2-1-2013

Evidence-Based Forecasting for Climate Change

Kesten C. Green

Willie Soon

J. Scott Armstrong
University of Pennsylvania, armstrong@wharton.upenn.edu

This is an unpublished manuscript.

Evidence-Based Forecasting for Climate Change

Abstract

Following Green, Armstrong and Soon's (*IJF* 2009) (GAS) naïve extrapolation, Fildes and Kourentzes (*IJF* 2011) (F&K) found that each of six more-sophisticated, but inexpensive, extrapolation models provided forecasts of global mean temperature for the 20 years to 2007 that were more accurate than the "business as usual" projections provided by the complex and expensive "General Circulation Models" used by the U.N.'s Intergovernmental Panel on Climate Change (IPCC). Their average trend forecast was .007°C per year, and diminishing; less than a quarter of the IPCC's .030°C projection. F&K extended previous research by combining forecasts from evidence-based short-term forecasting methods. To further extend this work, we suggest researchers: (1) reconsider causal forces; (2) validate with more and longer-term forecasts; (3) adjust validation data for known biases and use alternative data; and (4) damp forecasted trends to compensate for the complexity and uncertainty of the situation. We have made a start in following these suggestions and found that: (1) uncertainty about causal forces is such that they should be avoided in climate forecasting models; (2) long term forecasts should be validated using all available data and much longer series that include representative variations in trend; (3) when tested against temperature data collected by satellite, naïve forecasts are more accurate than F&K's longer-term (11-20 year) forecasts; and (4) progressive damping improves the accuracy of F&K's forecasts. In sum, while forecasting a trend may improve the accuracy of forecasts for a few years into the future, improvements rapidly disappear as the forecast horizon lengthens beyond ten years. We conclude that predictions of dangerous manmade global warming and of benefits from climate policies fail to meet the standards of evidence-based forecasting and are not a proper basis for policy decisions.

Keywords

complexity, conservatism, cost benefit analysis, decision making, extrapolation, public policy, uncertainty, validation

Disciplines

Business | Environmental Policy | Environmental Studies | Marketing | Policy Design, Analysis, and Evaluation | Statistics and Probability

Comments

This is an unpublished manuscript.

Evidence-based Forecasting for Climate Change

Kesten C Green, University of South Australia, Australia
Willie Soon, Harvard-Smithsonian Center for Astrophysics, U.S.A.
J Scott Armstrong, University of Pennsylvania, U.S.A.

DRAFT

1 February 2013

Following Green, Armstrong and Soon's (*IJF* 2009) (GAS) naïve extrapolation, Fildes and Kourentzes (*IJF* 2011) (F&K) found that each of six more-sophisticated, but inexpensive, extrapolation models provided forecasts of global mean temperature for the 20 years to 2007 that were more accurate than the "business as usual" projections provided by the complex and expensive "General Circulation Models" used by the U.N.'s Intergovernmental Panel on Climate Change (IPCC). Their average trend forecast was .007°C per year, and diminishing; less than a quarter of the IPCC's .030°C projection. F&K extended previous research by combining forecasts from evidence-based short-term forecasting methods. To further extend this work, we suggest researchers: (1) reconsider causal forces; (2) validate with more and longer-term forecasts; (3) adjust validation data for known biases and use alternative data; and (4) damp forecasted trends to compensate for the complexity and uncertainty of the situation. We have made a start in following these suggestions and found that: (1) uncertainty about causal forces is such that they should be avoided in climate forecasting models; (2) long term forecasts should be validated using all available data and much longer series that include representative variations in trend; (3) when tested against temperature data collected by satellite, naïve forecasts are more accurate than F&K's longer-term (11-20 year) forecasts; and (4) progressive damping improves the accuracy of F&K's forecasts. In sum, while forecasting a trend may improve the accuracy of forecasts for a few years into the future, improvements rapidly disappear as the forecast horizon lengthens beyond ten years. We conclude that predictions of dangerous manmade global warming and of benefits from climate policies fail to meet the standards of evidence-based forecasting and are not a proper basis for policy decisions.

Key words: complexity; conservatism; cost benefit analysis; decision making, extrapolation; public policy; uncertainty; validation

File: ISF/ISF-2012/ GAS_on_F&K-v81.doc
Latest draft available from http://www.kestencgreen.com/gas-improvements.pdf

The Intergovernmental Panel on Climate Change (IPCC) is responsible for projections of strongly increasing global mean temperatures and of substantial harm arising from increasing temperatures (IPCC AR5 reports from Working Group I, II and III due out in September 2013, March 2014, and April 2014, respectively). The IPCC attributes projections of increasing temperatures to carbon dioxide emissions from human activity. Lobby groups and politicians refer to the IPCC's reports to justify expensive policy recommendations such as abandoning coastal settlements, taxing carbon dioxide emissions, and subsidizing selected energy sources. Governments have implemented such policies and calls for further and more expensive policies continue.

We argue that, in every reasonable understanding of the term, the IPCC projections are forecasts. The projections are statements about what will happen in the future. The term forecast and predict, and their derivations, occur 127 times in the IPCC's AR4 report, and decision makers use what the IPCC refer to as estimates or projections as forecasts. The IPCC's forecasts dominate the media (e.g., see New Scientist 2007; Hansen and Sato 2012; *Washington Post* Editorial Board, 2013) and the policy agenda including future weather-related disaster preparedness (see for example USGCRP 2009 and Parris *et al*. 2012).

As with our previous research, we make comparisons with the IPCC's 1990 and 1992 forecasts of "future warming at rates of about 0.3°C/decade… over the next century" (p. 17, Houghton, Callander, and Varney 1992). By using this rate, it is possible to assess the accuracy of out-of-sample forecasts generated by the IPCC procedures for periods of up to 21 years at the time of writing. This 21-year maximum forecasting horizon is short compared to the century-long horizon—and now 300-year horizon (Caesar 2012)—for which the forecasts are provided.

The long horizons are important, because warming for the next decade or two at the IPCC forecast average rate of .03°C per annum would result in little change, and the cost of policies intended to reduce change would be wasted if the warming stopped or reversed beyond the forecast horizon independently of any human actions. Uncertainty increases greatly with longer horizons.

In contrast to the IPCC's efforts on climate projections, the efforts of the forecasting research community have been modest. We are therefore pleased to be able to build upon Robert Fildes and Nikolaos Kourentzes's (2011) contribution to the scientific forecasting of climate. Climate change is properly a forecasting problem and proper forecasting procedures are needed in order to determine any rational policy response.

Approach used by Fildes and Kourentzes

Green and Armstrong's (2007) audit of the procedures used by the IPCC to forecast temperatures found that they were not scientific, in the sense that they violated 72 of 89 relevant evidence-based principles. For example, IPCC lead authors famously resisted disclosing their data and methods (Michaels 2009).

In contrast, Fildes and Kourentzes (2011) (denoted below as F&K) demonstrated the proper application of many forecasting principles in deriving their forecasts of changes in global mean temperatures up to 20 years ahead. Given that they followed many forecasting principles, it is no surprise that their forecasts were much more accurate than the projections (henceforth "forecasts") of the IPCC. The errors from the least accurate of F&K's six forecasting methods were less than half the IPCC forecast errors for forecast horizons of 1 to 4 years, and for the 10-year ahead forecasts. The improvement was achieved using inexpensive methods.

F&K's results were similar to the ones we presented in Green, Armstrong, and Soon (2009), which we denote below as GAS. GAS showed that the IPCC forecast[1] errors were 45% larger than the forecasts from a naive no change model for one- to ten-year horizons, and for 99- to 100-year horizons the errors were nearly 13 times larger. Forecasting researchers know that, "a well-specified… model should forecast at least as well as the naïve no-change method" (p. 348, Allen and Fildes 2001).

F&K's approach provides a model for those doing forecasting research on climate change. Their approach followed evidence-based principles as described in Armstrong (2001). These include:

1. Use simple methods for a complex and uncertain situation
2. Compare reasonable alternative methods
3. Consider that different methods might be relevant for short-term forecasting
4. Combine forecasts from evidence-based methods

[1] The IPCC forecasts examined by GAS were not published by the IPCC, but were the product of the IPCC forecasting model that predicts an annual increase in global mean temperatures of 0.03°C when atmospheric concentrations of carbon dioxide increase at an exponential rate (Figure 2.3 on p. 38, IPCC 2007).

5. Conduct validation tests on data unknown to the models[2]
6. Use valid and relevant error measures
7. Provide full disclosure including peer review, publish commentary, and respond to questions about the research.

F&K tested our "naïve" (or no-change) forecasting model against six other alternatives to the IPCC's General Circulation Models, one GCM (Smith's Decadal Climate Prediction System referred to as DePreSys in F&K), and a combined forecast using the seven non-GCM methods. F&K refer to the naïve model as the random walk model or RW. As in GAS, we use the term "naïve model" in this paper. To test the forecasting models, F&K used a version of the same temperature series that we used in GAS (HadCRUT3), the U.K. Met Office's University of East Anglia Hadley Centre proxy global mean temperature anomaly series, HadCRUT3v. [3]

The naïve model we used in validation tests in GAS was probably overly simple in that it ignored the persistence of important influences on climate such as volcanic eruptions and El Nino weather patterns (Wu *et al*. 2011; Qian *et al*. 2011). Thus it seems reasonable to forecast, as the F&K methods do, that in the short-term, trends will continue or that recent levels will persist.

Results of Fildes and Kourentzes

We expected, then, that the forecasts from the F&K models would be relatively more accurate than forecasts from the GAS naïve model across the relatively short horizons examined by F&K. We were surprised to find that they were not. While some of F&K's methods provided modestly more accurate forecasts, for each horizon the best method was different.

F&K did not propose which method they expected to provide the most accurate forecasts, so we compared forecast errors from the naïve model with the average of those from F&K's six alternatives. We did not compare the naïve model forecast errors with those from the DePreSys GCM because there is only a small sample of forecasts from the one model and that was only for very short horizons. In any event, inclusion of those forecast errors in an average with those from F&K's six methods would have had little effect on the average accuracy.

To make the comparison, we summed all 176 absolute errors for each method's forecasts across the 1–4, 10, and 20-year time horizons that F&K used (66 observations from their Table 1, and 60 and 50 observations from Table 2; F&K p. 984 and p. 985). The sum of the absolute errors from the naïve model forecasts were 5% *smaller* than the average of the six F&K methods' sums of absolute errors.

F&K subsequently provided us with all the forecasts that they described in their paper: Namely, 1,190 forecasts for horizons from one to twenty years from each of their six methods. For the full set of forecasts, errors from the naïve model forecasts were 2% smaller than the average of the six F&K methods' errors.[4] We found that forecasts from F&K's models for their longer, 11 to 20 year, horizons were nearly 8% less accurate than the equivalent forecasts from the naïve model.

Approaches used by those who forecast dangerous warming

We are not aware of any efforts by the IPCC modelers to adopt evidence-based forecasting methods. They instead appear to have chosen to go down the route of further elaborating already complex and expensive models (e.g, Soon *et al*. 2001; Dawson *et al*. 2012; Jiang *et al*. 2012; Mauritsen *et al*. 2012; Turner *et al*. 2012). To do so, they have had to rely heavily on unaided judgment in selecting model variables and setting parameter values. In their section on "Simulation model validation in longer-term forecasting" (p. 969–973), F&K observe of the IPCC modeling procedures: "a major part of the model building is judgmental" (p. 970). They describe how judgment is heavily relied upon for:

- construction of a global temperature series from disparate selected local readings

---

[2] In the case of F&K, the modelers knew the data used for validation tests. We discuss this issue below.
[3] HadCRUT3 has at the time of writing been superceeded by HadCRUT4. We use HadCRUT4 in this paper, except where the earlier series are needed for consistency. The series all commence in 1850 and consist of adjusted composites of manual temperature readings from selected weather stations and ships (Morice *et al* 2012; Jones *et al*. 2012).
[4] Tables summarising our analysis of the F&K forecasts are available at http://kestencgreen.com/gas2012.pdf.

- inclusion and exclusion of variables
- representation of climate processes
- estimation of parameters
- selection of initial conditions
- selection and interpretation of model outputs

Consistent with Green and Armstrong (2007) and others, F&K noted of the IPCC's GCMs, "Because of the complexity of such models, the… costs of optimizing are… prohibitive. Even if it were feasible, given the large number of degrees of freedom and the limited observations, it is necessary to use judgment. Thus a major part of the model building is judgmental" (p. 970). Regarding the complexity of models, the first author of F&K and his co-author concluded their review of the evidence on quantitative causal forecasting methods with the principle that forecasters should "aim for a relatively simple model specification… always" (Allen and Fildes 2001, p. 348).

Regarding the use of judgment by experts in complex and uncertain situations, a review of the evidence on the accuracy of forecasts by Armstrong (1980) found that judgmental forecasts in such situations were no more accurate than those of novices or of guessing. In another study, 284 experts made 82,361 forecasts of political, social and economic developments over a 20-year period. Their forecasts were no more accurate than the forecasts of novices, or naïve forecasts (Tetlock 2005).

People often assume that the situation they are concerned with is somehow unique. Thus, they question whether the principles of forecasting apply to their situation. It is not surprising, then, that climate modelers raise the issue. The answer is that the principles are drawn from comparative tests of which methods work best under given conditions. Much comparative research has been conducted in the physical and social sciences. The principles, published in 2001, were initially due to the work of 39 scientists from different disciplines (psychology, economics, business, weather, criminology, etc) and different countries, aided by 123 reviewers. Following that, the forecastingprinciples.com site was established to update the principles. When global warming alarmists claimed that the principles (which include "fully disclose the methods and data used to make the forecasts") did not apply to them, we asked for a list of relevant principles that do not apply and the evidence that they do not. We have yet to receive any such evidence. Furthermore, the procedures used in the IPCC's Fourth Assessment Report in 2007 failed to cite any papers that validated their forecasting procedures. We received no replies to our requests for research that might validate the procedures.

In sum, we concur with F&K's statement that, "the structural weaknesses in the GCM identified here suggest that a reliance on the policy implications from the general circulation models, and in particular the primary emphasis on controlling global carbon dioxide (CO_2) emissions, is misguided" (p. 992).

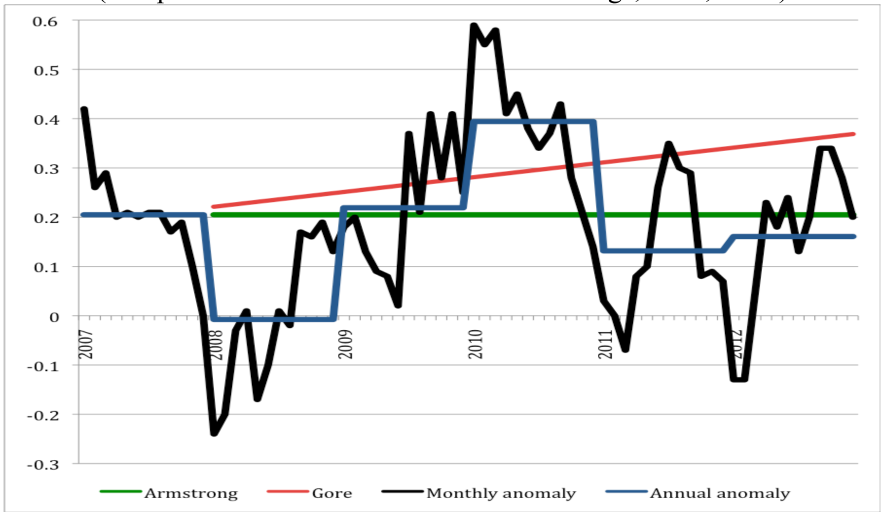
Testing forecast validity

Short-term trends provide little information as to long-term trends, so it is important to test long-term forecast validity and uncertainty *before* advocating policies with long-term effects. Nevertheless, in an effort to encourage interest in the idea that climate forecasting methods should be tested in an open and scientific manner before being declared fit for use, in 2007 one of us (Armstrong) challenged former U.S. Vice President Gore—a high-profile promoter of the IPCC's alarming 0.030°C per annum "business as usual" forecasts—to bet that those forecasts would be more accurate than naïve forecasts over the ten-year period from 2008 to 2017.

The bet was proposed to highlight the need for climate model validation. The claim in the global warming community was that we are at the point of rapid acceleration and that no time can be lost (Hansen 2007; Vermeer and Rahmstorf 2009; Hansen and Sato 2012; the latter proposed a very alarming scenario of 5-meter rise in global sea level by 2100 AD). Over such a short period, however, it is quite likely that temperatures could drift up by more than 0.015°C per annum, on average, as a result of random variation. Nevertheless, Armstrong's naïve forecast that global mean temperatures would remain at the 2007 average was more accurate than the IPCC's forecast for 40 out of the first 60 months of the bet (and four of the five years) and the cumulative absolute error for the naïve forecasts was 14% smaller than for the Gore-IPCC forecasts.

The progress of the proposed bet at the time of writing—to December 2012—is shown in Figure 1, and the latest information is available from theclimatebet.com.

Figure 1: First 5-years of a 10-year bet on global mean temperatures
(Temperature deviation from 1981-2010 average, in °C, UAH)



Implications for short-term climate forecasting

For the one-to-twenty year horizons examined by F&K, the errors of the F&K *combined forecasts* were 8% smaller than the naïve forecast errors in their tests, suggesting there are gains in accuracy to be had from methods that incorporate information on recent trends and causal forces, and from combining forecasts from different methods.

With these hints that improvements might be possible in mind, we suggest further testing of damped trend extrapolation, and testing of judgmental adjustments using inputs from Delphi panels to adjust for major events that are anticipated. We expect combinations of forecasts from these methods with forecasts from the naïve model will increase short-term forecast accuracy.

Suggestions for improving long-term climate forecasting

F&K is an important addition to the field. We hope that the authors will continue their research and that other forecasting researchers will join in this effort. To aid in these efforts, we propose possible further improvements for long-term climate forecasting:

1. Consider causal forces
2. Conduct validations with more and longer-term forecasts
3. Adjust validation data for known biases and use alternative validation data, and
4. Damp trends to fully compensate for the complexity and uncertainty of the situation.

We discuss each of these suggestions in turn.

Consider causal forces, and whether things are different now
If causal models are feasible, it is important to "consider all important causal variables based on guidelines from theory and earlier research [and] include difficult-to-measure variables and proxy variables" (p. 310, Allen and Fildes 2001).

To select causal forces it is not sufficient to find a correlation. For example, while a correlation of 0.86 between atmospheric CO_2 and global mean temperature for the period 1850 to 2008 might seem high, for the same period the correlation between U.S. Postal rates and temperatures was 0.84, and for the period 1970 to 2006 the correlation between National Oceanic and Atmospheric Administration (NOAA) expenditures and temperatures was 0.82.

Scientists dispute the nature and importance of causes of global mean temperature changes. The IPCC modelers assume that manmade changes in CO_2 and other trace gases in the atmosphere are major drivers of temperature and hence of climate changes (Randall *et al.* 2007). Other researchers dispute that assumption, concluding that there is no direct evidence of any human effect on global climate (e.g., Beenstock *et al*. 2012).

Leading climate scientists propose other sources for climate change such as variations in solar radiation reaching the Earth (see e.g., Soon *et al*. 2011), unexplained variations in cloudiness (Lindzen and Choi 2011), and variations in ocean and atmospheric currents (Lindzen 1994; Wunsch 2010). In contrast, there is little dispute that major volcanic eruptions can lead to cooling over several years (see e.g., Gao *et al*. 2008; Breitenmoser *et al*. 2012), or that the Sun is an important influence on climate (see e.g., Soon 2009; Soon and Legates 2013).

Very long temperature series (see, e.g., Figure 2, from Jouzel *et al*. 2007) suggest that in the absence of human influence (1) mean temperatures have changed substantially over long periods of time, (2) trends appear to reverse at irregular intervals, and (3) there is no underlying long-term trend. It is by no means clear that there is sufficient evidence to conclude that modern warming is different from, for example, the medieval warming, and that it will not reverse (see Figure 3, from Loehle and McCulloch 2008).

Figure 2: 800,000-year Record of Antarctic Temperature Change

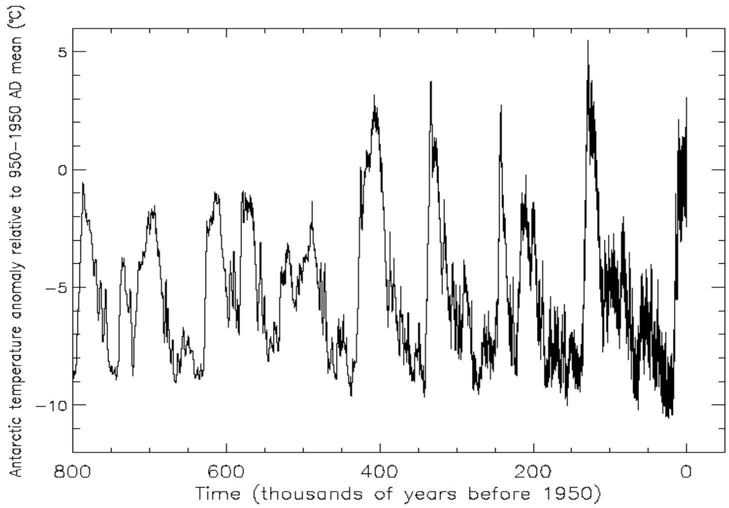
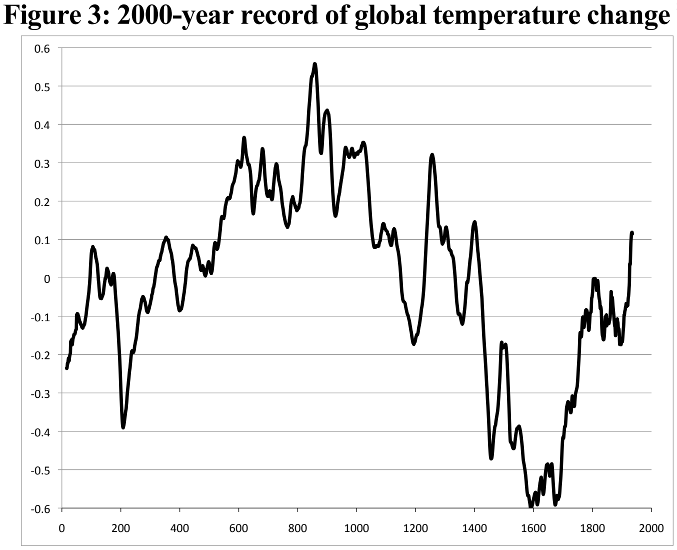


Figure 3: 2000-year record of global temperature change [5]



[5] We followed the conservative presentation of findings in Loehle and McCulloch (2008) by showing in Figure 3 the smoothed paleo-temperature series available from 16 AD to 1935 AD only. There is an inherent unresolvable question related to how one can connect these paleo-temperature series to a series based on thermometer readings (see Soon *et al*. 2003), not the least of which is how to correct for the non-climatic contamination of the thermometer records. We are nevertheless satisfied of the rebustness of Loehle and McCulloch's (2008) finding that "The peak value of the MWP [Medieval Warm Period] is 0.526 Deg C above the mean over the period (again as a 29 year mean, not annual, value). … While instrumental data are not strictly comparable, the rise in 29 year-smoothed global data from NASA GISS … from 1935 to 1992 (with data from 1978 to 2006) is 0.34 Deg C. Even adding this rise to the 1935 reconstructed value, the MWP peak remains 0.07 Deg C above the end of the 20th Century values…" (pp. 97-98).

Does a closer inspection of more recent data support the hypothesis that causal forces are operating differently now? We know that a regression line fitted to the Hadley series from its start in 1850 through to 2012 has a slope of .0046°C per year, but any sequence of as short a period as 160 or so years' temperatures from the pre-industrial period is also likely to slope modestly; sometimes upwards, and sometimes downwards. Over the 2000-year record shown in Figure 3, for example, 5.4% of 160-year sequences ending prior to 1800 have absolute slopes that exceed .0030°C, and half have slopes that are less than -.0016°C or greater than .0013°C. Moreover, as we discuss below, there are reasons to believe that the Hadley data is biased in the direction of a warming trend to the extent that the actual rate of warming over the period of the Hadley data might be less than half of the Hadley warming. If we assume that the actual average warming rate for the period were .0023°C, we find that 20% of 160-year sequences ending prior to 1800 warmed or cooled at a faster rate.

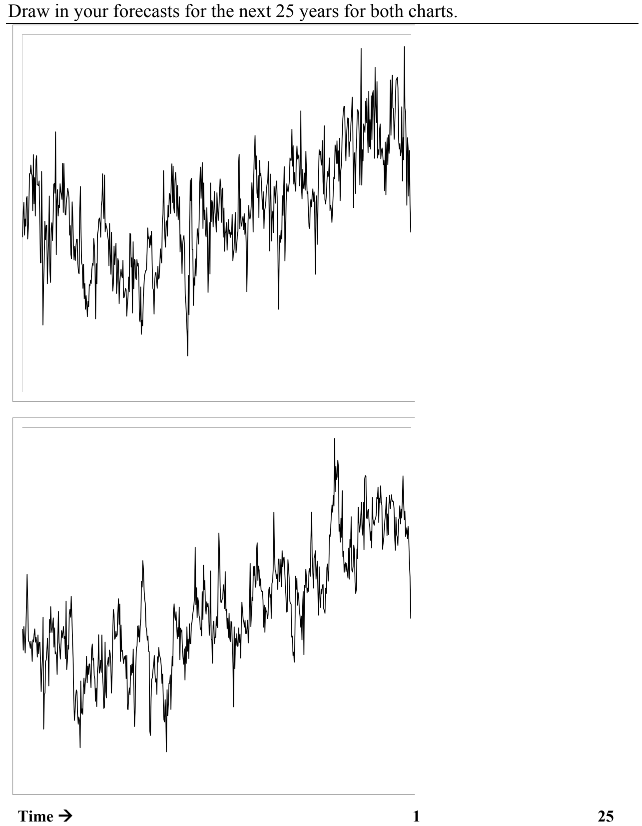
We asked forecasters at the 2012 International Symposium on Forecasting in Boston to forecast the next 25 years of two similar-looking 50-year sequences of monthly global mean temperatures from the Hadley data set (Figure 4). We told the forecasters that both sequences occurred during the industrial period. (In other words, both coincided with exponential increases in atmospheric carbon dioxide). In the event, nearly half, or 23 of the 51 respondents, made forecasts that were consistent with carbon-dioxide causality, while another 27 forecast little or no trend over the forecast period. Figure 5 shows the data in context and a trend line for the 25-year period that the forecasters were asked to forecast and for which there is data available. Note that the similarity of the two sequences is consistent with the null hypothesis that climate changes are no different now to those in earlier times.[6]

Figure 4: Forecasting quiz using Hadley (HadCRUt3) global mean temperatures

Test your forecasting skills:
Print this page and draw in your forecasts

Monthly global mean temperatures over two half centuries*

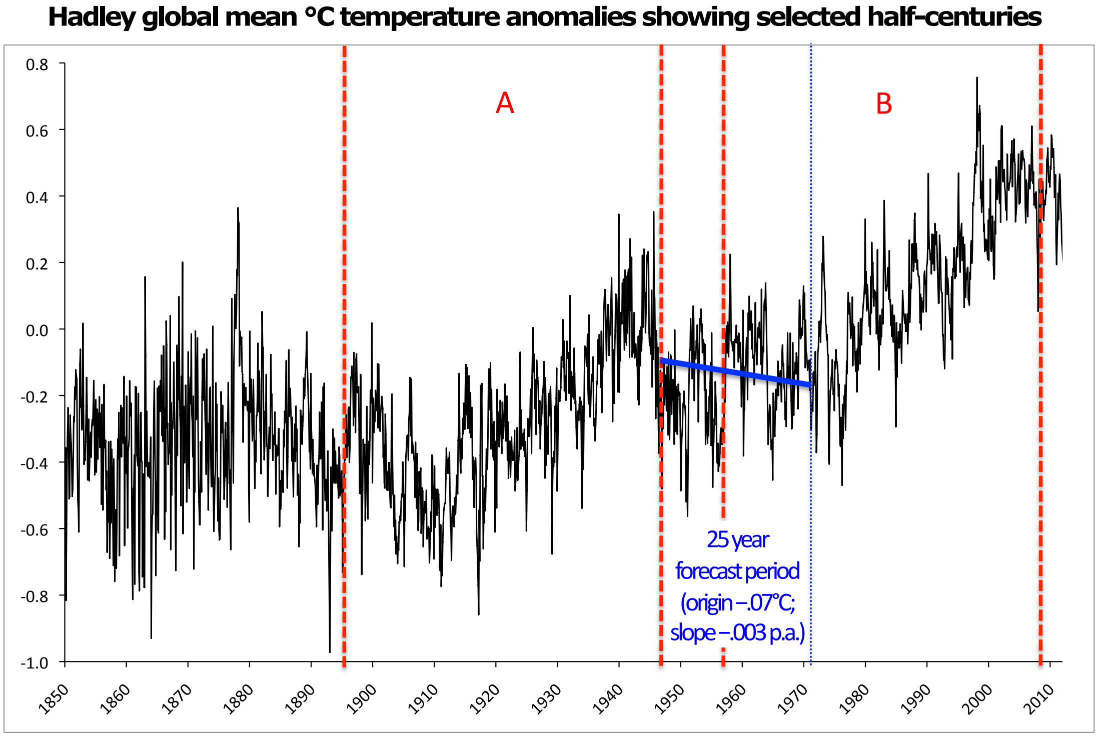
Draw in your forecasts for the next 25 years for both charts.



**Both during industrial era*

[6] The idea behind this comparison came from a presentation by Meyer (2009).

Figure 5: Forecasts for the two similar graphs



Given the uncertainty over the nature and extent of causal forces and the difficulty of forecasting at least some of the causal influences that have been proposed by researchers, we suggest that causal forces should be incorporated into forecasting models with great caution and in a simple and transparent manner.

In practice, the uncertainty about causes, direction of effects, effect sizes, timing and durations of effects, interactions, feedback issues, excluded variables, and the lack of a consistent long-term trend suggest that the conservative no-change model may well be better for long-term forecasting of global mean temperatures than can currently be achieved with causal models. At the least, the no-change model should be heavily weighted relative to causal and trend-forecasting methods, rather than equally weighted as it is in F&K's combined forecast.

We tested the accuracy of an average of the combination of forecasts from F&K's six methods (50%) and the naïve model (50%) over the one- to twenty-year horizons for which F&K forecasts are available. Bearing in mind that F&K's methods were applied to a small sample that was known to trend more in one direction (up) than the other, we expected that forecasts that incorporated a trend would perform better than the no-trend naïve forecasts for that period.

And so they do. We found that the cumulative relative absolute error (CumRAE; Armstrong and Collopy 1992) for the resulting 1,190 "conservative trend" forecasts was .922. The figure represents an error reduction of almost 8% relative to the naïve forecast, and a reduction of 10% relative to the forecasts from the average of F&K's six models.

F&K's AR models[7] are also conservative trend models. On examining the forecasts F&K provided us with, we found the models forecast trends of between 0.001°C and 0.007°C, with a mean of 0.0037°C or 0.37°C per century. That is roughly half the 0.70°C per century rate at which the Hadley data increased during the 1914 to 2006 period over which the models were variously estimated, and an eighth of the IPCC's projected warming rate. Moreover, the AR models are inconsistent with any claim of accelerating warming: models estimated with more recent data do not forecast faster warming than those estimated with older data.

[7] Autoregression models forecast the next period of a time series using earlier values of the series.

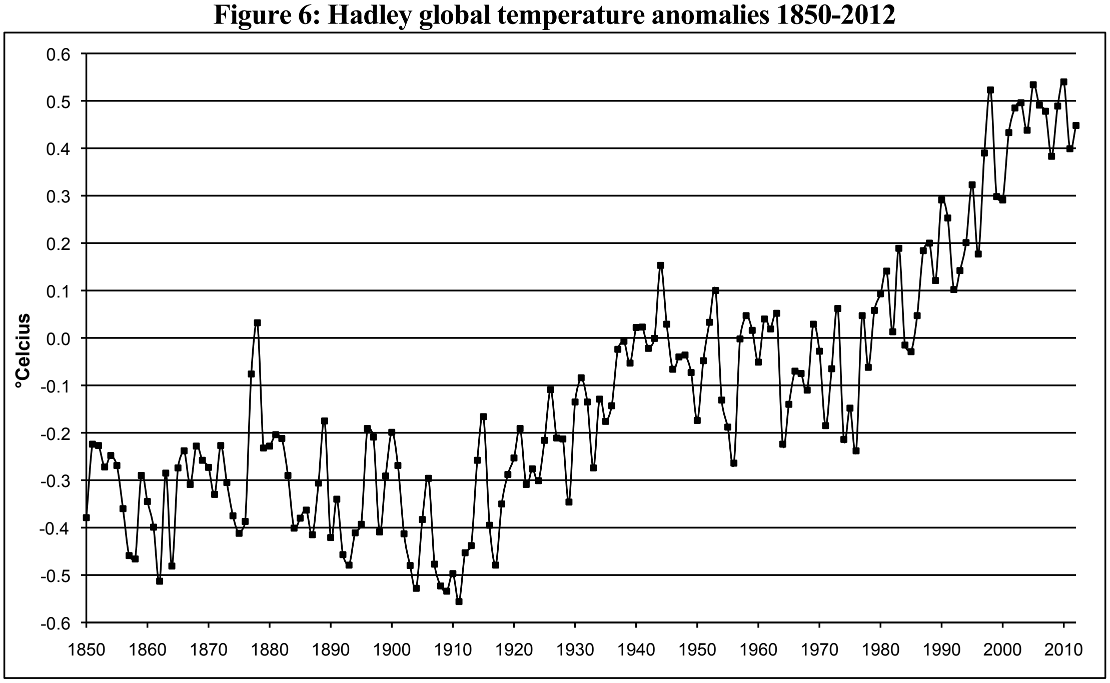
Validate methods with more and longer-term forecasts

One rule-of-thumb for forecasters is that the forecast horizon should be no longer than the length of the historical data. The rule is consistent with the principle that forecasters should be conservative when there is uncertainty in the situation. To properly test methods for long-term forecasting, then, large validation samples are needed. There are currently only 21 years of truly out-of-sample forecasts from the IPCC models for which actual values are available. In GAS, we got around this validation problem by applying the IPCC's forecast that exponentially increasing atmospheric CO_2 levels would cause global mean temperatures to increase by 0.03°C per year to forecast the Hadley temperature series from 1851 to 1975. (The period was one of exponentially increasing industrialization and energy use and is thus analogous to the IPCC's 0.03°C per year business-as-usual scenario.) By doing so we obtained 28 forecasts for 100-years-ahead, 76 forecasts for 50-year-ahead forecasts, and 106 forecasts for 20-years-ahead. The errors for those forecasts were six times greater than the naïve forecast errors.

We suggest that other researchers also use the *full* Hadley data set[8], which is shown in Figure 6. In GAS, we used the data from 1851 to 2007 for validating our naïve model. We were able to use all but the first observation of the series because our model assumed there was no trend in the data. This approach provided 10,750 forecasts for 1-year to 100-year horizons. In contrast, F&K's validation tests were based on 1,190 forecasts for shorter (1–20 year) horizons from each model. While it is well known that the Hadley data series trends broadly upwards the full series does include downward trending periods, and so it provides a stronger test.

Using the full Hadley data set does not entirely avoid the problem of lack of out-of-sample data because forecasters are aware of the upward trends in both CO_2 concentrations and (very broadly) temperatures over that time. This is contrary to the principle that one should "test all models for performance with data not used in estimation" (p. 340, Allen and Fildes 2001). To better avoid the problem, we suggest using temperature series that are much longer than the Hadley series, such as the roughly 2000-year Loehle and McCulloch (2008) series (Figure 3), and local and regional series, for validating long term forecasting methods.

Figure 6: Hadley global temperature anomalies 1850-2012



Adjust validation data for known biases, and use alternative validation data

Studies of the Hadley data series have found biases. These include the heat island effect, the reduction of the number of weather stations in isolated areas, poor maintenance of weather stations, and the increasing use of

---

[8] Hadley (HadCRUt4) "best estimate" annual average temperature differences from the U.K. Met Office Hadley Centre obtained from http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/time_series/HadCRUT.4.1.1.0.annual_ns_avg.txt. The graph of Hadley data to 2012 uses data obtained on 5 January 2013. Information on the series is available from http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/series_format.html and http://www.metoffice.gov.uk/hadobs/hadcrut4/data/versions/HadCRUT.4.1.1.0_release_notes.html

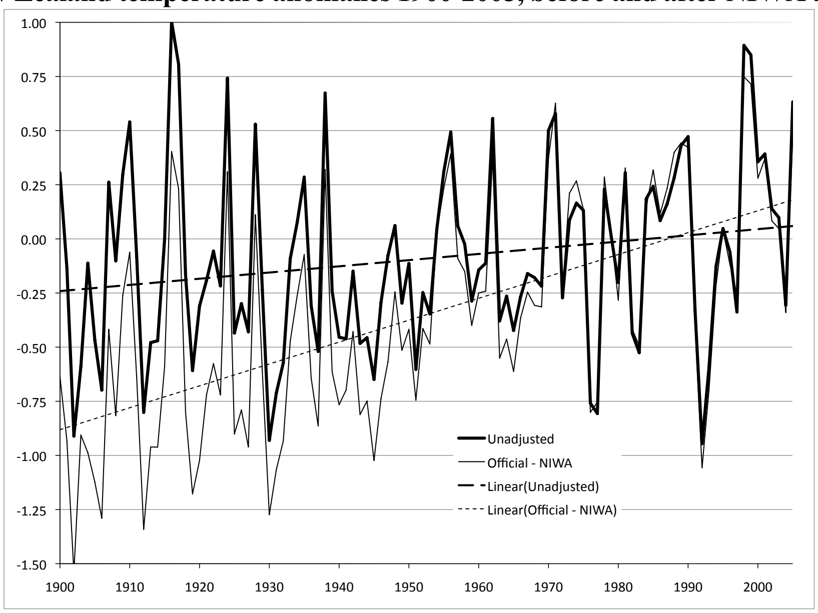
electronic thermometers, which can record very short duration highs that can arise from heat eddies (Fall *et al*. 2011; Watts *et al*. 2012). In some cases, recorded temperatures have been adjusted without explanation (see, for example, Figure 7).

Research on climate forecasting has been hampered by an anti-scientific reluctance from those who forecast global warming to respond to requests for information. For example, in 2005, Warwick Hughes wrote to Phil Jones of the University of East Anglia's Climate Research Unit, source of the HadCrut3 surface temperature data set, to ask for the original data. Jones replied, "Why should I make the data available to you, when your aim is to try and find something wrong with it?" (Michaels 2009).[9]

The biases all appear to be in the direction of producing a warming trend. One analysis led to the conclusion that the actual increase in temperatures from 1979 to 2002 was half as large as the increase in the Hadley data (McKitrick and Michaels 2007; McKitrick and Nierenberg 2010; Watts *et al*. 2012).

Researchers in New Zealand constructed an unadjusted temperature series using data from the same weather stations that are used in the pre-2010 official (adjusted) temperature series. The official series is also used in the construction of the Hadley temperature series. Their findings are summarized in Figure 7. The official series has an increasing trend of 1.0°C per century, whereas the unadjusted has a trend of 0.3°C per century. Prior to their 2010 review, the government agency responsible (NIWA) provided neither a record of the adjustments that were made in order to construct the official series, nor the reasons for them (The New Zealand Climate Science Coalition 2011).

Figure 7: New Zealand temperature anomalies 1900-2005, before and after NIWA adjustments[10]



The lower troposphere temperature anomaly data from the University of Alabama at Huntsville (UAH) is the most valid of the global mean temperature series currently available (Figure 8). The measurements involve using satellite-mounted instruments to detect microwave emissions. The measurement method avoids contamination from near-surface structures and other non-climatic influences by spatial sampling emissions from molecular oxygen, which is somewhat evenly distributed in the global atmosphere (Christy *et al.* 2010; McNider *et al*. 2012).

Unfortunately, the series started only in late-1978, and the roughly 30 years of the series is insufficient for adequate testing of long-term forecasting accuracy. Nevertheless, we reasoned that the data presents an opportunity to compare the relative accuracy of forecasts from F&K's six models and from the GAS naïve model.

---

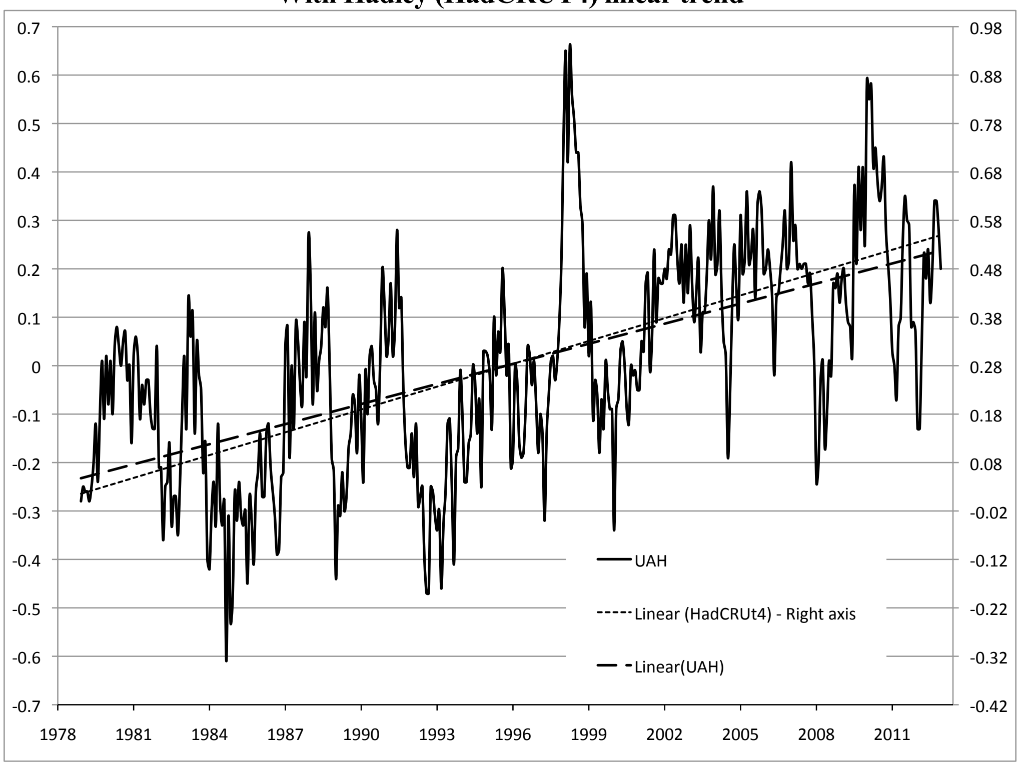
[9] The correspondence from a more recent attempt to obtain the original data is provided in Willis Eschenbach's post "An open letter to Dr. Phil Jones of the UEA CRU": http://wattsupwiththat.com/2011/11/27/an-open-letter-to-dr-phil-jones-of-the-uea-cru/

[10] Data from personal correspondence with Barry Brill of The New Zealand Climate Science Coalition.

We did not have access to the F&K models to derive forecasts based on the UAH data. To get around the problem, for each model's forecasts we compared the change from the previous year with the actual UAH changes in order to derive forecast errors. We expected that F&K's models would have an advantage in forecasting out to 10 and perhaps 20 years, but that the GAS naïve model would be a strong contestant for longer horizons when uncertainty is greater.

The first UAH annual change that can be calculated is for 1980 and the last F&K forecast was for 2007, there are 532 F&K forecasts per model for horizons from 2 to 10 years over this period. We found that the average absolute errors from the F&K model forecasts, expressed as annual temperature changes, were 1.1% smaller than the equivalent naïve model forecasts. The 190 errors from the F&K model forecasts for F&K's longer horizons (11 to 20 years), however, were 0.5% larger than the comparable naïve model forecast errors.

Figure 8: UAH's global lower troposphere temperature anomalies 1979-2012[11]
With Hadley (HadCRUT4) linear trend



Damp forecasted trends strongly to compensate for the complexity and uncertainty of the situation
We expect that damping longer-term forecasts would improve accuracy.

Complexity and uncertainty are both high for climate forecasting due to problems in obtaining valid and reliable estimates of important climate variables (such as global mean temperatures) for long time periods, and due to poor knowledge about causal relationships.

Forecasting procedures and forecasts should be conservative when complexity and uncertainty are high. The average of F&K's six models' 10-year ahead forecast errors (MAE) was .16°C and of the models' 20-year ahead MAE was .23°C (F&K's Table 2, by calculation). The more conservative naïve model forecast errors were smaller for both horizons at .15°C and .20°C. The pattern of errors suggests that forecasts should be damped to the naïve for horizons beyond 10 years.

To test the effect of damping, we damped the forecasts from each of the F&K models progressively, starting with the 11-year horizon, to equal the naïve model forecast by the 20-year horizon using the formula F_m *damped* $= F_m + [(F_N - F_m) / (21 - h)]$; where F_N is the naïve model forecast and h is the horizon from 11 to 20. The damped F&K forecast errors were nearly 1% smaller than the naïve model forecast errors; an improvement

[11] UAH data from http://vortex.nsstc.uah.edu/data/msu/t2lt/uahncdc.lt. The UAH data plotted are the entire series to December 2012. See latest discussion ("Readme") at http://vortex.nsstc.uah.edu/data/msu/t2lt/.

in accuracy of nearly 3% relative to the original F&K forecasts. We made no attempt to optimize the damping algorithm.

Policy implications

How should policy makers respond to calls for costly policies to combat climate changes?

Predictions of climate change have been made in the past but did not lead to expensive policy changes. In 1896 the Swedish Nobel Prize winner in chemistry, Svante Arrhenius, speculated about the effect of increases in atmospheric CO_2 and forecast that higher concentrations would cause warming (Baliunas and Soon 1999). Arrhenius's idea attracted little attention at the time, probably because he assumed that global warming was desirable.

In the 1960s, experts predicted that associated human emissions of soot and aerosols would more than overpower the effects of increases in atmospheric CO_2 thereby threatening the world with dangerous manmade global cooling (Rasool and Schneider 1971). Calls of alarm that the world was in danger of entering a new ice age faded when the declining trend in global mean temperatures stalled.

By 1988, NASA's James Hansen was warning of dangerous manmade global warming. The IPCC was established in the same year to seek evidence to support the manmade global warming hypothesis. As with the cooling alarm, temperatures have not changed in accord with the predictions of advocates of the dangerous warming hypothesis. Despite the lack of forecasting success, governments have implemented some of the expensive policies that have been advocated.

F&K's findings have reinforced our findings in Green and Armstrong (2007) and GAS that scientific forecasting leads to a conclusion that the hypothesis of dangerous warming over the 21st Century and beyond should be rejected. The additional analyses in this paper strengthen that conclusion.

As we pointed out in Armstrong, Green, and Soon (2008), it would not be sufficient to merely show that warming is likely. Policy actions require scientific forecasts to support the claims that the effects would be detrimental. No such forecasts exist. Finally, it would then be necessary to provide scientific forecasts showing that the policies would be cost effective. Again, no such forecasts exist.

Some people do believe that it is reasonable to make policy on the basis that a particular unlikely (and perhaps impossible) event will occur, and call this belief the "precautionary principle" [12]. We do not. The unscientific precautionary principle would lead to complete atrophy if its advice to stop or not start an activity just in case something bad happens were applied consistently. In practice, the concept is applied inconsistently to advocate a particular course of action. In the case of climate change, it would be just as reasonable to use the precautionary principle to advocate policies in response to possible dangerous global cooling[13].

Moreover, policy makers should properly determine whether welfare would be greater in practice with proposed policies than it would if no policy changes were implemented. In other words, if people were left to adapt as necessary and as they saw fit to any local changes that might occur. These are not trivial questions. For example, people in many places would benefit from a warmer climate and higher levels of CO_2.[14]

It is hard to imagine that people, given their diverse local climates and preferences, could possibly agree on an optimum global mean temperature and on how much they would be willing to pay for it, heroically assuming it were possible to achieve that optimum.

Despite the lack of scientific support for climate policies, vast resources have already been devoted to policies that are nominally intended to stop predicted dangerous manmade global warming (e.g., Hansen 2007; Hansen and Sato 2012). This apparently strange situation can be explained by the motivations of the parties involved and the incentives that they face. Governments and interest groups are attracted to possibilities for increasing tax revenues and the opportunities to allocate that revenue to favorite projects and causes, supporters, and important voting blocks. Businesses that can arrange their affairs to take advantage of government policies stand to make great profits. Individuals may not see that the new taxes and government borrowing affect them

[12] See Green and Armstrong's 2008 essay "Uncertainty, the Precautionary Principle, and Climate Change" at http://kestencgreen.com/uncertainty-precautionary.pdf

[13] See, for example, 'Russian Lake Vostok Scientists Say "New Ice Age is Unavoidable"' at http://notrickszone.com/2012/04/27/russian-lake-vostok-researchers-say-ice-age-is-unavoidable/

[14] See, for example, Robinson, Robinson, and Soon (2007) on the beneficial effects of elevated levels of atmospheric CO_2.

and by how much. When people are fearful that some bad thing may happen, such as a five meter sea level rise[15] (Hansen and Sato 2012), they tend to ignore that the probability the event will occur is low (Sunstein and Zeckhauser 2011). In addition, when fearful people can blame others, such as large corporations or foreigners, or have no control over the event, they tend to demand that those who they see as responsible should pay large sums of money to make their fear go away. In contrast, individuals are unwilling to spend their own money to make the fear go away. These normal human biases lead to voting and policy decisions that reduce welfare.

Conclusions

In GAS, we concluded that forecasts from the naïve model are so accurate for long-term forecasts (up to 100 years), that there would be no economic benefit from improving the accuracy of the forecasts. In our judgment, the F&K findings support our conclusion.

The forecasts from the six models that F&K tested are less accurate than naïve model forecasts for horizons beyond ten years, and F&K's longer forecasts can be considerably improved in accuracy by damping them toward the naïve model forecast. The findings are consistent with the principle that one should forecast conservatively in complex, poorly understood, and uncertain situations.

What if forecasts of manmade global warming were valid? Responsible policy makers would still need scientific long-term forecasts of the local effects of global warming, both good and bad, on people with diverse endowments and preferences. And then they would need scientific forecasts of the effects of feasible alternative policies, including taking no action. Only then would it be possible to conduct proper cost-benefit analyses in order to identify policy alternatives that could reasonably be expected to substantially improve human welfare.

References

Allen, P.G. & Fildes, R. (2001). Econometric forecasting. In Armstrong, J.S. (ed.) *Principles of forecasting,* Kluwer: Boston. 303–362.

Armstrong, J. S. & Collopy, F. (1992). Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting*, *8*, 69–80

Armstrong, J. S., Green, K. C., & Soon, W. (2011). Research on forecasting for the manmade global warming alarm. *Energy and Environment*, 22, 1091-1104.

Armstrong, J. S., Green, K. C., & Soon, W. (2008). Polar bear population forecasts: A public-policy forecasting audit. *Interfaces*, 38, 382-404.

Baliunas, S. & Soon, W. (1999). Pioneers in the greenhouse effect, *World Climate Report*, 4 (no. 19, June 14 issue), 1-5. (available http://www.worldclimatereport.com/archive/previous_issues/vol4/v4n19/cutting.htm).

Beenstock, M., Reingewertz, Y., and Paldor, XN (2012), Polynomial cointegration tests of anthropogenic impact on global warming, *Earth System Dynamics,* 3, 173-188.

Breitenmoser, P., Beer, J., Bronnimann, S., Frank, D., Steinhilber, F., & Wanner, H. (2012). Solar and volcanic fingerprints in tree-ring chronologies over the past 2000 years. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 313-314, 127–139.

Christy, J. R., Herman, B., Pielke, R., Sr., Klotzbach, P., McNider, R. T., Hnilo, J. J., Spencer, R. W., Chase, T., & Douglass, D. (2010). What do observational datasets say about modeled tropospheric temperature trends since 1979? *Remote Sensing*, 2, 2148-2169.

Dawson, A., Palmer, T.N., & Corti, S. (2012). Simulating regime structures in weather and climate prediction models. Geophysical Reseasrch Letters, **39**, doi 10.1029/2012GL053284.

Fall, S., Watts, A., Nielsen-Gammon, J., Jones, E., Niyogi, D., Christy, J.R., and Pielke Sr., R.A. (2011) Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. *Journal Geophysical Research*, vol. 116, doi:10.1029/2010JD015146.

Fildes, R. & Kourentzes, N. (2011). Validation and forecasting accuracy in models of climate change. *International Journal of Forecasting*, *27*, 968–995.

Gao, C., Robock, A., & Ammann, C. (2008). Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. J Geophys. Res. **113**, D23111, doi 10.1029/2008JD010239.

[15] In order to understand the extremely low or near-zero probability of a 5 meter global sea level rise scenario, we quote Pfeffer *et al*. (2008, p. 1340): "We find that a total sea-level rise of about 2 meters by 2100 could occur under physically possible glaciological conditions but only if all variables are quickly accelerated to extremely high limits. More plausible but still accelerated conditions lead to total sea level rise by 2100 of about 0.8 meter. These roughly constrained scenarios provide a "most likely" starting point for refinements in sea-level forecasts that include ice flow dynamics."

Green, K. C. & Armstrong, J. S. (2007). Global warming: forecasts by scientists versus scientific forecasts. *Energy and Environment*, *18*, No. 7+8, 995–1019.

Green, K. C., Armstrong, J. S. & Soon, W. (2009). Validity of Climate Change Forecasting for Public Policy Decision Making. *International Journal of Forecasting*, *25*, 826–832.

Hansen, J.E. (2007) Scientific reticence and sea level rise, *Environmental Research Letters*, 2, doi:10.1088/1748-9326/2/2/024002.

Hansen, J.E., and Mki. Sato, 2012: Paleoclimate implications for human-made climate change. In Climate Change: Inferences from Paleoclimate and Regional Aspects. A. Berger, F. Mesinger, and D. Šijački, Eds. Springer, pp. 21-48, doi:10.1007/978-3-7091-0973-1_2.

Houghton, J. T., Callander, B. A., and Varney, S. K. (eds.) (1992). *Climate Change 1992: The supplementary report to the IPCC scientific assessment*. Cambridge, U.K.: Cambridge University Press for Intergovernmental Panel on Climate Change.

IPCC (2007). *Climate Change 2007: Synthesis Report, Contribution of Working Group I, II and III to the Fourth Assessment Report to the Intergovernmental Panel on Climate Change*. Edited by Core Writing Team, Pachauri, R.K. and Reisinger, A., IPCC Geneva, Switzerland. [Available from: http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_synthesis_report.htm

Jiang, J.H., and 29 co-authors (2012) Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. Journal of Geophysical Research, 117, doi:10.1029/2011JD017237.

Jones, P.D., Lister, D.H., Osborn, T.J., Harpham, C., Salmon, M. and Morice, C.P., 2012: Hemispheric and large-scale land surface air temperature variations: an extensive revision and an update to 2010. Journal of Geophysical Research, 117, doi:10.1029/2011JD017139.

Jouzel, J., and 31 co-authors (2007). Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, 317, 793–796.

Lindzen, R.S. (1994) Climate dynamics and global change. *Annual Review of Fluid Mechanics,* vol. 26, 353-378.

Lindzen, R.S., and Choi, Y.-S. 2011. On the observational determination of climate sensitivity and its implications. *Asia-Pacific Journal of Atmospheric Sciences*, vol. 47, 377-390.

Loehle, C., & McCulloch, J. H. (2008). Correction to: A 2000-year global temperature reconstruction based on non-tree ring proxies. *Energy & Environment*, 19, 93–100.

Mauritsen, T., and 14 co-authors (2012), Tuning the climate of a global model, Journal of Advances in Modeling Earth Systems, 4, M00A01, doi:10.1029/2012MS000154.

McKitrick, R.R., and Michaels, P.J. (2007) Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data. *Journal Geophysical Research*, vol. 112, doi:10.1029/2007JD008465.

McKitrick, Ross R. and Nicolas Nierenberg (2010). Socioeconomic Patterns in Climate Data. *Journal of Economic and Social Measurement*, 35(3,4) pp. 149-175.

McNider, R.T., Steeneveld, G.J., Holstag, A.A.M., Pielke Sr., R.A., Mackaro, S., Pour-Biazar, A., Walters, J., Nair, U., and Christy, J.R. (2012) Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. *Journal Geophysical Research*, vol. 117, doi:10.1029/2012JD017578.

Meyer, W. (2009). Catastrophe denied: A critique of catastrophic man-made global warming theory. Annotated slides from a presentation in Phoenix, AZ, November 10. [Available from http://www.climate-skeptic.com/Climate%20Presentation%20Annotated%201-1-2010.pdf]

Michaels, P. J. (2009). The dog ate global warming. *National Review Online*, September 23, http://www.nationalreview.com/articles/228291/dog-ate-global-warming/patrick-j-michaels

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, J. Geophys. Res., 117, D08101, doi:10.1029/2011JD017187.

New Scientist (2007). Climate Catastrophe, 28 July issue, 30–34.

New Zealand Climate Science Coalition (2011). Statistical Audit of the NIWA 7-Station Review. The New Zealand Climate Science Coalition [Available from http://nzclimatescience.net/images/PDFs/main_rept_on_niwa_7ss_review%20v5.pdf]

Qian, C., Wu, Z., Fu, C., and Wang, D. (2011), On changing El Ninos: A view from time-varying annual cycle, interannual variability, and mean state, *Journal of Climate,* 24, 6486-6500.

Parris, A., P. Bromirski, V. Burkett, D. Cayan, M. Culver, J. Hall, R. Horton, K. Knuuti, R. Moss, J. Obeysekera, A. Sallenger, and J. Weiss. (2012). Global Sea Level Rise Scenarios for the US National Climate Assessment. NOAA Tech Memo OAR CPO-1. 37 pp.

Pfeffer, W. T., Harper, J.T., & O'Neel, S. (2008). Kinematic constraints on glacier contributions to 21st-century sea-level rise. *Science*, 321, 1340–1343.

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.

Rasool, S. I., & Schneider, S. H. (1971). Atmospheric carbon dioxide and aerosols: Effects of large increases on global climate. *Science*, 173 (3992), 138–141.

Robinson, A. B., Robinson, N. E., & Soon, W. (2007). Environmental effects of increased atmospheric carbon dioxide. *Journal of American Physicians and Surgeons*. 12, 79–90.

Soon, W. (2009) Solar Arctic-mediated climate variation on multidecadal to centennial timescales: Empirical evidence, mechanistic explanation, and testable consequences. *Physical Geography*, 30, 144−184.

Soon, W., Baliunas, S., Idso, C., Idso, S., and Legates, D.R. (2003). Reconstructing climatic and environmental changes of the past 1000 years: a reappraisal. *Energy & Environment*, 14, 233-296.

Soon, W., Baliunas, S. Idso, S.B., Kondratyev, K.Ya. & Posmentier, E.S. (2001). "Modeling climatic effects of anthropogenic carbon dioxide emissions: Unknowns and uncertainties." Climate Research **18**: 259-275.

Soon, W., Dutta, K., Legates, D.R., Velasco, V. and Zhang, W. (2011) Variation in surface air temperature of China during the 20th Century. *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 73, 2331-2344.

Soon, W. and Legates, D.R. (2013) Solar Irradiance Modulation of Equator-to-Pole (Arctic) Temperature Gradients: Empirical Evidence for Climate Variation on Multi-decadal Timescales. *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 93, 45-56.

Sunstein, Cass R. & R. Zeckhauser (2011), Overreaction to fearsome risks, *Environmental and Resource Economics*, 48, 435-449.

Tetlock, P.E. (2005). *Expert political judgement*. Princeton.

Turner, J., Bracegirdle, T., Philips, T., Marshall, G.J., and Hosking, J.S. (2012), An initial assessment of Antarctic sea ice extent in the CMIP5 models, *Journal of Climate,* in press, doi:10.1175/JCLI-D-12-00068.1.

USGCRP (2009) *Global Climate Change Impacts in the United States*, Thomas R. Karl, Jerry M. Melillo, and Thomas C. Peterson (eds), Cambridge University Press.

Washington Post Editorial Board (2013). A sweltering planet's agenda. January 13.

Vermeer, M. and Rahmstorf, S. (2009), Overreaction to fearsome risks, *Proceedings of the National Academy of Sciences*, 106, 21527-21532.

Watts, A., Jones, E., McIntyre, S., and Christy, J.R. (2012) An area and distance weighted analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. PREPRINT DRAFT DISCUSSION PAPER [available for open peer review here: http://wattsupwiththat.files.wordpress.com/2012/07/watts-et-al_2012_discussion_paper_webrelease.pdf]

Wu, Z., Huang, N.E., Wallace, J.M., Smoliak, B.V., and Chen, X. (2011), On the time-varying trend in global-mean surface temperature, *Climate Dynamics,* 37, 759-773.

Wunsch, C. (2010) Towards understanding the Paleocean. *Quaternary Science Reviews*, vol. 29, 1960-1967.