Global warming: Forecasts by scientists versus scientific forecasts

Green, Kesten C. and Armstrong, J. Scott

3 August 2007

Global Warming: Forecasts by Scientists versus Scientific Forecasts*

Kesten C. Green, Business and Economic Forecasting Unit, Monash University, Victoria 3800, Australia.
Contact: PO Box 10800, Wellington 6143, New Zealand.
kesten@kestencgreen.com; T +64 4 976 3245; F +64 4 976 3250

J. Scott Armstrong†, The Wharton School, University of Pennsylvania
747 Huntsman, Philadelphia, PA 19104, USA.
armstrong@wharton.upenn.edu

(This paper is a draft of an article that is forthcoming in *Energy and Environment*.)
Version 53 – August 3, 2007
We continue to work on this paper and we invite peer review

Abstract

In 2007, the Intergovernmental Panel on Climate Change's Working Group One, a panel of experts established by the World Meteorological Organization and the United Nations Environment Programme, issued its Fourth Assessment Report. The Report included predictions of dramatic increases in average world temperatures over the next 92 years and serious harm resulting from the predicted temperature increases. Using forecasting principles as our guide we asked: Are these forecasts a good basis for developing public policy? Our answer is "no."

To provide forecasts of climate change that are useful for policy-making, one would need to forecast (1) global temperature, (2) the effects of any temperature changes, (3) the effects of alternative policies, and (4) whether the best policy would be successfully implemented. Proper forecasts of all four are necessary for rational policy making.

The IPCC Report was regarded as providing the most credible long-term forecasts of global average temperatures by 31 of the 51 scientists and others involved in forecasting climate change who responded to our survey. We found no references to the primary sources of information on forecasting methods despite the fact these are easily available in books, articles, and websites. We audited the forecasting processes described in Chapter 8 of the IPCC's WG1 Report to assess the extent to which they complied with forecasting principles. We found enough information to make judgments on 89 out of a total of 140 forecasting principles. The forecasting procedures that were described violated 72 principles. Many of the violations were, by themselves, critical.

The forecasts in the Report were not the outcome of scientific procedures. In effect, they were the opinions of scientists transformed by mathematics and obscured by complex writing. Research on forecasting has shown that experts' predictions are not useful. We have been unable to identify any scientific forecasts of global warming. Claims that the Earth will get warmer have no more credence than saying that it will get colder.

Keywords: accuracy, audit, climate change, evaluation, expert judgment, mathematical models, public policy.

*Neither of the authors received funding for this paper.
† Information about J. Scott Armstrong can be found on Wikipedia.

“A trend is a trend,
But the question is, will it bend?
Will it alter its course
Through some unforeseen force
And come to a premature end?”
Alec Cairncross, 1969

Research on forecasting has been conducted since the 1930s. Empirical studies that compare methods in order to determine which ones provide the most accurate forecasts in given situations are the most useful source of evidence. Findings, along with the evidence, were first summarized in Armstrong (1978, 1985). In the mid-1990s, the forecasting principles project was established with the objective of summarizing all useful knowledge about forecasting. The knowledge was codified as evidence-based principles, or condition-action statements, in order to provide guidance on which methods to use when. The project led to the *Principles of Forecasting* handbook (Armstrong 2001): the work of 40 internationally-known experts on forecasting methods and 123 reviewers who were also leading experts on forecasting methods. The summarizing process alone required a four-year effort.

The forecasting principles are easy to find: They are freely available on forecastingprinciples.com, a site sponsored by the International Institute of Forecasters. The Forecasting Principles site has been at the top of the list of sites in internet searches for “forecasting”, for many years. A summary of the principles, currently numbering 140, is provided as a checklist in the Forecasting Audit software available on the site. There is no other source that provides evidence-based forecasting principles. The site is often updated as evidence on forecasting comes to hand. A recent review of new evidence on some of the key principles was published in Armstrong (2006).

The strength of evidence is different for different principles, for example some principles are based on common sense or received wisdom. Such principles are included when there is no contrary evidence. Other principles have some empirical support, while 31 are strongly supported by empirical evidence.

Many of the principles go beyond common sense, and some are counter-intuitive. As a result, those who forecast in ignorance of the forecasting research literature are unlikely to produce useful predictions. For example, here are some well-established principles that apply to long-term forecasts for situations involving of complex issues where the causal factors are subject to uncertainty (as with climate):

- *Unaided judgmental forecasts by experts have no value.* This applies whether the opinions are expressed in words, spreadsheets, or mathematical models. It also applies regardless of how much scientific evidence is possessed by the experts. Among the reasons for this are:
  - a) Complexity: People cannot assess complex relationships through unaided observations.
  - b) Coincidence: People confuse correlation with causation.
  - c) Feedback: People making judgmental predictions typically do not receive unambiguous feedback they can use to improve their forecasting.
  - d) Bias: People have difficulty in obtaining or using evidence that contradicts their initial beliefs. This problem is especially serious for people who view themselves as experts.

- *Agreement among experts is weakly related to accuracy*. This is especially true when the experts communicate with one another and when they work together to solve problems, as is the case with the IPCC process.
- *Complex models (those involving nonlinearities and interactions) harm accuracy because their errors multiply.* Ascher (1978), refers to the Club of Rome's 1972 forecasts where, unaware of the research on forecasting, the developers proudly proclaimed, "in our model about 100,000 relationships are stored in the computer. Complex models also tend to fit random variations in historical data well, with the consequence that they forecast poorly and provide misleading conclusions about the uncertainty of the outcome. Finally, when complex models are developed there are many opportunities for errors and the complexity means the errors are difficult to find. Craig, Gadgil, and Koomey (2002) came to similar conclusions in their review of long-term energy forecasts for the US made between 1950 and 1980.
- *Given even modest uncertainty, prediction intervals are enormous*. For example, prediction intervals (ranges outside which outcomes are unlikely to fall) expand rapidly as time horizons increase, so that one is faced with enormous intervals even when trying to forecast a straightforward thing such as automobile sales for General Motors over the next five years.
- *When there is uncertainty in forecasting, forecasts should be conservative*. Uncertainty arises when data contain measurement errors, when the series are unstable, when knowledge about the direction of relationships is uncertain, and when a forecast depends upon forecasts of related (causal) variables. For example, forecasts of no change were found to be more accurate than trend forecasts for annual sales when there was substantial uncertainty in the trend lines (e.g., Schnaars and Bavuso 1986). This principle also implies that forecasts should revert to long-term trends when such trends have been firmly established, do not waver, and there are no firm reasons to suggest that they will change. Finally, trends should be damped toward no-change as the forecast horizon increases.

The Forecasting Problem

In determining the best policies to deal with the climate of the future, a policy maker first has to select an appropriate statistic to use to represent the changing climate. By convention, the statistic is the averaged global temperature as measured with thermometers at ground stations throughout the world, though in practice this is a far from satisfactory metric (see, e.g., Essex et al., 2007).

It is then necessary to obtain forecasts and prediction intervals for each of the following:

1. *Mean global temperature in the long-term (say 20 years or longer).*
2. *Effects of temperature changes on humans and other living things.*
   If accurate forecasts of mean global temperature can be obtained *and* the changes are substantial, *then* it would be necessary to forecast the effects of the changes on the health of living things and on the health and wealth of humans. The concerns about changes in global mean temperature are based on the assumption that the earth is currently at the optimal temperature and that variations over years (unlike variations within days and years) are undesirable. For a proper assessment, costs and benefits must be comprehensive. (For example, policy responses to Rachel Carson's *Silent*

Spring should have been based in part on forecasts of the number of people who might die from malaria if DDT use were reduced).

3. *Costs and benefits of alternative policy proposals.*
   If reliable forecasts of the effects of the temperature changes on the health of living things and on the health and wealth of humans can be obtained *and* the forecasts are for substantial harmful effects, *then* it would be necessary to forecast the costs and benefits of alternative policy proposals.
4. *Whether the policy changes can be implemented successfully.*
   If reliable forecasts of the costs and benefits of alternative policy proposals can be obtained *and* at least one proposal is predicted to lead to net benefits, *then* it would be necessary to forecast whether the policy changes can be implemented successfully.

A policy proposal should only be implemented *if* reliable forecasts of policy implementation can be obtained *and* the forecasts show net benefits from the policy, *and* the policy can be successfully implemented. A failure to obtain scientifically validated forecasts at any stage would render subsequent stages irrelevant. Thus, we focus on the first of the four forecasting problems.

Is it necessary to use scientific forecasting methods? In other words, to use methods that have been shown by empirical validation to be relevant to the types of problems involved with climate forecasting? Or is it sufficient to have leading scientists examine the evidence and make forecasts? We address this issue before moving on to our audits.

On the value of forecasts by experts

Many public policy decisions are based on forecasts by experts. Research on persuasion has shown that people have substantial faith in the value of such forecasts. Faith increases when experts agree with one another.

Our concern is with what we refer to as unaided expert judgments. In such cases, experts may have access to empirical studies and other information, but they use their knowledge to make predictions without the aid of well-established forecasting principles. Thus, they could simply use the information to come up with judgmental forecasts. Alternatively, they could translate their beliefs into mathematical statements (or models) and use those to make forecasts.

Although they may seem convincing at the time, expert forecasts make for humorous reading in retrospect. Cerf and Navasky's (1998) book contains 310 pages of examples, such as Fermi Award-winning scientist John von Neumann's 1956 prediction that "A few decades hence, energy may be free". Examples of expert climate forecasts that turned out to be completely wrong are easy to find, such as UC Davis ecologist Kenneth Watt's prediction in a speech at Swarthmore College on Earth Day, April 22, 1970:

> If present trends continue, the world will be about four degrees colder in 1990, but eleven degrees colder in the year 2000. This is about twice what it would take to put us into an ice age.

Are such examples merely a matter of selective perception? The second author's review of empirical research on this problem led him to develop the "Seer-sucker theory," which can be stated as "No matter how much evidence exists that seers do not exist, seers will find suckers" (Armstrong 1980). The amount of expertise does not matter beyond a basic minimum level. There are exceptions to the Seer-sucker Theory: When experts get substantial well-summarized feedback about the accuracy of their forecasts and about the reasons why their forecasts were or were not accurate, they can improve their forecasting. This situation applies for short-term (up to five day) weather forecasts, but we are not aware of any such regime for long-term global climate forecasting. Even if there were such a regime, the feedback would trickle in over many years before it became useful for improving forecasting.

Research since 1980 has added support to the Seer-sucker Theory. In particular, Tetlock (2005) recruited 284 people whose professions included, "commenting or offering advice on political and economic trends." He asked them to forecast the probability that various situations would or would not occur, picking areas (geographic and substantive) within and outside their areas of expertise. By 2003, he had accumulated over 82,000 forecasts. The experts barely if at all outperformed non-experts and neither group did well against simple rules.

Comparative empirical studies have routinely concluded that judgmental forecasting by experts is the least accurate of the methods available to make forecasts. For example, Ascher (1978, p. 200), in his analysis of long-term forecasts of electricity consumption found that was the case.

Experts' forecasts of climate changes have long been popular. Anderson and Gainor (2006) found the following headlines in their search of the *New York Times*:

Sept. 18, 1924	"MacMillan Reports Signs of New Ice Age"
March 27, 1933	"America in Longest Warm Spell Since 1776"
May 21, 1974	"Scientists Ponder Why World's Climate is Changing: A Major Cooling Widely Considered to be Inevitable"
Dec. 27, 2005	"Past Hot Times Hold Few Reasons to Relax About New Warming"

In each case, the forecasts were made with a high degree of confidence.

In the mid-1970s, there was a political debate raging about whether the global climate was changing. The United States' National Defense University (NDU) addressed this issue in their book, *Climate Change to the Year* 2000 (NDU 1978). This study involved nine man-years of effort by Department of Defense and other agencies, aided by experts who received honoraria, and a contract of nearly $400,000 (in 2007 dollars). The heart of the study was a survey of experts. It provided them with a chart of "annual mean temperature, 0-80^0 N. latitude," that showed temperature rising from 1870 to early 1940 then dropping sharply up to 1970. The conclusion, based primarily on 19 replies weighted by the study directors, was that while a slight increase in temperature might occur, uncertainty was so high that "the next twenty years will be similar to that of the past" and the effects of any change would be negligible. Clearly, this was a forecast by scientists, not a scientific forecast. However, it proved to be quite influential. The report was discussed in The Global 2000 Report to the President (Carter) and at the World Climate Conference in Geneva in 1979.

The methodology for climate forecasting used in the past few decades has shifted from surveys of experts' opinions to the use of computer models. However, based on the explanations that we have seen, such models are, in effect, mathematical ways for the experts to express their opinions. To our knowledge, there is no empirical evidence to suggest that presenting opinions in mathematical terms rather than in words will contribute to forecast accuracy. For example, Keepin and Wynne (1984) wrote in the summary of their study of the International Institute for Applied Systems Analysis's "widely acclaimed" projections for global energy that, "Despite the appearance of analytical rigour… [they] are highly unstable and based on informal guesswork". Things have changed little since the days of Malthus in the 1800s. Malthus forecast mass starvation. He expressed his opinions mathematically. His mathematical model predicted that the supply of food would increase arithmetically while the human population grew at a geometric rate and went hungry.

International surveys of climate scientists from 27 countries, obtained by Brat and von Storch in 1996 and 2003, were summarized by Bast and Taylor (2007). Many scientists were skeptical about the predictive validity of climate models. Of more than 1,060 respondents, 35% agreed with the statement, "Climate models can accurately predict future climates," and 47% percent disagreed. Members of the general public were also divided. An Ipsos Mori poll of 2,031 people aged 16 and over found that 40% agreed that "climate change was too complex and uncertain for scientists to make useful forecasts" while 38% disagreed (Eccleston 2007).

Trenberth (2007) has claimed that the IPCC does not provide forecasts but rather presents scenarios or "projections." As best as we can tell, these terms are used by the IPCC authors to indicate that they provide "conditional forecasts." As it happens, the word "forecast" and its derivatives occurred 37 times, and "predict" and its derivatives occurred 90 times in the body of Chapter 8. Recall also that most of our respondents (29 of whom were IPCC authors or reviewers) nominated the IPCC report as the most credible source of forecasts (not "scenarios" or "projections") of global average temperature. We conclude that the IPCC does provide forecasts and that these forecasts are informed by the modelers' experience and by their models—but they are unaided by the application of forecasting principles.

An examination of climate forecasting methods

We searched for prior reviews of long-term climate forecasting processes and found nine independent reviews. We also assessed the extent to which those who have made climate forecasts used evidence-based forecasting procedures. We did this by conducting Google searches. We then conducted a "forecasting audit" of the forecasting process behind the IPCC forecasts. The key aspects of a forecasting audit that can be used to identify ways to improve the audited forecasting process are to:

- examine all elements of the forecasting process,
- use principles that are supported by evidence, or are self-evidently true and unchallenged by evidence, against which to judge the forecasting process,
- rate the forecasting process against each principle, preferably using more than one independent rater,
- disclose the audit.

To our knowledge, no one has ever published a paper that is based on a forecasting audit, as defined here. We suggest that for forecasts involving important public policies, such audits should be expected and perhaps even-required. In addition, they should be fully disclosed with respect to who did the audit, what biases might be involved, and what were the detailed findings from the audit.

Reviews of climate forecasts

We could not find any comprehensive reviews of climate forecasting efforts. With the exception of Stewart and Glantz (1985), the reviews did not refer to evidence-based findings. None of the reviews provided explicit ratings of the processes and, again with the exception of Stewart and Glantz, little attention was given to full disclosure of the reviewing process. Finally, some reviews ignored the forecasting methods and focused on the accuracy of the forecasts.

Stewart and Glantz (1985) conducted an audit of the National Defense University (NDU 1978) forecasting process that we described above. They were critical of the report because it lacked an awareness of proper forecasting methodology. Their audit was hampered because the organizers of the study said that the raw data had been destroyed and a request to the Institute for the Future about the sensitivity of the forecasts to the weights went unanswered. Judging from a Google Scholar search, climate forecasters have paid little attention to this paper.

Carter, et al. (2006) examined the *Stern Review* (Stern 2007). They concluded that the authors of the *Report* made predictions without reference to scientific validation and without proper peer review.

Pilkey and Pilkey-Jarvis (2007) concluded that the long-term climate forecasts they examined were based only on the opinions of the scientists. The scientists' opinions were expressed in complex mathematical terms without any evidence on the validity of chosen approach. The authors provided the following quotation on their page 45 to summarize their assessment: "Today's scientists have substituted mathematics for experiments, and they wander off through equation after equation and eventually build a structure which has no relation to reality (Nikola Telsa, inventor and electrical engineer, 1934)." While it is sensible to be explicit about beliefs and to formulate these in a model, forecasters must also demonstrate that the relationships are valid.

Carter (2007) examined evidence on the predictive validity of the general circulation models (GCMs) used by the IPCC scientists. He found that while the models included some basic principles of physics, scientists had to make "educated guesses" about the values of many parameters because knowledge about the physical processes of the earth's climate is incomplete. In practice, the GCMs failed to predict recent global average temperatures as accurately as simple curve-fitting approaches (Carter 2007, pp. 64 – 65). They also forecast greater warming at higher altitudes in the tropics when the opposite has been the case (p. 64). Further, individual GCMs produce widely different forecasts from the same initial conditions and minor changes in parameters can result in forecasts of global cooling (Essex and McKitrick, 2002). Interestingly, when models predict global cooling, the forecasts are often rejected as "outliers" or "obviously wrong" (e.g., Stainforth et al., 2005).

Roger Pielke Sr. gave an assessment of climate models in a 2007 interview (available at http://climatesci.colorado.edu/2007/04/30/interview-by-marcel-crok-of-roger-a-pielke-sr-jan-2007/):

> You can always reconstruct after the fact what happened if you run enough model simulations. The challenge is to run it on an independent dataset, say for the next five years. But then they will say "the model is not good for five years because there is too much noise in the system". That's avoiding the issue then. They say you have to wait 50 years, but then you can't validate the model, so what good is it?
>
> …Weather is very difficult to predict; climate involves weather plus all these other components of the climate system, ice, oceans, vegetation, soil etc. Why should we think we can do better with climate prediction than with weather prediction? To me it's obvious, we can't!
>
> I often hear scientists say "weather is unpredictable, but climate you can predict because it is the average weather". How can they prove such a statement?

Bellamy and Barrett (2007) found serious deficiencies in the general circulation models described in the IPCC's *Third Assessment Report*. In particular, the models (1) produced very different distributions of clouds and none was close the actual distribution of clouds, (2) parameters for incoming radiation absorbed by the atmosphere and for that absorbed by the Earth's surface varied considerably, (3) did not accurately represent what is known about the effects of CO_2 and could not represent the possible positive and negative feedbacks about which there is great uncertainty. The authors concluded:

> The climate system is a highly complex system and, to date, no computer models are sufficiently accurate for their predictions of future climate to be relied upon. (p. 72)

Trenberth (2007), a lead author of Chapter 3 in the IPCC WG1 report wrote in a *Nature.com* blog "… the science is not done because we do not have reliable or regional predictions of climate."

Taylor (2007) compared seasonal forecasts by New Zealand's National Institute of Water and Atmospheric Research (NIWA) with outcomes for the period May 2002 to April 2007. He found NIWA's forecasts of average regional temperatures for the season ahead were, at 48% correct., which was no more accurate than chance. That this is a general result was confirmed by New Zealand climatologist Dr Jim Renwick, who observed that NIWA's low success rate was comparable to that of other forecasting groups worldwide. He added that "Climate prediction is hard, half of the variability in the climate system is not predictable, and so we don't expect to do terrifically well." Dr Renwick is an author on Working Group I of the IPCC 4th Assessment Report, and also serves on the World Meteorological Organization Commission for Climatology Expert Team on Seasonal Forecasting. His expert view is that current GCM climate models are unable to predict future climate any better than chance (New Zealand Climate Science Coalition 2007).

Similarly, Vizard, Anderson, and Buckley (2005) found seasonal rainfall forecasts for Australian townships were insufficiently accurate to be useful to intended consumers such as farmers planning for feed requirements. The forecasts were released only 15 days ahead of each three month period.

A survey to identify the most credible long-term forecasts of global temperature

We surveyed scientists involved in long-term climate forecasting and policy makers. Our primary concern was to identify the most important forecasts and how those forecasts were made. In particular, we wished to know if the most widely accepted forecasts of global average temperature were based on the opinions of experts or on scientific forecasting methods. Given the findings of our review of reviews of climate forecasting, conclusion from our Google search that many scientists are unaware of evidence-based findings related to forecasting methods, we expected that the forecasts would be based on the opinions of scientists.

We sent a questionnaire to experts who had expressed diverse opinions on global warming. We generated lists of experts by identifying key people and asking them to identify others. (The lists are provided in Appendix A.) Most (70%) of the 240 experts on our lists were IPCC reviewers and authors.

The questionnaire asked the experts to provide references for what they regarded as the most credible source of long-term forecasts of mean global temperatures. We strove for simplicity to minimize resistance to our request. Even busy people should have time to send a few references, especially if they believe that it is important to evaluate the quality of the forecasts that will influence major decisions. We asked:

> "We want to know which forecasts people regard as the most credible and how those forecasts were derived…
>
> > In your opinion, which scientific article is the source of the most credible forecasts of global average temperatures over the rest of this century?"

We received useful responses from 51 people, 42 of whom provided references to what they regarded as credible sources of long-term forecasts of mean global temperatures. Interestingly, eight respondents provided references in support of their claims that no credible forecasts exist.

Of the 42 expert respondents who were associated with global warming views, 30 referred us to the IPCC's report.

Based on the replies to this survey, it was clear that the IPCC's Working Group 1 Report contained the forecasts that are viewed as most credible by the bulk climate community. These forecasts are contained in Chapter 10 of the Report and the models that are used to forecast climate are assessed in Chapter 8, "Climate Models and Their Evaluation" (Randall et al. 2007). Chapter 8 provided the most useful information on the forecasting process used by the IPCC to derive forecasts of mean global temperatures, so we audited that chapter.

We also posted calls on email lists and on the forecastingprinciples.com site asking for help from those who might have any knowledge about scientific climate forecasts. This yielded few responses, only one of which provided relevant references.

A forecasting audit for global warming

In order to audit the forecasting processes described in Chapter 8 of the IPCC's report, we each read it prior to any discussion. Chapter 8 was, in our judgment, poorly written. The writing showed little concern for the target readership. It provided extensive detail on items that are of little interest in judging the merits of the forecasting process, provided references without describing what readers might find, and imposed an incredible burden on readers by providing 788 references. We found the Chapter difficult to read. In addition, the Chapter reads in places like a sales brochure. In the three-page executive summary, the terms, "new" and "improved" and related derivatives appeared 17 times. Most significantly, the chapter omitted key details on the assumptions and the forecasting process that were used. If the authors used a formal structured procedure to assess the forecasting processes, this was not evident.

We each made a formal, independent audit of IPCC Chapter 8 in May 2007. To do so, we used the Forecasting Audit software on the forecastingprinciples.com site, which is based on material originally published in Armstrong (2001). To our knowledge, it is the only evidence-based tool for evaluating forecasting procedures.

While Chapter 8 required many hours to read, it took us each about one hour to rate the forecasting approach described in the Chapter using the Audit software. We have each been involved with developing the Forecasting Audit program, so other users would likely require much more time. Ratings are on a 5-point scale from -2 to +2. A rating of +2 indicates the forecasting procedures were consistent with a principle, and a rating of -2 indicates failure to comply with a principle. The Audit software also has options to indicate that there is insufficient information to rate the procedures or that the principle is not relevant to a particular forecasting problem.

Our initial overall average ratings were similar at -1.37 and -1.35. We compared our individual ratings for individual principles and discussed inconsistencies. In some cases we averaged the ratings, truncating toward zero. In other cases we decided that there was insufficient information or that the information was too ambiguous to rate with confidence. Our final ratings are fully disclosed in the Special Interest Group section of the forecastingprinciples.com site that is devoted to Public Policy (publicpolicyforecasting.com).

Of the 140 principles in the Forecasting Audit, we judged that 127 were relevant for auditing the forecasting problem addressed in Chapter 8. The Chapter provided insufficient information to rate the forecasting procedures that were used against 38 of these principles. For example, we did not rate the Chapter against Principle 10.2, "Use all important variables." At least in part, our difficulty in auditing the Chapter was due to the fact that it was abstruse. It was sometimes difficult to know whether the information we sought was present or not.

Of the 89 forecasting principles that we were able to rate, the Chapter violated 72. Adherence to some of the key principles is necessary for forecasts to be valid. We address four such principles, all based on strong empirical evidence: violation of any one of them would render the

IPCC climate forecasts invalid. All four of these key principles were violated by the forecasting procedures described in IPCC Chapter 8. We key these principles to their numbering in the Forecasting Audit software.

Consider whether the events or series can be forecasted (Principle 1.4)

This principle refers to whether a forecasting method can be used that would do better than a naïve method. A common naïve method is to assume that things will not change.

Interestingly, naïve methods are often strong competitors with more sophisticated alternatives. This is especially so when there is much uncertainty. To the extent that uncertainty is high, forecasters should emphasize the naïve method. (This is illustrated by regression model coefficients: when uncertainty increases, the coefficients tend towards zero.) Departures from the naïve model tend to increase forecast error when uncertainty is high.

In our judgment, the uncertainty in forecasting global mean temperature is extremely high. For example, there is controversy among climate scientists over the current trend. One researcher, Carter (2007, p. 67) wrote:

> …the slope and magnitude of temperature trends inferred from time-series data depend upon the choice of data end points. Drawing trend lines through highly variable, cyclic temperature data or proxy data is therefore a dubious exercise. Accurate direct measurements of tropospheric global average temperature have only been available since 1979, and they show no evidence for greenhouse warming. Surface thermometer data, though flawed, also show temperature stasis since 1998.

Global climate is complex. Scientific evidence on many key relationships is weak or absent; e.g., does increased CO_2 in the atmosphere cause high temperatures or do high temperatures increase CO_2 (e.g. Jaworowski 2007)? Measurements of key variables such as local temperatures and a representative global temperature are contentious in the case of modern measurements, because of the distribution of weather stations and possible artifacts such as the urban heat island effect, and often speculative in the case of ancient ones, such as those climate proxies derived from tree ring and ice-core data (Carter 2007). Finally, it is difficult to forecast the causal variables. Stott and Kettleborough (2002, p. 723) summarize:

> Even with perfect knowledge of emissions, uncertainties in the representation of atmospheric and oceanic processes by climate models limit the accuracy of any estimate of the climate response. Natural variability, generated both internally and from external forcings such as changes in solar output and explosive volcanic eruptions, also contributes to the uncertainty in climate forecasts.

The already high level of uncertainty rises rapidly as the forecast horizon increases.

While the authors of Chapter 8 claim that the forecasts of global mean temperature are well-founded, their language is imprecise and relies heavily on such words as “generally,” “reasonable well,” “widely,” and “relatively” [to what?]. The report makes many explicit references to uncertainty. For example, the phrases “. . . it is not yet possible to determine which estimates of the climate change cloud feedbacks are the most reliable” and “Despite advances since the TAR, substantial uncertainty remains in the magnitude of cryospheric feedbacks within AOGCMs” appear on p. 593. In discussing the modeling of temperature, the authors wrote, “The extent to which these systematic model errors affect a model’s response to external perturbations is unknown, but may be significant” (p. 608), and, “The diurnal temperature range… is generally too small in the models, in many regions by as much as 50%" (p. 609), and "It is not yet known why models generally underestimate the diurnal temperature range." The following words and phrases appear at least once in the Chapter: unknown, uncertain, unclear, not clear, disagreement, uncertain, not fully understood, appears, not well observed, variability, variety, difference, unresolved, not resolved, and poorly understood.

Given the high uncertainty regarding climate, the appropriate naïve method for this situation would be the "no-change" model. Prior evidence on forecasting methods suggests that attempts to improve upon the naïve model might increase forecast error. To reverse this conclusion, one would have to produce validated evidence in favor of alternative methods. Such evidence is not provided in Chapter 8 of the IPCC report.

We are not suggesting that climate change cannot be forecast, only that this has yet to be demonstrated. Methods such as the naïve model with drift, rule-based forecasting, well-specified simple causal models, and combined forecasts might prove useful. The methods are discussed in Armstrong (2001). To our knowledge, their application to long-term climate forecasting has not been examined to date.

Keep forecasting methods simple (Principle 7.1)

Complex methods involve the use of a large number of variables in forecasting models, complex interactions, and relationships that employ nonlinear parameters. Complex forecasting methods are only accurate when there is little uncertainty about the relationships now and in the future, where the data are subject to little error, and where the causal variables can be accurately forecasted. These conditions do not apply to climate forecasting. Thus, simple methods are recommended. We gained the impression from the IPPC chapters and from related papers that climate forecasters generally believe that complex models are necessary for forecasting climate and that forecast accuracy will increase with model complexity.

The use of complex models when uncertainty is high is at odds with the evidence from forecasting research (e.g., Allen and Fildes 2001, Armstrong 1985, Duncan, Gorr and Szczypula 2001, Wittink and Bergestuen 2001). The use of complex methods makes criticism difficult and prevents forecast users from understanding how forecasts were derived. One effect of this exclusion of others from the forecasting process is to reduce the chances of detecting errors.

Do not use fit to develop the model (Principle 9.3)

It was not clear to us to what extent the models described in Chapter 8 (or in Chapter 9 by Hegerl et al. 2007) are either based on, or have been tested against, sound empirical data. However, some statements were made about the ability of the models to fit historical data, after tweaking their parameters. Extensive research has shown that the ability of models to fit historical data has little relationship to forecast accuracy (See "Evaluating forecasting methods" in Armstrong 2001.) It is well known that fit can be improved by making a model more complex. The typical consequence of increasing complexity to improve fit, however, is to decrease the accuracy of forecasts.

Use out-of-sample (ex ante) error measures (Principle 13.26)

Chapter 8 did not provide evidence on the accuracy of *ex ante* long-term forecasts from the models used to generate the IPCC's forecasts of climate change. It would have been feasible to assess the accuracy of alternative forecasting methods for short- and medium-term forecasts by using "successive updating." This involves withholding data on a number of years, then providing forecasts for one-year ahead, then two-years ahead, and so on up to, say, 20 years. The actual years could be disguised during these validation procedures. Furthermore, the years could be reversed (without telling the forecasters) to assess back-casting accuracy. If, as is suggested by forecasting principles, the models were unable to improve on the accuracy of forecasts from the naïve method in such tests, there would be no reason to suppose that accuracy would improve for longer forecasts. "Evaluating forecasting methods" in Armstrong 2001 provides evidence on this principle.

Summary of audit findings

A list of the 72 violations of forecasting principles by the IPCC forecasting procedures is provided on the Public Policy Special Interest Group Page at forecastingprinciples.com. The many violations provide further evidence that the IPCC authors were unaware of evidence-based principles for forecasting. If they were aware of them, it would have been incumbent on them to present evidence to justify their departures from the principles. They did not do so. We conclude that because the forecasting processes examined in Chapter 8 overlook scientific evidence on forecasting, the IPCC forecasts of climate change are not scientific.

We invite others to provide evidence-based audits of Chapter 8. These can be posted on web sites to ensure that readers have access to the audits. As with peer review, we will require all relevant information on the people who conduct the audits prior to posting the audits. Prior to the publication of this paper, we invited other researchers, using messages to email lists and web sites, to replicate our audit by providing their own ratings. In addition, we asked for information about any relevant principles that have not been included in the Forecasting Audit. At the time of writing, we have received neither alternative ratings nor evidence for additional relevant principles.

Climate change forecasters and their clients should use the Forecasting Audit early and often. Doing so would help to ensure that they are using appropriate forecasting procedures. Outside evaluators should also be encouraged to conduct audits. The audit reports should be made available to both the sponsors of the study and the public by posting on an open web site such as publicpolicyforecasting.com.

Climate forecasters' use of the scientific literature on forecasting methods

Between April and July 2007, we tried to assess the extent to which climate modelers relied on scientific studies on the proper use of forecasting methods. In one approach, we used the Advanced Search function of Google Scholar to get a general sense of the extent to which climate forecasters refer to scientific studies on forecasting. When we searched for "global warming" and "forecasting principles," we found no relevant sites. Nor did we find any relevant sites for "forecastingprinciples.com" and "global warming." Nor were there any relevant citations for the relevant-sounding paper, "Forecasting for Environmental Decision-Making" (Armstrong 1999) published in a book with a relevant title: *Tools to Aid Environmental Decision Making*. A search for "global warming" and the best selling textbook on forecasting methods (Makridakis et al. 1998) revealed two citations, neither related to the prediction of global mean temperatures. Finally, there were no citations of the meta-analysis of research on econometric models by Allen and Fildes (2001).

Using the titles of the papers, we independently examined the references in Chapter 8. The Chapter contained 788 references. Of these, none had any apparent relationship to forecasting methodology. Our examination was not difficult as most papers had titles such as, "Using stable water isotopes to evaluate basin-scale simulations of surface water budgets," and, "Oceanic isopycnal mixing by coordinate rotation."

We also examined the 535 references in Chapter 9. Of these, 17 had titles that suggested the article might be concerned at least in part with forecasting methods. When we inspected the 17 articles, we found that none of them referred to the scientific literature on forecasting methods.

It is difficult to understand how scientific forecasting could be conducted without reference to the research literature on how to make forecasts. One would expect to see empirical justification for the forecasting methods that were used.

Conclusions

To provide forecasts of climate change that are useful for policy-making, one would need to prepare forecasts of (1) temperature changes, (2) the effects of any temperature changes, (3) the effects of proposed policy changes, and (4) whether the best policies would be successfully implemented. To justify policy changes based on climate change, policy makers need scientific forecasts for all four forecasting problems and they need those forecasts to show net benefits flowing from proposed policies. If governments implement policy changes without such justification, they are likely to cause harm to many people.

We have shown that failure occurs with the first forecasting problem: predicting temperature over the long term. Specifically, we have been unable to find a scientific forecast to support the currently widespread belief in "global warming." Climate is complex and there is much uncertainty about causal relationships and data. Prior research on forecasting suggests that in such situations a naïve (no change) forecast would be superior to current predictions. Note that recommending the naïve forecast does not mean that we believe that climate will not change; it means that we are not convinced that current knowledge about climate is sufficient to make useful long-term forecasts about climate. Policy proposals should be assessed on that basis.

Many policies have been proposed in association with claims of global warming. It is not our purpose in this paper to comment on specific policy proposals, but it should be noted that policies may be valid regardless of future climate. To assess this, it would be necessary to directly forecast costs and benefits assuming that climate does not change or, even better, to forecasts costs and benefits under a range of possible future climates.

Based on our literature searches, those forecasting long-term climate change have no apparent knowledge of evidence-based forecasting methods, so we expect that similar conclusions would apply to the other three necessary parts of the forecasting problem.

Public policy makers owe it to the people who would be affected by their policies to base them on scientific forecasts. Advocates of policy changes have a similar obligation. We hope that in future climate scientists with diverse views will embrace forecasting principles in order to provide policy makers with scientific forecasts of climate.

Acknowledgements

We thank Kay A. Armstrong, Robert Carter, Alfred Cuzán, Robert Fildes, Paul Goodwin, David Henderson, Jos de Laat, Ross McKitrick, Kevin Trenberth, and Tom Yokum for helpful suggestions on various drafts of the paper. This does not imply that the reviewers agreed with our findings.

References

Allen, P.G. and Fildes, R. (2001). Econometric Forecasting in Armstrong, J.S. ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer.

Anderson, R.W. and Gainor, D. (2006). Fire and Ice: Journalists have warned of climate change for 100 years, but can't decide weather we face an ice age or warming. Business and Media Institute, May 17. Available at http://www.businessandmedia.org/specialreports/2006/fireandice/FireandIce.pdf

Armstrong, J.S. (1980). The Seer-sucker theory: The value of experts in forecasting. *Technology Review* 83 (June-July), 16-24.

Armstrong, J.S. (1978; 1985). *Long-Range Forecasting: From Crystal Ball to Computer.* New York: Wiley-Interscience.

Armstrong, J.S. (1999). Forecasting for environmental decision-making, in Dale, V.H. and English, M.E. eds., *Tools to Aid Environmental Decision Making*. New York: Springer-Verlag, 192-225.

Armstrong, J.S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers.

Armstrong, J.S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting*, 22, 583-598.

Ascher W. (1978). *Forecasting: An Appraisal for Policy Makers and Planners*. Baltimore: Johns Hopkins University Press.

Bast, J. and Taylor, J.M. (2007). *Scientific consensus on global warming*. The Heartland Institute: Chicago, Illinois. Available http://downloads.heartland.org/20861.pdf. [The responses to all questions in the 1996 and 2003 surveys by Brat and von Storch are included as an appendix.]

Bellamy, D. and Barrett, J. (2007). Climate stability: an inconvenient proof. *Proceedings of the Institution of Civil Engineers – Civil Engineering, 160*, 66-72.

Carter, R.M. (2007). The myth of dangerous human-caused climate change. The Aus/MM New Leaders Conference, Brisbane May 3, 2007. Full text available from http://members.iinet.net.au/~glrmc/new_page_1.htm

Carter, R.M., de Freitas, C.R., Goklany, I.M., Holland, D. and Linzen, R.S. (2006). The Stern review: A dual critique: Part 1. *World Economics*, *7*, 167-198.

Cerf, C. and Navasky, V. (1998). *The Experts Speak*. New York: Pantheon. Baltimore, MD: Johns Hopkins University Press.

Craig, P.P., Gadgil, A., and Koomey, J.G. (2002). What Can History Teach Us? A Retrospective Examination of Long-Term Energy Forecasts for the United States. *Annual Review of Energy and the Environment, 27*, 83-118.

Duncan, G. T., Gorr W. L. and Szczypula, J. (2001). Forecasting Analogous Time Series, in Armstrong, J.S. ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer.

Eccleston, P. (2007). Public 'in denial' about climate change. *telegraph.co.uk*, 12:01 BST 03/07/2007. Available at http://www.telegraph.co.uk/core/Content/displayPrintable.jhtml;jse...MGSFFOAVCBQWIV0?xml=/earth/2007/07/03/eawarm103.xml&site=30&page=0

Essex, C., McKitrick, R. and Andresen, B. (2007). Does a global temperature exist? *Journal of Non-Equilibrium Thermodynamics*, *32*, 1-27. Working paper available at http://www.uoguelph.ca/~rmckitri/research/globaltemp/globaltemp.html

Essex, C. and McKitrick, R. (2002). *Taken by Storm. The Troubled Science, Policy & Politics of Global Warming*, Toronto: Key Porter Books.

Hegerl, G.C., F. W. Zwiers, P. Braconnot, N.P. Gillett, Y. Luo, J.A. Marengo Orsini, N. Nicholls, J.E. Penner and P.A. Stott (2007). Understanding and Attributing Climate Change, in Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jaworowski, Z. (2007). CO_2: The greatest scientific scandal of our time. *Executive Intelligence Review*, 34(11), 38-53. Available at http://www.warwickhughes.com/icecore/zjmar07.pdf

Keepin, B. and Wynne, B. (1984). Technical analysis of IIASA energy scenarios. *Nature*, 312, 691-695.

Makridakis, S., Wheelwright, S.C., and Hyndman, R.J. (1998). *Forecasting: Methods and Applications* (3rd ed.), Hoboken, NJ: John Wiley.

NDU (1978). *Climate Change to the Year 2000*. Washington, D.C.: National Defense University.

New Zealand Climate Science Coalition (2007). World climate predictors right only half the time. Media release 7 June. Available at http://www.scoop.co.nz/stories/SC0706/S00026.htm

Pilkey, O.H. and Pilkey-Jarvis, L. (2007). *Useless Arithmetic Why Environmental Scientists Can't predict the Future.* New York: Columbia University Press.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A. and Taylor, K.E. (2007). Climate Models and Their Evaluation, in Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L. eds., *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Schnaars, S.P. and Bavuso, R.J. (1986). Extrapolation models on very short-term forecasts. *Journal of Business Research*, 14, 27-36.

Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton, D., Smith, L.A., Spicer, R.A., Thorpe, A.J. and Allen, M.R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, *433*, 403-406.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*. New York: Cambridge University Press. Available from http://www.hmtreasury.gov.uk/independent_reviews/stern_review_economics_climate_change/sternreview_index.cfm

Stewart, T.R. and Glantz, M.H. (1985). Expert judgment and climate forecasting: A methodological critique of '*Climate Change to the Year 2000'*. *Climate Change*, 7, 159-183.

Stott, P.A. and Kettleborough, J.A. (2002). Origins and estimates of uncertainty in predictions of twenty-first century temperature rise, *Nature, 416*, 723-726.

Taylor, M. (2007). An evaluation of NIWA's climate predictions for May 2002 to April 2007. Climate Science Coalition. Available at http://www.climatescience.org.nz/assets/2007691051580.ClimateUpdateEvaluationText.pdf Data available at http://www.climatescience.org.nz/assets/2007691059100.ClimateUpdateEvaluationCalc.xls.pdf

Tetlock, P.E. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

Trenberth, K.E. (2007). Predictions of climate. *Climate Feedback: The Climate Change Blog, Nature.com*, June 4. Available at http://blogs.nature.com/climatefeedback/2007/06/predictions_of_climate.html

Vizard, A.L., Anderson, G.A., and Buckley, D.J. (2005). Verification and value of the Australian Bureau of Meteorology township seasonal rainfall forecasts in Australia, 1997-2005. *Meteorological Applications*, *12*, 343-355.

Wittink D., and Bergestuen T. (2001). Forecasting with Conjoint Analysis, in Armstrong, J.S. ed. Principles of Forecasting: A Handbook for Researchers and Practitioners. Norwell, MA: Kluwer.

Appendix A: People to whom we sent our questionnaire (* indicates a relevant response)

IPCC Working Group 1

Myles Allen, Richard Alley, Ian Allison, Peter Ambenje, Vincenzo Artale, Paulo Artaxo, Alphonsus Baede, Roger Barry, Terje Berntsen, Richard A. Betts, Nathaniel L. Bindoff, Roxana Bojariu, Sandrine Bony, Kansri Boonpragob, Pascale Braconnot, Guy Brasseur, Keith Briffa, Aristita Busuioc, Jorge Carrasco, Anny Cazenave, Anthony Chen*, Amnat Chidthaisong, Jens Hesselbjerg Christensen, Philippe Ciais*, William Collins, Robert Colman*, Peter Cox, Ulrich Cubasch, Pedro Leite Da Silva Dias, Kenneth L. Denman, Robert Dickinson, Yihui Ding, Jean-Claude Duplessy, David Easterling, David W. Fahey, Thierry Fichefet*, Gregory Flato, Piers M. de F. Forster*, Pierre Friedlingstein, Congbin Fu, Yoshiyuki Fuji, John Fyfe, Xuejie Gao, Amadou Thierno Gaye*, Nathan Gillett*, Filippo Giorgi, Jonathan Gregory*, David Griggs, Sergey Gulev, Kimio Hanawa, Didier Hauglustaine, James Haywood, Gabriele Hegerl*, Martin Heimann*, Christoph Heinze, Isaac Held*, Bruce Hewitson, Elisabeth Holland, Brian Hoskins, Daniel Jacob, Bubu Pateh Jallow, Eystein Jansen*, Philip Jones, Richard Jones, Fortunat Joos, Jean Jouzel, Tom Karl, David Karoly*, Georg Kaser, Vladimir Kattsov, Akio Kitoh, Albert Klein Tank, Reto Knutti, Toshio Koike, Rupa Kumar Kolli, Won-Tae Kwon, Laurent Labeyrie, René Laprise, Corrine Le Quéré, Hervé Le Treut, Judith Lean, Peter Lemke, Sydney Levitus, Ulrike Lohmann, David C. Lowe, Yong Luo, Victor Magaña Rueda, Elisa Manzini, Jose Antonio Marengo, Maria Martelo, Valérie Masson-Delmotte, Taroh Matsuno, Cecilie Mauritzen, Bryant Mcavaney, Linda Mearns, Gerald Meehl, Claudio Guillermo Menendez, John Mitchell, Abdalah Mokssit, Mario Molina, Philip Mote*, James Murphy, Gunnar Myhre, Teruyuki Nakajima, John Nganga, Neville Nicholls, Akira Noda, Yukihiro Nojiri, Laban Ogallo, Daniel Olago, Bette Otto-Bliesner, Jonathan Overpeck*, Govind Ballabh Pant, David Parker, Wm. Richard Peltier, Joyce Penner*, Thomas Peterson*, Andrew Pitman, Serge Planton, Michael Prather*, Ronald Prinn, Graciela Raga, Fatemeh Rahimzadeh, Stefan Rahmstorf, Jouni Räisänen, Srikanthan (S.) Ramachandran, Veerabhadran Ramanathan, Venkatachalam Ramaswamy, Rengaswamy Ramesh, David Randall*, Sarah Raper, Dominique Raynaud, Jiawen Ren, James A. Renwick, David Rind, Annette Rinke, Matilde M. Rusticucci, Abdoulaye Sarr, Michael Schulz*, Jagadish Shukla, C. K. Shum, Robert H. Socolow*, Brian Soden, Olga Solomina*, Richard Somerville*, Jayaraman Srinivasan, Thomas Stocker, Peter A. Stott*, Ron Stouffer, Akimasa Sumi, Lynne D. Talley, Karl E. Taylor*, Kevin Trenberth*, Alakkat S. Unnikrishnan, Rob Van Dorland, Ricardo Villalba, Ian G. Watterson*, Andrew Weaver*, Penny Whetton, Jurgen Willebrand, Steven C. Wofsy, Richard A. Wood, David Wratt, Panmao Zhai, Tingjun Zhang, De'er Zhang, Xiaoye Zhang, Zong-Ci Zhao, Francis Zwiers*

Union of Concerned Scientists

Brenda Ekwurzel, Peter Frumhoff, Amy Lynd Luers

Channel 4 "The Great Global Warming Swindle" documentary (2007)

Bert Bolin, Piers Corbyn*, Eigil Friis-Christensen, James Shitwaki, Frederick Singer, Carl Wunsch*

Wikipedia's list of global warming "skeptics"

Khabibullo Ismailovich Abdusamatov*, Syun-Ichi Akasofu*, Sallie Baliunas, Tim Ball, Robert Balling*, Fred Barnes, Joe Barton, Joe Bastardi, David Bellamy, Tom Bethell, Robert Bidinotto, Roy Blunt, Sonja Boehmer, Andrew Bolt, John Brignell*, Nigel Calder, Ian Castles*, George Chilingarian, John Christy*, Ian Clark, Philip Cooney, Robert Davis, David Deming*, David Douglass, Lester Hogan, Craig Idso, Keith Idso, Sherwood Idso, Zbigniew Jaworowski, Wibjorn Karlen, William Kininmonth, Nigel Lawson, Douglas Leahey, David Legates, Richard Lindzen*, Ross Mckitrick*, Patrick Michaels, Lubos Motl*, Kary Mullis, Tad Murty, Tim Patterson, Benny

Peiser*, Ian Plimer, Arthur Robinson, Frederick Seitz, Nir Shaviv, Fred Smith, Willie Soon, Thomas Sowell, Roy Spencer, Philip Stott, Hendrik Tennekes, Jan Veizer, Peter Walsh, Edward Wegman

Other sources
Daniel Abbasi, Augie Auer, Jonathan Boston, Daniel Botkin*, Reid Bryson, Robert Carter*, Ralph Chapman, Al Gore, Kirtland C. Griffin*, David Henderson, Christopher Landsea*, Bjorn Lomborg, Tim Osborn, Roger Pielke*, Henrik Saxe, Thomas Schelling*, Matthew Sobel, Nicholas Stern*, Brian Valentine*, Antonio Zichichi.