Simulation and Calibration: Mitigating Uncertainty

Deborah Haar*†

Calibrating a simulation is a crucial step for certain kinds of simulation modeling, and it results in a simulation that is epistemically different from its pre- or uncalibrated counterpart. This article discusses how simulation model builders mitigate uncertainty about model parameters that are necessary for modeling through calibration and argues that the simulation outcomes after calibration are physically meaningful and relevant. When evaluating the epistemic status of computer simulations, comparisons between computer simulations and traditional experiments need to consider this important methodological step.

1. Introduction. One main area of philosophical investigation of computer simulations concerns their epistemic strength. This research has been advanced by considering their features and methodology but also through comparisons with traditional experiments. In this article, I bring together these two strands to argue that aspects of computer simulation, such as its 'motley' nature (Winsberg 1999) and the iterative aspect of simulation model building (Humphreys 2004), have epistemic implications for the comparisons between computer simulation and experiment. I show that the epistemic strength of a simulation

*To contact the author, please write to: Department of Philosophy, University of Illinois at Chicago; e-mail: dhaar2@uic.edu.

†Many thanks to those at the University of Illinois at Chicago's Beyond Spacetime research group, Women in Philosophy at Northwestern University, Models and Simulations 8 at the University of South Carolina, and the German Research Foundation Research Training Group 2073's October 2019 Research Colloquium at Universität Bielefeld, where versions of this article were presented. I would also like to thank the PSA's anonymous reviewers, as well as Martin Carrier, Torsten Wilholt, and Mathias Frisch, for their comments and suggestions. I am very grateful to Robert Will for our many conversations about reservoir simulations, to Daniel Skibra for numerous discussions, and to Nick Huggett for his input on many drafts of this article.

Philosophy of Science, 88 (December 2021) pp. 985–996. 0031-8248/2021/8805-0020$10.00

depends crucially on how it is constructed, notably whether the model is calibrated. This article thus considers in detail the process of simulation calibration and how it can yield information about the target system that is physically meaningful and relevant. Understanding the impact of calibration on simulations shows that comparisons between computer simulation and experiment need to be sensitive to the distinction between calibrated and uncalibrated or precalibrated models.

A point of comparison between simulations and traditional experiment centers on the notion of 'materiality' (Guala 2002, 2005; Morgan 2003; Parker 2009). Arguments focused on materiality consider whether simulations have the kind of relationship with the target found between experiments and the target and, if not, why the relationship found in simulations is epistemically relevant for modeling. Guala (2002) states that since they are materially similar to the target system, experiments do not have to model the system that is being studied and can black box the details. Experiments are thus superior to simulations because experiments have the same material causes as the system that is being studied. A computer simulation, in contrast, would not be materially similar but only formally similar.

Parker (2009) argues that simulations are material experiments since they are run on a physical system and we intervene on them by initializing the system. However, rather than the materiality of a simulation or experiment, relevant similarities justify the inferences made about the target system. Furthermore, Parke (2014) pushes back on the idea that differences between simulations and experiment could serve as the basis for general epistemic evaluations about simulations. Whether an inquiry is classified as an experiment, simulation, or some hybrid of the two, this is not relevant to our evaluation of its epistemic power. What is relevant to the quality of the inferences made using the tool are the details of the inquiry and how well our tool of inquiry captures the relevant features of the target system. In other words, context matters.

Although similarly critical of arguments appealing to mere materiality, Roush (2018) argues for a contextual difference between experiment and simulation in which materiality can play a role. She argues that computer simulations are always epistemically inferior to experiment in cases where relevant information about the target system is missing. When comparing the epistemic value of simulations and experiments under an epistemic 'other things equal' (OTE) condition, Roush recognizes that the epistemic value of a simulation or experiment depends on the specifics of the system being investigated. Roush situates her argument in the case in which there is enough information to begin constructing a simulation but there are some crucial unknowns about the target of investigation. In this case, Roush argues, experiments are epistemically superior to simulations because experimenters can black box the relevant unknowns. That is, if they can assure the internal validity of the experiment, they can run the experiment without having to speculate about the unknowns.[1]

The simulator cannot simply refrain from including the unknown features of the system in the simulation since these crucial unknowns influence the outcomes of the simulation. To account for the unknowns, Roush states that they could make some arbitrary choice about the unknown structure, but this would make the results meaningless. Or they could program multiple simulations based on different hypotheses about the unknown structure, but this would result in a simulation model that would give theoretical results based on the model. Neither option would tell us what happens in the system we are investigating. Another possibility would be to use information that was not available to the experimenter, but then we would not be comparing simulation and experiment under the OTE condition. Rather than materiality, background knowledge is the "crucial and neglected factor that must be held equal" (Roush 2018, 4891).

This article looks at simulation model building and how the modeler deals with uncertainty about the target. Broadly speaking, during the model-building process, modelers go from a general theoretical model to a dynamic computational model by discretizing continuous theoretical equations that are not analytically tractable, rendering them solvable through numeric methods. Details relative to the target system as well as ad hoc elements are added to the model during this process. This description of simulation methodology may make it seem like modeling is a single linear two-step process of first discretizing equations and then adding details to the model. It also suggests that we have all the information we need at the outset of the modeling task.

However, simulation model building is an iterative process.[2] Simulation model building requires the use of the model while it is still in construction to make adjustments based on its output and to generate information about the target system to be included in the simulation model. These tasks are part of the process of calibration that will be discussed in detail in section 2. During calibration, modelers mitigate uncertainty about the target system. I present a case study of the calibration of a simulation of a hydrocarbon reservoir in petroleum engineering, which shows how new information about the target system is gained during the model's construction. In section 3, I argue that a well-calibrated model can provide physically meaningful information about the target. In section 4, I argue that epistemic comparisons under epistemic background conditions such as proposed by Roush (2018) are

1. Roush later argues that the justification for black boxing is the assumption that the experimental object and the target system are of the same natural kind.

2. See Morgan (1999), Winsberg (1999), and Humphreys (2004) for descriptions of simulation model building.

not epistemically relevant comparisons of simulation and experiments because comparison under these conditions forces us to misconstrue simulation model building. There is an epistemic difference between a calibrated model and pre- or uncalibrated model, and comparisons between simulation and experiments need to account for this distinction.

2. Simulations and Uncertainty. Computer simulations are dynamical computer-based models capable of representing complex natural systems such as galaxy formation and the earth's climate. Computer simulations are especially useful when the system of interest cannot be intervened on experimentally (e.g., galaxies). Computer simulations are also used in cases where information is sparse (Winsberg 1999). In these cases, constructing a computer simulation model is partly a process of including what we already know about the system into a computational model, such as any theoretical models of the system (mathematical equations describing the dynamics) and specific measurement data about the system, but it is also a process of mitigating uncertainty with respect to the unknowns.

Different types of unknowns can affect simulation model building. There may be uncertainty about whether all the relevant properties of the target system have been included in the model. This kind of uncertainty concerns what I will refer to as a model's *completeness*.[3] There is another kind of uncertainty that reflects imprecision of value (Oberkampf and Roy 2011; Roy 2019).[4] This kind of uncertainty concerns what I will refer to as the *parametrization* of the model. If Roush (2018) is correct, then the only way to get meaningful information about the system when there are unknowns is to conduct an experiment. However, simulators mitigate uncertainty during model building by calibrating their models to data of the target system.

Measurement instruments have been considered the traditional target of calibration. More recently, the target of calibration has been extended to include objects such as certain procedures of data analysis, economic models, and simulations (Soler et al. 2013). A measurement instrument is calibrated by comparing the known value of a property of an object with the value measured by the measurement instrument. The results of the comparison of the calibration object and the measurement output inform us about the reliability of the measurement instrument (Franklin 1997).

3. A model is never fully complete; the target system can never be fully represented by the model, but the modeler endeavors to include all the relevant parameters.

4. Roy (2019) defines uncertainty in the second sense only. It can be the case that we do not have any information about the value of a given parameter. That situation is also included in the parametrization of a model.

Simulation model calibration involves comparing the simulation output for a set of properties to the values for the same properties measured in the target system. To discuss more concretely how calibration may be used to mitigate uncertainty, I will use an example of the application of numerical flow simulation for the optimization of a subsurface hydrocarbon reservoir. Although this example is drawn from the field of petroleum engineering, it is relevant to the present discussion. Computer simulation methodology is advanced in this field, and simulations are used precisely in cases where there exist the types of unknowns that are of interest.

In petroleum engineering, computer simulations of subsurface hydrocarbon reservoirs are used to model the reservoir to solve predictive-type questions such as 'what is the number and placement of wells required for commercially optimal extraction of this resource'? Computer simulations are used to model reservoirs because of their capacity to quantify uncertainty in predictions resulting from unknowns or poorly understood features of the reservoir. A key feature of computer simulation methodology is that building a simulation is an iterative process; constructing the simulation requires running it to aid in its own construction. Running the simulation helps the modeler determine the level of detail justified in the model and the fidelity of model output that might be reasonably expected.

During the construction of a hydrocarbon reservoir simulation, modelers must address the uncertainty about the model's completeness and parametrization to have a simulation that can be applied to the kinds of problems it is being built to solve. While much empirical data about a subsurface reservoir can be gained through measurements, due to their complex nature, and financial constraints, important features of the reservoir remain unknown to modelers when they construct the simulation. To address these unknowns, the modelers calibrate the simulation. This 'tuning' of the model is carried out with two purposes in mind. One is to determine the model's completeness: Are there any properties, such as a geological fault, missing from the model or that are superfluous? A second purpose is to better parametrize the model: to make better estimates of the values of certain properties that exist in the target system and must be specified in the model. The examples I will use to illustrate parametrization are porosity and permeability.[5]

A reservoir simulation model is composed of a static model and a transient model. The static model is typically a three-dimensional geological representation constructed from direct and indirect measurements. The transient model describes the reservoir's hydrodynamic states of pressure and saturation within the particular spatial domain and time period of interest.

5. The porosity of the rock is the percentage of void space in the rock, and permeability is the capacity of the rock to transmit fluids.

The simulation is based on fundamental equations that provide a system of partial differential equations (PDEs), the solution of which (finite difference or finite element) describes the flow within the reservoir, the reservoir 'response'. This procedure gives us approximate solutions to exact equations.[6]

Geologic structures and petrophysical properties that are important features of the reservoir and relevant for modeling can be unknown. A good example of an uncertainty that affects model completeness is a fault representing a flow barrier. Empirical data about the presence of faults can sometimes, but not always, be gained from seismic surveys or intersecting wells. However, these data alone are often not enough to prove the existence of a fault or whether a fault impedes fluid flow and therefore should be included in the model with an associated value of transmissivity. An example of parameter uncertainty is the uncertainty concerning the parameter values of the porosity and permeability in various locations in the reservoir. Reservoir simulations must contain values for these two essential features of the reservoir, but it is technically and financially not feasible to take measurements of these properties throughout the reservoir.

Mitigating the impacts of uncertainties is a task that is undertaken during calibration, which in reservoir simulation is called production data history matching.[7] History matching is achieved by first forming an 'objective' function by taking the difference between field measurements and model predictions. In the objective function (OF)

$$S_\varepsilon = \sum_{i=1}^{n_{\text{parm}}} \left[w_i (X_{io} - X_{is})^2 \right],$$

X_i refers to the observation being matched, X_{io} are the measured values for that observation and X_{is} are the simulated values, and w_i is the weight assigned to the observational data. For instance, you could have information about pressure from two different wells, and w_i indicates how much each one contributes to the error function.[8] Then the parameter sensitivity is explored throughout the parameter uncertainty space. A goal of history matching is to find parameter values that minimize the OF. The lower the OF value, the less mismatch between the historical data and the simulated output of the parameters being matched. The OF is then minimized through systematic

6. The general description here is one from a mathematical model (the PDEs) to a numerical model (the PDEs discretized nonlinear algebraic equations) and finally to a computer model (discretized and linearized).

7. Before the use of simulations, history matching was done manually.

8. To avoid overfitting during history matching, some of the observational data are reserved for the verification and validation of the simulation in the final stages of the simulation model's development.

adjustment of the selected parameters on the basis of uncertainty and impact using one of many optimization techniques.

During history matching, ideally only data that are least accurately known are updated, and the updating is constrained within limits that are defined in advance by experts of that area (Ertekin, Abou-Kassem, and King 2000, 350). The selection of parameters to match depends on the available empirical data and the objective of the calibration. Pressure and flow rates are common parameter values used to history match when adjusting permeability and porosity values. As part of the optimization process, changes are made to the probability distribution function (PDF) initially assigned to the parameters for which there is uncertainty. Such adjustments result in updated ('posterior') parameter distributions.

There are different techniques and strategies at the engineer's disposal for sampling model parameter uncertainty space: Monte Carlo method is one example. When using Monte Carlo methods, the history matching process looks like this: (1) randomly sample model parameters using the initial PDFs, (2) run the simulation with these values for the historical period, (3) compute the OF, (4) store the value of the OF and all parameters, and (5) repeat 1–4 until the space is properly sampled. Each parameter set generated using the Monte Carlo method is a model of the target system. Using Monte Carlo methods, the modeler does not converge on one model. The modeler selects a cutoff value for the OF.[9] All parameter sets that result in an OF value that meets this cutoff value form a subset of the reservoir model parameter space. This set is analyzed further by looking at the mean and standard deviation of the values. The analysis of the subset of parameters is used to update the unknown parameter values (e.g., for porosity and permeability). These updated values are referred to as the parameters' posterior PDFs.

The reservoir simulation case demonstrates how simulators deal with uncertainty about unknowns in the system. By comparing the model output with historical data from the target system, the simulation is used to generate empirically informed estimates of the unknown parameters as well as indicate whether some feature of the target system is missing. The quality of these estimates determines the ability of the simulation to make meaningful predictions about the target system.

3. The Meaningfulness and Relevance of Simulation Outcomes. Simulators must model the salient properties of the target system including any unknowns that influence the system. According to Roush, the simulators have two options to model the unknowns while not violating the OTE condition.

9. Many factors determine the cutoff value selected, such as the uncertainty associated with the empirical data.

They could make some arbitrary choice about the unknown structure, but this would make the results meaningless. Or they could program multiple simulations based on different hypotheses about the unknown structure, but this would just give us the mathematical consequences of the theoretical model. Neither option would tell us what happens in the system being modeled. In Roush's example, she essentially is considering a simulation that has not been calibrated.

However, the outcomes of a calibrated simulation model, when put to answer the question the simulation was designed to solve, are neither meaningless nor merely theoretical. That is, properly calibrated models can tell us something about what is the case about the physical system being simulated. There are two reasons for this. First, in the reservoir simulation case presented, the features being calibrated are known to be real properties of physical reservoir.[10] Second, the updated estimates for the parameter values and inclusion or exclusion of features to the model are supported empirically by the historical data, as well as the functioning of the simulation itself (this justification rests on background knowledge of the team putting together the simulation). The posterior PDFs for porosity and permeability are not set arbitrarily but depend on the set of simulation models that fall under the OF cutoff. These values are generated from the models that produce results that are most like the historical data gathered from the reservoir. Postcalibration the simulation is intended to be a better representation of the physical reservoir and better reflect the uncertainties associated with the reservoir, precisely so that it can answer the investigative question.

Calibration is also not merely a theoretical exercise. A theoretical exercise would involve developing a theoretical model and solving the foundational equations for some initial values. However, the problem that uncertainties create for simulators is a type of inverse problem. This problem arises when a model describing a system cannot be fully parametrized given the information that is known. Solving inverse problems starts with the data about the system, and then we work backward to find the best set of parameters for the model. In the case of a reservoir simulation, calibrating the model by history matching is the modeler's way of solving the inverse problem.

Calibration requires that we run the simulation with a particular set of parameter values to generate outcomes (X_{is}) that will be compared to the historical data (X_{io}). It could be argued that the way we generate the new parameters during the parameter space exploration is a kind of theoretical exercise since we are picking parameters to see what happens when they are modeled.

10. This is not the case for all simulation calibrations or all elements of a simulation model. In climate simulations, for example, some of the elements included in the model do not correspond to any real feature of the system under study.

However, the parameter space is not entirely theoretical; the parameter space is constrained by known physical limits of geological and petrophysical properties as well as the properties of the particular reservoir being modeled. Furthermore, it has been convincingly demonstrated that simulating is not necessarily a theoretical exercise since plenty of nontheoretical information can be included in a simulation model (Winsberg 2003; Lenhard 2007).

4. Calibrated Simulations: Background Knowledge and 'Other Things Equal'. I have shown that simulators are able to mitigate uncertainty in a meaningful and empirically relevant way by calibrating a simulation to historical data. Contrary to what Roush (2018) asserts, when there are unknowns it is not the case that we can only turn to experiments to answer a determinate question; the simulator could calibrate the simulation model. What does this mean for the Roush (2018) claim about the epistemic superiority of experiments? I agree with the modified claim that, when there are unknowns, experiments are epistemically superior to any simulation precalibration. However, in the kind of simulation I am considering, an uncalibrated simulation is one that is only partially constructed. What we should compare are experiments to calibrated simulations. Yet comparing them under the epistemic OTE condition requires assumptions about simulation methodology that are false.

In the initial statement of the epistemic OTE condition, Roush asserts that in comparisons between simulation and experiment background knowledge must be held equal (2018, 4891). This suggests that to fulfill the epistemic OTE condition, the simulator and experimenter must merely possess the same background knowledge. In the article she indicates further what the epistemic OTE condition amounts to by stating how it can be violated. To model the system, Roush says that the simulator would have to use more knowledge that the experimenter needed or actually existed (4895). Later she states, "Holding other things equal, the simulator must make more specific commitments about the unknown structure and dynamics of the world in order to give an answer to the question" (4895). So while the epistemic OTE condition stipulates that the simulator and experimenter simply must have shared background knowledge, we see that it is in fact a stronger epistemic condition. It is violated if the simulator either uses more knowledge or makes more commitments about the unknowns than the experimenter.

As a first pass, it is worth noting that the OTE condition appears to be violated in the most trivial way by the simulator and the experimenter. Both the simulator and experimenter will require the application of knowledge and experience that is particular to their method of investigation. Simulators need to know details about computer programming that will be beyond the scope of the programming knowledge the experimenters will need to do their analysis, and likewise for the experimenter and experimental setup. Simulators must discretize the equations that are being used to represent the target system. Experimenters need to know how to use their particular measurement devices. Considering all the information that is particular to simulation model construction and experimental setup, on this reading of the epistemic OTE condition we could never epistemically compare simulations to experiments because they would never meet the OTE condition.

However, suppose we ignore knowledge that is particular to each investigation type and just consider the knowledge that is available about the target system. Roush states that the OTE would be violated if the simulators use more knowledge in their model. The epistemic OTE condition could be understood then as one that stipulates that they cannot use information that is not included in the set body of knowledge possessed by both the simulator and experimenter. This point is partially echoed in the second example of how the epistemic OTE is violated by having to make assumptions.[11] Considering that the calibrated simulation contains estimates about the unknown parameters, would comparing the calibrated simulation to experiments not be an epistemic OTE scenario?

Comparing a calibrated simulation to an experiment seems to violate the OTE condition because a calibrated simulation uses more knowledge than the experimenter, but comparing experiment to an uncalibrated simulation, the simulation is not even fully constructed. Calibration takes place during the model-building phase of the simulation, and unlike the model-building phase of an experiment, it already yields new insights into the target system and the existing data. The historical data necessary for calibration would be available to the experimenter before the experiment, and any new information gained about the target system (such as the posterior PDFs for porosity and permeability) could in principle be shared with the experimenter before any experimental run. The experimenter would have to determine whether or how to incorporate this information into the experimental setup. In the case in which the experimenter chooses not to incorporate this information, it does not seem correct to argue that the calibrated simulation and experiment are not epistemically equal since they have access to the same information. Although we are assuming the simulator and experimenter are investigating the same question, it does not mean they have to proceed in exactly the same way.

Penalizing calibrated simulations for using empirically based estimates for the unknowns generated during calibration reveals a faulty assumption about the two methodologies—namely, that there is a clear epistemic and methodological distinction between the setup of an experiment and the run of the experiment and likewise for simulations. Methodologically and epistemically

11. That simulators must make assumptions at all should not be an epistemic strike against them unless assumptions are always liabilities. The process of calibration can be viewed an effort to put the assumptions on surer footing.

we can distinguish the experimental setup from the experimental run in that new information is gained experimentally from experimental runs, not from the experimental setup. However, this is not exactly the case for simulations. Simulation will be run numerous times during the model-building phase. It is not run merely to check for computational matters such as whether the simulation terminates but, as we see during calibration, to generate new information about the target system. Rejecting calibrated simulations from the comparison on the basis of the epistemic OTE condition then demands that we misconstrue simulation methodology (by rejecting calibration) and relegates the generation of new information to only the running of an experiment or the final version of the simulation.

5. Conclusion. In the case where there are unknowns about the system under investigation, investigators have more options than just running an experiment; they can calibrate a simulation that can then be used to yield meaningful results about the target system. Simulators deal with uncertainty about important unknowns of the target system through calibration during the simulation model's construction. This process yields empirically informed estimates about unknowns. When epistemically comparing simulations to experiments, the proper comparison is between calibrated simulations and experiments. Comparing calibrated simulations to experiments should be allowed on the epistemic OTE condition since calibrated simulations make use of information that is also available to the experimenter. Arguing that calibrated simulations violate the OTE depends on a misunderstanding of the methodology of simulations. Given that calibration is a method to deal with uncertainty before running the final simulation, it is not obvious that experiments will always be epistemically superior to simulations when there are unknowns.

REFERENCES

Ertekin, Turgay, J. H. Abou-Kassem, and G. R. King. 2000. *Basic Applied Reservoir Simulation*. Richardson, TX: Society of Petroleum Engineers.

Franklin, Allan. 1997. "Calibration." *Perspectives on Science* 5 (1): 31–80.

Guala, Francesco. 2002. "Models, Simulations, and Experiments." In *Model-Based Reasoning: Science, Technology, Values*, ed. L. Magnani and N. Nersessian, 59–74. New York: Kluwer.

———. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Humphreys, Paul. 2004. *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. New York: Oxford University Press.

Lenhard, Johannes. 2007. "Computer Simulation: The Cooperation between Experimenting and Modeling." *Philosophy of Science* 74 (2): 176–94.

Morgan, Mary S. 1999. "Learning from Models." In *Models as Mediators: Perspectives on Natural and Social Science*, ed. Mary S. Morgan and Margaret Morrison, 347–88. Cambridge: Cambridge University Press.

———. 2003. "Experiments without Material Intervention." In *The Philosophy of Scientific Experimentation*, ed. Hans Radder. Pittsburgh: University of Pittsburgh Press.

Parke, Emily C. 2014. "Experiments, Simulations, and Epistemic Privilege." *Philosophy of Science* 81 (4): 516–36.

Parker, Wendy S. 2009. "Does Matter Really Matter? Computer Simulations, Experiments, and Materiality." *Synthese* 169:483–96.

Roush, Sherrilyn. 2018. "The Epistemic Superiority of Experiments to Simulation." *Synthese* 195 (11): 4883–906.

Roy, Christopher J. 2019. "Errors and Uncertainties: Their Sources and Treatment." In *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, ed. Claus Beisbart and Nicole J. Saam, 119–41. Cham: Springer.

Roy, Christopher J., and William L. Oberkampf. 2011. "A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing." *Computer Methods in Applied Mechanics and Engineering* 200 (25): 2131–44.

Soler, Léna, Frédéric Wieber, Catherine Allamel-Raffin, Jean-Luc Gangloff, Catherine Dufour, and Emiliano Trizio. 2013. "Calibration: A Conceptual Framework Applied to Scientific Practices Which Investigate Natural Phenomena by Means of Standardized Instruments." *Journal for General Philosophy of Science* 44 (2): 263–317.

Winsberg, Eric. 1999. "Sanctioning Models: The Epistemology of Simulation." *Science in Context* 12 (2): 275–92.

———. 2003. "Simulated Experiments: Methodology for a Virtual World." *Philosophy of Science* 70 (1): 105–25.