

NASA

When doubt is a sure thing

Is it possible to adopt a more rigorous approach to the communication of scientific uncertainty? Jim Giles talks to the climatologists whose pursuit of this goal has seen them dubbed the 'uncertainty cops'.

Stephen Schneider is very certain in his views about uncertainty. As a leading climate researcher at Stanford University in California, he has plenty of experience in dealing with it. Global temperatures are set to rise — but no one is sure by how much. And although this warming will influence variables from biodiversity to economic productivity, exactly how it will do so remains unclear.

Schneider cannot eliminate uncertainties from his work, but he has little time for fellow scientists who are sloppy in expressing just how unsure they are. He argues, for instance, that the phrase “low confidence” should have a precise, quantitative meaning. He espouses the use of graphical tools to illustrate where scientific uncertainties come from. And he is adamant that statistical confidence levels should be attached to even the most complex scientific predictions.

Together with Richard Moss, currently executive director of the Global Change Research Program in Washington DC, which coordinates the US government’s climate-change research, Schneider has been pestering his colleagues about these ideas for more than five years. With plans for a major new international climate assessment just getting started, Schneider and Moss are gearing up to pro-

mote their agenda once again. They also believe that other subjects that engender both public concern and scientific uncertainty could benefit from taking a similar approach.

Schneider and Moss began their campaign after witnessing the furore that greeted the 1995 publication of the second climate assessment by the Intergovernmental Panel on Climate Change (IPCC). The report was the first IPCC publication to state unequivocally that human activities are having a discernible impact on the Earth’s climate. Environmental pressure groups seized upon this statement, while lobbyists for the fossil-fuel industry and a minority of sceptical climate scientists complained bitterly that uncertainties in the science had been played down.

Dazed and confused

In part, claim Schneider and Moss, the rancour stemmed from the confused way in which the report dealt with uncertainties. Moss cites the example of its estimate of climate sensitivity — the increase in average global temperatures that is expected to occur if carbon dioxide concentrations were to reach twice their pre-industrial-revolution levels. This was given as a range of 1.5–4.5 °C. Environmental groups naturally quoted figures at the top of the range, which are

Clouded world view: data on future climate trends are shrouded in scientific uncertainty.

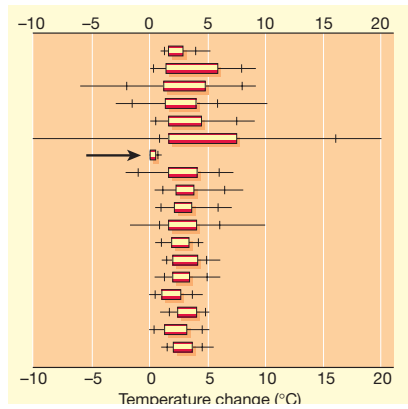
actually less likely to materialize than those near the middle. This point, which was often missed in the arguments that followed the report’s publication, would have been much clearer if probabilities had been assigned to different parts of the range, says Moss.

Reluctance to use probabilities in this way is common in climate-change research, and stems from the nature of the predictions involved. In related fields, results that can be calculated in a straightforward manner often have probabilities attached to them. Weather forecasters, for example, use computer models to predict whether certain atmospheric conditions will bring rain or shine. Because their models have been tested using real data from past weather systems, the forecasters know how accurate their predictions are.

But models that simulate long-term climate changes have no comparable reality check, so climate modellers often shy away from using probabilities. Instead, they qualify their results by discussing how well understood a model is, or by noting the availability of observational data. It was precisely this qualitative approach, argue Schneider and Moss, that led to the arguments over how the

USGS

J. ALLEN/MODIS/UNIV. MIAMI/RSMIS



Cops on patrol: Richard Moss (top, left) and Stephen Schneider want climatologists to use four-axis plots to show where the uncertainty in their predictions comes from. The examples (above) are for temperature change (left) and flooding risk. The pair also advocate plotting individual scientists' predictions (top right) — this example shows that one expert (arrowed) used different methods.

1995 IPCC report should be summarized.

Schneider and Moss tried to make sure that this mistake was not repeated in the IPCC's next report. In 1996, they held a session on the presentation of uncertainty at the Elements of Change conference at the Aspen Global Change Institute in Colorado. They also consulted risk-communication experts and pressured the IPCC's leaders to recognize the problem. The result was a paper¹, published by the two 'uncertainty cops' in 2000 as the IPCC prepared its huge third assessment for publication the following year.

Confidence boost

At the heart of Schneider and Moss's paper lay a call for researchers to use a common language when describing their results. The pair argued that the phrase "low confidence", for example, should only be used in connection with confidence ratings of between 5% and 33%. In all, Schneider and Moss recommended five categories, ranging from "very low confidence" (less than 5%) to "very high confidence" (95–100% confidence).

If a lack of data prevents uncertainty from being calculated using standard statistical techniques, Schneider and Moss suggested that authors should assign subjective probabilities to their results, taking into account their knowledge of the models and data behind them. This method, known as bayesian statistics, is not as woolly as it sounds. Initial confidence levels may be subjective, but they can be modified as more data about a system become available. The degree to which the initial assumption biases the final confidence level can also be assessed. "Bayesian

statistics allow scientists to indicate their degree of belief in a result given the information available to them," explains Moss.

For those who are uncomfortable with using subjective estimates, Schneider and Moss suggested a descriptive scale for assessing the "state of knowledge" about a result. The term "well established", for example, should describe a result for which the models involve known processes, the observations are largely consistent with the models, or the finding is supported by several lines of evidence.

Did the IPCC's authors listen? Well, yes and no. When the third assessment was published last year, the panel's Working Group I, which focuses on the atmospheric and oceanographic science behind climate change, embraced the terminology for describing confidence levels, and even added a new category — greater than 99% confidence was termed "virtually certain". Working Group II, which seeks to assess the impact of climate change on ecosystems and human activities, used the scale in its original form — perhaps unsurprisingly, as Schneider was one of the lead authors of the group's report. But Working Group III, which is charged with outlining techniques for tackling climate change, did not use the terminology at all.

Schneider ascribes this experience to differences in culture between the disciplines that make up the working groups. Many of the climate researchers of Working Group I were used to dealing with predictions that carried high degrees of confidence, and so were comfortable with the ratings — and were happy to

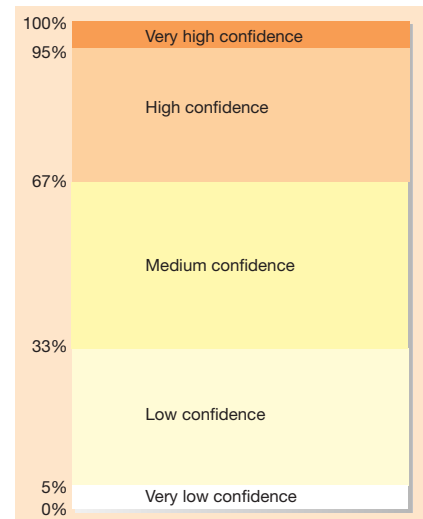
add another at the top of the scale. But such high confidence is rarely shared by the sociologists and economists who were involved in the other two working groups, hence their more circumspect approach.

Many of the IPCC's authors — particularly the members of Working Group I — were suspicious of bayesian statistics, preferring instead to rely on Schneider and Moss's qualitative scale when data from similar past events were not available. These attitudes also stem from the authors' backgrounds, argues Schneider. "The economists understood it was not irresponsible," he says. "But it was tougher for your average natural scientist." Mike Hulme, executive director of the Tyndall Centre for Climate Change Research at the University of East Anglia in Norwich, UK, agrees. "It is not a normal approach in the lab," he says. "But it is useful when communicating results to policy-makers."

Get in shape

Schneider and Moss have also developed graphical techniques to summarize how confidence, or lack of it, arises. One scheme involves plotting points on four axes — corresponding to confidence in the theory, the observations, the models and the consensus within a field — arranged like the points on a compass. The points are then joined to create a shape — its area indicates the overall degree of confidence in the result, and its outline describes how that confidence arises (see figure, above left). The technique was used by Mike Scott of the Pacific Northwest National Laboratory in Richland, Washington, in his section of Working Group II's report dealing with the impact of climate change on human settlements. "I could kiss him," says a clearly delighted Schneider.

Other graphical techniques in Schneider and Moss's paper were intended for use in the absence of clear consensus over a result — such as for climate-sensitivity estimates.



Schneider and Moss's proposed confidence bands.

Rather than averaging these estimates, Schneider and Moss proposed a graphical approach that was used in a 1995 paper by climate-policy experts Granger Morgan, of Carnegie Mellon University in Pittsburgh, and David Keith, then at Harvard University².

Morgan and Keith asked several climate experts to estimate the range of climate sensitivity, and found that one of them made a very different prediction from those of the others. Averaging these results would have obscured the fact that this researcher was using different assumptions and models to generate the prediction (see figure, previous page).

Morgan–Keith plots were not used in the IPCC’s 2001 report. But Schneider and Moss say that they will be pushing for their ideas to be used more consistently when researchers begin work on the panel’s fourth assessment, due to be published in 2007. Moss is particularly keen for researchers to adopt another of his proposals — he wants complex predictions depending on a series of results to come with a “traceable account” that includes details of what lines of evidence were used to generate each result, and the uncertainties associated with that evidence.

Estimates vary

But the uncertainty cops will have to overcome some resistance — as Schneider found last time around. Before producing its third report, the IPCC asked a group of academic researchers, environmentalists, industry representatives, engineers and economists to consider how emissions of greenhouse gases are likely to change over this century. These emissions scenarios were used in parts of the third assessment — Working Group I used them to produce estimates of how different emission levels would affect the climate, and Working Group II used the resulting estimates to consider the likely impact of the predicted climate changes.

At a meeting of the IPCC’s scenarios task group, held in 1997 at the International Institute for Applied Systems Analysis in Laxenburg, Austria, Schneider attempted to persuade delegates to assign relative probabilities to the emission scenarios they had conceived. The idea was controversial, as emissions depend on many factors, from new technologies to changes in population. Delegates held widely different views on these issues, and most believed that simply identifying the various scenarios was as much as they could do.

That’s all very well, says Schneider, but it means that policy-makers can assume that all emissions scenarios are equally likely. Special-interest groups can then use whichever scenario fits their agenda when making claims about climate change, and politicians can pick those that help to justify their policies. So Schneider says that he will continue to push for probabilities to be attached to the IPCC’s emissions scenarios.

Given that climate change is just one of



On shaky ground: researchers studying GM crops and seismic data are also plagued by uncertainty.

many issues that are plagued by scientific uncertainty, could any of the techniques espoused by Schneider and Moss find use in other fields? The use of bayesian statistics in clinical trials and toxicity assessments has been growing in recent years, as researchers incorporate prior knowledge into their analyses. Advocates of this approach argue that it can also help policy-makers in other areas. “I live ten kilometres away from a major fault,” says Elisabeth Paté-Cornell, an engineer and risk analyst at Stanford University. “I don’t have enough historical information to say when an earthquake is going to come. But bayesian statistics help make policy decisions when the perfect science is not available.”

Cunning plots

The plots pioneered by Morgan and Keith are also routinely used in medicine to compare results from different clinical trials. But using these plots in other fields could be difficult. Joe Perry, an ecologist at Rothamsted Research, an agricultural research institute in Harpenden, north of London, is working on large-scale studies of the impact of genetically modified (GM) crops on biodiversity. Morgan–Keith plots are interesting, says Perry, because they allow unlikely but worrying outcomes to be considered alongside more probable possibilities.

Consideration of such improbable but severe outcomes is important in the debate over GM crops, says Perry. “The public is very concerned about the small chance that something very serious will go wrong, such

as the possibility that introducing genes for virus resistance could lead to the development of super-weeds,” he argues. But he says that it is difficult to think of a single variable that can be used to represent these fears, as would be required for a Morgan–Keith plot. “The method may be useful for precisely phrased issues,” says Perry, “but GM debates have rarely restricted themselves in this way.”

Perry is also concerned about Schneider and Moss’s four-axis plots. He warns that the area of the plotted shape may not accurately represent the overall confidence in the result if the uncertainties associated with the different axes are not independent. “Theory comes from observations, and both of these feed into models, so the different axes may depend on each other,” says Perry. “If they are dependent, the size of the shape will not be representative of the total probability.”

Moss admits that the theory behind the four-axis plots is not as precise as that which underlies more established areas of statistics. He says that he and Schneider are “blazing a bit of a trail”, adding that “this kind of process is new”. Rather than seeing the plots as the finished article, Moss hopes that they will evolve after being used by other researchers.

Schneider accepts that he has some way to go in convincing his peers to embrace his approach to the communication of scientific uncertainty. But he certainly cannot be accused of failing to practise what he preaches. When he was diagnosed with lymphoma, Schneider says that he used the bayesian approach to help make decisions about his various treatment options. “My recovery is doing great,” he adds.

With many senior politicians still unconvinced by the IPCC’s consensus on global warming, Schneider and Moss believe that it is more important than ever to get the world’s climate scientists signed up to their uncertainty manifesto. “There is no other way to advise policy-makers,” claims Schneider. ■

Jim Giles is Nature’s associate news and features editor.

1. Moss, R. H. & Schneider, S. H. in *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC* (eds Pachauri, R., Taniguchi, T. & Tanaka, K.) 33–51 (World Meteorol. Org., Geneva, 2000).
2. Morgan, M. G. & Keith, D. *Env. Sci. Tech.* **29**, 468–476 (1995).