



Article

Validity and Validation of Computer Simulations—A Methodological Inquiry with Application to Integrated Assessment Models

Alan Randall ^{1,*} and Jonathan Ogland-Hand ²

¹ Sustainability Institute and Department of Agricultural, Environmental & Development Economics, The Ohio State University, Columbus, OH 43210, USA

² Carbon Solutions LLC, Saint Paul, MN 55105, USA

* Correspondence: randall.3@osu.edu

Abstract: Our purpose is to advance a reasoned perspective on the scientific validity of computer simulation, using an example—integrated assessment modeling of climate change and its projected impacts—that is itself of great and urgent interest to policy in the real world. The spirited and continuing debate on the scientific status of integrated assessment models (IAMs) of global climate change has been conducted mostly among climate change modelers and users seeking guidance for climate policy. However, it raises a number and variety of issues that have been addressed, with various degrees of success, in other literature. The literature on methodology of simulation was mostly skeptical at the outset but has become more nuanced, casting light on some key issues relating to the validity and evidentiary standing of climate change IAMs (CC-IAMs). We argue that the goal of validation is credence, i.e., confidence or justified belief in model projections, and that validation is a matter of degree: (perfect) validity is best viewed as aspirational and, other things equal, it makes sense to seek more rather than less validation. We offer several conclusions. The literature on computer simulation has become less skeptical and more inclined to recognize that simulations are capable of providing evidence, albeit a different kind of evidence than, say, observation and experiments. CC-IAMs model an enormously complex system of systems and must respond to several challenges that include building more transparent models and addressing deep uncertainty credibly. Drawing on the contributions of philosophers of science and introspective practitioners, we offer guidance for enhancing the credibility of CC-IAMs and computer simulation more generally.

Keywords: methodology; deep uncertainty; computer simulation; integrated assessment modeling; climate change integrated assessment models



Citation: Randall, A.; Ogland-Hand, J. Validity and Validation of Computer Simulations—A Methodological Inquiry with Application to Integrated Assessment Models. *Knowledge* **2023**, *3*, 262–276.

<https://doi.org/10.3390/knowledge3020018>

Received: 22 February 2023

Revised: 19 April 2023

Accepted: 17 May 2023

Published: 22 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The debate on the scientific status of integrated assessment models (IAMs) of global climate change is spirited. For example, the *Review of Environmental Economics and Policy* 2017 symposium on IAMs focusing on climate change [1–3] may leave the reader at a loss as to what should be believed. Metcalf and Stock [1] argue that complicated IAMs, while in need of continuing improvement, are essential to informed policy making concerning climate change; Pindyck [2] sees CC-IAMs as crucially flawed, fundamentally misleading, and in essence mere rhetorical devices; and [3] Weyant sees value in CC-IAMs especially for “if . . . , then . . . ” analysis to explore the implications of alternative model structures, parameterizations, and driver settings (see also [4]).

To this point, the debate has been conducted mostly among climate change modelers and users seeking guidance for climate policy. However, there is a considerable literature on the methodology of computer simulation—where methodology is used in its original meaning: the study of research methods and their associated background assumptions—engaging philosophers of science as well as introspective simulation modelers and users of

model output. With this as background, we ask whether and how computer simulations contribute to knowledge, how well IAMs serve as methods of knowing, and how they might be improved from that perspective. Here we address what we consider the key issues in methodology of CC-IAM, drawing judiciously from the simulation modeling and methodology literature. We begin by recognizing that the complexity of integrated assessment modeling calls for a reconsideration of the idea of validity, and we suggest a plausible and workable concept of validity, credence. The core of our argument comes in two parts: Sections 2 and 3, which elaborate on the uncertainties encountered in CC-IAM, for example, and strategies for addressing deep uncertainty; and Sections 4 and 5, which engage with the methodological literature on simulation and its implications for enhancing credence. Section 6.1 provides conclusions regarding integrated assessment modeling, and Section 6.2 addresses conclusions about the prospects for validation of computer simulations.

In the end, we endorse the optimistic view—that CC-IAMs are potentially key contributors to informed climate policy—and conclude that there is scope for improving the validation of CC-IAMs, the transparency of these models, and the way in which the whole modeling exercise and its real-world implications are communicated. In a very uncertain world, improved modeling and communication of uncertainty is a key component of a comprehensive strategy to increase the validity of CC-IAM.

1.1. Validity, Confidence, and Credence

CC-IAMs can be enormously complex but their complexity is dwarfed by the complexity of the real world. Because abstraction is essential to building tractable models, validity is not so much about realistic depiction of the system at hand: it makes more sense when applied to the propositions (e.g., projections) yielded by such models. The key question is: under what conditions might a human evaluator justifiably have confidence in the accuracy, reliability, etc., of propositions emerging from IAM simulations? This is a question of credence (noun, of people and their state of mind): mental conviction of the truth of a proposition or reality of a phenomenon; justified belief. A CC-IAM is valid to the extent that belief in the propositions it generates is justified.

Several implications follow. Rather than a single decisive test (e.g. of logic for a deductive model, or of empirical outcomes against a prior null hypothesis), the process of judging credence often involves weighing evidence of various kinds and qualities. Credence is in practice an ordinal, rather than absolute, concept. While a model that merits absolute credence is aspirational but unattainable, the practicable aspiration should be to build models that merit more rather than less credence. The ordinality of credence implies that validation is not binary (valid/not), but a matter of degree. A model may be more or less valid and, other things equal, it makes sense to seek more rather than less validity. In the case of CC-IAMs, the direct objects of justified belief are the projections obtained using the model. Projections from IAMs are conditional on model structure and parameter values, but also on the settings of drivers that include exogenous influences on the system and policies that might be applied to the system. IAMs often are constructed for the explicit purpose of exploring the impacts of alternative driver settings. It follows that credence in IAMs is inherently conditional: justified belief that the suite of model projections is conditionally valid.

Given that validity is conditional, at least three kinds of conditions are relevant: (i) the epistemological integrity of the model, the validity of its parameter values, and its faithful representation in programming and computation meet reasonable expectations given the state of the art and the knowledge of the day; (ii) the suite of driver settings examined represents adequately the exogenous influences likely to be encountered and the policies likely to be considered, insofar as they can be foreseen; and (iii) the inevitable uncertainties in modeled relationships, parameter values, the values of exogenous drivers, and future policy settings have been considered carefully and addressed to the extent feasible given the state of the art and the knowledge of the day.

In a highly uncertain world, the credence framework explicitly incorporates the treatment of uncertainty: credence requires validation, and validation criteria include adequate treatment of uncertainty. We may be inclined to think of adequacy as judged in terms of some external benchmark(s), but some methodologists propose a “fit for purpose” criterion [5].

1.2. Challenges to Credence

Challenges to credence may arise from complexity, incomplete knowledge and weaknesses in modeling. The possibility of ordinary human error in modeling and parameterization suggests the need for *verification*: the process of making sure that the modelers’ intentions are implemented accurately, precisely, and completely. The likelihoods that the system under study is indeterminate and/or the modelers possess incomplete knowledge of the system suggest the need for *validation*. Furthermore there is always the possibility that the uncertainties impinging on the system have received inadequate consideration, often due to real difficulties in modeling cascading and interacting uncertainties.

The complexity and relative non-transparency of many CC-IAMs place a substantial burden on the modelers to provide adequate verification and validation—users of modeling output are at a considerable disadvantage in assessing model quality. Credence may also be impaired by implausible driver settings, and/or a suite of driver settings that fails to span the range of plausible future values of exogenous and policy drivers. Users are better equipped to identify these weaknesses because driver settings are more transparent than embedded model structure and parameterization, features that may also be influenced by modelers’ worldviews and habits of thought.

2. Chance and Uncertainty in IAM

2.1. The Distinction between Epistemic and Aleatory Uncertainty

Epistemic uncertainty. In a deterministic, non-chaotic system, there is by definition no role for chance, but there is the possibility of human ignorance. The perception of chance may arise from the incompleteness and imperfection of our knowledge—uncertainty is epistemic; we are unsure of how the system works. With no good model of the system, we may perceive arbitrariness or randomness in the data despite the determinism of the system that produced it. There are two kinds of epistemic uncertainty: structural and parametric. *Structural uncertainty* arises from imperfect mental models of the mechanisms involved. In IAMs, structural uncertainty may pertain to the complex interrelationships in the system under study (a concern arguably unique to complex systems modeling), and to matters familiar in other kinds of empirical/numerical work (e.g., functional forms of key relationships). As we learn more about the structure of the system, epistemic uncertainty is reduced. *Parametric uncertainty* in deterministic systems arises from our inadequate empirical knowledge to fully and accurately parameterize the system we are modeling. More knowledge and observations of the system tend to reduce this kind of epistemic uncertainty.

Aleatory uncertainty. In a well-understood but stochastic system, there is, by definition, no epistemic uncertainty. Uncertainty is entirely *aleatory*: we face chance because we are not prescient, e.g., despite knowing the relevant probability distribution(s), we cannot know the next draw. Ordinary risk analysis is designed to address exactly this kind of chance, i.e., aleatory uncertainty; and it is taught routinely via examples drawn from games of chance in which the system is well-understood and the odds can be calculated precisely.

A system may exhibit both kinds of uncertainty. If the system is buffeted by chance and not well understood, our statistical methods typically have difficulty isolating the contributions of epistemic and aleatory uncertainty to this unsatisfactory state of affairs. If the system is non-stationary, the drivers of regime shifts may have systematic properties but are likely also to be influenced by chance. Approaching such a system, there is no a priori reason to believe that the chance we encounter is entirely aleatory. Applying convenient stochastic specifications in this situation conflates more complex kinds of chance with

ordinary risk. The crucial assumption, seldom given the attention it deserves, is that the system is fully understood or (equivalently) that the game is fully specified. Frequentist statistical logic, being addressed to the interpretation of data about the occurrence or not of specific events as the outcome of a stochastic process, is entirely about aleatory uncertainty. Probability is, to a frequentist, the frequency of a particular outcome if the experiment is repeated many times. Because so many statistical applications are aimed at learning about parameters and so reducing epistemic uncertainty, it is common in frequentist practice that some (reducible) epistemic uncertainty is analyzed, purists would say inappropriately, as aleatory [6].

Statisticians have long understood this dilemma. Carnap distinguished probability¹ (credence, i.e., degree of belief) in contrast to probability² (chance, which is mind independent, objective, and defined in terms of frequency) [7]. Bayesian reasoning, being addressed to statements about the degree of belief in propositions, allows adjustment of probabilities in response to improved theories of how things work, better interpretations of empirical observations (e.g., better statistical models), and more observations. Decision theorists use probability to address our imperfect knowledge, as well as the indeterminism of the systems we study. Not surprisingly, many decision theorists are attracted to Bayesian approaches where less prominence is accorded to the distinction between aleatory and epistemic uncertainty. Probability is interpreted in terms of degree of belief in a proposition. For each proposition, there is a prior belief, perhaps well-informed by theory and/or previous observation but perhaps no more than a hunch. The prior belief is just the beginning; probabilities are adjusted repeatedly to reflect new evidence. The process by which we interpret what we are learning—especially whether we attribute patterns in the data to properties of the system or merely of the data at hand—typically combines pragmatism with more formal procedures.

2.2. Uncertainty Involves More Than Stochasticity

Uncertain circumstances include:

- Risk—in classical risk, the decision maker (DM) faces stochastic harm. The relevant pdf is known and stationary, but the outcome of the next draw is not. The uncertainty is all aleatory.
- Ambiguity—the relevant probability distribution function is not known. Ambiguity piles epistemic uncertainty on top of ordinary aleatory uncertainty.
- Deep uncertainty, gross ignorance, unawareness, etc.—the DM may not be able to enumerate possible outcomes, let alone assign probabilities. Inability to enumerate possible outcomes suggests a rather serious case of epistemic uncertainty, but aleatory uncertainty is likely also to be part of the picture.
- Surprises—in technical terms, the eventual outcome was not a member of the ex ante outcome set. The uncertainty that generates the possibility of a surprise is entirely epistemic—we failed to understand that the eventual outcome was possible. However, there likely are aleatory elements to its actual occurrence in a particular instance.

In IAM, we deal typically with complex systems. It follows that we would expect to encounter the above sources of epistemic and aleatory uncertainty, and two additional kinds of uncertainty: regime shifts and policy uncertainty. Regime shifts are imperfectly anticipated discrete changes in the systems under study. The uncertainty likely includes epistemic and aleatory components. The epistemic component includes failure to comprehend the properties of the particular complex system, but it likely also that aleatory uncertainty adds noise to the signals in the data that, properly interpreted, might warn of impending regime shifts. A policy is a suite of driver settings intended to achieve desired outcomes, and decentralized agents experience policy uncertainty as epistemic—the “policy generator” works in ways not fully understood—but perhaps also aleatory if there are random influences on driver settings. Incomplete transparency muddies the perception of uncertainty and its attribution to epistemic and aleatory causes.

All of the above kinds of uncertainty may exist in the real world that we are modeling and affect the performance of the system. There is recognition in the IAM literature that probabilities fail to represent uncertainty when ignorance is deep enough [8,9].

3. Getting Serious about Uncertainty in IAM

3.1. Uncertainty as a Challenge to Credence

Uncertainty in the future values of system drivers is a serious concern, but uncertainty within the real-world system itself is an even greater concern. If the task of the model is to make accurate projections about real-world outcomes in response to various driver settings, that task is much harder to accomplish when there is epistemic uncertainty about how the system works and real world outcomes are subject also to aleatory uncertainty. Validation also becomes much more difficult; for example, how do we interpret history matching if history is itself the outcome of uncertain processes such that the fact of a particular outcome does not mean it had to happen that way?

Uncertainty within the system is often epistemic, because (i) complex systems structures are often opaque, (ii) even if we knew the skeletal structure of the system, we likely would not know the functional form of the probability distributions for key variables (which implies that there is little we can say with confidence about the magnitude of worst-case outcomes), (iii) parameter uncertainty is endemic, and (iv) there is always the prospect of cascading uncertainties.

If we have not made every effort to ensure that the treatment of uncertainty in our models is conceptually correct and empirically well-informed, and that our characterization of the uncertainty that exists is plausible and consistent with the evidence, we are poorly placed to ask for credence in our projections. Pindyck takes these concerns seriously enough to suggest that formal modeling is unhelpful and perhaps misleading, and we would do better to simply sit with decision makers and discuss the issues and the attendant uncertainties frankly [2].

3.2. Scenario Analysis to Address Uncertainty

Scenario analysis frequently is offered by IAM modelers, to facilitate policy analysis by projecting system outcomes under a range of hypothetical policy settings. To address uncertainty in future values of exogenous drivers, several scenarios are constructed representing a range of driver settings, and each is modeled as deterministic. To explore the potential impacts of policy options, policy settings are chosen purposefully to create a range of scenarios that permits projections of system futures under alternative policies of ex ante interest. Scenario analysis also enables a crude response to uncertainty regarding model structure and parameters, by exploring the implications of a range of possible specifications and parameter values.

Example: The *Australian National Outlook* study is attentive to climate and climate policy in the broader context of Australian futures [10]. It used deterministic models with constructed scenarios around a set of global drivers considered exogenous to Australia (global economic demand, climate, and greenhouse gas abatement effort), and a set of Australian drivers (Australian resource efficiency, hours worked annually, proportion of “experience goods” in the consumption mix, agricultural productivity, and increases in land use for bioenergy production, conservation, carbon sequestration). Each of these drivers could be set at any of four (in a few cases, six) levels. The levels could result from different combinations of ground-up trends and purposeful policies—note that ground-up trends are adjusted by tweaking model parameters, while policy drivers are adjusted directly. With this many global and Australian drivers, and four to six settings of each, thousands of scenarios (combinations of driver settings) are mathematically possible. Given that some combinations of driver settings are internally inconsistent, the remaining menu includes hundreds of plausible scenarios. The modelers elaborated 20 scenarios, and then identified four that stake-out a plausible spectrum of options for Australia through 2050: resource-intensive, business as usual, a middle-way response to climate and conservation

concerns, and full-bore lean and green. For these four scenarios, projections are reported and discussed in detail. Fruition for any of these scenarios would require a complex interaction of purposeful Australian policies, ground-up trends in Australia, and global driver settings over which Australia would have relatively little influence.

We have labeled scenario analysis a crude response to uncertainty because the models are deterministic—which means we are comparing projections of alternative certainties—and the scenarios are relatively few in number and chosen in thoughtful but ultimately ad hoc fashion; notions of the relative likelihood of the various scenarios are informal and ad hoc; and validation of projected outcomes relies heavily on their intuitive plausibility. If scenario analysis is our only attempt to address uncertainty, it does not amount to an adequate response.

3.3. The Challenge of Better Capturing the Real-World Uncertainties within the Deterministic, Multiple Scenarios Framework

A more complete set of scenarios would help to characterize the uncertainties in the system, as would a more systematic sampling of the possibilities space. A systematic sampling of possible outcomes would require attention to model structure, key parameters, and policy drivers, all of which are susceptible to uncertainty.

Uncertainty is not something that strikes after we have committed ourselves to a course of action, shocking us out of our complacent assumption of a certain world. Rather, we live with uncertainty all of the time, and it influences our behavior in many ways, e.g., inducing purposeful strategies to manage uncertainty, but also more passive responses such as procrastination and indecisiveness [11,12]. It follows that the baseline conditions, the “hard data” component upon which projections of future outcomes are based, were generated by decision makers who faced uncertainty in real time. If future uncertainties are expected to be much like those in the recent past, experienced performance may serve as a good starting point for projecting future performance. However, it is possible, and in some cases expected, that future uncertainties might be quite different in kind and degree from those experienced recently.

It seems that the ambition to achieve a more realistic representation of uncertainties in CC-IAM calls also for a more realistic representation of how people behave under uncertainty, how they manage uncertainties and how they adapt to new uncertainties. Ideally, these representations should be fine-grained enough to distinguish the behavioral consequences of alternative driver settings.

3.4. Introducing Stochasticity in a Few Variables Thought Ex Ante to Be Sensitive

Recent IAM work has achieved some notable successes in introducing stochasticity in a few key variables thought ex ante to be influential on outcomes and susceptible to uncertainty. Cai and various colleagues consider uncertain technological growth and the existence of stochastic tipping points for global climate, regional water quality [13,14], or both by accounting for interactions [15]. Advances in computational methods to support these stochastic model specifications include a nonlinear certainty equivalent approximation method [16] and parallel dynamic programming [17].

It is important to give credit where credit is due: these approaches make significant advances in modeling, and do in fact capture some of the major, or perhaps most obvious, uncertainties in the systems under study. However, the converse is also true: all but the major uncertainties are ignored, or perhaps captured only in scenario analysis. Uncertainties that are not among the most obvious today may turn out to be crucial in the future; and modeling just a few key uncertainties precludes consideration of the full extent of uncertainty propagation, i.e., the combined effect of several uncertain variables on a function and, then, of several uncertain functions on a model. Furthermore, even when particular uncertainties are captured in the model, we capture only the impacts of those uncertainties on computed outcomes, thereby ignoring the possibility that these uncertainties may change human behavior and decision making.

3.5. How Might IAMs Be Restructured to Better Address the Range of Real-World Uncertainties?

State-of-the-art CC-IAMs fall far short of capturing the uncertainties in real-world systems. However, IAM modelers face real constraints imposed by the mathematics of complex systems and limits to computational resources. The uncertainty propagation that is to be expected when many variables and functions are uncertain compounds the challenge: uncertainty propagation can be modeled [18], but so doing adds to the computational burden.

Computational burden is one of the motivations for possibility-based modeling approaches [19,20] including imprecise probabilities, binary or interval specifications [20], and ranking schemes [21]. Possibility-based modeling approaches are, in principle, less demanding of computational resources. These approaches demand less information than is implicit in frequentist probabilities and reduce the temptation to substitute structure for information. However, they address uncertainties in binary or interval (i.e., lumpy) fashion. Viewed this way, possibility-based approaches offer a clear trade-off: we could address a much broader set of real-world uncertainties but address each less precisely. The trade-off may tilt in favor of possibility-based approaches if one concedes, as do Pindyck [2,22] and Heal [23], that the CC damage function is understood very poorly, especially toward the extremes of the distribution. There is perhaps too much scope for modeler discretion in substituting structure for information when specifying the damage function and choosing the relative weights to place on most-likely versus extreme cases. Possibility-based (or robust) approaches have the advantage of consistency with threat-avoidance strategies [15,17], including precautionary approaches.

Given that the default approach to risk among economists—and many others, too—is to invoke stochastic methods, it is important to point out the common logical foundations of possibility and probability approaches: there are convergence theorems relating possibility theory and probability theory [24]. More pragmatically, one might ask whether decision makers seek, or even demand, probability-based formulations. Perhaps, for some kinds of issues, outcome propositions expressed in binary or interval fashion are more credible, especially to decision makers.

3.6. What Can Be Gained in Validity by Improving Our Characterization of Uncertainty in IAM?

Suppose modelers were able to improve substantially our expression of the uncertainties and their likely consequences. What would be gained? Modeling more of the uncertainties likely to be encountered would produce, most likely, a broader spread and a finer-grained mapping of potential system outcomes; and perhaps more fully articulated notions of the likelihoods of alternative scenarios and projected outcomes. It is possible also that improved modeling of decision maker response to uncertainty would improve CC-IAMs, resulting in more realistic model structures, more reliable model parameters, and perhaps greater capacity to model the effects of new uncertainties likely to emerge along with the new most-likely outcomes.

4. Validation and Credence in IAM Output

How can the claim that outputs of a simulation merit justified belief be evaluated? First, we attend briefly to a few rather sweeping assertions that claims of validation for CC-IAMs (or simulations in general) are inherently misleading.

4.1. Arguments That Validation Claims Re IAM Are Inherently Misleading

Oreskes et al. argued in 1994 that the idea of validation is misplaced in simulation and claims of validation [25] are misleading: (paraphrasing) *to claim validation is to risk misleading the reader, who might place unwarranted trust in the model and its projections as they pertain to the real world.* In a similar vein, Konikow and Bredehoeft argue that (again, paraphrasing) *emphasizing validation deceives society with the impression that, by expending sufficient effort, uncertainty can be eliminated, and absolute knowledge be attained* [26]. Since these sweeping claims are founded on a binary concept of validity (valid/not), we have

already rejected them (Section 1.1). Validation makes sense only as an ordinal, rather than absolute, concept and validation in practice is a matter of degree. While a model that merits absolute credence is aspirational but unattainable, the practicable aspiration should be to build models that merit more rather than less credence.

This defense of the idea of validation does not let IAM modelers off the hook. With respect to information and expertise, there remains a massive asymmetry between modelers and their audience, presumably policy makers and the informed public. Informed critics can help narrow that gap, as they have in fact been doing rather well in recent years in the CC case. The risk of being interpreted as knowing more than we do surely is present in simulation, but that risk is always present in research, and the scientific community has well-established ways—the ethic of modesty in researchers’ claims and the tradition of robust review and critique—of moderating it. Nevertheless, a burden of transparency rests heavily on the shoulders of CC-IAM modelers. For example, the justification of CC-IAMs as “if . . . , then . . . ” analyses that facilitate exploring the implications of alternative model structures, parameterizations, and driver settings entails an obligation of transparency about the nature of the exercise.

Pindyck leveled an additional charge: CC-IAMs are little more than rhetorical devices because they can be manipulated so readily to achieve results congenial to the researchers [2]. The heft of this charge depends on the meaning attached to “rhetoric”. As McCloskey insisted, to call CC-IAMs devices that advance the art of reasoned argumentation [27] is very different from calling them devices for inflated and bombastic oratory, as suggested by the *Word* dictionary.

It is unsurprising that, within limits, dueling modelers might be able to get results they each find congenial by tweaking the assumptions. For example, in modeling the potential benefits of climate change mitigation, modelers have some discretion over assumptions about the discount rate and the weight placed on unlikely but high-damage projected outcomes. The challenge is to place that fact in context. The “if . . . , then . . . ” (or scenario) analysis perspective helps clarify what is at stake. Pindyck’s particular concerns, the discount rate and the weight placed on unlikely but high-damage outcomes [2], are exactly the kinds of issues that should be debated vigorously when assessing the benefits of mitigating climate change. That debate began in earnest with Stern and Nordhaus as protagonists in 2007 [28–30] and is in full swing now; among others, Millner et al. [31], Traeger [32] and Dietz et al. [33] have shown that the DICE model can be re-worked with lower discount rates, greater sensitivity to uncertainty, and ambiguity rather than pure risk, in each case obtaining conclusions more favorable to aggressive CC mitigation. One may wonder why it took so long. We speculate that the DICE structure—it maximizes intergenerational welfare—impeded transparency by embedding the discount rate and the treatment of uncertainty within the model. Millner, Traeger, Dietz and their co-authors had to work hard to obtain their results. This raises a dilemma for economists engaged in CC-IAM. The appeal of maximizing intergenerational welfare is obvious, since it conforms so well with the notion of weak sustainability. Alternatives are available, though: models can be closed by market-clearing constraints, which would generate projected outcomes in terms of prices and quantities. Welfare assessment then would proceed, more transparently, by imposing explicit discount rates and weights on avoiding worst-case outcomes.

We find the “if . . . , then . . . ” defense of CC-IAMs convincing but, again, transparency really matters. “If . . . , then . . . ” conveys a very different message than “our model projects . . . ”, and conclusions about the welfare implications of CC mitigation convey a yet more amped-up sense of authority.

4.2. Does Simulation Per Se, as Compared to Other Established Ways of Doing Science, Pose Special Problems for Validation?

Rodney Brooks, a recognized pioneer in artificial intelligence, has been credited with the aphorism “The problem with simulations is that they are doomed to succeed” [34]. That is, there is something really different about simulation that masks failure and undermines

quality control. Winsberg attempted to provide a methodological foundation for the contention that simulation really is different in some important way(s), compared to scientific methods that are more venerable and perhaps better accepted [35]. He argued that inferences from simulations have the following three properties. They are *downward*, i.e., *originating in theory*. In simulation, our inferences about particular features of phenomena are commonly drawn (at least in part) from high theory. This contrasts with the standard (but contentious, we would interject) claim of empiricism—still the consensus methodology among practicing scientists—that theory is developed by generalizing from observed particulars. Simulations are *motley*. In addition to theory, simulation results typically depend on many other model ingredients and resources, including parameterizations (data driven or not, as the case may be), ad hoc assumptions, numerical solution methods, function libraries, mathematical approximations and idealizations, compilers and computer hardware, and a lot of trial and error. Finally, simulations are *autonomous*. Much of the knowledge produced by simulation, e.g., projections of future outcomes, is autonomous in the sense that there is no observable counterpart that provides a clear standard for validation [36].

None of these three conditions is original to computer simulation [37]—they apply also to pen-and-paper modeling—but Winsberg argued in 2013 that it is the simultaneous confluence of all three features that is new and daunting in simulation [38]. An applied environmental economist must disagree: nonmarket valuation, for example, is theory-driven, autonomous (why do it, if valid values were readily observable?), and collection of data and its econometric analysis are quite motley, as anyone attempting a meta-analysis is sure to notice [39]. It should be noted also that by 2022, Winsberg seemed less sure that simulation is uniquely challenged in this respect [38].

Nevertheless, relative to more standard ways of doing science, the difficulty of validating CC-IAM simulations is, if not strictly different in kind, surely different in degree. The most basic intuitive criteria for model validation are “is it built right?” and “does it predict well?” Winsberg’s first two properties of simulations—reliance on theoretical propositions, often in lieu of evidence, and input from a variety of sources that vary widely in terms of their epistemological status—raise substantial impediments to “is it built right?” tests, increasing the burden on prediction tests. However, autonomy restricts the applicability of prediction tests.

4.3. *Is it Built Right? The Emergence of Regional and Local CC-IAMs*

There has emerged a proliferation of regional and local CC-IAMs, some down to the watershed level, encouraged by government initiatives (e.g., the Food, Energy, and Water Systems program of the National Science Foundation in the US, and similar initiatives in China and, more recently, Europe). While the regions studied are in fact embedded in the global economy and atmospheric carbon-greenhouse-gases regime, it is unreasonable to expect builders of regional and local CC-IAMs to construct global models from scratch.

The IPCC, seeking to anticipate how future developments in global economy and society might impact climate forcing, has produced sets of shared socio-economic pathways (SSPs) and representative concentration pathways (RCPs) using a process that leans heavily on large panels of experts [40]. Early commentators, e.g. [41], cautioned that the SSPs were not intended for direct use as scenarios in CC-IAM policy modeling, but such uses began to proliferate. The O’Neill et al. review in 2020 [40] tacitly accepted the trend toward treating the SSPs as scenarios, and focused on developing recommendations to better coordinate SSPs and RCPs, and distinguish between appropriate and inappropriate uses in IAM.

SSPs and RCPs are both free-standing but incomplete: RCP-generated climate projections are not matched to specific societal pathways, while the SSPs are alternative societal futures independent of climate change policies and impacts. It is left to the builders of regional and local CC-IAMs to select and combine particular SSPs and RCPs for assessing climate risks and adaptation or mitigation strategies. This process is not without its challenges. O’Neill et al., while mostly supportive of SSP and RCP efforts thus far, offer a

series of recommendations for developing and using this framework [40]: improve integration of societal and climate conditions; improve applicability to regional and local scales; extend the range of reference scenarios that include impacts and policy; capture relevant uncertainties; keep scenarios up to date; and improve relevance of climate change scenario applications for users (e.g., policy makers).

In particular, the recommendation to focus on scenarios that include impacts and policy seems crucial. The independence of SSPs and RCPs from each other and their agnosticism regarding climate change policies and impacts is not quite credible: surely the SSPs and RCPs carry policy and impact baggage that is not transparent but is nevertheless built into the baseline for policy simulations.

4.4. Critiques of Validation as Practiced

Having argued that validation is not an inherently misleading concept in the context of CC-IAM, we turn to questions concerning the nature of validation for simulation models. What is claimed when we say that a model has been validated? The following list provides a succinct summary of the norm for validation of simulation models, i.e., an agreed standard that, while perhaps less stringent than the ideal, is aspirational for practitioners. (i) The model structure and parameterization represent in a computationally tractable way the essence of what is known, understood, and plausibly conjectured about the system under study and the conditions under which it might be expected to operate during the time period under consideration. (ii) Model implementation has been verified to ensure the absence of mistakes in programming and data entry, and failures in computation. (iii) The model has been refined in an iterative process involving calibration, i.e., testing and adjustment in response to test outcomes. (iv) The resulting model has been subjected to a suite of validation tests and has performed reasonably well. (v) All of the above has been reported in a manner that informs independent evaluation and critique.

Taking this list as a norm for validation of simulation models, we now consider two serious challenges to common practice.

In validation, often a matter of tracking and matching exercises, the bar frequently has been set too low. So, how is the IAM community doing, when it comes to prediction tests? Parker claimed that too much of what passes for validation of simulation models lacks rigor and structure because it consists of little more than side-by-side comparisons of simulation output and observational data, with little or no explicit argumentation concerning what, if anything, these comparisons indicate about the capacity of the model to provide evidence for specific hypotheses of interest [42].

First, note Parker's Popperian stacking of the deck. A Popperian might assume without hesitation that hypothesis testing is the sine qua non of science, but IAM is commonly undertaken in order to explore likely and alternative futures, and future-oriented hypotheses that are testable now are hard to come by (Winsberg's autonomy issue). Furthermore, it should be noted that Popper's strict falsificationism is itself a methodology that Popper eventually abandoned, and many philosophers of science find problematic and ultimately unconvincing [43]. However, Parker's challenge nevertheless has some heft. Surely tracking and matching exercises should accompany numerical displays with explicit argumentation as what they portend for validation. Setting a low bar for empirical/numerical validation does little to enhance credence in IAM.

The criterion for validation should be survival of severe testing. Parker [42], following Mayo [44], defines severe tests as those that have a high probability of rejecting H iff H is false, a definition very much in the "bold hypotheses, severe tests" tradition of Popper in his middle years when he was skeptical of "confirmationism". Parenthetically, in his later years, Popper was more inclined to attach some credence to well-corroborated hypotheses [43]. Parker recognizes that severe tests are aspirational (e.g., there are few chances to test climate projections for decades into the future against the reality), but she clearly regards the aspiration as virtuous. She offers a list of potential errors in IAMs, including errors that

would be exposed by verification as well as those requiring validation, and follows it with several pages of detailed discussion of severe tests that should be applied.

More recently, Parker has been open to exploring the possibility that computer simulation is capable of providing evidence for hypotheses about real-world systems and phenomena, a kind of symbiosis between data from the real world and output from simulation models [45], while maintaining that evidence from simulation is of a different kind than that typically obtained from observation and experiment [46].

Katzav et al. [47] critique the IPCC-style confidence-building approach to validating CC-IAMs [48] from a severe testing perspective. They recognize problems with severe testing, but find confidence building—which leans heavily upon corroboration, tracking, history matching, etc.—even more problematic. They worry specifically that climate models share too much structure and data—and hence many of the same imperfections [49]—to provide convincing convergence tests. Furthermore, they charge, models are typically “tuned”, i.e., calibrated in extended test/revise/re-test routines, thereby invalidating many of the correspondence tests that are offered as evidence of validity. Grim et al., however, defend calibration as essential to debugging the model [34]. This discussion of tuning exhibits analogies to the debate about specification search in econometrics [50].

Lloyd argues for a relatively rigorous confirmation process [51]. For example, more independent corroborations should carry greater weight than fewer, different kinds of confirmations should count more than just one kind, etc.—here again we see parallels with the nonmarket valuation literature [38]. Lloyd, while offering substantive suggestions for improving validation practice in CC-IAM, concludes that climate models have been tested more rigorously than critics have recognized [51].

All of this suggests a two-stage process: calibration to improve the model, followed by testing that is independent of the preceding calibration—again, the analogy to the specification search debate in econometrics springs to mind.

Grim et al., seeking to undermine the Brooks aphorism that simulations are doomed to succeed, catalog a multitude of ways simulations can not only fail but can be seen to fail [34]. In the end, they see mutual obligations between modelers and critics: modelers should attempt to make claims of correspondence as explicit as possible. At the same time, however, critics of a simulation must specify how lapses in correspondence constitute relevant failures. Current procedure often leaves it to the reader to simply ‘see’ the relevant correspondences.

5. Validation Criteria for IAMs

The literature offers many checklists of validation tests. Examples mentioned already include Parker’s list of potential errors that should be submitted to severe testing [42], the Grim et al. list of possible simulation failures, along with suggestions, or at least broad hints as to the kinds of tests that might be appropriate [34]. Several additional authors offer lists. Sargeant’s is perhaps typical: a lengthy checklist [52]. Roy and Oberkampff offer “a complete framework to assess the predictive uncertainty of scientific computing applications”, addressing indeterminism and incomplete knowledge [18]:

- Address aleatory (or random) uncertainties in model inputs using cumulative distribution functions.
- Treat epistemic uncertainties as intervals.
- Propagate both types of uncertainties through the model to the system response quantities of interest.
- Estimate numerical approximation errors using verification techniques.
- Quantify model structure uncertainties using model validation procedures.
 - Compare experimental data, calibrate.
 - Extrapolate the uncertainty structure beyond experimental data.
- Communicate the total predictive uncertainty to decision makers.

Roy and Oberkampf are aerospace engineers for whom an obvious reference case is simulation models for space flights. Such simulations are simpler than many IAMs, with fewer but clearer objectives, and the need for accuracy of projections is more crucial. Importantly, these simulations are less autonomous than many others in the sense of Winsberg [38], because theory and measurement in astronomy and astrophysics are sufficiently well developed to provide a rather precise set of expectations with which to compare simulation results. In contrast, many IAMs are exploratory—aiming to provide a relatively big-picture sense of the metaphorical terrain, e.g., of global wellbeing under various climate change scenarios, while achieving lesser standards of accuracy in projections—and autonomous. Impressed as we are with the work of Roy and Oberkampf [18]—and we agree with them concerning the need to take epistemic uncertainties and uncertainty propagation seriously—calculating and communicating the total predictive uncertainty would be asking too much of IAMs.

Criteria for credence in IAMs should be established in the context of the complexity that is common among IAMs, the provisional nature of the models and the resulting projections, and the mutual understanding among modelers and users that “if . . . , then . . . ” analysis ranks high among the services that the models are intended to provide.

6. Conclusions

6.1. Conclusions Re Validation Criteria for IAMs

We conclude that, while perfection in validation is aspirational, credence in model results may be enhanced by:

- Constructing models that capture the relevant features of the real world, including its uncertainties, in convincing fashion.
- Addressing uncertainty in structural equations and parameter values in the model and its estimation.
- Verifying that the modelers’ intentions are implemented accurately, precisely, and completely.
- Confirming the representations of variation in parameters by applying appropriate statistical measures and tests.
- Testing and calibrating model performance using history matching, tracking, and prediction tests, given near-median and extreme values of key variables. If real-world experience does not yield observable responses to extreme driver values, test whether the model response to extreme driver values accords with expectations informed by theory.
- Sequential updating of model structure and parameterization to reflect what is learned in the calibration process, thereby improving model structure and parameterization.
- Exposing the resulting model to validation tests that are independent of prior calibration.
- To the extent that the model has evolved through sequential learning and updating, communicating this process to end users.
- Communicating results in a manner that conveys the nature of the exercise—in many cases, “if . . . , then . . . ” analysis of how alternative settings for exogenous and policy drivers may affect future outcomes—and fully reflects the remaining epistemic and aleatory uncertainties.

This list of validation criteria is consistent with the agreed norm summarized at the beginning of Section 4.3, but it is more complete and more detailed. Given the trend toward regional and local IAMs, we support the O’Neill et al. [40] agenda of better integrating SSPs, RCPs and policy simulations, and a more transparent interface between models/modelers and policy makers. Where our list immediately above departs from the earlier list, it provides greater specificity as to how credence may be enhanced. It is informed by the literature on severe testing but, consistent with our review and critique of that literature, the tests we endorse are more feasible and perhaps a little less severe. Many of the items on our list endorse practices that are being adopted by leading modelers. The exceptions involve epistemic uncertainty—uncertainty re model structure and uncertainty in structural

equations and parameter values—and uncertainty propagation, which reflects the difficulty of addressing these kinds of uncertainty within the familiar modeling conventions.

Addressing the current CC-IAM controversies, we endorse the optimistic view that CC-IAMs are potentially key contributors to informed climate policy and argue that the methodological literature, properly understood, offers grounds for that kind of optimism. Nevertheless, there is substantial scope for improving the validation of CC-IAMs, the transparency of these models—which has been challenged anew by the emergence and widespread use of the SSPs and RCPs—and the way in which the whole modeling exercise and its real-world implications are communicated. In a very uncertain world, improved modeling and communication of uncertainty is a key component of a comprehensive strategy to enhance the validity of CC-IAM.

6.2. Conclusions Re Computer Simulation

The methodological literature on computer simulation was mostly skeptical at the outset: see, e.g., the critique that validation claims are inherently misleading since validation is impossible [25,26]. Pindyck's appraisal of CC-IAMs echoes these earlier complaints about simulations generally: they are inherently flawed, fundamentally misleading, and in essence mere rhetorical devices because they can be manipulated so readily to achieve results congenial to the researchers [2]. Given the framework we have proposed, these critiques are no longer credible: validity is best viewed as aspirational and, other things equal, it makes sense to seek more rather than less validation. The critics fail to recognize the extent to which calibration serves to discipline simulations when future outcomes are unobservable, and to understand that exploring the implications of a range of assumptions is an essential component of "if . . . , then . . . " analysis.

Mayo in the 1990s [44] and Parker a decade later [42] took simulation more seriously than, say Oreskes et al. [25], but argued for harsh tests in the style of mid-career Popper. More recently, Parker's view has become much more nuanced: she has been open to exploring the possibility that computer simulation is capable of providing evidence for evaluating hypotheses about real-world systems and phenomena, and in that sense shares a symbiotic relationship with real-world data [45], while maintaining that evidence from simulation is of a different kind than typically is obtained from observation and experiment [46].

Drawing on several strands of literature, and understanding that perfection in validation is aspirational, we have suggested steps toward validation that would enhance credence in results from CC-IAMs and, by extension, computer simulations more generally. Progress is being made in most of the directions we recommend, but the stubborn outlier is addressing epistemic uncertainty in model structure and specification and parameterization of structural equations, and the potential for uncertainty propagation in complex systems. The ultimate goal must be to incorporate deep uncertainty in credible ways.

Author Contributions: Conceptualization, A.R.; early drafts, submitted version and revisions A.R. and J.O.-H.; investigation, deep uncertainty, J.O.-H. and A.R.; investigation, integrated assessment modeling, A.R. and J.O.-H.; investigation, philosophy of simulation, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: Research support was provided by National Science Foundation Innovations at the Nexus of Food Energy Water Systems grant INFEWS #1739909 and the National Institute for Food and Agriculture award #2018-68002-27932.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Shaohui Tang for excellent research assistance and members of the IAM workshop at The Ohio State University and this journal's reviewers for helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Metcalf, G.E.; Stock, J.H. Integrated Assessment Models and the Social Cost of Carbon: A Review and Assessment of U.S. Experience. *Rev. Environ. Econ. Policy* **2017**, *11*, 80–99. [CrossRef]
2. Pindyck, R. The use and misuse of models for climate policy. *Rev. Environ. Econ. Policy* **2017**, *11*, 100–114. [CrossRef]
3. Weyant, J. Contributions of integrated assessment models. *Rev. Environ. Econ. Policy* **2017**, *11*, 115–137. [CrossRef]
4. Nordhaus, W. Estimates of the social cost of carbon: Concepts and results from the DICE-2013R model and alternative approaches. *J. Assoc. Environ. Resour. Econ.* **2014**, *1*, 273–312. [CrossRef]
5. Winsberg, E. Computer Simulations in Science. In *The Stanford Encyclopedia of Philosophy*, Summer 2015 ed.; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2015. Available online: <https://plato.stanford.edu/archives/sum2015/entries/simulations-science/> (accessed on 19 May 2023).
6. O'Hagan, T. Dicing with the Unknown. *Significance* **2004**, *1*, 132–133. Available online: http://www.stat.columbia.edu/~gelman/stuff_for_blog/ohagan.pdf (accessed on 19 May 2023). [CrossRef]
7. Carnap, R. *Logical Foundations of Probability*; University of Chicago Press: Chicago, IL, USA; London, UK, 1950.
8. Halpern, J. *Reasoning about Uncertainty*; MIT Press: Cambridge, MA, USA, 2003.
9. Norton, J. Ignorance and indifference. *Philos. Sci.* **2008**, *75*, 45–68. [CrossRef]
10. Hatfield-Dodds, S.; Schandl, H.; Adams, P.D.; Baynes, T.M.; Brinsmead, T.S.; Bryan, B.A.; Chiew, F.H.S.; Graham, P.W.; Grundy, M.; Harwood, T.; et al. Australia is 'free to choose' economic growth and falling environmental Pressures. *Nature* **2015**, *527*, 49–53. [CrossRef]
11. Madrian, B.C.; Shea, D.F. The power of suggestion: Inertia in 401 (k) participation and savings behavior. *Quart. J. Econ.* **2001**, *116*, 1149–1187. [CrossRef]
12. Thaler, R.; Sunstein, C. *Nudge: Improving Decisions about Health, Wealth, and Happiness*; Penguin Books: New York, NY, USA, 2009.
13. Cai, Y.; Judd, K.; Lenton, T.; Lontzek, T.; Narita, D. Environmental tipping points significantly affect the cost-benefit assessment of climate policies. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4606–4611. [CrossRef]
14. Cai, Y.; Sanstad, A. Model uncertainty and energy technology policy: The example of induced technical change. *Comput. Oper. Res.* **2016**, *66*, 362–373. [CrossRef]
15. Cai, Y.; Lenton, T.; Lontzek, T. Risk of multiple interacting tipping points should encourage rapid CO₂ emission reduction. *Nat. Clim. Chang.* **2016**, *6*, 520–525. [CrossRef]
16. Cai, Y.; Steinbuks, J.; Judd, K.L.; Jaegermeyr, J.; Hertel, T.W. Modeling Uncertainty in Large Natural Resource Allocation Problems. *World Bank Policy Res. Work. Pap.* **2020**, *20*, 9159.
17. Cai, Y.; Golub, A.A.; Hertel, T.W. Developing long-run agricultural R&D policy in the face of uncertain economic growth. In Proceedings of the 2017 Allied Social Sciences Association (ASSA) Annual Meeting, Chicago, IL, USA, 6–8 January 2017.
18. Roy, C.; Oberkampf, W. A complete framework for verification, validation, and uncertainty quantification in scientific computing. *Comput. Methods Appl. Mech. Eng.* **2011**, *200*, 2131–2144. [CrossRef]
19. Dubois, D.; Prade, H. Possibility Theory. In *Computational Complexity*; Meyers, R., Ed.; Springer: New York, NY, USA, 2012.
20. Dubois, D.; Prade, H. Possibility theory and its applications: Where do we stand? In *Springer Handbook of Computational Intelligence*; Kacprzyk, J., Pedrycz, W., Eds.; Springer: Berlin, Germany, 2015; pp. 31–60.
21. Gerard, R.; Kaci, S.; Prade, H. Ranking Alternatives on the Basis of Generic Constraints and Examples—A Possibilistic Approach. *IJCAI* **2007**, *7*, 393–398.
22. Pindyck, R. Climate change policy: What do the models tell us? *J. Econ. Lit.* **2013**, *51*, 860–872. [CrossRef]
23. Heal, G. The economics of climate. *J. Econ. Lit.* **2017**, *55*, 1046–1063. [CrossRef]
24. Huber, F. Formal Representations of Belief. In *The Stanford Encyclopedia of Philosophy*, Spring 2016 ed.; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2016. Available online: <https://plato.stanford.edu/archives/spr2016/entries/formal-belief> (accessed on 19 May 2023).
25. Oreskes, N.; Shrader-Frechette, K.; Belitz, K. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science* **1994**, *263*, 641–646. [CrossRef]
26. Konikow, L.; Bredehoeft, D. Groundwater models cannot be validated. *Adv. Water Resour.* **1992**, *15*, 75–83. [CrossRef]
27. McCloskey, D. The rhetoric of economics. *J. Econ. Lit.* **1983**, *21*, 481–517.
28. Stern, N. *The Economics of Climate Change: The Stern Review*; Cambridge University Press: Cambridge, UK, 2007.
29. Stern, N. The economics of climate change. *Am. Econ. Rev. Pap. Proc.* **2008**, *98*, 1–37. [CrossRef]
30. Nordhaus, W. A Review of the Stern Review on the economics of climate change. *J. Econ. Lit.* **2007**, *45*, 686–702. [CrossRef]
31. Millner, A.; Dietz, S.; Heal, G. Scientific ambiguity and climate policy. *Environ. Resour. Econ.* **2013**, *55*, 21–46. [CrossRef]
32. Traeger, C.P. Why uncertainty matters: Discounting under intertemporal risk aversion and ambiguity. *Econ. Theory* **2014**, *56*, 627–664. [CrossRef]
33. Dietz, S.; Gollier, C.; Kessler, L. *The Climate Beta*; Working Paper 190; Grantham Institute: London, UK, 2015.
34. Grim, P.; Rosenberger, R.; Rosenfeld, A.; Anderson, B.; Eason, R. How simulations fail. *Synthese* **2013**, *190*, 2367–2390. [CrossRef]
35. Winsberg, E. Simulations, Models, and Theories: Complex Physical Systems and their Representations. *Philos. Sci.* **2001**, *68*, S442–S454. [CrossRef]

36. Ramsey, J. Towards an expanded epistemology for approximations. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, East Lansing, MI, USA, 1 January 1992*; Cambridge University Press: Cambridge, UK, 1992; Volume 1, pp. 154–166.
37. Frigg, R.; Reiss, J. The philosophy of simulation: Hot new issues or same old stew. *Synthese* **2009**, *169*, 593–613. [[CrossRef](#)]
38. Winsberg, E. Computer Simulations in Science. In *The Stanford Encyclopedia of Philosophy*, Winter 2022 ed.; Zalta, E.N., Nodelman, U., Eds.; Stanford University: Stanford, CA, USA, 2022. Available online: <https://plato.stanford.edu/archives/win2022/entries/simulations-science> (accessed on 19 May 2023).
39. Randall, A. What practicing agricultural economists really need to know about methodology. *Am. J. Agric. Econ.* **1993**, *75*, 48–59. [[CrossRef](#)]
40. O'Neill, B.C.; Carter, T.R.; Ebi, K.; Harrison, P.A.; Kemp-Benedict, E.; Kok, K.; Kriegler, E.; Preston, B.L.; Riahi, K.; Sillmann, J.; et al. Achievements and needs for the climate change scenario framework. *Nat. Clim. Change* **2020**, *10*, 1074–1084. [[CrossRef](#)]
41. Kriegler, E.; Edmonds, J.; Hallegatte, S.; Ebi, K.L.; Kram, T.; Riahi, K.; Winkler, H.; Van Vuuren, D.P. A new scenario framework for climate change research: The concept of shared climate policy assumptions. *Clim. Change* **2014**, *122*, 401–414. [[CrossRef](#)]
42. Parker, W.S. Computer Simulation through an Error-Statistical Lens. *Synthese* **2008**, *163*, 371–384. [[CrossRef](#)]
43. Caldwell, B. Clarifying Popper. *J. Econ. Lit.* **1991**, *29*, 1–33.
44. Mayo, D. *Error and the Growth of Experimental Knowledge*; The University of Chicago Press: Chicago, IL, USA, 1996.
45. Parker, W.S. Evidence and knowledge from computer simulation. *Erkenntnis* **2020**, *3*, 1–8. [[CrossRef](#)]
46. Parker, W.S. Local Model-Data Symbiosis in Meteorology and Climate Science. *Philos. Sci.* **2020**, *87*, 807–818. [[CrossRef](#)]
47. Katzav, J.; Dijkstra, H.; de Laat, A. Assessing climate model projections: State of the art and philosophical reflections. *Stud. Hist. Philos. Mod. Phys.* **2012**, *43*, 258–276. [[CrossRef](#)]
48. IPCC. *Climate Change 2007: Synthesis Report*; Contribution of Working Group I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change; IPCC: Geneva, Switzerland, 2007; p. 104.
49. Allen, M.; Ingram, W. Constraints on future changes in climate and the hydrologic cycle. *Nature* **2002**, *419*, 224–232. [[CrossRef](#)]
50. Heckman, J. Haavelmo and the birth of modern econometrics. *J. Econ. Lit.* **1991**, *30*, 876–886.
51. Lloyd, E.A. Varieties of support and confirmation of climate models. In *Aristotelian Society Supplementary Volume*; Oxford University Press: Oxford, UK, 2009; Volume 83, pp. 213–232.
52. Sargeant, R. Verification and validation of simulation models. *J. Simul.* **2013**, *7*, 12–24. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.