Uncertainty Analysis Comes to Integrated Assessment Models for Climate Change…and Conversely

Roger M. Cooke
Resources for the Future and
Dept. Mathematics, Delft University of Technology
Cooke@rff.org
Feb. 24, 2012, May 15, 2012

Abstract: This article traces the development of uncertainty analysis through three generations punctuated by large methodology investments in the nuclear sector. Driven by a very high perceived legitimation burden, these investments aimed at strengthening the scientific basis of uncertainty quantification. The first generation building off the Reactor Safety Study introduced structured expert judgment in uncertainty propagation and distinguished variability and uncertainty. The second generation emerged in modeling the physical processes inside the reactor containment building after breach of the reactor vessel. Operational definitions and expert judgment for uncertainty quantification were elaborated. The third generation developed in modeling the consequences of release of radioactivity and transport through the biosphere. Expert performance assessment, dependence elicitation and probabilistic inversion are among the hallmarks. Third generation methods may be profitably employed in current Integrated Assessment Models (IAMs) of climate change. Possible applications of dependence modeling and probabilistic inversion are sketched. It is unlikely that these methods will be fully adequate for quantitative uncertainty analyses of the impacts of climate change, and a penultimate section looks ahead to fourth generation methods.

1. Introduction

Downside risk drives the policy concern with uncertainty, and without uncertainty there is no risk, only adversity. Uncertainty analysis and risk analysis were born under the same star at the same time: The Reactor Safety Study (Rasmussen Report, WASH−1400, 1975) was the first comprehensive risk analysis, and also the first major uncertainty analysis….but it was triplets. The Rasmussen report also marks the first attempt to use quantitative expert judgment as scientific input into a very large engineering study. This confluence is no coincidence. Risk analysis was invented to quantify the probability and consequences of catastrophic events that

had not yet happened. Complex systems are functions of their components, and some data at the component level are always available, but not enough. Gaps must be filled with expert judgment. In addition, there are always questions about the models of the systems. Hence, risk analysis, uncertainty analysis and structured expert judgment are joined at the hip. The nuclear experience has been impelled by a very high perceived legitimation burden, driving uncertainty quantification to become science−based to the extent possible. Climate change poses perhaps the greatest risk humanity has created for itself, and the complexity of the system in question − the whole planetary climate − dwarfs that of the systems with which uncertainty analysis, risk analysis and structured expert judgment have developed.

This article traces the development of uncertainty analysis since its inception through three generations, punctuated by large investments in methodology in the nuclear sector. The goal in tracing these developments is to gather momentum for the challenges of the fourth generation that lies ahead. Society does not have time to re−learn painful lessons of the past 50 years. Those lessons may be encapsulated as:
1. Quantify uncertainty as expert subjective probability on uncertain quantities with operational meaning
2. Apply probabilistic inversion to obtain distributions on unobservable variables
3. Assess dependence between uncertain quantities
4. Assess expert performance in panels of independent experts

It is unlikely that these methods will be fully adequate for quantitative uncertainty analyses of the impacts of climate change, and a penultimate section looks ahead to fourth generation methods.

1. First Generation

Throughout the 1950's, the US Atomic Energy Commission pursued a philosophy of risk management based on the "maximum credible accident". Because "credible accidents" were covered by plant design, residual risk was estimated by studying the hypothetical consequences of "incredible accidents." An early study (AEC 1957) focused on three scenarios of radioactive releases from a 200 megawatt nuclear power plant operating 30 miles from a large population center. Regarding the probability of such releases, the study concluded that "no one knows now or will ever know the exact magnitude of this low probability." Successive design improvements were intended to reduce the probability of a catastrophic release of radioactivity. Such improvements could have no visible impact on the risk as studied with the then current methods. On the other hand, plans were being drawn for reactors in the 1000 megawatt range located near population centers, and these developments would certainly have a negative impact on the consequences of the "incredible accident".

The desire to quantify and evaluate the effects of these improvements led to the introduction of *probabilistic* risk analysis (PRA). Whereas the earlier studies had dealt with uncertainty by making conservative assumptions, the goal now was to provide a *realistic*, as opposed to conservative, assessment of risk. A realistic risk assessment necessarily involved an assessment of the uncertainty in the risk calculation. The basic methods of PRA developed in the aerospace program in the 1960s found their first full-scale application, including accident consequence analysis and uncertainty analysis, in the Reactor Safety Study of 1975. Costed at 70 man-years

and four million 1975 dollars, the study was charged to "consider the uncertainty in present knowledge and the consequent range in the predictions". The study's appendix on failure data contains a matrix of assessments of failure rates by 30 experts for 61 components (many cells are empty). The study team represented the failure rates as log normal distributions meant to cover both the variability in failure rates for components in different systems and also the uncertainty in the data. Its method for doing this was not transparent. Monte Carlo simulation propagated these (independent) distributions through the risk model. This was the first study of this magnitude to (a) treat expert probability assessments as scientific data, (b) quantify a large engineering model with subjective probability distributions and (c) propagate these distributions through the model with Monte Carlo simulation.

The Reactor Safety Study caused considerable commotion in the scientific community, so much so that the US Congress created an independent panel of experts, the Lewis Committee (Lewis 1979), to review its "achievements and limitations". Citing a curious method for dealing with model uncertainty (the so-called square root bounding model[1]), whose arbitrariness "boggles the mind," the panel concluded that the uncertainties had been "greatly understated", leading to the study's withdrawal[2]. However, the use of structured expert judgment and uncertainty quantification were endorsed and became part of the canon. Shortly after the 1979 accident at Three Mile Island, a new generation of PRAs appeared in which some of the methodological defects of the Reactor Safety Study were avoided. The U.S. NRC released the PRA Procedures Guide in 1983, which shored up and standardized much of the risk assessment methodology. An extensive chapter devoted to uncertainty and sensitivity analysis laid out the basic distinctions between sensitivity analysis, interval analysis and uncertainty analysis, and addressed interpretational and mathematical issues. In as much as a boggled mind is an uncertain mind, the contretemps over the square root bounding model may be seen as the first foray into model uncertainty. The Reactor Safety Study served as model for many probabilistic risk analyses with uncertainty quantification in the nuclear, aerospace and chemical process sectors.

The main features of first generation uncertainty analyses are
- *Structured expert judgment with point estimates*
- *Distinction of physical variability and uncertainty*
- *Propagation of uncertainty with Monte Carlo simulation*

2. Second Generation

In 1987 the US Nuclear Regulatory Authority issued a draft Reactor Risk Reference Document (USNRC 1987 NUREG−1150) focused on processes inside the containment, in the event of breach of the reactor vessel. Extensive use was made of expert judgment, not unlike the methods in the Reactor Safety Study. Peer review blew the draft document out of the water,

---

[1] The probability of one control rod failing to insert was assessed as 10^{-3} per demand. If three adjacent rods fail to insert, overheating may occur. If the failures are completely independent the probability of three failures is 10^{-9}, if they are completely dependent it is 10^{-3}. The square root bounding model uses the root of the product, i.e. 10^{-6}.

[2] "…in light of the Review Group conclusions on accident probabilities, the Commission does not regard as reliable the Reactor Safety Study's numerical estimate of the overall risk of reactor accident"(US NRC press release 1979).

triggering a serious investment in expert judgment methodology. In 1991, a suite of studies, known as NUREG-1150 (USNRC 1991), appeared that set new standards in a number or respects. Most importantly, experts were identified by a traceable process and underwent extensive training in subjective probability assessment. Instead of giving point estimates, experts quantified their uncertainty on the variables of interest (Hora, S. and Iman, R 1989). Expert probability distributions were combined with equal weighting. In addition, experts were extensively involved in elaborating alternative modeling options and providing arguments pro and con. Latin hypercube sampling was implemented as a quasi−random number sampling technique with greater efficiency than simple Monte Carlo sampling. Perhaps the most significant innovation in this period was the decision to elicit expert subjective probabilities only on the results of possible observations, and not on abstract model parameters. Indeed, this study had to contend with the poorly understood physical processes inside the reactor containment after breach of the reactor vessel. There was great uncertainty regarding the physical models to describe these phenomena, and opinions were sharply divided. It would be impossible to ask an expert to quantify his/her uncertainty on a parameter in a model to which (s)he did not subscribe.

The NUREG−1150 methods found traction outside the nuclear engineering community. The National Research Council has been a persistent voice in urging the US government to deal with uncertainty. "Improving Risk Communication" (NRC 1989) inveighed against minimizing uncertainty. The landmark study "Science and Judgment" (NRC 1994) gathered many of these themes in a plea for quantitative uncertainty analysis as "*the only way to combat the 'false sense of certainty' which is caused by a refusal to acknowledge and (attempt to) quantify the uncertainty in risk predictions."* Following the NUREG−1150 lead, in 1996 the National Council on Radiation Protection and Measurement issued guidelines on uncertainty analysis with expert judgment. The US EPA picked this up shortly thereafter in Guidelines for Monte Carlo Analysis (1997) and cautiously endorsed the use of structured expert judgment. In 2005 the US EPA's Guidelines for Carcinogen Risk Assessment (EPA, 2005) (p. 3−32) advised that *"...the rigorous use of expert elicitation for the analyses of risks is considered to be quality science.*" EPA's first expert judgment study concerned mortality from fine particulates and was completed in 2006 (Industrial Economics 2006).

The work of Granger Morgan's Carnegie Mellon group fits best in this category. In a widely cited passage, Henrion and Morgan (1990, p.50) endorse Ron Howard's clairvoyant test: "*Imagine a clairvoyant who could know all facts about the universe, past, present and future. Could she say unambiguously whether the event will occur or had occurred, or could she give the exact numerical value of the quantity? If so, it is well-specified.*" If a model contains terms that do not pass the clairvoyant test, then the model is ill−specified and should go back to the shop.

In the heat of the fray the clairvoyant test proves a bit too severe. A simple example illustrates the point, and tees up a later discussion of probabilistic inversion. A power law describes the lateral dispersion $\sigma(x)$ of a contaminant plume at downwind distance x as

1) $\sigma(x) = Ax^B$.

In a constant wind field, the dispersion coefficients *A* and *B* have a clear physical meaning as constants in the solution of the equations of motion for fluids. In reality, the wind field is not constant, there is vertical wind profile, Coriolis force, plume meander, surface roughness, etc. There are no such constants, no omniscient entity could determine their values and *(A,B)* fail the clairvoyant test. In practice such power laws are used with coefficients that depend inter alia on surface roughness, release height and atmospheric stability, which may be characterized in a variety of ways. The important thing is not how an omniscient being would determine these values (which is impossible), but how WE would determine them. We must specify a virtual measurement procedure which we would use to determine their values. If the virtual measurement procedure is not one with which experts are familiar and comfortable, then expert elicitation may will not be possible on *(A,B)* directly, and inversion procedures described in section 4 will be required. The term "operational definition" was proposed by the physicist P.W. Bridgeman (1927); let us relax the clairvoyant test to the "Bridgeman test[3]": Every term in a model must have operational meaning, that is, the modeler should say how, with sufficient means and license, the term would be measured.

Morgan and Henrion (1990) also advise against combining expert judgments in the event of disagreement. Rather, the analyst should seek to understand the reasons for the disagreements and if these cannot be resolved, should present all viewpoints. This position is echoed in a report of the U.S. Climate Change Science Program (Morgan et al 2009). In second and third generation studies, there is ample documentation of experts' rationales, from which one quickly realizes that expert disagreement is inherent in the partial state of scientific knowledge. The scientific method is what creates agreement among experts - if the science is not there yet, then experts should disagree. Indeed, if scientists all agreed in spite of partial knowledge, science would never advance. Expert agreement cannot be a legitimate goal of an expert judgment methodology. Preserving and reporting the individual opinions is essential, but in complex studies combining the expert's opinions is also essential. In the third generation study reported in the next section, not combining experts would have meant presenting the problem owner with 67 million different Monte Carlo exercises. Important applications to aspects of the climate change problem are published in (Morgan and Kieth 1995, Zickfeld et al 2010).

The main features of second generation uncertainty analyses are
- *Experts quantify their uncertainty as subjective probability*
- *Experts' distributions are combined with equal weighting, or not combined*
- *Extensive 'accounting trail' traces the expert judgment process, including expert recruitment, training, and interviewing, documentation of rationales, data sources, and models.*
- *Use of Latin Hypercube sampling.*

3. Third Generation

---

[3] The philosophy of science has studied 'theoretical terms' at great length, and shown that scientific theories relate to observations in ways which are much more complex than term−wise operational definitions. None the less, this simple Bridgeman criterion is adequate for a first pass.

From 1990 through 2000 a joint US-European program, hereafter called the Joint Study, quantified uncertainty in the consequence models for nuclear power plants. Expert judgment methods were further elaborated, as well as methods for screening and sensitivity analysis. European studies spun off of this work perform uncertainty analysis on European consequence models and provide extensive methodological guidance. With 2,036 elicitation variables assessed by 69 experts spread over 9 panels and a budget of $7 million 2010 dollars (including expert remuneration of $15,000 per expert), this suite of studies established a benchmark for expert elicitation, performance assessment with calibration variables, expert combination, dependence elicitation and dependent uncertainty propagation (see SOM for details and references). The expert judgments within each panel were combined with both equal weighting and performance based weighting. Since the panel outputs feed into each other sequentially, to arrive at a distribution for the study's endpoints without combining the experts, there would be 67 million possible sequences.

2nd generation studies trained experts in being 'good probability assessors' but no validation of actual performance was undertaken. In the joint study experts assessed their uncertainty on seed or calibration variables from their field in addition to the variables of interest. Considerable effort went into finding suitable calibration variables (for examples, see SOM). Performance measurement in terms of statistical likelihood and informativeness has the multiple benefit of (a) raising awareness that subjective probabilities are amenable to objective empirical control, (b) enhancing credibility in the combined assessments and (c) enabling performance based combinations of expert distributions as an alternative to equal weighting.

In earlier studies the distributions of uncertain variables are assumed to be independent. In complex models this is untenable. Consider a few examples from consequence models:

- the uncertainties in effectiveness of supportive treatment for high radiation exposure in people over 40 and people under 40;
- the amount of radioactivity after one month in the muscle of beef and dairy cattle; and
- the transport of radionuclides through different soil types.

The format for eliciting dependence was to ask about joint exceedence probabilities: "Supposing the effectiveness of supportive treatment in people over 40 was observed to be above the median value, what is the probability that the effectiveness of supportive treatment in people under 40 would also be above its median value?" Experts quickly bought into this format. Dependent bivariate distributions were found by taking the minimally informative copula that reproduced these exceedence probabilities.

One of the most challenging innovations was the use of probabilistic inversion. Simply put, probabilistic inversion denotes the operation of inverting a function at a (set of) (joint) distribution(s). Whereas standard parameter estimation fits models to data, probabilistic inversion may be conceived as a technique for fitting models to expert judgment.   A simple illustration is given in section 4.2.

The important features of 3rd generation uncertainty analyses are

- *Measurement of expert performance in terms of statistical likelihood and informativeness*
- *Enabling performance based combinations of expert distributions*
- *Dependence elicitation and propagation*
- *Probabilistic inversion for obtaining distributions on unobservable model parameters*

Compared to the resources available for nuclear safety, the uncertainty analyses of integrated assessment models are a cottage industry. All the greater is our debt to those who have pursued this work. The social priorities expressed by this fact merit sober reflection. The following section sketches how 3ʳᵈ generation techniques might be applied in climate change uncertainty analyses.

4. Application of third generation techniques

In striving for science based uncertainty quantification, the nuclear experience yields four lessons that may be directly applied to uncertainty quantification with regard to climate change.

4.1 Operational meaning

Independent experts can assess uncertainty with respect to possible physical measurements or observations. In doing so, they may apply whatever models they like, and must not be constrained to adopt the presuppositions of any given model. Operational meaning is most pertinent to Integrated Assessment Models (IAMs) with regard to discounting. The Supplementary Online Material (SOM) contains a derivation of the Social Discount Factor (SDF) and Social Discount Rate (SDR) as

2) $\text{SDF} = e^{-t(\rho + \eta G(t))}; \ \text{SDR} = \rho + \eta G(t)$

where ρ is the rate of pure time preference, η is the coefficient of constant relative risk aversion and $G(t)$ is the time average growth rate of per capita consumption out to time t.

It is generally recognized that the discount rate is an important driver, if not *the* most important driver, in IAMs. Some (Stern 2008) see a strong normative component. Others infer values for ρ and η from data. (Evans and Sezer 2005). Nordhaus (2008) equates the SDR to the observed real rate of return on capital with a constant value for $G(t)$, and sees ρ and η as "unobserved normative parameters" (p. 60) or "taste variables" (p. 215) which are excluded from uncertainty quantification. Pizer (1999) assigns distributions to ρ and η. Nordhaus and Popp (1996) put a distribution on ρ. Weitzman (2001) fits a gamma distribution to SDR based on an expert survey. Frederick, Lowenstein and O'Donoghue (2002, p 352) note that "virtually every assumption underlying the DU [discounted utility] model has been tested and found to be descriptively invalid in at least some situations". They also cite the founder of discounted utility, Paul Samuelson: "It is completely arbitrary to assume that the individual behaves so as to maximize an integral of the form envisaged in [the DU model]"(p. 355) Weitzman (2001) and Pizer and Newell (2003) show that uncertainty in the discount rate drives long term rates down.

We may distinguish variables according to whether their values represent

   a) Policy choices
   b) Social preferences
   c) Unknown states of the physical world.
Uncertainty quantification is appropriate for (b) and (c) but not for (a). We see that various
authors assign time preference and risk aversion to both (a) and (b). Given the paramount
importance of discounting for the results of IAMs, resolving disagreement about whether this
can be uncertain deserves high priority. A proposal based on probabilistic inversion is advanced
below.

4.2 Probabilistic inversion

In 3rd generation nuclear studies, experts were often unwilling or unable to quantify uncertainty
on model parameters without clear operational meaning with which they were familiar.
Examples included the dispersion coefficients mentioned earlier, and also transfer coefficients
in environmental transport models. The simple power law for dispersion coefficients is most
intuitive. Asking experts for their joint distribution over *(A, B)* would require the experts to
propagate that distribution through the power law Ax^B to optimally approximate his/her
uncertainty in the lateral or vertical dispersion at downwind distances *x* from 500 meters to 10
km. That would be an excessive cognitive burden. The solution is to ask experts to quantify
their uncertainty on the spreads $\sigma(x_i)$ at several downwind distances $x_1,...x_n$, using whatever
model they like, from which the analyst finds a distribution over *(A, B)* that best reproduces the
experts' uncertainty. This operation is termed probabilistic inversion and has become a
mainstay in uncertainty analysis. The joint study applied it extensively to obtain distributions
over coefficients in environmental transport models. Note that the $\sigma(x_i)$ are directly observable,
whereas the *(A, B)* are not.

This technique might be applied to the Social Discount Rate $\rho+\eta G(t)$ (eq.2). In Weitzman's
expert survey (2001) experts are asked for their "professionally considered gut feeling" over the
"real interest rate …to discount over time the (expected) benefits and (expected) costs of
projects to mitigate the effects of climate change" (p. 266). The results are summarized as a
gamma distribution with mean 3.96% and standard deviation 2.94%. The SDR depends on time
through the average growth rate. If the dependence on growth were captured in the elicitation,
then the example would closely resemble the dispersion coefficient case discussed in section 2.
Suppose for the sake of illustration that experts were asked for their "professional gut feeling"
for the real interest rate given that the time average growth rate of consumption was fixed at
respectively 1.5, 2.5 and 3.5. Suppose that their answers can be represented as gamma
distributions $\Gamma(2.5, 1.5)$, $\Gamma(3, 2)$ and $\Gamma(5, 2.5)$ with where $\Gamma(\mu, \sigma)$ denotes the gamma
distribution with mean μ and standard deviation σ. The inversion problem is then:

Find a distribution over (ρ, η) such that
   $\rho + 1.5\eta$ is distributed as $\Gamma(2.5, 1.5)$
   $\rho + 2.5\eta$ is distributed as $\Gamma(3,2)$
   $\rho + 3.5\eta$ is distributed as $\Gamma(5, 2.5)$

A joint distribution over (ρ, η) approximately satisfying these constraints can be found with the
iterative proportional fitting algorithm. The SOM provides explanation and background.

4.3 Dependence modeling

Dependence between uncertain parameters often makes a relatively small contribution to overall uncertainty, but sometimes the contribution is large, and for this reason it cannot be categorically ignored. Even small correlations, when they affect a large number of parameters, can have a sizeable effect.

Dependence in forcing factors of climate sensitivity is an example where the effect of dependence might be large. Following (Roe 2009), the equilibrium change in global mean temperature relative to a reference scenario is proportional to $1/(1-\sum f_i)$, where f_i are feedbacks from climate fields not in the reference scenario. There is considerable scatter in estimates of feedbacks f_i . Positive dependence between the f_i could inflate the variance of $\sum f_i$, pushing probability mass toward one and greatly enlarging the uncertainty of climate sensitivity. On the economic side, dependence may be suspected between total factor productivity, the depreciation rate of capital, the price of a backstop technology and the rate of decarbonization, as all could be influenced by technological change. If potential dependencies are identified, they can be quantified in a structured expert elicitation. For unobservable parameters such as time preference and risk aversion, probabilistic inversion offers more perspective.

The subject of dependence modeling involves three questions:
   i.   How do we obtain dependence parameters?
   ii.   How do we represent dependence in a joint distribution?
   iii.   How do we sample a high dimensional joint distribution with dependence?

These questions are closely related, since the way we sample must determine how we represent dependence in a joint distribution, and that in turn drives the sort of dependence information we acquire. This subject merits a book unto itself (Kurowicka and Cooke 2006); this article can do no more than posit answers to the above questions:
   i.   Dependence parameters are obtained by expert elicitation or by probabilistic inversion
   ii.   Dependence between pairs of variables is represented by rank correlations, together with a copula
   iii.   The pairs of correlated variables are linked in structures which can be sampled

Explaining the reasons behind these choices would take us very far afield (the SOM gives background). The importance of the choice of copula merits brief discussion. A bivariate copula is a joint distribution of the ranks or quantiles of two random variables. Different types of copula having the same rank correlations can give different results, even in simple examples. Let *X,Y* and *Z* be three exponential variables with unit expectation and pair wise rank correlation of 0.7. Different copulae may be chosen to realize these correlations. The normal copula is the rank distribution of the bivariate normal distribution. It is always tail independent: for any non-degenerate correlation, the probability of two variables exceeding a high percentile approaches the independent probability as the percentile becomes greater. It is the only copula implemented in commercial uncertainty analysis packages, and its uncritical use has been blamed for reckless risk taking on Wall Street (Salmon 2009). The "mininf" copula (Bedford and Cooke 2002) is the copula realizing the stipulated correlations which is minimally informative with respect to

the independent copula. The Gumbel copula has upper tail dependence. Consider the product $X \times Y \times Z$; its mean relative to the independent distribution increases by factors 3.9, 2.9 and 4.8 for the normal, min inf and Gumbel copulae respectively. Bearing in mind that this is a very simple example, the "copula effect" is significant, though of course the greatest effect would be ignoring dependence altogether. The SOM provides background and graphs.

4.4 Performance assessment

Weitzman (2001) and Nordhaus (1994) conducted expert judgment exercises on the discount rate, and the damages from climate change, respectively. These are not fully 3rd generation studies, but they are very valuable. For the most part, however, uncertainty analysis has been performed by the modelers themselves putting distributions on the parameters of their models. That this is not a good idea became starkly evident when the Joint Study results were compared with previous "in house" uncertainty quantifications done by the modelers themselves. The in−house confidence bands were frequently narrower by an order of magnitude or more (see SOM).

The Joint Study applied a scheme of performance measurement for probabilistic assessors previously developed in Europe (Cooke 1991), both to experts and to combinations of experts. Experts quantify their uncertainty by giving median values and 90% central confidence bands for uncertain quantities, including quantities from their field whose true values are known post−hoc. Experts and combinations of experts are treated as statistical hypotheses. Performance is measured in terms of statistical accuracy and informativeness. Statistical accuracy is the p−value at which the experts' probabilistic assessments, considered as a statistical hypothesis, would be falsely rejected. These measures are combined to form weights so as to satisfy a scoring rule constraint: an expert maximizes his/her expected long run weight by stating percentiles corresponding to his/her true beliefs. The only way to game the system is by being honest. The nuclear experience with structured expert judgment is representative of the larger pool of applications (Cooke and Goossens 2008, Aspinall 2010). Some experts give stunningly accurate and informative assessments. Although many 'expert−hypotheses' would be rejected at the 5% significance level, the combinations generally show good statistical performance (see SOM for details).

5. **The Fourth Generation: model uncertainty**

Opponents of uncertainty quantification for climate change claim that this uncertainty is "deep" or "wicked" or "Knightian" or just plain unknowable. We don't know which distribution, we don't know which model, and we don't know what we don't know. Yet, science−based uncertainty quantification has always involved experts' degree of belief, quantified as subjective probabilities. There is nothing to not know. The question is whether the scientific community has a role to play in the quantification of climate uncertainties. If we cannot endorse a science based quantification of uncertainty for climate change, should we be building quantitative models at all?

That said, model uncertainty seems much greater and more pervasive for climate change than in the first three generations. Nonetheless, existing techniques for capturing model uncertainty

appear applicable. This is not the place to elaborate and evaluate these options. Instead, this section illustrates two techniques, stress testing and model proliferation; the SOM contains further suggestions in this direction.

5.1 Stress testing models: Bernoulli dynamics

Stress testing means feeding models extreme parameter values to check that their behavior remains credible. In physics stress tests in the form of thought experiments play an important role in model criticism. When Einstein imagined accelerating a classical particle to the speed of light, using the Lorentz transformation, he found that time would stop in the particle's reference frame. Failing this stress test motivated the development of special relativity. We apply an illustrative stress test to the economic growth dynamics underlying many IAMs.

Economic growth dynamics is often based on a differential equation solved by Jacob Bernoulli in 1695 (see SOM). Neglecting climate damages and abatement costs, a Cobb-Douglas function combines total factor productivity $A(t)$, capital stock $K(t)$ and labor $N(t)$ at time t:

3) $Output(t) = A(t)K(t)^{\gamma}N(t)^{1-\gamma}$; $0 < \gamma < 1$.

Capital in the next time period is capital in the previous time period depreciated at constant rate δ, plus investment, given as a fraction ϕ of output:

4) $K(t+1) = (1-\delta)K(t) + \phi(t)Output(t)$

Take A, ϕ and N as constant. Substituting (4) into (5) yields a differential equation whose solution is

5) $K(t) = [(1-\gamma)\,\phi AN^{1-\gamma} \int_{x=o..t} e^{-(1-\gamma)\delta x}\, dx \ + e^{-(1-\gamma)\delta t}\, K(0)^{(1-\gamma)}]^{1/(1-\gamma)}$.

This growth dynamics makes strong assumptions; the rate of change of capital depends only on the current values of the variables in (3). There is no other "stock variable" whose accumulation or depletion affects capital growth. Capital cannot decrease faster than rate $(1-\delta)$, and it forgets its starting point at an exponential rate. To visualize, Figure 1 sets A and N at their initial values in the IAM DICE2009, takes values of δ and γ from DICE and plots two capital trajectories. The solid trajectory starts with an initial capital of 1\$, that is, \$1.5×10^{-10} for each of the earth's 6.437×10^6 people. The dotted trajectory starts with an initial capital equal to ten times the DICE2009 initial value. The limiting capital value is independent of the starting values – with a vengeance: the two trajectories are effectively identical after 60 years.
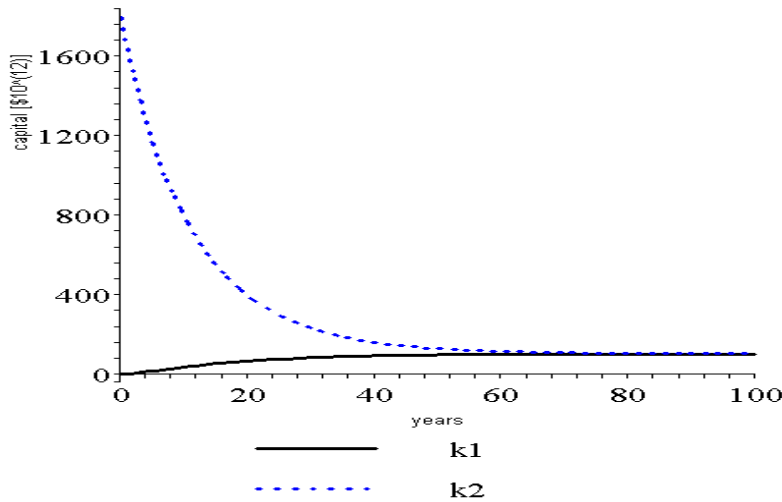
Figure 1: Two capital trajectories with DICE with default values, no temperature rise, no abatement   K1(0) = 1\$ and K2(0) = 1800 trillion \$

With one dollar, the earth could not afford one shovel, and the consequences of such a shock would surely be very different than those suggested in Figure 2. It is often noted that simple models like the above cannot explain large differences across time and geography between different economies, pointing to the fact that economic output depends on many factors not present in such simple models (Barro and Sala−i−Martin 1999, chapter 12). Negative stress test results raise the question, heretofore unasked, for which initial values is the model plausible? They also enjoin us to consider other plausible models.

5.2 Model proliferation: Lotka Volterra dynamics

Are there other plausible dynamics? One variation is based on the following simple idea: Gross World Production (GWP) produces pollution in the form of greenhouse gases. Pollution, if unchecked, will eventually destroy necessary conditions for production. This simple observation suggests that Lotka-Volterra, or predator-prey type models. Focusing on GWP, such a model offsets the growth of GWP, at rate α, by damages caused by global temperature rise T above pre industrial levels. Dell, Jones and Olken (2009) argue that rising temperature decreases the growth rate of GWP. Using country panel data, within-country cross-sectional data and cross country data they derive a temperature effect which accounts for adaptation: Yearly growth, after adaptation, is lowered by $\beta = 0.005$ per degree centigrade warming. For fixed T, the damage would be proportional to current GWP, and for fixed current GWP the damage would be proportional to T. The difference equation would then be:

6) $GWP(t+1) = (1+\alpha)GWP(t) - \beta T(t)GWP(t).$

Other equations would capture the dynamics of T(t) in terms of global carbon emissions and the carbon cycle. Without filling in these details, we can see where the non−linear dynamics of (7) will lead. As temperature rises, eventually $\beta T(t) = \alpha$. According to the World Bank, GWP has grown at 3% yearly over the last 48 years; substitution $\alpha = 3\%$ yields a tipping point at T =

6°C. Beyond this point, GWP's growth is negative. Given the inertia in the climate system, it can stay negative for quite some time.

The SOM distinguishes inner and outer measures of climate damages, and a more data driven model for the effects of climate change is sketched.

6. Conclusion

There is one simple conclusion: uncertainty quantification of the consequences of climate change should be resourced at levels at least comparable to major projects in the area of nuclear safety. Fourth generation uncertainty analyses will hopefully apply the lessons learned in the first three generations. Uncertainty should be assessed on outcomes of possible observations by independent panels of experts, whose performance is validated on variables from their field with values known post hoc. Dependence must be addressed either through direct assessment or through probabilistic inversion, and a rich set of models should be developed to deal with model uncertainty.

References

Aspinall, W. (2010) "A route to more tractable expert advice" Nature, vol. 463, 21 January.

Barro, R.J. and Sala-i-Martin X. (1999) Economic Growth MIT Press, Cambridge Ma.

Bedford, T.J., & Cooke, R.M., (2002) "Vines - a new graphical model for dependent Random Variables"- in Annals of Statistics, vol. 30, No.4, 1031-1068.

Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan.

Cooke R.M. (1991), Experts in Uncertainty; Opinion and Subjective Probability in Science, Oxford University Press; New York Oxford.

Cooke, R.M., Goossens, L.H.J. (2008) TU Delft Expert Judgment Data Base, Special issue on expert judgment Reliability Engineering & System Safety, 93, 657-674, No. 5.

Dell, M. Jones, B.F., and Olken, B.A. (2009) "Temperature and Incomd: reconciling new cross-sectional and panel estimates" *American economic Review Papers and Proceedings*

Evans, D. J. and Sezer, H, (2005), Social discount rates for member countries of the European Union, *Journal of Economic Studies*, **32**, issue 1, p. 47-59.

Fredrick, S. Loewenstein, G., O'Donoghue, T. (2002)"Time discounting and time preference: a critical review" Journal of Economic Literature vol. 40, no. 2, 351−401

Industrial Economics (2006) "Expanded Expert Judgment Assessment of the Concentration-Response Relationship Between PM2.5 Exposure and Mortality Final Report" September 21, prepared for: Office of Air Quality Planning and Standards U.S. Environmental Protection Agency Research Triangle Park, NC 27711.

Kurowicka D. and Cooke R.M..(2006) Uncertainty Analysis with High Dimensional Dependence Modeling. Wiley,

Lewis, H. et al. (1979) Risk Assessment Review Group Report to the US Nuclear Regulatory Commission NUREG/CR-04000.

Morgan, G.M. Dowlatabadi,H. Henrion,M. Keith,D. Lempert,R. McBride,S. Small,M. and WilbanksT.(2009) "Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decisions", U.S. Climate Change Science Program Synthesis and Assessment Product 5.2.

Morgan, M.G., Henrion, M (1990). Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis, New York: Cambridge University Press.

Morgan, M.G., Keith, D. (1995) "Subjective judgments by climate experts" Environ. Sci. Technol. 29:468A–476A.

National Council on Radiation Protection and Measurement:NCRP Commentary No. 14,(1996) "A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination," National Committee on Radiation Programs, Scientific Committee 64-17.

Newell, R.G. and Pizer, W.A. (2003) "Discounting the distant future: how much do uncertain rates increase valuations?" J. of Environmental Economics and Management 46, 53-71.;

Nordhaus W. and Popp, D. (1996) "What is the value of scientific knowledge: an application to global warming using the PRICE model"   Cowles Foundation discussion paper no.1117.

Nordhaus, W. D. (2008). A Question of Balance: Weighing the Options on Global Warming Policies. New Haven, CT, Yale University Press.

Nordhaus, W.D. (1994) "Expert opinion on climatic change" American Scientist, vol. 82 p. 45-51.

Pizer, W.A. (1999) "The optimal choice of climate change policy in the presence of uncertainty" Resource and energy economics, 21 255-287.

Roe, G. (2009) "Feedbacks, Timescales,and Seeing Red", Annu. Rev. Earth Planet. Sci. 37:93–115

Salmon, F. (2009) "Recipe for disaster" The formula that killed Wall Street" Wired Magazine, 17.03, 02.23.

Stern, N. (2008). "The Economics of Climate Change " American Economic Review 98(2): 1-37.

U.S. EPA (1997) "Guiding Principles for Monte Carlo Analysis" EPA/630/R-97/001.

U.S.AEC (1957), Theoretical Possibilities and consequences of Major Accident in Large Nuclear Power Plants, U.S. Atomic Energy Commission, WASH-740.

US Environmental Protection Agency, (2005), Guidelines for Carcinogen Risk Assessment, EPA/630/P-03/001F, Risk Assessment Forum, Washington, DC (March); http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=116283.

US Nuclear Regulatory Commission (1979), Nuclear Regulatory Commission issues policy statement of Reactor Safety Study and Review by the Lewis Panel, NRC press release, no. 79-19, 19 January.

US Nuclear Regulatory Commission (1983) PRA Procedures Guide, NUREG/CR-2300, 1983.

US Nuclear Regulatory Commission (1991) Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants, Vol. 1 - Final Summary Report; Vol. 2, Appendices A, B & C Final Report; & Vol 3, Appendices D & E. NUREG-1150.

US Nuclear Regulatory Commission, (1987) Reactor Risk Reference Document NUREG-1150.

Weitzman, M.L. (2007)"Structural uncertainty and the value of statistical life in the economics of catastrophic climate change" http://www.economics.harvard.edu/faculty/Weitzman/files/ValStatLifeClimate.pdf

Weitzman, Martin L. (2001) "Gamma Discounting" Source: The American Economic Review, Vol. 91, No. 1 (Mar., 2001), pp. 260-271

Zickfeld, K. Morgan M.G., Frame, D.J. and Keith D.W. (2010)"Expert judgments about transient climate response to alternative future trajectories of radiative forcing" PNAS July 13, fol. 107, no. 28, 12451–12456.