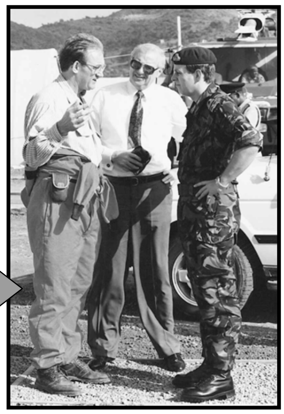
MVO structured elicitation procedure for scientific advice



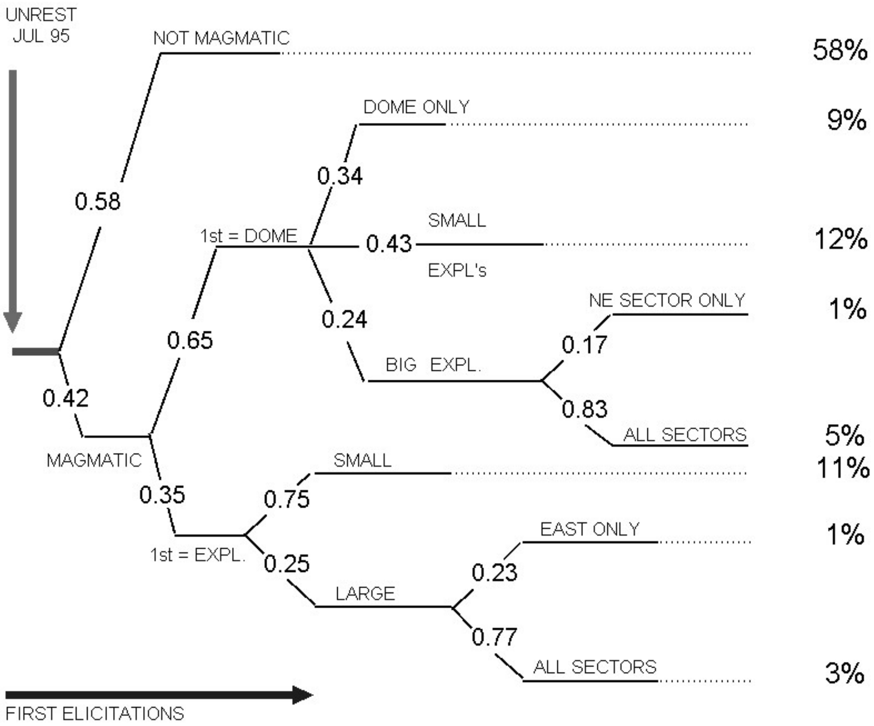
$$DM_i = \Sigma W_j * Q_i$$

Synthesised group "Decision-Maker"

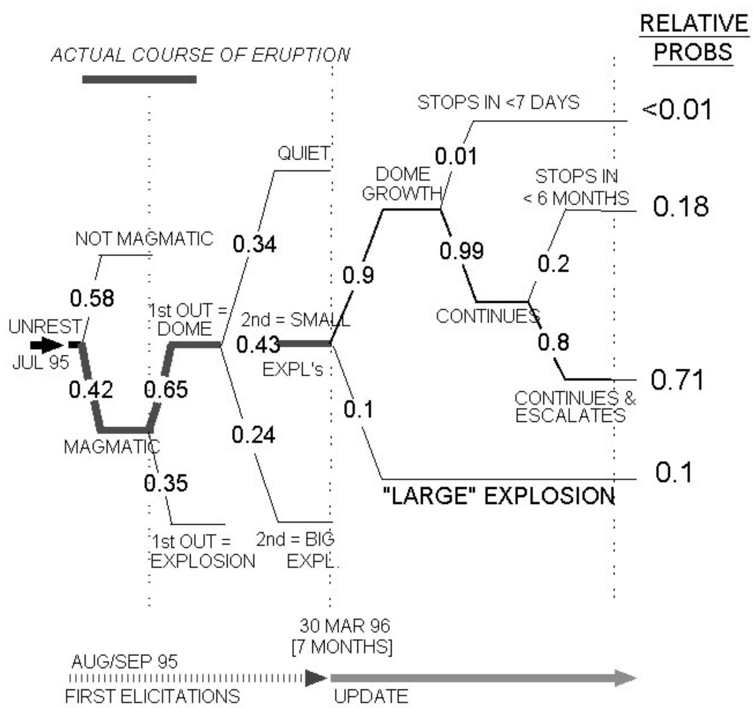




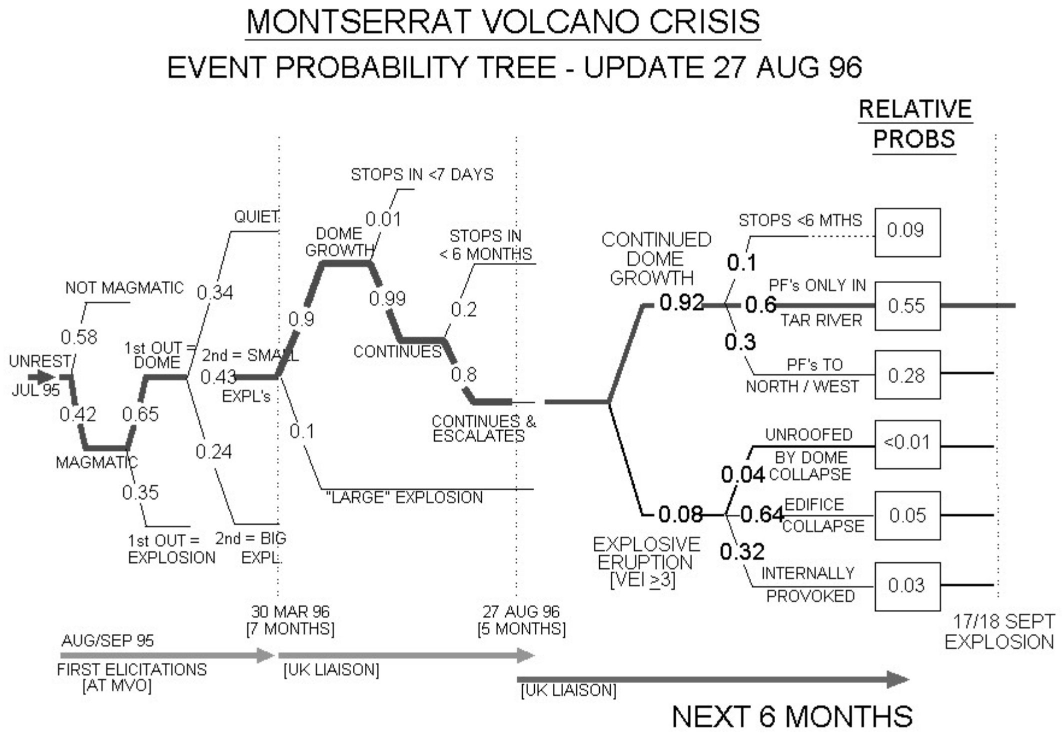
"…that's not the precise number I would have used, but I can't argue with it"

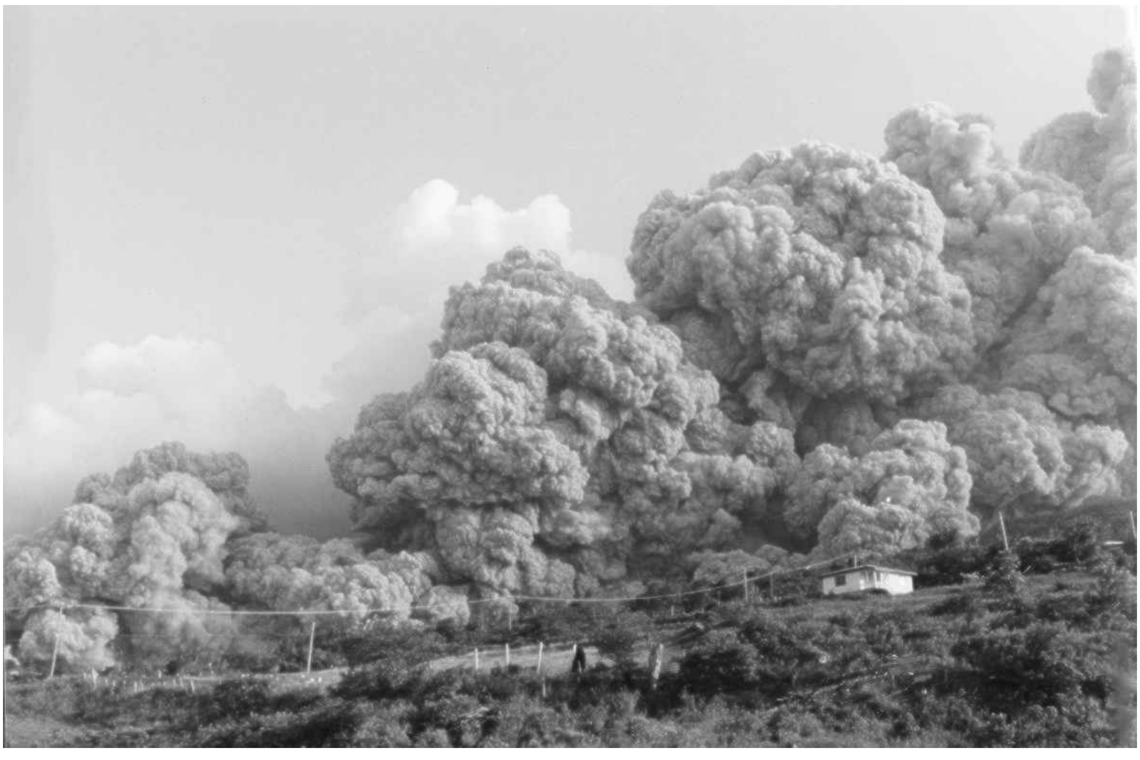


VOLCANO CRISIS: INITIAL PROBABILITY TREE



MONTSERRAT: EVENT PROBABILITY TREE - 1st UPDATE

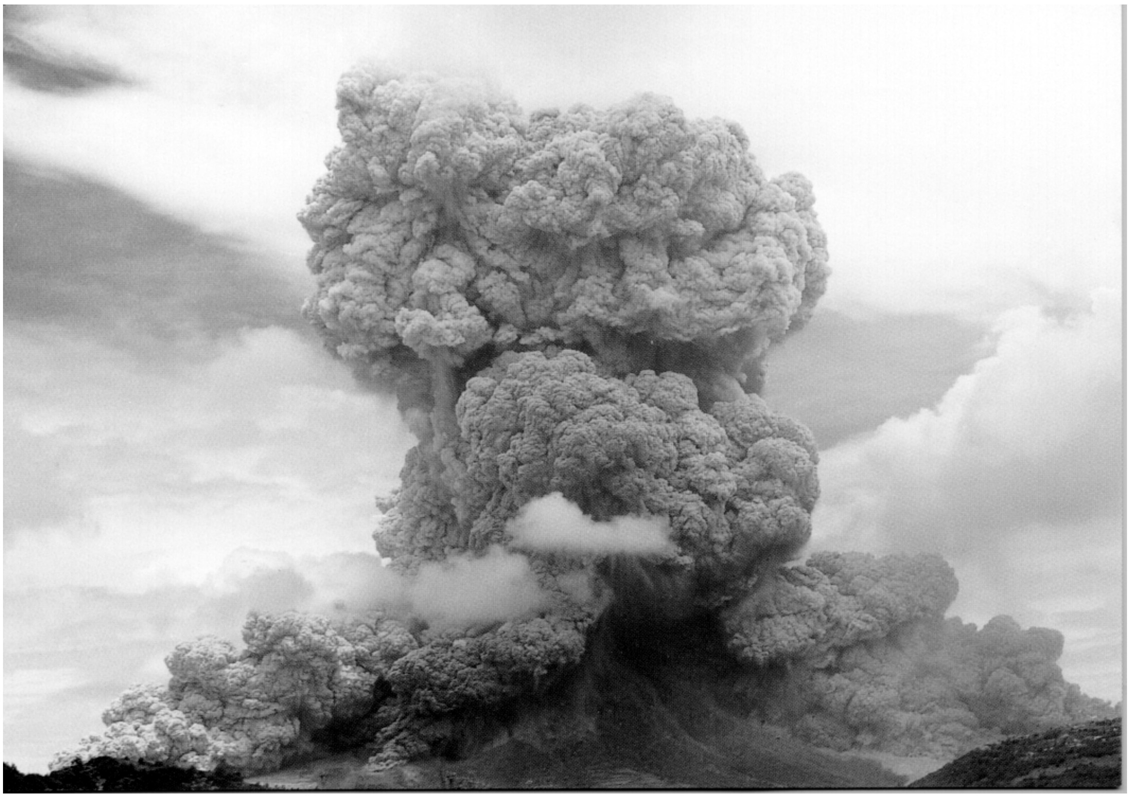




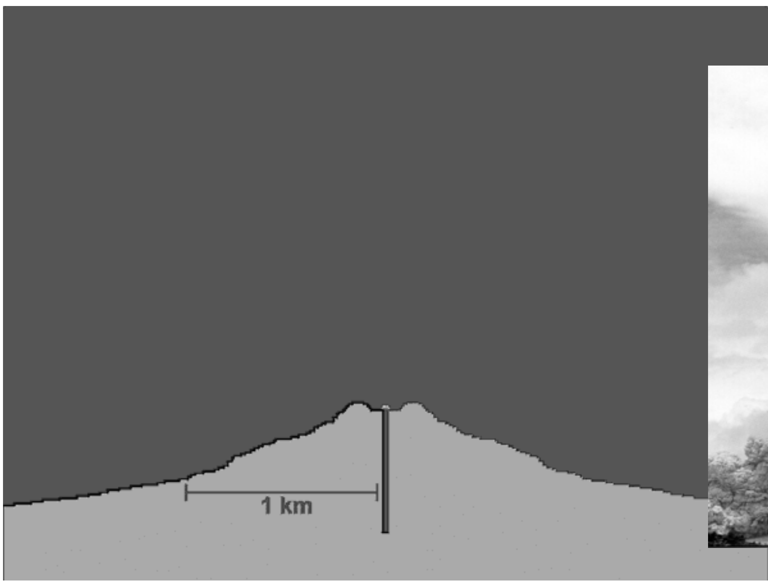
The pyroclastic flows get bigger, ………

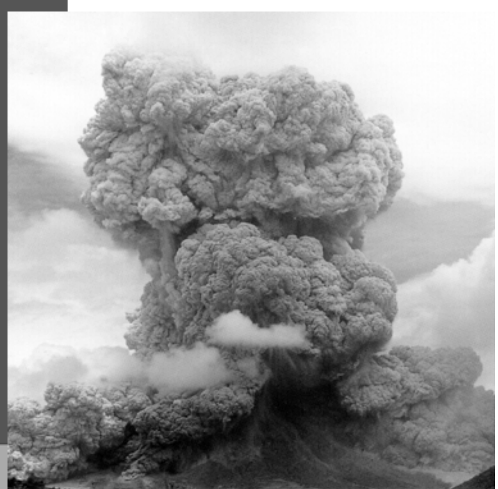
The ‘big one’ starts…….

Nineteen people die in the Danger Zone

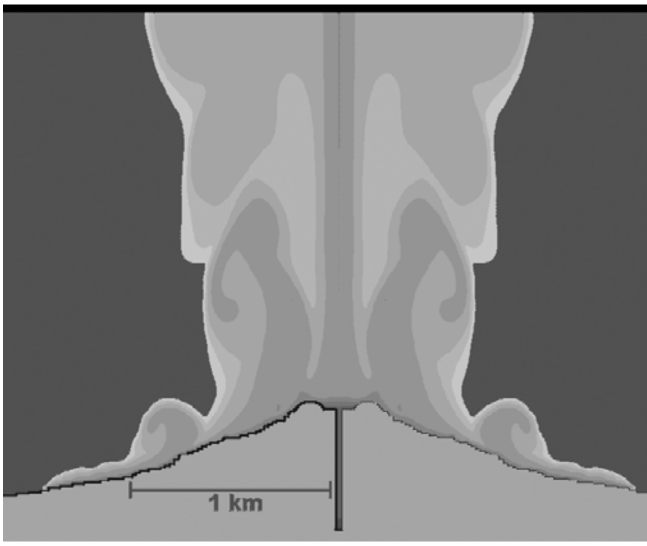


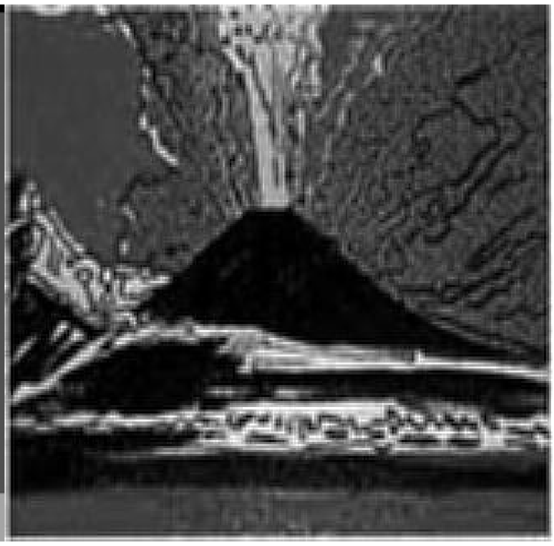
August 1997, eruption style turns more explosive


1 km



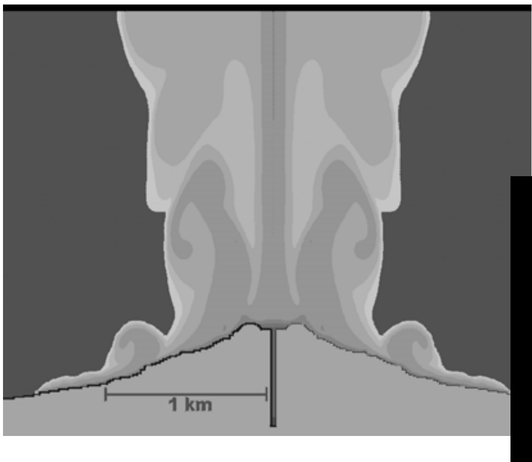
The scientist's view…..





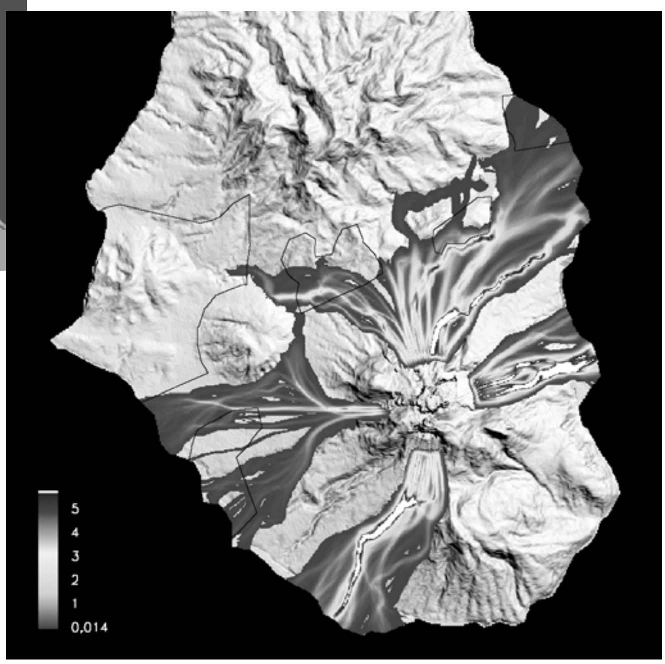
…………and the artist's view

Andy Warhol 1985

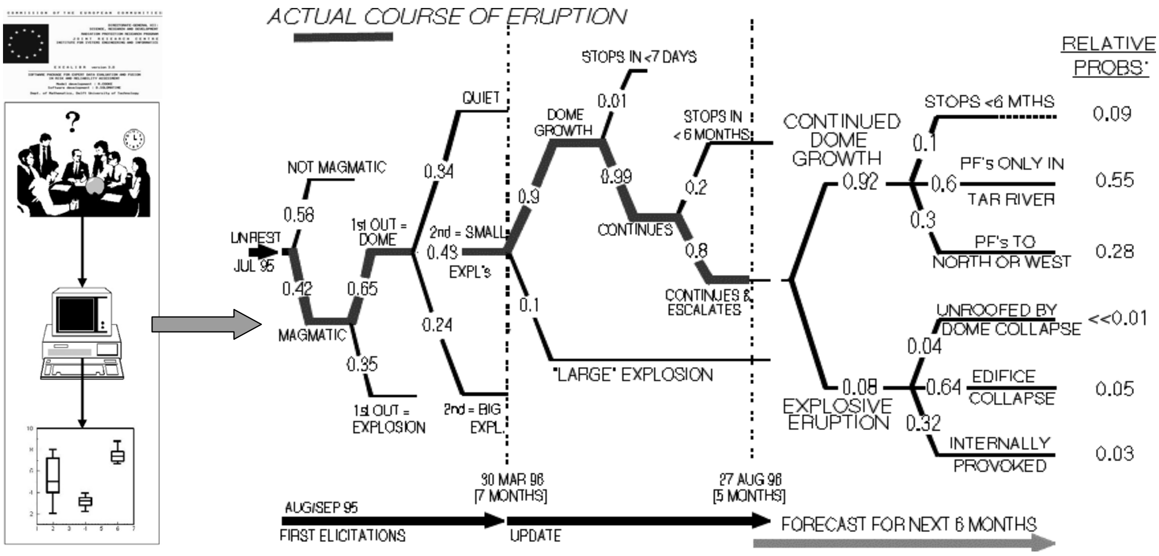


…..from computer simulation…

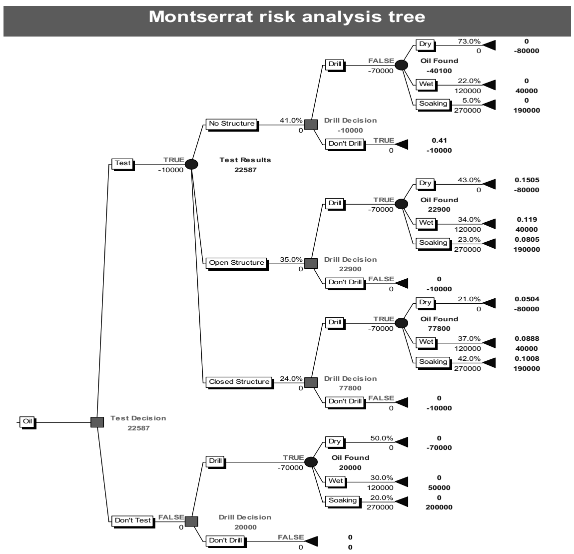
……. to area risk map



Structured elicitations used to construct and update volcanic event probability trees



……which can be linked to specific localities…………….



Monte Carlo simulation of numbers of potential casualties using parameter uncertainty distributions in a logic tree formulation - with distributions derived through elicitation

Event tree hazard scenarios

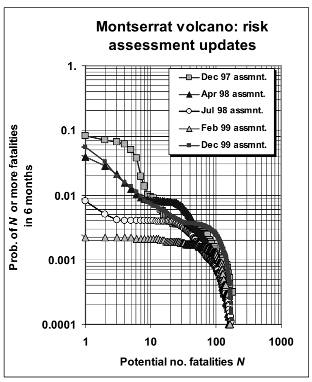
intensity/magnitude – probability

impact areas

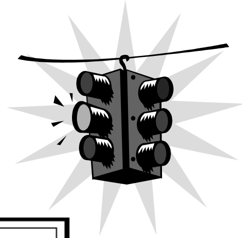
population vulnerability

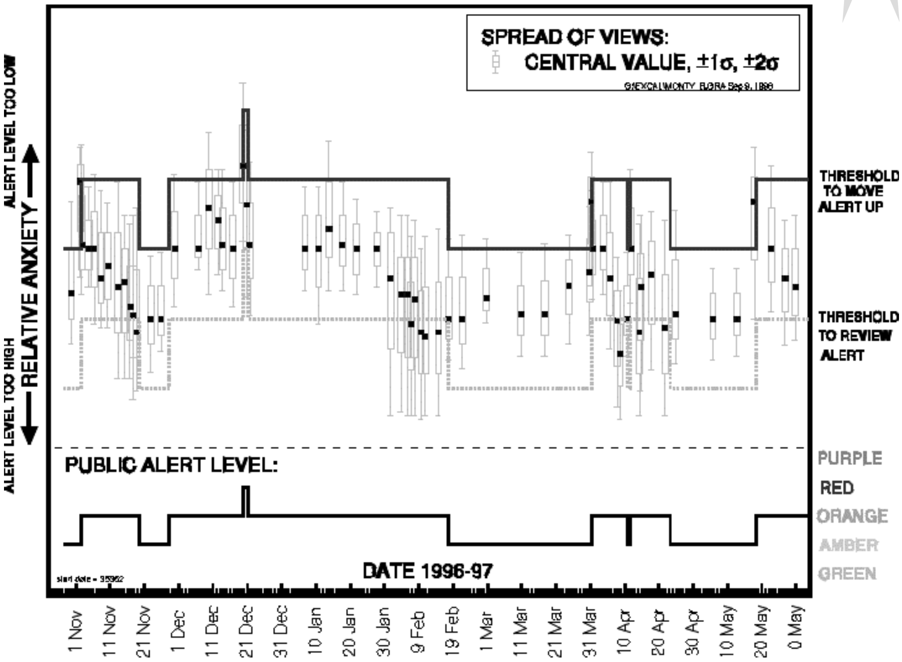
F-n casualty curves

(probability-exceedance)

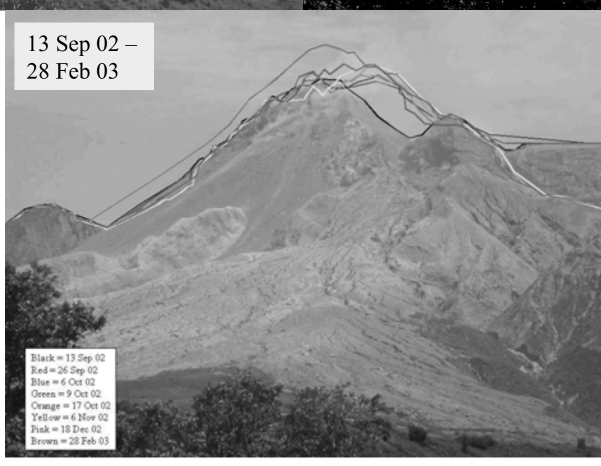


Elicitations also used in Montserrat to make prompt and traceable decisions on public alert levels

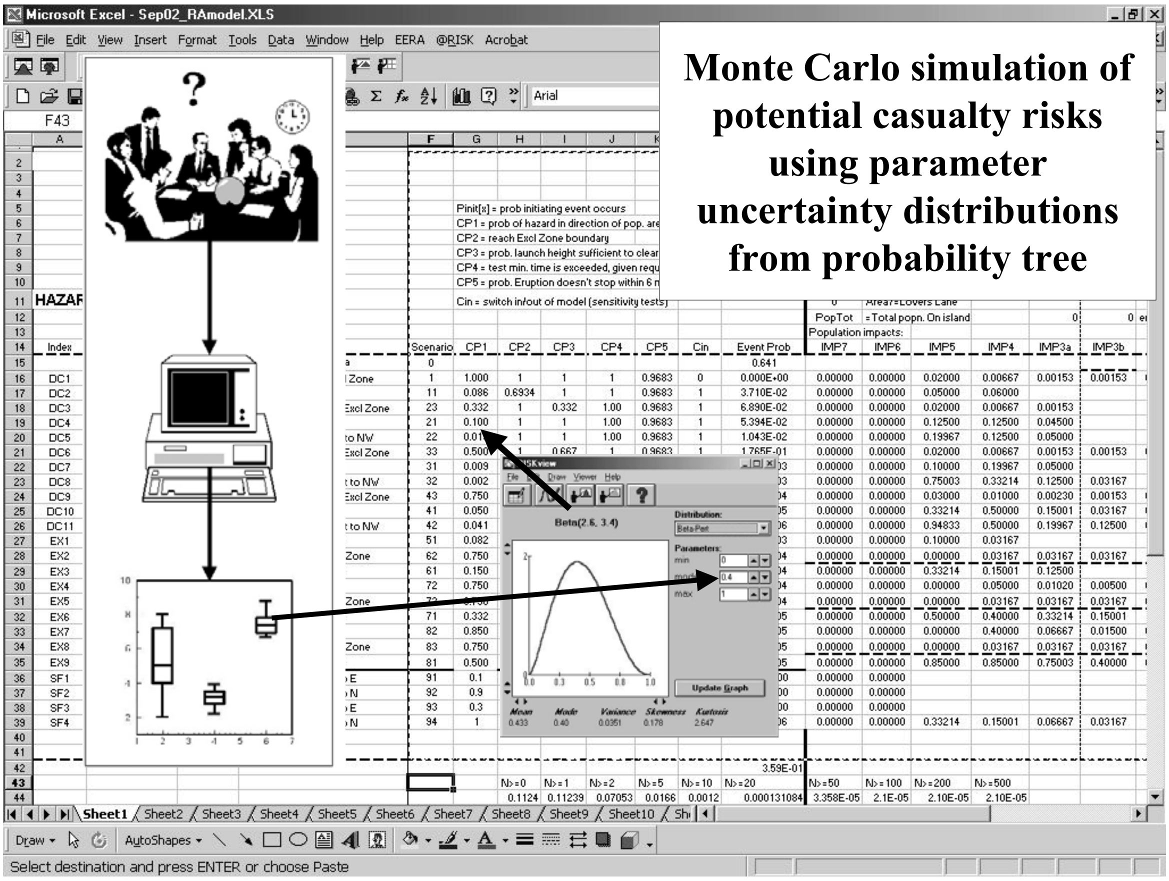


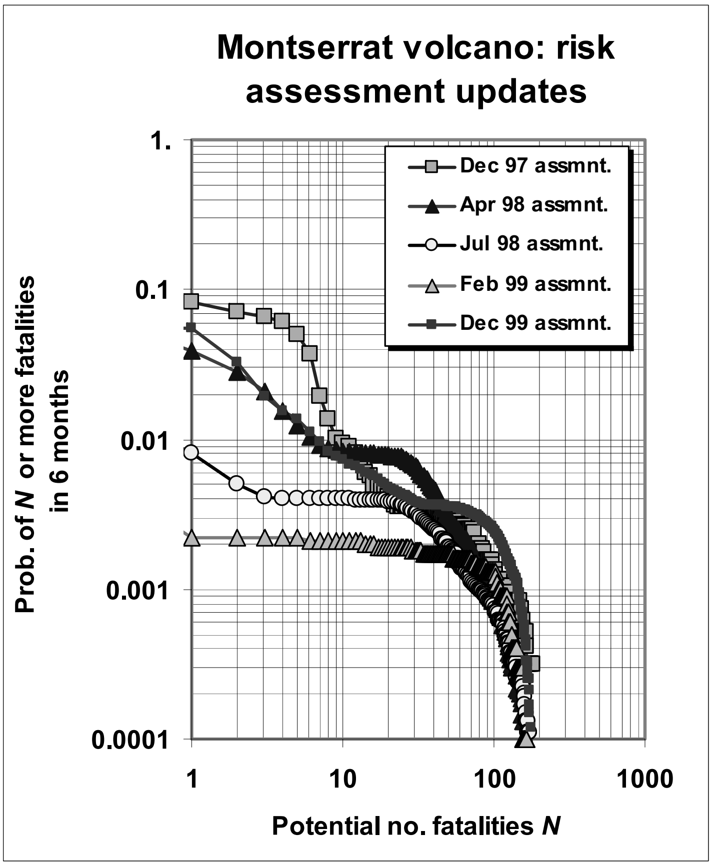


Profile changes:



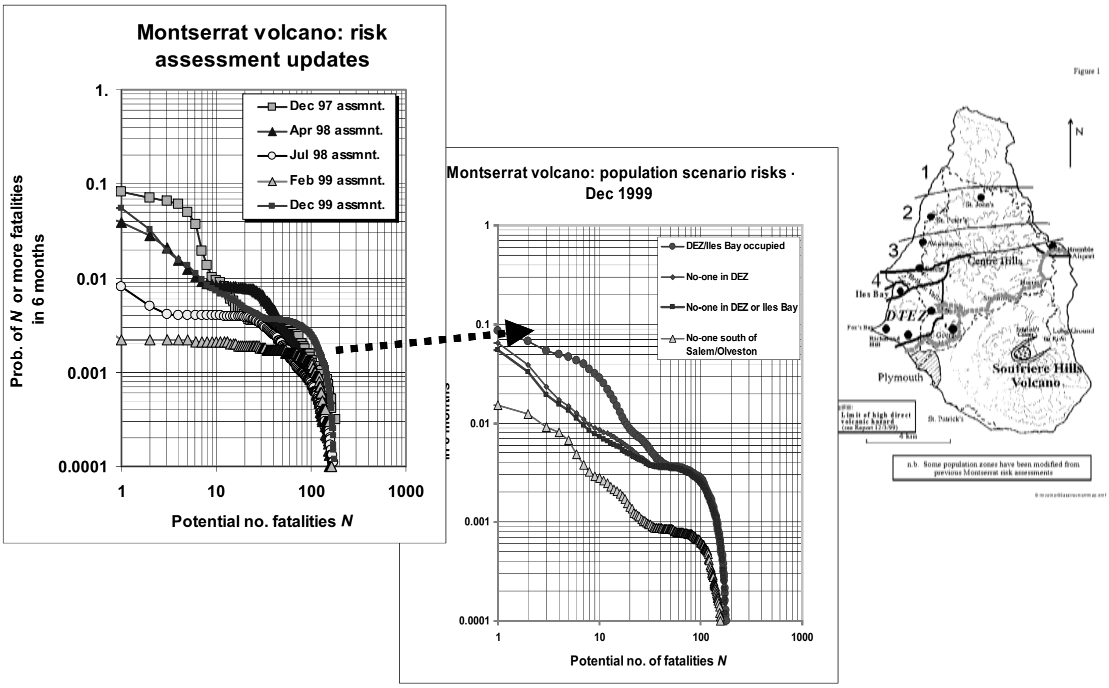
Monte Carlo simulation of potential casualty risks using parameter uncertainty distributions from probability tree



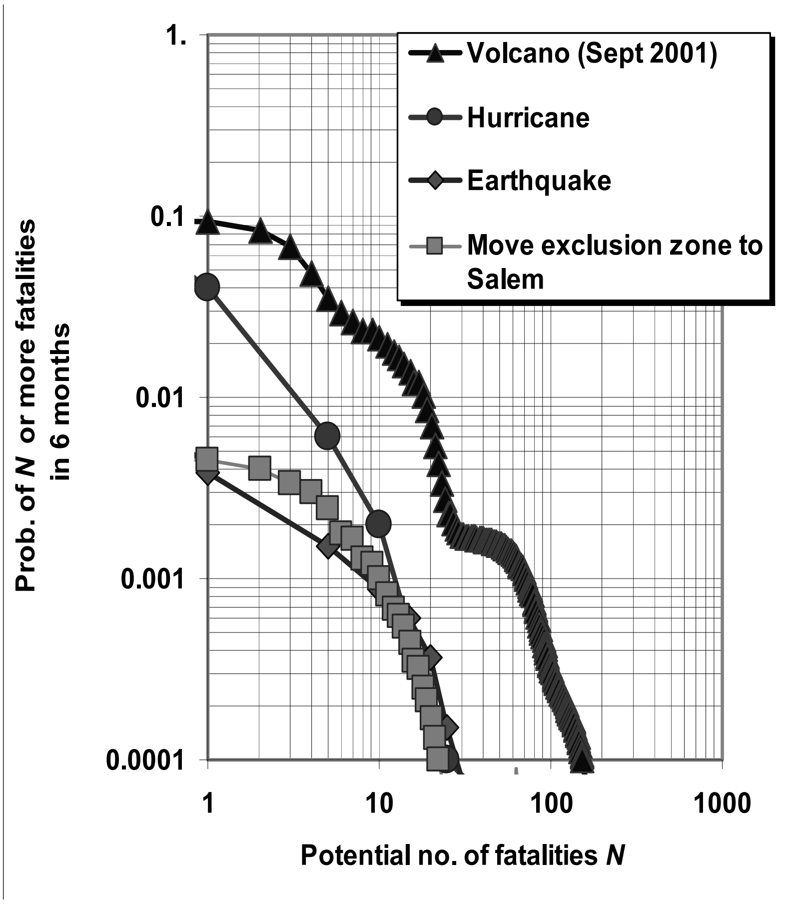


Typical results of Monte Carlo simulations of potential casualties in Montserrat

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



Comparison of volcanic risks with other natural hazards encountered in Montserrat



2002 – 2003, the biggest dome…… so far

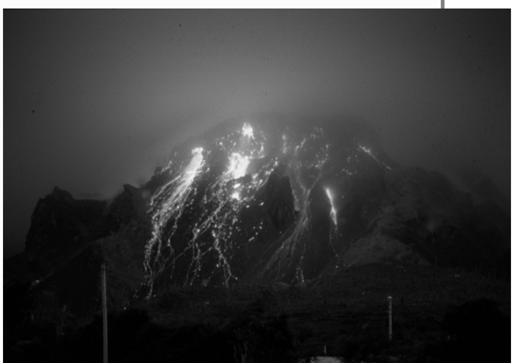
some 'O' level physics

Magma:			
density		**2500 kg/cu m**	
total vol of dome	**2.00E+08 cu m**		
total mass	**5.00E+11 kg**		500,000,000 tonnes

some 'O' level physics

Magma:			
density	**2500**	**kg/cu m**	
total vol of dome	**2.00E+08**	**cu m**	
total mass	**5.00E+11**	**kg**	500,000,000 tonnes
Say, collapse volume =	**1.00E+08**	**cu m**	
collapse mass	**2.50E+11**	**kg**	250,000,000 tonnes

some 'O' level physics



Magma:			
density	**2500**	**kg/cu m**	
total vol of dome	**2.00E+08**	**cu m**	
total mass	**5.00E+11**	**kg**	
Say, collapse volume =	**1.00E+08**	**cu m**	
collapse mass	**2.50E+11**	**kg**	
dome temp	**770**	**degC**	
(degC above ambient)			
spec heat of magma	**1254**	**J/kg/degC**	
dome collapse time	**10800**	**sec**	
3 hrs			
Power dissipated	**22,359,944**	**MW**	
Sizewell B	**1400**	**MW**	
Output equivalent to:	**15,971**	**Sizewells**	
annual UK electricity energy consumption	**324**	**TW-hr, or**	1.1664E+18 J
Dome heat =	**41**	**% annual UK energy**	

Helping to bridge the gap….

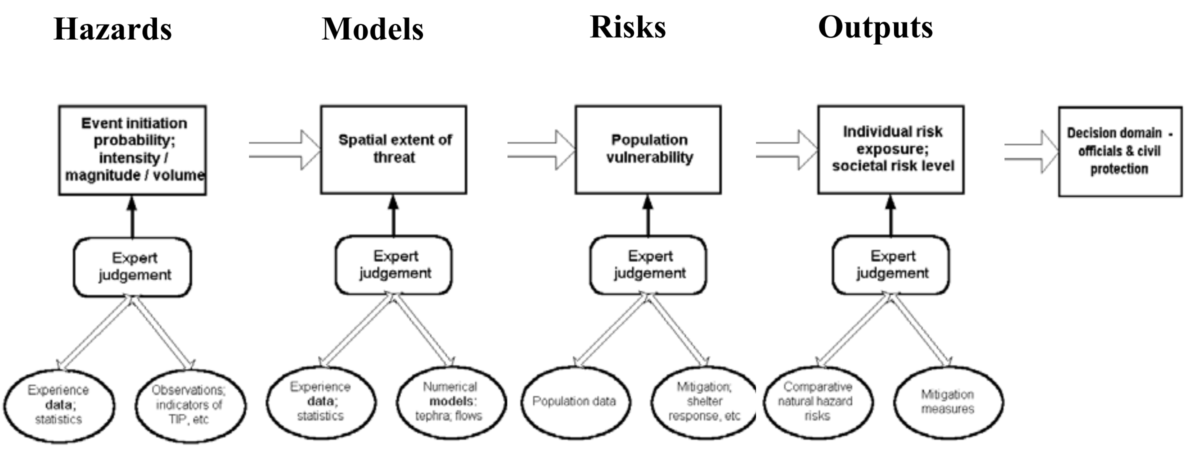
DOMAINS		
HAZARD	RISK	DECISION
events probabilities	vulnerability 'utility'	socio-economic consequences
scientific input ------------→		
		--------→political output

Risk management gap

…………..we have used the EXCALIBR structured expert opinion elicitation procedure with some success in the Montserrat crisis….

………innovations in expert judgement elicitation methodology offer assistance for finding a pathway to rational decision-making in many areas……..

The increasing complexity of volcanic hazard modelling and risk estimation



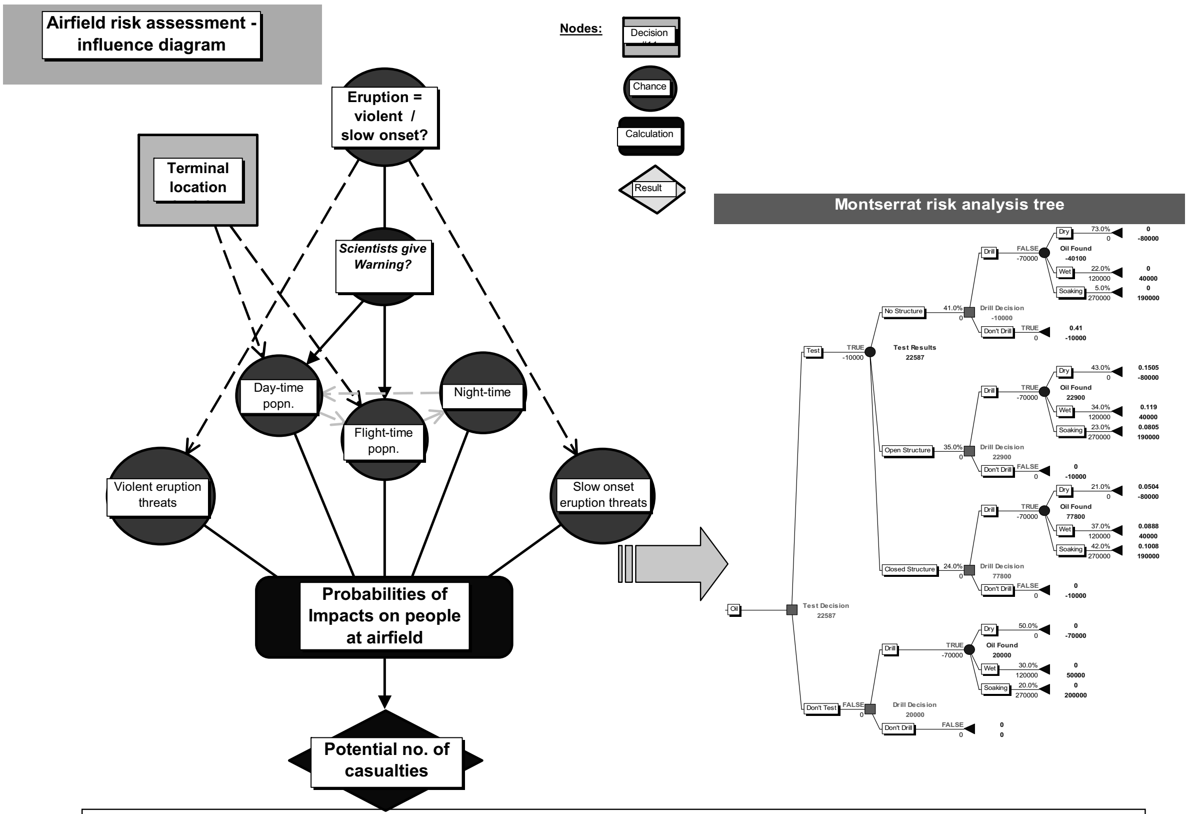
Applications

- **Space**
  (propulsion system reliability)
- **Space**
  (space debris impact)
- **Space**
  (strength of composites)
- **Industrial**
  (flange connection failures)
- **Industrial**
  (fuelling crane failure)
- **Hydrology**
  (predicting groundwater contamination; reservoir safety)
- **Meteorology**
  (flood forecasting)
- **Seismology**
  (earthquake engineering for nuclear power stations)
- **Volcanology**
  (hazard mitigation)

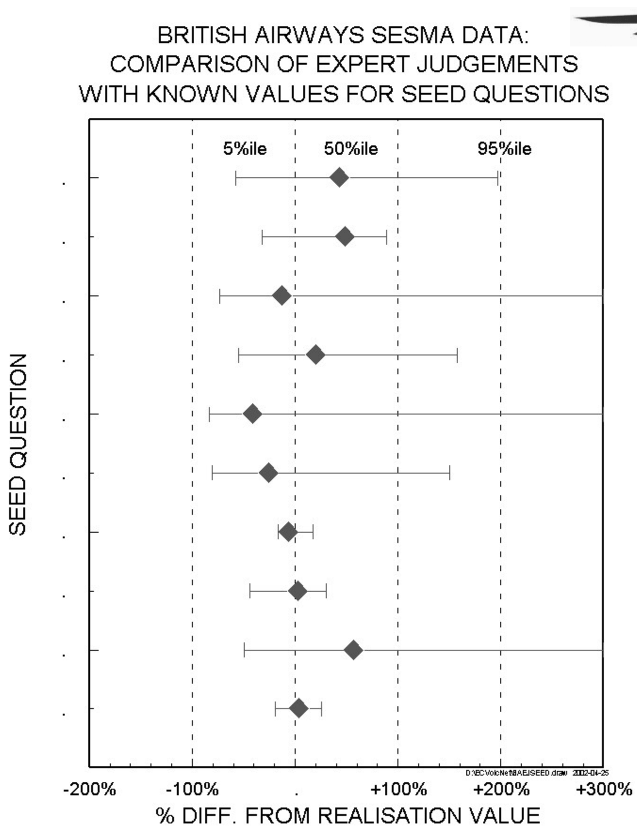
The eruption of Mount Pinatubo, 1990….

and effects at Clark AFB

….from volcanoes to civil aviation……



Influence diagram for estimating volcanic eruption risks to airfield operations, depending upon siting decision

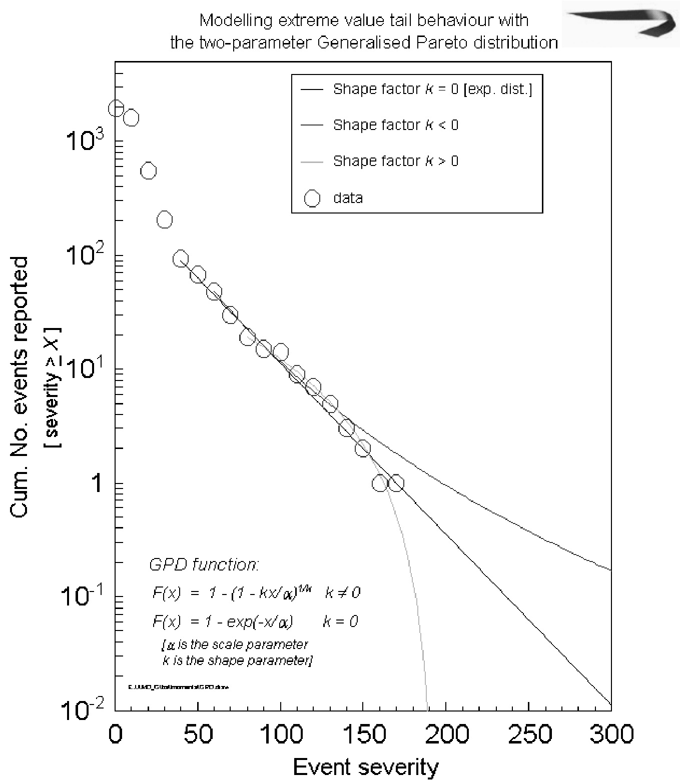


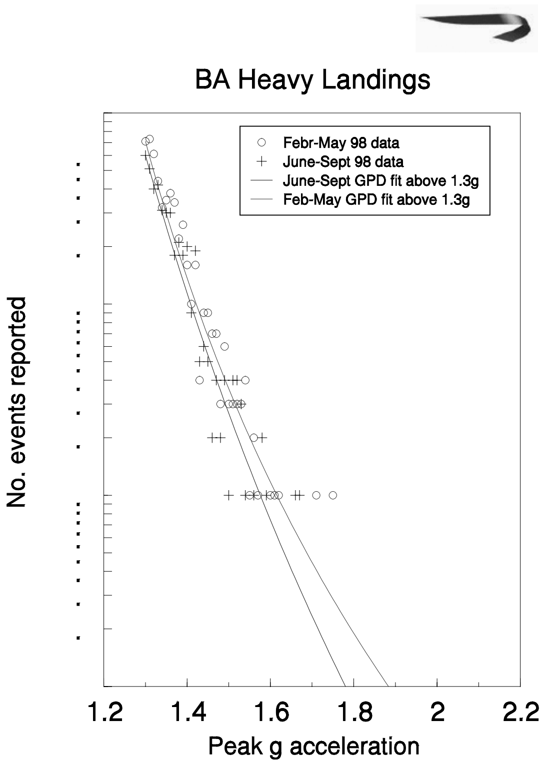
The tendency for the pooled results to converge towards the 'correct' answer - from a quick and simple elicitation of expert opinions - was of considerable interest to the airline's management.

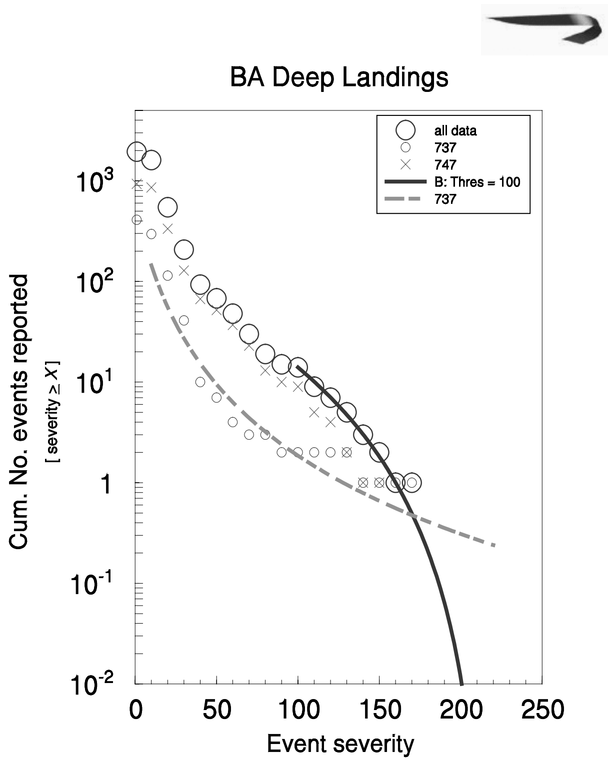
The method is being taken up by other companies, in other areas…..

British Airways:

Objective: to construct a numerical 'severity index' for ranking the relative significance of rare or extreme flight operation 'events' - by eliciting judgements from the most experienced senior captains.

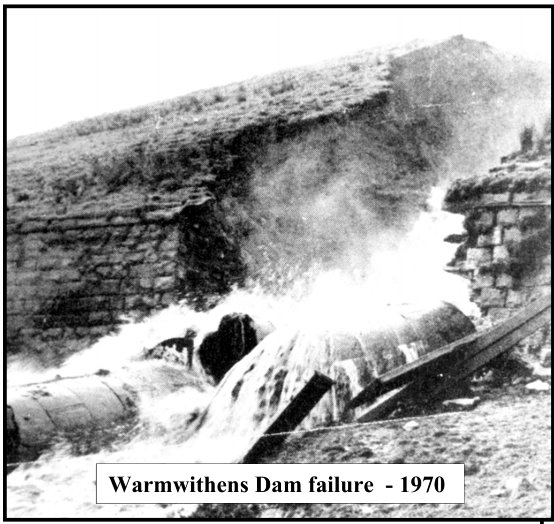






Moral: don't let sales people dictate operational procedures without checking the consequences….

From air to water………

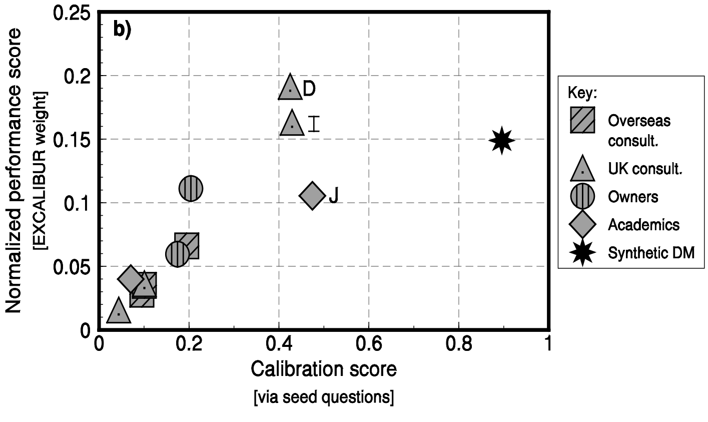

Warmwithens Dam failure - 1970

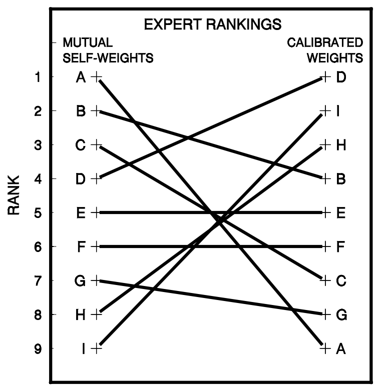

Cowlyd Reservoir inspection party - 1917

..risk assessment and reservoir safety in the UK

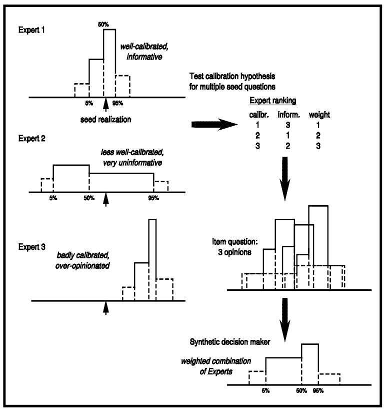
Objective: to developing a generic quantitative model for accelerated internal erosion in Britain's population of 2,500 ageing dams, using elicited quantities for key variables

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

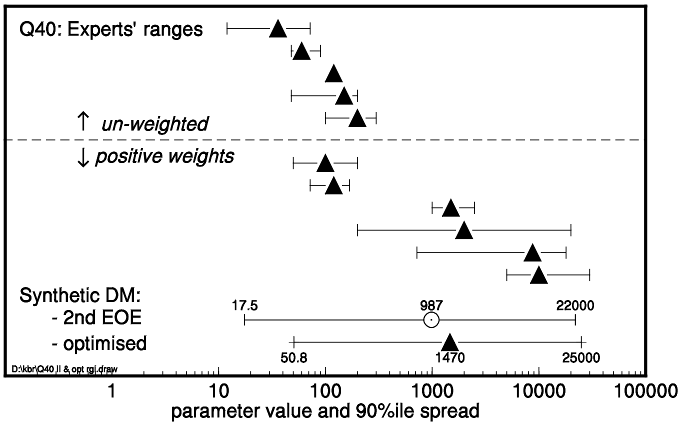




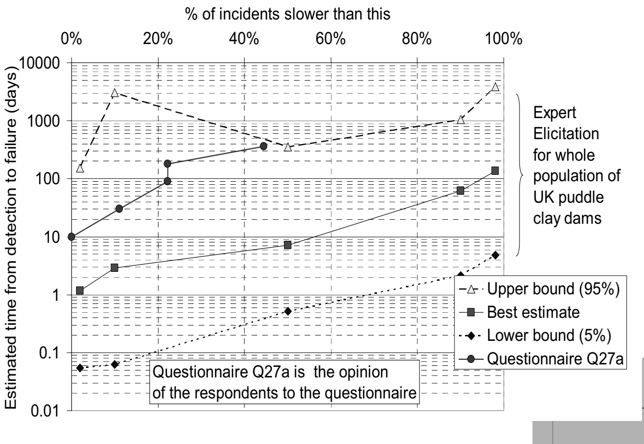
Example of the experts' spreads of opinion for one parameter of interest, and the outcomes obtained by alternative ways of pooling the weighted opinions



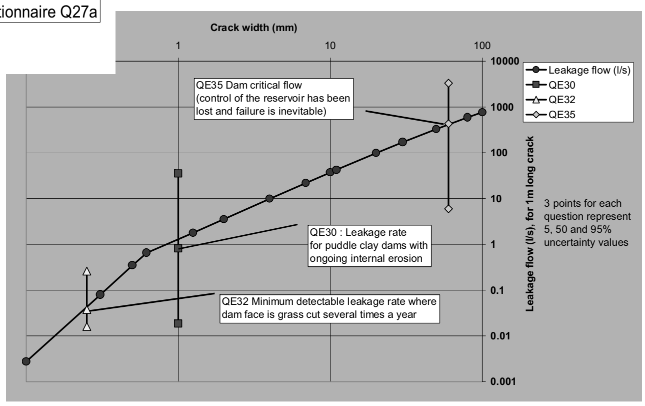
Experts' opinions on the time-to-failure (in days from first detection) of the 10%ile slowest cases



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



Reservoir erosion model: relating elicited uncertainty spreads to experience data, and to theoretical leakage models………



Input of expert judgement is essential for the parameterisation of models of complex uncertain processes……..

….back to the atmosphere, and climate change……..

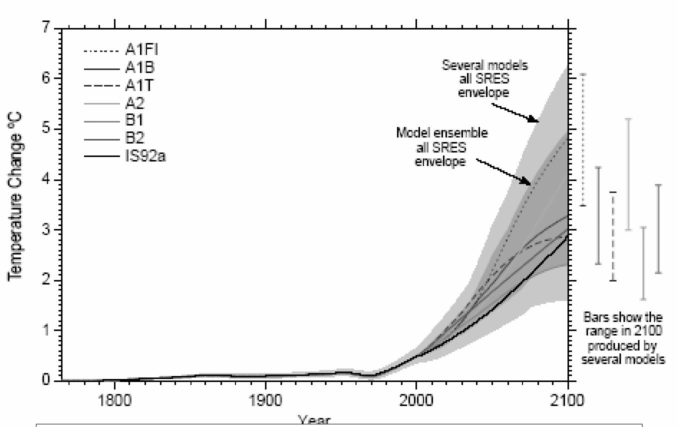
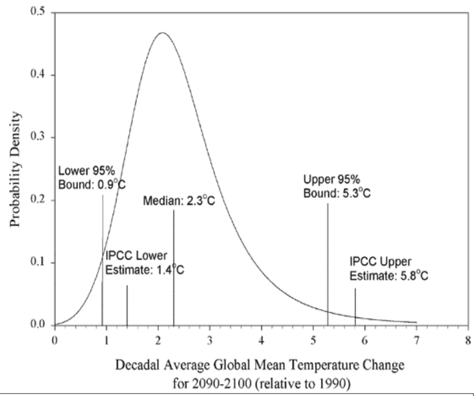


Fig 3.2 from: Climate adaptation: Risk, uncertainty and decision-making. UK Climate Impacts Programme - Technical Report May 2003



from Webster et al.,

MIT Joint Program on the Science and Policy of Global Change - Report #73. Uncertainty Analysis of Global Climate Change Projections, 2001

…climate change modelling is one candidate for more formal techniques of structured elicitation of expert judgement

Structured opinion elicitation for decision-making - positive remarks

Cooke's EXCALIBR procedure relies on cornerstones of the scientific method:

Empirical control - ***evaluates weights for experts on basis of actual or possible observations***

Accountability - ***inputs are traceable in terms of scientific inputs of individuals***

Reproducibility - ***can replicate and review all calculations used***

Advantages:

Impartiality - ***experts are treated equally prior to calibration***

Equity – ***individual experts' scores are maximised by stating true scientific views***

Diagnostic - ***procedure can highlight discrepancies in reasoning or inconsistencies in interpretation***

……the approach produces a "rational consensus", and sits squarely within the Bayesian paradigm for decision-support

Pros & cons of structured opinion elicitation for decision-making - for volcanic crisis management

Advantages:

Inclusive: *can involve whole team in decision-making process*

Un-biased: *individual polling procedure encourages optimal expression of true opinion*

Exhaustive: *all sources of uncertainty are treated fully and explicitly*

Neutral: *de-personalises provision of scientific advice*

Transparent: *approach accords with new British government guidelines for scientific advice, and requirements to pool wide range of expertise*

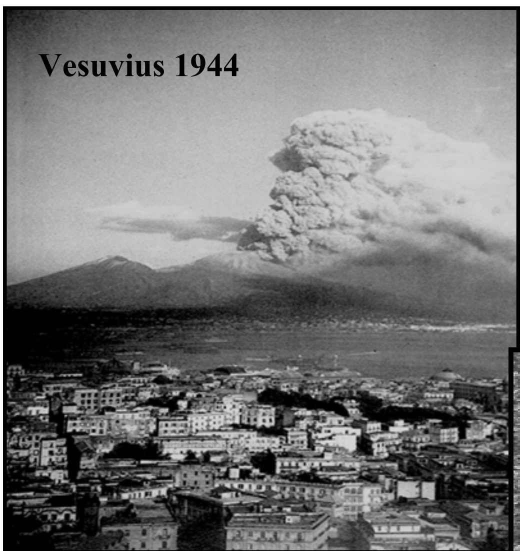
Disadvantages:

Concept and principles of subjective probability are not familiar to many scientists

Individual "calibrations" are more difficult to justify in context of volcanological hazard assessment than in some other disciplines

Requires specialist "facilitator" to ensure correct implementation

In a volcanological context, what next??

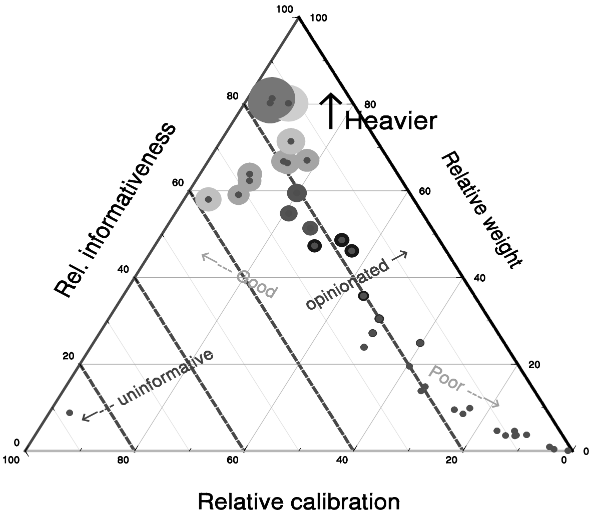


… the next eruption of Vesuvius.. ……is likely to cause massive problems on the ground, and in the air……and difficult decisions will have to be made!

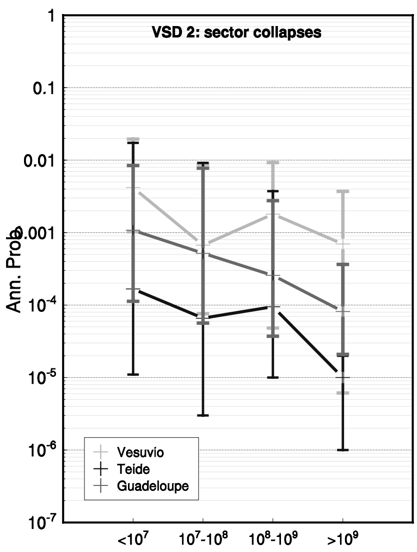
The European Community is sponsoring a major multi-disciplinary study into assessing the risk from the next eruption of Vesuvius,

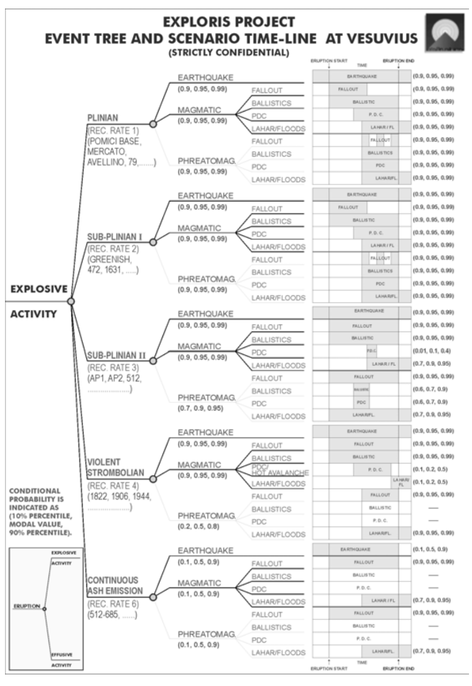
The EXCALIBR approach is being used…..

Relationship of experts relative weights to informativeness and calibration

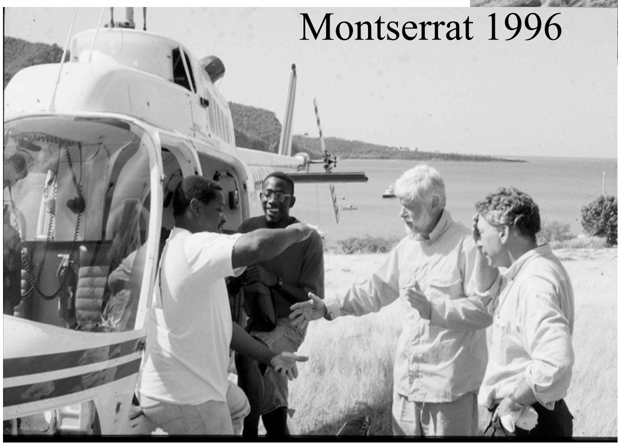


In EXPLORIS, we are deriving eruption size – frequency relationships (similar to Gutenberg-Richter) and scenario Event Tree representations for hazard and risk modelling purposes:





For over 100 years, volcanologists have been facing the challenge of making life or death decision-making in the presence of uncertainty…..



….the issues are very similar for other safety-critical industries, dams, NPP installations, etc…..

One perspective:

We cannot stop volcanoes from erupting…….

…but we should be able to prevent NPP's from exploding !!!

Using expert judgement for PSHA…..

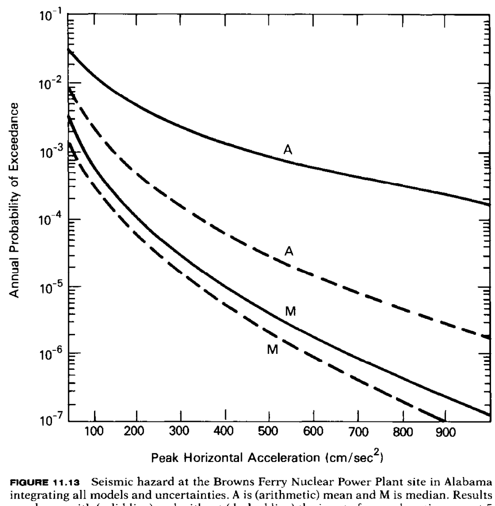
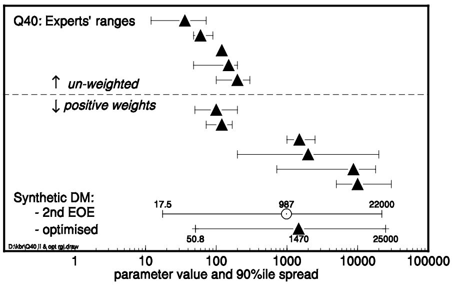
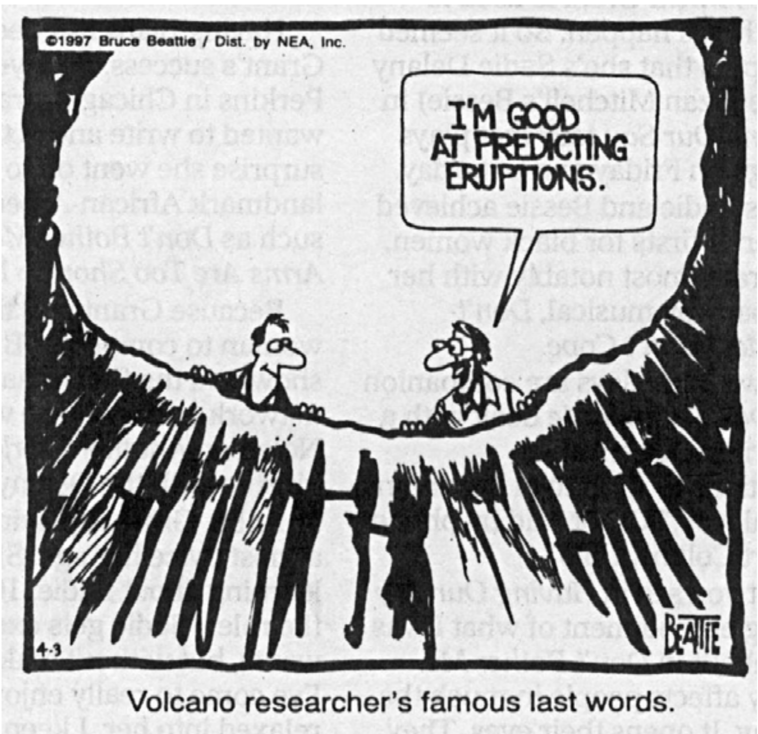


FIGURE 11.13 Seismic hazard at the Browns Ferry Nuclear Power Plant site in Alabama integrating all models and uncertainties. A is (arithmetic) mean and M is median. Results are shown with (solid line) and without (dashed line) the input of ground motion expert 5 (after Bernreuter and others 1989).

Many problems in quantifying and parameterising PSHA models - such as how best to accommodate the influence of 'Expert 5' - can be formally and addressed by adopting the structured elicitation procedure we have just described…………



Where expert judgements must be used, this procedure provides an auditable trail, - item by item - producing a "rational consensus" in a suitable and accessible form for peer review, regulatory inspection, and public confidence.



Volcano researcher's famous last words.

Thank you!

References

Bonano EJ et al., 1990 Elicitation and use of expert judgement in performance assessment for high-level waste repositories. ***NUREG/CR-5411,*** **Washington DC.**

Cooke RM 1991 ***Experts in Uncertainty*****. Oxford University Press.**

Coppersmith KJ, Youngs RR 1990 Probabilistic seismic hazard anallysis using expert opinion: an example from the Pacific North-West. ***Geol. Soc. Amer., Reviews of Engng Geol. VIII, 29-45.***

Reiter R 1990 ***Earthquake Hazard Analysis*****. Columbia University Press, NY.**

TERA 1980 Seismic hazard analysis: solicitation of expert opinion. ***NUREG/CR-1582, 3,*** **Washington DC.**

Woo G 1992 Calibrated expert judgement in seismic hazard analysis. In: ***Proc. 10th World Conf. Earthq. Engng*****, Balkema, Rotterdam, 333-338.**

Woo G 1996 Kernel estimation methods for seismic hazard area source modelling.
Bulletin of the Seismological Society of America 86(2): 353-362.