SAND2011-8232C

Comparison of Climate Forecasts: Expert Opinions vs. Prediction Markets

Mark Boslough
Discrete Math & Complex Systems
Sandia National Laboratories
Albuquerque, NM 87185
mbboslo@sandia.gov

Expert elicitation is a method that is often used to estimate probability density functions for risk assessment. A very diverse set of opinions can sometimes be used to reach a consensus. The Third Santa Fe Conference on Global and Regional Climate Change brings together experts with a remarkably broad range of opinion about climate sensitivity. The poster session provides an opportunity to elicit and quantify the resulting breadth of expert opinion about the probability of various levels of future global warming. As an experiment, we will ask participants to go on the record with estimates of probability that the global temperature anomaly for calendar year 2012 will be equal to or greater than x, where x ranges in increments of .05 °C from .30 to 1.10 °C (relative to the 1951-1980 base period, and published by NASA GISS). The result will be a matrix of probabilities, a set of 17 for each participant. These estimates will be compared to those aggregated by prediction market contracts.

Introduction

Probabilistic risk assessment is a widely-used technique that was originally applied to quantify the uncertainty and consequences of potential failure modes of advanced engineered systems such as nuclear reactors, waste repositories, aircraft, and space vehicles. It is particularly applicable to systems in which the risk is dominated by low-probability, high-consequence events. One highly successful application has been to assess and reduce the hazard from asteroid impacts, for which large but highly unlikely globally-catastrophic events have a disproportionate contribution (*Chapman and Morrison*, 1994). Probabilistic risk assessment was used to inform a policy aimed at reducing the uncertainty in the number of asteroids on potential collision courses by funding observational programs to discover, catalog, and characterize asteroids in orbits that cross Earth's orbit. The Spaceguard Survey was created with the goal of discovering 90% of Earth-crossing asteroids with diameters greater than 1 km. Climate models and expert opinions suggest that impacts of this size can cause a global climate catastrophe, leading to a billion or more deaths due to environmental, agricultural, and economic collapse. The asteroid impact threat, at its core, is primarily a climate change threat. Due to the success of probabilistic risk assessment and the resulting Spaceguard Survey, the assessed risk of impact-induced climate catastrophe—with the potential to bring down civilization—has been reduced by an order of magnitude (Harris, 2008).

Uncertainty is a key component of risk assessment, and is often separated into "epistemic" (lack-of-knowledge) and "aliatory" (stochastic) uncertainty. Such a separation is more methodological than fundamental, and depends on context. For example, the known size distribution of near-Earth asteroids can be converted to a probability density function (PDF) using a methodology based on aliatory uncertainty (Boslough, 2011) but in reality the uncertainty is epistemic. The asteroids are all in well-defined, deterministic orbits and either will or will not collide with the Earth. We model them as if they were stochastic because we lack knowledge of their orbital elements and their whereabouts. Likewise, uncertainties in the fundamental properties of the Earth's climate system are due to lack of knowledge, but can be treated as stochastic. It can be argued that probabilistic risk assessment for climate change should be treated no differently than that for asteroids or engineered systems, in which the primary focus of uncertainty reduction should be placed on reducing the probability of rare events.

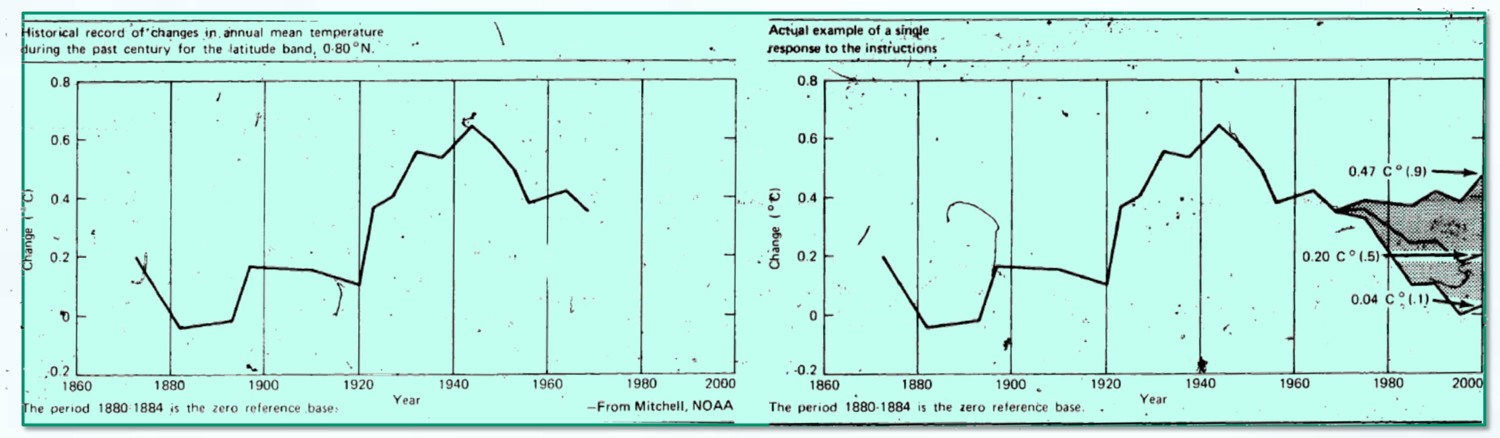
For systems in which uncertainty is dominated by epistemic uncertainty, opinions of subject-domain experts are elicited and combined mathematically, for example by adding more weight to those with higher level-of-expertise ratings by their peers. These methods were pioneered in the 1970s for reports related to reactor safety, climate effects of stratospheric ozone depletion , and climate change (e.g. National Defense University, 1978). Expert elicitation can be a highly effective way to quantify uncertainty and assess risk if properly implemented, but can also generate misleading results (Arkes, 1997). Climate change is a subject that evokes a high degree of emotion and political polarization. The lack of formal rigor and numerous methodological pitfalls associated with expert elicitation suggests that any result would immediately be dismissed as "just opinion" and criticized as "elitist" for putting more weight on the answers of acknowledged subject-domain experts. In the current political climate, it is unlikely to inform policy.

There are two potential modifications that address this issue: 1) Rather that elicit estimates of probabilities of rare events, ask participants to make actual forecasts that are sufficiently short term to be compared to an actual outcome, but sufficiently long-term to have a measureable climate component in the long run; 2) Implement a prediction market that consists of an ensemble of contracts that represent the forecasts associated with the expert forecasts. Prediction markets have been shown to be highly successful at forecasting the outcome of events ranging from elections to box office returns. In prediction markets, traders can take a position on whether some future event will or will not occur. These positions are expressed as contracts that are traded in a double-auction market that aggregates price, which can be interpreted as a consensus probability that the event will take place. Since climate sensitivity cannot directly be measured, it cannot be predicted. However, the changes in global mean surface temperature are a direct consequence of climate sensitivity, changes in forcing, and internal variability. Viable prediction markets require an undisputed event outcome on a specific date. They are capable of distinguishing the level of acceptance of the various global warming hypotheses, even by their respective proponents. Moreover, they can be used to determine a consensus estimate of future warming and climate variability that is weighted according to level of risk taken on by those providing the estimates, while filtering out the opinions of individuals unwilling to accept any financial risk associated with being wrong.

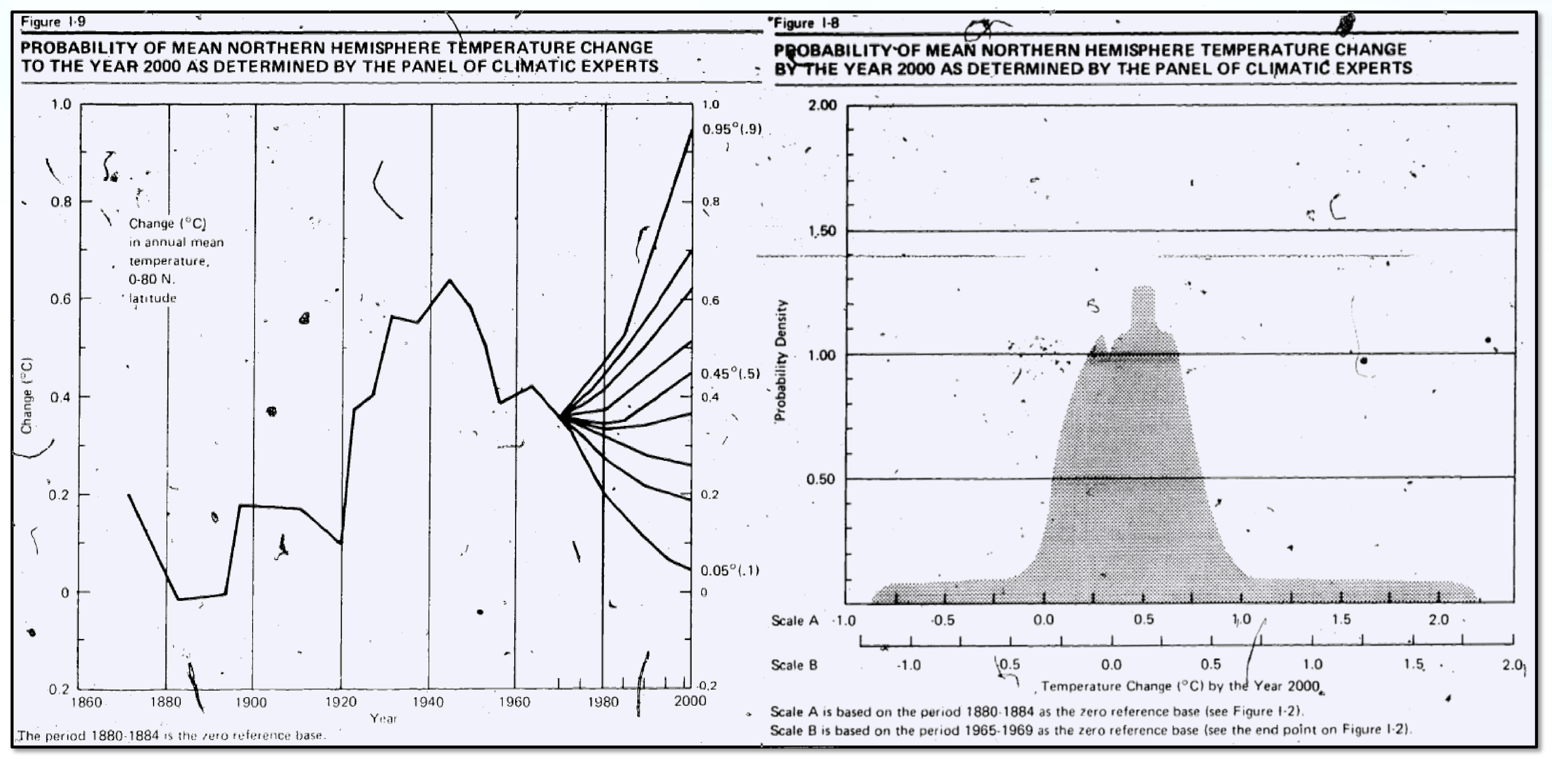
Expert Opinions and Climate Forecasting Circa 1978

Forecasts of anthropogenic global warming in the 1970s (e.g. Broecker, 1975) were taken seriously by policy makers. At that time, climate change was already broadly recognized within the US defense and intelligence establishments as a threat to national and global security, particularly due to climate's effect on food production. There was a high degree of uncertainty about the degree of global warming, and media-hyped speculation about global cooling that confused the public. Because science-informed policy decisions needed to be made in the face of this uncertainty, the US Department of Defense funded a study by National Defense University (NDU) called "Climate Change to the Year 2000" in which a panel of experts was surveyed. Stewart and Glantz (1985) provide a detailed critical analysis of the NDU report in which they discuss major pitfalls of expert elicitation. The following figures summarize the questions and results of the global warming part of the NDU survey.

According to the report summary, "Question 1, dealing with possible changes in global mean surface temperature [for the purpose of this study, 'global temperature' is used as equivalent to annual mean temperature between 0 and 80 north latitude], was a pivotal question because perceptions of global mean temperature greatly influence perceptions with respect to the climate variables treated in subsequent questions."



Question 1 referred to the left-hand graph in the above figure, which showed the annual mean temperature (NOAA) for the previous 100 years.. "Each respondent was asked to provide three estimates of the future course of possible changes in global temperature to the year 2000. The first estimate was to be a temperature path to the year 2000 such that there was only one chance in 10 that the actual path would be even lower. The second estimate was to be a path with an even chance that temperature could be either lower or higher, and the third was a path based on 1 chance in 10 that it could be even higher." This method, in effect, defines a crude PDF. The right-hand side of the figure shows a sample response, suggesting that such a PDF could be constructed for any year between 1978 and 2000.



The left-hand panel of the above figure shows a graph the aggregated percentiles, from which the aggregate cumulative density function (CDF) could be determined as a function of year between 1978 and 2000. Notably, the entire CDF is consistent with global warming since the base period (1880-1884), contradicting the modern blog-fueled myth that scientists in the 1970s were predicting "global cooling" and an impending ice age. The CDF was also transformed to an aggregate PDF (right-hand panel).

References

Arkes, H, et al., (1997), Combining expert opinions. *Science* 275, 461-465.

Boslough. M, et al., in review, (2011), Arguments and evidence against a Younger-Dryas impact. In *Climates, Past Landscapes and Civilizations* (AGU Chapman Conference Volume), Editors: Giosan, L; DQ Fuller; K Nicoll; R Flad and PD Clift.

Broecker, WS, (1975), Climatic change: Are we on the brink of a pronounced global warming? *Science* 189, 460-463.

Chapman, CR, and D Morrison, (1994), Impacts on the Earth by asteroids and comets: assessing the hazard. *Nature* 367, 33-40.

Harris, A, (2008), What Spaceguard did. *Nature* 453, 1178-1179.

Lindzen, RS and Y-S Choi, (2009), On the determination of climate feedbacks from ERBE data. *Geophys. Res. Lett*. 36., L16705.

National Defense University, (1978), *Climate change to the year 2000*. Washington, DC, 137 pp.

Stewart, TR and M.H. Glantz, (1985), Expert judgment and climate forecasting: a methodological critique of "Climate change to the year 2000." *Climatic Change* 7, 159-183.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Call for 2012 Forecasts

If you wish to participate in this forecasting experiment, send a message to Mark Boslough (mbboslo@sandia.gov). Participation is entirely voluntary. Forecasts received on or before Nov. 30, 2011 may be included in a poster to be presented at the American Geophysical Union Fall Meeting, Dec. 5-9, 2011, San Francisco (session GC11A. Climate Modeling: Uncertainty Quantification and its Application to Climate Change I Posters, 8:00 AM – 12:20 PM, Monday, Dec. 4, 2011).

If you wish to make an "unofficial" forecast now for discussion purposes, please use the space below, or attached sheets. Please write your estimate of the probability that the global temperature anomaly (°C) for calendar year 2012 (J-D as published by the NASA Goddard Institute for Space Studies) will be greater than or equal to the figure at the top of the column. Please use two significant figures. These estimates will be compared to the aggregate estimates determined by prediction markets at Intrade.com.

Name	≥.30	≥.35	≥.40	≥.45	≥.50	≥.55	≥.60	≥.65	≥.70	≥.75	≥.80	≥.85	≥.90	≥.95	≥1.00	≥1.05	≥1.10

Using Prediction Markets for Climate Forecasting

One axiom of the verification and validation community is, "codes don't make predictions, people make predictions." This is a statement of the fact that subject domain experts generate results using assumptions within a range of epistemic uncertainty and interpret them according to their expert opinion. Different experts with different methods will arrive at different PDFs. For effective decision support, a broadly-accepted consensus PDF would be useful. We suggest that market methods are a superior way to aggregate an ensemble of opinions into a single distribution that expresses the consensus.

Typical climate contracts predict the probability of a specified future temperature, but not the probability density or best estimate. One way to generate a probability distribution is to create a family of contracts over a range of specified temperatures and interpret the price of each contract as its exceedance probability. The resulting plot of probability vs. anomaly is the market-based cumulative density function. The best estimate can be determined by interpolation, and the market-based uncertainty estimate can be based on the spread.

Intrade is an online trading exchange that has created a family of 17 climate prediction markets for an experiment to be presented at the AGU Fall 2011 Meeting. Each contract can be described as "Global temperature anomaly for 2012 to be greater than x °C or more," where the figure x ranges in increments of .05 from .30 to 1.10 (relative to the 1951-1980 base period), based on data published by NASA GISS. Each market will settle at $10.00 if the published global temperature anomaly for 2012 is equal to or greater than x, and will otherwise settle at $0.00.

Global warming hypotheses can be cast as probabilistic predictions for future temperatures. The first modern such climate prediction is that of Broecker (1975), whose temperatures are easily separable from his CO_2 growth scenario—which he overestimated—by interpolating his table of temperature as a function of CO_2 concentration and projecting the current trend into the future.

For the currently-expected 2012 concentration of 395 ppm, Broecker's equilibrium temperature anomaly prediction relative to pre-industrial is 1.05 °C, or about 0.75 °C relative to the GISS base period. His neglect of lag in response to the changes in radiative forcing was partially compensated by his low sensitivity of 2.4 °C, leading to a slight overestimate. Simple linear extrapolation of the current trend since 1975 yields an estimate of .65 ± .09 °C (net warming of .95 °C) for anthropogenic global warming with a normal distribution of random natural variability.

To evaluate an extreme case, we can estimate the prediction Broecker would have made if he had used the Lindzen & Choi (2009) climate sensitivity of 0.5 °C. The net post-industrial warming by 2012 would have been 0.21 °C, for an expected change of -0.09 from the GISS base period. This is the temperature to which the Earth would be expected to revert if the observed warming since the 19th century was merely due to random natural variability that coincidentally mimicked Broecker's anthropogenic change prediction for the past 36 years.

Assertions made outside the scientific literature can also be cast into predictions for 2012 temperatures, for example Carter's (2006) argument for a lack of warming since 1998 can be extrapolated to a 2012 value of 0.56 °C (net warming of .86 °C), and Easterbrook's (2010) claim of global cooling can be extrapolated to a 2012 value of .42 °C (net warming of .72 °C).

All 17 contracts in the newly-opened market ensemble are consistent with net warming from pre-industrial temperatures. They are also capable of distinguishing the level of acceptance of the various global warming hypotheses, even by their respective proponents. Moreover, they can be used to determine a consensus estimate of future warming and climate variability that is weighted according to level of risk taken on by those providing the estimates, while filtering out the opinions of individuals unwilling to accept any financial risk associated with being wrong.