Effective Science Arbitration:

Some Lessons from Recent Scientific Assessments

Roger A. Pielke, Jr.

20 October 2006

Abstract

This paper focuses on three case studies in which scientific assessments failed to accurately portray relevant scientific literature. Assessments are a form of "intelligence" in decision processes and thus have well established criteria for evaluation. Specifically, this paper asks two questions in the three case studies: Does the assessment faithfully represent the existing literature? If not, what does the literature actually say? The first of these cases examines how the Intergovernmental Panel on Climate Change (IPCC) presented conclusions on the role of human-caused climate change in the trend of escalating disaster losses. The second case discusses the IPCC's treatment of sea level rise impacts related to storm surge. The third case discusses the treatment of seasonal hurricane forecasting in an assessment report published by the American Geophysical Union in the summer of 2006. The three cases discussed here provide the basis for suggesting some general guidance for the production of scientific assessments for decision makers. First is the importance of adhering to quality standards in assessment production. Second is the continuing importance of the peer reviewed literature on which an assessment is based. Finally, effective science arbitration would benefit from a clear

presentation of criteria of policy relevance which are used to judge the inclusion or presentation of particular scientific findings or conclusions to decision makers.

Introduction

Scientific assessments are a common mechanism for bringing knowledge to decision makers. For example, the periodic assessment reports of the Intergovernmental Panel on Climate Change (IPCC) have been among the most well-known assessments and are widely viewed as an authoritative perspective on climate science. But assessments also can be controversial. The IPCC faced criticism from a number of observers for its prominent featuring of the so-called "hockey stick" temperature reconstruction in its 2001 Summary for Policymakers. Such criticism led the U.S. Congress to request from the National Research Council an assessment of the IPCC's treatment of the "hockey stick." The subsequent NRC report led to a reinterpretation of the relevant science underlying the IPCC's claims and, to some degree, a reexamination of how information is conveyed to decision makers via assessments (see, e.g., NRC, 2006). Perhaps predictably, the NRC's findings were themselves contested.

This paper focuses on three case studies in which scientific assessments failed to accurately portray relevant scientific literature, two of which focus on the IPCC. None of the cases discussed in this paper have been as prominent or politically significant as the "hockey stick" controversy (to say the least), but together they do offer some lessons for the production of effective scientific assessments according to criteria developed from literature on science in policy and politics (e.g., McNie, 2007). These lessons may have broader relevance for the design of effective assessments in the decision making process. In particular, these lessons are intended to contribution to the reconciliation of the production of scientific knowledge and its use (Pielke, 2007).

This paper begins by justifying the simple criterion used subsequently to evaluate the three case studies based on the broader literature on evaluation of scientific assessments and knowledge in decision making more generally. It then presents the three case studies in each of which the actual practice of preparing an assessment failed to meet the standard set forth in the criterion. The paper's final section distills three overarching lessons from the cases, which unsurprisingly reinforce the broader literature on how to produce effective assessments in support of decision making. Perhaps ironically, the experiences related in the three cases suggest that assessors of assessments face their own challenges in providing useful knowledge to the producers of assessments. This paper focuses on the challenge of developing such useful guidance.

Criteria for Evaluation of Scientific Assessments

A scientific assessment refers to "the process of synthesizing, evaluating, and communicating scientific knowledge to inform a policy or decision process" (Dessler and Parson, 2005, p. 43; cf. Morgan et al. 2005; Mitchell et al. 2006). Dessler and Parson (p. 42) distinguish between positive and normative questions, defining positive questions as those focused on "scientific knowledge about the world" and normative questions involve "questions of our values, desires, and political principles." Dessler and Parson's conception of scientific assessment is conceptually equivalent to what Pielke (2007) describes as "science arbitration" (see also Lackey, 2007):

> The Science Arbiter seeks to focus on issues that can be resolved by science, which may originate in questions raised by decision makers or debate among

> decision makers. In practice, such questions are sent for adjudication to the scientist(s), who may be on an assessment panel or advisory committee, who then renders a judgment and returns to the policy makers scientific results, assessments or findings.

Pielke (2007) discusses other modes of interaction between scientists and policy makers that need not necessarily focus on positive questions, but which could also take the form of assessments. A broad literature supports the conclusion that separation of positive and normative questions is impossible in many settings (e.g., Jasanoff, 1990). The focus of this paper however is on assessments that seek, in principle if not in practice, to provide answers to positive questions, that is to serve the role of science arbitration.

For such assessments by what criteria might they be evaluated? Assessments are a form of "intelligence" in decision processes and policy scholars have recommend criteria which might be used to evaluate the effectiveness of intelligence. Clark (2002), for example, draws upon the work of Harold Lasswell to suggest the following criteria for the evaluation of an intelligence process: dependability (factual), comprehensiveness (complete), selectivity (targeted), creativeness (in finding facts), and availability (to everyone). Cash et al. (2002) focus on science assessments specifically and suggest that assessments should be credible, legitimate, and salient, where credibility refers to standards of scientific quality, legitimacy refers to political and procedural fairness, and salience refers to the relevance to a decision maker's needs. Dessler and Parson (2005, p. 44) suggest that scientific assessments "must be able to communicate clearly to a non-scientific audience without sacrificing scientific accuracy."

These various criteria share some obvious overlaps, and as a set go well beyond the focus of this paper which focuses only on a subset of the criteria of dependability or credibility (for a comprehensive review of the inter-relationship of the supply of knowledge and its use see McNie, 2007). Specifically, for each of the three case studies discussed in this paper two questions are asked:

- Does the assessment faithfully represent the existing literature?
- If not, what does the literature actually say?

The basis for selecting this subset of the relevant criteria for assessment evaluation is that the ability of assessments to offer accurate information to decision makers will be impossible if existing literature is not represented in a credible and dependable manner, threatening the entire intelligence process. If assessments do not meet this most straightforward criterion, then it will be unlikely to succeed with respect to broader criteria such as usefulness and legitimacy. Thus the faithful representation of existing peer-reviewed literature is a necessary but not sufficient condition for intelligence to effectively contribute to a functioning decision process (Clark, 2002; Cash et al., 2002; Sarewitz et al., 2000).

Three Case Studies

This section discusses three instances in which assessments failed to present accurate information with respect to the existing literature on which the assessment was based. The first two cases involve the Third Assessment Report of the Intergovernmental Panel on Climate Change, and specifically its report of Working Group II. The first of these cases examines the IPCC's conclusions on the role of human-caused climate change in the trend of escalating disaster losses. The second case discusses the IPCC's treatment of sea level rise impacts related to storm surge. The third case discusses a conclusion on seasonal hurricane forecasting in an assessment report on hurricanes published by the American Geophysical Union in the summer of 2006 in response to the devastating hurricane season of 2005 and hurricane Katrina in particular. The experiences documented in the three cases suggest some practical guidelines for the implementation of scientific assessments for decision makers.

Climate Change and Disaster Trends

Subsequent to the publication of the Third Assessment Report of the Intergovernmental Panel on Climate Change a debate took place about the possible role of human-caused climate change in the increasing economic toll of natural disasters (see Mills, 2005a; Pielke, 2005; Mills, 2005b). One important reason for this debate is a claim made by the IPCC Working Group II attributing some part of the trend of increasing disaster losses to changes in climate (McCarthy et al., 2001). Specifically, the IPCC concluded:

> Demographic and socioeconomic trends are increasing society's exposure to weather-related losses. Part of the observed upward trend in historical disaster losses is linked to socioeconomic factors such as population growth, increased wealth, and urbanization in vulnerable areas, and part is linked to climatic factors such as observed changes in precipitation, flooding, and drought events. Precise attribution is complex, and there are differences in the balance of these two causes

> by region and by type of event. Notably, the growth rate in the damage cost of non-weather-related and anthropogenic losses was one-third that of weather-related events for the period 1960-1999 (Munich Re, 2000). Many of the observed upward trends in weather-related losses are consistent with what would be expected under human-induced climate change.

However, upon closer examination, the claim of attribution was then unfounded and not justified by the literature (cf. Höppe and Pielke, 2006).

Specifically, the IPCC relied on a report published in 2000 by Munich Re that found that global disasters resulted in $636 billion in losses in the 1990s compared with $315 billion in the 1970s, after adjusting for changes in population and wealth (Munich Re, 2000). The Munich Re report concluded that disaster costs have increased by a factor of two (i.e., 636/315), independent of societal changes, and the IPCC subsequently concluded that anthropogenic climate change is therefore responsible for the difference. The IPCC's conclusions were then uncritically repeated in a range of subsequent publications (e.g., Mills, 2005b)

Methodologically, the IPCC's conclusion is suspect for a number of reasons. First, the Munich Re report provided neither methods nor data, and does not address the issue of attribution. Second, Munich Re admits that data on changes in wealth are not available around the world and changes in GDP are not always a good proxy for data on wealth. Third, Munich Re's data apparently includes weather and non-weather events (e.g., it appears to also include earthquake damages).

But assuming that all of the issues raised above can be overcome, and in the end there remains a 2-to-1 ratio between the decade, the fact is that the large decadal variability in disaster losses makes it quite difficult to assert a trend by comparing two different ten-year periods over a period of 30 years. This can be illustrated with an example from work on hurricane losses over time (i.e., based on Pielke and Landsea, 1998). When hurricane loss data are adjusted to account for trends in population, wealth, and inflation to 2005 values it allows for a comparison across decades (data can be seen in Figure 1). First, the ratio of the 1990s to the 1970s is quite similar to the Munich Re analysis, 2.1 ($87B/$41B). But other decadal comparisons picture looks quite different:

- 1990s to 1940s = 0.8 ($87B/$110B)
- 1990s to 1920s = 0.5 ($87B/$184B).

Thus, the 2000 Munich Re analysis, which provided some valuable insights on disasters to be sure, says nothing about the attribution of the causes for increasing disasters, yet its results were used by the IPCC WGII to suggest otherwise. This is particularly of note given that the IPCC WGI was unable to detect or attribute global trends in the occurrence of weather extremes (cf. Pielke, 2005).

A 2006 workshop took a close look at this issue and concluded (Höppe and Pielke, 2006):

> Because of issues related to data quality, the stochastic nature of extreme event impacts, length of time series, and various societal factors present in the disaster loss record, it is still not possible to determine the portion of the increase in

> damages that might be attributed to climate change due to GHG emissions. In the near future the quantitative link (attribution) of trends in storm and flood losses to climate changes related to GHG emissions is unlikely to be answered unequivocally.

The workshop also acknowledged that the IPCC anticipates that human-caused climate change will lead to an increased incidence and intensity of extreme events. If so, these will necessarily add to the ever-escalating toll of weather-events. However, in 2001 the IPCC TAR WGII reached a premature conclusion based on an insufficient examination of the relevant literature and reaching conclusions beyond what the literature could support. Rather than summarizing the relevant literature the IPCC offered its own hypothesis as a conclusion. This had the effect of potentially misleading decision makers, and also mischaracterizing the state of research on trends and causes of increasing disasters.

Storm Surge and Sea Level Rise

The IPCC Working Group II (WGII) Third Assessment Report's Summary for Policymakers (SPM), p. 13 includes the following sentence describing the potential future impacts of sea level.

> Model-based projections of the mean annual number of people who would be flooded by coastal storm surges increase several-fold (by 75–200 million, depending on the adaptive response) for mid-range scenarios of a 40-cm sea-level rise by the 2080s relative to scenarios with no sea-level rise (IPCC, 2001a).

The sentence has been widely quoted in the years since the Third Assessment. But a review of how it has been interpreted shows considerable confusion about what it actually means. For example:

- A peer-reviewed article in the British Medical Journal came to one conclusion (Patz and Kovatz, 2002):

  > The number of people at risk from flooding by coastal storm surges is projected to increase from the current 75 million to 200 million in a scenario of mid-range climate changes, in which a rise in the sea level of 40 cm is envisaged by the 2080s.

- A report by a group of scientists in an oil industry magazine came to a second interpretation (Cannell et al., 2001):

  > Even a somewhat conservative scenario of a 40-cm [15.8-in.] sea-level rise by the 2080s would add 75 to 200 million people to the number currently at risk of being flooded by coastal storm surges.

- The Red Cross offered still another interpretation (Pearce, 2001),

  > The average annual number of people whose houses are flooded by storm surges along coastlines is expected to increase from a few million each year to between 75 and 200 million by the year 2080, estimates the IPCC.

Such differing perspectives are characteristic of this particular conclusion of the IPCC. No one, it seems, knows what this sentence means. Now this might not be surprising; after all, the IPCC is written by committee and it would not be surprising to see an occasional confusing sentence come out of such a process. However, it appears that the IPCC did not explain what the sentence means in its full report, and going back to the original scientific source for the statement does not allow one to arrive at a clear interpretation of what it actually means.

The IPCC WGII SPM points to Section 4.5 of the full assessment report as the source for the sentence (IPCC, 2001b). In that section of the full report, one finds absolutely nothing related to the sentence. However, the trail does not end there. Chapter 7 of the IPCC includes this passage (IPCC, 2001c):

> Worldwide, depending on the degree of adaptive response, the number of people at risk from annual flooding as a result of a 40-cm sea-level rise and population increase in the coastal zone is expected to increase from today's level of 10 million to 22-29 million by the 2020s, 50-80 million by 2050s, and 88-241 million by the 2080s (Nicholls et al., 1999). Without sea-level rise, the numbers were projected at 22-23 million in the 2020s, 2732 million in the 2050s, and 13-36 million in the 2080s. The 40 cm sea-level rise is consistent with the middle of the range currently being projected for 2100 by Working Group I. In 2050, more than 70% (90% by the 2080s) of people in settlements that potentially would be flooded by sea-level rise are likely to be located in a few regions: west Africa, east Africa, the southern Mediterranean, south Asia, and southeast Asia. In terms

> of relative increase, however, some of the biggest impacts are in the small island states (Nicholls et al., 1999).

However, nowhere in this passage is it clear where the SPM arrived at the 75-200 million range, or what it might mean. But the passage does point the reader to Nicholls et al. (1999). A close reading of Nicholls et al. (1999) lead to its Table 7 which is reproduced as Figure 2.

One possible interpretation of where the IPCC arrived at its figures is that the IPCC confused the cumulative people affected by a 1 in 1000 year storm surge event (People to Respond or PTR in the Table 7) with the "mean annual number of people who would be flooded by coastal storm" (AAPF in the Table 7) because these numbers -- 70-205 million – are very close to those in the SPM (75-200 million) and it would perhaps be easy to confuse the last digits. If so then this would be a huge error, conflating a cumulative number with an annual number. These values are highlighted in red in Figure 2.

Robert Nicholls, the lead author of the original study, graciously responded to an email query suggesting that there was another possible interpretation that got one reasonably close to the IPCC numbers (Nicholls, 2006). That possibility would involve taking the Average Annual People flooded under "Evolving Protection" for the 2080s – 88 million – and subtracting the 13 million for the 2080s under evolving protection with no sea level rise, leaving one with 75 million (these values are highlighted in blue in Figure 2). However, there is no parallel calculation that leads to the value of 200 million.

The analogous calculation under the "Constant Protection" column is 228-36 = 192 million. One could confuse the different models to arrive at 237(+/-4)-36 = 205 or 197 (the midpoint gives 201), but this would be an obvious methodological error (comparing across different models), and then compounded with a typo, rounding, or other means to arrive at 200 (these values are highlighted in green in Figure 2).

Where exactly the numbers came from remains a mystery. The inscrutable sentence matters for several reasons. First, it would have violated the procedures of the IPCC for the numbers to have been generated through some procedure outside the peer-reviewed literature. And if the number simply represents a mistake, it could have consequences. When I first searched for information about the sentence in February, 2006 I found that it had been used in official documents related to climate policy by the Japanese[1], Canadian (Environment Canada, 2006), and British (Environment Agency, 2006) governments. Because the IPCC SPM is used to justify policy actions on climate change, is important that all of its information be accurate and understandable. In this case, it appears that neither criterion was met.

The IPCC has great potential to inform policy makers, however its credibility rest on being accurate and faithful to the literature. Errors will inevitably occur, but in this case an error in one of the IPCC's most important summaries was used uncritically in policy documents and academic studies, its imprecision or incorrect reflection of the literature on which it was based apparently not having been noticed.

[1] http://www.env.go.jp/en/global/cc/050901.pdf

An accurate summary of Nicholls et al. (1999) than found in the IPCC TAR would in fact leads to conclusions at odds with that distilled by those who have misinterpreted the IPCC sentence. In fact, Nicholls et al. provide considerable support for the argument that increasing coastal habitation is by far more important than sea level rise when it comes to the factors responsible for projected storm surge impacts. A more accurate summary of Nicholls et al. (1999) would start by observing that the paper used a scenario that had that number of people subjected to 1 in 1000 or greater risk of coastal storm surges increasing from 197 million in 1990 to 575 million in the 2080s without any rise in sea level, due only to population growth and demographic changes. Under this scenarios, with sea level rise (and no societal reaction) the 2085 population-at-risk would be 630-640 million. With no sea level rise Nicholls et al. (1999) calculate an increase in the average annual number of people flooded rising from 10 million (1990) to 36 million (2085) with 1990 levels of protection. With protection that evolves as a function of GDP (but not otherwise changed due to sea level rise) the number of people flooded increases only by 3 million, from 10 million (1990) to 13 million (2085), even though population-at-risk increased by 378 million people. This suggests that only 0.8% of the additional people who inhabit the coastal zone by 2085 with a 1 in 1000 or greater risk would on average experience an annual flood. This compares with 5.1% of inhabitants who experience a flood in 1990 based on the assumptions of Nicholls et al.

Putting this together, according to Nicholls et al. (1999) population-at-risk increases by 438 million by 2085 while the average annual number of people flooded increases by 78 million people, from 10 million in 1990 to 88 million in 2085 (and by 3 million, from 10 million to 13 million under the no sea level rise scenario). So the IPCC should have reported something like the following:

> Nicholls et al. estimate that of the 438 million additional people expected to inhabit the coastal zone by 2085 with a 1 in 1000 or greater rise of storm surge, about 0.7% - 18% of them, or on average between 3 million and 78 million additional people annually would experience the effects of coast flooding under the assumptions of the sensitivity analysis.

The difference between assuming that only 0.7% of the additional people who inhabit the coastal zone at risk to flooding under a scenario of no sea level rise versus 18% under a scenario of sea level rise results from the fact that the methods assumes an appropriate adaptive response in the first case but not the latter. For instance, to reach the 18% portion requires than many people would move to locations far riskier than they choose to inhabit today. Using the same ratio of people who experience floods to those at risk (i.e., assuming that people adapt where they live according to the contemporaneous risk) as Nicholls et al. use for 1990 (i.e., 5.1%) would suggest an additional 22.3 million people would be exposed to floods with a population increase of 438 million. Using the ratio suggested for 2085 under no sea level rise results in an additional 3.5 million people who experience floods. Nicholls et al. provide no reason to expect that society will be significantly more risk averse in 2085 than today, nor why the presence of sea level rise would change the acceptable risk from a scenario of no sea level rise. The IPCC thus contributed to the reaching of conclusions diametrically at odds with what the original literature concluded.

Skill of Seasonal Hurricane Forecasting

The American Geophysical Union (AGU) released an assessment report in June, 2006 titled "Hurricanes and the US Gulf Coast" which was the result of a "Conference of Experts" held in January, 2006 (AGU, 2006). One aspect of the report illustrates why it is so important to have such assessments carefully balanced with participants holding a diversity of legitimate scientific perspectives. When such diversity is not present, it increases the risks of misleading or false science being presented as definitive or settled, which can be particularly problematic for an effort intended to be "a coordinated effort to integrate science into the decision-making processes."

The AGU Report includes the following claim:

> There currently is insufficient skill in empirical predictions of the number and intensity of storms in the forthcoming hurricane season. Predictions by statistical methods that are widely distributed also show little skill, being more often wrong than right.

Such seasonal predictions are issued by a number of groups around the world, and are also an official product of the U.S. government's Climate Prediction Center. If these groups were indeed publishing forecasts "more often wrong than right" or demonstrating "no skill" in the jargon of forecast evaluation, then there would be good reason to ask them to cease immediately and get back to research, lest they mislead the public and decision makers.

The claim by the AGU is incorrect, or at a minimum, is a minority view among the relevant expert community. According to groups responsible for providing seasonal forecasts of hurricane activity, their products do indeed have skill, which is defined as the relative improvement of a forecast over some naïve baseline. Consider the following perspectives that are contrary to that presented in the AGU assessment:

- Tropical Storm Risk, led by Mark Saunders, finds that their (and other) forecasts of 2004 and 2005 demonstrated excellent skill according to a number of metrics (Lea and Saunders, 2006; Saunders and Lea, 2005).

- Phil Klotzbach (2006) of Colorado State University, now responsible for issuing the forecasts of the William Gray research team, writes in an email[2]:

  > All three of our monthly forecasts have shown skill with respect to the previous five-year monthly mean of NTC using MSE (mean-squared error as our skill metric). Here are our skills (% value is the % improvement over the previous five-year mean):
  >
  > - August Monthly Forecast: 38%
  > - September Monthly Forecast: 2%
  > - October Monthly Forecast: 33%

[2] Klotzbach also provided a link to the original data used to calculate skill: http://tropical.atmos.colostate.edu/Includes/Documents/Publications/seasonalskill.xls

- NOAA's Chris Landsea provided two figures (Figures 3a and b) which show NOAA's seasonal forecast performance. He commented of NOAA's forecast skill:

  > You can see that we do okay in May (4 out of 7 seasons correctly forecasting number of hurricanes for example), but better in August (6 of 8 seasons correct). (Landsea, 2006)

  Landsea also pointed to a peer reviewed paper evaluating seasonal hurricane forecast skill (Owens and Landsea, 2003).

From this information provided by three different seasonal forecasting teams, it is unambiguous that there exist strong arguments in support of the skill of seasonal hurricane forecasts and relevant analyses in the peer-reviewed literature. None of these arguments or references appeared in the AGU report. Because the AGU assessment did not acknowledge the diversity of perspectives on the skill of seasonal forecasts, it did not accurately reflect the comprehensive, literature-based perspective that one might expect to find in an assessment report.

The AGU provided no references to analyses or scientific literature to support is claim of the lack of skill in seasonal forecasting. This left the AGU open to a wide range of perceptions about the basis for including an unsupported claim at odds with the existing community consensus. In particular, by not providing a scientific basis for its claim the AGU left itself vulnerable to the perception -- whatever the reality -- that the issue of seasonal hurricane forecasting has gotten caught up in the "climate wars" over hurricanes and global warming. William Gray of Colorado State University is the originator of seasonal hurricane forecasts and has loudly dismissed the notion of human-caused global warming, much less a connection to hurricanes. One of the lead authors of the AGU assessment is a former student of Bill Gray's who happens also to be a strong advocate of a human role in recent hurricane activity. In the time period of the drafting of the AGU report he was in a nasty public feud with Bill Gray (Bauerlein, 2006). One month after the AGU Assessment was released, the alleged lack of skill in seasonal hurricane forecasts was cited in Congressional testimony as evidence of global warming influence on hurricanes, citing unpublished work by the AGU contributor as the basis for the claim (Curry, 2006). Thus, by not rigorously adhering to the existing scientific literature and diversity of views within the community, the AGU left itself open to charges that it had made a serious error, a perception that some mischief had occurred in its assessment process, or both.

Lessons for Effective Science Arbitration

The three cases discussed here provide the basis for suggesting some general guidance for the production of scientific assessments for decision makers. First is the importance of adhering to quality standards in assessment production. In each of the three cases arguably existing quality standards were violated. Second is the continuing importance of the peer reviewed literature on which an assessment is based. The original literature is important both for the process of assessment, but also the interpretation of assessments subsequent to their production. Finally, effective science arbitration would benefit from a clear presentation of criteria of policy relevance which are used to judge the inclusion or presentation of particular scientific findings or conclusions to decision makers.

Adherence to quality standards

The three cases discussed here were not selected through some random procedure, but happened to be instances in which I observed problems in the assessment process while doing research. Thus it is difficult to assess how widespread the issues discussed here might be in the assessment literature. However broad the problem is, as the IPCC prepares to publish its fourth assessment report, and scientists and policy makers continue to emphasize the importance of assessments, it seems critical to carefully evaluate procedures for accuracy, and for users of assessments to understand the strengths and limits of assessments. Ultimately, the distilling of complex, nuanced research into one-sentence sound bites that perhaps inevitably cannot accurately capture what is to be found in a lengthy scientific article.

Adhering to clear standards of quality control can militate against perceptions that extra-scientific factors play a role in the production of an assessment report. There are always strong incentives for such factors to play a role, and effective science arbitration necessitates a strict focus on positive questions (Pielke, 2007). Scientists typically bristle at the notion that extra-scientific factors like personal feuds might play a role in their actions, whether or not such perceptions are well-founded. Extra-scientific factors may or may not have played a role in the case of the AGU report, only the actors know for certain. But knowing motivation matters little from the standpoint of the effects of the misleading information contained in the AGU assessment for both science and policy. When claims far outside the bounds of the existing body of scientific knowledge are made in an assessment seeking to reflect a community consensus for policy makers, the alternative to mischief in the assessment process is simply that an egregious mistake was made. Neither prospect is particularly appealing to consumers of assessments. As was the case with the IPCC's conclusions about disasters and climate change, erroneous conclusions make take years to correct in the literature.

In all three cases, the best way to avoid errors creeping into efforts to arbitrate science is for a rigorous adherence to established standards or quality control. The IPCC (2003) provides clear guidance on how to support scientific claims in an assessment report:

> Contributions should be supported as far as possible with references from the peer-reviewed and internationally available literature, and with copies of any unpublished material cited. Clear indications of how to access the latter should be included in the contributions.

In the case of disaster losses the IPCC cited a source for the claims being made but that source proved to be a dead end, no further information was available. The IPCC clearly went beyond what the literature could support. In the case of the AGU report no reference was made to support the scientific claim advanced in the report. In addition, the report referenced three papers as being "submitted" (i.e., not yet having gone through peer review) and with no indication for how those works might be obtained. Assessment reports are important to policy makers as the case of sea level rise and the IPCC indicates. In addition, assessment report are often cited in subsequent peer-reviewed literature, under a presumption that they have been rigorously peer-reviewed, and the claims made in assessments gain a further standing in the literature. In the case of sea level rise the error has propagated so thoroughly that correcting it may be impossible and it may simply have to be superseded by more recent work.

It is certainly the case that the AGU assessment was produced on nothing like the scale and effort of the IPCC reports. Nonetheless, quality standards in assessments should be upheld in all contexts where science is being brought to the attention of policy makers as relevant to their decisions. The experiences in the IPCC discussed here suggest that scale and scope are no guarantee of the upholding of existing quality standards.

Importance of primary literature

Each of the three cases discussed in this paper reinforce the continuing importance of the conventional peer-reviewed literature. In the case weather disasters the IPCC reached a major conclusion about the influence of anthropogenic climate change on disaster losses based on a very simplistic and misleading analysis. Such a claim would all but certainly not have survived a conventional peer review process. More recent research reinforces the challenges faced in detecting and attributing trends in disaster losses (Höppe and Pielke, 2006). Arguably, the IPCC's reliance on an analysis which did not support its conclusion has led to frequent misinterpretations of significance of science of disaster losses for policy options (compare the exchange between Mills, 2005a; Pielke, 2005; Mills, 2005b).

In the case of the IPCC's claims about the projected impacts of sea level rise, at some point in the process the IPCC's claims became untethered from what was reported in the original peer-reviewed study. This led to a claim that was either ambiguous, and thus was interpreted in a range of different ways, or misinterpreted in ways contrary to that reported in the original study. While assessments can serve as a useful "shortcut" to researchers, particularly for areas outside their direct expertise, it is appropriate for researchers to continue to rely on original literature in their scientific work, rather than to simply depend on assessments as accurate means to convey scientific findings. Inevitably, assessments must simplify, in the process losing much of the nuance and uncertainties that characterize any complex scientific study.

For example, the details of Nicholls et al. (1999) on storm surge are very complicated and to understand them requires a careful examination of the primary literature. Clearly, what appeared in the 2001 IPCC bears only a distant relation to what the original study actually said. For this reason, scientific assessments cannot replace the primary literature, and some thought should be given by scholars to how best to deal with knowledge that is highly simplified through assessment and then recirculated into academic inquiry in subsequent peer-reviewed literature. Like the children's game of telephone, this is a recipe for miscommunication, mischaracterization of scientific research, and a foundation of knowledge that rests a few feet above the ground. A more appropriate role for assessments would be to focus more explicitly on the information needs of policy makers and focus attention on a wide range policy options and their possible consequences. Efforts to summarize complex science may not ultimately prove useful to policymakers if they result in oversimplifications and mischaracterizations.

Provenance of policy relevance

Both the IPCC and AGU assessments discussed here were prepared to inform decision making by policy makers. But neither assessment provides information for understanding what criteria of policy relevance it used to select among existing scientific literature for inclusion in its report. Arguably, many of the problems with assessments found here could be avoided if assessments answered a straightforward question with respect to the information that it presents – "So what?"

Asking an assessment to distill the potential relevance for action, or at a minimum to specify criteria of policy relevance, would not necessarily require abandonment of a focus on positive questions. An assessment built upon questions provided by policymakers would create a close tie between the information demanded by decision makers and that being produced in assessments (cf. McNie, 2007). Neither the IPCC nor the AGU have such an explicit connection to the information needs of decision makers, leaving one to guess or speculate why it is that information on disaster losses, sea level rise, or seasonal hurricane forecasting is thought to be important enough to highlight via assessment.

A more explicit tie to the information needs of decision makers has been central to studies of the design of effective assessments. Cash et al. (2002) conclude that:

> efforts to connect knowledge to action are effective only if they are sufficiently salient, credible, AND legitimate with multiple audiences simultaneously. In other words, such efforts are often undermined by perceived deficiencies in a single attribute – thus the danger of only focusing on credibility at the expense of salience or legitimacy.

In each of the three cases discussed here, shortfalls in credibility have potential threaten the assessment legitimacy. And both credibility and legitimacy could be enhanced through a more explicit focus on assessment salience, which was lacking in all three instances.

Conclusion

Scientific assessments have great potential to serve as a useful resource to decision makers. However, the cases discussed here suggest that the practice of designing and implementing assessments has some ways yet to go with respect to the understandings of criteria for assessment success found in the academic literature. Shortfalls in assessment performance lead one to the somewhat ironic conclusion that assessors of assessments also face some challenges in sharing the knowledge that they have gleaned with producers of assessments.

This paper suggests three straightforward lessons that might help to better reconcile knowledge of effective assessment practice with the practice assessment implementation. Assessments should adhere to existing standards of quality control. Such standards exist for a reason and when violated problems ensue. The scientific literature, although imperfect, remains the basis for assessments, and assessments do not substitute for that literature. Such original literature remains important even subsequent to the production of an assessment. Over-reliance on assessments, rather than original literature, can allow errors to propagate in the literature and in policy discussions, much longer than might be the case with a focus on original studies. Above all, assessments seeking to provide useful information to decision makers should engage decision makers in the assessment process. By simply asking the question – “so what?” – of any particular scientific conclusion, some of the error, ambiguity, and potential mischief in an assessment process might be avoided. At a minimum a more explicit focus on the significance of assessed knowledge for action would enhance the prospects for effective science arbitration.

Acknowledgements

Useful comments, inputs, and suggestions were provided on earlier versions of arguments presented in this paper by Daniel Collins, Phil Klotzbach, Chris Landsea, Richard J. T. Klein, and Robert Nicholls. Ami Nacu-Schmidt expertly assisted with the preparation of the manuscript. All responsibility for content lies with the author. This material is based upon work supported by the National Science Foundation under Grant No. 0345604. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Figure 1. Historical hurricane losses adjusted to 2005 values. Source: R. Pielke, Jr. (work in progress)

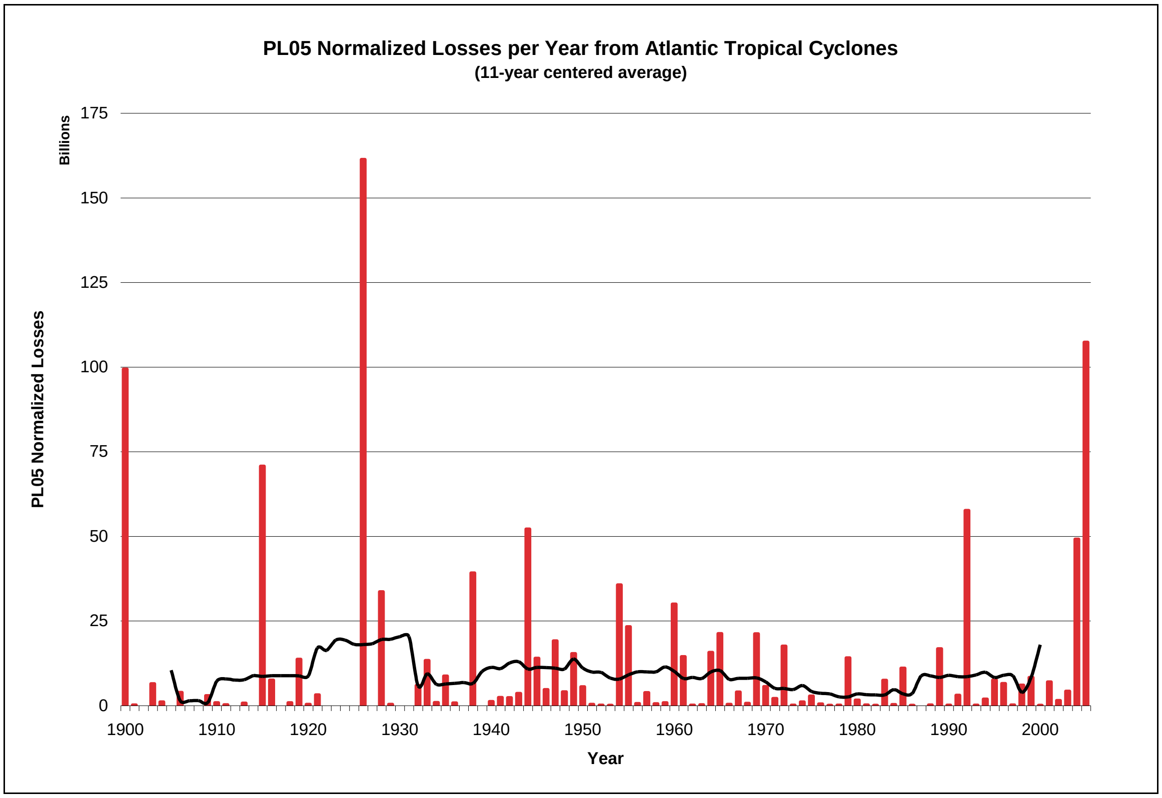


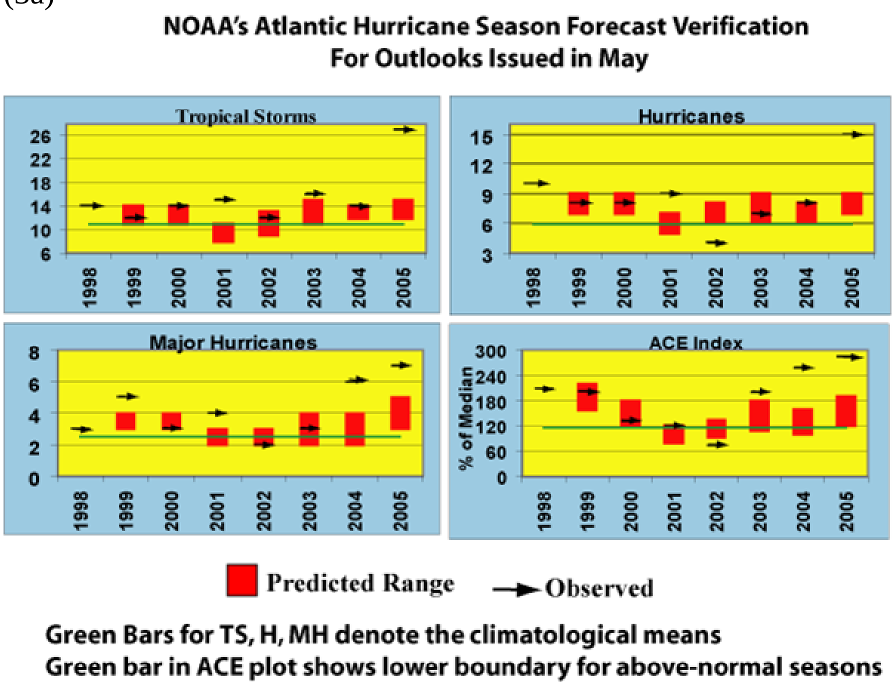
Figure 2. Reproduction of Table 7 from Nicholls et al. (1999).

Table 7
People in the hazard zone (PHZ), average annual people flooded (AAPF) and cumulative people to respond (PTR) for the different sea-level rise and protection scenarios. The results for the HadCM2 ensemble include 95% confidence intervals.

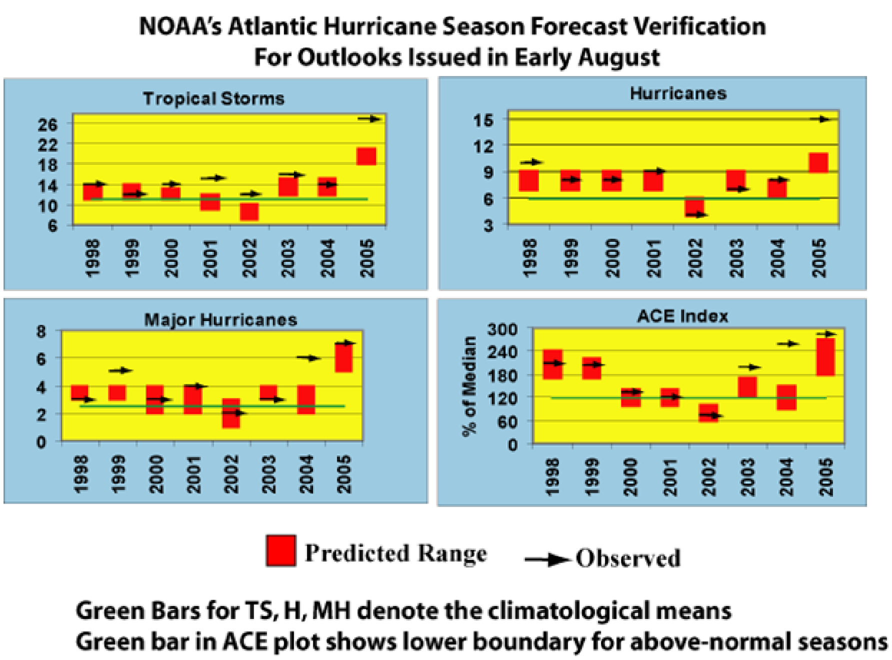
Scenario	Time	PHZ	AAPF	PTR	AAPF	PTR
			Constant protection		Evolving protection	
		(In millions of people)				
Reference (no global sea-level rise)	1990	197	10	0	10	0
	2020s	399	23	0	22	0
	2050s	511	32	0	27	0
	2080s	575	36	0	13	0
HadCM2 GGa ensemble (mean ± 95% confidence interval)	2020s	410	30	5	26	3
	2050s	542 ± 1	79 ± 1	46 ± 1	51 ± 1	20 ± 1
	2080s	636 ± 1	237 ± 4	205 ± 4	93 ± 2	70 ± 2
HadCM3 GGa	2020s	409	29	5	26	2
	2050s	542	78	45	50	20
	2080s	634	228	195	88	65

Figures 3a and b. NOAA's official annual seasonal hurricane forecasts 1998-2005, comparison of predicted and observed activity. ACE refers to "Accumulated Cyclone Energy" which is a measure of intensity. Figures courtesy of C. Landsea, NOAA/NHC.

(3a)



(3b)



References

American Geophysical Union (AGU). 2006. Hurricanes and the U.S. Gulf Coast: Science and Sustainable Rebuilding. 19 June. http://www.agu.org/report/hurricanes/.

Bauerlein. V. 2006. Hurricane debate shatters civility of weather science, Worsened by global warming? Spats are so tempestuous, sides are barely talking, Charge of 'brain fossilization,' The Wall Street Journal. (2 February): A1.

Cannell. M., Filas. J., Harries. J., Jenkins. G., Parry. M., Rutter. P., Sommeland. L. and Walker. J. 2001. Global warming and the E&P industry. Oilfield Review. (Autumn): 44-59. http://www.slb.com/media/services/resources/oilfieldreview/ors01/aut01/global.pdf.

Cash. D., Clark. W. C., Alcock. F., Dickson. N., Eckley. N. and Jäger. J. 2002. Salience, credibility, legitimacy and boundaries: Linking research, assessment and decision making. KSG Working Papers Series RWP02-046. (November). http://ssrn.com/abstract=372280.

Clark, T. 2002. The Policy Process: A Practical Guide for Natural Resources Professionals. New Haven, CT: Yale University Press, New Haven, CT. 215 pp.

Curry, J. 2006. Global warming and hurricanes. Testimony presented to the House Committee on Government Reform. 20 July. http://reform.house.gov/UploadedFiles/GT%20-%20Curry%20Testimony.pdf.

Dessler, A. E. and Parson, E. A. 2006. Science and Politics of Global Climate Change: A Guide to the Debate. New York: Cambridge University Press. 190 pp.

Environment Agency. 2006. Your environment: Melting ice and sea level rise. http://www.environment-agency.gov.uk/yourenv/639312/641094/642206/642371/642369/?version=1&lang=_e.

Environment Canada. 2006. CEPA Environmental Registry: CEPA Review. http://www.ec.gc.ca/CEPARegistry/review/default.cfm.

Höppe, P. and Pielke, Jr., R. A. 2006. Report of the Workshop on Climate Change and Disaster Losses: Understanding and Attributing Trends and Projections. Center for Science and Technology Policy Research. Boulder: University of Colorado. http://sciencepolicy.colorado.edu/sparc/research/projects/extreme_events/munich_workshop/full_workshop_report.pdf.

Intergovernmental Panel on Climate Change (IPCC). 2001a. Summary for Policymakers. Climate Change 2001: Impacts, Adaptation, and Vulnerability. Working Group II. http://www.ipcc.ch/pub/wg2SPMfinal.pdf.

Intergovernmental Panel on Climate Change (IPCC). 2001b. Climate Change 2001: Working Group II: Impacts, Adaptation and Vulnerability. Chapter 4: Hydrology and Water Resources. http://www.grida.no/climate/ipcc_tar/wg2/179.htm.

Intergovernmental Panel on Climate Change (IPCC). 2001c. Climate Change 2001: Working Group II: Impacts, Adaptation and Vulnerability. Chapter 7: Human Settlements, Energy, and Industry. http://www.grida.no/climate/ipcc_tar/wg2/310.htm#72214.

Intergovernmental Panel on Climate Change (IPCC). 2003. Procedures for the preparation, review, acceptance, adoption, approval and publication of IPCC reports. http://www.climatescience.gov/Library/ipcc/app-a.pdf.

Jasanoff, S. 1990. The Fifth Branch: Science Advisers as Policymakers. Cambridge, MA: Harvard University Press. 302 pp.

Klotzbach, P. 2006. email correspondence with author. 23 June.

Lackey. R. T. 2007. Science, scientists, and policy advocacy. Conservation Biology. (21): accepted.

Landsea, C., 2006. email correspondence with author. 21 June.

Lea, A. S. and Saunders, M. A. 2006. How well forecast were the 2004 and 2005 Atlantic and U.S. hurricane Seasons? In: Proceedings of the 27th Conference on Hurricanes and Tropical Meteorology. Monterey, USA: 24-28 April. http://tsr.mssl.ucl.ac.uk/docs/AMS2006Forecast.pdf.

McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. and White, K. S. (eds.). 2001. Climate Change 2001: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge, UK: Cambridge University Press. 1032 pp.

McNie. E. 2007. Reconciling the supply of scientific information with user demands: An analysis of the problem and review of the literature. Environmental Science & Policy. 10(1): in press.

Mills. E. 2005a. Insurance in a climate of change. Science. 309(5737):1040-1044.

Mills. E. 2005b. Response. Science. 310(5754):1615-1616.

Mitchell, R. B., Clark, W. C., Cash, D. W. and Dickson, N. M., (eds.). 2006. Global Environmental Assessments: Information and Influence. Cambridge, MA: MIT Press. 344 pp.

Morgan. M. G., Cantor. R., Clark. W.C., Fisher. A., Jacoby. H. D., Janetos. A. C., Kinzig. A. P., Melillo. J., Street. R. B. and Wilbanks. T. 2005. Learning from the U.S. national assessment of climate change impacts. Environ. Sci. Technol. 39(23): 9023-9032.

Munich Re. 2000. Topics, 2000: Natural Catastrophes, The Current Position. Munich Reinsurance. Munich, Germany. http://www.munichre.com/publications/302-02354_en.pdf?rdm=80335.

National Research Council (NRC). 2006. Surface temperature reconstructions for the last 2,000 years. Washington, DC: National Academies Press. http://fermat.nap.edu/catalog/11676.html.

Nicholls, R. J. 2006. email correspondence with author. 5 May.

Nicholls. J. N., Hoozemans. F. M. J. and Marchand. M. 1999. Increasing flood risk and wetland losses due to global sea-level rise: regional and global analyses. Global Environmental Change. 9(Supplement 1): 69-87.

Owens. B. F. and Landsea. C. W. 2003. Assessing the skill of operational Atlantic seasonal tropical cyclone forecasts. Weather and Forecasting. 18(1): 45-54.

Patz. J. A. and Kovats. R. S. 2002. Hotspots in climate change and human health. British Medical Journal. 325(73727372): 1094-1098.

Pearce. F. 2001. Heating up! Red Cross and Red Crescent: The Magazine of the International Red Cross and Red Crescent Movement. 2: http://www.redcross.int/EN/mag/magazine2001_2/heating.html.

Pielke. Jr. R. A. 2005. Attribution of disaster losses. Science. 310(5760): 1615.

Pielke, Jr., R. A. 2007. The Honest Broker: Making Sense of Science in Policy and Politics. Cambridge, UK: Cambridge University Press. in press.

Pielke. Jr. R. A. and Landsea. C. W. 1998. Normalized hurricane damages in the United States: 1925-95. Weather and Forecasting. 13 (September): 621-631.

Sarewitz, D., Pielke, Jr., R. A. and Byerly, Jr., R. (eds.). 2000. Prediction: Science, Decision Making and the Future of Mature. Washington, DC: Island Press. 405 pp.

Saunders. M. A. and Lea. A. S. 2005. Seasonal prediction of hurricane activity reaching the coast of the United States. Nature. 434(21 April): 1005-1008.