

# A protocol to articulate and quantify uncertainties in climate change detection and attribution

James S. Risbey<sup>1,\*</sup>, Milind Kandlikar<sup>1</sup>, David J. Karoly<sup>2</sup>

<sup>1</sup>Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

<sup>2</sup>CRC for Southern Hemisphere Meteorology, Monash University, Clayton 3168, Australia

**ABSTRACT:** This work develops and describes a formal probabilistic protocol via which the process of identifying lines of evidence for climate change, assessing likely causes for changes in the evidence, and combining the lines of evidence to make overall attributions of cause to greenhouse gases can be made. This open and detailed model of the detection and attribution process is designed to identify issues at stake in detection and attribution, and to facilitate scrutiny and understanding on this contentious issue in broader communities. The protocol provides a convenient means to make each of the judgements in this issue explicit. These judgements are characterized via expert elicitation techniques in both quantitative and qualitative form. The protocol focuses on detecting climate change and attributing causes to the enhanced greenhouse effect rather than on more general anthropogenic change, because the former is more consequential. Major uncertainties identified in the protocol relate to characterization of natural variability for each line of evidence, non-greenhouse forcings, and the climate response to forcing. The relative roles of uncertainty in climate sensitivity and climate forcings are still unclear in making determinations of attribution. Measures of attribution need to account for both the amount of signal explained by a postulated cause as well as its associated probability. When combining lines of evidence to form overall measures of attribution, the level of dependence assumed among lines of evidence is critical. Finally, the protocol highlights the need to reevaluate standards of evidence for attributing greenhouse climate change.

**KEY WORDS:** Detection · Attribution · Climate change · Expert judgement

*Resale or republication not permitted without written consent of the publisher*

## 1. INTRODUCTION

The issue of whether, and to what extent, human-induced increases in greenhouse gases (primarily carbon dioxide) have caused climate changes is contentious. Considerable effort has been put into detecting climate change from amidst the background fluctuations of natural variability. This is the 'detection' problem for climate change, which seeks to determine if the changes observed this century are unusual or not. If they are simply natural fluctuations of the climate system, one might not expect them to be as persistent as compared to the case when some systematic

forcing is driving the change. If the changes are unusual, one would like to know what is causing them, for this will also have implications for their persistence. If the cause is related to something that is fluctuating naturally, such as the sun, it will have different implications than if the cause is increased concentrations of greenhouse gases. In the latter case, the changes could be expected to continue in response to increases in greenhouse gas concentrations. The problem of attributing causal blame for any unusual climate changes is known as the 'attribution' problem.

The chapter dealing with detection and attribution issues (Santer et al. 1996a) in the Intergovernmental Panel on Climate Change (IPCC) second assessment report (IPCC 1996) was one of the major foci of public attention when the report was released. Controversy

\*E-mail: ris@cmu.edu

arose in part over whether the science was sufficiently well developed to justify a conclusion that humans have begun to have a discernible influence on the climate system (Edwards & Schneider 1997). Controversies such as this one have arisen because of political interest in the issue of climate change, because of the open-ended nature of the science itself, and because of difficulties of communicating uncertainties, both within and outside the immediate climatological community.

Past efforts to describe and communicate uncertainty in climate change knowledge have tended to resist quantification and the temptation to be too specific.<sup>1</sup> But this has left outsiders with few concrete insights into the nature of the difficulties. This has also led some to perceive a lack of transparency in the process by which scientific consensus on climate change is reached. As a result, some critics have made claims that the processes of generating the uncertainty estimates are biased and misleading (e.g. Seitz 1996). In any event, the process of communicating uncertainties on issues such as detection and attribution of climate change would benefit from more transparent and formal methods. To this end, we have developed a formal uncertainty protocol in which each of the steps and assumptions required to make judgements on detection and attribution is made as explicit as possible. This paper presents the rationale for such a protocol, along with an outline and a discussion of its utility and limitations. Some preliminary results from use of the protocol are reported, but only for the purpose of clarifying its structure and function.

### 1.1. Evolution of detection and attribution studies

Formal studies of detection and attribution of climate change go back to at least the early 1980s (Clark 1982). The early studies of detection and attribution focused on trends of globally averaged mean surface air temperature (e.g. Wigley & Raper 1990). This is not surprising because global mean temperature change is a good single aggregate indicator of the climate system, and sufficient data exists to monitor it for over a century now (Jones 1994). More recent efforts have expanded their focus to include other data sets and the use of new methodologies based on general circulation models (GCMs), such as fingerprinting or pattern correlation approaches (Karoly et al. 1994, 1999, Santer et al. 1995, 1996b, Hegerl et al. 1997). These approaches have introduced additional metrics for climate change beyond the global mean temperature record. Most

have tended to focus on other aspects of the 3-dimensional temperature structure of the atmosphere, for example, the surface geographical (Santer et al. 1995) or vertical patterns (Santer et al. 1996b) of temperature change. Some have focused on temporal features of the temperature record such as the annual cycle (Karoly et al. 1999) and autocorrelation structure (Wigley et al. 1998). Recent summaries of detection and attribution studies are given by Barnett et al. (1999) and Mitchell & Karoly (2000).

The different lines of evidence employed include different physical quantities, have varying spatial and temporal coverage and resulting uncertainties, and use data from a variety of sources—from global measurement networks to data synthesized from GCMs. They pose new methodological challenges of finding systematic ways for combining their influence in order to evaluate the hypothesis that human activities, and not other climate forcing such as changes in solar intensity or internal variability generated by ocean-atmospheric interactions, are causing the observed changes in global climate.

When considering multiple lines of evidence, scientists have tended to make implicit and holistic assessments of their subjective confidence judgements in the overall likelihood of detection and attribution of climate change. In the 1995 IPCC second assessment report, the detection and attribution chapter reviewed the evidence for climate change from a number of sources, and in summarizing reached the conclusion that ‘the balance of evidence suggests that there is a discernible human influence on global climate’. Though consensual, this conclusion was tentative because it was the first major effort of its kind, and was based on a rather informal process for weighing the influence of the different lines of evidence. There have been few formal studies where the process of making such inferences is made explicit. An important next step is to develop formal models for making assessments of detection and attribution, where uncertain knowledge is represented in a coherent probabilistic manner.

### 1.2. Why develop formal methods to represent uncertainty?

The motivation to develop formal methods to represent uncertainty in the detection and attribution process is derived from both the scientific arena and the policy arena, and in particular from the intersection of the two. The scientific motivation is derived from the need described above for methods that can cope with multiple lines of evidence, their uncertainties, and associated inferences. As a public policy issue, climate

<sup>1</sup>There are recent exceptions to this (e.g. Varis & Kuikka 1997, Kheshgi & White 2000)

change lies squarely in the 'post-normal' domain characterized by Funtowicz & Ravetz (1994), where facts are uncertain, values are in dispute, stakes are high, and decisions urgent. In such settings, uncertainties can play a central rhetorical role, and there is likely to be increased scrutiny of scientific findings as we have seen. In response to this and to calls from the second assessment report,<sup>2</sup> Moss & Schneider (1999) developed a guidance paper for consistent assessment and reporting of uncertainties in the IPCC third assessment report (scheduled for 2001). This guidance document grew out of a need to produce a 'traceable account of how estimates are constructed' throughout the IPCC report (Pearce 1999). So long as the methods used in constructing such traceable accounts are relatively transparent, they afford a means by which to communicate scientific findings to broader communities such as the IPCC and its audience.

### 1.3. How should it be done?

It is now widely held that there is a need for formal methods to describe uncertainty in communications about climate change. However, the form that such formal methods should take has been only loosely prescribed. Moss & Schneider (1999) emphasize the need for careful quantification of statements about levels of certainty and likelihood in describing outcomes from the science. Such statements will ultimately be expressed probabilistically. Further, they stress that the probability metrics will need to be 'Bayesian' or 'subjective' in characterizing uncertainty rather than 'classical' or 'frequentist' (see Hasselmann [1998] for a comparison of these approaches to detection and attribution). The reason is that it is often impossible to produce the many realizations of an experiment (climate change) that are required for generating a statistical sample in the frequentist paradigm. Perhaps the only way to generate frequentist statistics is by performing model experiments of the real system. However, the reliability of climate models is still not well known (IPCC 1996, Risbey & Stone 1996), and some element of subjective interpretation would still be required to translate the frequentist climate model statistics into statements of uncertainty.

Beyond the ability to produce quantitative subjective probability judgements, formal methods should have the following characteristics: In order to produce a traceable account of the method, each step in the process should be made as explicit as possible. The various assumptions employed should be articulated, along with the elements of judgement. For transparency, the methods should be coherent and consistent with the science. Judgements that need to be made should be tractable and meaningful. Further, where there is a range of disagreement in making judgements based on reasonable arguments, that range should be captured in representing the judgements.

To summarize, the kinds of formal methods used for developing assessments on detection and attribution and highlighting uncertainties should:

- make explicit as many of the relevant judgements in the process of detection and attribution as is reasonable/possible;
- allow for quantitative representations of those judgements;
- be cast in subjective probability terms; and
- capture the range of reasonable opinion on these judgements.

Given these requirements, we have developed a protocol where the problem of detection and attribution is structured and each of its component steps outlined. We tried to capture as much as possible of the full chain of logic in detection and attribution studies in developing the protocol. We classify this as a protocol rather than a survey because of the attempt to provide structure and complete detail in setting out the relevant questions. It is not a survey of overall opinions on detection and attribution, but a tool to identify judgements and provide diagnostic guidance. We use expert elicitation methods (Morgan & Henrion 1990) to fill out the set of detailed judgements in the protocol. We do this because the judgements that need to be made have subjective components on which experts can deliberate. These deliberations are asked for in both qualitative and quantitative form. The qualitative explanations provide a means for the experts to articulate the reasoning behind each of their judgements. This helps in diagnosis of disagreements among experts and it also helps each expert in providing quantitative responses to the protocol. The quantitative responses are given in the form of probability distributions to capture the subjective nature of the judgements and to represent uncertainties.

A range of experts (approximately 20) will complete the protocol in order to sample from among the spectrum of opinion on detection and attribution issues. It will not be distributed as a remote survey to a large group of climatologists because the detection and attri-

<sup>2</sup>For example, chapter 11 in the IPCC second assessment report notes that 'IPCC projections are often expressed with upper and lower limits. These ranges are based on the collective judgement 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'

bution problem requires a high degree of familiarity with the underlying data and methods applied. Such familiarity is not widely distributed outside the community dedicated to this issue. Further, the protocol is designed to elicit the reasoning underlying each of the judgments made, and this is more easily understood in personal interviews than remote surveys. Our selection of experts contains some geographic bias toward North America and Europe (to minimize travel costs). The selection includes about 8 of the authors from the IPCC detection and attribution chapters, a mix of observationalists, theoreticians, modellers, and generalists, as well as scientists with a diverse range of opinions on the severity of greenhouse climate change issues. All of the selected experts publish work relevant to detection and attribution issues. Further details on the experts will be provided in a forthcoming companion paper describing results from the protocol.

A more 'objective' alternative formal approach to the one we have outlined would be to attempt to make each of the determinations in the detection and attribution chain of logic in a climate model (rather than using expert judgement). But to do that still requires that the steps in the process be identified so that diagnostics can be formulated and calculated in a model. A protocol of the kind we outline here is a useful first step in developing any formal approach to describing uncertainties in the manner outlined above. In later work the protocol will include results from a variety of climate models to compare models with one another and with the subjectively determined results.

The remainder of this paper is arranged as follows: First we describe the elements of a formal protocol for making judgements on detection and attribution. Then we describe some of the judgement issues at stake and the role they play. Examples of these issues are provided with reference to one of the key lines of evidence in the detection and attribution protocol. After describing the protocol, we then discuss its utility, making particular reference to the manner in which it has been received in the climate change community and the functions it is intended to perform. The principal purpose of this presentation of the protocol is to describe the *issues* at stake in detection and attribution studies, not their results.

## 2. ELEMENTS OF A DETECTION AND ATTRIBUTION PROTOCOL

A detection and attribution protocol provides a systematic and formal approach for analyzing the problem of detection and attribution of climate change. The protocol must use current scientific understanding, be transparent without sacrificing the complexity of the

issue, and allow for a range of expert opinions. The protocol also needs to provide an intuitively appealing structure to the detection and attribution problem and to account for multiple lines of evidence.

The protocol uses as inputs, probabilistic judgements by climate experts regarding the different lines of evidence, and their interactions. Experts consider the problem in its component steps, which include selecting lines of evidence, and making judgements regarding the detection and attribution of climate change for each line of evidence in isolation, and in combination with other lines of evidence. For each line of evidence, the experts provide probability distributions for the accuracy of observations, and for the size of naturally occurring fluctuations. This allows calculation of the degree to which observed trends exceed estimates of naturally occurring fluctuations. Experts then provide probability distributions for the representation of climate forcings from different causes (including both human activities [e.g. greenhouse gas and aerosol emissions] and natural processes [e.g. solar fluctuations and volcanic eruptions]) and the representation of climate model generated responses to the forcings. They also provide measures of attribution to different forcings. This is represented by the fraction of change in a line of evidence caused by different candidate forcings, conditioned on the actual values of the forcing.

The components in the detection and attribution protocol are as follows:

*Determination of evidence* entails the selection of relatively independent lines of evidence for detection and attribution of climate change. Each line of evidence must be expressed with a quantifiable metric and examined over a specified period of interest.

*Evaluation of evidence* relates to the confidence one has in the validity of the data used to represent the line of evidence. This is an assessment of the likelihood that the data represent the 'true' underlying variable and is represented by a probability distribution for each line of evidence.

*Detection of change* is the process of identifying a signal of change in a climate variable that exceeds the range of background noise of natural climate variability on the relevant time scales. We define natural variability as internally generated, i.e. not induced by external changes in radiative forcing, regardless of the source. Detection requires an assessment of the probability distribution for natural variability in a line of evidence. By comparing that distribution with the distribution identifying changes in a line of evidence, one can assess the probability that the observed change exceeds natural variability. This does not identify the cause of any climate change, but merely identifies that a change has been detected.

*Attribution of cause* for any residual signal (in excess of natural variability) of climate change for each line of evidence must be made among the candidate forcings. Thus the potential causes of any residual signal must be identified. These will typically include changes in greenhouse gas forcing, changes in aerosol forcing, changes in solar forcing, changes in volcanic forcing, and changes in ozone forcing. Since changes in radiative forcing are also uncertain, attribution requires an assessment of the probability distributions for each of the different types of forcing, and an assessment of the fraction of change in the residual signal that can be attributed to each forcing.

*Inferences about attribution* for lines of evidence taken in combination, accounting for possible dependence among different lines.

Each step in the process above can be represented probabilistically. The first step entails selection and characterization of each of the multiple lines of evidence,  $E_i$ ,  $i = 1, I$  ( $I$  is the total number of lines of evidence, typically 4 here). The probability density function (pdf) for the observed change over a specified period for each line of evidence, given by  $f(E_i)$  for line of evidence  $i$ , is elicited from the expert. This is followed by an elicitation of the pdf for the natural variability  $f(N_i)$  over the same time period for that line of evidence. In some cases the distributions are drawn, and in some cases the distributions are estimated by discretizing across the range of the distribution and asking the experts to assign probability masses within each specified interval.

The probability of detection,  $P(D_i)$ , is the probability that the observed change in a line of evidence,  $f(E_i)$ , exceeds the natural variability,  $f(N_i)$ . That is:

$$P(D_i) = P(E_i - N_i > 0) = \int_0^{\infty} f(z_i) dz_i \quad (1)$$

where  $f(z_i) = f(E_i - N_i)$  is the pdf for the residual signal. It follows that  $P(D_i) > 0.5$  implies that the observed change in a line of evidence exceeds the natural variability and some residual signal has been detected. Thus, for example,  $P(D_i) = 0.9$  implies that there is only a 10% chance that the observed change in a line of evidence is a result of natural variability.  $P(D_i) < 0.5$  implies that the signal cannot be distinguished from the noise. By defining detection in this way one has effectively ruled out natural variability as a potential cause of any residual signal, and one must choose among other factors (climate forcings) in making an attribution of cause. In the attribution stage, the probability judgments assume that a change has been detected.

Attribution of causation requires an evaluation of probability density for the history of each candidate forcing,  $F_{i,j}$ . For each line of evidence,  $i$ , the experts identify the relevant forcing mechanisms,  $j$ , that influence attribution. Each of the forcings have some uncer-

tainty associated with them over the time and space scales pertinent to the particular line of evidence. This uncertainty can be represented by a pdf,  $f(F_{i,j})$ .

Once the relevant forcings have been characterized, the task of attributing the residual signal to these forcing mechanisms remains. Note that we are not particularly interested in the probability of attribution per se. In the limiting case a very small fraction of the residual signal may be attributed to a forcing mechanism with probability close to 1, telling us little about the role of that forcing mechanism. We are really interested in the extent to which a forcing mechanism can explain the observed residual signal. Thus, there are 2 elements to attribution: (1) the fraction of residual signal that can be attributed to a given mechanism and (2) the probability with which this attribution can be made. For example, greenhouse forcing may contribute a small fraction to the signal with a high probability (say 20% to the residual with probability 0.95), or a large fraction with a low probability (say 95% to the residual with probability 0.2). In principle, we could ask the experts to produce a pdf accounting for the probability with which each fraction (0 to 1) of the residual signal can be attributed to each forcing mechanism (integrated across the uncertainty in the forcing distributions). In practice this is cumbersome, and we simplify by asking for the expert assessment of just the expected fraction of the residual signal accounted for by each forcing. Thus, our simple overall index for attribution for a given line of evidence is  $Q_{i,j}$ , which is the expected fraction of the residual signal that can be causally attributed to a given forcing mechanism,  $j$ .

Ideally, one would then assess the conditional density  $f(Q_{i,j}|F_{i,j})_{j=1,J}$  for each  $E_i$  and each forcing mechanism  $F_j$ .  $f(F_{i,j})_{j=1,J}$  is the joint density function for each of the  $J$  competing forcing mechanisms for line of evidence,  $i$ . Since the forcing mechanisms are assumed to be independent of each other,

$$f(F_{i,j})_{j=1,J} = \prod_{j=1}^J f(F_{i,j}) \quad (2)$$

Thus,

$$Q_{i,j} = \int f_i(Q_{i,j} | f(F_{i,1})f(F_{i,2}) \dots f(F_{i,J})) dF_{i,1} dF_{i,2} \dots dF_{i,J} \quad (3)$$

In practice, we did not elicit the conditional density  $f(Q_{i,j}|f(F_{i,j}))_{j=1,J}$ , since our pilot test experts found this to be a very difficult task. We simplified the analysis by discretizing the range of values that each forcing  $f(F_{i,j})$  can take into  $m_j$  ranges (typically,  $m_j = 2$  or 3). The experts then partition the probability mass for each forcing amongst the  $m_j$  ranges. This procedure results in a combination of  $K$  different forcing cases to consider, where  $K = \prod_{j=1}^J m_j$ . If forcing  $f(F_{i,1})$  is split into 3 ranges ( $m_1 = 3$ ) and  $f(F_{i,2})$  is split into 2 ranges ( $m_2 = 2$ ), this yields  $K = 6$  forcing combinations to

assign fractional attributions. For each of the resulting  $k = 1, K$  forcing combinations, we elicit the quantity  $Q_{i,j,k}$ .  $Q_{i,j,k}$  is an average value for  $Q_{i,j}$  over the  $k$ th discrete combination. This reduces Eq. (3) to a simple summation:

$$Q_{i,j} = \sum_{k=1}^K Q_{i,j,k} P_k \quad (4)$$

where

$$P_k = \int_j f(F_{i,1}) \times \dots \times f(F_{i,j}) dF_{i,1} \dots dF_{i,j} \quad (5)$$

is approximated by  $P_k = \prod_{j=1}^J P_{j,k}$ , where  $P_{j,k}$  are the probability masses assigned to each of the forcing ranges.

Eq. (4) considerably reduces the difficulty of the elicitation. The illustrative example that follows shows how. Consider the change in the global mean surface air temperature over the past century as the line of evidence ( $i = 1$ ), with greenhouse forcing ( $j = 1$ ), solar forcing ( $j = 2$ ), and all other forcing ( $j = 3$ ) as the competing mechanisms for attribution. Note that radiative forcing from aerosols is not a competing mechanism for attributing the change, since aerosols provide mostly negative forcing in this case, which would not cause the positive response of the temperature signal. However, forcing from aerosols can affect the attributions that the expert makes for the competing forcing mechanisms, since uncertainties in the negative forcing can change one's view of uncertainties in the positive forcing. Hence aerosol forcing is included in the list of forcings characterized by the experts.

The first step is to divide the relevant forcing changes into appropriate ranges. Since greenhouse forcing is relatively well known compared to the other forcings we ignore the uncertainty in greenhouse forcing and do not elicit a density function  $f(F_{1,1})$  or its equivalent discretized representation. For the case of solar forcing, we divide the change in forcing over the past century into ranges of  $<0$ ,  $0-1$ , and  $>1 \text{ W m}^{-2}$  ( $m_2 = 3$ ). Similarly we divide aerosol forcing into ranges  $< -1.5$  and  $> -1.5 \text{ W m}^{-2}$  ( $m_3 = 2$ ). The experts assign a probability to the likelihood that the forcing is in each of these ranges. Next, assuming a particular forcing combination as given, we ask the expert to estimate the fraction of residual signal explained by each of the forcing mechanisms (greenhouse, solar, and all other forcings). Thus the expert answers questions such as: 'Given that solar forcing is  $<0 \text{ W m}^{-2}$  and aerosols  $< -1.5 \text{ W m}^{-2}$ , what fraction of the residual signal can be attributed to greenhouse forcing?' Similar questions are asked for each of the 6 forcing combinations ( $K = m_{\text{sol}} \times m_{\text{aer}}$ ), and answers corresponding to  $Q_{i,j,k}$  in Eq. (4) are obtained. The probability mass  $P_1, \dots, P_6$

associated with each of the 6 forcing combinations can be calculated by taking the appropriate product of the probabilities assigned to each of the discrete range of values of the forcing. A simple summation is now used to determine  $Q_{i,j}$ . Note that the fractional attribution measure  $Q_{i,j}$  is not a probability. Rather, it is a measure of the expected fraction of the residual signal that can be attributed to each candidate forcing averaged over the uncertainties in the different forcings.

The final task is to combine the analysis of detection of change and attribution to cause to provide an overall measure for detection and attribution for a given line of evidence. This is done by multiplying the probability of detection,  $P(D_i)$ , and fractional attribution measure,  $Q_{i,j}$ , to produce an overall measure for attribution  $P(A_{i,j})$ , i.e.

$$P(A_{i,j}) = P(D_i) Q_{i,j} \quad (6)$$

$P(A_{i,j})$  is an overall confidence measure for detection and attribution, and makes intuitive sense. If probability of detection  $P(D_i)$  is high, and fractional attribution measure  $Q_{i,j}$  is high, then  $P(A_{i,j})$  is high, i.e. one has high confidence that a change in the line of evidence has been detected (exceeding natural variability) and that the change can be attributed to the postulated forcing. Conversely, if either  $P(D_i)$  or  $Q_{i,j}$  are low, then one has low overall confidence in attribution—even if either detection or attribution probabilities are individually high. For example, if  $P(D_i)$  for change in global mean surface air temperature is high, yet much of the change can be attributed to climate forcing from the sun, then the value of  $P(A_{i,j})$  from greenhouse forcing is low. In the case that it is difficult to distinguish the change in global mean surface air temperature from its natural variability, then even attributing the entire residual signal to greenhouse forcing would do little to increase the overall confidence in attribution of this change to greenhouse forcing. Note that  $P(A_{i,j})$  is strictly speaking not a probability. It is a measure of confidence in detecting and attributing change in line of evidence  $i$  to forcing mechanism  $j$ . Hence,  $1 - P(A_{i,j})$  is a confidence measure that characterizes the state where either no change in  $i$  is detected or the detected change is not attributed to mechanism  $j$ .

After each line of evidence has been analyzed in the above manner, we then need to consider the attribution outcomes for the combination of lines of evidence. The combined (net) confidence for attribution of the 'balance of evidence' can be calculated as

$$P(A_{\text{net},j}) = P(\cup A_{i,j}) \quad (7)$$

where  $\cup$  is the union operator and  $A_{i,j}$  is the attribution for line of evidence  $i$  and climate forcing  $j$ . Making an attribution from combined evidence requires that careful attention be paid to whether the lines of evidence

are independent or not. Since each individual line of evidence may not be independent of others, dependence among the different pieces of evidence must be accounted for. In practice, the protocol can be simplified somewhat by selecting a finite set of relatively independent lines of evidence. This is the approach we have taken here. A smaller set of lines of evidence also reduces the number of combinations that must be accounted for in assessing dependence. Ultimately, we calculate results using 3 different assumptions about the degree of dependence: (1) Assuming independence among lines of evidence (providing an upper bound for attribution). In this case

$$P(A_{\text{net},j}) = \sum_i P(A_{i,j}) - \sum_{i=1, l=1, i \neq l, i < l}^{I, I} P(A_{i,j})P(A_{l,j}) + \prod_i P(A_{i,j}) \quad (8)$$

(2) Assuming complete dependence among lines of evidence (providing a lower bound for attribution). In this case

$$P(A_{\text{net},j}) = \max[P(A_{i,j})] \quad (9)$$

(3) By taking account of the expert's assessment of any dependence between lines of evidence.

The spread between the independent and dependent assumption cases provides a measure of the importance of judgements on dependence. The manner of selection of lines of evidence is discussed in the following section.

### 3. LINES OF EVIDENCE

A variety of different lines of evidence can be brought to bear on the question of human influence on climate change (Santer et al. 1996a). These include:

- the 20th Century long global mean surface air temperature increase;
- sea level rise;
- water vapor and evaporation increases over tropical oceans;
- high latitude precipitation increases;
- the reduction in diurnal temperature range in many locations;
- better assessments of natural climate variability;
- increasing sophistication and success of pattern matching exercises relating geographical, vertical, and seasonal patterns of modeled greenhouse gas and aerosol climate change with observations;
- the increase in statistical significance of such pattern fits with time; and
- the inconsistency with solar and volcanic forcing in the vertical pattern of temperature.

Some of the items in this list are not lines of evidence as such. For example, assessments of natural climate vari-

ability are relevant to detection for a given line of evidence, but are not lines of evidence in their own right. A number of experts cite the millennial scale temperature reconstructions (e.g. Mann et al. 1998, 1999) as lines of evidence for greenhouse climate change. Again, we do not consider this to be a line of evidence in its own right. Part of it is simply definitional, in that we focus on possible causes of climate changes over the past 100 yr. Taking this time scale as given, it follows that the millennial series provide bounds on natural variability, but do not describe the kinds of climate changes we are interested in. We are interested in greenhouse-induced climate change, and the millennial series cover a period in which the changes in greenhouse forcing were relatively small (until the recent 100 yr period that we focus on). Thus, in terms of greenhouse climate change, there is not much that one could hope to detect and attribute in that series. In this protocol, a 'line of evidence' only qualifies as such if one can hypothesize some change in the metric for that line of evidence that might usefully be related to greenhouse (and other) forcing.

Lines of evidence are used as descriptors of climate change in the past century. No one line of evidence suffices to describe this climate change uniquely. A set of lines of evidence provides a richer picture of that climate change than any one line of evidence. However, given the nebulous nature of climate and climate change (Lorenz 1995), any set of lines of evidence is still only an approximation. For the development of this protocol, we chose not to be exhaustive in including all conceivable lines of evidence. If we did that, the number of questions posed would be too great to sustain the attention of the climate experts, and the problem would be too unwieldy to work through. It is more fruitful to account for the major issues at stake, and stop at a point where each additional line of evidence provides only marginal gains in confidence in detection and attribution. For example, some lines of evidence are mostly derivative of the fact that 20th Century global mean temperatures have increased. This would remove sea level rise and water vapor and evaporation increases over tropical oceans, which are substantially derivative of the temperature increase. The point at which one stops adding lines of evidence is determined empirically in consultation with each of the experts.

We conducted an initial survey with a number of experts to assess the degree to which they converge in this manner. In practice, experts largely agree on a core set of about 4 distinct lines of evidence as markers of greenhouse climate change. Few added additional lines of evidence beyond the following: (1) the global mean surface air temperature increase during the 20th Century; (2) the reduction in diurnal temperature

range over land areas; (3) the geographical/spatial pattern of changes in surface temperature; and (4) the vertical pattern of changes in global temperature. These lines of evidence form the core set for the protocol. However, the protocol allows experts to select their own lines of evidence and to add to this set (or subtract from it) until they feel they have accounted for the important lines of evidence.

The above lines of evidence were selected as diagnostics of greenhouse climate change, not of more general anthropogenic climate change. This is because the search for anthropogenic climate change, irrespective of cause, can be readily established from the impact of ozone depletion on stratospheric temperatures (Chanin & Ramaswamy 1999). Further, the problem of climate change in the longer run is mostly a problem of controlling emissions of greenhouse gases. While aerosol emissions and ozone depletion have important impacts on current climate, their role is expected to diminish in the future relative to the greenhouse gas contribution. While the protocol is structured to detect greenhouse climate change, the contributions from other natural and human forcings are explicitly included at each step as well.

The selection of lines of evidence is also constrained to some degree by the sources, quality, and quantity of available data. Thus more weight tends to be placed on lines of evidence that can be characterized by reliable long-term observational records. This favors the global surface air temperature record (e.g. Jones 1994) over the much shorter upper air satellite record (e.g. Christy & Spencer 1995) for example. It also favors temperature-based lines of evidence over hydrologically based lines of evidence (e.g. precipitation; Karl & Knight 1998) which have less homogeneity or have been less well observed (Wigley 1999). This bias presents the possibility that searches for greenhouse climate change will 'lamppost' (Ravetz 1997) around available data. This is difficult to overcome save for maintenance of a general awareness of the problem.

#### 4. GLOBAL MEAN SURFACE TEMPERATURE EXAMPLE

To provide the flavor of the set of judgements identified in responding to the protocol we trace through some of the elements of the protocol for a single line of evidence. We have selected the century-long trend in the global mean surface temperature for this example ( $i = 1$ ) because it is the best known and studied among the different lines of evidence. Perhaps not surpris-

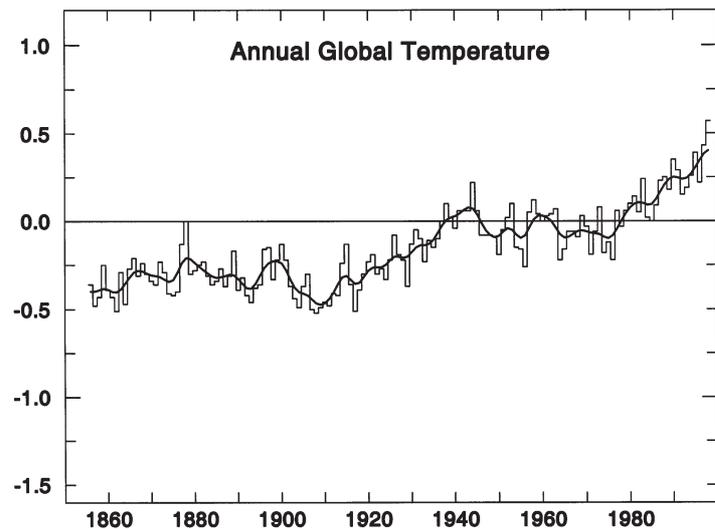


Fig. 1. Long-term global mean surface temperature anomalies. The thicker line represents a five-year running mean. The data are from Jones (1994) and Jones et al. (1999), and provided via <http://www.cru.uea.ac.uk>

ingly, preliminary work with the protocol reveals that it is also the line of evidence with the highest levels of confidence for detection and attribution. A representation of the time series of the global mean surface temperature is shown in Fig. 1. This series shows a century-long warming on the order of 0.5°C. To clarify the questions pertaining to this line of evidence we will provide responses from one of the experts (Expert 11) who completed the protocol. We stress that results from this expert are given to highlight specific issues that arise in detection and attribution, but not for the purpose of drawing conclusions about the data. A full set of results from all the experts for all lines of evidence will be reported elsewhere.

##### 4.1. Eliciting probability density for evidence: $[f(E_i)]$

The first confidence judgement to make is in each line of evidence as such. The evidence itself will be subject to various limitations. For example, observations of the century-long global mean surface air temperature record are subject to a variety of limitations related to: sparse observations with incomplete coverage, especially over oceans; changes in locations of measuring stations; changes of instrumentation/method for determining sea surface temperatures; and urban heat island effects (Jones & Wigley 1990). The question being asked at this point is the extent to which the evidence is 'real'. This judgement is expressed in the form of the expert's pdf assessing the possible values for the global mean surface tempera-

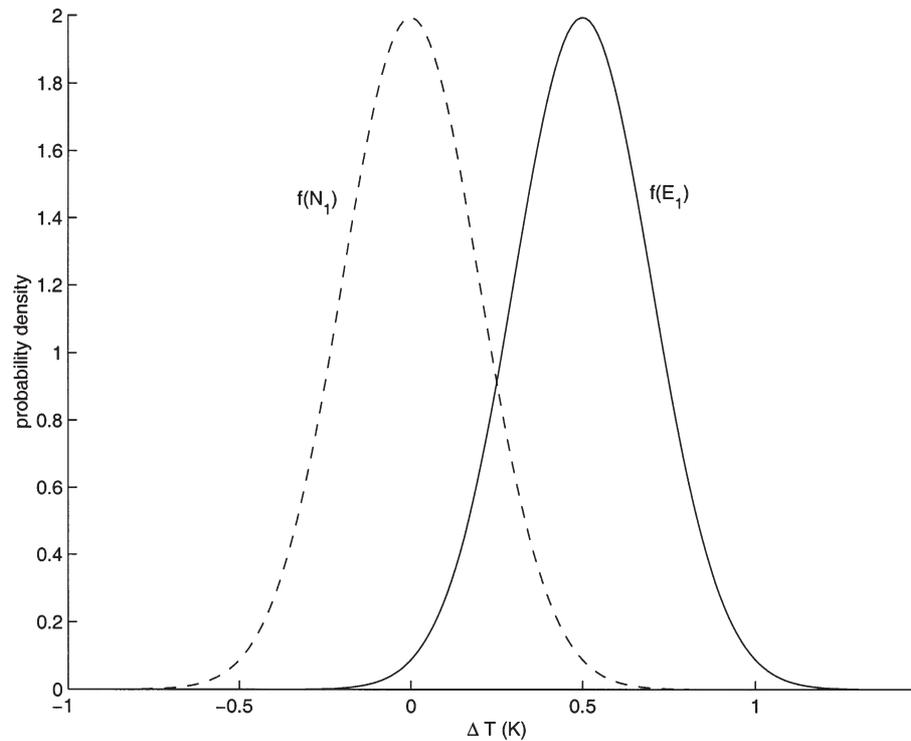


Fig. 2. Expert 11's probability density function for an estimate of the century-long trend in global mean surface air temperature (solid line) and for the century-long trend in natural variability (dashed line)

ture change over the period. This will typically take the form shown by the solid line in Fig. 2, which is the pdf for  $f(E_1)$  given by Expert 11. This expert believes that the most likely value for the century-long change is 0.5 K, but allows some uncertainty and provides a standard deviation of 0.2 K about this value. Note that, while this expert used a normal distribution for the pdf, any distribution could be given. The tails of the distribution indicate that, given this experts estimate of the uncertainty, there is a small probability that the actual temperature change could be as low as zero or as high as 1 K.

Once confidence in the evidence has been assessed, the next task is to characterize natural variability in the line of evidence and to make judgements on the significance of any trends relative to natural variability. This is the detection stage.

#### 4.2. Eliciting probability density for natural variability: [ $f(N_i)$ ]

Each expert's judgement of the magnitude of natural variability of the global mean surface temperature change over the past century is also represented by a pdf,  $f(N_1)$ . The experts are first asked about the processes responsible for setting natural variability on this time scale, and about their confidence in understanding of the pertinent processes. For global mean surface air temperature, the natural variability is set by the

internal dynamics of energy exchange among reservoirs with different time constants. For example, storms in the atmosphere pump energy into and out of the ocean, and the ocean moves heat around between the surface and deeper layers via a variety of circulation mechanisms with different time constants. Experts tend to agree on the processes driving natural variability, but are less sanguine about the ability to model them. After a discussion of these kinds of processes the experts are asked to specify a distribution for  $f(N_1)$ .

Natural variability,  $f(N_1)$ , is difficult to characterize from observations for most lines of evidence. This is because we do not have reliable long-term measurements of the 3-dimensional temperature structure of the atmosphere. Direct observations of surface temperature cover only the period since the industrial revolution began (and greenhouse gas concentrations started to increase), while measurements above the surface are even more recent. Thus detection studies usually make use of representations of natural variability in climate models as a proxy for the actual natural variability (e.g. Santer et al. 1996b, Tett et al. 1996). The question arises as to how good the model-generated estimates of natural variability are. The protocol contains questions asking the experts to provide confidence judgements in the representation of natural variability in models to ascertain whether there are likely to be significant biases introduced by use of model proxy measures of natural variability. It is generally considered that the models underestimate the

'true' natural variability, though there is little consensus on the degree of underestimation. Expert 11 estimates that there is about a 30% chance that the true natural variability is between 0.5 and 1.5 times the value typically estimated by models (i.e. models about right) and about a 70% chance that the true natural variability is greater than 1.5 times the value typically estimated by models.

In specifying the spread of the distribution for  $f(N_1)$  of the global mean temperature over the past 100 yr, experts frequently make reference to proxy reconstructions of this quantity over multiple century time scales (e.g. Mann et al. 1998, Briffa & Osborn 1999). While proxy reconstructions are limited in global coverage and infer temperatures indirectly, they are invoked to provide loose bounds on the potential magnitude of century-scale natural variability. The expected mean trend in century-long series of (internally generated) natural variability is typically zero unless there are reasons to expect a prolonged cooling or warming on this scale. The proxy records indicate some natural cooling over the past millenium (Mann et al. 1999), though this is loosely attributed to astronomical forcing, which is an external forcing. Since the mean of the expected trend in natural variability is close to zero on 100 yr time scales, the spread and tails are the more critical aspects of this distribution for detection studies. Expert 11's distribution for natural variability,  $f(N_1)$ , is shown by the dashed line in Fig. 2. It has a mean of zero (assumes no natural long-term warming or cooling) with a standard deviation of 0.2 K. From this distribution we would conclude that century-long temperature excursions of +0.5 K can occur naturally, but with low probability.

#### 4.3. Calculating probability of detection: $[P(D_1)]$

By comparing the expert's pdfs for  $f(E_1)$  and  $f(N_1)$  one can calculate the probability of detection,  $P(D_1)$ . This is the probability that the observed trend of global mean surface temperature exceeds natural variability of this quantity on the century time scale. By specifying probability distributions for both the size of the trend in global mean surface temperature change and the expected value of natural variability over the same period the experts are effectively answering the question: Is the trend unusual, and, given the uncertainty of the estimates, would we know it if it was? For the global mean surface air temperature, a preponderance of results from the literature and the expert assessments with this protocol suggest that the trend is indeed unusual compared to natural variability. The detection probability for Expert 11, calculated from the distributions given in Fig. 2, is 0.95.

Once a trend in global mean surface temperature has been detected at some level of confidence, one has now effectively ruled out natural variability as the source of any residual signal. Of course, if an expert believes there is effectively no residual signal, that would be accounted for by a very low estimate in the detection probability  $P(D_1)$ . For the global mean surface air temperature, detection levels tend to be uniformly high across experts, though 3 of the 16 experts who have currently completed the protocol yielded  $P(D_1) < 0.95$ . Once the probability of detection has been calculated, the next step is to attribute causes to any residual signal.

#### 4.4. Identifying and characterizing the forcing: $[f(F_{i,j})]$

The first issue to grapple with in determining the cause of any residual signal is to identify the set of processes that might be responsible. This amounts to an identification of possible climate forcings (causes) over the relevant period and requires that probability distributions be specified for the possible values of each forcing,  $f(F_{i,j})$ . For each postulated forcing of a residual signal, one needs to characterize the forcing history in space and time in order to generate a response which can be compared against the observed response of the system. There are a variety of plausible forcings for each line of evidence and each will typically contribute in varying amounts.

For the global mean surface temperature trend there are many different forcing processes that could play some role, although the dominant ones are thought to be increases in greenhouse gases, aerosols, and solar variability. These different forcings are known with different degrees of confidence. Fig. 3 shows one esti-

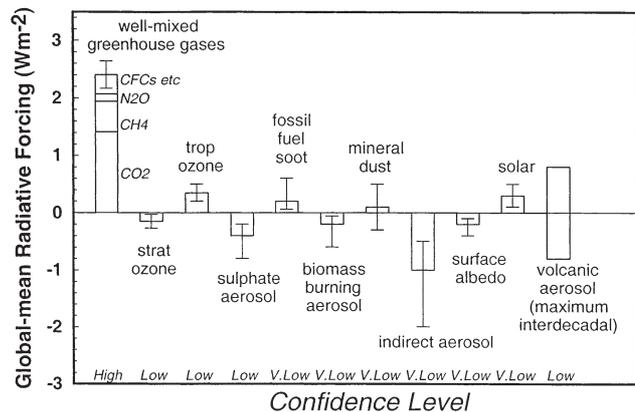


Fig. 3. Global mean radiative forcing for the period from pre-industrial times to the present. From Shine & Forster (1999)

Table 1. Partitioning of probability mass between the nominated forcing ranges for the global mean surface air temperature line of evidence for Expert 11

Solar forcing (Wm <sup>-2</sup> )	Probability mass	Aerosol forcing (Wm <sup>-2</sup> )	Probability mass
<0	0.15	0 – -1.5	0.8
0 – 1	0.80	-1.5	0.2
>1	0.05		

mate of the range of relevant forcings, together with an assessment of the confidence in each forcing from Shine & Forster (1999). The forcing due to increases in carbon dioxide (the most important anthropogenic greenhouse gas) is reasonably well known over the past 100 yr and the forcing is relatively homogeneous across the planet. While the forcing due to the sun is also relatively homogeneous in space, it is very poorly known as a function of time before direct measurements began in the satellite era (ca 1978) (Kelly & Wigley 1990). For aerosols the forcing is heterogeneous in space and the magnitude of the forcing due to indirect processes (interactions of aerosols with clouds) is poorly known (Kiehl & Briegleb 1993), even now (Shine & Forster 1999). Other forcing factors that are thought to play a role in setting global mean surface temperature trends are cloud processes, changes in ozone, stratospheric aerosols from volcanic eruptions, and land surface physics.

Expert 11 identified each of the above forcings as relevant for the century-long change in global mean surface air temperature, though attributed smaller magnitudes to the latter 3. The forcing from volcanic aerosols was judged to be slightly negative on the time scale of the past century. Land use changes were argued to be either slightly negative due to albedo changes, or slightly positive due to changes in the surface energy budget. For aerosols, this expert feels that there is an 80% probability that the true aerosol forcing lies in the range between 0 and -1.5 W m<sup>-2</sup>, and allows a 20% probability that it may be even stronger negative than that (see Table 1). For solar forcing, this expert allocates 80% probability mass to the range between 0 and 1 W m<sup>-2</sup>, 5% to the possibility of stronger positive forcing, and the remainder to the possibility of negative solar forcing. This expert reasoned that present proxy reconstructions for solar forcing are highly uncertain and allow for the possibility of zero, or even negative, forcing.

Once the candidate forcings have been characterized, they must each be translated into a response of the climate system to compare the responses to those forcings with the actual response of the system (global mean surface temperature trends in this case). Since the translation of forcing to response is usually done by climate models, we ask the experts to describe what the relevant physics are for translation from forcing to response, and to assess how well the model represents the relevant processes. This is done for each of the candidate forcings. Different physical processes will be called into play for different lines of evidence. For example, characterization of the global mean surface temperature response relies on good representations of temperature-dependent feedback processes in the models (clouds, snow and ice, water vapor). The characterization of the vertical pattern temperature response in climate models relies on having good representations of radiative and convective processes in the model, which are important in redistributing heat through the vertical atmospheric column. Atmospheric circulation processes are also important for the vertical pattern. For the geographical pattern of temperature, ocean circulation processes will be important in addition, since they will influence the sea surface temperature distribution as well as adjacent land temperatures. For characterization of the diurnal temperature response it is important that the model has a good representation of boundary layer and land surface physics, and cloudiness.

**4.5. Eliciting fractional attribution: [Q<sub>i,j</sub>]**

Once the forcing history of the various candidate forcings has been characterized and the translation of

Table 2. Fractional attribution values for Expert 11 for the global mean surface air temperature. This is the fractional responsibility of greenhouse forcing, solar forcing, and all other forcing in explaining any residual signal in century-long surface temperature change that exceeds natural variability. This partitioning is carried out given that the forcing by aerosols and the sun is in the indicated ranges

Solar forcing (W m <sup>-2</sup> )	Aerosol forcing between 0 and -1.5 W m <sup>-2</sup>	Aerosol forcing stronger than -1.5 W m <sup>-2</sup>
<0 Wm <sup>-2</sup>	Q <sub>1,ghg,1</sub> = 0.6	Q <sub>1,ghg,2</sub> = 0.6
	Q <sub>1,solar,1</sub> = 0.0	Q <sub>1,solar,2</sub> = 0.0
	Q <sub>1,other,1</sub> = 0.4	Q <sub>1,other,2</sub> = 0.4
0 – 1	Q <sub>1,ghg,3</sub> = 0.6	Q <sub>1,ghg,4</sub> = 0.6
	Q <sub>1,solar,3</sub> = 0.1	Q <sub>1,solar,4</sub> = 0.1
	Q <sub>1,other,3</sub> = 0.3	Q <sub>1,other,4</sub> = 0.3
>1	Q <sub>1,ghg,5</sub> = 0.6	Q <sub>1,ghg,6</sub> = 0.6
	Q <sub>1,solar,5</sub> = 0.1	Q <sub>1,solar,6</sub> = 0.1
	Q <sub>1,other,5</sub> = 0.3	Q <sub>1,other,6</sub> = 0.3

those forcings into a response has been assessed, the next step is to attribute and partition any residual signal among the candidate forcings. This assessment is carried out for various combinations of uncertainty ranges in characterization of the forcing,  $f(F_{1,j})$ , so that we can appropriately take into account the effect of this uncertainty in the attribution to each forcing.

The fractional attribution values for Expert 11 are shown in Table 2. The pdf for aerosol forcing has been discretized into the 2 ranges shown, and that for solar forcing into the 3 ranges shown. That means that there are  $K = 6$  ( $2 \times 3$ ) different forcing cases of the expert to consider. The fractional attributions in each of the 6 cases are weighted by the probability masses allocated to those ranges indicated in Table 1. The highest weighting ( $0.8 \times 0.8 = 0.64$ ) corresponds to the case where aerosol forcing is between 0 and  $-1.5 \text{ W m}^{-2}$  and solar forcing is between 0 and  $1 \text{ W m}^{-2}$  ( $k = 3$ ). In this case, Expert 11 partitions most of the responsibility for the residual century-long warming trend to greenhouse forcing ( $Q_{1,\text{ghg},3} = 0.6$ ), with solar forcing accounting for  $\frac{1}{10}$  of the residual trend and all other forcing accounting for the remaining  $\frac{3}{10}$ .

For the case where aerosol forcing is stronger than  $-1.5 \text{ W m}^{-2}$  and solar forcing is between 0 and  $1 \text{ W m}^{-2}$  ( $k = 4$ ; 0.16 weighting), Expert 11 has not changed the fractional attributions ( $Q_{1,j,3} = Q_{1,j,4}$ ). This determination relates to assumptions about the relative role of uncertainties in knowledge of climate sensitivity and the climate forcings. In this case the aerosol forcing is stronger than in the previous case, meaning that there is more negative radiative forcing in the system. Some experts argue that in order to produce the observed warming trend with more negative forcing present, there must be more positive forcing of the system than previously thought to be the case. In focusing on the uncertainties in the forcing, the experts note that there is more uncertainty associated with the estimates of 'solar' and 'other' forcing than with greenhouse forcing. Thus, there is more scope to adjust the estimates of solar and other forcing upwards to counteract the additional negative aerosol forcing. These experts therefore increase the fractional attribution to solar and other forcings and decrease the fraction accounted for by greenhouse forcing in this case. Alternatively, some experts (including Expert 11) argue that one does not need to adjust ones estimates of the forcings in this case to account for the observed warming trend. There could simply be less net positive forcing of the system, and it may be the understanding of climate sensitivity that is in error. With less net positive forcing, one assumes that the climate sensitivity is higher than previously thought in order to account for the observed warming. With this argument the fractional attributions for each of the  $j$  forcing candidates do not need to

be adjusted between cases  $k = 3$  and  $k = 4$ . Expert 11 believes that errors in understanding of climate sensitivity are more likely to be important in this case than uncertainty in the forcing.<sup>3</sup>

From the information in Tables 1 & 2 we can use Eq. (4) to calculate Expert 11's fractional attribution to each of the forcings,  $Q_{1,j}$ , for the global mean surface air temperature line of evidence. This yields  $Q_{1,\text{ghg}} = 0.6$ ,  $Q_{1,\text{solar}} = 0.09$ , and  $Q_{1,\text{other}} = 0.31$ , confirming the lead role of greenhouse forcing (for this expert) when uncertainty in the forcing is taken into account. The overall fractional attribution to the forcings,  $P(A_{1,j})$ , can then be calculated using Eq. (6). Because the probability of detecting a residual signal from natural variability,  $P(D_1)$  is so high (0.95) for Expert 11 for this line of evidence, the overall fractional attributions are almost identical to the fractional attributions. Thus this expert attributes about  $\frac{6}{10}$  of the responsibility for the change in global mean surface air temperature to increased greenhouse forcing.

By the time the final confidence estimate in attribution of the change in a line of evidence to a specific forcing is made, each expert has explicitly considered each step in the chain from characterizing the initial evidence to get to this point. By structuring the problem in this manner, the assumptions and reasoning are made clear, the questions are specific enough to have meaning to the experts, and one can assess where confidence is high or low at specific steps in the process.

#### 4.6. Assessing dependence

After each line of evidence is considered in isolation, a set of judgements are then made regarding combinations of the lines of evidence. For example, Expert 11 identified 3 major lines of evidence: the century-long trend in global mean surface air temperature, the change in geographical pattern of temperature over the past 30 yr, and the change in diurnal cycle of temperature over land areas over the past 30 yr.<sup>4</sup> In order

<sup>3</sup>A third possibility is to conclude that uncertainty in climate forcings and sensitivity is less important than uncertainty in the assessment of natural variability. One expert noted that if the forcing turned out to be different he would first assume that natural variability was in the tails of his estimated distribution in accounting for the observed climate change

<sup>4</sup>For consistency among lines of evidence they should all cover the same time period. Thus, ideally the geographical pattern of temperature change and diurnal temperature change would also span the last century. In practice, there is insufficient observational data to construct reliable trends in these lines of evidence over more than the past 30 yr or so. Detection studies on these lines of evidence have therefore used the shorter time period. For consistency with these studies we have also adopted the shorter period

Table 3. Correlations between the global mean surface temperature and the geographical pattern of temperature change (represented by land-ocean mean temperatures over the globe where data is available), the diurnal temperature range over land, and the annual cycle of temperatures over land. Correlations are based on decadal low-pass filtered data. In the top row none of the time series has been detrended. In the bottom row each of the time series has been detrended by removing a linear trend over the period. The period of the data used is 1880 to 1994, except for the diurnal temperature range where it is 1890 to 1994. Sources of data are Jones (1994), Parker et al. (1994) and Peterson & Vose (1997)

	Geographical pattern	Diurnal temperature range	Annual cycle
Global temperature	0.47	-0.82	-0.57
Global temperature (detrended)	-0.07	-0.14	0.02

to account for possible dependence among these lines of evidence, we elicit the conditional probabilities among the different lines of evidence. An example of this would be a probability assessment that a change in geographical temperature pattern (land areas warming faster than ocean areas) is ‘true,’ given that a global mean surface air temperature change has been observed as ‘true.’ We have structured the questions so as to maximize the analytical benefits of conditional independence. We expect that many lines of evidence will either be roughly independent of one another, since the lines of evidence have been chosen at least in part on the basis of their relative independence, or be conditionally independent of one another given the global mean surface temperature change during the 20th Century.

One measure of the degree of dependence between the different lines of evidence is given by their correlations. These are shown in Table 3 for several lines of evidence correlated with global mean surface temperature. The correlations are shown with and without detrending of the data. Detrending effectively removes any century-long trends in the data, whether the result of natural variability or external forcing. As seen in Table 3, the detrended lines of evidence are effectively uncorrelated on decadal time scales, while those with trends do exhibit substantial correlation. In this case the question of the degree of dependence among the lines of evidence hinges in part on whether the trends should be removed from the data—which depends in part on whether you view the trend as the result of forced changes (remove it) or natural variability (leave it in). In practice it is a combination of both.<sup>5</sup> The precise partitioning between these contributions makes a big difference to the resultant correlation. This is an example of the kind of judgement that is integral to detection and attribution (whether it is made explicit or not), which is critical to the results, but for which there is no unique ‘right’ answer at present.

The ‘true’ correlation between 2 lines of evidence is dependent on the degree to which the data underlying each of them are unique. The degree of independence for attribution is also a function of the extent to which the lines of evidence refer to different climate processes that are not simple derivatives of one another.<sup>6</sup> This boils down to an assessment of the degree to which each line of evidence is able to provide differentiation between candidate forcings beyond that which is provided by other lines of evidence. In the protocol, subjective judgements about dependence among lines of evidence are expressed in probabilistic

form. For Expert 11, as for many other experts, the degree of dependence among lines of evidence is assessed to be uniformly low. The one exception for Expert 11 is for the case of the global mean and geographical pattern, where the expert reasoned that there is some overlap in the data, but also considerable spatial structure that is not connected to the global mean. Having assessed the degree of dependence among the lines of evidence, the final step is to develop a combined measure of attribution across all lines of evidence for each forcing.

#### 4.7. Assessing combined attribution: $[P(A_{\text{net},j})]$

The fractional attributions to greenhouse forcing for each line of evidence,  $P(A_{i,1})$  can be given separately or combined to form a net attribution,  $P(A_{\text{net},1})$ . There are a variety of ways in which the evidence can be combined, both quantitatively and qualitatively. One way is to set a threshold for the lines of evidence,  $P_T$ , and calculate the number of lines of evidence for which  $P(A_{i,1}) > P_T$ . This facilitates the use of qualitative measures such as ‘balance of evidence’ to describe the overall level of attribution. Examples of other qualitative terms for net attribution are discussed in Section 5.5. Expert 11’s fractional attributions to greenhouse forcing in explaining the

<sup>5</sup>Note that high correlations among the variables could also result from an internal physical driver common among the variables. Such a relationship would normally apply on different time scales and should therefore survive detrending. The fact that the correlations plummet when the data is detrended suggest that this is not the case

<sup>6</sup>For example, though sea level rise is a complex phenomenon, one would expect that thermal expansion in response to increasing ocean temperatures would result in sea level rise. Thus, global mean surface temperatures and sea level rise should be related

changes in global mean surface temperature, the geographical pattern, and diurnal cycle are 0.6, 0.5, and 0.25 respectively. If  $P_T$  were set at 0.5, for instance, to account for lines of evidence where greenhouse forcing is always the dominant factor, then 2 out of 3 lines of evidence qualify for Expert 11. If  $P_T$  were set at 0.95, no lines of evidence would count for Expert 11. Setting  $P_T = 0.95$  is equivalent to saying that greenhouse forcing accounts for 95% of the century-long climate change. This standard is needlessly stringent, since greenhouse forcing does not need to account for almost all of the signal in past climate change to be important (or even dominant) in future climate changes.

To provide a quantitative assessment of  $P(A_{\text{net},1})$  we combine each of the  $P(A_{i,1})$  using different assumptions about the degree of dependence among lines of evidence. In the case that the additional lines of evidence are completely dependent on the global mean surface temperature change,  $P(A_{\text{net},1}) = P(A_{i,1}) = 0.6$  (Eq. 9) for Expert 11. In the case that all lines of evidence are independent of one another, Eq. (8) yields  $P(A_{\text{net},1}) = 0.85$  for Expert 11. Thus, the confidence in combined attribution to greenhouse forcing changes from 60 to 85% for this expert, depending on the degree of dependence among the data. When the expert's own assessment of dependence among the data is taken into account, this figure is 75%.

Translating these numbers back into qualitative form, one could say that this expert has 'medium' confidence (see Table 4) in attribution of a greenhouse gas role to climate change as represented by the global mean surface air temperature alone. When the expert's other 2 lines of evidence are included in the representation of climate change, confidence either remains in the 'medium' category if the lines of evidence are completely dependent or increases to the 'high' level if the expert's assessment of dependence is taken into account. This further underscores the importance of dependence issues in making overall judgements across lines of evidence. Whatever the appropriate levels of dependence are, the protocol provides a way to make confidence levels explicit, along with the influence of critical assumptions on the overall levels of confidence.

Table 4. Scale for assessing state of knowledge in translating between quantitative probabilities and qualitative uncertainties. From Moss & Schneider (1999)

Probability value	Qualitative ranking
1.00–0.95	Very high confidence
0.95–0.67	High confidence
0.67–0.33	Medium confidence
0.33–0.05	Low confidence
0.05–0.00	Very low confidence

## 5. FUNCTION AND SCOPE OF THE PROTOCOL

The protocol described here for detection and attribution has been distributed among the community of scholars working in this field, and has tended to provoke either quite favorable or quite unfavorable responses. This polarization may reflect differently held paradigms of the proper function and role of science for policy, and it may also reflect different views about the intended function and application of the protocol. We explore both these issues in this section.

### 5.1. Questioning expertise

The protocol is in some sense very narrow and pertains specifically to the question of detecting and attributing possible greenhouse-induced climate changes. Yet the process of answering this question calls forth a large number of judgements (which the protocol is designed to make explicit) covering a range of areas (radiative forcing, climate modeling, observational analysis, statistical methods), which tend to be inhabited by specialists because of the detailed nature of each of these areas. This raises the question as to whether any single expert is competent to complete the entire protocol. In fact, some experts have already communicated to us that they would not venture to make some of the judgements. We regard it as an open question as to whether single experts are able to make meaningful judgements across the protocol, and will make efforts to test this by administering specific parts of the protocol to acknowledged sub-disciplinary experts (in say radiative forcing), so that we may compare the judgements of those who work directly in a particular area with one another and with those who don't work in the area. If it turns out that only specialists seem capable of making meaningful judgements on their 'own' data, that finding has implications for the way in which groups like the IPCC working groups attempt to come to consensus on issues that span multiple specialist domains. Preliminary results from the protocol suggest that spread among experts reflects a general lack of understanding on some specific issues (e.g. characterization of the diurnal cycle) more so than it reflects the ignorance of non-specialists per se.

### 5.2. Identifying important issues

While the response to the protocol may have implications for the process of achieving consensus on issues like detection and attribution, it is not designed as a tool for building consensus. Its goal is to help identify areas of uncertainty and indicate where agree-

ment exists and where it does not. It attempts to highlight the elements of judgement on this issue and explore how different experts make these judgements. That may in fact be an impediment to building consensus in making the areas of disagreement explicit. However, by identifying precisely where in the chain of logic from identifying evidence to attributing greenhouse climate change the disagreements occur, one can at least develop a sense of whether the disagreements are fundamental or not. Some disagreements among scientists are germane to policy responses and some are not. For example, if disagreement focuses on the vertical temperature pattern, but if that turns out to be a less critical line of evidence for setting each expert's overall attribution to greenhouse gases (as seems to be the case), then it would be useful to know that.

### 5.3. Highlighting process

The protocol is designed such that it provides for quantitative probabilistic responses for each of the steps in the detection and attribution chain, and calculates overall probabilities of attribution in a mathematically consistent manner. However, the focus of interest is not entirely on the final attribution probabilities that emerge at the end. Admittedly, the magnitudes of the attribution probabilities and their spread across experts are of genuine interest. If the attribution probabilities are uniformly high and the spread is low, that sends a signal about confidence on this issue that is very different than if the converse is true. But it is also possible that even in this case the agreement across experts on overall attribution probabilities may be fortuitous. For example, it might be the result of compensating differences in probabilities of detection and attribution, or of differences between different lines of evidence from expert to expert. The overall attribution results are much more meaningful when considered in conjunction with the steps that were taken to reach them. If there is agreement on the series of explicit judgements underlying the overall attribution probabilities, that is much less likely to be fortuitous. The information content from the protocol resides more so in the articulation of process than it does in the final results.

### 5.4. Highlighting subjectivity

In our interactions with researchers in this field, some have objected to the explicitly subjective framing of the protocol. For example, one critic stated that 'this expert elicitation weights the science by people's prej-

udices and is of no merit. If it were an *objective* approach it would be very relevant to IPCC.' This comment weighs in on one side of the long-standing paradigm dispute over the relative roles of facts and values in scientific research and analysis. Our critic presumably supports the view that facts and values are separable, and that there are facts which are not value-laden. By contrast, we believe that some subjectivity is unavoidable, but that it is by no means fatal. As noted by Shrader-Frechette (1984), just 'because our observations may be seen in *different* ways, it does not follow that they may be seen in *any* way.'

Subjective judgements are an integral part of the process of contemporary detection and attribution studies, whether they are made *explicit* in a probabilistic protocol such as we have described or whether they are *implicit* in the choices used to construct climate models for detection studies for instance. The notion that there are purely objective ways to carry out detection and attribution studies is a convenient fiction. For many of the relevant lines of evidence for greenhouse climate change, one *fact* that all experts can agree on is that existing data are inadequate to completely characterize the natural variability of the climate system on the time scales of relevance to the detection problem. This means that any detection analyst must use some judgement as to what the natural variability has been. Without long time series of observational data, studies frequently use climate models to characterize the natural variability. Sometimes the analyst will articulate reasons why they think the natural variability of the model is under-representing the true natural variability of the system. For instance, the models do not include all the processes occurring in the real system, and available comparisons of model spectra with observations suggest that the models lack power at the lowest frequencies (Barnett 1999). Thus analysts will sometimes increase the magnitude of natural variability in the model to a value that seems like it might better represent the natural variability of the real system. There are sensible choices to make in doing this and some indications from observations that will guide the choice, but there is no single 'objective' way to match a figure whose true value is unknown. Thus, one is always engaged in making some subjective judgement in a detection and attribution study, whether or not this subjectivity is identified and articulated.

Of course one might wait until we have long-enough time series of observational data to unambiguously determine the natural variability on the time scales pertinent for greenhouse response (though even this is difficult, because the time series is already 'corrupted' by the forcing). But by then the greenhouse 'experiment' will have already been performed and the notion of a 'detection' study rendered moot. Thus one can

adopt a 'subjective' probability framework and provide detection and attribution assessments before the fact, or one can adopt the 'frequentist' (or 'objective') statistical framework and provide input largely after the fact. But it is a fallacy to argue for frequentist approaches as alternatives to subjective approaches for contemporary analyses with existing limited data.

### 5.5. Communicating results

By highlighting the process used to obtain detection and attribution results, the protocol allows for more transparency and nuance in communication of the results. The result is not just a single number or statement, but a map of the terrain of knowledge and uncertainty on this issue. Further, the probabilistic results from the protocol can be readily translated into qualitative statements of confidence using formulations such as that presented by Moss & Schneider (1999). Their 5 point scale translates between probabilities and levels of confidence as shown in Table 4. The exact break up of the scale is not so important as the fact that one uses a common and well-articulated scale for such translations.

Terms like 'balance of evidence' are also more amenable to definition when using the protocol. With probabilities of attribution calculated for each line of evidence one can define thresholds for attribution ( $P_T$ ) and calculate the number of lines of evidence exceeding the thresholds. For example, when all lines of evidence exceed the threshold one could speak of a 'consensus of evidence.' When more exceed the threshold than not, one could refer meaningfully to a 'balance of evidence' and, when fewer exceed than not, one could use terms like 'some evidence.' In each case, it would be clear exactly what the qualitative label meant in quantitative terms. The level at which the threshold,  $P_T$ , is set can be varied to set the appropriate burden of responsibility for greenhouse forcing.

If formal frameworks like the detection and attribution protocol were used periodically at something like the frequency of IPCC assessments, they could be appropriate vehicles for reflecting improvements in scientific understanding. In any given period the protocol could be used to prioritize research by highlighting where uncertainties or disagreements are large, and more importantly, where they are likely to be critical in terms of overall ability to detect and attribute greenhouse climate changes.

## 6. CONCLUSIONS

This paper describes a protocol that provides a coherent structure to analyse issues of detection and

attribution of climate change. The problem being addressed is essentially one of developing a methodology for coping with multiple causation in the face of considerable uncertainty. In this case there are a variety of competing causes that may be responsible in part or whole for climate changes that have been observed. The attribution among competing causes is important for gauging the potential extent of future possible climate changes. While the science underlying detection and attribution studies has been improving in sophistication and scope, the results presented to date have reflected either single isolated studies of one part of the issue, or have provided holistic assessments lacking detail of the underlying reasoning. The protocol presented here provides a synthesis of both these kinds of studies by allowing for overall assessments of attribution while preserving a transparent trail of the process used to arrive at those assessments.

There are a number of advantages to preserving a map of the detection and attribution process. First, the areas of relative agreement and disagreement among experts can be identified. The relative importance of any loci of disagreement or uncertainty can be assessed by examining what the contribution of those steps are in the detection/attribution process to the overall probability of attribution. This is possible because the framework is both explicit and quantitative. By isolating critical aspects of the detection/attribution process in this manner, research priorities can be set accordingly. Further, in sorting critical from inconsequential disagreements (in terms of their contribution to overall attribution) the policy community will be better informed about the role of uncertainties that may matter. Formal detection/attribution frameworks also provide meaningful and unambiguous ways to translate between quantitative expressions of probability and qualitative expressions of uncertainty. They may also provide useful means for monitoring and diagnosing progress in the field.

A number of issues in detection and attribution studies are brought to the surface in the protocol. The first one identified here is whether one seeks to detect and attribute causes for anthropogenic climate change in general or greenhouse climate change in particular. We focus on the latter because we argue that it is a less trivial problem and has more direct implications for future climate changes. The selection of lines of evidence to represent climate change on the time scale of greenhouse forcing and response (the past century) is subjective and approximate and is limited by the availability of suitable observations. Similarly, the characterization of natural variability for each line of evidence is limited by the lack of data and entails an element of choice, whether characterized by modellers or experts. Carrying out attribution of cause for

detected climate changes requires a characterization of climate forcings and of the translation of forcing to response. The uncertainty in the latter step is difficult to include in attribution studies and is rarely made explicit. The attributions of cause for each line of evidence have 2 components: the fraction of signal that is attributed to the forcing, and the probability associated with that fraction. It is useful to specify the fraction attributed, since one can have a high probability of attribution to a vanishingly small contribution of the forcing to the postulated climate change—revealing little about the role of that forcing.

The judgements entailed in detection and attribution studies are often complex. When experts are used to make these judgements, some cognitive limitations are apparent. For example, it is difficult to make conditional attributions to different forcings when uncertainty in each of the forcings is represented by pdfs. These judgements can be rendered tractable by approximating the forcing pdfs by discrete ranges. When experts make attributions conditional on the forcing, the contributions of both forcing uncertainty and uncertainty in climate sensitivity must be weighed. Neither of these uncertainties has been given much attention in detection and attribution studies, and it is not surprising that experts are divided over the relative importance of these uncertainties.

Another important issue for detection and attribution studies is the role of dependence among lines of evidence. This comes into play when combining attributions from separate lines of evidence into net measures of attribution. The higher the level of dependence among lines of evidence, the lower the net attribution to a given forcing. The 'burden of proof' for different possible causes of climate change can be established by setting thresholds for each line of evidence. In the detection and attribution protocol presented here, attributions are measured in fractional terms. That is, the attribution assesses the relative contributions of each possible forcing candidate to the climate change (accounting for the role of natural variability). When attribution is measured in this way, it makes little sense to set thresholds at the 95% level. If we are concerned about future climate changes, we want to take account of the most important processes driving the change, whether the result of natural variability or externally forced. If the dominant process is both external and persistent (as in the case of greenhouse forcing), there are implications for the persistence of the change. Dominance among attributions need not be near total (95%) to be important.

The problem of detecting unusual environmental changes and attributing the changes to various uncertain causes is not unique to climate change. A range of environmental problems from endocrine disruption

(Colborn 1997) to the disappearance of frogs (Phillips 1995) share the problem of identifying probable causes from a range of potential natural and anthropogenic perturbations. It is conceivable that some of these problem domains could benefit from formalizations of the type we have described here.

*Acknowledgements.* We are grateful to all those who have given us feedback on the protocol described here. They include Mark Handel, Haroon Kheshgi, Benoit Morel, Stephen Schneider, Urmila Diwekar, Granger Morgan, Marina Pantazidou, Wendy Parker, Gerard Roe, and the audiences of seminars at Carnegie Mellon and Texas A & M Universities. We are indebted to all the experts who gave of their time and wisdom in completing the protocol. We are also grateful for the detailed comments of several anonymous reviewers. This work was funded by the NOAA Climate and Global Change Program. Figs. 1 & 3 and are gratefully reproduced with permission from Phil Jones (Fig. 1) and Keith Shine and Piers Forster (Fig. 3).

#### LITERATURE CITED

- Barnett T (1999) Comparison of near-surface air temperature variability in 11 coupled global climate models. *J Clim* 12(2):511–518
- Barnett T, Hasselmann K, Chelliah M, Delworth T, Hegerl G, Jones P, Rasmusson E, Roeckner E, Ropelewski C, Santer B, Tett S (1999) Detection and attribution of recent climate change: a status report. *Bull Am Meteorol Soc* 80(12): 2631–2659
- Briffa KR, Osborn TJ (1999) Perspectives: climate warming—seeing the wood from the trees. *Science* 284:926–927
- Chanin ML, Ramaswamy V (1999) Trends in stratospheric temperatures. In: *Scientific assessment of ozone depletion: 1998*. WMO Global Ozone Research and Monitoring Report No. 44, Geneva, p 5.1–5.59
- Christy J, Spencer R (1995) Assessment of precision in temperatures from the microwave sounding units. *Clim Change* 30(1):97–102
- Clark W (ed) (1982) *Carbon dioxide review*. Oxford University Press, Oxford
- Colborn T (1997) *Our stolen future: Are we threatening our fertility, intelligence, and survival?—a scientific detective story*. Plume, New York
- Edwards P, Schneider S (1997) The 1995 IPCC report: broad consensus or 'scientific cleansing'? *Ecofable/Ecosci* 1(1): 3–9
- Funtowicz S, Ravetz J (1994) The worth of a songbird: ecological economics as a post-normal science. *Ecol Econ* 10: 197–207
- Hasselmann K (1998) Conventional and Bayesian approach to climate-change detection and attribution. *Q J R Meteorol Soc* 124(552):2541–2565
- Hegerl GC, Hasselmann K, Cubasch U, Mitchell JFB, Roeckner E, Voss R, Waszkewitz J (1997) On multi-fingerprint detection and attribution of greenhouse gas- and aerosol forced climate change. *Clim Dyn* 13:613–634
- IPCC (1996) *Climate change 1995: The science of climate change*. Houghton et al. (eds). Cambridge University Press, Cambridge
- Jones PD (1994) Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *J Clim* 7: 1794–1802

- Jones PD, Wigley TML (1990) Global warming trends. *Sci Am* August, 84–91
- Jones PD, New M, Parker D, Martin S, Rigor I (1999) Surface air temperature and its changes over the past 150 years. *Rev Geophys* 37:173–199
- Karl T, Knight R (1998) Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull Am Meteorol Soc* 79 (2):231–241
- Karoly DJ, Cohen JA, Meehl GA, Mitchell JFB, Oort AH, Stouffer RJ, Wetherald RT (1994) An example of fingerprint detection of greenhouse climate change. *Clim Dyn* 10:97–105
- Karoly DJ, Braganza K, Hirst A, Power S (1999) Climate change detection and attribution using simple global indices. In: *Proceedings of the 10th Symposium on Global Change Studies*, 10–15 January 1999. American Meteorological Society, Boston, p 240–241
- Kelly PM, Wigley TML (1990) The influence of solar forcing trends on global mean temperature since 1861. *Nature* 347:460–462
- Kheshgi HS, White BS (2000) Testing distributed parameter hypotheses for the detection of climate change. *J Clim* (in press)
- Kiehl JT, Briegleb BP (1993) The relative roles of sulfate aerosols and greenhouse gases in climate forcing. *Science* 260:311–314
- Lorenz EN (1995) *Climate is what you expect*. MIT, Cambridge, MA
- Mann ME, Bradley RS, Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* 392:779–787
- Mann ME, Bradley RS, Hughes MK (1999) Northern hemisphere temperatures during the past millennium: inferences, uncertainties, and limitations. *Geophys Res Lett* 26 (6):759–762
- Mitchell JFB, Karoly DJ (2000) Detection of climate change and attribution of causes. Intergovernmental Panel on Climate Change Working Group 1 Third Assessment Report. Cambridge University Press, Cambridge
- Morgan MG, Henrion M (1990) *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, New York
- Moss R, Schneider SH (1999) Towards consistent assessment and reporting of uncertainties in the IPCC TAR: initial recommendations for discussion by authors. IPCC Guidance Paper distributed to lead authors. Cambridge University Press, Cambridge
- Parker D, Jones PD, Folland C, Bevan A (1994) Interdecadal changes of surface temperature since the late nineteenth century. *J Geophys Res* 99:14373–14399
- Pearce F (1999) Rivers of doubt. *New Scientist* 161(2174):48
- Peterson TC, Vose RS (1997) An overview of the global historical climatology network temperature data base. *Bull Am Meteorol Soc* 78:2837–2849
- Phillips K (1995) *Tracking the vanishing frogs: an ecological mystery*. Penguin, New York
- Ravetz J (1997) Integrated environmental assessment forum: developing guidelines for 'good practice'. Working paper ULYSSES, Darmstadt University of Technology
- Risbey J, Stone P (1996) A case study of the adequacy of GCM simulations for input to regional climate change assessments. *J Clim* 9(7):1441–1467
- Santer BD, Taylor K, Wigley TML, Penner J, Jones PD, Cubasch U (1995) Towards the detection and attribution of an anthropogenic effect on climate. *Clim Dyn* 12:77–100
- Santer B, Wigley TML, Barnett T, Anyamba E (1996a) Detection of climate change and attribution of causes. In: Houghton et al. (eds) *Climate change 1995: the science of climate change*. Cambridge University Press, Cambridge, p 406–443
- Santer B, Taylor K, Wigley TML, Johns T, Jones P, Karoly D, Mitchell JFB, Oort A, Penner J, Ramaswamy V, Schwarzkopf M, Stouffer R, Tett S (1996b) A search for human influences on the thermal structure of the atmosphere. *Nature* 382:39–46
- Seitz F (1996) A major deception on global warming. *Wall Street Journal*, June 12, New York
- Shine K, Forster PdeF (1999) The effect of human activity on radiative forcing of climate change. A review of recent developments. *Global Planet Change* 20(4):205–225
- Shrader-Frechette K (1984) Science policy, ethics, and economic methodology: some problems of technology assessment and environmental impact analysis. Reidel, Dordrecht
- Tett S, Mitchell JFB, Parker D, Allen M (1996) Human influence on the atmospheric vertical temperature structure: detection and observations. *Science* 274:1170–1173
- Varis O, Kuikka S (1997) BeNe-EIA: a Bayesian approach to expert judgment elicitation with case studies on climate change impacts on surface waters. *Clim Change* 37(3): 539–563
- Wigley TML (1999) The science of climate change: global and U.S. perspectives. Pew Center on Global Climate Change, Arlington, VA
- Wigley TML, Raper SCB (1990) Natural variability of the climate system and detection of the greenhouse effect. *Nature* 344:324–327
- Wigley TML, Smith R, Santer B (1998) Anthropogenic influence on the autocorrelation structure of hemispheric-mean temperatures. *Science* 282:1676–1679

*Editorial responsibility: Hans von Storch, Geesthacht, Germany*

*Submitted: November 19, 1999; Accepted: May 28, 2000  
Proofs received from author(s): October 13, 2000*