Abating Carbon Dioxide Emissions from Electric Power Generation: Model Uncertainty and Regulatory Epistemology

Author(s): Alan H. Sanstad

Source: *The Journal of Legal Studies*, Vol. 44, No. S2, Developing Regulatory Policy in the Context of Deep Uncertainty: Legal, Economic, and Natural Science Perspectives (June 2015), pp. S423-S445

Published by: The University of Chicago Press for The University of Chicago Law School

Stable URL: https://www.jstor.org/stable/10.2307/26458525

Abating Carbon Dioxide Emissions from Electric Power Generation: Model Uncertainty and Regulatory Epistemology

Alan H. Sanstad

ABSTRACT

Computational modeling of natural, economic, and technological systems is a primary analytical methodology in US energy and environmental regulation. Validating or otherwise evaluating such models and analyzing the uncertainties involved in their regulatory applications have become both more important and more challenging. This paper reviews these issues in the context of an important recent example involving energy, the US Environmental Protection Agency's (EPA's) development of regulations to reduce carbon dioxide emissions from electric power plants using a numerical model of the US electric power system. Following a summary of background information about greenhouse gas abatement policy, the paper discusses the agency's general computational model evaluation philosophy; the history of, and current practices in, energy model evaluation; the specific model used by the EPA and its application to carbon dioxide regulation; and the concept of fundamental model uncertainty and its significance for this modeling domain.

1. INTRODUCTION

Computational modeling has become a primary regulatory methodology in the decades since the modern American environmental policy regime was established in the late 1960s and early 1970s. Models of environmental processes, energy systems, and economic impacts have proliferated and have become ever more critical to environmental policy making and regulation, while also becoming increasingly complex. Accordingly,

ALAN H. SANSTAD is an Affiliate Staff Scientist at the University of California, Berkeley. I would like to thank David Weisbach for inspiration and support and an anonymous referee for comments. I would also like to thank the University of Chicago Law School, the Center for Robust Decision Making on Climate and Energy Policy, and the Energy Policy Institute for the opportunity to develop, present, and discuss these ideas. All errors of omission or commission are my responsibility.

[*Journal of Legal Studies,* vol. 44 (June 2015)]
 0047-2530/2015/4402-0031$10.00

the problems of validating or otherwise evaluating such models and characterizing and measuring the uncertainties associated with their use in regulatory applications have become both more important and more challenging.

One of the most important current applications of computational modeling in environmental policy is the development of policies to reduce emissions of greenhouse gases (GHGs), particularly carbon dioxide (CO_2), from the production and consumption of energy. In the long-running climate science and policy debate, the validity of and uncertainties associated with numerical models of the global climate system—general circulation models—and their projections of future climate trends have been the subject of considerable controversy in policy circles and the political arena. But the corresponding issues in the realm of energy, economic, and policy modeling for GHG abatement policy have received much less attention.

This paper reviews issues of model validity, evaluation, and uncertainty with reference to an important recent example, the US Environmental Protection Agency's (EPA's) development of regulations to reduce CO_2 emissions from electric power plants. These were based on a quantitative analysis using an established numerical energy model, employed by the EPA for over a decade and maintained and operated in a manner consistent with the agency's guidelines for model evaluation and quality control.

Validating or evaluating computational models is a complicated subject, with theoretical frameworks and quantitative methods varying widely across disciplines; there are no definitions of these terms that apply universally. Nonetheless, in the physical and engineering sciences, these activities have traditionally been firmly although not exclusively grounded in the concepts of empirical fidelity and predictive accuracy (Oberkampf and Roy 2010; NRC Committee on Mathematical Foundations 2012). By contrast, as discussed in this paper, these concepts have come to be all but absent in the field of energy modeling and in key regulatory applications. This paper highlights the fact that a rigorous alternative epistemology for energy modeling and its regulatory applications has yet to be developed.

The issues discussed here offer a certain perspective on practical compliance, in the case of energy policy, with the requirements of the White House Office of Management and Budget's Circular A-4 on cost-benefit analysis of federal regulations (Office of Management and Budget 2003).

Many policy problems, and prospective regulations, are sufficiently complex that computational models of one form or another are needed to perform the calculations stipulated by Circular A-4. In the domain of energy policy, this is in fact the rule. Circular A-4 includes guidelines for the treatment of uncertainty. A reasonable interpretation of these guidelines is that they reflect the view that dealing with uncertainty is an elaboration, albeit an important one, on the basic elements of cost-benefit analysis. As described in this paper, however, there are forms of uncertainty attending the design and application of energy models that are fundamental but currently neither well understood nor the subject of active research. Thus, the reliance on computational modeling for cost-benefit analysis raises the question of how well such analysis, in the case of energy, complies with the Circular A-4 requirements.

The paper is organized as follows. The next section provides background on national and international attempts to implement GHG emissions abatement policy over the past several decades and then summarizes the EPA's recent actions on CO_2 emissions from electric power generation. Next is an overview of the Integrated Planning Model (IPM), the computational model used by the EPA for this purpose. The paper then turns to the history of validation and quantification of uncertainty in energy modeling in general and describes current standard practices in the application of these models to policy analysis. Following a further discussion of the IPM, the idea of fundamental model uncertainty is introduced. The paper ends with concluding remarks.

2. BACKGROUND ON GREENHOUSE GAS EMISSIONS ABATEMENT POLICY

The current US federal efforts to regulate GHGs follow more than 2 decades of national and international analyses of, deliberations on, and attempts to implement policies to address the long-term risks of global climate change resulting from human society's GHG emissions, particularly from the production and consumption of fossil fuel sources of energy. The first World Climate Conference was held in 1979 under the auspices of the World Meteorological Association, and the Intergovernmental Panel on Climate Change (IPCC) was established in 1988 to systematically gather, assess, and disseminate scientific knowledge about climate change, including human influences. An international negotiating committee was convened in 1991, and in 1992 it adopted the United Nations Framework Convention on Climate Change (UNFCCC), which called for "stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system" (1771 U.N.T.S. 107, art. 2). More than 150 countries, including the United States, signed the convention in the year following its introduction, and it entered into force in 1994.

In 1997, the Kyoto Protocol was adopted as a mechanism for implementing the UNFCCC. The protocol stipulated an international regime of GHG emissions reductions and was ultimately endorsed (signed and ratified) by nearly 200 countries. The United States signed, but did not ratify, the protocol. Since 1997, the IPCC has released its third, fourth, and fifth assessment reports, and international meetings and negotiations under the umbrella of the UNFCCC have continued. However, coordinated and sustained international action to reduce anthropogenic GHG emissions has yet to occur.

The Clinton administration was centrally involved in the development of the Kyoto Protocol but failed to secure its ratification by the Senate. The issue of climate change was deemphasized by the second Bush administration, and Congress engaged in several unsuccessful efforts to develop and pass GHG reduction legislation, including Senate Bill 2191, the Lieberman-Warner Climate Security Act of 2007. There were, however, a number of successful state and regional efforts to develop and implement large-scale GHG emissions reductions during the early to mid-2000s. These included California Assembly Bill 32, the California Global Warming Solutions Act, passed by the state legislature and signed by Governor Arnold Schwarzenegger in 2006. The bill calls for a comprehensive system of regulatory and market initiatives to reduce the state's GHG emissions, notably a cap-and-trade system, which was developed by the state's Air Resources Board and began operating in January 2013.

Climate change and GHG emissions abatement reemerged as national priorities during the first Obama administration, which supported House of Representatives Bill 2454, the American Clean Energy and Security Act of 2009—the Waxman-Markey Bill. This legislation would have established a national policy and regulatory system for emissions reduction, including a cap-and-trade system. The legislation was approved by the House of Representatives but was defeated in the Senate.

In 2006 the EPA issued electric utility generating unit (EGU) pollutant standards that did not address CO_2 (71 Fed. Reg. 9866 [February 27, 2006]). Two groups (one a consortium of US states, the other of environmental groups) filed petitions for judicial review of this ruling, contending that it incorrectly omitted CO_2 from the standards. In 2007, the Supreme Court ruled in *Massachusetts v. EPA* (549 U.S. 497 [2007]) that GHGs are air pollutants under the Clean Air Act (CAA) and that the EPA therefore had the authority to regulate their emissions under CAA, section 111. Ultimately, the EPA negotiated a settlement with the petitioners and undertook development of CO_2 regulations, the first of which were proposed in March 2012.

The proposed regulations took the form of new source performance standards (NSPSs) applying to all new fossil-fuel-fired EGUs of greater than 25 megawatt (MW) capacity (77 Fed. Reg. 22,392 [April 13, 2012]).[1] These units would be required to emit no more than 1,000 pounds of CO_2 per megawatt-hour annually. The EPA based this criterion on the emissions profile of natural gas combined-cycle (NGCC) power plants.

The EPA's 2012 regulatory impact analysis (RIA) found that the standard would be met by current and new NGCC units and in principle by coal plants using carbon capture and sequestration technology (although this still has yet to be fully developed and demonstrated, much less commercially deployed) (EPA Office of Air Quality Planning and Standards 2012). The RIA also found that, given current and anticipated market conditions and technologies, new fossil-fuel-fired generating capacity through 2020 would most likely be NGCC plants. Given that the most likely to be installed units are non-fossil-fuel fired—that is, they use renewable energy—and would a fortiori meet the standard, the EPA concluded that "technologies planned for new sources currently envisioned by owners and operators of EGUs will meet the regulatory requirements of this NSPS or are not covered by the NSPS" (EPA 2012, p. ES-3). Thus, it was projected that the costs of the regulations, and their effects on electricity prices, would be nil.

Following the comment period, the EPA withdrew the proposal and subsequently issued a revised version that entailed separate standards for NGCC plants and other fossil-fuel-fired plants (79 Fed. Reg. 1430 [January 8, 2014]). Finally, in June 2014, the agency proposed a complementary set of rules for existing power plants that undergo modification or reconstruction (79 Fed. Reg. 34,960 [June 18, 2014]). The RIA of the

1. By way of comparison, in 2011 there were 1,400 coal-fired units in the United States with an average capacity of 227 megawatts (MW) and 5,574 natural gas units with an average capacity of 74.5 MW (EIA 2013a).

2014 rules found that while the regulations would result in nonnegligible costs, these would be exceeded by their benefits; in 2030, for example, net benefits (in 2011 dollars) would range from $46 billion to $82 billion, depending on discounting and implementation assumptions (EPA Office of Air Quality Planning and Standards 2014).

3. ABOUT THE INTEGRATED PLANNING MODEL

The RIA for electricity-sector cost and energy estimates was developed by the EPA using the IPM, a proprietary model developed, maintained, and operated for the EPA by a private company, ICF International.[2] Its primary use has been air quality regulation.

The IPM is a detailed, deterministic model of the electric power sector in the continental United States of the linear programming (LP) type. Mathematically, LP models are optimization models in which some objective is minimized or maximized subject to a system of constraints defining the set of feasible options for operating some system. Among theoretical and computational approaches to optimization, the defining characteristic of LP models is that both the objective and the constraints are linear functions of their inputs. The LP concept originated in the 1930s in the work of the Russian mathematician Leonid Kantorovich as an approach to implementing central planning. With the advent of broadly accessible and affordable computer hardware and software, it has become the most widely used form of practical numerical optimization, with applications in industry, research, and policy analysis. (The key to this broad adoption is the linearity assumption, which has made LP in general considerably more tractable than most other forms of optimization, especially in large-scale applications.)

With time horizons up to the year 2050, the IPM calculates least-cost solutions to the problem of running the power system subject to constraints describing supply, demand, engineering and technical aspects, and costs. The model has about 2 million decision variables, namely, variables that are determined in the optimization and on the order of 200,000 constraints. It contains information about more than 15,000 existing and planned power plants. In brief, to analyze the CO_2 NSPS, the IPM was used to compute two solutions to the electric power planning

2. The use of proprietary models for regulation is an important issue for the evaluation of models but will not be addressed here.

problem (using a time horizon of 2020), one without the standards and one with them, and their accompanying costs, emissions, technology deployment patterns, and other quantities, which were compared to gauge the effects of the regulations.

4. EVALUATION OF MODELS IN THE ENVIRONMENTAL PROTECTION AGENCY

Among environmental and energy regulatory agencies, the EPA's efforts to grapple in a sustained and serious way with the problems of validating or evaluating computational models are a noteworthy exception. Much of this work has been driven by the agency's Science Advisory Board (SAB), established in the late 1970s with a mandate to advise the EPA on a range of technical issues, including evaluation of models (see, for example, SAB Environmental Engineering Committee 1989; SAB Modeling Peer Review Subcommittee 1993; EPA Science Policy Council 1999). Pursuant to the SAB's recommendations, the agency established its Council on Regulatory Environmental Modeling in 2000 to "[promote] scientific integrity and defensibility in the modeling principles, practices, and guidance which inform environmental and public health regulatory decision-making and research applications."[3]

Notwithstanding such efforts, a 2007 report of the National Research Council (NRC) (NRC Committee on Models 2007) found the EPA's processes and procedures for computational model evaluation lacking. While recognizing that the "EPA is a global leader in advancing and using models in the environmental regulatory decision process," the NRC stated that "the agency has not sufficiently leveraged opportunities to improve its regulatory decisions by adopting a comprehensive strategy for periodically evaluating and refining its models" (NRC Committee on Models 2007, p. 1). The NRC provided a review of the underlying issues and delineated a set of guidelines for the agency to follow in this area. The council's perspective was heavily influenced by the views in Oreskes (1998) and Oreskes, Shrader-Frechette, and Belitz (1994, p. 641), in which the authors argue that "verification and validation of numerical models of natural systems is impossible" and advocate instead the goal of "evaluation" of such models. The NRC characterized this as follows: "Model evaluation is the process of deciding whether and when a model is suit-

3. See Environmental Protection Agency, Environmental Modeling (http://www.epa.gov/crem/index.html).

able for its intended purpose. This process is not a strict validation or verification procedure but is one that builds confidence in model applications and increases the understanding of model strengths and limitations" (NRC Committee on Models 2007, p. 3).

This perspective is reflected in the report's six core recommendations to the EPA. Of these, three address substantive as opposed to process or procedural issues: first, describe the model and its intended uses; second, describe the relationship of the model to data, including the data for both inputs and corroboration; and third, describe how such data and other sources of information will be used to assess the ability of the model to meet its intended task (NRC Committee on Models 2007, p. 4).

In a 2009 report, the EPA presented guidelines based on the NRC's recommendations. Defining a model as "a simplification of reality that is constructed to gain insights into select attributes of a particular physical, biological, economic, or social system," the report notes that "[t]he challenge facing model developers and users is determining when a model, despite its uncertainties, can be appropriately used to inform a decision. Model evaluation is the process used to make this determination. . . . [M]odel evaluation is defined as the process used to generate information to determine whether a model and its analytical results are of a quality sufficient to serve as the basis for a decision" (EPA Council for Regulatory Environmental Modeling 2009, pp. 45, 19).

The available documentation indicates that the IPM has been subject to review procedures that, in the view of the EPA and stakeholders, have built confidence in the model and confirmed its suitability for use: "Quality assurance and verification of code is routinely performed by ICF. . . . Model inputs and results are corroborated through extensive review and comment by EPA stakeholders. . . . IPM is regularly used in comparative modeling exercises. . . . Regularly scheduled peer review is performed on key elements and assumptions in EPA's application of IPM" (EPA 2012, p. 5).

The detailed examples of review fully documented by the EPA are two 2003 workshops on the model's coal and natural gas supply assumptions. These entail extremely detailed expert assessment of the model's inputs, outputs, and processes governing the use of the fuels in the electric power system. These assessments did not, however, include any validation analysis as this is understood and practiced in scientific and engineering computation: some form of systematic, quantitative comparison of the model's outputs with empirical data.

5. ENERGY MODELING: HISTORY AND ASPECTS OF CURRENT PRACTICE

Numerical modeling of energy systems emerged in the 1960s and was well established by the end of the 1970s. Of particular relevance for the present discussion are models based on economic and/or optimization principles, designed to analyze policies including energy or environmental taxes, technology subsidies, or technology standards. A key example is the model created by the Federal Energy Administration for Project Independence in 1974 to support the development of a national energy policy (Hogan 1975). This model (the Project Independence Evaluation System) was of the LP type (like its successor the IPM).

As such models proliferated—reflecting both that era's focus on energy issues and the advent of more widely accessible and affordable computer hardware and software—model validation and uncertainty quantification were recognized as critical issues and were given considerable attention by both researchers and decision makers. More than 100 entries on energy and electric power models are listed in a 1978 bibliography on validating computer models in policy analysis and the social sciences (Gruhl and Gruhl 1978). In 1979 and 1980, the National Bureau of Standards held several large, multidisciplinary workshops on validation of energy models (Gass 1980).

It is fair to say that such efforts raised more questions than they answered, which to a large extent reflects the difficulty of the underlying problems. Subsequently, activities of this type were attenuated, particularly with the end of the era of energy crises, as of the mid-1980s. The modeling itself, however, continued and over time expanded. Energy models have become the dominant analytical tools not only for energy policy but for important areas of environmental policy as well, particularly GHG emissions abatement, as exemplified by the use of the IPM. The hallmark of this field is the creation of extremely detailed deterministic models, parameterized through nonstatistical procedures and not formally evaluated or tested as this concept is commonly understood in scientific and engineering circles.

A key factor in the emergence of this paradigm is the rise of calibration rather than estimation as a standard modeling practice. This terminology refers to the following distinction. In econometrics or statistics, unknown model parameters are estimated—that is, assigned numerical values—by applying such methods as least squares or maximum likelihood to fit the model to data. This approach also quantifies uncertainty intrinsically, as it were, by providing various well-defined error estimates such as confidence intervals. By contrast, calibration refers to assigning parameter values in a nonstatistical manner to models that are constructed a priori on assumptions based on first principles.[4] The calibrationist philosophy is discussed by Dawkins, Srinivasan, and Whalley (2001); although focused on so-called applied (or computable) general equilibrium models in economics, their characterization well applies to energy models of the economic or optimization type more generally:[5] "[M]odellers typically see their simulations largely as numerical implementations of theoretical structures. To them, the widespread use of a particular structure in the theoretical literature is an indication of its worth, so that they seek less to test or validate models and more to explore the numerical implications of a particular model, conditional on having chosen it. . . . [T]he focus of micro modellers is to generate insights about the effects of policy or other changes conditional on a particular theoretical structure, rather than to test theory itself" (Dawkins, Srinivasan, and Whalley 2001, p. 3672).

This conditionality is not confined to research applications, however; it also describes how such models are applied to policy analysis. For such applications, the standard experimental design, as it were, is to compute a reference or baseline case representing the future trajectory of the system in the absence of the policy in question and then a policy case that includes the intervention. The interpretation of the results of such calculations of course depends significantly on how one interprets the reference or baseline case. Modelers are at some pains to deny that such cases constitute forecasts or predictions, instead characterizing them as projections and emphasizing their conditionality. A representative statement of this stance is provided by the Energy Information Administration (EIA), which maintains and runs the National Energy Modeling System (NEMS), which is in effect the Department of Energy's, and thus the federal government's, de facto official model of the national energy system: "Projections [generated by NEMS] are not statements of what will happen but of what might happen, given the assumptions used for any particular scenario. . . . [E]nergy models are simplified representations of [the energy system]. Projections are highly dependent on the data, methodologies, model structures, and assumptions used in their development.

4. It should be noted that the meaning of model calibration varies across disciplines, especially between economics and policy analysis, on the one hand, and physical and engineering sciences, on the other hand. The characterization given here is specific to the former.

5. It is also the case that some energy models are of the numerical general equilibrium type.

[These] projections are subject to much uncertainty. Many . . . events that shape energy markets are random and cannot be anticipated. In addition, future developments in technologies, demographics, and resources cannot be foreseen with certainty" (EIA 2010).

This statement is a justified and informative disclaimer. But the acknowledgment of these myriad uncertainties belies the practice of positing a single baseline or reference case, or at most a very small number of the numerous possibilities. Clarke et al. (2007, p. 59) acknowledge, in a multimodel study of global GHG stabilization paths, that "[e]ach of the modeling groups could have created a range of other plausible reference scenarios by varying assumptions about rates of economic growth, the cost and availability of alternative energy options, assumptions about non-climate environmental regulations, and so forth."

However, calculation of this range of plausible alternatives is extremely rare.[6] As Morgan and Keith (2008) point out, it is difficult to avoid the interpretation that modelers are implicitly endorsing their reference cases as having greater plausibility or likelihood than possible alternatives. Yet there are no clearly articulated concepts of uncertainty or validity that might formally support this approach.

Instead, the notion that the models are intended to generate insights, not numbers, is a cornerstone of contemporary energy modeling epistemology (Peace and Weyant 2008). Defining what constitutes an insight, however, is problematic. A commonly proffered example, of the cost-lowering benefits of geographic and temporal flexibility in abating GHG emissions, may on the contrary simply reflect the mathematical structure of the models in question: enlarging the set of feasible solutions of a concave optimization model necessarily results in improvements in the optimum.

In fact, the history of energy modeling can instead be interpreted as revealing that ever-increasing levels of detail per se are implicitly taken as an indicator of increasing verisimilitude or accuracy, despite there being few if any theoretical or empirical foundations for concluding that this is in general an outcome of greater model complexity. Prima facie, it might seem obvious that greater detail would indeed yield better accuracy. It is important to note, however, that the modelers themselves do

6. Lempert et al. (2006) develop an approach to integrated assessment of global climate change based on the computation of combinatorially large numbers of scenarios and an analytic framework to interpret them. Their methodology stands in contrast to that of almost all other energy and integrated assessment modelers.

not make this claim as such—the appeal to modeling for insight is in fact a disclaimer of sorts in this regard. This stance is consistent with the calibrationist position. But one of its consequences is that the modeling community has not developed concepts of accuracy or validity that could be applied to gauge whether or how an increase in detail improves a model.

It is useful to consider in concrete terms the example of modeling the electric power system, specifically the inclusion of information about thousands of individual power plants in the IPM. Asking whether and why the model is more accurate with this level of detail than it would be without it allows us to explore the question absent a standard or generally accepted definition. To begin, including a specific plant means representing it in terms of a small number of parameters such as output capacity, thermal efficiency, and cost. The actual power output of the plant, however, is determined in a model such as the IPM by a global optimization of the entire system: the model operates each plant in such a way as to minimize the total system cost of meeting exogenously specified demands. Thus, in contrast to what might be the case with a simulation model, the modeled output of the plant would not be expected to correspond to its actual output even in a base year for which this quantity may be known—for example, from survey data. In this respect, then, the model does not represent the plant accurately, nor is it designed to do so.

These considerations highlight the fact that computation using a model such as the IPM is normative, not positive, and thus focus our attention on what "accuracy" should mean in a normative context and how a high level of detail contributes to its achievement or does not. (The optimization assumption can be rationalized positively by appealing to mathematical demonstrations of the equivalence of centralized optimization and distributed rational behavior under ideal market conditions, but this argument is rarely if ever made for electric power modeling, because of the general deviation from those conditions.) Given that the basic use of the model is determining the least-cost means of operating the system, one might argue that including information about a large number of individual plants gives a better estimate of this cost than would be obtained with a less detailed model. This simply shifts the question, however, since it is not possible to run the real-world system optimally, in a manner corresponding to the model's representation, in order to study how well the outcome is replicated by computational models of varying levels of complexity.

It could be argued instead that the primary practical application of the model is to estimate how the cost of operating the system would be affected by potential policy interventions—such as the GHG regulations—and that this purpose is better served with a higher level of detail. That is, the point is to estimate the change, not the actual level. To examine this reasoning, consider the following thought experiment. Suppose that the model is static—that is, it represents only the present-day system rather than the evolution of the system over the next several decades—and is being used to analyze some hypothetical policy. If, as expected, the modeled (determined by the optimization) power outputs of individual plants do not correspond to their actual outputs in the base or prepolicy case, why would the presence of plant-level detail increase confidence that this detail improves the modeled response of the system to policy relative to a model with less detail?

In reality, the IPM and other models have an intertemporal structure. In this setting, the fact that the overall computation represents a scenario rather than a forecast or prediction is a reminder of the acknowledged conditionality of the exercise. Here, accuracy and validity are not merely undefined, they are explicitly disavowed. So the question of what is improved by a greater level of detail in a model becomes even more vexing. Accepting the epistemological principle of modeling for insight, one could conjecture that greater detail yields greater insight, but as noted above, this term itself is challenging to define.

The intertemporal structure of models such as the IPM highlights another reason to question the value of increased complexity: the importance of technological change in determining the future evolution of the energy system. Technological change in general has proven difficult or impossible to predict, but over sufficiently long timescales it is one of the primary drivers of energy use. In aggregate models, its effects on a large scale can at least be approximated by exogenous parameters that extrapolate past trends. In energy modeling, these are analogous to labor productivity parameters. Just as the latter capture the trend of increasing economic output per unit of labor input over time without representing the actual mechanisms underlying this phenomenon, so aggregate energy productivity parameters represent, in reduced form, the historically observed trend of increasing energy service output (from the energy industry) per unit of fuel input over time. Although it implicitly treats technical innovation as a black box, this approach has the important advantage of indirectly incorporating the effects of unanticipated future technical progress without requiring representation of the details. The modeler must make an assumption about the rate at which it will occur but not exactly how it will come about.

But this technique does not apply readily, if at all, when individual technology units such as power plants are explicitly represented. In an intertemporal model with individually represented existing and planned electric power plants (for example), the state of electricity generation technology is directly embodied in these units. The model thus implicitly locks technology into a combination of its current and anticipated states. In other words, it precludes unpredictable innovation. For this reason, if it is assumed that technological progress will continue to occur during the next several decades, a high degree of technology disaggregation makes the model less plausible over this planning horizon than might be the case with less detail. The problem of representing technological change in energy models is discussed further in Section 6.

A complementary issue associated with increasing complexity of the model is the possibility that increasing detail per se increases the uncertainty embedded in a model's output. This risk is suggested by analogy to the trade-off between bias and variance in statistics, econometrics, machine learning, and related disciplines. This term refers to the fact that, while accuracy with which a modeled quantity is represented can be improved by increasing a model's dimensionality or number of parameters, beyond a certain point this will necessarily result in greater uncertainty in the representation as well. In empirical modeling—that is, when statistical procedures are used to estimate parameters—this phenomenon can be explicitly quantified and analyzed. By contrast, the widespread reliance on the deterministic approach and on calibration (rather than estimation) in energy modeling, along with improvements in computational hardware and software, has allowed models' dimensionality to increase without formal empirical or theoretical analysis of the degree to which uncertainty may be simultaneously increasing.

6. DETAIL AND ACCURACY IN THE INTEGRATED PLANNING MODEL

It was noted above that the IPM has been and continues to be subject to reviews and other procedures that reflect the EPA's philosophy of confidence or credibility building as a key aspect of evaluating models. In addition, the high level of electric power system detail in the IPM is emphasized in the model's documentation and in reports on regulatory analyses in which it has been applied. The model illustrates the phenomenon noted in Section 5 of detail per se being claimed as evidence of validity or verisimilitude. The following example allows me to explore this hypothesis.

Although never enacted, the Clear Skies legislation, introduced by the Bush administration in 2002, would have amended the CAA to establish emissions cap-and-trade systems to reduce sulfur dioxide, nitrogen oxide, and mercury emissions from electric power plants. Extensive analyses of successive versions of the proposed legislation were among the first applications of the IPM. The 2002 analysis used IPM version 2.1, for which (as with successive versions of the model) extensive, multipart documentation was prepared (EPA Clean Air Markets Division 2002).

Model validation and evaluation are mentioned nowhere in this documentation. However, an ancillary document accompanying the IPM output files for the legislative analysis contains the following statement: "Projections for individual plants are based on data currently available and modeling parameters which are simplifications of the real world. It is likely that future actions regarding individual plants will differ from model projections of actions; however, the aggregate impacts are expected to be appropriately characterized by the model" (EPA 2002, p. 1).

This is a highly suggestive characterization of the model's accuracy but is apparently nowhere elaborated on by ICF International or the EPA. Precisely which impacts are at issue, and what "appropriately characterized" might mean, are left unspecified. For example, are the impacts simply the future trajectories of the national electricity supply, including the fuel mix? Or are they the response of the system to policy interventions? Or both? And is it being stated that the model predicts such impacts with some sort of empirical accuracy? More important, the statement raises the question of how the modelers and the EPA view the relationship between detail and accuracy in the IPM. One interpretation is that it is being claimed that the model's appropriate characterization of aggregate impacts is due to the high degree of detail about individual plants and that the model will be right on average in a statistical sense. But given that the model is completely deterministic and is used to generate only a small number of scenarios, the grounds for such a claim are unclear.

In terms of the discussion in Section 5, the implicit claim here may in fact be that the high level of detail increases the accuracy of the model in a deterministic rather than statistical sense. The justification for such an assertion is if anything even less clear than if it is interpreted statistically.

In energy modeling in general, it is in certain cases possible to compare past model-based projections with actual events. For example the EIA regularly releases retrospective assessments of NEMS projections that include discussions of reasons for discrepancies with subsequently observed phenomena. It is possible to make such a comparison for the IPM using the Clear Skies study of 2002, and this comparison highlights the problem of assuming that high levels of detail ensure model verisimilitude.

The IPM base case for the Clear Skies 2002 analysis projected that electric power CO_2 emissions would be 2,317 million metric tons (mmt) in 2005 and 2,429 mmt in 2010 (EPA Clean Air Markets Division 2002, table 9.3). The EIA reports that actual emissions were 2,454 mmt in 2005 and 2,313 mmt in 2010, which implies forecast errors of roughly 6 percent in 2005 and 5 percent in 2010 (EIA 2012, table 11.2e). These are quite reasonable errors for the forecast intervals involved and might inspire confidence in the model's predictive capability. However, the IPM projection for 2012 CO_2 emissions is 2,472 mmt,[7] while actual emissions in that year were 2,048 mmt (EIA 2013b, table 12.6), an error of more than 20 percent.

With the exception of the electric power CO_2 emissions reduction due to the general economic contraction following the financial crisis of 2008, this is the largest decrease (that is, 2011 to 2012) in these emissions since the EIA began collecting the data in 1973. The basic reason for this change was the rapid and significant shift from coal to natural gas in the electric power sector, due essentially to declines in gas prices resulting from technological advances in drilling—specifically, hydraulic fracturing, or fracking. While the magnitude of ultimately recoverable domestic natural gas supplies remains uncertain, there are few if any grounds for projecting a supply contraction and consequent price increase sufficient to reverse this fuel shift. In other words, a technology regime change has occurred. In retrospect, the version of the IPM that produced these projections (which went to 2020), had it been used to analyze CO_2 policy, would have yielded results based on large and systematic biases in the projections starting around halfway to the forecast horizon. It is worth noting that the issue here is indeed bias and not high variance; that is, the model was not in some sense right on average.

This example can be taken as one of numerous cautionary tales regarding technological change and energy modeling. As Parker and Yacobucci (2008, pp. 73, 15) emphasize in a review of model-based analyses

7. The Integrated Planning Model projections are reported in 5-year time steps; the 2012 figure is an interpolation.

of prospective national climate policy (that is, large-scale GHG emissions abatement): "[U]nforeseen events (such as technological breakthroughs) loom as critical issues which cannot be modeled. . . . [L]ong-term [model-generated] cost projections are at best speculative, and should be viewed with attentive skepticism." Furthermore, "The uncertainty about the future direction of the basic drivers of greenhouse gas emissions and the economy's responsiveness (economically, technologically, and behaviorally) illustrate the inability of models to predict the ultimate macroeconomic costs of reducing greenhouse gases."

These points highlight the fundamental weakness of the idea that high levels of detail promote some form of model validity, particularly in projecting future paths of the energy system. The complexity of version 2.1 of the IPM did not enable it to project the single biggest shift in electric power CO_2 emissions in the past 40 years, nor will it enable the current version to project any future such shifts. As argued in Section 5, the complexity may in fact decrease the capacity of the model to capture, even in aggregate terms, the overall trajectory of technical change in electric power generation relative to models using a very reduced-form representation of technical change. This is contrary to the assertion by the EPA quoted above regarding the IPM's capability to characterize aggregate impacts. There is a basic illusion-of-precision problem embedded in the modeling philosophy represented by the IPM.

7. A MODEL UNCERTAINTY PERSPECTIVE

The discussion thus far has focused on single models. However, a broader perspective is needed to fully capture the basic uncertainties attendant to energy modeling and can be motivated by the results of a multimodel study of the potential costs to the US economy of implementing the Kyoto Protocol. This study was conducted by Stanford University's Energy Modeling Forum (EMF), the leading center for structured scenario analyses using energy and integrated assessment models, including intercomparisons of multiple-model results. The study, conducted in 1999, used 11 models analyzing the protocol, including its impact on the United States over a 3-decade horizon (to 2020) (Weyant and Hill 1999).

Figure 1 shows the model-by-model marginal cost curves for US compliance with the treaty, in terms of the carbon tax required to meet a range of percentage of emissions reductions from each model's baseline

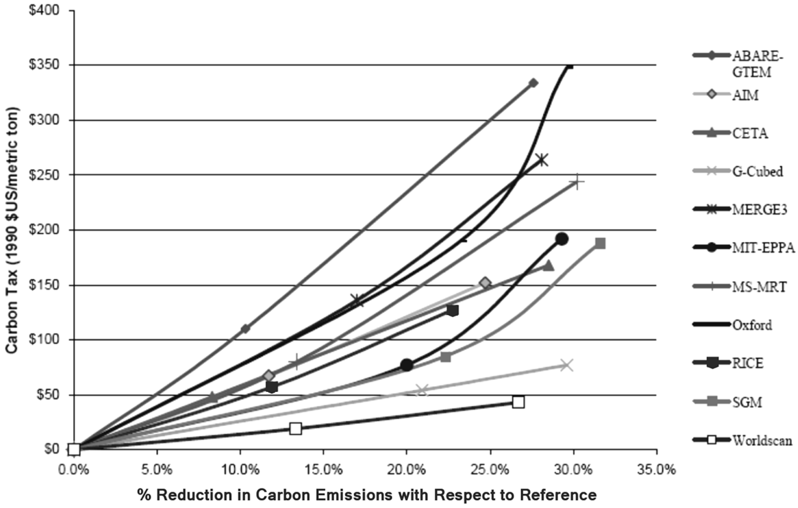

Figure 1. Model predictions of marginal abatement costs for US compliance with the Kyoto Protocol.

(Fischer and Morgenstern 2006).[8] The basic result is that estimated costs for full compliance—an approximately 30 percent reduction—vary by a factor of five among the models used in the study. This prompted Fischer and Morgenstern (2006) to observe that such a degree of uncertainty was a substantial impediment to policy makers' willingness to implement policy, and the researchers attempted to account for the sources of variation in a statistical analysis.

It is critical to note that this uncertainty was manifested after nearly 3 decades of computational model development, evolution, and application in the energy field. There are few indications that there has been a reduction in this type of uncertainty in the years since this study was conducted. This and similar results in many other studies are evidence of fundamental model uncertainty, under which the appropriate underlying mathematical description of a system cannot be specified uniquely so that more than one such description can be justified, quantified, and used as the basis for computational modeling.

Model uncertainty constitutes a basic indeterminacy in the mathematical and computational representation of systems. It therefore defines lim-

8. Figure 1 (Fischer and Morgenstern [2006], figure 1) shows the abatement percentages and marginal costs derived from four regions and 11 models for two policy scenarios. See Fischer and Morgenstern (2006) for details.

its to validation, as this term has been conventionally understood. If the state of knowledge in some domain allows for more than one equally plausible or defensible model of a phenomenon, then the concept of validation cannot entail a criterion, even implicitly, that a validated model is the best representation among candidate models, as is the case, for example, in econometric or statistical analysis. But this type of uncertainty also has profound implications for more general concepts of model evaluation, including the characterizations in the NRC and EPA documents discussed previously. If a given model is found, to paraphrase, to be credible and useful, but other models of the same phenomenon or system also meet these criteria, how should the relative validity, verisimilitude, or usefulness of any one of the models in decision making be defined and measured?

This problem is directly analogous to the issue of ambiguity in decision theory. This term refers to situations in which a decision maker is uncertain about which of several probability distributions or models correctly describes a given process, system, or set of potential choice outcomes. In the energy modeling context, the models are with few exceptions deterministic. But in very practical terms, energy and environmental policy makers and regulators are routinely dealing with multiple, coexisting models of the energy system that they are essentially compelled to consider as equally plausible but that produce varying decision-relevant outputs. The modeling and analytical communities have provided no formal or quantitative means of weighing, ranking, or otherwise combining such information nor more generally a theoretical framework or practical procedures for dealing with multiple, equally plausible models.

In the case of the EPA and the IPM, a single model has been established as the analytical platform. The evaluation process, such as it is, addresses what can be considered elements of internal validity—that is, internal consistency, input quality and extent, whether the model correctly implements its underlying assumptions, and so forth. But as previously noted, there is no expectation or requirement that the model meet some objective standard of verisimilitude—in colloquial terms, that it be demonstrated to be right even conditional on its assumptions and inputs. Moreover, as Fischer and Morgenstern (2006) and other such studies show, it is a given that the same underlying information—such as electric power plant detail—that drives the IPM would yield different but equally justifiable policy-relevant outputs under different but equally plausible modeling assumptions. Thus, by what amounts to institutional and procedural design, the model and its regulatory applications also avoid dealing with the fundamental model uncertainty that attends computational representations of the energy system.

8. CONCLUSION

A literal reading of the NRC evaluation guidelines, and their implementation by the EPA, is that they address the utility of models, their credibility in the eyes of regulators and stakeholders, and processes to establish both, but not validity or verisimilitude as such. In fact, the guidelines are based on an epistemology in which validation as it has traditionally been understood in physical science and engineering computation is a categorically inappropriate concept for complex numerical models. The IPM fully reflects this paradigm, which as I have discussed is firmly established in computational energy modeling more generally. But here as in other cases, the model's construction, development, and documentation strongly indicate that the modelers' and agency's view is that a high degree of detail per se is indicative of verisimilitude, notwithstanding the absence of theoretical or empirical grounds for this belief.

The stance of the NRC and now of the EPA is essentially that computational models and their use in regulatory analysis should be viewed from a decision-theoretic perspective. This is an entirely appropriate way of framing the matter. But neither entity has articulated its position this way, nor has it recognized or acknowledged that reliance on a single, extremely complex deterministic model is at odds with this philosophy. It is quite clear that the IPM is primarily being used to generate numbers, not insights, much less results that could be used in a formally constructed decision-theoretic analysis.

A useful and pragmatic step forward would be to rigorously state and analyze the claims made by the EPA, quoted in Section 6, regarding the relationship between the IPM's high level of detail on the one hand and its capability for characterizing aggregate impacts on the other hand. Such an effort could entail quantitative testing of how plant-level detail affects the model's aggregate response to various policy interventions, especially with respect to system cost. In principle, a comparison could be made to a form of the model with more aggregated technology representation. Some type of analysis along these lines is necessary to justify claims regarding the accuracy or usefulness of the model that are based, even implicitly, on its level of complexity.

More generally, model uncertainty as described in Section 6 is a form of deep uncertainty. Groundbreaking work by macroeconomists has analyzed model uncertainty and decision or policy rules for the modeled system that are robust against it (West, Durlauf, and Brock 2003; Brock, Durlauf, and West 2007; Hansen and Sargent 2007). In the case of CO_2 emissions abatement, an approach built on the recognition of fundamental model uncertainty would be a large step toward regulatory modeling that reflects the general state of knowledge—and ignorance—that underlies our capacity to project the future of the energy system and how it might respond to policies. At a minimum, such a paradigm would depart from the standard regulatory agency practice—of which the EPA's use of the IPM is but one example—of reliance on single, highly complex models. Developing and implementing such an approach would require the creation of methods suitable for deterministic, high-dimensional energy modeling; the optimal control-based techniques of Hansen and Sargent (2007) are not directly applicable. Undertaking such an initiative, however, is not solely a technical modeling problem per se. It would require in addition a framework for designing, not just analyzing, regulations that take account of the fundamental uncertainties. This is a daunting but critical—and ultimately unavoidable—challenge for both researchers and regulators.

REFERENCES

Brock, William A., Steven N. Durlauf, and Kenneth D. West. 2007. Model Uncertainty and Policy Evaluation: Some Theory and Empirics. *Journal of Econometrics* 136:629–64.

Clarke, Leon E., James A. Edmonds, Henry D. Jacoby, Hugh M. Pitcher, John M. Reilly, and Richard G. Richels. 2007. *Scenarios of Greenhouse Gas Emissions and Atmospheric Concentrations*. Synthesis and Assessment Product 2.1a by the Climate Change Science Program and the Subcommittee on Global Change Research. Washington, DC: Department of Energy.

Dawkins, Christina, T. N. Srinivasan, and John Whalley. 2001. Calibration. Pp. 5:3653–3703 in *Handbook of Econometrics,* edited by James J. Heckman and Edward E. Leamer. Amsterdam: North-Holland.

EIA (Energy Information Administration). 2010. Annual Energy Outlook 2010—with Projections to 2035. DOE/EIA-0383(2010). Washington, DC: EIA.

———. 2012. *Annual Energy Review 2011*. Report No. DOE/EIA-0384(2011). Washington, DC: EIA.

———. 2013a. *Electric Power Annual 2011*. Washington, DC: EIA.

———. 2013b. *Monthly Energy Review*. Report No. DOE/EIA-0035(2013/03). Washington, DC: EIA.

EPA (Environmental Protection Agency). 2002. *Parsed Results Using EPA Modeling Applications (v.2.1) of the Integrated Planning Model.* Washington, DC: EPA.

———. 2012. *Integrated Planning Model—EPA Applications*. Washington, DC: EPA.

EPA (Environmental Protection Agency) Clean Air Markets Division. 2002. *Documentation of EPA Modeling Applications (V.2.1) Using the Integrated Planning Model*. Report No. EPA 430/R-02-004. Washington, DC: EPA.

EPA (Environmental Protection Agency) Council for Regulatory Environmental Modeling Office of the Science Advisor. 2009. *Guidance on the Development, Evaluation, and Application of Environmental Models*. Document No. EPA/100/K-09/003. Washington, DC: EPA.

EPA (Environmental Protection Agency) Office of Air Quality Planning and Standards Health and Environmental Impacts Division. 2012. *Regulatory Impact Analysis for the Proposed Standards of Performance for Greenhouse Gas Emissions for New Stationary Sources: Electric Utility Generating Units*. Document No. EPA-452/R-12-001. Washington, DC: EPA.

———. 2014. *Regulatory Impact Analysis for the Proposed Carbon Pollution Guidelines for Existing Power Plants and Emissions Standards for Modified and Reconstructed Power Plants*. Document No. EPA-452/R-14-002. Washington, DC: EPA.

EPA (Environmental Protection Agency) Science Policy Council and Model Acceptance Criteria and Peer Review White Paper Working Group. 1999. *White Paper on the Nature and Scope of Issues on Adoption of Model Use Acceptability Guidance*. Washington, DC: EPA.

Fischer, Carolyn, and Richard D. Morgenstern. 2006. Carbon Abatement Costs: Why the Wide Range of Estimates? *Energy Journal* 27(2):73–86.

Gass, Saul I., ed. 1980. *Validation and Assessment Issues of Energy Models: Proceedings of a Workshop Held at the National Bureau of Standards, Gaithersburg, Maryland, January 10–11, 1979*. National Bureau of Standards Special Publication No. 569. Washington, DC: US Department of Commerce.

Gruhl, J., and N. Gruhl. 1978. Methods and Examples of Model Validation: An Annotated Bibliography. Working Paper MIT-EL 78-022WP. Massachusetts Institute of Technology Energy Laboratory, Cambridge, MA.

Hansen, Lars Peter, and Thomas J. Sargent. 2007. *Robustness*. Princeton, NJ: Princeton University Press.

Hogan, William W. 1975. Energy Policy Models for Project Independence. *Computers and Operations Research* 2:251–71.

Lempert, Robert J., David G. Groves, Steven W. Popper, and Steve C. Bankes. 2006. A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science* 52:514–28.

Morgan, M. Granger, and David W. Keith. 2008. Improving the Way We Think about Projecting Future Energy Use and Emissions of Carbon Dioxide. *Climatic Change* 90:189–315.

NRC (National Research Council) Committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification; Board on Mathematical Sciences and Their Applications; and Division on Engineering and Physical Sciences. 2012. *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. Washington, DC: National Academies Press.

NRC (National Research Council) Committee on Models in the Regulatory Decision Process. 2007. *Models in Environmental Regulatory Decision Making*. Washington, DC: National Academies Press.

Oberkampf, William L., and Christopher J. Roy. 2010. *Verification and Validation in Scientific Computing*. Cambridge: Cambridge University Press.

Office of Management and Budget. 2003. *Regulatory Analysis*. Circular A-4. Washington, DC: Office of Management and Budget.

Oreskes, Naomi. 1998. Evaluation (Not Validation) of Quantitative Models. *Environmental Health Perspectives* 106(supp. 6):1453–60.

Oreskes, Naomi, Kristin Shrader-Frechette, and Kenneth Belitz. 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, February 4, pp. 641–46.

Parker, Larry, and Brent Yacobucci. 2008. *Climate Change: Costs and Benefits of S. 2191*. CRS Report No. RL34489. Washington, DC: Congressional Research Service.

Peace, Janet, and John Weyant. 2008. Insights Not Numbers: The Appropriate Use of Economic Models. White paper. Pew Center on Global Climate Change, Washington, DC.

SAB (Science Advisory Board) Environmental Engineering Committee. 1989. *Resolution on the Use of Mathematical Models by EPA for Regulatory Assessment and Decision-Making*. Report No. SAB-EEC-89-012. Washington, DC: Science Advisory Board.

SAB (Science Advisory Board) Modeling Peer Review Subcommittee and Environmental Engineering Committee. 1993. *Review of Draft Agency Guidance for Conducting External Peer Review of Environmental Modeling*. Report No. EPA-SAB-EEC-LTR-93-008. Washington, DC: Science Advisory Board.

West, Kenneth D., Steven N. Durlauf, and William A. Brock. 2003. Policy Evaluation in Uncertain Economic Environments. *Brookings Papers on Economic Activity*, pp. 235–322.

Weyant, John P., and Jennifer N. Hill. 1999. Introduction and Overview. *Energy Journal* (Special Issue: Costs of the Kyoto Protocol: A Multi-model Evaluation) 20:vii–xliv.