Background Materials for a Workshop on:

Methods to Address Uncertainty in Forecasting Future Values of Key Social, Economic and Resource Variables

A workshop organized by the
Center for Climate and Energy Decision Making
which is anchored at Carnegie Mellon University
http://cedm.epp.cmu.edu/

2013 March 18–19

American Association for the Advancement of Science (AAAS) Building
1200 New York Avenue, NW
Revelle Conference Room
Washington, DC

Support for this workshop is provided by the Center for Climate and Energy Decision Making (CEDM), through a cooperative agreement between the National Science Foundation and Carnegie Mellon University (SES-0949710).

BACKGROUND NOTE:

Workshop on methods to address uncertainty in forecasting future values of key social, economic and resource variables.

Insight about how the future value of key social, economic or resource variables,[1] and how they will evolve over time, would be extremely valuable to many public and private parties engaged in planning or faced with making decisions that have long-term irreversible consequences. As a consequence, many individuals and organizations are in the business of trying to produce projections, forecasts,[2] scenarios or other products designed to help inform such decisions.

While some of these parties are careful to frame their outputs with cautionary language, most take the form of single value time series that run into the future. However, the product is framed, many users promptly drop or ignore the cautionary language, and treat the time series as a prediction. This is especially true of products produced by organizations such as the U.S. Energy Information Agency (EIA) and the International Energy Agency (IEA).[3]

Descriptions of how the future may unfold have been especially important in the climate and energy research and policy communities. While recent decades have witnessed enormous improvements in global circulation climate models, that now include full dynamically coupled models of the ocean and atmosphere, one of the largest uncertainties in the outputs of these models is introduced by ambiguity about how anthropogenic emission of greenhouse gasses, atmospheric concentrations of fine particles, and land use and land cover, will evolve over coming decades. To address these issues, groups like the IPCC have created scenarios which are intended to help users think about the range of futures that might be possible – but as Morgan and Keith (2008) argue in one of the attached readings, there are a variety of problems with the use of such scenarios.

Things that make forecasting and prediction difficult or impossible

There are some future quantities and events that can be predicted with remarkable precision. For example, thanks to Newtonian mechanics we can confidently say that there will be a total eclipse of the Sun on September 4, 2100. Indeed, we can even say that to within a fraction of a second, the moment of maximum eclipse will occur at 16:57:52 GMT. But given the time series shown in Figure 1, nobody in their right mind could

---

1 Examples include the future value of quantities such as population, GDP, energy consumption, prices of key commodities, and the availability of key resources - either at discrete moments in the future or in the form of time series.

2 In this background note, we use the terms forecast and projection interchangeably and have not used the word prediction. For a discussion of how the IPCC uses these and other terms, see Attachment 1.

3 A few organizations, such as the Population Division of the United Nations Department of Economic and Social Affairs, have begun to produce high and low scenarios together with their best estimate projections, but such treatment is still rare, and whether and how users make use of such ranges is not yet well understood. See, for example: http://www.un.org/esa/population/publications/longrange2/WorldPop2300final.pdf.

plausibly argue that we can predict U.S. oil or gas prices to ±50% in 10, let alone 50, years.

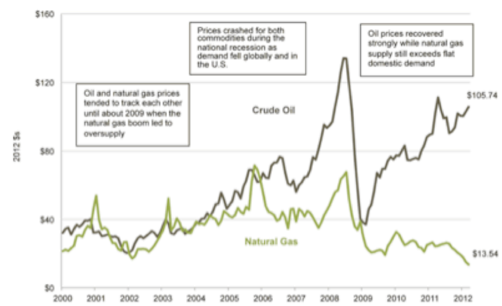

Fig 1: Time series of past prices of crude oil and natural gas in the U.S. Source: headwaterseconomics.org

Yet many groups make such deterministic forecasts all the time, and many economists, policy analysts and decision makers, use such forecasts as though they are entirely reliable. Of course, as several papers in the attached readings (e.g., Smil (2000), Craig et al. (2002)) make clear, when one looks back at how well these forecasts have actually performed, they have often done very poorly. For example, Figure 2 shows a retrospective look at a variety of forecasts of U.S. primary energy consumption in the year 2000 that were made between 1960 and 1980. Even those that reported fairly wide uncertainty bands did not include the value that actually obtained in 2000.[4]

More robust forecasts are typically possible when there are good data on past performance, when the statistics of the underlying processes are stationary, and when the structure of the underlying causal models are stable.

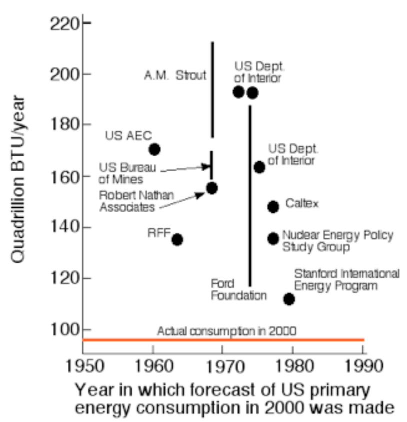

Fig 2: Comparison of a variety of forecasts of U.S. primary energy consumption in the year 2000 that were made between 1960 and 1980. Redrawn from a figure compiled by Vaclav Smil (2003).

A variety of factors can contribute uncertainty to the future value of key social, economic and resource variables. These include:

- random and/or chaotic physical processes;
- future choices made by key decision makers;
- the emergent consequences of many individual "agents;"
- the introduction of new technology.

Brownian motion is an example of random physical processes. Similarly, the evolution of weather over periods of several weeks depends on the chaotic unfolding of atmospheric processes. Two good examples of a critical choice made by key decision makers are the decision by Vannevar Bush and his colleagues to organize the U.S. research establishment before the county's entry into World War II and the decision by Winston Churchill to share all of the UK's advanced defense

4 Readers familiar with the literature on overconfidence, a topic discussed in a later section of this background note, will not be surprised by the fact that these ranges did not include the value that actually obtained in 2000.

technology with U.S. scientists while the U.S. was still neutral.[5] Perhaps the U.S. would have developed the equivalent of the Radiation Laboratory at MIT and developed high frequency radar, the proximity fuse, and a host of other technologies without the choices made by these two men. But, such developments would almost certainly been slower in coming without their choices, and might not have played the pivotal role they did in the allied victory.

In addition to choices made by key individuals, there are also the emergent consequences of choices made by millions of individuals. Such choices drive the evolution of markets. They are also critical in the development of social and political movements, such as the rise of environmentalism in the 1960s or the Arab Spring that has played out in recent years.

Technological change is another factor that makes it difficult or impossible to forecast the future values of key social and economic variables. Consider, for example, the impossibility in 1900 of forecasting the value of any number of key economic, demographic and cultural variables for the latter part of the 20th century without knowing about the development of aircraft, penicillin, the transistor, computers and microelectronics.

> *DISCUSSION QUESTIONS 1:* We have listed four factors that can make it difficult or impossible to forecast the value of some key social and economic variables. Are there other factors that we have missed? Could we build a useful taxonomy of when, and for what variables, various factors become most important?

The basic problem

Figure 3 provides an abstract representation of a forecast of the value of a variable of interest. While many forecasts involve just a single attribute (e.g., annual electricity sales) here for generality we show two, and though it is not easy to plot, some quantities of interest might involve multiple attributes. For example, in this figure, Attribute 1 might be U.S. retail electricity sales and Attribute 2 might be the average cost of electricity per kWh.

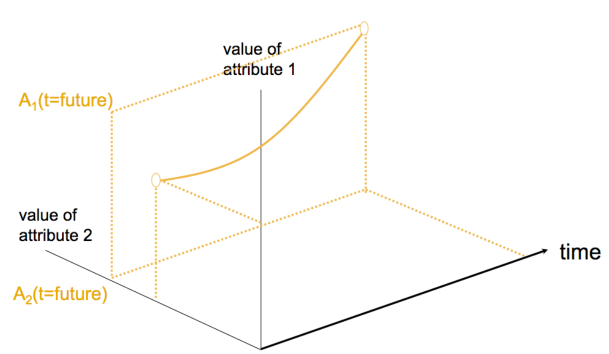

Fig 3: Example of a deterministic forecast over time for a quantity that involves two attributes.

While we have shown the forecast as a single line in this space, of course, as soon as one moves very

[5] See for example: G. Pascal Zachary, *Endless Frontier: Vannevar Bush, engineer of the American century*, The Free Press, 518pp., 1997; Vannevar Bush, *Pieces of the Action*, William Morrow Co., 366pp., 1970; Robert Buderi, *The Invention that Changed the World: How a small group of radar pioneers won the second world war and launched a technical revolution*, Simon and Schuster, 575pp., 1996.

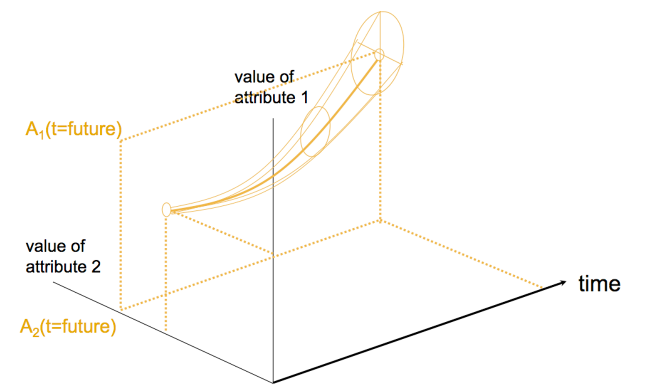

Fig 4: A more realistic version of the forecast shown in Fig 3, in which the cone of uncertainty about the value of the quantity being forecast grows as one moves further into the future.

far into the future, the values become uncertain. Thus, a more realistic representation is that shown in Figure 4, in which a cone of uncertainty grows larger as one moves further into the future.

These figures can be used to illustrate a key point that underlay the argument about attaching probabilities to the SRES scenarios that were used in the IPCC fourth assessment. Developers of those scenarios advanced several arguments for why they could not attach probabilities to those scenarios (see discussion in the paper by Morgan and Keith in the attached reading). However, at least to our knowledge they never made the one argument against assigning probabilities to their scenarios that is strictly correct. The SRES scenarios essentially consist of a line through an N-space of the sort shown in Figure 3. For conventional quantities, no probability attaches to a single point in a space, nor does a line through such a space carry a finite probability. However, one *can* attach probabilities to *regions* in such a space. Thus, it does make sense to talk about the probability that a quantity of interest will fall in each of the regions at t=T in Figure 5 or in any family of cones that lead from the present to those or similar regions. The same sort of arguments can be made about the Representative Concentration Pathways or RCPs that are being used in the fifth assessment.[6]

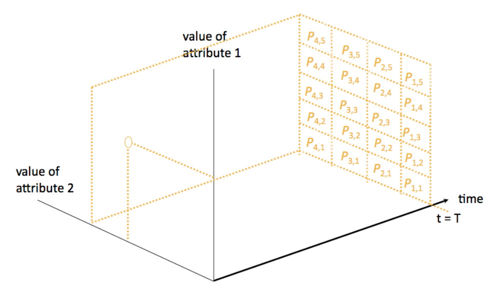

Fig 5: While one cannot attach a probability to the line through the space shown in Fig 1, one can attach a probability to any cone or family of cones through that space and to any array of end states such as those shown at t = T in this figure.

How well do laypeople understand uncertainty?

One possible argument against using strategies to describe the future that consist of something more than a single value trajectory, is that laypeople do not understand and cannot deal with uncertainty. However, we do not find that argument compelling. Essentially all important decisions that people face in their private lives involve considerable uncertainty: where to go to college; what career to enter; who to marry. We believe that while they may not fully understand the extent of the uncertainty that is involved, most people do understand that the outcomes associated with such choices are deeply uncertain.

---

[6] For details see http://sedac.ipcc-data.org/ddc/ar5_scenario_process/index.html.

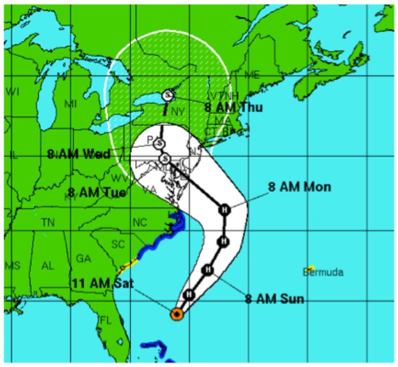

Fig 6: Example (for Hurricane Sandy) of the cone of uncertainty that NOAA now uses routinely in forecasting the uncertain future track of hurricanes.

More aligned with the issues we will address in this workshop, people also routinely deal with probabilistic weather forecasts, betting odds, and point spreads for sporting events. They may not understand, or process the uncertainty in strict compliance with the laws of probability, but most do recognize that outcomes cannot be predicted with precision. Even the idea of a spreading cone of future uncertainty is a concept that most people now grasp, witness the "cone of uncertainty" that NOAA's National Hurricane Center, and all major news outlets, now uses routinely (Figure 6).

Why do folks continue to make single-value forecasts?

We can think of six reasons why people and organizations might continue to make single-value time series forecasts. Specifically, folks persist in making deterministic projections because:

1. They are lazy – they find it too hard to do anything else;
2. They can get away with it because most people who get their forecasts have no appreciation of how bad past performance has been;
3. They believe (erroneously) that those who use the forecasts are not capable of understanding or dealing with uncertainty;
4. Deterministic forecasts are more persuasive in arguments than forecasts that come with any acknowledgment of uncertainty;
5. They believe that if they include uncertainty, people will perceive them to be less expert;
6. They have no idea what else they could do.

Clearly, not all of these reasons apply to all people and organizations making forecasts.

> *DISCUSSION QUESTIONS 2:* Are there other arguments that people and organizations might advance for persisting in making single-value time series forecasts? What is your view of the relative merits of these arguments?

Alternative strategies for estimating and describing future time series

A key objective of this workshop is to identify and discuss a range of strategies that might be adopted to assess and communicate time series of the future values of key social, economic and resource variables. In the readings, we have included six papers that outline a few such strategies. These include:

- Casman et al. (1999), which proposes the use of bounding analysis when uncertainty about model functional form becomes very high;
- Morgan and Keith (2008), which outlines a strategy to construct scenarios that are not lines but rather span a space of interest;

- Schweizer and Morgan (submitted paper), which performs a bounding analysis for US electricity demand in 2050;
- Raftery et al. (2012), which demonstrates the application of a Bayesian strategy to perform probabilistic population projections.

> *DISCUSSION QUESTIONS 3:* What are the strengths and limitations of the several methods outlined in the readings? Can you suggest additional methods? What are the relative advantages and disadvantages of using these various methods – both in terms of applying the methods, and in terms of the usability of the resulting forecasts?

The importance of cognitive heuristics and of ubiquitous over confidence[7]
While our brains are very good at doing many tasks, we do not come hard-wired with statistical processors. Over the past several decades, experimental psychologists have begun to identify and understand a number of the "cognitive heuristics" we use when we make judgments that involve uncertainty.

The first thing to note is that people tend to be systematically overconfident in the face of uncertainty – that is, they produce probability distributions that are much too narrow. Actual values, once they are known, often turn out to lie well outside the tails of their previous distribution. This is well illustrated with the data in the summary table reproduced in Figure 7. This table reports results from laboratory studies in which, using a variety of elicitation methods, subjects were asked to produce probability distributions to indicate their estimates of the value of a variety of quantities whose value is known, but which most people would not carry in their head (e.g., the length of the Panama Canal).

If the respondents were "well calibrated," then the true value of the judged quantities should fall within the 0.25 to 0.75 interval of their probability distribution about half the time. The frequency with which the true value actually fell within that assessed interval is termed the interquartile index. Similarly, the frequency with which the true value lies below the 0.01 or above the 0.99 probability values in their distribution is termed the "surprise index." Thus, for a well-calibrated respondent, the surprise index should be 2%.

In these experimental studies, interquartile indices typically were between 20 and 40% rather than the 50% they should have been, and surprise indices ranged from a low of 5% (2.5 times larger than it should have been) to 50% (25 times larger than it should have been).

Overconfidence is not unique to non-technical judgments. Henrion and Fischhoff (1986) have examined the evolution of published estimates of a number of basic physical constants, as compared to the best modern values. Here too the results show consistent

---

7 Portion of the text for this section are slightly edited excerpts from a discussion that originally appeared in M. Granger Morgan et al., *Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Decisionmaking*, CCSP 5.2, A Report by the Climate Change Science Program and the Subcommittee on Global Change Research. National Oceanic and Atmospheric Administration, Washington, DC, 96pp., 2009.

overconfidence. For examples, see Figure 2 in Koomey et al. (2003) and Figure 7 in Morgan and Keith (2008) in the attached readings.

	Number of assessments *N*	Interquartile index (ideal 50%)	Surprise index (ideal 2%)
Alpert & Raiffa (1969)			
Group 1-A	880	33	46
Group 2 & 3	1,670	33	39
Group 4	600	36	21
Hession & McCarthy (1974)			
Fractiles	2,035	25	47
Selvidge (1975)			
Five fractiles	400	56	10
Seven fractiles	520	50	7
Schaefer & Borcherding (1973)			
Fractiles	396	23	39
Hypothetical sample	396	16	50
Pickhardt & Wallace (1974)			
Group 1	?	39	32
Group 2	?	30	46
Seaver, von Winterfeldt, & Edwards (1978)			
Fractiles	160	42	34
Odds-fractiles	160	53	24
Probabilities	180	57	5
Odds	180	47	5
Log-odds	140	31	20
Stael von Holstein (1971)			
Fixed intervals	1,269	27	30
Murphy & Winkler (1974 & 1977)			
Fixed intervals	132	45	27 (ideal 25)
Fractiles	432	54	21 (ideal 25)
Schaefer (1976)			
Fixed interval	660	27	25
Lichtenstein & Fischhoff (1978)			
Fractiles	924	33	41
Seaver (1978)			
Parameters of beta dist.	3,200	29	25

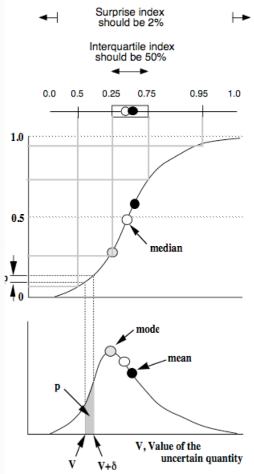

Fig 7: Summary of data from different studies in which, using a variety of methods, people were asked to produce probability distributions on the value of well known quantities (such as the distance between two locations), so that their distributions can be subsequently checked against true values. The results clearly demonstrate that people are systematically overconfident (*i.e.*, produce subjective probability distributions that are too narrow) when they make such judgments. The table is reproduced from Morgan and Henrion (1990) who, in compiling it, drew in part on Lichtenstein *et al.* (1982). Definitions of interquartile index and surprise index are shown in the diagram on the right.

Several cognitive heuristics[8] contribute to these findings, and otherwise distort or bias the judgments that people make about the value or probability of future events. Of these, the heuristic of availability[9] is probably the most relevant to our concerns in this workshop. When people judge the frequency of an uncertain event they often do so by the ease with which they can recall such events from the past, or can imagine such events occurring. A more detailed discussion of this heuristic and its consequences, especially in the context of detailed story lines in scenarios, is provided on pages 200 to 205 of Morgan and Keith (2008) in the attached readings. Problems we are addressing here are not helped by systematic evidence of "hindsight bias."[10] After the fact, people routinely judge their prospective judgments or forecasts to have been more accurate than they actually were. In this context, Fischhoff[10] has inferred from experimental studies he has run that:

> When we attempt to understand past events, we implicitly test the hypotheses of rules we use to both interpret and anticipate the world around us. If in hindsight, we systematically underestimate the surprises, which the past held and holds for us [as his other experimental results have demonstrated], we are subjecting those hypotheses to inordinately weak tests and, presumably finding little reason to change them. Thus, the very outcome knowledge which

---

8 Other heuristics include "representativeness" and "anchoring and adjustment." For a review, see D. Kahneman, P. Slovic, A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and biases*, Cambridge, 555pp., 1982.

9 A. Tversky and D. Kahneman, "Availability: A heuristic for judging frequency and probability," *Cognitive Psychology*, *4*, 207-232, 1973.

10 B. Fischhoff, "Hindsight ≠ Foresight: The effect of outcome knowledge on judgment under uncertainty," *Journal of Experimental Psychology*, *1*, 288-299, 1975; and S.A. Hawkins and R. Hastie, "Hindsight: Biases judgment of past events after the outcomes are known," *Psychological Bulletin*, *107*, 311- 327, 1990.

> gives us the feeling that we understand what the past was all about may prevent us from learning anything from it…

> *DISCUSSION QUESTIONS 4:* How should the existence of these cognitive heuristics shape our thinking about the problems we are exploring in this workshop? For example, are the concerns about their role in scenarios that are raised by Morgan and Keith (2008) overblown? Is the existence of hindsight bias likely to make it more difficult to persuade users of the need to develop more probabilistic strategies for describing forecasts?

Promoting adoption

It is fine to come up with improved methods for developing and communicating forecasts of key social, economic and resource variables that more adequately deal with and communicate our limited ability to make statements about the future. However, if nobody ends up using these methods, there is little point to working on the issue.

Potential creators and users of forecasts include government agencies; corporations, senior public and private leaders (Government Ministers, CEOs, etc.); economists and financial advisors; climate, energy and other modeling communities. If methods that better communicate forecast uncertainty are to begin to see wider use, the climate, energy and other modeling communities are the most likely to become earlier adopters.

Fifty years ago, virtually nobody performing risk assessments was doing serious uncertainty and sensitivity analysis. A number of folks in the analysis community, many of them with decision analytic backgrounds, began to focus on this issue, providing both worked examples demonstrating how it was possible to do uncertainty analysis and also advancing arguments for why it was important to treat uncertainty. Today, no federal agency, regulatory organization, or corporation would perform a risk analysis without explicit and systematic treatment of uncertainty. If we can develop improved methods for describing and dealing with uncertainty in forecasts, let's hope it does not take another 50 years to see them become widely adopted.

Attachment 1: A Note on Nomenclature

In this note, we have used the terms forecast and projection interchangeably and have not used the word prediction. Note, however, that the IPCC employs the following definitions:[11]

Projection: The term "projection" is used in two senses in the climate change literature. In general usage, a projection can be regarded as any description of the future and the pathway leading to it. However, a more specific interpretation has been attached to the term "climate projection" by the IPCC when referring to model-derived estimates of future climate.

Forecast/Prediction: When a projection is branded "most likely" it becomes a forecast or prediction. A forecast is often obtained using deterministic models, possibly a set of these, outputs of which can enable some level of confidence to be attached to projections.

Scenario: A scenario is a coherent, internally consistent and plausible description of a possible future state of the world. It is not a forecast; rather, each scenario is one alternative image of how the future can unfold. A projection may serve as the raw material for a scenario, but scenarios often require additional information (e.g., about baseline conditions). A set of scenarios is often adopted to reflect, as well as possible, the range of uncertainty in projections. Other terms that have been used as synonyms for scenario are "characterisation", "storyline" and "construction".

Scenario Family: One or more scenarios that have the same demographic, politico-societal, economic and technological storyline.

Storyline: A narrative description of a scenario (or a family of scenarios), highlighting the main scenario characteristics and dynamics, and the relationships between key driving forces.

Baseline/Reference: The baseline (or reference) is any datum against which change is measured. It might be a "current baseline", in which case it represents observable, present-day conditions. It might also be a "future baseline", which is a projected future set of conditions excluding the driving factor of interest. Alternative interpretations of the reference conditions can give rise to multiple baselines.

Exposure Unit: An exposure unit is an activity, group, region or resource exposed to significant climatic variations.

[11] See http://www.ipcc-data.org/ddc_definitions.html#anchorDefScenario.

Background Readings

The following papers are reproduced under the "fair use" exclusion for the non-profit educational use of participants in this workshop. Please do not circulate.

1. Papers that look at how well past projections have performed.

1.1 Paul P. Craig, Ashok Gadgil and Jonathan G. Koomey, "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*, pp. 83-118, 2002.

1.2 Vaclav Smil, "Perils of Long-Range Energy Forecasting: Reflections on looking far ahead," *Technology Forecasting and Social Change*, *65*, pp. 251-264, 2000.

1.3 Hans H. Landsberg, "Energy in Transition: View from 1960," *The Energy Journal*, *6*(2), pp. 1-18, 1985.

1.4 Hans Linderoth, "Forecast Errors in IEA-Countries' energy Consumption," *Energy Policy*, *30*, pp. 53-61, 2002.

1.5 James J. Winebrake and Denys Sakva, "An Evaluation of Errors in U.S. Energy Forecasts: 1982-2003," *Energy Policy*, *34*, pp. 3475-3483, 2006.

1.6 Brian C. O'Neill and Mausami Desai, "Accuracy of Past Projections of U.S. Energy Consumption," *Energy Policy*, *33*, pp. 979-993, 2005.

1.7 Hilliard G. Huntington, "Oil Price Forecasting in the 1980s: What went wrong?" *The Energy Journal*, *15*(2), pp. 1-22, 1994.

1.8 Bent Flyvbjerg, Mette K. Skamris Holm and Søren L. Buhl, "How Common and How Large are Cost Overruns in Transport Infrastructure Projects?," *Transport Reviews*, *23*(1), pp. 71-88, 2003.

1.9 Excerpts from two RAND Project Air Force Studies of Weapons System Cost Growth.

1.10 Jonathan Koomey, Paul Craig, Ashok Gadgil and David Lorenzetti, "Improving Long-Range Energy Modeling: A plea for historical retrospectives," *The Energy Journal*, *24*(4), pp. 75-92, 2003.

2. Papers that discuss or suggest various strategies for incorporating uncertainty into forecasts.

2.1 M. Granger Morgan and David W. Keith, "Improving the Way We Think About Projecting Future Energy Use and Emissions of Carbon Dioxide," *Climatic Change*, *90*, pp. 189-215, 2008.

2.2 Vanessa J. Schweizer and M. Granger Morgan, "Bounding U.S. Electricity Demand in 2050," paper in review at *Energy Policy*.

2.3 Adrian E. Raftery, Nan Li, Hana Sevciková, Patrick Gerland and Gerhard K. Heilig, "Bayesian Probabilistic Population Projections for all Countries," *Proceedings of the National Academy of Sciences*, 13915-13921, 2012.

2.4 Robert Lempert, "Scenarios that Illuminate Vulnerabilities and Robust Responses," *Climatic Change*, published on line, October 2012; doi:10.1007/s10584-012-0574-6.

2.5 Elizabeth A. Casman, M. Granger Morgan and Hadi Dowlatabadi, "Mixed Levels of Uncertainty in Complex Policy Models," *Risk Analysis*, *19*(1), pp. 33-42, 1999.

2.6 Vanessa Jine Schweizer and Elmar Kriegler, "Improving Environmental Change Research with Systematic Techniques for Qualitative Scenarios," *Environmental Research Letters*, *7*, 2012; doi:10.1088/1748-9326/7/4/044011.

NOTE: We have had difficulty finding additional examples. If you can suggest any, please let us know and if possible bring a copy to the workshop.