# Improving Long-Range Energy Modeling: A Plea for Historical Retrospectives

*Jonathan Koomey*, Paul Craig**, Ashok Gadgil*** and David Lorenzetti****

*One of the most striking things about forecasters is their lack of historical perspective. They rarely do retrospectives, even though looking back at past work can both illuminate the reasons for its success or failure, and improve the methodologies of current and future forecasts. One of the best and most famous retrospectives is that by Hans Landsberg, which investigates work conducted by Landsberg, Sam Schurr, and others. In this article, written mainly for model users, we highlight Landsberg's retrospective as a uniquely valuable contribution to improving forecasting methodologies. We also encourage model users to support such retrospectives more frequently. Finally, we give the current generation of analysts the kind of guidance we believe Landsberg and Sam Schurr would have offered about how to do retrospectives well.*

---

*It's better to be the drafter of a well-constructed plan,*
*For spending lots of money for the betterment of Man*
*But audits are a threat, for it is neither games nor fun*
*To look at plans of yesteryear and ask 'What have we done?'*
*And learning is unpleasant when we have to do it fast*
*So it's pleasanter to contemplate the future than the past.*

Section of *A Ballad of Ecological Awareness*, Kenneth Boulding, as quoted in Farvar and Milton (1972).

## INTRODUCTION

Hans Landsberg and Sam Schurr were giants in their field – founders of the post-World War II era of long-range energy and resource forecasting. We are proud to dedicate this paper to their contributions, which have stood the test

*The Energy Journal*, Vol. 24, No. 4. Copyright © 2003 by the IAEE. All rights reserved.

* Author for correspondence: MAP/Ming Visiting Professor of Energy and Environment, Stanford University, and Staff Scientist, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Building 90-4000, Berkeley, CA 94720, USA. E-mail: JGKoomey@lbl.gov

** Professor Emeritus, University of California, Davis, CA, USA.

*** Environmental Energy Technologies Division, Ernest Orlando Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA.

of time. Their work remains valuable today in part because they took great care to articulate their assumptions. Thus it is possible to look back and see just why they got it right or wrong, and why they were sometimes right for the wrong reasons.

One of the key lessons from the work of Landsberg and Schurr is the importance of retrospective analysis of forecasts. Looking with hindsight can both illuminate problems in past work and improve the methodologies of current and future forecasts. Other studies have described the general philosophy of improving modeling and forecasting, such as Armstrong (1978, 2001), Greenberger et al. (1976), Greenberger and Richels (1979), and Huntington et al. (1982), but none to our knowledge have explicitly examined the benefits of retrospective analysis in detail.

One of the best and most famous retrospectives is by Landsberg (1985), which investigates work by Landsberg, Schurr, and others. Almost two decades after Landsberg's retrospective, we set out to determine whether forecasters had taken his recommendations to heart. In particular, we wanted to learn whether retrospectives had become more widely used. We found only a handful of published retrospectives, few of them recent, comprehensive, or ongoing. In general, forecasters and their sponsors show a distinct lack of interest in conducting such studies and in evaluating the reasons for the success or failure of their past work.

Our interest in retrospectives is motivated in part by a belief that long-range forecasts are most useful and insightful when they clearly articulate underlying principles and fundamental driving forces. Retrospective analysis contributes to understanding these factors and hones instincts for interpreting future forecasts. Retrospective analysis is especially important for model users, since it helps them understand and constructively critique the work of model builders. We believe model users would be well-served by devoting more effort to retrospective analysis, as well as by developing systematic ways to undertake it.

In this article, we review analyses by Landsberg and Schurr (Landsberg 1985; Landsberg et al. 1963; Schurr et al. 1979; Schurr et al. 1960; Schurr et al. 1977). We discuss the benefits and limitations of forecasts, drawing on Hodges and Dewar's (1992) distinction between validatable and non-validatable models to suggest ways forecasts can be improved. Finally, we offer analysts the kind of guidance we believe Hans Landsberg and Sam Schurr would have given about how to do retrospectives well.

It is worth noting here that Paul Joskow's article in this same issue (Joskow, 2003) emphasizes the policy context within which Landsberg and Schurr analyzed emerging energy problems and prospective energy trends. From time to time the two articles overlap, but our own focus, as noted, is directed specifically towards the need for retrospectives, while Joskow gives insights relevant to the broader policy landscape.

## II. LANDSBERG'S RETROSPECTIVE

*Resources in America's Future* published in 1963 by the then decade-old Resources for the Future (RFF), was a landmark assessment of the demand and supply of all major U.S. resources from 1960 to 2000 (Landsberg et al. 1963).[1] The study combined economic and technical analysis. Economic factors were drawn primarily from U.S. government reports. The authors relied heavily on bottom-up trend analysis, supplemented by their professional judgment. Some assumptions were grounded in the laws of thermodynamics, but most energy technologies were so far from fundamental limits that these laws provided minimal constraint. Rather, technological innovation and human behavior were the dominant factors, and these factors proved hard to anticipate.

Landsberg revisited the report two decades later (Landsberg, 1985) in a 1984 luncheon address to the International Association of Energy Economists, published in this Journal. His stated goals were modest and, characteristically, a bit whimsical: “I figured I could use the occasion to satisfy my own curiosity, to pay the price of admission, and to give you a chance to relax.”

Landsberg’s perspective was philosophical: “One is a captive of the time of writing or calculating, typically without realizing it.” In his retrospective review he remarked on the failure to anticipate the oil embargoes of the 1970s. Actual energy growth, he noted, was higher from 1960 to 1980 than in the RFF forecast, and slowed dramatically thereafter (Figure 1, taken from Landsberg (1985), illustrates these trends). The period included the time of the OAPEC (Organization of Arab Petroleum Exporting Countries) embargo, when a structural shift occurred in energy use patterns.[2] The RFF study showed no such “break-point.” It assumed steady growth at a rate that led, fortuitously, to about the right outcome in 1980. The RFF forecasts become increasingly high in the 1980 to 2000 period as actual energy use continued to lag projected use (141 EJ primary energy demand in 2000 in the "medium" projection versus 103 EJ actual).

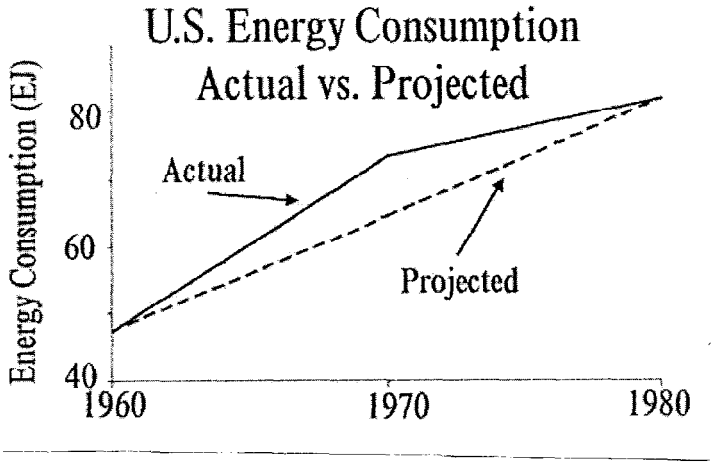
In the post-World War II era, energy growth was strongly correlated with economic growth. The assumption that this correlation would continue without alteration was responsible for many overestimates of energy use (Craig et al. 2002). For the period 1960-1980, Landsberg noted, his predictions of the various components of energy use erred

> in opposite directions and in roughly offsetting proportions. We underprojected coal and hydro, overprojected gas and nuclear, and got oil about right. We underestimated personal consumption expenditures but overestimated both investment and government purchases, with GNP coming out a little lower than reality (Landsberg, 1985).

1. The study also drew heavily upon a previous analysis by Schurr et al. (1960).

2. Oil shortages at this time were attributed to the embargo, but retrospective analysis has made a compelling case that hoarding and excessive inventory build-up were the real causes (Bohi, 1989).

**Figure 1. Schematic diagram illustrating how the 1963 study *Resources in America's Future* correctly predicted energy use in 1980, owing to compensating errors**




Notes: The forecast energy growth rate was too low in the pre-embargo years, but the oil embargoes of the 1970s led to a reduction in actual growth rate.
Source: The figure is reproduced from Landsberg (1985).

Even with the wisdom of hindsight it's hard to see how the RFF team might have been more accurate. Annual population growth had been roughly 1.8% in the 15 years prior to their work, but slowed thereafter. The labor force grew far faster than anticipated, driven by the decisions of women to enter the work force and an offsetting decline in worker productivity. Housing construction was underestimated due to failure to anticipate the rise in single-person households. Forecast and actual industrial sector growth rates were reasonably similar, masking large errors in individual components. For example, rubber and plastics were expected to grow by 200% over the two-decade period, but actually grew by 400%. Commercial energy consumption was underestimated due to a postulated correlation with residential electricity.

Landsberg's review leads to little hope for improving the accuracy of forecasting. However, as we emphasize throughout this article, accuracy is not a particularly compelling reason to undertake long-range energy forecasts. Rather, long-range forecasts achieve success when they illuminate the consequences of policy choices, and thereby help inform decisions in the present, a point that also emerges from Joskow (2003) and Huntington et al. (1982). Thus, we find ourselves in strong resonance with Hans Landsberg's

concluding remark: "If more of us once in a while took time out to review our past output, it might produce a healthy sense of self-discipline in our profession."

## III. SCHURR'S PERSPECTIVE

Sam Schurr and his colleagues were pragmatists. They sought to forecast on a time frame matched to the policy issues with which they were concerned. In their pioneering work *Energy in the American Economy: 1850-1975* (Schurr et al. 1960, p.3) they wrote

> In selecting a target year… the authors were strongly influenced by the fact that the uncertainties involved… argue in favor of not looking too far ahead. … Yet, in planning this study, we felt it necessary to look ahead far enough that there would be ample time to adopt those policy measures, in industry and government, which might be necessary to cope with supply problems that the future could bring … The selection of 1975 – fifteen years beyond the publication of this book … reflects our judgment that impending problems in energy supply, if any were to be disclosed, could be readily met through actions taken in the intervening fifteen years.

Their focus was near-term policy, but their work reflects a clear concern with long-term forecasting as a tool for public and private sector decision-making:

> This is not a forecast that the United States will satisfy its energy demands domestically – it is not even doing so at present – nor is it a recommendation that the country should choose to do so. Neither is it a forecast that energy prices (in constant dollar terms) will not rise. Policy decisions in industry and government, lying beyond the scope of this analysis, will be important in determining the actual course of events in these matters. We do not attempt to predict what those decisions will be, or to judge what they should be. Rather, our purpose is to provide basic information about energy consumption and availability which will be helpful in public debate and in guiding those responsible for reaching policy decisions (Schurr et al. 1960, p.4).

Schurr and his co-authors recognized that mechanical forecasting rarely illuminates. Judgment is required. A footnote (Schurr et al. 1979, footnote 12, p.193): makes this point explicit: "In the sources used, high or low projections were, in some cases, not merely a function of price

assumptions but of policy, technological and behavioral factors as well. Thus, no quantitatively tidy specification of 'driving forces' is possible."

Judgment is always required if forecasts are to provide insight. It is judgment that distinguishes brilliance from mediocrity, and Sam Schurr exhibited judgment in abundance.

## IV. WHAT DO WE MEAN WHEN WE USE THE TERM 'RETROSPECTIVES'?

We define the term "retrospective" to mean any careful analysis of a forecast that is conducted after the period covered by the forecast has become history.[3] Retrospective examination allows for a comparison between historical events and the predictions of the modelers, and gives model users and model builders insights as to the strengths and weaknesses of particular models. It also hones the instincts of model users regarding the forces that affect the future, which is one of the key purposes of long-run forecasting models.

Useful retrospectives must do more than simply compare total forecasted and actual energy use. Detailed comparisons by sector, fuel, and end-use (if possible) are essential for truly understanding how and why a forecast differs from actual events. Thus retrospectives can include econometric investigations of those events, as well as comparisons between the predicted and observed driving forces underlying a model. Where key assumptions about population growth, economic activity, weather, industrial production, and the like differ from actual events, the model can be rerun using the historical data, in order to determine how much of the error in the original forecast resulted from structural problems in the model itself and how much from incorrect specification of the fundamental drivers of the forecast.

Retrospectives may differ in character and purpose depending on who creates them. Greenberger and Richels (1979) distinguish among model developers, model users, third party model analysts, and joint efforts between developers, users, and third party analysts. Model developers may use retrospectives to help get the input data right, while model users and other analysts may find investigations of model algorithms affecting the results from policy analysis more directly relevant to their work. In this article, we focus mainly on the perspective of model users, both because they actually use the models to achieve programmatic goals and because they exercise

3. We distinguish here also between retrospectives and prospective model comparisons. The latter can be useful (see for example Huntington et al. 1982) but offer insights of a different character than those available from the retrospective analyses we advocate here.

disproportionate influence over how models are designed and used (they pay the bills).

## V. WHAT OTHER RETROSPECTIVES HAVE BEEN DONE?

When we searched for other retrospectives, we found a patchwork of such analyses. Many retrospectives were out of date, and few of them were as thorough as Landsberg's. We don't undertake the job of summarizing and evaluating each of these efforts, although that would be a worthy topic for future work. Instead, our goal here is to list the key references so that interested readers can investigate as they see fit.

One of the most comprehensive and impressive retrospectives is the 1978 book by Ascher (1978), which reviews dozens of forecasts from a variety of fields, including energy. Other retrospectives (not exclusively from the energy forecasting field) include those by Huss (1985a, 1985b, 1985c), Nelson and Peck (1985), Dominguez et al. (1988), Nelson et al. (1989), Schnaars (1989), Huntington (1994), Shlyakhter et al. (1994), US DOE (1999), DeCanio (2003), Laitner et al. (2003), and Sanstad et al. (2003).

Special attention to retrospectives is warranted for forecasts that are conducted on an ongoing basis, so here we make an exception to our "no evaluation" rule and delve into a bit more detail about the limited but ongoing retrospectives conducted by the Energy Information Administration for the *Annual Energy Outlook* forecasts. The most recent comprehensive retrospective conducted by EIA is contained in US DOE (1999), and can be found on the web at http://www.eia.doe.gov/oiaf/issues/forecast.html. This study compares consumption, production, and price results by fuel from the *Annual Energy Outlook* (AEO) forecasts (AEO 1982 to AEO 1999) to actual historical values, and explores some of the reasons for the differences between history and the forecasts. Updated tables containing the latest such comparisons are described briefly at http://www.eia.doe.gov/oiaf/analysispaper/forecast_eval.html, but this latest report contains much less discussion of differences between history and the forecasts than does the 1999 version.

These retrospectives, while useful as far as they go, could and should go beyond simply reporting percentage errors compared to actual values by fuel and should discuss issues related to model structure and assumptions in detail. Even a simple correction to account for weather fluctuations would yield important insights, and including sector and end-use detail, while difficult, would further help to disentangle predicted and actual values.

The model used to generate the AEO (the National Energy Modeling System, or NEMS) is a rich source of insight into the driving forces underlying the forecasts. EIA should take advantage of the detailed structure

of NEMS to undertake more elaborate retrospectives, both for its own institutional benefit as well as that of outside analysts seeking to more fully understand the U.S. energy system.

## VI. MAKING MODELS "BETTER"

Retrospectives are an important method for making models and forecasts "better." By "better," we do *not* mean that long-term energy forecasts will become more accurate, because that goal is beyond our reach (Craig et al. 2002). Instead, we believe that retrospectives can help make models more useful, and we elaborate below on how they can do so.

To understand why long-range energy forecasts cannot be expected to predict the future accurately, we draw on the classification system of Hodges and Dewar (1992). Their system describes a fundamental but not well known distinction between "validatable" and "nonvalidatable" models. In their terminology, validatable models have the potential to yield predictions of the future in which one can have high confidence. While nonvalidatable models can have many useful features, they are likely to yield low accuracy and low precision. They will also embody unquantifiable errors.

Validatable models apply only to situations that

1. are observable;

2. exhibit constancy of structure in time;

3. are not affected significantly by exogenous factors that the model assumes, either explicitly or implicitly, do not matter; and

4. permit collection of ample and accurate data.

Validatable models can potentially forecast precisely and confidently. Astronomical and satellite orbital predictions are a clear example. Satellite orbits can be calculated with enormous precision because the laws of orbital mechanics are characterized by the four properties listed above.

Unfortunately, the situations modeled by long-range energy forecasting tools do not meet criteria 2, 3, and 4 in the list above, and sometimes fail to meet criterion 1 as well.[4] Consequently, long-range energy

4. For example, human motivations and expectations are sometimes explicitly included in models. These cannot be directly observed but must be inferred.

forecasting models are not validatable in Hodges and Dewar's sense, and reproducible accuracy can never be expected from such models (except by chance).

Why do energy forecasting models fail to meet these criteria? Long-run forecasting models generally assume that underlying structural relationships in the economy vary in a gradual fashion. The real world, in contrast, is rife with discontinuities and disruptive events, and the longer the time frame of the forecast, the more likely it is that pivotal events will change the underlying economic and behavioral relationships that all models attempt to capture.

Energy forecasting necessarily makes assumptions about human behavior (including social, institutional, and personal) and human innovation. Institutional behavior evolves, individual behavior changes, and pivotal events occur, affecting outcomes in ways we cannot envision. Models cannot anticipate the long-term evolution of the real world, not just because their data and underlying algorithms are inevitably flawed, but because the world sometimes changes in unpredictable and unforeseeable ways. Further, data are always limited and incomplete. Important characteristics of the energy/economy system may not be measured, or existing measurements may not be published, making accurate characterization of the system difficult or impossible.

Despite their inability to forecast accurately, nonvalidatable models can be made more useful and successful (and thus better). We list here the most important ways models can achieve success (adapted from Craig et al. 2002):

1. as bookkeeping devices;

2. as aids in selling ideas or achieving political ends;

3. as training aids;

4. as aids to communication and education;

5. to understand the bounds or limits on the range of possible outcomes; and

6. as aids to thinking and hypothesizing.

This broad range of uses for models reflects the many different situations in which they find application. In each case, retrospectives can help improve the model – or at least the way the model is applied to the task at hand.

## VII. WHAT ARE RETROSPECTIVES GOOD FOR?

### They Help Make Models More Useful

We listed above the various ways that models can be useful. Retrospectives will usually enhance each of these.[5] In each case, retrospective analysis can help illuminate the key assumptions that lead to the results. The value of such models is not that they can accurately forecast the future (because they can't). Rather, these models can help hone the users' intuition about the key forces affecting the future. For this purpose, having retrospectives is invaluable.

### They Reinforce Modesty

We need to be humble in the face of our modest abilities to foresee the future (Craig et al. 2002). This caution is especially warranted when assessing effects of technological choices on the environment, but it applies equally well to most energy forecasts. Fundamental limitations on our ability to foresee consequences have important implications for the ways we use forecasts in our planning, and historical retrospectives can help us avoid the most common pitfalls in the future.

### They Teach You About Biases and Embedded Assumptions

Computer models depend on data and assumptions that are based in large part on analysts' judgment. Most models contain dozens or hundreds of such assumptions. This heavy reliance on assumptions for unknown parameters is most prevalent in models used in business and public policy, but even models of the physical world sometimes rely on poorly understood assumptions.

Almost all models of the physical world are validatable and the erroneous assumptions are uncovered and corrected sooner or later. However, there are good reasons to conduct retrospectives, even for energy forecasting models (which are not validatable). Historical retrospectives can lay bare key assumptions to scrutiny and evaluation, and can stimulate debate and discussion in the modeling community about the validity of these assumptions.

5. It is true that retrospectives can sometimes be anethema to those trying to sell particular ideas if those ideas are not supported by the evidence, but we do not consider such cases here.

## They Help Uncover and Explain Uncertainties

Uncertainty in models can arise in three main areas: experiment, theory, and computation. Computational uncertainty, which results from roundoff and discretization errors and the like, is not a major source of uncertainty for forecasting models, so we do not treat it here. Experimental uncertainty can come from random fluctuations, over time, in measured quantities. For example, the unsteady variations in fluid velocity due to turbulence cause large deviations about the average velocity (Hinze, 1975). In addition, experimental results always contain uncertainty due to random and systematic measurement errors (Lyons, 1998, p. 4). Finally, experimental uncertainty includes the effects of non-observable or uncontrolled quantities in the system of interest. In economic forecasting, consumer expectations and worker productivity exemplify this form of uncertainty. In cases where information exists about the probability distribution of errors, uncertainty analysis offers a way to incorporate it in forecasting models.

Long-range energy forecasts have proven particularly vulnerable to systematic errors. The understanding of systematic error is one of the most interesting results of retrospective analysis (Craig et al. 2002). Systematic errors appear ubiquitous. Particularly interesting examples occur in the measurement of fundamental constants of nature. We discuss one such example, the lessons from which have broad applicability to measurement of economic parameters as well.

The reported value of Planck's constant, *h*, illustrates how systematic errors can overwhelm experimental uncertainty. Figure 2 shows the evolving estimate, over time, of both the value and the one-standard-deviation uncertainty bounds for this fundamental physical constant. By 1947, its value was pretty well known, with an estimated uncertainty of about 500 parts per million. Over the next several years, as measurements improved, the accuracy – as evidenced by the decreasing width of the error bands – was believed to have improved as well. However, by 1952 new high-quality measurements had yielded estimates several standard deviations above the previously accepted value. In due course, the new value was accepted, and the standard deviation again decreased as measurement methods improved. However, by 1963 the accepted value had jumped outside the error bounds again. In each case, wholly unsuspected errors had compromised the earlier measurements. This kind of situation has occurred many times in the history of physics, right up to the present.[6] Given the difficulties of accurately

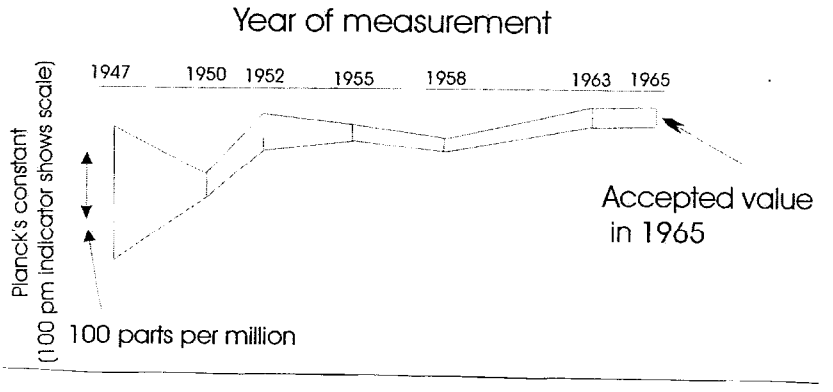
6. The most interesting current example is Newton's gravitational constant, where the most recent credible data differs from the accepted value by about 40 standard deviations (Mohr and Taylor, 2000).

measuring parameters in systems that meet the criteria for validatable models, it is important to view with skepticism any attempts to supply "confidence intervals" for the results of non-validatable models like those used in long-run energy forecasting.

Beyond experimental uncertainty, theoretical uncertainty can arise because the governing equations are not known, in which case engineering correlations may be used. Alternately, the governing equations may be simplified in order to reduce the computational burden of solving them.[7] In such cases, the model purposely does not represent the full complexity of the physical world. Thus it applies only to a particular problem domain (Roache, 1998).

The danger of using a model outside of a defensible domain (model mis-specification) becomes especially acute with nonvalidatable models. If simplifications related to structure are embedded in such a model, models based on historical trends may have limited predictive capability.

**Figure 2. Estimates of Plank's Constant "*h*" Over Time**




Notes: In this "validatable" system, researchers repeatedly underestimated the error in their determinations. At each stage uncertainties existed of which the researchers were unaware. The problem of error estimation is far greater in long-range forecasting.
Source: Simplified from Cohen and DuMond (1965).

7. For example, Newtonian mechanics represent a mathematical reduction of the complete laws of motion. Similarly, assuming incompressible flow simplifies the Navier-Stokes equations for fluid dynamics (Ferziger and Peric, 2002, p. 12).

## VIII. HOW SHOULD RETROSPECTIVES BE DONE?

This section gives our interpretation of the guidance Landsberg and Schurr might have given to analysts creating future retrospectives, if these two visionaries were alive today. More extensive guidance of this nature is contained in Craig et al. (2002) and Koomey (2001).

### Disentangle Input Data Issues from Modeling Issues

It is important to distinguish errors in input data from errors in model specification. Assumptions (both for data and for algorithms) should be made explicit. Ideally, the model's documentation identifies all assumptions, but more often than not they have to be exposed through examination of the model's code or direct discussions with the modeler.

Next, it is important to identify where data and modeling assumptions caused the forecast to diverge sharply from actual events. Examining such differences can illuminate the effects of key drivers like population, GDP, and prices, as well as those of the underlying input assumptions (and structural aspects of the model). Quantifying the reasons for differences between forecasted and actual results is the ultimate goal, and such quantification is almost always possible to some degree. Usually some part of the variation can't be explained, but for a model to be compelling most of the variation should be explainable.

For example, Landsberg's retrospective explicitly showed how compensating errors led to a fortuitously accurate estimate of total energy demand in 1980. He quantified the effects of various assumptions and algorithms in the model and was able to determine how unanticipated compensating changes in driving forces led to the 1980 result. This is one of the reasons why Landsberg's retrospective was exemplary.

Landsberg's review also highlights the importance of maintaining long-run consistent statistical series for doing historical retrospectives. When data collection methods change, historical data either become useless or vastly diminished in value. The recent shift from the old U.S. Standard Industrial Classification System (SIC) industrial categories to the newer North American Industrial Classification System (NAICS) system is one example where years of historical data became much less useful because the SIC categories do not map precisely onto the NAICS system.[8]

8. For details on SIC and NAICS see http://www.census.gov/epcd/www/naics.html

## Use Historical Decomposition Techniques

There is an extensive literature on retrospective decomposition of historical trends in energy use, energy efficiency, weather, and emissions (Adams et al. 1985; Davis et al. 2002; Schipper et al. 2001; US DOE, 1995). Techniques used in such studies (especially divisia decomposition and related methods) are directly relevant to conducting retrospective evaluations of forecasts, and they can contribute greatly to understanding the fundamental driving forces affecting the actual and forecasted evolution of the energy system (and should be applied for that purpose). The best of these studies do not simply compare total forecasted energy use with historical data, but instead explore sectoral and end-use detail to truly unravel underlying trends.

## Document Everything

The importance of clear and complete documentation to successful retrospectives (and to credible forecasting efforts in general) cannot be overestimated. All assumptions should be recorded in a form that can be evaluated, reproduced, and used by others. Functional forms should be stated explicitly. Unless assumptions are explicit, others can't evaluate their reasonableness, and creating useful retrospectives is all but impossible. Of course, there are other good reasons for creating detailed documentation, and those justify documentation efforts all by themselves.

Because of the ease with which they can be documented, simpler and more transparent models are often superior in usefulness and equivalent in accuracy to large and complex ones. Simpler models are more amenable to peer review of underlying data and assumptions. Simpler models are easier to communicate to model users, and often allow one to make a more compelling case. A model that the audience can actually grasp is inherently more persuasive than a "black box" that no one outside of a small circle of analysts understands. Transparent models for which the input data and assumptions are also well documented are even more compelling, but are all too rare.

## Identify and Assess Discontinuities

One of the biggest unsolved issues in forecasting relates to the treatment of discontinuities or tipping points. In the analysis of climate change, for example, many climate models assume linear responses to perturbations in greenhouse gas concentrations. Unfortunately, there is an unknown but nonzero probability that the climate system may respond in a discontinuous manner to rapid changes in greenhouse gas concentrations. For example, there may be thresholds beyond which the climate "snaps" to a new equilibrium level that is far from the current one, which could include

substantially different ocean circulation and temperature patterns (Broecker, 1997; Taylor, 1999). This rapid shift could lead to a multitude of effects, such as flooding or changes in agricultural patterns (IPCC 2001, p. 93, Ch. 19). Non-linearities and instabilities occur frequently in ecology and are of concern in virtually all ecological forecasting (Clark et al. 2001).

Discontinuities can also affect economic systems. The oil crises in the 1970s radically changed the way the future unfolded in the 1980s, affecting human and institutional behavior, technology choices, and related political developments. Time will tell if the U.S. response to the terrorist attacks in the Fall of 2001 will also represent such a discontinuity. Discontinuities are inherently difficult or impossible to predict, but they remain important to consider in whatever imperfect ways scenario developers can do so.

## IX. CONCLUSIONS

This article celebrates the work of Hans Landsberg and Sam Schurr. Today's forecasters would do well to follow their examples, both to understand the limitations of current work and to improve it. By creating detailed historical retrospectives they can best accomplish both those goals.

One distinguishing characteristic of scientific activity is the repeated testing of models and theories against experimental observations (McCloskey, 2000). Such testing either increases the level of confidence in the validated models or spurs attempts to improve those models and theories that do not measure up to experimental facts. Thus, making predictions from models and moving on to making more predictions without any retrospective examination of past failures or successes (and reasons for these) is inconsistent with the scientific method.

This criticism holds even though the models used for energy forecasting are not validatable in the sense discussed above. As a community, energy analysts cannot expect their "art" to get better over time unless lessons are learned from past successes and failures. Vigorous retrospective examinations are a necessity for any endeavor to be considered truly scientific, and long-term forecasting is no exception. Sam Schurr and Hans Landsberg's work was largely directed to the policy community, that is, to model users. Following their lead, we believe model users can best make use of retrospective analysis as a technique for providing insight and understanding, and as a vehicle for assessing the merits of forward-looking analysis. Allocation of additional resources to this area will, we believe, pay off in more and greater insights gleaned from each modeling dollar.

## ACKNOWLEDGEMENTS

In searching for published examples of retrospective forecasts, we requested input from many colleagues in the forecasting field, including Jae Edmonds (Pacific Northwest National Laboratory, PNNL), Andrew Nicholls (PNNL), Skip Laitner (Environmental Protection Agency), Stephen DeCanio (UC Santa Barbara), Dan Kammen (UC Berkeley), Joe Eto (LBNL), John Cymbalsky (Energy Information Administration, EIA), Erin Boedecker (EIA), Steve Wade (EIA), Hill Huntington (Stanford University), John Weyant (Stanford University), Nebojsa Nakicenovic (IIASA), Stephen Peck (Electric Power Research Institute), and Kenneth Train (UC Berkeley). Barry Taylor at NIST contributed his insights and ideas on our discussion of errors in the fundamental constants in nature. The paper benefited substantially from the excellent comments of several reviewers, including Hill Huntington, Geoff Pearce, and Campbell Watkins. We also benefited from the thoughtful guidance of the editor of this special issue, Joel Darmstader, and we are grateful for his contributions. Errors or omissions are of course the responsibility of the authors alone.

This work was supported in part by the Office of Atmospheric Programs of the U.S. Environmental Protection Agency. Prepared in part for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

## REFERENCES

Adams, R. C., D.B. Belzer, J.M. Fang, K.L. Imhoff, D.H. Lax, R.J. Moe, J.M. Roop, and A.R. Wusterbarth (1985). *A Retrospective Analysis of Energy Use and Conservation Trends: 1972-1982*. Richland, WA: Pacific Northwest National Laboratory. PNL-5026. June.

Armstrong, J. Scott (1978). *Long Range Forecasting: From Crystal Ball to Computer*. Chichester: Wiley-Interscience.

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

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

Bohi, Douglas R. (1989). *Energy Price Shocks and Macroeconomic Performance*. Washington, DC: Resources for the Future.

Broecker, Wallace S. (1997). "Thermohaline circulation, the Achilles heel of our climate system: Will man-made $CO_2$ upset the current balance?" *Science* 278: 1592-1588.

Clark, James S., S.R. Carpenter, M. Barber, S. Collins, A. Dobson, J.A. Foley, D.M. Lodge, M. Pascual, R. Pielke Jr, W. Pizer, C. Pringle, W.V. Reid, K.A. Rose, O. Sala, W.H. Schlesinger, D.H. Wall, and D. Wear (2001). "Ecological Forecasts: An Emerging Imperative." *Science* 293(657): 657-660.

Cohen, E. Richard, and Jesse M. DuMond (1965). "Our Knowledge of the Fundamental Constants of Physics and Chemistry in 1965." *Review of Modern Physics* 37(4): 537-594.

Craig, Paul, Ashok Gadgil, and Jonathan Koomey (2002). "What Can History Teach Us?: A Retrospective Analysis of Long-term Energy Forecasts for the U.S." In

*Annual Review of Energy and the Environment 2002*. Edited by R. H. Socolow, D. Anderson and J. Harte. Palo Alto, CA: Annual Reviews, Inc. (also LBNL-50498).

Davis, W. Bart, Alan H. Sanstad, and Jonathan G. Koomey (2002). "Contributions of Weather and Fuel Mix to Recent Declines in U.S. Energy and Carbon Intensity." *Energy Economics* (also LBNL-42054) 25(4): 375-396.

DeCanio, Stephen J. (2003). *Economic Models of Climate Change: A Critique*. Basingstoke, UK: Palgrave-Macmillan (forthcoming).

Dominguez, Kathryn M., Ray C. Fair, and Matthew D. Shapiro (1988). "Forecasting the Depression: Harvard Versus Yale." *American Economic Review* 78(4): 595-612.

Farvar, M. Taghi, and John P. Milton, ed. (1972). *The careless technology: ecology and international development* (Conference on the Ecological Aspects of International Development ). Garden City, N.Y.: Natural History Press.

Ferziger, Joel H., and Milovan Peric (2002). *Computational Methods for Fluid Dynamics*. Berlin, Germany: Springer-Verlag.

Greenberger, Martin, Matthew A. Crenson, and Brian L. Crissey (1976). *Models in the Policy Process : Public Decision Making in the Computer Era*. New York: Russell Sage Foundation: distributed by Basic Books.

Greenberger, Martin, and Richard Richels (1979). "Assessing Energy Policy Models: Current State and Future Directions." *Annual Review of Energy* 4: 467-500.

Hinze (1975). *Turbulence*. New York, NY: McGraw Hill.

Hodges, James S., and James A. Dewar (1992). *Is it you or your model talking? A framework for model validation*. Santa Monica, CA: RAND. ISBN 0-8330-1223-1.

Huntington, Hillard G. (1994). "Oil Price Forecasting in the 1980s: What Went Wrong?" *The Energy Journal* 15(2): 1-22.

Huntington, Hillard G., John P. Weyant, and James L. Sweeney (1982). "Modeling for Insights, Not Numbers: The Experiences of the Energy Modeling Forum." *Omega: The International Journal of Management Science* 10(5): 449-462.

Huss, William R. (1985a). "Can Electric Utilities Improve Their Forecast Accuracy? The Historical Perspective." In *Public Utilities Fortnightly*. Dec.26, 1985: 3-8.

Huss, William R. (1985b). "Comparative Analysis of Company Forecasts and Advanced Time Series Techniques Using Annual Electric Utility Energy Sales Data." *International Journal of Forecasting* 1: 217-239.

Huss, William R. (1985c). "What Makes a Good Load Forecast?" In *Public Utilities Fortnightly*. Nov. 28, 1985: 3-11.

IPCC (2001). *Climate Change 2001--Impacts, Adaptation, and Vulnerability*. Contribution of Working Group II to the Third Assessment Report to the Intergovernmental Panel on Climate Change. New York, NY: Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, UK.

Joskow. Paul L. (2003). "Energy Policies and Their Consequences After 25 Years." *The Energy Journal* 24(4): 17-49.

Koomey, Jonathan (2001). *Turning Numbers into Knowledge: Mastering the Art of Problem Solving*. Oakland, CA: Analytics Press. Available from World Wide Web: (http://www.numbersintoknowledge.com)

Laitner, John A. "Skip", Stephen J. DeCanio, Jonathan G. Koomey, and Alan H. Sanstad (2003). "Room for Improvement: Increasing the Value of Energy Modeling for Policy Analysis." *Utilities Policy* (also LBNL-50627) 11: 87-94.

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

Landsberg, Hans H., Leonard L. Fischman, and Joseph L. Fisher (1963). *Resources in America's Future: Patterns of Requirements and Availabilities 1960-2000*. Baltimore: Johns Hopkins Press for Resources for the Future.

Lyons, Louis (1998). *A Practical Guide to Data Analysis for Physical Science Students*.Cambridge, UK: Cambridge University Press.

McCloskey, Deirdre N. (2000). "How to be Scientific in Economics." *Eastern Economic Journal* 26(2): 241-246.

Mohr, P. J., and B. N. Taylor (2000). "CODATA Recommended Values of the Fundamental Constants: 1998." *Review of Modern Physics* 72(2): 351-496.

Nelson, Charles R., and Stephen C. Peck (1985). "The NERC Fan, A Retrospective Analysis of the NERC Summary Forecasts." *Journal of Business and Economic Statistics*. 3(3): 179-187.

Nelson, Charles R., Stephen C. Peck, and Robert G. Uhler (1989). "The NERC Fan in Retrospect and Lessons for the Future." *The Energy Journal* 10(2): 91-107

Roache, Patrick J. (1998). *Verification and Validation in Computational Science and Engineering*. Albuquerque, NM: Hermosa Publishers.

Sanstad, Alan H., Jonathan G. Koomey, and Skip Laitner (2003). "Back to the Future: A Note on Long-Range Energy Price and Quantity Projections in Retrospect." Submitted to *Contemporary Economic Policy* (also LBNL-48107).

Schipper, Lee J., Scott Murtishaw, and Fritjof Unander (2001). "International Comparisons of Sectoral Carbon Dioxide Emissions Using A Cross-Country Decomposition Technique." *The Energy Journal* (also LBNL-48783) 22(2): 35-75.

Schnaars, Steven (1989). *Megamistakes: Forecasting and the Myth of Rapid Technological Change*. New York, NY: Free Press.

Schurr, Sam H, Joel Darmstadter, Harry Perry, William Ramsay, and Milton Russell (1979). *Energy in America's Future: The Choices Before Us*. Baltimore, MD: Johns Hopkins University Press for Resources for the Future.

Schurr, Sam H., Bruce C. Netschert, Vera F. Eliasberg, Joseph Lerner, and Hans H. Landsberg (1960). *Energy in the American Economy 1850-1975*. Baltimore, MD: Johns Hopkins Press for Resources for the Future.

Schurr, Sam H., Bruce C. Netschert, Vera F. Eliasberg, Joseph Lerner, and Hans H. Landsberg (1977). *Energy in the American Economy 1850-1975*. Westport, CT: Greenwood Press.

Shlyakhter, Alexander I., Daniel M. Kammen, Claire L. Broido, and Richard Wilson (1994). "Quantifying the credibility of energy projections from trends in past data: The US Energy Sector." *Energy Policy* 22(2): 119-130.

Taylor, Kendrick (1999). "Rapid Climate Change." *American Scientist* 87(4): 320-28.

US DOE (1995). *Energy Efficiency in the US Economy (Report 3)--Energy Conservation Trends: Understanding the Factors Affecting Energy Conservation Gains and Their Implications for Policy Development*. Washington, DC: US Department of Energy, Office of Policy. DOE/PO-0034. April.

US DOE (1999). *Issues in Midterm Analysis and Forecasting 1999*. Washington, DC: Energy Information Administration, U.S. Department of Energy. DOE/EIA-0607(99). August.