Correctness criteria for models' validation – A philosophical perspective

Ijeoma Sandra Irobi, Johan Andersson, and Anders Wall
Department of Computer science & Computer Engineering (IDT)
Mälardalen University,
P.O. Box 883
721 23, Västerås
Sweden.
iii03001@student.mdh.se,
johan.x.andersson@mdh.se,
anders.wall@mdh.se

Abstract

Valid models are central to the existence of Computer science as in most other disciplines, but at what point can one say that a model is valid and hence correct?

A model is often taken to be an abstraction and simplification of reality (of the system being modelled) but reality (the nature of measured data, environmental and human factors) in itself, has a nature of abstract complexity, hence a 'correct' model could at best be judged as one which is ***'closest'*** *in representation to the real system, but the question is* ***just exactly how close should 'closest' be to be correct?***

In this paper, we shall examine some common and general correctness criteria for models validation and seek to relate them to various philosophical perspectives to see how much information the basis of acceptance of such valid models could give (content and truth).

We shall also strongly explore and consider the salient philosophical angle, which presents validation only as a method to improve the level of confidence in a model and not a demonstration of its 'truth' content. Models should not be used as a substitute or sole basis for critical thoughts, considerations or major decisions but should be viewed just as a tool for improving judgement and intuition.

1 Introduction

Over time, simulation models have gained grounds increasingly in being used in solving problems and aiding decision-making in several disciplines.

The use of a simulation model can be viewed as a surrogate for experimenting with an actual system whether it exists or is a mere proposal, which could be disruptive, not cost-effective or just impossible. Developers, users of these models and decision makers who make use of information obtained from results of these models are all concerned with whether a model and its results are correct.

However, the simulation model of any system could only be an approximation of the actual system no matter how much time or money is spent on the model building. Hence if the model produced is not a 'close' enough approximation to this actual system, conclusions derived from such model are likely to be divergent and erroneous, leading to possible costly decision mistakes been made.

More so, it is important to note that there is no such thing as a absolute model validity since a model is supposed to be a mere abstraction and simplification of reality. Therefore the definition that would be accorded model validation in the context of this paper would be that of:

"Substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" [1].

From the above definition, it follows that a simulation model should always be developed for a particular set of objectives. In fact, a model, which is valid for one objective, may not be valid for another.

There are several 'scientific' grounds on which models are built. There are even much more techniques on which their validation rest. Philosophically, the premises on which these validations are based tend to raise more questions than answers in their forms of correctness. It is this philosophical angle that would constitute the nerve of the discussion of this paper.

The section 2 of our paper deals with validation definitions in related work and section 3, with forms of validation while section 4 handles some techniques for developing and validating models. In section 5 we shall discuss some sources of errors in models, while in section 6 we shall raise and analyse the philosophical aspects of the validation criteria in the light of information provided by the models (content and truth). In section 7 we shall make point out areas of possible future work and conclude in section 8.

2 Defining validation

Much research work has been done with respect to simulation models.

In their paper, Robert G. Sargent et al states three basic approaches used in deciding whether a simulation model is valid or invalid. These approaches are:

- The development team takes the decision as to whether the model in question is valid. This is a subjective decision based on the outcome of different tests and evaluations done as part of the model development process.
- The use of a third party to decide if the model is valid, independent of the model developers and users. This method is often used when cost associated with the problem the simulation model is needed to address is high and also in terms of certification of credibility.
- The use of scoring models in which scores are determined subjectively when conducting various aspects of the validation process and then added together to determine the category/overall scores for the simulation model. In this case, a simulation model would be considered valid if its total scores and category scores are higher than the passing scores.

Averill M. Law et al [2] reinstates that validation can be done for all models regardless of whether their corresponding systems exists presently or would be built in future.

Also in their paper [3], Jack P.C. Kleijnen et al give insight on validation of models using statistical techniques and reasoned that the technique that should be applied would depend on the availability of data in the real system. Regarding this data availability, they distinguish three scenarios namely:

- No real-life data available
- There is only data on the real output (not the corresponding input or scenario)
- Besides the output data, the corresponding input is also know

They agreed that in the event that no real-life data is available, strong validation claims remain impossible! In this case then, *sensitivity analysis could be used to support validation, which can be defined as a systematic investigation of the reaction of the simulation responses to extreme values of models' input or to drastic changes in the models' structure.*

However, these kinds of analyses show if *factors* have effects that agree with experts' prior substantial knowledge. Unfortunately, in actual practice, it is not all simulation models that have effects with known signs; still many models do have *factors* with known factors.

Kleijnen et al defined their problem entity as the system (real or proposed) that is to be modelled. The conceptual model would be the mathematical/logical representation of this entity for a given study, while the computer model is that obtained through a computer programming and implementation phase. Inferences and conclusions are therefore drawn by conducting experiments on the computerized model.

3 Forms of validation

There are many forms of validation. It could be seen that validation of *conceptual models* is determining that the theories and assumptions underlying the conceptual models are correct and its representation of the problem entity reasonable for a given purpose. The big question here is whether the conceptual model contains all the necessary details to meet the given objectives.

The *operational validity* is referred to as ensuring that the model's output behaviour has enough accuracy for its intended purpose on its domain of applicability whereas data validity is defined as determining that the necessary data for model construction, evaluation and testing are adequate and correct.

Data validation deals with determining that the data needed for building the model, experimentation and validation are adequately sufficient.

White-box validation is the process of determining that constituent parts of the model represent the corresponding real-world elements with adequate accuracy. The big question here is whether each part of the model represents the real world with enough accuracy.

The *Black-box* validation is concerned with determining that the total (entire) model is an adequately accurate representation of the real world.

Several validation techniques are used. According to [4] there is *no algorithm* or particular pattern to select a given technique to use. However, there are several factors, which affect the choice of the techniques that one would use.

4 Validation Techniques

In this section we shall look at several validation techniques often used, and we shall later consider their philosophical interpretations.

Comparison to other models: Different outputs of the simulation model being validated are compared to those of other 'valid' models

Degenerate test: This has to do with appropriately selecting values of the input and internal parameters to test the degeneracy of the model's behaviour. For instance, to test to see if the average number in the queue of a single server continue to increase with respect to time when the arriving rate is larger than the service rate.

Events validity: The events of occurrences of the simulation model are compared to those of the real system to see if they are similar.

Face validity: This is often used to know if the logic used in the conceptual model is correct and if the input-output relationship is reasonable. This has to do with asking knowledgeable people if the system model behaviour is reasonable.

Historical Data validation: If data was collected on a system for building or testing the model, part of the data are used to build the model and the remaining data are used to test if the model behaves in the same way the system does.

Predictive validation: Here the model is used to forecast the system's behaviour and the model's forecast to determine if they are the same.

Traces: Specific entities in the model are followed through the model to know if the logic of the model is correct and if the necessary accuracy is obtained.

'Turing tests': Knowledgeable experts on the system are asked if they can differentiate between the output of the system and model.

Schellenberger's Criteria: This include *technical* validation which has to do with identifying all divergences between the model assumptions and perceived reality as well as the validity of the data used, *operational* validity which addresses the question of how important these divergences are and dynamic validation which ensures that the model will continue being valid during its lifetime.

5 Sources of errors in models

There are several sources of errors in models, which may not necessarily be independent. These sources include:

- *Model-structure*. In both the conceptual model and the mathematical model important physical phenomena might be omitted or overlooked, and mathematical simplifications might be inadequate for capturing complex dynamics.

- *Numerical solution*. The solution of the numerical model might differ dramatically from the (unknown) ideal solution of the mathematical model.

- *Calibration*. Residual uncertainty about values of model parameters remains after calibration.

- *Input values*. Proper numerical values of the code inputs that describe the scenario for prediction might be known only approximately.

6 Philosophical perspectives on models validation

Considering the questions: '*what does it mean to validate concepts? Or what are the criteria? Both philosophers and scientists have been unable to agree about the answers to them.* [Adapted from Shannon, 1975, p. 211]'.

In this section, we shall examine crucial questions arising from the validation criteria of models that have been mentioned above.

The computer science (or information science in general) is faced with this difficulty more so than social sciences because of its diverse constituents, ever-changing contextual environment (technology), and relatively short life span.

Validation assures that a model (or each construct in a conceptual model) contains the features imputed to it in their individual definitions or description. In other words, validity implies that it is well grounded, sound or capable of being justified.

The response of a computer science empiricist to the question "How do we validate?" could be to design *an experiment or build a prototype and test your concept or conceptual model*. But, a fundamental problem with this approach, notwithstanding the assumptions inherent in statistical experimental design, is the presupposition of the "validity" of a concept or conceptual model. That is, a belief in the notion that mere definition implies that a concept has "face validity." If simply using a "term" made it acceptable to a discipline, one would never reach an agreement on commonly held truisms or knowledge of that discipline.

Simulation models are believed across disciples to give *information* on the real system. In [5], a 21st century philosopher Luciano Floridi defines information as basically comprising 'content and truth'.

In philosophy, there is a huge difference between truth and correctness. While truth is an absolute, correctness is relative to the system. For example, if you read a book on the philosophy of mathematics,
"truth" is not the issue because mathematics does not deal in truth but deals with provability. Maybe physics deals in truth, because the job of science and engineering is to understand the world as it is. Thus the issue for consideration here is *correctness*.
An important question to ask in this context would be: *can simulation models yield knowledge about the real world?*

The epistemological importance of this question is such that if the answer is *no*, then what many scientists are doing nowadays is just playing with computers, not creating new knowledge!

However, considering the practical importance of the question, if no is still the answer, it means that the several policies which are now based on simulation models would grossly be misguided. It is interesting to note however that even in the field of philosophy, varying opinions do exist about whether verification and validation are possible or not.

In [6], an interesting philosophical argument issues between Oreskes et al and Fredrik Suppe in trying to proffer solution to this seeming deadlock.

Oreskes et al strongly argues that simulation models cannot be verified and hence scientists cannot obtain knowledge from simulation modelling. On the contrary, Fredrik Suppe retorts that simulation models can be verified in some sense and hence knowledge could be obtained from them. Some important issues that readily comes to mind in this case would be a deep consideration of some epistemological questions such as

- What (and how) do we learn from experience?
- What is the correct way of learning from experience?

There are also several traditional philosophical views, which include Inductivism (enumerative induction, inference to the best explanation and Bayesianism) and Falsificationism.

However, Oreskes et al argues the above, utilizing traditional philosophical debate over inductivism. Their criticism of the traditional view in three different areas stemmed from Hume's problem of induction, which says that

- All inductive reasonings are based on the assumption of uniformity: What we have observed and what we haven't yet are basically similar. According to him, the question would be: 'why can we rely on such an assumption?' Nothing we have observed until today does not assure that the same regularity will hold tomorrow (unless we use induction --- this is a circular argument).
- **Underdetermination**

  - Given any amount of evidence, there are mutually incompatible theories, which equally fit with the evidence
  - when a prediction from a theory contradicts with the observation; there are various mutually incompatible ways for making the theory compatible with the evidence.
- **Theory-ladenness of observation**
  These philosophical views presuppose that our observation is somewhat independent from our scientific theory. But what we see is strongly influenced by our background knowledge and assumptions. A common example would be asking a zoologist and a social scientist to give interpretations of a diagram of a rabbit.

Why do we care about theory-ladenness of observation?
This is because a conflict between two incompatible theories is supposed to be settled by doing some experiment or observation. However, Theory-ladenness can cause a serious problem with such a procedure.

Considering the Underdetermination vs. Theory-ladenness, the difference between the underdetermination thesis and theory-ladenness can be summarized as follows:
Underdetermination
Same evidence -> Incompatible theories

Theory-ladenness
Incompatible theories -> Different evidence.

In the actual sense, arguments by Oreskes et al. are an application of these traditional criticisms of induction to simulation models.

6.1 Degrees of certainty

However, an interesting categorization was projected by Oreskes et al in which they made the following distinctions: They inferred that there were various degrees of certainty:

- Absolutely true (logical truth) ie verification

- Plausible, probable (in terms of evidence) - confirmation

- Consistent (not contradictory) - validation

Therefore from the philosophical analogies given above it can be deduced that:

(a) Models *cannot* be verified in that there is no logical proof that a model is true.

(b) Models *can* be validated, this means that we can prove that a model does not contain a detectable flaw and thus internally consistent.

This can be evident in:

Comparisons

If two totally different ways of solving the same problem give the same answer, these ways of solution may be reliable.

Calibration

Adjust initial values so that the model can accommodate known data. These procedures are far from *verifying* the model.

Confirmation

Models may yield predictions that match with observation, but this means only that the model is probable, not that the model is true.

Therefore from the above analysis, Oreskes et al concludes that:

- The primary value of a simulation model is heuristic, that is, to give evidence to strengthen what may already have been partially established through other means, for instance, sensitivity analysis, or even challenging existing formulations.
- A simulation model is a `fiction'. It is never a `real thing'. (Cartwright).

In contrast to the above views, Suppe assumes a less strict philosophical stance as follows:

(1) It is true that we cannot logically prove that a model is true. But maybe their way of defining 'verify' is too strict. Do we really want that absolute certainty? That makes all empirical knowledge impossible.

(2) Extra factors can affect the result. But still a simulation model is creating knowledge about the real world when the system is isolated or other factors are negligible.

(3) Don't take underdetermination too seriously. Often it is hard to find even one reasonable solution.

(4) Don't take assumption-ladenness of simulation models too seriously, either.

(5) An important aspect of modelling is the mapping relationship between three systems. As far as this mapping relation holds, a simulation model is a representation of that aspect of the real world, not just a heuristic tool.

With view to the above two major open and highly contestable areas, one could strike some good balance by answering the following questions:

- What level of certainty do we want for scientific knowledge?
- Can simulation models provide that level of certainty?

6.2 Possible integrations?

In [7] Khazanchi attempts to integrate notions from the philosophy of social sciences, the information systems (IS) field and its referent disciplines and sets forth a framework for the validation of IS concepts. The proposed philosophical framework for validation of concepts and conceptual models consists of a set of "criteria for validation" of concepts.

He asserts that as a concept fulfils each succeeding criteria its potential ability to have inherent "truth content" with regard to its general acceptance in the field strengthens. After all, "... concept formation and theory formation in science go hand in hand.... The better our concepts, the better the theory we can formulate with them, and in turn, the better the concepts available for the next improved theory." [8].

The following are his suggested criteria for such validation:

1. Is it plausible? A concept or conceptual model is plausible if it has face validity. Plausibility establishes that this model is more than just a belief. This criterion is useful to assess the apparent reasonableness of an idea and could be demonstrated by deduction from past research or theories, or, it could be developed on the basis of observation or induction.

2. Is it feasible? This criterion dictates that a concept or conceptual model, at the least, has the quality of being workable. Added to plausibility, a feasible concept or conceptual model would be operational in that it would be amenable to verbal, graphical, mathematical, illustrative, prototypical characterization.

3. Is it effective? This criterion deals with the question: How effectively does the model describe the phenomena under study? Also an effective concept or conceptual model has the potential of serving our scientific purposes [Kaplan, 1964]. It also guides and stimulates other scientific inquiries.

4. Is it pragmatic? The pragmatism criterion dictates that a concept or conceptual model should not be restrictive to the extent of logically excluding previously valid models. Thus, this criterion provides that concepts or conceptual models should subsume, for obviously practical reasons, any conceptual structures that previously explained related

phenomenon. Hunt [1990] demonstrates this criterion with the example of Newton's law. He argues that simple pragmatism would require that any new conceptual development could not preclude Newton's laws (as in the case of Relativity, where these laws are a special case subsumed within relativity). In effect this criterion emphasizes that concepts and conceptual models should have some degree of abstract, logical self-consistency or coherence with other concepts and conceptual models in the discipline.

5. Is it empirical? (Does it have empirical content?) Empirical content implies that a concept or conceptual model must have "empirical testability" [9]. In this vein, Dewey also affirms that although concepts can be developed without reference to direct observation, and although this logical conceptual development is indispensable to the growth of science, the ultimate test of a concept or conceptual model lies in having the ability to empirically collect data to "corroborate" it. According to Dewey [10], "Elaboration by reasoning may make a suggested idea very rich and very plausible, but it will not settle the validity of that idea.

6. Is it predictive? (Does it explain a phenomenon that is expected to occur?) We can better understand the meaning of this criterion through words of Rashevsky: "A theory or theoretical concept is considered the more convenient or useful, the better it enables us to predict facts that hitherto have not been observed... The scientist constructs theories, theoretical concepts or theoretical frames of reference that are isomorphic with the world of observable phenomena. This isomorphism is never complete, never covers the whole range of observable phenomena... wider the range of isomorphism, the greater predictive value of the theory." Thus, a concept or conceptual model that is predictive would, at the least, demonstrate that given certain antecedent conditions, the corresponding phenomenon was somehow expected to occur [9].

7. Is it intersubjectively certifiable? Hunt, Nagel, and several others are of the opinion that all scientific knowledge, and in consequence, concepts or conceptual models "must be objective in the sense of being *intersubjectively certifiable*." This criterion provides that concepts or conceptual models must be "testable by different investigators (thus inter-subject)." Investigators with differing philosophical stance must be able to verify the imputed truth content of these concepts or conceptual structures through observation, logical evaluation, or experimentation.

8. Is it intermethodologically certifiable? In addition to being intersubjectively certifiable, this related criterion provides that investigators using different research methodologies must be able to test the veracity of the concept or conceptual model and predict the occurrence of the same phenomenon.

7 Conclusions

We have explored validation of models and the general criteria on which it is based. We have also considered common techniques available for these validation and common error prone areas.

Most importantly, we were able to look into several burning philosophical issues, views and opinions held in this area and have come to the vital conclusion that even though models' verification is still highly contestable, model validation is seen even philosophically as a 'can-do'. The eight Khazanchi's postulated criteria further gave insight as to how to 'test' a models' inherent 'truth content'.

An important question remains that even with an agreed level of abstraction for correctness, are we interested in *correct* models or models that *yield* knowledge and information (content and truth)?

8 Future works

The questions that have been explored in this paper have by no means conclusive answers.

In future, we like to view and explore model's validity philosophically as a measure of the model's *absolute* truth content, not just theoretically but experimentally as well.

Another future area of interest would be to compare methodologies of validation across several disciplines.

9 References

[1] Robert G. Sargent: Verification and Validation of Simulation models

[2] Averill M. Law, Michael G. McComas: How to Build Valid and Credible Models – *proceedings of the 2001 Winter Simulation Conference*

[3] Jack P. C Kleijnen: Validation of Models: Statistical Techniques and Data Availability – *Proceedings of the 1999 Winter Simulation Conference. P. A. Farrington, H. B. Nembhard, D. T. Sturrock and G. W Evans, eds.*

[4] Stewart Robinson, Simulation Model Verification and Validation: Increasing the User's Confidence – *Proceedings of the 1997 Winter Simulation Conference, ed. S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson.*

[5] Luciano Floridi: Philosophy of Information as a new field of Research.

[6] Suppe vs. Oreskes et al. Verification, Validation, and Confirmation of Simulation Models: May 7, 1998 PHIL 250
[7] Deepak Khazanki, A Philosophical Framework for the Validation of Information Systems Concepts.

[8] Kaplan, A. The conduct of inquiry, Scranton, PA: Chandler Publication. Co., 1964, p. 52-54.

[9] Hunt, S., Marketing theory: The philosophy of marketing science, Homewood, IL: Richard D. Irwin, Inc, 1990.

[10] Dewey, J. How We Think, Boston: D. C. Heath & Co., 1910 (1933 reprint, p. 183).