****

[Moss, R.H. and Schneider, S.H., 2000: Uncertainties in the IPCC TAR: Recommendations to lead authors for more consistent assessment and reporting. In: Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC [eds. R. Pachauri, T. Taniguchi and K. Tanaka], World Meteorological Organization, Geneva, pp. 33-51.]

****

UNCERTAINTIES IN THE IPCC TAR: Recommendations To Lead Authors For More Consistent Assessment and Reporting

Prepared by

Richard H. Moss[1] **Battelle Pacific Northwest National Laboratory (PNNL), USA**

Stephen H. Schneider[2] **Stanford University, USA**

CONTENTS

---

[1] Richard Moss is a scientist in the Washington, DC office of the Global Climate Group of Battelle Pacific Northwest National Laboratory (PNNL). He is currently on assignment as Director of the Office of the US Global Change Research Program. PNNL is operated by Battelle Memorial Institute (BMI) for the US Department of Energy under contract DE-AC06-76RLO 1830. Work on this paper was carried out under contract 12062. The views and opinions expressed are those of the author and do not represent positions of the United States Government, BMI, or PNNL.

[2] Stephen Schneider is a professor in the Department of Biological Sciences (Mail Stop 5020), a Senior Fellow at the Institute for International Studies, and Professor by Courtesy in the Department of Civil Engineering at Stanford University, Stanford, CA 94305 USA. Work on this paper was supported, in part, by a grant from the Winslow Foundation.

Note

This paper contains recommendations to writing teams of the IPCC Third Assessment Report on how to improve consistency of assessment and reporting of key uncertainties. Part I (Introduction) provides background information, including the rationale for this guidance. Part II lists the specific recommended steps—concisely in Box 1, and more fully in the text itself. Annex 1 contains examples from each of the contributions of the three IPCC Working Groups to the Second Assessment Report (SAR) to illustrate the diversity of approaches used to assess and characterize uncertainty in previous IPCC efforts. This “final” revision responds to comments received during three previous rounds of review. We use quotes around the adjective “final” to describe this published version because we know that work on guidelines such as these will never truly be completed. The chapter writing teams will undoubtedly offer additional comments and criticisms based on their experience in preparing their contributions to the TAR, as the debate on this topic continues. We welcome further opportunities to work with some of those authors in revising the guidance after the TAR is complete, perhaps generalizing the recommendations for use in other assessments.

Acknowledgements

We acknowledge and thank those who have commented on one or more of the versions of this document: Bryson Bates, Tom Beer, Richard Berk, Graciela Chichilnisky, Don Forbes, Filippo Giorgi, Dave Griggs, Roger Jones, David Karoly, Lee Ann Kozak, Neil Leary, Bo Lim, Mike MacCracken, Martin Manning, Pim Martens, Luis Jose Mata, Tony McMichael, Bert Metz, Granger Morgan, Adil Najam, Rajendra Pachauri, Jiahua Pan, Anand Patwardhan, Joyce Penner, Michel Petit, Barrie Pittock, Michael Prather, V. Ramaswamy, John Schellnhuber, Mike Scott, Serguei Semenov, Susan Solomon, John Stone, Rob Swart, T. Taniguchi, Alistair Woodward, and Kenji Yamaji. The content of the draft (including any remaining errors) remains our responsibility. We deeply appreciate the editing help we received from Kristin Kuntz-Duriseti and the efforts of Kanako Tanaka in getting the final version ready for publication.

1. Introduction

> “The IPCC function is to assess the state of our understanding and to judge the confidence with which we can make projections of climate change and its impacts. These tentative projections will aid policymakers in deciding on actions to mitigate or adapt to anthropogenic climate change, which will need to be re-assessed on a regular basis. It is recognized that many remaining uncertainties need to be reduced in each of [many] disciplines, which is why IPCC projections and scenarios are often expressed with upper and lower limits. These ranges are based on the collective judgment of the IPCC authors and the reviewers of each chapter, but it may be appropriate in the future to draw on formal methods from the discipline of decision analysis to achieve more consistency in setting criteria for high and low range limits.”
>
> Climate Change 1995: The Science of Climate Change, Chapter 11 (McBean et al., 1996)

One of the major challenges in preparing the IPCC Third Assessment Report (TAR) is that authors will need to present a clear snapshot of information on climate change, potential impacts, and response options, when the extent of what we know is continuously evolving. Given the needs of decision-makers to weigh potential responses to the risks of climate change before all uncertainties can be resolved, the available information, imperfect as it may be, must be synthesized, evaluated, and presented in a responsible and informative manner. To do this, lead authors will be reviewing the published literature, documenting the ranges and distributions of findings and estimates in the literature, assessing the scientific merit of this information, and explicitly distinguishing and communicating which findings are well understood, which are somewhat understood, and which are speculative. In short, assessment of the relative credibility of a variety of processes and outcomes is a major goal of the Reports.

In the past, writing teams preparing different IPCC reports and methodologies (e.g., IPCC Guidelines for National Greenhouse Gas Inventories) have used a variety of approaches for developing best estimates and ranges, as well as a number of terms for describing the "state of science" or level of certainty attached to a particular finding. Terms such as "almost certain," "probable," "likely," "possible," "unlikely," "improbable," and "doubtful," have been used, along with variations on "high, medium, and low confidence." These terms have not been carefully—or quantitatively—calibrated and thus have been used differently across chapters and reports, let alone in the interpretations of these materials by the general public and the media (e.g., Moss and Schneider, 1997).

The purpose of this guidance paper is to recommend the elements of a common approach for assessing, characterizing, and reporting uncertainties in a more consistent—and to the extent possible, quantitative—fashion across the various chapters of the TAR. It is hoped that the recommendations will enable authors to be more systematic in characterizing the types and sources of uncertainty. In turn, this should help improve communication between the research community, decisionmakers, and interested publics regarding what is known and unknown (and to what degree) about key dimensions of the climate issue.

Attempts to achieve more consistency in assessing and reporting on uncertainties have not received much attention. Some researchers have expressed concern that it is difficult to even know how to assign a distribution of probabilities for outcomes or processes that are laced with different types of uncertainties (a number of studies since the SAR do use probability distributions, e.g., Morgan and Dowlatabadi, 1996;, see also the citations in Schneider, 1997). However, the scientific complexity of the climate change issue and the need for information that is useful for policy formulation present a large challenge to researchers and policymakers alike—it requires both groups to work together towards improved communication of uncertainties. Reaching this goal is especially challenging in an assessment process such as the IPCC, where writing team "group dynamics" adds a great deal of complexity. Note, for example, that uncertainty within a group resulting from conflicting strongly held individual views is qualitatively different from that which exists within a group of uncertain individuals, and that knowledge of this qualitative component of the uncertainty may be valuable to audiences of the report. The research community must also bear in mind that users of IPCC reports often assume for themselves what they think the authors believed to be the distribution of probabilities when the authors do not specify it for themselves. For example, integrated assessment specialists may have to assign probabilities to alternative outcomes (even if only qualitatively specified by natural scientists) since many integrated assessment tools require estimates of the likelihood of a range of events in order to calculate efficient policy responses. We believe *it is more rational for scientists debating the specifics of a topic in which they are acknowledged experts to provide their best estimates of probability* distributions and possible outliers based on their assessment of the literature *than to have users less expert in such topics make their own determinations* (e.g., Morgan and Henrion, 1990).

The term "uncertainty" can range in implication from a lack of absolute sureness to such vagueness as to preclude anything more than informed guesses or speculation. Sometimes uncertainty results from a lack of information, and on other occasions it is caused by disagreement about what is known or even knowable. Some categories of uncertainty are amenable to quantification, while other kinds cannot be sensibly expressed in terms of probabilities (see Schneider *et al.*, 1998, for a survey of the recent literature on characterizations of uncertainty). It is important to note that uncertainty is not unique to the domain of climate change research. Even researchers in areas of science confined to the laboratory must confront uncertainties that arise from such factors as linguistic imprecision, statistical variation, measurement error, variability, approximation, subjective judgment, and disagreement. However in climate research, as in other areas such as seismic hazard prediction, ozone depletion, and hazardous wastes, these problems are compounded by additional characteristics. These include their global scale, long time lags between forcing and response, low frequency variability with characteristic times greater than the length of most instrumental records and the impossibility of before-the-fact experimental controls. Moreover, because climate change and other complex, socio-technical policy issues are not just scientific topics but also matters of public debate, it is important to recognize that even good data and thoughtful analysis may be insufficient to dispel some aspects of uncertainty associated with the different standards of evidence and degrees of risk aversion/acceptance that individuals participating in this debate may hold (Casman *et al.*, 1999, and Morgan, 1998).

A final note before turning to the specific recommendations themselves—the paper assumes that for most instances in the TAR, a "Bayesian" or "subjective" characterization of probability will be the most appropriate (see, e.g., Edwards, 1992, for a philosophical basis for Baysian methods; for applications of Bayesian methods, see e.g., Anderson, 1998; Howard *et al.*, 1972). The Bayesian paradigm is a formal and rigorous language to communicate uncertainty. In it, a "prior" belief about a probability distribution (typically based on existing evidence) can be updated by new evidence, which causes a revision of the prior, producing a so-called "posterior" probability. Applying the paradigm in the assessment process involves combining individual authors' (and reviewers') Bayesian assessments of probability distributions and would lead to the following interpretation of probability statements: the probability of an event is the degree of belief that exists among lead authors and reviewers that the event will occur, given the observations, modeling results, and theory currently available. When complex systems are the topic, both prior and updated probability distributions usually contain a high degree of (informed) subjectivity. Thus in the TAR, we expect Bayesian approaches to be what is most often meant when probabilities are attached to outcomes with an inherent component of subjectivity or to an assessment of the state of the science from which confidence characterisations are offered.

Some scientists have expressed concern that scientific investigation requires a long sequence of observational records, replicable trials, or model runs (e.g., Monte Carlo simulations) so that authors can adopt a frequentist approach to characterise the significance of the results. In other words, results should always be specified by a formal statistical characterization of the frequency and frequency distribution of the outcomes being assessed. Such objective or frequentist probabilities are not always possible in the context of scientific assessment intended to help with the policy process. In fact, even in a research setting, the idea of a limitless set of identical and independent trials that is "objectively out there" is a heuristic device that we use to help us quantify uncertainty in one particularly rigorous way—while there may be a large number of trials in some cases, this is not truly the same as a "limitless" number, and rarely are these trials truly identical or independent.

It is certainly true that "science" itself strives for objective empirical information to test theory and models. But at the same time "science for policy" must be recognized as a different enterprise than "science" itself, since science for policy (e.g., Ravetz, 1986) involves being responsive to policymakers' needs for expert judgment at a particular time, given the information currently available, even if those judgments involve a considerable degree of subjectivity. The methods outlined below are designed to make such subjectivity both more consistently expressed (linked to quantitative distributions when possible) across the TAR, and more explicitly stated so that well-established and highly subjective judgments are less likely to get confounded in policy debates. The key point is that *authors should explicitly state what sort of approach they are using in a particular case*: if frequentist statistics are used the authors should explicitly note that, and likewise if the probabilities assigned are subjective, that too should be explicitly indicated. Transparency is the key in all cases.

2. Options for improving consistency and clarity

This section provides specific recommendations intended to increase the consistency with which authors assess and communicate uncertainties in the TAR. Given the diverse subject areas, methods, and stages of development of the many areas of research to be assessed in the TAR, the paper cannot provide extremely detailed procedures that will be universally applicable. Therefore, this document provides *general* guidance; writing teams will need to formulate their own detailed approaches for implementing the guidance while preparing their chapters and summaries. The recommended steps are summarized in Box 1 and discussed more extensively in the text.

Note that these steps are intended to be applied to a relatively small number of the major conclusions and/or estimates (of parameters, processes, or outcomes) that will be developed in each of the chapters, for example the main findings discussed in a chapter executive summary or forwarded for inclusion in the summary for policymakers. It is not intended that authors must follow all of these steps every time they use a term such as “likely” or “unlikely” or "medium confidence" in the main text of their chapters or every time a specific result is given. However, one recommendation that should be applied throughout the report is that care should be taken to avoid vague or very broad statements with “medium confidence” that are difficult to support or refute. For example, if we know very little, we often are indifferent to whether climate change will cause a positive or negative response in some variable. In this trivial case, we would actually have at least medium confidence (i.e., near 50%--as defined below in Fig 3) that "warming could alter biodiversity". That says nothing profound unless we add quantitative modifiers on the amount of warming and the direction and severity of the biodiversity change. The point is to phrase all conclusions so as to avoid nearly indifferent statements based on speculative knowledge. In addition, all authors—whether in Working Group I, II or III—should be as specific as possible throughout the report about the kinds of uncertainties affecting their conclusions and the nature of any probabilities given.

*Moreover, not all chapters are expected to be able to implement all of the steps for each of the conclusions and/or estimates that will be developed in their assessment of the literature***.** For example, step 7 in Box 1 (“Use formal probabilistic frameworks…”) may be appropriate for only a few key issues in a limited number of chapters.

Box 1
Summary of steps recommended for assessing uncertainty in the TAR

1. For each of the major findings you expect to be developed in your chapter, *identify the most important factors and uncertainties that are likely to affect the conclusions*. Also specify which important factors/variables are being treated exogenously or fixed, as it will almost always be the case that some important components will be treated in this way when addressing the complex phenomena examined in the TAR.

2. *Document ranges and distributions in the literature*, including sources of information on the key causes of uncertainty. Note that it is important to consider the types of evidence available to support a finding (e.g., distinguish findings that are well established through observations and tested theory from those that are not so established).

3. Given the nature of the uncertainties and state of science, *make an initial determination of the appropriate level of precision*—is the state of science such that only qualitative estimates are possible, or is quantification possible, and if so, to how many significant digits? As the assessment proceeds, recalibrate level of precision in response to your assessment of new information.

4. Quantitatively or qualitatively *characterise the distribution of values that a parameter, variable, or outcome may take*. First identify the end points of the range that the writing team establishes, and/or any high consequence, low probability outcomes or “outliers.” Particular care needs to be taken to specify what portion of the range is included in the estimate (e.g., this is a 90% confidence interval) and what the range is based on. Then provide an assessment of the general shape (e.g., uniform, bell, bimodal, skewed, symmetric) of the distribution. Finally, provide your assessment of the central tendency of the distribution (if appropriate).

5. Using the terms described below, *rate and describe the state of scientific information* on which the conclusions and/or estimates (i.e. from step 4) are based.

6. *Prepare a “traceable account”* of how the estimates were constructed that describes the writing team’s reasons for adopting a particular probability distribution, including important lines of evidence used, standards of evidence applied, approaches to combining/reconciling multiple lines of evidence, explicit explanations of methods for aggregation, and critical uncertainties.

7. OPTIONAL: *Use formal probabilistic frameworks for assessing expert judgment* (i.e. decision-analytic techniques), as appropriate for each writing team.

1. *For each of the major findings you expect to be developed in your chapter, identify the most important factors and uncertainties that are likely to affect the conclusions.* Writing teams are most likely fairly far along in this process, but it is worth stressing that this identification process is important and may require several iterations within the writing team to develop a set of well-posed questions or issues. The most important factors affecting the findings could include processes, variables, parameters, different types of data (experimental, observational, historical, field, etc.), and interdependencies that are likely to have a significant bearing on the estimates. In identifying the sources of uncertainty, it is important to consider the types of evidence available to support a finding, for example distinguishing between findings that are well established through observations or well-tested theories versus those that are not so well-established. Identification of the factors and uncertainties affecting the outcomes/estimates is important for being able to gauge the degree of uncertainty that is likely to affect the estimates. Typologies of uncertainties are available in the literature of different disciplines. We will not survey here the many such typologies that exist to classify uncertainty (e.g., see the review and citations in Schneider, et al., 1998), but rather list in Box 2 examples of common types encountered by IPCC authors.

Box 2

Examples of sources of uncertainty

Problems with data

1. Missing components or errors in the data
2. "Noise" in the data associated with biased or incomplete observations
3. Random sampling error and biases (non-representativeness) in a sample

Problems with models

4. Known processes but unknown functional relationships or errors in the structure of the model
5. Known structure but unknown or erroneous values of some important parameters
6. Known historical data and model structure, but reasons to believe parameters or model structure will change over time
7. Uncertainty regarding the predictability (e.g., chaotic or stochastic behavior) of the system or effect
8. Uncertainties introduced by approximation techniques used to solve a set of equations that characterize the model.

Other sources of uncertainty

9. Ambiguously defined concepts and terminology
10. Inappropriate spatial/temporal units
11. Inappropriateness of/lack of confidence in underlying assumptions
12. Uncertainty due to projections of human behavior (e.g., future consumption patterns, or technological change), which is distinct from uncertainty due to "natural" sources (e.g., climate sensitivity, chaos)

In phenomenon that are as complex and multifaceted as those related to climate change, it is likely that some of the processes, variables, and parameters that introduce uncertainty will be treated exogenously (i.e., as assumptions or givens that are inputs) in order to make the problem tractable. This situation is likely to arise frequently in the TAR because of the increased attention to linkages among different subject areas and the use of scenarios developed in one area of research to examine sensitivities and possible outcomes in others. Thus many estimates or outcomes will be affected not only by uncertainties in their immediate substantive domain, but also by uncertainties in the scenarios or parameters generated in other areas of research. For example, in assessing possible effects of climate change on agriculture, it may be useful to assess responses of crop yields to specified changes in climate and to present information on the level of confidence in the projected yield changes. The specified changes in climate may well be treated as exogenous to the impacts analysis in such cases, and it is important in the assessment of uncertainty to note this. Another option available to writing teams (particularly chapters dealing with integration and synthesis) is to “composite” the uncertainties in all aspects of the problem, breaking down the composite uncertainty into its sources, range, and distribution. This step will require coordination across the writing teams and TSUs of different working groups in some cases. Some chapter writing teams may believe that their limited assessment of uncertainties might be misinterpreted as representing a composite distribution by some readers. In such cases authors can both clearly identify the uncertainties they present as limited to a partial set of factors and can provide cross-references to similar partial analyses of uncertainties in other TAR chapters in which parameters or scenarios exogenous to the original chapter contain their own analyses of uncertainty for those exogenous factors. Of course, chapters assigned the task of integration and synthesis will have to consider performing an explicit compositing exercise of their own (see traceable account guidance below for suggestions on making such composite distributions).

A single aggregated damage function or a “best guess” climate sensitivity estimate is a very restricted representation of the wide range of beliefs available in the literature or among lead authors about either climate sensitivity or climate damages. If a causal chain includes several different processes, then the aggregate distribution might have very different characteristics than the various distributions that comprise the links of the chain of causality (see Jones, 2000). Thus, poorly managed projected ranges in impact assessment may inadvertently propagate uncertainty. The process whereby uncertainty accumulates throughout the process of climate change prediction and impact assessment has been variously described as a "cascade of uncertainty" (Schneider, 1983) or the "uncertainty explosion" (Henderson-Sellers, 1993). When an assessment progresses from the biogeochemical cycle to radiative forcing and climate sensitivity calculations through to economic and social outcomes, including valuations of climate damages, considerable uncertainty can be accumulated (see Figure 1).

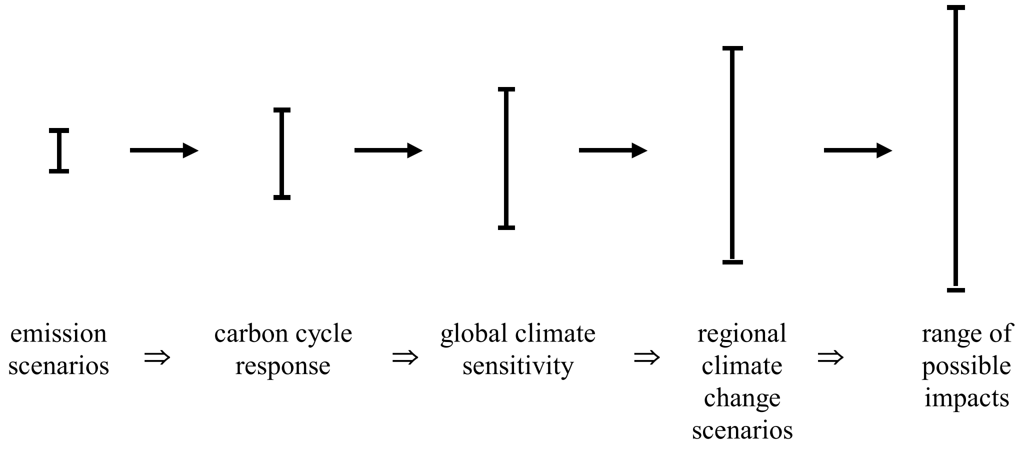


Figure 1. Range of major uncertainties typical in impact assessments showing the “uncertainty explosion” as these ranges are multiplied to encompass a comprehensive range of future consequences, including physical, economic social and political impacts and policy responses (modified after Jones, 2000, and the "cascading pyramid of uncertainties" in Schneider, 1983).

Finally, another source of uncertainty involves ambiguously defined concepts and terminology. An excellent example of this sort of uncertainty concerns the ambiguously defined concept of “surprises” (see Box 3). We suggest that TAR authors be cognizant of the ambiguities of such terms and be as precise as possible in their definitions. When one potential definition of a term is adopted over others, it is important for authors to briefly contrast their preferred definition with the others in the literature. Perhaps early in the assessment process, *a glossary of terms with the potential for inconsistent definitions across various chapters and working groups should be constructed* to reduce the likelihood of such inconsistent usage. We recommend that Coordinating Lead Authors let their respective Technical Support Units (TSUs) know about such terms as soon as they are identified to improve the consistency of terminology across the TAR as soon as practicable./

Box 3
Surprises

Strictly speaking, a surprise is an unanticipated outcome. However, in the IPCC Second Assessment Report (SAR), "surprises" were defined as rapid, non-linear responses of the climatic system to anthropogenic forcing, and analogies to paleo-climatic abrupt events were cited to demonstrate the plausibility of such a possibility. Moreover, specific examples of such non-linear behaviors that the authors could already envision as plausible were given (e.g., reorganization of thermohaline circulation, rapid deglaciation, fast changes to the carbon cycle). Strictly speaking, it would be better to define these as *imaginable abrupt events*. Finally, the WGI SAR concluded its SPM with the statement that non-linear systems when rapidly forced are particularly subject to unexpected behavior. Here is an example of unknown outcomes (i.e., true surprises) but *imaginable conditions for surprise* (Schneider et al., 1998).

2. *Document ranges and distributions in the literature*, including sources of information on the key causes of uncertainty. Writing teams should document ranges and distribution of estimates in the published literature, describing how the ranges and distributions are constructed, and clearly specifying what they signify, e.g., a 2 sigma range, results not related to a particular confidence interval, etc. Note that it is important to consider the types of evidence available to support a finding (e.g., distinguish findings that are well established through observations and tested theory from those that are not so established). As part of the process of assessing the literature and drafting the chapter, it is critical to characterise not just a single estimate, but a range of estimates and associated probability distributions. This should include attention not only to the central tendency, but also to the end points of the range of outcomes, possible outliers, the likelihood that outcomes beyond the end points of the range might occur, and the type of distribution of potential outcomes, e.g., normal, bimodal, etc.

3. Given the nature of the uncertainties and state of science, *make an initial determination of the appropriate level of precision*—is the state of science such that only qualitative estimates are possible, or is quantification possible, and if so, to how many significant digits? As the assessment proceeds, recalibrate the appropriate level of precision in response to your assessment of new information. For example, in some cases, lead authors may determine, after summarizing recent literature, that statements such as "increase" or "decrease" may be all that can be justified. In other cases, they may determine that quantification is possible at a high level of precision. Note that it is not necessary early in the group's deliberations to try to determine the estimates or numerical ranges themselves, simply the type of estimate that appears appropriate given the prior information available to the authors about the types and expected levels of uncertainty.

4. Quantitatively or qualitatively *characterise the distribution of values that a parameter, variable, or outcome may take*. First identify the end points of the range, and/or any high consequence, low probability outcomes or "outliers." Particular care needs to be taken to specify what portion of the range is included in the estimate (e.g., this is a 90% confidence interval) and what the range is based on. Then provide an assessment of the general shape (e.g., uniform, bell, bimodal, skewed, symmetric) of the distribution. Finally, provide your assessment of the central tendency of the distribution (if appropriate). Characterisations may be qualitative and/or quantitative, depending on the authors' assessment of the state of knowledge.

Following this particular order (i.e., identifying end points of the range of probability distributions and possible outliers before best estimates) is important because of the well-documented tendency of assessors to overstate the confidence with which outcomes are likely to lie near the central tendency. This has been shown to happen because of reliance on different subconscious rules of thumb in reaching these judgments. The idea, spelled out in Kahneman et al. (1982), is that, due to limited mental processing capacity, humans rely on strategies of simplification, or mental heuristics, to reduce the complexity of judgment tasks. While facilitating decision-making, these procedures are vulnerable to systematic error and bias. These heuristic devices include "availability" (relating the probability of this outcome to previous occurrences or ease with which one could imagine such occurrences); and "anchoring and adjustment" (judgment of probability of this outcome is overly influenced by the starting estimate, which becomes an "anchor" for subsequent estimates—which may lead to overconfidence in central tendencies or underestimates of range outliers, as mentioned earlier).

Overconfidence is a cognitive illusion that has been reported to bias experts' judgments. A considerable amount of evidence has been amassed for the view that people suffer from an overconfidence bias (Kahneman and Lovallo, 1993, Kahneman and Tversky, 1979, 1996 Kahneman et al., 1982, Tversky and Kahneman, 1974, 1983). The common finding is that respondents are correct less often than their confidence assessments imply. However "ecological" theorists (cf. McClelland and Bolger, 1994) claim that overconfidence is an artefact of the artificial experimental tasks and the non-representative sampling of stimulus materials; thus the appearance of overconfidence may be an illusion created by research and not a cognitive failure by respondents. They (Gigerenzer, 1994, 1996, Gigerenzer *et al.*, 1991, and Juslin, 1994) claim that individuals are well adapted to their environments and do not make biased judgments. Furthermore, in cases of judgments of repeated events (weather forecasters, horse race bookmakers, tournament bridge players), experts make well-calibrated forecasts. In these cases, respondents might be identifying relative frequencies for sets of similar events rather than judging likelihood for individual events (e.g. Wright and Ayton, 1992).

Writing teams should be clear what sort of range and confidence interval they are constructing, or what sorts of possible outcomes are included in the range. For example, do the endpoints (or outliers beyond them) include potential known or imaginable non-linear rapid events? Is the range given by the authors the one in which the "true" value would fall with a two out of three chance (or some other probability)? Or is the range the one in which two thirds of modelled outcomes available in the literature would lie? These are all very different statements, and care should be taken in clarifying exactly what is meant. One suggestion made by several authors in the review comments on earlier drafts of this guidance paper is to establish a uniform standard for probability distributions (e.g., 67%, or one-sigma if a normal distribution), and then to allow the rating of confidence in this distribution to fluctuate to reflect the quality of the evidence available (see step 6 for a discussion of assessment of the quality of scientific information).

On the other hand, different reviewers noted that for some parameters or outcomes, continuous estimates of probability are available in the literature from extensive data analyses, Monte Carlo model runs, or formal decision-analysis elicitations (e.g., see Figure 2). In those cases, cumulative distribution functions (CDFs) can be drawn. These provide the probability that an estimate of a parameter would be less than or equal to given numerical value of that parameter.

It is important to note that by providing only a truncated estimate of the full range of outcomes (e.g., not specifying outliers that include "surprises", and thus making the range of outcomes described smaller), one is not conveying to potential users a representation of the full range of uncertainty associated with the estimate. This has important implications regarding the extent to which the report accurately conveys uncertainties. Some authors are likely to feel uncomfortable with the full range of uncertainty, because the likelihood of a "surprise" or events at the tails of the distribution may be extremely remote or essentially impossible to gauge experimentally, and the range implied could be extremely large. Thus there may be a case to be made for providing a truncated range in addition to outliers for a specific case, provided that it is clearly explained what the provided range includes and/or excludes. It should be stressed that if a truncated range is provided, it is important that authors specify how likely it is that the answer could lie outside the truncated distribution, and what was the basis for specifying such possibilities.

If possible, it would be useful for writing teams to provide an assessment of the shape of the distribution. Is it roughly uniform, such that values that are close to the mean are no more or less likely than values that are slightly more distant from the mean (e.g., we believe that the state of the science is not well-established and thus the climate damages from a few degrees C warming in developed countries aggregated across all market sectors would likely have a uniform distribution within plus or minus a few percentage points of GNP)? Or roughly bell shaped, such that the true value is more likely to lie near the estimated mean than in an interval that is distant from the mean? Is the distribution thought to be symmetric or skewed? If skewed, how?

In deciding whether there is a "best estimate," writing teams should evaluate the shape of the distribution of outcomes, i.e., the chance that different outcomes would occur. Depending on the approach being taken in the chapter, the “best guess” could be the mean, the median, the mode, or some other value. If all outcomes within the range seem equally probable (i.e., the probability distribution is uniform), writing teams may not consider it appropriate to make a “best guess” regarding the outcome. An important consideration is evaluation of the logic from which a best estimate may be based. For example, the values within the range might have been determined by assumptions from equally plausible “what-if” scenarios, or, alternatively, they might be derived by various methodologies whose reliability cannot be simply evaluated. Either way, this should be reported, and no best estimate offered.

In developing a best estimate, authors need to guard against aggregation of results (spatial, temporal, or across scenarios) if it hides important regional or inter-temporal differences. It is important not to combine automatically different distributions into one summary distribution. For example, most participants or available studies might believe that the possible outcomes are normally distributed, but one group might cluster its mean far from the mean of another group, resulting in a bimodal aggregate distribution. In this case, it is inappropriate to combine these into one summary distribution unless it is also indicated that there are two (or more) “schools of thought.”

“Costs of mitigation” is an example. Perhaps one sub-group using tools which assume perfect markets would place the costs of achieving some specific GHG concentration targets and timetables at an average 1-2% decrease in annual GDP, whereas another group might estimate that the costs for achieving such targets and timetables would actually be zero or even slightly negative due to the existence of “no regrets” technological options.

Climate sensitivity is another example, as seen in Figure 2. Here scientists 2 and 4 offer a very different estimate of range outliers (i.e., values below the 5th percentile tick mark on the left end of each “box and whisker” plot or the 95th percentile estimate at the right hand end of each box and whisker plot) for imaginable abrupt events. But the means and variance of scientists 2 and 4 are quite similar to 13 of the 14 remaining scientists in this decision analytic survey, the exception being scientist 5. This is an example where it would likely be inappropriate to aggregate all respondents distributions into a single composite estimate of uncertainty since scientist 5 has a radically different mean and variance estimate than the other 15 scientists. If this were to occur in the literature or among lead authors reflecting on the literature, it is not appropriate to aggregate such “schools of thought” into a single distribution, but rather to show the two “paradigms” and mention the amount of support expressed for each distribution.

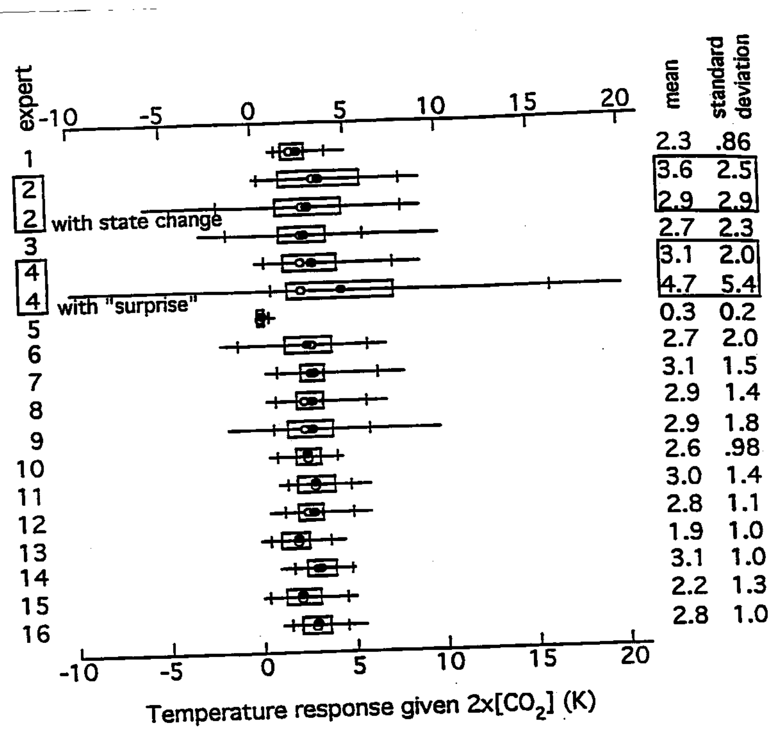


Figure 2. Climate Sensitivity under 2xCO$_2$ forcing, °K. Box plots of probability distributions (elicited from 16 climate scientists) of the change in global average surface temperature resulting from a doubling of CO$_2$. The horizontal lines denote the range from minimum (1st percentile) to maximum (99th percentile) assessed possible values. Vertical marks indicate the locations of the lower 5th and upper 95th percentiles. The boxes indicate the interval spanned by the 50% confidence interval. The solid dots indicate the mean and open dots, the median. Source: Morgan and Keith, 1995.

This is not simply a reporting of values available in the literature, but rather the assessment of the Lead Authors' about the relative likelihood that different values in the literature represent accurate estimates or descriptions. Writing teams will need to guard against the potential for "gaming" or strategic behavior, in which participants might select outlier estimates to compensate for what they consider to be over- or under-estimates by some of their colleagues or estimates in the literature.

5. Using the terms described below, *rate and describe the state of scientific information* on which the conclusions and/or estimates (i.e. from step 4) are based. This assessment of the state of knowledge should reflect both the type/amount of evidence (e.g., observations, interpretation of model results, or expert judgement) and the level of peer acceptance/consensus. The text should distinguish between confidence statements based on well-established, "objective" findings versus those based on subjective judgements. Care should be taken not to fall into the trap of widening a confidence interval to take account of outliers and then describing the confidence in the conclusions as low (e.g., as in Figure 2 of the Technical Summary of the Working Group I SAR). If the confidence interval is sufficiently wide, there should be moderate to high confidence that the true value will lie within it.

Conclusions should be phrased in such a way to avoid statements of indifference that are not illuminating. Clear, precise statements with assessed confidence levels are preferable. In particular, meaningless "medium" confidence labels should not be used. For example, it would indeed be "medium confidence," the way we've defined it in Fig. 3 (after 3 rounds of IPCC peer reviews), to say that "global warming could increase El Nino frequency". Knowing virtually nothing more than El Nino is partly driven by large scale forcings makes that statement--or one which replaces the "increase" with "decrease"--an even bet. A much more meaningful statement would be qualified. For instance, "Climate change of more than two degrees warming would cause a substantial increase in the El Nino frequency." Personally, we have low confidence in that clearly bounded conclusion.

An expert assessment might disagree, but at least the statement isn't referring to an indifferent outcome (i.e., increase or decrease) from an unspecified climate change. In short, assessors should strive to avoid using language that expresses indifference (change in either direction – increase or decrease – is equally likely) and then assign what amounts to an essentially meaningless "medium confidence" label to the conclusion.

In addition, the language of the text should be consistent with the level of confidence – specifically, avoid using double qualifiers that undermine confidence in the conclusion. For example, if words like could or might are included, then the implication in that the statement is very likely to be true and should not carry an indifferent "medium confidence" label; a "high confidence" label is more consistent with the language. If authors are uneasy about using anything but conditional statements, then they should either include no confidence level label or an appropriately high one, since the conditional language implies the statement is very likely.

Previous drafts of this guidance paper have suggested a variety of options for describing the state of knowledge. In response to calls for a straightforward approach, we suggest that the IPCC agree to test a single set of terms. As the assessment progresses, it will be critical to review these terms and the consistency with which they are applied by various writing teams. As noted by many reviewers of earlier drafts of the guidance paper, consistency in the use of confidence descriptors is critical, and a clear way to assure this is to have a discrete quantitative scale such as that suggested below (Figure 3). Without such a discrete quantitative scale, there is strong experimental evidence that the same uncertainty words often have very different meanings for different people in different circumstances (e.g., Morgan and Henrion, 1990).

(1.00) "Very High Confidence" (0.95)
(0.95) "High Confidence" (0.67)
(0.67) "Medium Confidence" (0.33)
(0.33) "Low Confidence" (0.05)
(0.05) "Very Low Confidence" (0.00)

Figure 3. Scale for Assessing State of Knowledge

We realise from the comments on earlier drafts that some may be uncomfortable with having only one option, and thus we propose a set of qualitative uncertainty terms that can be used to supplement the five point scale and explain why a writing team may express high, medium, or low confidence in a particular finding (see Figure 4). We propose this as a supplement rather than as an alternative because these qualitative terms do not always map well onto a quantitative scale, increasing the likelihood of inconsistent usage.

HIGH

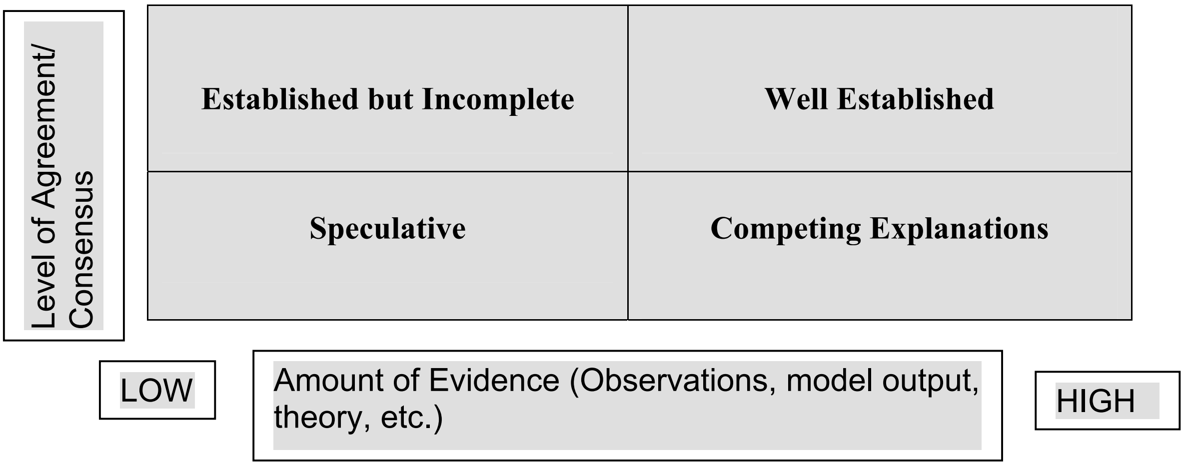


Figure 4. Supplemental Qualitative Uncertainty Terms.

Key to qualitative "state of knowledge" descriptors:

***Well-established*: models incorporate known processes; observations largely consistent with models for important variables; or multiple lines of evidence support the finding)**

***Established but Incomplete:* models incorporate most known processes, although some parameterizations may not be well tested; observations are somewhat consistent with theoretical or model results but incomplete; current empirical estimates are well founded, but the possibility of changes in governing processes over time is considerable; or only one or a few lines of evidence support the finding**

***Competing Explanations*: different model representations account for different aspects of observations or evidence, or incorporate different aspects of key processes, leading to competing explanations**

***Speculative*: conceptually plausible ideas that haven't received much attention in the literature or that are laced with difficult to reduce uncertainties or have few available observational tests**

Graphical approach to communicating the state of knowledge. Several experts have suggested development of a graphical approach for communicating evaluations of the state of knowledge. Simple approaches used in previous assessments have included systems of symbols such as asterisks to denote different levels of confidence in findings.

We explore here another option: a "radar plot" or "snowflake chart" that signifies increasing confidence as it increases in size. For some users, the size of the graphic would be the only interpretation required, but for others who wish to understand more about why a particular level of confidence was assigned, the axes of the plot would provide additional information about the major sub-components of the evaluation of the state of knowledge. In the following examples (Figures 5A-C), produced in a spreadsheet, the overall size of the graphic varies with hypothetical writing teams' assessments (on a scale of 1 to 10) of the amount/quality of theory, observations, and model results available to support a finding. The degree of consensus in evaluations of the members of the writing team is calculated and normalised using a non-parametric statistic. If there were interest in this approach among writing teams, further refinement of the concept would be required. One possible modification would be to drop the "consensus" axis and provide a box and whisker plot on each of the other three axes, thus allowing users to see the actual distribution of evaluations (having an odd number of axes would make it clearer that "consensus" and "observations," for example, are not opposites). The main advantage of such an approach is that it does not attempt to compress several judgments that go into the evaluation of the state of knowledge into one dimension, yet, the graphic also has a simple interpretation. It conveys more information than a single number or word, yet is simpler than the n-dimensional hypercube that would be required to portray independently all the dimensions of the state of science evaluations that writing teams will be making.

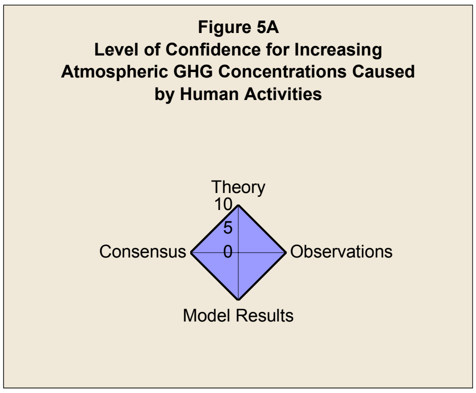


Figure 5A
Level of Confidence for Increasing Atmospheric GHG Concentrations Caused by Human Activities

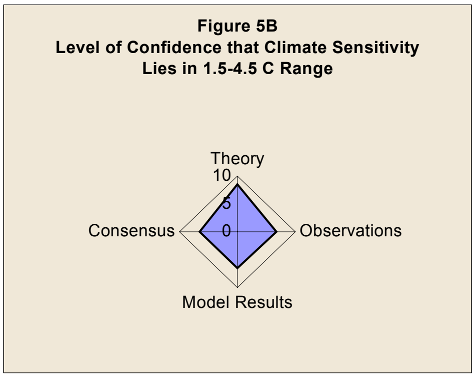
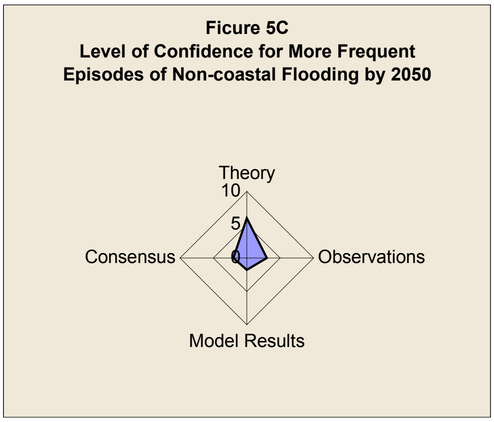


Figure 5B
Level of Confidence that Climate Sensitivity Lies in 1.5-4.5 C Range



Ficure 5C
Level of Confidence for More Frequent Episodes of Non-coastal Flooding by 2050

6. *Prepare a "traceable account"* of how the estimates were constructed that describes the writing team's reasons for adopting a particular probability distribution, including important lines of evidence used, standards of evidence applied, approaches to combining/reconciling multiple lines of evidence, explicit explanations of methods for aggregation, and critical uncertainties. In constructing the composite distributions, it is important to include a "traceable account" of how the estimates were constructed. For example, if the selection of outliers was based on the results of modeled output, but it is known that the models do not incorporate certain specific processes that are known to operate and that would, in the collective judgment of the authors, increase the range by a certain percentage, the outliers should reflect this information, and the text of the chapter should include a description of how the range was constructed. Or, if regionally heterogeneous distributions are aggregated into hemispheric or global averages, for instance, the data from which the aggregation was made should be given or a reference cited so others can perform different aggregations—a possibility only if such a traceable account is constructed. Or if costs and benefits of some climatic impact event are aggregated into a single measure of value or "numeraire"--typically a monetary unit--but different individuals or cultures might evaluate different numeraires (e.g., loss of species, changes in the distribution of effects across income groups or loss of life) differently, then it is inappropriate to aggregate across these numeraires without first providing a traceable account of how each numeraire was valued before the aggregation obscured those valuation assumptions (e.g., see the "Five Numeraires" discussion in Schneider et al, 2000).

In addition, attention should be given to the logical links generating main conclusions. Assigning confidence levels to various links in a chain of logic that leads to an important conclusion can identify areas of disagreement or limited knowledge. A single confidence level attached to the conclusion may be misleading, especially if confidence in the conclusion is low because one of the links is speculative while the others are well established.

7. OPTIONAL: *Use formal probabilistic frameworks for assessing expert judgment* (i.e. decision-analytic techniques), as appropriate. In a few cases (e.g., "detection and attribution"), lead authors

may wish to use more formal decision analytic techniques to achieve a more consistent assessment of the subjective probability distribution for a particularly important outcome (e.g., as seen on Figure 2 for climate sensitivity estimates). In addition to describing such formal studies (if any) as may already exist in the current literature, some writing teams may choose to formalize development of quantitative cumulative probability distributions of a group of lead authors or other experts they might identify. In this case, decision analysis experts from outside the chapter writing team who are familiar with techniques for and pitfalls associated with elicitation of consistent expert judgments and construction of cumulative probability distributions should be invited to assist the writing team.

In this approach, the outside experts would work with the team of authors to frame a number of explicit questions in which ranges of estimates of outcomes or parameters would be called for from authors and/or other participants (see Nordhaus, 1994; Morgan and Keith, 1995; Titus and Narayanan, 1996 for examples of decision analytic elicitations of climate effects and impacts; see Roughgarden and Schneider, 1999, and Morgan and Dowlatabadi, 1996, for examples of how such elicited subjective probability distributions can be incorporated into integrated assessment models examining "optimal" policies). Formal one-on-one interviews, group interviews, or mail-in questionnaires (perhaps with follow-up contacts) could be used to establish subjective probability distributions for a few key parameters or issues. In addition to soliciting outcomes (*e. g.*, likelihood of biodiversity losses or changes in gross domestic product), processes and major uncertainties in data or theory could also be elicited. Furthermore, questions could be designed to improve the consistency between outcome estimates and process uncertainty estimates made by each individual or group. When the responses of each author are compared, first privately for each respondent and then later as a group, the discussions that follow often very quickly focus on the main points of agreement or disagreement. These discussions can identify literature that some Lead Authors may not have been aware of in their initial responses, and thus can speed up the process of consistent assessment of outcomes or processes. The results of such formal elicitations (with or without re-elicitations) can be used in several ways, including formally in the chapter, but the actual responses of each author/outside expert (or even a group aggregate) need not necessarily be published as part of the final report. The primary goal is to improve the assessment process.

3. Graphical Communication of Uncertainty

More careful approaches to assessing and characterizing uncertainty will increase the clarity of conclusions in the IPCC TAR. The communication of this improved assessment of uncertainty associated with key findings and estimates will be aided by improved graphical representation of the results. Selection of specific approaches for graphical presentation of uncertain quantitative information is left to individual writing teams. However, it may be productive for the TSUs to facilitate an exchange of ideas or approaches at about the middle of the drafting process, so that particularly effective ideas can be propagated across the report. A number of potentially useful displays are available, involving trade-offs between simplicity and sophistication, particularly in the choice of the number of dimensions to use in presenting the information.

One example already discussed in this paper is the use of Tukey “box and whisker” plots, like those on Figure 2 which are meant to represent CDFs of subjective probability of each of 16 scientists on the magnitude of climate sensitivity. The tick marks at the left and right hand ends of the whiskers represent respondents 5th and 95th percentile estimates, respectively, and thus the lines extending beyond the tick marks represent range outliers (the end points of the whiskers being first and 99th percentiles in Figure 2). The box represents 25th and 75th percentiles and the dots mean and median of the distributions drawn by each of the 16 respondents. This graphical representation contains a remarkable amount of information about scientists’ opinions about the uncertainties associated with climate sensitivity in an accessible form. Such a graphical representation of uncertainty is a very convenient means to convey information once a writing team has agreed that a quantitative representation of uncertainty is appropriate in their particular application.

Annex. Statement of the problem: developing self-consistent collective assessments of the state of knowledge

The following examples from each of the contributions of the three IPCC Working Groups to the Second Assessment Report (SAR) illustrate the diversity of approaches used and point to the need for more explicit and consistent treatment of uncertainties in future assessments for all working groups.

Case 1: "Climate Sensitivity"

"Climate sensitivity"—the globally averaged surface temperature response that eventually can be expected to occur (i. e., at equilibrium) if the CO2 concentration were to double from pre-industrial levels and remain at this concentration indefinitely—was first estimated in a 1979 U.S. National Research Council report to be within the range of 1.5-4.5°C. While some changes in the underlying science have occurred (e. g., new formulations of cloud or biophysical parameterizations have been developed), small (±0.5°C) changes to the outliers (end points) of the ranges have seemed unjustified—even frivolous—to many in view of the absence of fundamental new data or a tested theory. Thus, the estimate of the range has remained the same over the first two IPCC reports (IPCC 1996a).

The process of achieving more consistent aggregate scientific judgments is critical to establishing more meaningful and credible ranges of potential outcomes like climate sensitivity. More consistent estimates of the endpoints of a range (e.g., as on Figure 2) for any variable would minimize misunderstandings and reduce the likelihood that interest groups could misunderstand or misrepresent the findings.

Case 2: Impacts of Climate Change on Agriculture and Food Security

A controversial aspect of the WG II SAR (IPCC 1996b) concerned the adaptive potential of agriculture to climatic change and the potential implications for regional and global food security. This assessment involves a wide range of issues, including inherent climate variability (i.e., climatic noise) that masks long term trends (making it difficult for farmers to know to what to adapt); carbon dioxide fertilization effects; crop/climate/insect interactions; the economic modeling of agricultural trade; and the socio-economic and political conditions under which hunger occurs. In the SAR, estimates of the impact of equilibrium, doubled-CO2 climate conditions were made using crop-climate models, and the assessment indicated a wide range of yield changes, from large and positive to large and negative, when compared to results under current climate conditions. The variation in results is accounted for by a variety of factors, including different assumptions about the factors above, as well as potential surprises and different assumptions in different studies regarding the extent to which adaptation is possible.

While the SAR writing team developed a definition of “vulnerability” based on the probability density function for climate change and a damage function that relates impacts to varying levels of climate change, in practice, specific climate model results played little role in the determination of vulnerability. These determinations were based on other factors, such as identifying those populations judged as being vulnerable to hunger or famine, or identifying potential thresholds for different crops.

Clearly, assessing the vulnerability of different populations will not be possible simply by using equilibrium, CO2-doubled climatic change scenarios—transient runs with various aerosol forcings from several modeling groups will also be needed (and are now available through the IPCC Data Distribution Centre established by the Task Group on Climate Scenarios for Impacts Assessment--TGCIA). Because the literature may not yet contain results from a large number of crop-climate model studies using these scenarios for many regions of the world, TAR authors will need to assess the causes of uncertainty explicitly (e.g., those arising from the factors above as opposed to those from neglect of transients or alternative aerosol scenarios). At the same time, they will need to consider (although great caution will likely need to be expressed in their conclusions) the subjective likelihood that crop yields will be altered given the available knowledge in the literature about transient scenarios and aerosol effects. By explicitly specifying the nature of uncertainties regarding agricultural impacts, the authors should be able to develop a carefully hedged set of subjective estimates of their collective judgment of possible impacts given the limited studies available to them (as well as establishing a template for future assessments). Of overarching importance is the need to convey clearly that the collective judgments of the TAR are based on an evolving hierarchy of rather heterogeneous studies, not on a linear progression beginning with a full range of plausible scenarios fed into state-of-the-art transient atmosphere/ocean/biosphere/cryosphere coupled models driving a comprehensive set of impacts models. While the latter may be a desirable goal, in practice assessors will be required to use judgments since such a comprehensive integrated modeling activity is not yet feasible (and would almost certainly have elements of unpredictability embedded).

Case 3: Aggregate Economic Impacts of Climate Change and Emissions Abatement

Working Group III of the SAR estimated “damages” resulting from climate change to range from \$5-\$125 per ton of carbon emitted, the range stemming from differences in estimation techniques as well as different assumptions about the appropriate "discount rate" to use in assessing future impacts in current monetary terms.[3] The WG III SAR presented a range of “best guess” estimates of aggregate damages. Globally, annual impacts under doubled CO2 equilibrium climate were estimated to range from 1-2 percent of GDP. Regionally, the annual impacts on GDP were estimated to range from slightly positive (for the former USSR and perhaps China), to mildly negative (for OECD countries), to as high as a 10 percent loss (for Africa).

It is extremely important to note that ranges associated with these impacts estimates simply represented the range of best guesses of the authors, not their estimates of the full range of potential damages from low to high. No estimation of uncertainties with regard to the full social costs of climate change was made, and in fact, no systematic calculations are available in the literature as yet, as far as we know. But generally, the uncertainties are known to be large and difficult to quantify formally because of such issues as valuing non-market impacts in monetary terms. The lack of estimation techniques made the existing uncertainties difficult to communicate.

In the opinion of one SAR lead author, Chapter 6 of WG III (see the discussion in Moss and Schneider, 1997) conveys a message that knowledge is better developed than in fact it is, and that uncertainties are smaller than they actually are. It does seem likely that the ranges would not be nearly as small as portrayed in the chapter had procedures to estimate, say, tenth and ninetieth percentile range outcomes (let alone subjective probability distributions within and beyond those range endpoints) been more formal, thus allowing high and low estimates of benefits/damages to be included with the range of best guesses provided. Since decision makers often devise policies to deal with low probabilities of very consequential outcomes, authors need to provide such estimates to the extent the state of the science permits.

References

---

[3] Although the choice of discount rate is a source of significant dispute, it is not a source of uncertainty, because it is a value choice, and the effects of different discount rates on the estimates of climate impacts can be calculated with precision. In fact, such value choices should be treated parametrically so that decision makers can see the implications of adopting different value judgments. However, in another sense, estimating how future generations might assign various discount rates in their assessments is a source of uncertainty, for which only subjective opinions can be offered by this generation of assessors. Advising lead authors of the TAR of the need to make such aspects of uncertainty explicit is part of the purpose of this guidance paper.

Anderson, Judith L. 1998. Embracing uncertainty: The interface of Bayesian statistics and cognitive psychology. Conservation Ecology[online] 2(1): 2. Available from the Internet at URL: http://www.consecol.org/vol2/iss1/art2

Casman, E.A., Morgan, M..G. and Dowlatabadi, H. 1999. Mixed Levels of Uncertainty in Complex Policy Models. *Risk Analysis* 19(1): 33-42.

Edwards, A.W.F. 1992. Likelihood. Baltimore, MD: The Johns Hopkins University Press. (Originally published 1972, Cambridge, UK: Cambridge University Press.)

Gigerenzer, G. 1996. On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky. Psychological Review 103: 592-596.

Gigerenzer, G. 1994. Why the distinction between single event probabilities and frequencies is important for Psychology and vice-versa. In G. Wright and P. Ayton, eds. Subjective Probability. Chichester: Wiley.

Gigerenzer, G., Hoffrage, U. and Kleinbölting, H. 1991. Probabilistic mental models: A Brunswikian theory of confidence. Psychological Review 98: 506-528.

Henderson-Sellers, A. 1993. An Antipodean climate of uncertainty. *Climatic Change* 25: 203–224.

Howard, R..A., Matheson, J.E. and North, D.W. 1972. The Decision to Seed Hurricanes. *Science* 176: 1191-1202.

IPCC Guidelines for National Greenhouse Gas Inventories, pg 2

Intergovernmental Panel on Climatic Change (IPCC). 1996a. *Climate Change 1995. The Science of Climate Change: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change.* Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Kattenberg, A., and Maskell, K., eds. Cambridge: Cambridge University Press.

Intergovernmental Panel on Climatic Change (IPCC). 1996b. Climate Change 1995. Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change. R.T. Watson, M..C. Zinyowera, R.H. Moss (eds.). Cambridge: Cambridge University Press.

Jones, R.N. 2000. Managing uncertainty in climate change projections: Issues for impact assessment. Climatic Change 45(3-4): 403-419.

Juslin, P. 1994. The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. Organizational Behavior and Human Decision Processes 57: 226-246.

Kahneman, D. and Lovallo, D. 1993. Timid choices and bold forecasts. A cognitive perspective on risk taking. Management Science 39: 17-31.

Kahneman, D. and Tversky, A. 1979. Intuitive prediction: Biases and corrective procedures. Management Science 12: 313-327.

Kahneman, D., Slovic, P. and Tversky, A., eds. 1982. Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press.

McBean, G.A., Liss, P.S. and Schneider, S.H. 1996. Advancing our understanding. In J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, K. Maskell (eds.). Climate Change 1995. The Science of Climate Change: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 517-531.

McClelland, A.G.R. and Bolger, F. 1994. The calibration of subjective probabilities: Theories and models 1980-1994. In G. Wright P. and Ayton, eds. Subjective Probability. New York: Wiley.

Morgan, M.G. 1998. Uncertainty Analysis in Risk Assessment. *Human and Ecological Risk Assessment* 4(1): 25-39.

Morgan, M.G. and Dowlatabadi, H. 1996. Learning from Integrated Assessment of Climate Change, *Climatic Change* 34(3-4):337–368.

Morgan, M.G. and Henrion, M. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, Cambridge.

Morgan, M.G. and Keith, D.W. 1995. Subjective Judgments by Climate Experts. Environmental Science and Technology 29: 468A-476A.

Moss, R.H. and Schneider, S.H. 1997. Characterizing and Communicating Scientific Uncertainty: Building on the IPCC Second Assessment. In Elements of Change 1996, S.J. Hassol and J. Katzenberger, eds. (Aspen, CO: Aspen Global Change Institute), 90-135.

Moss, R.H.2000. Cost Estimation and Uncertainty. Pacific Asia Journal of Energy, forthcoming.

Nordhaus W D. 1994. Expert opinion on climatic change. American Scientist 82: 45-52.

Ravetz, J.R. 1986. Usable Knowledge, Usable Ignorance: Incomplete Science with Policy Implications. In Clark and Munn (eds.). Sustainable Development of the Biosphere, New York, Cambridge University Press, pp. 415-432.

Roughgarden, T. and Schneider, S.H. 1999. Climate change policy: quantifying uncertainties for damages and optimal carbon taxes. Energy Policy 27(7): 415-429.

Schneider, S.H. 1983. CO_2, Climate and Society: A brief overview. In Chen, R. S., Boulding, E. and Schneider, S. H. (eds.) *Social Science Research and Climate Change: In Interdisciplinary Appraisal*, D. Reidel, Boston, 9–15.

Schneider S.H. 1997. Integrated assessment modeling of global climate change: transparent rational tool for policy making or opaque screen hiding value-laden assumptions? Environmental Modeling and Assessment 2(4): 229-248.

Schneider, S.H., Kuntz-Duriseti, K. and Azar, C. 2000. Costing Nonlinearities, Surprises and Irreversible Events. Pacific Asia Journal of Energy, forthcoming.

Schneider, S.H., Turner, B.L., and Morehouse Garriga, H. 1998. Imaginable Surprise in Global Change Science. Journal of Risk Research 1(2): 165-185.

Titus, J. and Narayanan, V. 1996. The Risk of Sea Level Rise: A Delphic Monte Carlo Analysis in which Twenty Researchers Specify Subjective Probability Distributions for Model Coefficients within their Respective Areas of Expertise. Climatic Change 33 (2): 151-212.

Tversky, A. and Kahneman, D. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review 90: 293-315.

Tversky, A. and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and Biases. Science 185: 1124-1131.

Wright, G and Ayton, P. 1992. Judgmental Forecasting in the immediate and medium term. Organizational Behavior and Human Decision Processes 51: 344-363.