Model Evaluation for Policy Insights: Reflections on the Forum Process

Huntington, Hillard G.

Stanford University

1 June 2021

Model Evaluation for Policy Insights: Reflections on the Forum Process

Hillard G. Huntington

Energy Modeling Forum, Stanford University
hillh@stanford.edu

EMF OP74

June 2021

Forthcoming in *Energy Policy* 156, September 2021, 112365.

Energy Modeling Forum

Stanford University

Huang Engineering Center

475 Via Ortega

Stanford, CA 94305-4121

Model Evaluation for Policy Insights: Reflections on the Forum Process

Hillard G Huntington
Stanford University

June 2021 (Revised)

Abstract

Model evaluation is best considered as a process for communicating with the policymaking or policy-advising community. Six decades of energy modelling have witnessed increasing complexity in these systems, a situation that raises a number of important challenges in using them effectively in policymaking organizations. When used as a learning rather than forecasting tool, these systems can be evaluated individually one by one or through joint efforts to compare them in multi-model exercises. After summarizing the evolution of energy modelling and efforts to evaluate them since the first oil embargo, this essay provides a guide to future evaluation collaborations by highlighting a few challenges that would improve the value of these studies for the policymaking community. These challenges range broadly and cover topics such as enhancing the engagement of the model user, ventilating the models' complexity with intuitive insights, using simple models to demonstrate key parameters or responses, applying judicious occasional meta-analysis when there is value added, reporting model responses and calibrating them for decisionmakers, considering retrospective evaluation for a past period (when possible), selecting standardized or modeler-choice baseline conditions, selectively developing policy or diagnostic alternative cases, and institutionalising the model evaluation process for a specific topic or region.

JEL Classification Codes: C02; Q41; Q48; Q54

Keywords: Energy Policymaking; Model Evaluation; Model Comparison

Acknowledgments: The author would like to acknowledge a number of useful discussions over the years with his colleague, John Weyant, and past participants in Energy Modeling Forum studies. He also thanks two referees for their helpful comments. The views expressed in this paper are his alone.

Model Evaluation for Policy Insights: Reflections on the Forum Process

Hillard G Huntington
Stanford University

1. Introduction

Energy modelling and policymaking have been closely linked for most of the last five decades, beginning in the 1970s (Greenberger, Brewer, Hogan, *et.al.*, 1983). These models essentially track how the energy system responds to shifts in energy supply or demand conditions brought about changed economic, environmental, technical or policy conditions. Policy developers use these systems to aid in their choices between alternative strategies for meeting certain policy goals such as adapting to increased electrification or mitigating damages from greenhouse gas emissions. In discussing the early days of developing and using the Project Independence Evaluation System within the United States Federal Energy Administration, Hogan (2002) describes policy-relevant energy modelling as a process for learning and communicating about energy systems rather than as a final product. In a similar manner, it is more fruitful to consider evaluations of models for developing policies as a process rather than a product. Evaluations will be more valuable for those engaged in developing policy if they focus on how various models behave when applied to a particular problem than if they focus more narrowly on the detailed model traits and structure alone.

This essay expands upon the theme of model evaluation as a process for communicating with the policymaking or policy-advising community. Although model evaluation has been a very old interest, this process in now being applied to energy models that have undergone dramatic changes over the last five decades. A long-term perspective on these developments can help immeasurably in efforts to improve model evaluation. In order to know how to proceed with future model evaluation efforts, it is essential to understand the evolution of energy modelling and efforts to evaluate them since the first oil embargo. Although this history described in the next two sections is well known to experienced energy modellers, it is important background information for anyone within the energy policymaking community who wants to understand the prospects for evaluating models in the current environment. Based upon this experience, this essay provides a guide to future evaluation collaborations by highlighting a few challenges that would improve the value of these studies for the policymaking community. By design, the included topics are selective rather than comprehensive and have been chosen based upon how they aid policymaking discussions based upon informal comments from past participants rather than on any formal polling process. They are discussed to provide examples of what might be appropriate approaches for expanding the use of models in policymaking organizations.

The discussion assumes that the policymaking organizations want to use models to help them understand energy market conditions and environmental tradeoff options. The next section begins with a brief description of how energy models have evolved over time, emphasizing their increasing complexity. After section 3 considers how models can be used and misused, section 4 contrasts the approach for verifying a single model with a process for comparing multiple models. The latter has become increasingly popular in recent times. When this form of multiple model comparisons has a significant user orientation, Greenberger and Richels (1979) label this effort as the forum approach. It views model comparison not only as a vehicle for understanding differences between models but also as a means for expanding knowledge for the policymaking community. With this background appropriately articulated, the discussion in section 5 focuses upon nine critical challenges that future studies are likely to encounter based upon the experiences of the Energy Modeling Forum at Stanford University. A final section concludes with a brief overview of the principal points for policymakers that were made in the previous sections.

2. Six Decades of Energy Modelling

Energy models have evolved over the last six decades with the rise of new issues and with technological advancements that have transformed the modelling process. Early research from the 1970s focused upon the dependence of energy systems on uncertain oil supplies. Later efforts beginning the 1990s shifted attention to developing strategies for energy transitions that would reduce damages from greenhouse gas emissions. This new policy focus combined with enormous expansion in computer power transformed relatively simple frameworks into much more complex systems with much richer detail. This transformation towards more complex modelling systems create very new challenges for efforts engaged in communicating model results and their differences to the model-using community. Section 4 discusses these challenges.

Several pioneering modelling efforts in these early years stand out. Nordhaus (1973) developed a framework to identify how society can use scarce resources optimally over the long run. He discussed the high oil prices of the 1970s and their relationship to this long-run path. An important contribution was the concept of an uncertain but abundant backstop energy technology that would be available in some future decade to satisfy energy demands at a sustained higher price. Manne (1976) developed the Energy Technology Assessment (ETA) framework to delve into the opportunities to replace U.S. oil and natural gas dependence with a more diversified mix of energy sources. These academic research projects were complemented by a massive U.S. government effort to develop the Project Independence Evaluation System (PIES) model. Although originally designed to explore how the U.S economy could become totally independent of foreign oil supplies, the Federal Energy Administration staff quickly learned that this goal was unrealistic (Hogan, 2002). Nevertheless, this effort became the core of this agency's analytical efforts. It codified data that was previously organized in printed tables and structured the policy discussion about how to diversify away from uncertain foreign oil supplies.

Many early models quickly experienced the curse of dimensionality: the lack of computer power that limited a single model's ability to represent spatial, temporal, technical and interfuel dimensions in a realistic manner. Often, separate models were developed to analyze issues involving coal, natural gas, petroleum and electric power markets, as is evidenced by many of the early model-evaluation studies performed by the Massachusetts Institute of Technology (1978) or by the Energy Modeling Forum (Huntington et al., 1982). These models were often partial equilibrium, representing a single key energy market like crude oil or electric power and excluding all other energy or nonenergy markets.

Later modelling emphasized strategies for abating climate change damages as this issue gained prominence in international policy circles. Edmonds and Reilly (1983) provided an earlier modelling effort to evaluate carbon dioxide emissions within the broader energy transition trends. Nordhaus (1993a; 1993b) and Manne and Richels (1991) shifted their modelling efforts to the new policy arena, building upon their past experiences in evaluating energy market transitions during previous decades. With the rise of the climate change issue, policy interest shifted from evaluating key single energy markets like oil and power to more integrated systems balancing all fuels. Over time, this transformation also resulted in more complex models for representing the whole energy system.

The shift to more complicated models was also reinforced by expanded computing power and the electronic availability of new data sources. General modelling algorithms like GAMS gained popularity by both engineers and economists. Computer costs declined very sharply relative to the labor costs for monitoring electronic computations. Trends in million instructions per second (MIPS) measure a computer's processor speed and neatly capture the dramatic rise in computer power. Since the 1940s, $ per MIPS declined by a factor of ten roughly every five years (AI Impacts, n.d.; Nordhaus, 2007). This cost decline represents a 36.9% reduction on a compounded average rate. Meanwhile, median weekly earnings for full-time operations research analysts increased by 1.01% per year in real terms (deflated by urban consumer price index) between 2000-2008 before declining slightly after 2008.[1]

The shift towards bigger datasets and substantially larger models improves the ability to represent some important details that could improve policy decisions. The downside of this development is that it becomes much more difficult to compare and evaluate models in a way that communicates directly with the policy developer. When the model incorporates complex interactions between important variables, it is no longer an easy task to “tell the story” in a compelling manner. It becomes much more challenging to explain the precise reasons for why different models differ from each other. This issue is addressed more fully in later sections.

Two quite different approaches currently exist for evaluating combined energy markets: (1) a top-down approach focusing upon the major economic drivers and relationships between the energy sector and other goods and services, and (2) a bottom-up approach describing the resources and technologies for producing and consuming energy in various applications.

The top-down method relates energy market conditions directly with aggregate macroeconomic conditions such as real GDP or sectorial activities or with a general equilibrium perspective that seeks prices at which supply equals demand in every market goods, factors, and foreign markets (Wing, 2009). By integrating the energy system and the rest of the economy, these frameworks are useful for estimating the welfare or GDP losses associated with competing energy or environmental policy strategies. These approaches remain popular when the analyst wants to represent basic economic tradeoffs, although some important limitations include the difficulty of representing key engineering processes with economic formulae and the lack of empirical support for some of the key responses.

[1] US Bureau of Labor Statistics is the source for both median wages and the consumer price index and are easily available from Federal Reserve Economic Data, Federal Reserve Bank of St. Louis.series. Median wages can be accessed at https://fred.stlouisfed.org/series/LEU0254531300A and urban consumer price index can be accessed at https://fred.stlouisfed.org/series/CPIAUCSL.

The bottom-up framework focuses on the energy system frequently with little tie to the remainder of the economy (DeCarolis et al., 2017; Pfenninger et al., 2014). Energy system models emphasize the process that supply energy-services to end-users. They include all elements of the energy system, such as fuel resources, technologies for producing energy, infrastructure for carrying energy sources, end-use technologies and processes, and energy service demand by sector. Systems are often optimized by choosing the least cost option to satisfy a given demand target over a specified planning horizon (MARKAL[2] and its successor, TIMES[3]). This approach is particularly helpful for research and development planning because it evaluates the relative costs of competing technologies serving end-use service demands. The framework's assumption of optimizing behavior may be at odds with real-market behavior; Trutnevyte (2016) has evaluated the potential biases of this assumption and how modelers might adjust for it.

Some system models can be simulated without forcing an immediate supply-demand equilibrium or optimizing behavior. One example of this latter approach is systems dynamics, which views the whole system as comprised of numerous circular and connecting parts that can relate to each other with time delays (Naill, 1992).

Due to the limitations of each approach, hybrid systems have been proposed to combine the best features of energy-systems and general equilibrium approaches into a single model or integrated system of models. An early effort along these lines was the linking of Hudson-Jorgenson general economic model with the Brookhaven energy system model (Hoffman and Jorgenson, 1977). Another was ETA-Macro (Manne, 1979), which explored the dual linkage between a systems approach for representing explicit technologies within the energy sector and the remaining aggregate U. S. economy

[2] MARKet Allocation selects the least expensive combination of technologies to satisfy inputs for energy demands and emissions reductions.

[3] The Integrated MARKAL-EFOM System adopts the MARKAL approach but includes several additional features. An important technique is the ability to allow flexible time slices for all energy sources rather than for electricity and heat alone.

within the confines of a single integrated model. Rivers and Jaccard (2005) provide a more recent example that combines detailed discrete choice models for consumer behavior with a comprehensive energy-economy model.[4]

Hybrid systems represent some sectors with multiple market equilibria or macroeconomic conditions and other sectors with greater technological or behavior detail. A widespread current application is to link technology choice within the electric power generation and dispatch system to the general equilibrium conditions in the remainder of the economy. A prominent example of such a system is the US-REGEN model that has been applied to evaluate a number of policy issues (e.g., see Bistline and Blanford, 2020).

Energy models are an integral feature of climate change frameworks because they represent the energy system costs associated with a transition from fossil-based fuels to carbon-free sources. Evaluating greenhouse mitigation options, however, also requires information that incorporates potential benefits from reducing emissions like preventing reductions in crop growth, land inundated by sea level rise, and additional deaths from heat stress. A further differentiation in model type concerns how these integrated systems incorporate these climate change impacts (Weyant 2017). The more disaggregated detailed process (DP) integrated assessment models (IAMs) represent climate change impacts for different regions and sectors that may include the physical impacts or some monetary measure. In contrast, benefit–cost (BC) IAMs represent climate change impacts in more aggregated terms in a single monetary metric to conduct cost-benefit analysis.

3. Uses and Misuses of Models

This remainder of this essay discusses the process of evaluating models assuming that the public policymaking organization is committed to using models, perhaps combined with other knowledge, to

[4] Discrete choice models explain the probability that competing options will be chosen based upon different fuel prices, activity levels, and personal characteristics of the decisionmaker.

develop its strategy. One should begin the process by asking when and how should one use a model. For concreteness, it is helpful to begin initially with models focused upon a single topic (in this case, climate change policy) and close with comments about using models applied to any public policy problem.

Integrated assessment models (IAMs) for evaluating the adjustments to different restrictions on greenhouse gas concentration levels have generated considerable controversy. Pindyck (2013, 2017) criticizes the numerical estimates for carbon-induced damages, their dependence upon the intertemporal discount rate, and a failure to capture low-probability but catastrophic climate events. Metcalf and Stock (2017) believe that despite their limitations, the initial social cost estimates based upon these models provide one of several possibilities for guiding public policy decisions and represent a useful first step for understanding the public policy tradeoffs. Nordhaus (2014) and Weyant (2017) acknowledge some of the limitations of existing estimates but believe that the true value of these models lies beyond the numerical estimates they provide.

Essentially, the issue comes down to whether one wants a numerical estimate for future conditions or about an important metric, or one wants a framework that provides a compelling narrative about how the system works.[5] This situation means that models will be "a glass half empty" when a value is needed for a very uncertain outcome, such as climate change benefits that depend upon such factors as intertemporal discount rates, temperature responses to increasing concentrations, monetary damages for physical impacts, and catastrophic climate events. When conceptual learning may be more valuable than a numerical outcome, the models may be "a glass half full." A well-developed framework for integrated assessment may provide a valuable guide for future research on such key unknowns as the appropriate discount rate and damages (Nordhaus 2014). Furthermore, in their current versions,

[5] Several pithy quotes capture the spirit of the second argument: "all models are wrong, but some are useful (George Box)," "computing for insights, not numbers" (H.K. Hamming) or "a good decision is based on knowledge and not on numbers" (Plato).

they have improved one's understanding of the tradeoff between emissions mitigation and other social goals like biological research to prevent pandemics, the role of flexible market-oriented policy instruments applied broadly in cost-effective solutions, and the value of information that reduces the uncertainty about advanced technologies and scientific progress.

The debate about if, when and how to use models permeates many public policy topics and is not limited to the energy and climate change discussions (e.g., see Kay and King, 2020, chapter 20). Drawing upon their discussion, models can be misused in a variety of ways:

1. They can often be applied to problems or choices that are disparate from the original purpose for developing the framework.
2. They require numbers for policy evaluation where sometimes these estimates are highly uncertain or subjective.
3. They usually assume that the process under investigation does not change when shifted in time.
4. As these systems become more complex, they often fail to communicate the bounds of uncertainty for the model outcomes they represent. They may not reveal which future options may become available and which future options may be removed with a given strategy.
5. Their complexity may preclude them from communicating with all interested parties and thereby preventing collaboration about possible options. Instead, the easiest response often becomes commissioning a new study that adopts a new set of assumptions, setting in place a war between modelling studies.

At the same time, models can be used to define more precisely a conceptual narrative about how things work:

1. They should embody relatively simple conceptual ideas to identify the critical factors that shape the decision and potential tradeoffs.

2. They should foster future research that will provide more reliable estimates for the critical factors and parameters.
3. Flexibility in model structure allows one to consider robust strategies that produce acceptable outcomes in a range of different outlooks for future conditions or scenarios (Lempert 2019).
4. The framework should allow additional metrics to evaluate the outcome of a strategy when new issues or decisions arise. As an example, systems focused upon the least-cost strategy for ameliorating climate change impacts may need to be revised to incorporate the growing interest in their distributional impacts.

4. Alternative Approaches for Evaluating Models

Communicating critical insights derived from models into the public policy process has always been difficult, but it appears particularly challenging given the complexity of current frameworks. As public policymakers used energy models more frequently, they began to ask more difficult questions that required greater complexity and detail to provide useful insights. Today models cover a wider set of issues and can easily expand along numerous dimensions – fuels, sectors, regions, and time periods. Each expansion introduces new interactions that must be incorporated, such that many integrated assessment (IAM) models used for climate change analysis bring together energy, the economy, land, climate and other critical sectors.

Complexity makes it much more difficult to understand what causes the observed behavior in a model. How does one test for different parameters or responses in a model where one sector or concept depends upon what happens in the larger more complex system. If it became helpful to understand how responsive energy consumption was to changes in energy prices or economic activity, it is not straightforward how one would adopt assumptions that would appropriately control supply conditions.

Evaluating models typically takes one of two approaches: (1) single model evaluation or verification including open-source model sharing, and (2) inter-model comparison that may or may not involve active participation by the user community. Before focusing upon the second approach, we discuss both below because each provides relevant information to the policymaking organization.

4.1 Single Model Evaluations

Once an organization has committed to a single framework, it becomes difficult to broaden the focus to consider what other modelling systems might provide. Nevertheless, there are some standard ways to evaluate the selected system and judge whether it is providing reliable results for the specific decision of interest. These evaluations may include model verification, validation, ventilation (or transparency), forecasting accuracy, backcasting, replication by other researchers and usability by those making policy decisions. All these efforts can be time consuming.

Audits verify that the model does what it intends to do. They confirm that results are correctly calculated and identify any potential errors of significance. If conflicts arise between the model developers and the auditing group, however, the process can become combative. The modeler feels picked upon and the auditor fears that the discussions are not sufficiently open to be useful.

Model validation efforts test the behavior of the model, whether it produces sensible results and if it tracks real world energy markets. Often, insights about important energy or environmental responses and conditions are more important than precision in forecasting the correct number (Huntington et al., 1982), particularly when there is deep uncertainty about responses and future conditions (Lempert, 2019). But a poor forecast of reference conditions could seriously bias assessments of carbon targets based upon historical levels. More rapid reference emissions growth leads to much more severe carbon constraints to return the future emissions levels back to historical values. Another problem with validating a model by its forecasting accuracy concerns the choice of metrics. Seldom does one model's forecast perform better than other models on all important variables. The choice of variable or concept will depend very much on the policy decision under consideration.

Another useful technique is to apply *ex post* evaluation of a model's output to ascertain whether the estimates approximate or deviate noticeably from actual history. In these backcasting exercises, the analyst tries to use actual values for any exogenous variables to track past values more closely. Comparing the backcasted and actual values of key dependent variables provides an opportunity to inform policymakers how accurate the system's forecast have been over past experiences. When applied to evaluating the effects of past policies, this technique often finds that climate policies are not as burdensome as depicted in many models (Cropper et al., 2018). This result happens because real-market behavior can be more flexible than how it is represented in formal models.

Ventilation is the process of increasing the model's transparency or openness. Does the model cast sunlight on a complicated issue? Does it explain the results in a straightforward way? Transparency often involves a leader of the modelling team who can explain results in an interesting, engaging, and truthful manner.

Users may accept a model more readily if other researchers can replicate its findings. DeCarolis et al. (2012) reviewed twelve models with respect to their transparency. The paper suggests that in most cases, replication of model results is currently impossible. Replication raises some important issues because modelling teams may be reluctant to release data and code that took years to organize and process. The software Excel does not release its code, but users have faith in the calculations because so many groups use the program for a variety of projects. It standardizes the routines in a way that makes results believable.

Open source modelling frameworks often do the same thing. An early effort along these lines was the open source modelling system for long-run integrated assessment and energy planning (OSeMOSIS) developed by Howells et al. (2011). This platform has also benefitted from a close relationship with two United Nations agencies. Pfenninger et al. (2018) discuss other similar projects and explain the general process and existing challenges in developing these systems. Groissböck (2019) argues that this approach can often be useful but also notes several promising topics for further elaboration.

Model usability by the policymaking community can be encouraged by a well-written model documentation that explains what is included in a system and how to use the model. Although these documents can be enormously useful for an agency that wants hands-on experience with operating the model, many model documentations do not provide information about how much key variables respond to policy shocks or changed energy and economic conditions. Although it may be too burdensome to report all fuel and sector price and income elasticities to demonstrate individual responses, policymakers would find it useful to have some summary metrics about how the overall system responds (e.g., to a carbon fee with lump-sum redistribution).

4.2 Multiple Model Evaluations

Another evaluation approach involves working with more than one system. Comparisons of model structures based upon system documentation or model polls can provide important information about the types of activities and responses that are covered in different models. By itself, however, structure does not necessarily reveal much about the behavior of a model when applied to a particular problem. Some models may have similar structures but respond quite differently to the implementation of a policy instrument like a carbon tax or cap-and-trade agreement. A policymaker needs to know when they use a model that appears very much more or less responsive than other available frameworks.

Comparisons of model outputs from standardized scenarios represent another approach that can reveal important information about how different models behave. Modelers have worked cooperatively with each other in different disciplines for a number of decades. Supported by the National Science Foundation, the National Bureau of Economic Research (NBER) organized a number of seminars to compare U.S. macroeconomic forecasts and model responses (Fromm and Klein, 1973). The (National Research Council, 1978) established the Modeling Resource Group (MRG) to evaluate strategies for satisfying future U.S. energy demand. Similar suggestions for joint modelling activities to communicate to policymakers have been offered for health care policy modelling (Ringel et al., 2010).

By applying common assumptions to different models, intermodal comparisons can show how models respond to alternative conditions and reveal the most important factors contributing to these different responses. Productive intermodal comparisons depend critically upon the development of the common rules used by participating model teams. Intermodel comparisons do not necessarily require active user participation to produce interesting and innovative results, although often modelling teams are familiar with the key policy issues.

When these comparisons are made in the presence of and with joint participation by the model-using community, Greenberger and Richels (1979) describe this effort as forum analysis. Comparisons of model outputs from standardized scenarios can be used not only as a platform for modelling teams to share relevant data and best practices but also as a communication bridge with model users. In concept, forum analysis offers model users an important role in defining, using and communicating the study results. Forum analysis can provide important perspectives for resolving critical policy issues. Users can help define important policy issues in ways that are more meaningful for policy developers than for energy modelers. In return, modelers can help shape the views of policy developers to ask questions in a way that generates more useful insights. Forum analysis will be most effective when two-way communication between the two groups provides a richer understanding than operating within a group of model developers or users alone.

4.3 Multiple Model Evaluations in Practice

The forum approach was pioneered by the Energy Modeling Forum at Stanford University in 1976 (Huntington et al., 1982). Improving both the use and usefulness of existing energy models was this activity's principal mission. *Ad hoc* working groups of policy developers and modelling teams conducted each study with the following goals: (1) improve understanding of important energy system dynamics and problems, (2) contrast differences, advantages and limitations of various approaches, and (3) guide future improvements, linkages, and extensions of energy models and relevant data sets. These groups conducted each study by following a set of five design principles ranging from user orientation to decentralized analysis. See Table 1. Examples of what policymakers learned from the first six EMF studies are provided in this early discussion about the process. Studies that provided a believable narrative about how things worked tended to be more successful. Efforts that tried to anticipate how future energy markets would evolve were less successful due to the large uncertainty associated with assumptions about the conditioning factors.

Participants recognized early in the process that informal communication between model developers and users was, at a minimum, as important as any formal communication through reports, journal articles, testimonies, or workshops. Some participants in government organizations revised the price elasticity of energy demand used in their policy models and analysis, once they had the benefit of discussing this issue with other analysts (Huntington, et al, 1982). In a later effort, economists in the US government learned that they had used a model that produced economic costs of reducing greenhouse gas emissions that were close to the average of ten different models and that were not low-ball estimates biasing their results towards accepting the mitigation strategy (Frankel 2008). Although these studies leave many important issue unresolved, this informal communication can be a powerful reason why many government organizations find value in these types of analysis.

Over the last decade or two, multi-model comparisons have expanded dramatically, initially within the United States and Europe but now also in Asia and Latin America. The efforts have focused upon large-scale integration assessment models for evaluating national and international greenhouse gas emissions strategies. Weyant and Kriegler (2014) provide a useful summary of some but not all major efforts including the Innovation Modeling Comparison Project (IMCP), the Climate Change Science Program (CCSP) Product 2.1 (a) study, the Asian Modeling Exercise (AME), the Roadmaps towards Sustainable Energy future (RoSE) project, the Low Climate Impact Scenarios and The Implications of Required Tight Emission Control Strategies (LIMITS) project, the Assessment of Climate Change Mitigation Pathways and Evaluation of the Robustness of Mitigation Cost Estimates (AMPERE) project, the Program on Integrated Assessment Modeling Development, Diagnostic and Inter-Comparisons (PIAMDDI), and the Latin American Modeling Project (LAMP) project. In addition, modelling teams have found it very useful to organize their efforts through the Integrated Assessment Modeling Consortium (n.d.). These collaborations within and outside the EMF umbrella have produced a number of important journal articles and have greatly facilitated the Intergovernmental Panel on Climate Change process on climate change mitigation (IPCC, 2014).

In a recent and important study organized by the U.S. government, IWG (Interagency Working Group, 2010, 2016)[6] simulated three integrated assessment models using standardized scenarios in order to estimate the social cost of carbon, which is the monetized value of additional carbon emissions that is not incorporated in private decisions. This value for social costs determines how aggressive the United States should be in reducing emissions, with higher estimates warranting more costly strategies. In contrast to many previous model-comparison studies, this effort focused upon deriving a specific value for a critical input in environmental decision making. The IWG estimates have generated considerable controversy about the uses and misuses of models for addressing the benefits from adopting climate change policies that were discussed in section 3 above.

[6] This effort continues with updated assumptions but these references provide a useful description of the IWG process.

5. Remaining Challenges for the Model Evaluation Process

More than 40 years ago, a thorough review of the model analysis process applied to energy systems and markets closed by posing an interesting query:

> The most interesting question to us is whether development along the main path will become well enough received, especially by policymakers and model users generally, to lead to institutionalisation of model analysis in the form of a recognized professional discipline with laboratories and centers in many application areas and in many regions of the country ((Greenberger and Richels, 1979, p. 499).

In hindsight, this comment when applied to forum analysis appears quite prescient. As the development and application of energy system models have expanded multiple times over, many have recognized the critical need to evaluate these systems and to increase their application in meaningful ways.

What has change over the last 40 years has been the increasing complexity and sophistication of any modelling system that policymaking organizations are likely to use for making decisions or identifying strategies. This development has made it increasingly more difficult to maintain a communication bridge between modelling teams wanting to use the most recent analytical techniques and the model-using community desiring a clear and transparent explanation about how the results are derived. The forum analysis identified by Greenberger and Richels (1979) has been a major workhorse for bringing these groups together, but there are some important challenges confronting any effort to foster continued communication. Table 2 briefly annotates nine major challenges to increased communication that have surfaced during recent efforts to compare models. They are not the only hurdles, but they represent a minimum set that can guide future efforts to use models as a learning rather than a forecasting tool.

5.1 Model User Perspective

First, above all, the effort should directly engage the model user. The process should integrate the model user into the process of comparing models whenever possible. Since the models are always reaching for more complexity, it is not easy to accomplish this objective. Ignoring this step altogether, however, severely limits the usefulness of a model-comparison exercise. The model-using community is likely to learn more about the systems when they participate in the process of identifying key results than from after-the-fact briefings that detail the final results or describe the physical characteristics of the participating models. In addition, users often bring to the table an understanding of the most critical policy issues, tradeoffs and debates. Participants sometimes ask whether the discussions should focus upon models or policies. The appropriate response appears to be both. Avoiding detailed consideration of the policy implications or sufficient background on the nature of the models shuts down the two-way communication.

Second, when the model-using community is an important consumer, the model analysis process should ventilate the model's complexity with intuitive insights. The user can enjoy the car ride without understanding the motor's interior design. Ventilation is the process of increasing the model's transparency or openness.

Third, closely related is the use of simple models for demonstrating the importance of key parameters or responses. Typically, one develops a benchmark curve or frontier that displays how results change with the critical factor. Model results are then plotted in relationship to this benchmark in order to explain the range of results.

The first study conducted by the Energy Modeling Forum (1977) used this approach with considerable success to explain how restrictions on energy availability caused reductions in long-run economic output. It emphasized the important role played by the substitution response between energy and non-energy factors of production, which is closely related to the price elasticity of energy demand. Another important factor was the effect of decreased energy availability on capital formation.

A more recent example occurred during the EMF24 study (Fawcett et al., 2014), which developed a policy-efficient frontier that shows the cumulative costs (e.g., like net present value of the loss in the consumption component of gross national product) as a function of cumulative carbon dioxide emissions reduction. These frontiers are based upon the cap-and-trade scenarios, which represent cost-effective strategies for implementing different emission reduction strategies because they allow complete flexibility in choosing the lower cost options. Placed upon these frontiers are the model results from various technology strategies (e.g., like optimistic nuclear conditions) that allow flexibility in only one of many options. Distances between the technology cases and the frontiers provide the policymaker with an intuitive perspective on the relative efficiency of each set of technology conditions.

This very effective approach is often underutilized in many model-comparison efforts for several reasons. First, simplifying complex systems is often an art that may ignore factors that some participants view as critical. Second, it requires an indepth understanding of all the models prior to developing the scenarios if one wants to quantify the factors for use in the benchmark.

5.2 Simplifying Model Responses

Fourth, occasional use of meta-analysis may sometimes contribute additional value added (Fischer and Morgenstern, 2006). In this approach, researchers use statistical analysis to combine the results of multiple original studies, or in this case, models. This analysis reveals how changes in various model attributes and variables influence the reported results for important policy variables like energy use or emissions reductions. When one knows a priori which factors are most important and all systems can report the required data, these techniques are useful for summarizing responses. Often, however, a forum analysis may involve models with very different formulations based upon varying disciplines. Any meta-analysis should be carefully integrated and coordinated with the study's organization as well as with the evolving discussion of the study results as key findings are revealed.

5.3 Calibrating Model Responses

A fifth challenge involves summarizing important responses in ways that help policymakers understand an individual model or group of models. An important model trait is often the degree of substitution between fuels and between energy and nonenergy inputs. Although it may be cumbersome to report multiple responses in a digestible way, a good summary for models evaluating climate change policy might be a standard generic carbon tax or fee case with explicit rules for recycling revenues like lump-sum redistribution. Moreover, once this case has been specified and tested, it might serve as a useful benchmark for any future model to report results in any documentation that it provides. This procedure would allow policymakers the opportunity to quickly understand how any new model compares with others.

There may also be opportunities to report model responses for key underlying factors and calibrate them with historical trends where possible. Although future trends are likely to differ from past experience, the latter provides an effective benchmark for understanding the model's projections. Key responses may be the price and income elasticities of energy (or major fuel) demand or the response of economic growth or emissions reductions to carbon or greenhouse gas prices. By focusing upon a single fuel or sector, (Energy Modeling Forum, 2013. pp. 23-25) developed scenarios for deriving responses for natural gas, while Edelenbosch et al. (2017) performed similar analyses for transportation fuels. More challenging is deriving system-wide responses, but EMF Group Working 4 (1981) presents a comprehensive statement of the procedure and some early estimates.

A sixth and related concept is how well the system "backcasts" a past period. Instead of projecting future market outcomes, backcasting simulates past experiences using correct assumptions that match actual values for the exogenous variables. This procedure potentially allows one to view how well the system tracks what is known to have happened in a previous period. This task can be very difficult to implement in a complicated framework with many various interactions and processes that may require some calibration. It may also introduce many arbitrary adjustments when finetuning the backcast to match historical fact.

Backcasting can also be conducted for analyzing past policies or events.[7] In this case, the outcome of the policy case is known to be the *ex-post* set of conditions that have already been observed. Perhaps more challenging is to define the counterfactual conditions for representing what the outcome would be in the absence of the policy. Rose and Blomberg (2010) describe a multi-model study that suggests that U.S. resilience in the face of the 9/11 attack on the World Towers may have partially explained why the repercussions were not as large as many had feared.

Although the modelling team can learn much from a carefully implemented and reasoned backcast of their system, it may be best done by the individual group. If done within a multiple-model forum analysis, there will be constant competition between groups trying to duplicate the known historical experience.

Although multi-model backcasting comparisons are seldom done, several studies suggest the approach that could be undertaken. Huntington (1994) applies this technique for evaluating where oil price projections were wrong over the 1980s. In addition to lower economic growth within the developing world, this analysis demonstrated the role played by oil demand that was less responsive to falling prices after 1982 than they were to rising prices during the 1970s. Without direct access to the models themselves, the author conducted the backcast by using the simple oil market model developed during the model-comparison study.

[7] Useful discussions with Richard Tol and Adam Rose helped the author to appreciate how one might engage in backcasting when comparing models.

In another backcasting study applying multiple models, Tol (2014) compares ten different frameworks for their ability to predict the estimated reductions in carbon emissions from past European climate policy. Without direct access to the models themselves, the author developed statistical functions to derive model-specific relationships between carbon taxes and emissions reductions from output representing future scenarios. These statistical emulators were used to simulate historical climate policies in an effort to ascertain the potential biases of each system as well as of all systems collectively. He concludes that models often used by the European Union tended to perform less well than others.

5.4 Policy-Specific and Diagnostic Scenarios

Other challenges relate to the process of selecting and specifying high-value scenarios. While comparisons of forecasts can be done with a single projection from each group, policy comparisons require one to compare the policy simulation with a baseline or reference case. A seventh challenge involves selecting standardized assumptions for the baseline conditions. If one performs a physics experiment on a material in the laboratory, one wants to control carefully for temperature, pressure and other external effects in order to establish a conditional relationship. Defining this relationship may have considerable merit, but it will not capture the underlying uncertainty of the outcome when temperature, pressure and other external factors can also change. Many model comparisons focused upon climate change strategies or targets often allow participating frameworks to use their own internal outlooks for how the future economic and energy trends will evolve under baseline or reference conditions. This approach is attractive for capturing the uncertainty in the policy outcomes due to differences in the underlying baseline conditions, such as economic growth, world oil prices or restrictive future policies discouraging coal or nuclear use in electric generation. The principal disadvantage in this approach is that the underlying baseline conditions are not the same in all models and it will be more difficult to explain differences in the model responses to the policy or technology shocks. The process will need to select the best option for providing the most relevant information for the target group of policymakers.

Specifying alternative policy, economic, or technology scenarios for comparison with the baseline is an eighth challenge that depends upon the ultimate mission. It is often underappreciated that policy relevant information can be generated from a wide range of possible alternative scenarios. The process should use policy-specific cases when the goal is to evaluate a specific policy action or strategy under current consideration by a government entity. Coordinating these specific assumptions with multiple modelling groups can be quite challenging because the policy details are often constantly shifting and being redefined by the government rulemaking process. Another drawback of this approach is that it can cast the specific policy as a winning or losing strategy that may alienate some policymakers and make them distrust the "biased" results of the participating models. A less precise but potentially useful approach would be to relax assumptions on all but a few key mechanisms that would inform policymakers about general principles without attempting to directly evaluate the proposed policy instrument.

The U.S. Clean Power Plan proposed by the Obama administration generated a lot of policy interest as a way to mitigate carbon dioxide emissions within the generation sector. Rather than anticipating the precise rules that would eventually be finalized, Energy Modeling Forum (2015) was able to generate initial results by specifying a generic technology performance standard (TPS) for the electric power sector. In a similar vein, a later EMF study (Bistline et al., 2018) captured the main ingredients of the same policy instrument with a power sector mass-based cap-and-trade simulation. Although these cases did not precisely measure the exact policy options that were eventually adopted, they added considerable value added to the policymaking decisions.

A number of model-comparison studies evaluate policy-general cases that are not tied directly to specific proposals or legislation. Included in these efforts are those that specify an initial level and rate of increase for a carbon tax in real or inflation-adjusted terms. Even when policymakers are unlikely to adopt a pure carbon tax, these cases can be a quite useful benchmark, because under certain assumptions they can represent the least-costly approach for mitigating greenhouse gases. Model-comparison studies should consider policy-general cases to provide important insights about key factors that may shape the response to any specific policy instrument. It is not always necessary to represent each policy alternative for the process to be useful.

A final alternative scenario would be diagnostic cases that communicate a key response in various models that critically determine their outcomes. Diagnostic cases focus on the key relationships between variables that do not directly address a policy concern. Although they initially may appear to be less policy relevant, they can often be an important method for providing insights that help to improve the understanding of how energy markets might respond to a number of different conditions. For example, future long-run oil market prices and quantities are very dependent upon several basic economic conditions including future oil demand trends. Energy Modeling Forum (1991) developed structural scenarios to decompose the trends in oil consumption projections into several components, including the effects of future prices, income (GDP) growth, technical trends and momentum effects caused by past price shocks. In a later study on the emerging market dynamics for natural gas in the wake of the shale revolution, Energy Modeling Forum (2013, pp. 23-25) reports estimates for the price elasticities of natural gas supply and demand that can be useful for anticipating future market outcomes caused by changes in a range of different conditions like economic growth, policies, or resource availability. Price elasticities of supply were computed by comparing the price and quantity responses to high economic growth case relative to reference growth, where both cases assumed the same underlying conditions defining the supply curve or schedule.[8] Price elasticities of demand were

[8] Supply curves hold constant resource availability, extraction costs and technology trends that allow natural gas production quantities to change only if market prices vary.

calibrated by comparing the price and quantity responses to low natural gas supply conditions (more expensive resources combined with less technical improvements) relative to the reference supply conditions, where both cases assumed the same underlying reference demand conditions defining the demand curve or schedule.[9]

5.5 Institutionalising the Process

Given the current interest in communicating with policymakers, it is likely that more organizations will consider the prospects for adopting the forum approach. If such efforts are undertaken, a ninth challenge is to institutionalise the model evaluation process whenever possible. In organizing a forum study, the lead people who guide the process should combine a knowledge of the existing models and the frontier for best practices with an understanding of the emerging policy issues circulating within the government agencies. The process should be nonpartisan without favoring a particular policy strategy or instrument or a specific energy technology or option. This balanced approach should also apply to the participating models too. A proprietor of an existing model undoubtedly possesses a valuable perspective and knowledge of the often complicated modelling techniques, but there may be some concerns if that person also leads the forum analysis. When possible, it is usually better to avoid any possibility of one system being favored over another. Other modelling teams are more likely to participate in the study if there is a perception that all models will be treated fairly.

A related institutional issue is the financial support to allow frameworks to participate in forum studies. Models provide a framework for learning iteratively. Using and comparing models provide considerable value added, but it is important to update the learning from these experiences and to allow each effort to build upon the findings of past studies. Model-using organizations should

[9] Demand curves hold constant economic growth, energy prices other than natural gas, and technology trends that allow natural gas consumption quantities to change only if market prices vary.

sufficiently fund, support and reward the model evaluation process, including participation by modelling teams in group forum studies to compare multiple frameworks. Model evaluation should be planned for when initiating a modelling project and integrated into the overall process and budget.

6. Policy Implications and Conclusion

Model evaluation is best considered as a process for communicating with the policymaking or policy-advising community. It is important to appreciate the regional, temporal and sectoral coverage of any model, but policy developers may be more interested in the model's behavior when it is applied to critical policy options. This paper has emphasized model-evaluation approaches for policymakers to use when they are interested in understanding model behavior.

Six decades of energy modelling have witnessed increasing complexity in these systems reinforced by expanded computing power and the electronic availability of new data sources. With the rise of the climate change issue, policy interest shifted from evaluating key single energy markets like oil and power to more integrated systems balancing all fuels and markets. This transformation also resulted in more complex models for representing the whole energy system.

Increasing complexity may allow these systems to capture more interactions that influence the model results. However, as these systems become more complex, evaluating them to improve their usability will be difficult. When used as a learning rather than forecasting tool, these systems can be evaluated individually one by one or through joint efforts to compare them in multi-model exercises. Efforts that integrate closer interactions with the policymaker or policy advisors are often called forum analysis.

Model forum evaluation often reverses the process of model building. To use an analogy with home construction, good model developers often begin with the cellar and build upward with a solid theoretical basis, justifying each assumption as one proceeds. Inputs are carefully scrutinized before evaluating the results and key conclusions. Policymakers initially are often less interested in the theoretical justification and multitude of conceptual assumptions, preferring instead to focus on model outputs. Useful model evaluation begins with model outputs to identify the reasons for the variation in results from different models. If the discussion can be shaped effectively in this way, it brings the policymaker through the front door directly to the “crown jewels” rather than through the back door and into the cellar.

Joint efforts to combine models into one study have become very popular in recent years. As a guide to future collaborations, this essay closed by highlighting a few remaining challenges that would improve the value of these studies for the policymaking community. These challenges call for these efforts to engage the model user, ventilate the models’ complexity with intuitive insights, use simple models to demonstrate key parameters or responses, judicious occasional use of meta-analysis when there is value added, report model responses and calibrate them for decisionmakers, consider the system “backcast” for a past period (when possible), select standardized or modeler-choice baseline conditions, selectively develop policy or diagnostic alternative cases, and institutionalise the model evaluation process.

References

AI Impacts, n.d. Trends in the cost of computing [WWW Document]. URL https://aiimpacts.org/trends-in-the-cost-of-computing/ (accessed 12.6.20).

Bistline, J.E., Hodson, E., Rossmann, C.G., Creason, J., Murray, B., Barron, A.R., 2018. Electric sector policy, technological change, and U.S. emissions reductions goals: Results from the EMF 32 model intercomparison project. Energy Econ. 73, 307–325. https://doi.org/10.1016/J.ENECO.2018.04.012

Bistline, J.E.T., Blanford, G.J., 2020. Value of technology in the U.S. electric power sector: Impacts of full portfolios and technological change on the costs of meeting decarbonization goals. Energy Econ. 86. https://doi.org/10.1016/j.eneco.2020.104694

Cropper, M.L., Morgenstern, R.D., Rivers, N., 2018. Facilitating retrospective analysis of environmental regulations. Rev. Environ. Econ. Policy 12, 359–370. https://doi.org/10.1093/reep/rey011

DeCarolis, J., Daly, H., Dodds, P., Keppo, I., Li, F., McDowall, W., Pye, S., Strachan, N., Trutnevyte, E., Usher, W., Winning, M., Yeh, S., Zeyringer, M., 2017. Formalizing best practice for energy system optimization modelling. Appl. Energy 194, 184–198. https://doi.org/10.1016/j.apenergy.2017.03.001

DeCarolis, J.F., Hunter, K., Sreepathi, S., 2012. The case for repeatable analysis with energy economy optimization models. Energy Econ. 34, 1845–1853. https://doi.org/10.1016/j.eneco.2012.07.004

Edelenbosch, O.Y., van Vuuren, D.P., Bertram, C., Carrara, S., Emmerling, J., Daly, H., Kitous, A., McCollum, D.L., Saadi Failali, N., 2017. Transport fuel demand responses to fuel price and income projections: Comparison of integrated assessment models. Transp. Res. Part D Transp. Environ. 55, 310–321. https://doi.org/10.1016/j.trd.2017.03.005

Edmonds, J., Reilly, J., 1983. A long-term global energy- economic model of carbon dioxide release from fossil fuel use. Energy Econ. 5, 74–88. https://doi.org/10.1016/0140-9883(83)90014-2

Energy Modeling Forum, 2015. North American Natural Gas Markets in Transition. Stanford University, Stanford, CA.

Energy Modeling Forum, 2013. Changing the Game? Emissions and Market Implications of New Natural Gas Supplies. Stanford University, Stanford, CA.

Energy Modeling Forum, 1991. International Oil Supplies and Demands. Stanford University, Stanford, CA.

Energy Modeling Forum, 1981. Aggregate elasticity of energy demand. Energy J. 2, 37–75. https://doi.org/10.5547/issn0195-6574-ej-vol2-no2-3

Energy Modeling Forum, 1977. Energy and the Economy. Stanford University, Stanford, CA.

Fawcett, A.A., Clarke, L.E., Rausch, S., Weyant, J.P., 2014. Overview of EMF 24 Policy Scenarios. Energy J. 35, 33–60. https://doi.org/10.5547/01956574.35.SI1.3

Fischer, C., Morgenstern, R.D., 2006. Carbon abatement costs: Why the wide range of estimates? Energy J. 27, 73–86. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol27-No2-5

Frankel, J. (2008), Serious research balances economic costs & environmental benefits of climate policy, Views on the Economy and the World (blog), Aug. https://www.belfercenter.org/publication/serious-research-balances-economic-costs-environmental-benefits-climate-policy

Fromm, G., Klein, L.R., 1973. A comparision of eleven econometric models of the United States. Am. Econ. Rev. 63 (2), 385–393.

Greenberger, M., Brewer, G.D., Hogan, W.W., Russell, M., 1983. Caught Unawares: The Energy Decade in Retrospect. Ballinger, Cambridge,MA.

Greenberger, M., Richels, R., 1979. Assessing energy policy models: current state and future directions. Annu. Rev. Energy 4, 467–500. https://doi.org/10.1146/annurev.eg.04.110179.002343

Groissböck, M., 2019. Are open source energy system optimization tools mature enough for serious use? Renew. Sustain. Energy Rev. 102, 234–248. https://doi.org/10.1016/J.RSER.2018.11.020

Hoffman, K.C., Jorgenson, D.W., 1977. Economic and technological models for evaluation of energy policy. Bell J. Econ. 8, 444–466. https://doi.org/10.2307/3003296

Hogan, W.W., 2002. Energy modeling for policy studies. Oper. Res. 50, 89–95. https://doi.org/10.1287/opre.50.1.89.17803

Howells, M., Rogner, H., Strachan, N., Heaps, C., Huntington, H., Kypreos, S., Hughes, A., Silveira, S., DeCarolis, J., Bazillian, M., Roehrl, A., 2011. OSeMOSYS: The Open Source Energy Modeling System: An introduction to its ethos, structure and development. Energy Policy 39, 5850–5870. https://doi.org/10.1016/j.enpol.2011.06.033

Huntington, H.G., 1994. Oil price forecasting in the 1980s: What went wrong. Energy J. 15, 1–22.

Huntington, H.G., Weyant, J.P., Sweeney, J.L., 1982. Modeling for insights, not numbers: the experiences of the energy modeling forum. Omega 10, 449–462.

Integrated Assessment Modeling Consortium, n.d. IAMC [WWW Document]. URL https://www.iamconsortium.org/ (accessed 12.6.20).

IPCC, 2014. Climate Change Mitigation. Fifth Assessment Report (AR5), Working Group III, Intergovernmental Panel on Climate Change.

IWG (Interagency Working Group). 2010. Interagency Working Group on Social Cost of Carbon, US Government, technical support document: Social cost of carbon for regulatory impact analysis under Executive Order 12866. http://www.epa.gov/oms/climate/regulations/scc-tsd.pdf

IWG (Interagency Working Group). 2016. Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis, US Government, technical support document under Executive Order 12866. https://www.epa.gov/sites/production/files/2016-12/documents/sc_co2_tsd_august_2016.pdf

Kay, J., King, M., 2020. Radical uncertainty: decision-making beyond the numbers, New York: W.W. Norton and Company, 544pp.

Lempert, R.J., 2019. Robust decision making (RDM), in: Marchau, V., Walker, W., Bloemen, P., Popper, S. (Eds.), Decision Making under Deep Uncertainty. Springer, Cham., pp. 23–51. https://doi.org/10.1007/978-3-030-05252-2_2

Mai, T., Bistline, J., Sun, Y., Cole, W., Marcy, C., Namovicz, C., Young, D., 2018. The role of input assumptions and model structures in projections of variable renewable energy: A multi-model perspective of the U.S. electricity system. Energy Econ. 76, 313–324. https://doi.org/10.1016/j.eneco.2018.10.019

Manne, A.S., 1979. ETA-MACRO: A model of energy-economy interactions, in: ACS Symposium Series. JAI Press, pp. 205–233.

Manne, A.S., 1976. ETA: A model for energy technology assessment. Bell J Econ 7, 379–406. https://doi.org/10.2307/3003263

Manne, A.S., Richels, R.G., 1991. Buying greenhouse insurance. Energy Policy 19, 543–552. https://doi.org/10.1016/0301-4215(91)90034-L

Massachusetts Institute of Technology, 1978. Independent assessment of energy policy models : two case studies. Cambridge, MA.

Metcalf, G.E., Stock, J.H., 2017. Integrated assessment models and the social cost of carbon: A review and assessment of U.S. experience. Rev. Environ. Econ. Policy 11(1):80–99. https://doi.org/10.1093/reep/rew014

Naill, R.F., 1992. A system dynamics model for national energy policy planning. Syst. Dyn. Rev. 8, 1–19. https://doi.org/10.1002/sdr.4260080102

National Research Council, 1978. Energy Modeling for an Uncertain Future. National Academies Press, Washington, DC. https://doi.org/10.17226/19960

Nordhaus, W., 2014. Estimates of the Social Cost of Carbon: Concepts and Results from the DICE-2013R Model and Alternative Approaches. J. Assoc. Environ. Resour. Econ. 1(1/2):273–312. https://doi.org/10.1086/676035

Nordhaus, W., 1993. Optimal greenhouse-gas reductions and tax policy in the "Dice" model. Am. Econ. Rev. 83(2), 313–317.

Nordhaus, W.D., 2007. Two centuries of productivity growth in computing. J. Econ. Hist. 67, 128–159. https://doi.org/10.1017/S0022050707000058

Nordhaus, W.D., 1993. Rolling the "DICE": an optimal transition path for controlling greenhouse gases. Resour. Energy Econ. 15, 27–50. https://doi.org/10.1016/0928-7655(93)90017-O

Nordhaus, W.D., 1973. The allocation of energy resources. Brookings Pap. Econ. Act. 1973, 529–576. https://doi.org/10.2307/2534202

Pfenninger, S., Hawkes, A., Keirstead, J., 2014. Energy systems modeling for twenty-first century energy challenges. Renew. Sustain. Energy Rev. 33, 74–86. https://doi.org/10.1016/j.rser.2014.02.003

Pfenninger, S., Hirth, L., Schlecht, I., Schmid, E., Wiese, F., Brown, T., Davis, C., Gidden, M., Heinrichs, H., Heuberger, C., Hilpert, S., Krien, U., Matke, C., Nebel, A., Morrison, R., Müller, B., Pleßmann, G., Reeg, M., Richstein, J.C., Shivakumar, A., Staffell, I., Tröndle, T., Wingenbach, C., 2018. Opening the black box of energy modelling: Strategies and lessons learned. Energy Strateg. Rev. 19, 73–81. https://doi.org/10.1016/j.esr.2017.12.002

Pindyck, R.S., 2013. Climate change policy: What do the models tell us? J. Econ. Lit. 51(3):860–72. https://doi.org/10.1257/jel.51.3.860

Pindyck, R.S., 2017. The use and misuse of models for climate policy. Rev. Environ. Econ. Policy 11(1):100–114. https://doi.org/10.1093/reep/rew012

Ringel, J.S., Eibner, C., Girosi, F., Cordova, A., McGlynn, E.A., 2010. Modeling health care policy alternatives. Health Serv. Res. 45, 1541–1558. https://doi.org/10.1111/j.1475-6773.2010.01146.x

Rivers, N., Jaccard, M., 2005. Combining top-down and bottom-up approaches to energy-economy modeling using discrete choice methods. Energy J. 26, 83–106. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol26-No1-4

Rose, A.Z., Blomberg, S.B., 2010. Total economic consequences of terrorist attacks: Insights from 9/11. Peace Econ. Peace Sci. Public Policy 16, 1–14. https://doi.org/10.2202/1554-8597.1189

Sweeney, J.L., Weyant, J.P., 1979. The Energy Modeling Forum: Past, present and future. Energy Modeling Forum, Stanford University, EMF PP 6.1, Stanford, CA.

Tol, R.S.J., 2014. Ambiguity reduction by objective model selection, with an application to the costs of the EU 2030 climate targets. Energies 7, 6886–6896. https://doi.org/10.3390/en7116886

Trutnevyte, E., 2016. Does cost optimization approximate the real-world energy transition? Energy 106, 182–193. https://doi.org/10.1016/j.energy.2016.03.038

Weyant, J., 2017. Some contributions of integrated assessment models of global climate change. Rev. Environ. Econ. Policy 11(1):115–137. https://doi.org/10.1093/reep/rew018

Weyant, J., Kriegler, E., 2014. Preface and introduction to EMF 27. Clim. Change 123, 345–352. https://doi.org/10.1007/s10584-014-1102-7

Wing, I.S., 2009. Computable general equilibrium models for the analysis of energy and climate policies, in: Evans, J., Hunt, L. (Eds.), International Handbook on the Economics of Energy. Edward Elgar Publishing Ltd., pp. 332–366. https://doi.org/10.4337/9781849801997.00019

Table 1. Design Principles for Energy Modeling Forum

• **User Orientation.** The EMF should work to improve the use and usefulness of energy models, approaching the studies from the user perspective and maintaining an active user involvement.
• **Model Comparison.** The EMF studies should compare the capabilities and limitations of many models, and these comparisons should be descriptive rather than normative. This is a unique contribution that the EMF can make, and it avoids the difficult problem of model validation.
• **Issue Focus.** For the general model user, abstract model comparison should be conducted in the context of the application to an important energy issue. This will provide a direction and discipline for the model tests.
• **Broad Participation.** The communication objectives of the EMF are best served if there is a wide participation in the selection of study topics, the formation of the working groups, and the dissemination of the study results.
• **Decentralized Analysis.** Existing energy models are often complex and require skillful application by the model developer. Despite the inherent advantages of third-party analysis, the EMF must rely on model tests as reported by the individual research group.

Source: Sweeney and Weyant (1979).

Table 2. Remaining Challenges for the Forum Evaluation Process

1. engage the model user
2. ventilate the models' complexity with intuitive insights
3. use simple models to demonstrate key parameters or responses
4. judicious occasional use of meta-analysis when there is value added
5. report model responses and calibrate them for decisionmakers
6. consider the system "backcast" for a past period (when possible)
7. select standardized or modeler-choice baseline conditions
8. selectively develop policy or diagnostic alternative cases
9. institutionalise the model evaluation process