A Primer on Validation and Accreditation

D. E. Stevenson
442 R. C. Edwards Hall
Department of Computer Science
Clemson University
PO Box 341906
Clemson, SC 29634-1906
steve@cs.clemson.edu

May 10, 2000

Abstract

Keywords. Validation, physics, logic, numerical analysis.

1 Motivation

All living beings constantly model their world. It is believed by biologists that predators actually solve the equations of motion before attacking their prey. But would they be able to solve these equations on a different planet? Probably not. Herein lies the problem: just because some model works in one case does not mean it works in all cases.

For most of recorded history, scientific and engineering models were simple enough to be solved by the tools of mathematical analysis. These models were very intuitive and, after a problem or two, their results were quite usable. In the 1960's this changed when Edward Lorenz stumbled into chaos, literally and figuratively. The problem is nonlinearity. These systems behave in very unintuitive ways, making the task of validation very difficult. As nonlinear models became more prevalent, the only possible solutions were those produced on the computer. Similar requirements arose in such areas as economics, finance, and transportation.

Let's agree to call *simulation* any use of computation to solve the behavior posed by a model. For much of its history, simulation has been a second choice for developing and testing products and services. In the 1990s, this changed, as the costs of business inefficiency became prohibitive in both time and resources. The United States Department of Defense is seriously considering a procurement policy based on computer simulations, not prototypes [21]. Science, too, is using simulations in a more pivotal way. Experiments in science have become ever more dangerous or completely impossible, examples being nuclear winter and the maintenance of the U. S. nuclear weapons stockpile.

While the above examples point primarily to continuous simulation, discrete event simulations suffer from the same growth in complexity. Consider any computer communications system in this 24/7 world of the Internet. Demands for speed and reliability make every model expensive due to detail and data gathering. Attempts to control the complexity of these simulations has — as it always seems to — been centered on the magic of language; in this case the Higher Level Architecture (HLA) developed by Mitre and the Object Management Group. In fact, the Warsaw conference had a interactive discussion with the developers of HLA.

However, it makes no difference whether we are dealing with continuous or discrete models. We must insure that any decision based on the output of a simulation is grounded in reality. Our purpose is to introduce the reader to two processes: validation and accreditation. Informally, validation is a process we go through to match our models of the world with the world itself. We constantly validate very mundane and not so mundane theories. The next time you trip and fall, you are performing one more experiment which validates the universal law of gravity. Due to the inductive nature of science, complete validation is impossible. A more practical approach to validation is to establish the correctness of specific uses of a model or simulation with restricted parameters. This second approach is called accreditation. We take the view that validation and accreditation are two aspects of the same thing.

Part I
The Practice of Validation and Accreditation

2 Preliminaries

2.1 Modeling is Key

Modeling comes from our need to understand our world and the problems it presents us. How deeply we need to understand, and, hence, how intricate our model, is a function of the costs and risks posed by the problem to be investigated. The costs and risks in turn generate a need to make decisions (which, recursively, is a problem). Therefore, modeling starts with the real world and a problem and ends in decisions with real world ramifications. Modeling is the process of studying a system and then writing down descriptions of the objects and their relationships. This is a fundamental cognitive process. If there were no realities to our world, then there would be no question of the relative merits of the model produced. We do endorse the view expressed by John Kemeny that "Science must start with facts and end with facts, no matter what theoretical structures it builds in between." Modelers must decide what is important to the problem and ignore other aspects. As Richard Bellman said, "...the scientist, like the pilgrim, must wend a straight and narrow path between the pitfalls of oversimplification and the morass of oversimplification."

2.2 Model Validation and Accreditation

We use the term validation to mean the process by which we determine the degree to which our model predicts reality. On a purist note, a model can only be validated on real phenomena data although one often hears the term applied to theoretical models (such as engineering design models) that are not tested against reality. This confusion arises because the terms validation and verification are used interchangeably. There are several terms used in the context of validation. Since their meaning is not universal accepted, we define them informally here.

1. Calibration is putting numbers on dials. That is, the values of parameters to produce results that are correct to the units used.
2. Verification is justifying that one has followed the rules correctly. Verification is "did we build the system right?"
3. Validation is testing our models against the real phenomena. Validation is "did we build the right system?" Validation is about credibility.
4. Accreditation is testing the capability and relevance in a context narrower than total validation. Accreditation is "Are we assured that the system is right for this problem and this particular parameter setting?" These terms are not independent of one another.

We take the view that, methodologically, each accreditation is actually a warrant for validation. Therefore, we can think of validation as the set of all relevant accreditations. Therefore, the overall scheme of a validation plan is to 1. Establishing the warrants for credibility, capability, and relevance. 2. Establishing a plan to collect the data specified. 3. Comparing the values produced to the observations.

2.3 Basics of Validation

Reality includes intended use, costs, risks, and benefits; this is true for the scientists as well as business executives. Validation would be trivial if we lived in a perfect, Platonic world. In such a world

1. All observations were perfectly taken without any error or imprecision;
2. All parameters were perfectly known;
3. All physical relationships were perfectly known and functional;
4. All calculations could be done at infinite precision.

Alas, we do not live in such a world. What is important is that the nature, depth, and breadth of information are outside the application's control. Therefore, we must establish that the model as proposed (and perhaps implemented in a simulation) has the credibility, capability, and relevance to allow us to make *correct decisions.*

3 Simulation

3.1 Simulation Implements Models and Is Not Modeling Itself

Most useful models are eventually too complicated to be explored by hand. Even assuming that one has a model that completely captures the system under study, it is by no means assured that one can write a computer simulation for all possible situations described by the model. One issue is the finite capabilities of the computer, which leads to problems of imprecision. The larger problem is the whole development of the software [ste2, ste1]. Software development is the purview of software engineering that seems not interested in working on scientific and engineering software. There are some techniques, as we discuss in the Best Practices.

3.2 Why should I be afraid?

Here are three recent reasons that VV&A is a crucial part of science and engineering.

1. On February 25, 1991, an American Patriot missile missed an incoming Scud. The Patriots use a floating point value to accumulate time. This causes a round-off error to occur if the battery remains active for approximately eight hours. This particular battery had been on line for over 100 hours. The round-off error caused the missile to miscalculate the closing speed; the Scud completely slipped by the Patriots.
2. On June 4, 1996, the inaugural Ariane 5 rocket, costing $500 milliondollarsandtheresultofa 7$ billion development, was destroyed 30 seconds into its maiden voyage. The proximal cause of the failure was an aborted conversion between 64 bit floating point and 16 bit integer numbers. Ultimately, the cause was attributed to the use of software developed for another rocket; this software was not properly checked out for the Ariane 5.
3. On Dec 3, 1999, the Mars Polar Lander disappeared while descending to the surface of Mars. The problem was immediately apparent: the developers had made a sophomore error in units: one group worked in feet and the other in meters and failed to account for the conversions.

4 The Role of Uncertainty

While one might try to minimize the point by saying these are all human errors, the point is that they would have been caught by any serious attempt to validate. One might think, for example, that using IEEE floating point would eliminate imprecision errors. In the same vein, one might think that the use of formal methods would eliminate errors introduced during development. Unfortunately, uncertainty comes about in other ways.

4.1 Sources of Uncertainty

There is some tendency for workers to equate uncertainty with numerical error. This may be because of the emphasis on numerical error in numerical analysis and numerical methods texts. The Suppes' quote is more on the mark: You cannot map uncountable sets into finite sets without a great loss in information. Uncertainty is the lack of information. Uncertainty arises throughout the conduct of science. In model validation, the study of uncertainty is a central issue. But uncertainty about what?

1. Observations: precision of data, completeness, erroneous data;
2. Predictions: functional form;
3. Decisions: are all factors accounted for? Are all criteria accounted for? Science and engineering add another problem: phenomena usually have multiple explanations (theories).

4.2 Conditions for Validation

One way to think about validation is to consider the subjective syllogism. Observation produces the concrete data after the appropriate statistical procedures have been applied. Predictions come from the simulation. The predictions then must be compared to real data. Because of uncertainty, we must make the Zeigler decision with uncertainty, probably statistically. Goals for Research in VV&A Framework must hand the two different reasoning styles in science and mathematics: (1) inductive reasoning and (2) deductive reasoning. To fully grasp the confusion and controversy, one must first understand the differences in scientific and mathematical *truth*.

Science use inductive reasoning (not the same induction as used in high school induction proofs). In inductive theories, we have little or no formal logic help. In effect, scientific theories are *proof by affirming the consequent* which is a fallacy in deductive logic. Inductive systems are such that the can find new information from the premises. On the other hand, deductive logic, so familiar to computer scientists, mathematicians and engineers, cannot find new information. In fact, the premises must hold all the keys to proving the consequences.

Thus, to understand validation, we must look to the scientists view.

4.3 Understanding the Scientists" View

The major distinction between science and mathematics is simply this: the scientists look for explanations that can be related to the "common man." Mathematics is used as a tool, but the equations have intrinsic meaning whereas to a mathematician, an equation is a set of symbols with no intrinsic meaning. Here, in Feynman's words, is the crux of the matter. "But, for example, in Faraday's law we don't really need mathematics. In this case, Faraday's law is 'obvious' manifestations of atoms. But for Newton's law of gravitation, we only have mathematics because the underlying physical causes are not known in detail. Mathematics is needed for detailed phenomena in large, complex systems" [10].

Knowledge, then, is our ability to explain. But that explanation must be accepted by the scientific community as a whole. Despite the popular perception of the lone researcher, the truth is that the methodology of science is an agreed upon set of principles used to build a coherent set of explanations. Knowledge requires that we justify our conclusions using the methodology. Unfortunately, the bosses' view of what makes scientific work acceptable and the communities' view of acceptance are vastly different [26].

4.4 Explanations and the Multiple Assumptions of Science

Open any science and engineering textbook. A quick reading will convince the reader that there are multiple sets of competing assumptions about any subject. We call these sets of assumptions explanations. In theory, there is a language system for each such explanation.

On the other hand, the need for the scientist is to consider competing explanations as a system. That is, we must be able to reason about the various views resulting from a particular set of observations.

This is a note to myself.

1. Physicists like mechanism!

2. There is lots of stuff in this book. But the main point here is that the physicists look for consistency in a model-based sense. Because of their synthetic viewpoint, we want to support that by some sort of methodology that emphasizes consistency. One such mechanism is to insist on topological ideas, or at least partial order ideas.

3. Continuous and open is the trick. In other words, we want to start with data that forms a topological space and then we have continuous maps to preserve the topological-ness of the data.

4.5 Real Explanations

Explanations are not justification. "Explanation is a basic feature of science. Science seeks to organize and systematize our knowledge of what goes on in the world on the basis of explanatory principles that can afford answers to why questions" [24]. Explanations must elucidate linkages between the *explanans* (the thing doing the explanation) and the *explanandum* (the thing being explained). It should elucidate how the connections among objects actually function. They should be at least coherent with current understanding; that is, they should integrate laws into a corpus, an accepted scheme of things.

Explanations must make the subject intelligible, usually by reducing it to the familiar. The understandability of the explanation is the key concept. In the scheme of things, it may be necessary to add concepts of process: the how of the subject. Explanations should provide the principal considerations bearing on correctness and relative strength of the explanation. The explanation should be complete — it should explain all observations. Explanations must clarify predictions.

The explanations must subsume observed events within accepted, established generalizations and laws; i.e., the scientific methods of the discipline. Here we get into the probabilistic explanations versus universal law. Develop the sufficiency and necessity versus the modal probability. "A scientific explanation will be a subsumption argument whose major premise is a suitably lawful proposition" [24, p. 13]. Consistency shows here: Subsumtion is a synonym for "is a subset of"' — exactly the topological situation. Regardless of the form of the explanation, Feynman sets the goal of validation: "Every theory that you make up has to be analyzed against all possible consequences." [10, p. 39].

Therefore, validation of model is relative to the underlying explanation it implements. Digress into what consistency is and how topology might be a good framework.

5 Scientific Reasoning

Mathematical thinking can actually be a hindrance because equivalent formulations in physics add insight; mathematical equivalence (isomorphisms) does not. "Mathematics is a language plus reasoning — Mathematics is a tool for reasoning. It is in fact a big collection of the results of some person's careful thought and reasoning. But with mathematics it is possible to connect one statement to another" [10, p. 40]. Physical laws are not just statements. "Laws of theory extend beyond the range of their deductions." [10, p. 50]

But the big difference is that in physics, equations have true semantic meaning — their physical manifestation. "We can deduce often from one part of physics, like the Law of Gravitation, a principle which turns out to be much more valid than the derivation (angular momentum). ... We have these wide principles which sweep across the different laws, and if we take the derivation too seriously, and feel that one is only valid because another is valid, then we cannot understand the interconnections of the different branches of physics" [10, p. 49]

6 Best Practices

6.1 Preliminaries

A problem is a given and presents a need for a decision. The decision sets the need for information. The two taken together set the modeling requirements. The model sets the need for accuracy and sensitivity. The model also sets the outputs. Validation would require that everything be tested against reality.

In the real world, there are not the resources to do a complete validation study. Accreditation requirements detail the type and depth of information that is needed to ensure that a model meets the modeling requirements; i.e., can present a credible defense of the model. We need to validate observational data as well as simulation data/operation. [ASP-III is the buzzword in DMSO docs.]

Accreditation sets capability, credibility, and relevance (CC&R) requirements. These are budgeted to three areas: functional requirements, fidelity requirements, and operating requirements.

Accreditation also must deal with the dual realities of risks and budget.

6.2 Establishing Best Practices

The U. S. Department of Defense (DOD) is perhaps the world's single largest consumer of models and simulations. Because of the unique status of DOD and its unique services, there is a requirement for uniformity of approach to modeling and simulation. The Defense Modeling and Simulation Office (DMSO) is a clearinghouse for information. One role for DMSO is the documentation of practices within the U. S. defense establishment and its customers. To complete our summary we look at the DMSO Best Practices list [6].

6.3 Comments on the DMSO List

Of the techniques listed, the three columns named Informal, Static, and Formal have to be disregarded because the do not include real data input. We have discussed how the formal column should be disregarded since all the elements listed are deductive and not inductive. This leaves of the section of the table listed as Dynamic as usable. Two other techniques also remain: sensitivity analysis and statistical methods.

7 A Look at Detailed Concepts

Generating a validation study is a difficult task. If there is a simulation involved, then this task becomes even more difficult because of the interdisciplinary nature of the product and the team. This interdisciplinary team operates in, what has become a standard term, computational science and engineering. Generically, these teams are constituted from the application organizations, the algorithm developers (both mathematical sciences and computer sciences) and the architecture groups such as hardware, compiler, software developers and so on. What guidance can we give this team?

7.1 Basics

The application group should certainly be expected to come up with the critical parameters of the application, but the algorithm and the architecture may interact in a meaningful way. For example, in [stev-siam] I used a real problem to motivate verifying the arithmetic parameters and their suitability for the task at hand. Another issue that is primarily the purview of the application is the set of critical scenarios that must be simulated before any validation or accreditation decisions can be made. From these scenarios, the team would develop a set of metrics for the decisions to be made. The scenarios also can be tested for both costs and risks. Such evaluation should order the scenario execution.

The encoding of the decision process is the most crucial, since this is where the real observations and the computed predictions meet. The critical development would set the thresholds for the decisions. It is here that the sensitivity analysis plays a role. Sensitivity analysis would identify those values that most effect the outcomes. One must be careful to not over-focus on the model and ignore operating (machine) and physical environmental issues.

The development of the decision criteria will define relationships and from these relationships the metrics that are crucial to the decision. These metrics are often called the metrics of merit. These measures can be categorized into two categories: primary measures and dependent measures. The primary measures are used to compute the dependent measures. The outputs can then be categorized in the same manner. The costs and risks go with the primary outputs. For example, the age-old observation about nuclear winters: it is very costly to annihilate the Earth to validate a nuclear winter simulation.

At the end of the design process, the most important question to be answered is, "Do we have consensual agreement that the process here is acceptable to the community?" The answer to this question

The DMSO Best Practices

Informal Techniques	Static Techniques	Formal Techniques
Audit	Cause-Effect Graphing	Induction
Desk Checking	Control Analysis	Inference
Face Validation	Calling Structure	Logical Deduction
Inspections	Concurrent Process	Calculus (unreadable)
Reviews	Control Flow	Predicate Calculus
Turing Test	State Transition	Predicate Transformations
Walkthroughs	Data Analysis	Proof of Correctness
	Data Dependency	
	Data Flow	
	Fault/Failure Analysis	
	Interface Analysis	
	Model Interface	
	User Interface	
	Semantic Analysis	
	Structural Analysis	
	Symbolic Evaluation	
	Syntax Analysis	
	Traceability Assessment	

	Dynamic Techniques	
Acceptance Testing	Alpha Testing	Assertion Checking
Beta Testing	Bottom Up Testing	Comparison Testing
Compliance Testing	Debugging	Execution Testing
Fault/Failure Insertion Testing	Field Testing	Functional Testing
Graphical Comparisons	Interface Testing	Object-Flow Testing
Partition Testing	Product Testing	Predictive Validation
Product Testing	Regression Testing	Sensitivity Analysis
Special Input Testing	Statistical Techniques	Structural Testing
Sub-Model / Module Testing	Symbolic Debugging	Top-Down Testing
	Visualization / Animation	

Compliance Testing	Execution Testing	Interface Testing
Authorization	Monitoring	Data
Performance	Profiling	Model
Security	Tracing	User Standards
Special Input Testing	White Box Testing	
Boundary Value Equivalence Partitioning	Branch	
Extreme Input	Condition	
Invalid Input	Data Flow	
Real-Time Input	Loop	
Self-Driven Input	Path	
Stress	Statement	
Trace-Driven		

Figure 1: The DOD Best Practices List

Assertion Checking	Comparison Checking	Execution Testing
Fault/Failure Testing	Field Testing	Functional Testing
Predictive Validation	Regression Testing	Sensitivity Analysis
Special Boundary Testing	Statistical Testing	Structural Testing
Sub-Model Testing	Top-Down Testing	

Figure 2: Validation Usable Techniques

revolves around agreement on three issues: (1) functional requirements, (2) fidelity requirements, and (3) operational requirements.

7.1.1 Functional Requirements

The drivers here are the requirements to identify experiments and outputs (both model and experiments) to validate functioning. The functional test plan is a verification plan that ensures the correct functional capability of the simulation to produce the necessary output but *we cannot yet call it 'correct.'* The functional capability of the simulation allows the team to determine the exact parameters for necessary experiments. These experiments are determined by back planning from the validation conditions.

7.1.2 Fidelity Requirements

Fidelity requirements define the degree of correlation between model and reality. There are two separate issues: (1) accuracy and precision and (2) sensitivity analysis. Accuracy and precision are mostly verification issues. Sensitivity analysis is important both for planning and for the final decision. The basic design issues are the budgeting of computational error onto the measures of merit. The budgeting process will indicate how sensitive the system can be.

The term sensitivity analysis is not uniquely defined mathematically, but revolves around investigations of the linear term of the behavior function. How this is conventionally carried out varies from discipline to discipline.

7.1.3 Operating Requirements

Operating requirements might well be considered under another common term: performance benchmarking. This refers to the practical operating constraints under which the simulation code must run. For example, does this application need a supercomputer or can it run on a workstation? Enough has been written in this area to obviate discussion here.

7.2 The Interaction of Decisions and Verification

The development of the scenarios also put demands on the verification process by specifying which functional capabilities must work. Logical verification is the massaging of the assumptions, limitations, known errors, and approximation to determine if the model can reasonably be expected to produce results realistic enough to satisfy the information needs. The functional testing can be used to run some validation values or to set the calibration as well as to test.

Software Quality Assurance. Well, are you going to say anything here?

Verification and validation/accreditation are thus intertwined about

1. The correlated decision analysis and functional tests;
2. Audits: overflow and underflows; array bounds;
3. Verified outputs versus hand calculation or verified alternative; and
4. Testing output as model predictions.

7.3 Putting It All Together

The eventual test plan must describe a theoretical test that could be used to measure the actual parameters required to validate a particular functional entity. To this end, it must address the following issues:

1. Calibration,
2. Sensitivity analysis,
3. Topological analysis,
4. Statistical analysis and preparation of observations and simulation output,
5. Statistical Decisions
6. Analytical significance: the model has to be analytically useful. That is, it must actually be good for the purpose.

8 Preliminary Conclusion

Validation and accreditation are completely separate concepts from verification. In verification we attempt to justify that our simulation fits the requirements implied by the scientific, engineering, or other model. Validation and accreditation seek to determine how well the model itself fits reality. The three processes are intertwined in practice, so it is imperative that there be close coordination between the developers and the testers. Simulations are moving targets since their use in the decision making process of science, engineering, and management is moving. The use of an accredited model today may be inappropriate tomorrow. In the past, the day to day situation has not been changing as quickly as today's. The rapidly changing conditions force us to rely on outputs of simulations, perhaps as the sole predictive tool. Therefore, validation and accreditation methodology should be seen as a critical part of the model; indeed, it must become a critical part of the education of the modelers themselves.

We now turn to an example of some of these concepts. The theory we address is the theory of the æther and its elimination as a viable physical theory by the Michelson-Morley experiment.

Part II
The Michelson-Morley Experiment

9 Introduction

The *luminiferous æther* was posited by Augustin Fresnel (1788–1827) to explain the wave properties of light. By the 1850s, the physics community had hypothesized that the luminiferous æther as present throughout all of space. In 1881, Albert Michelson (1852–1931) performed an experiment, originally performed by Armand Fizeau (1819–1896), using a new apparatus which was to measure the parameters of the ether. The results of the experiment were generally taken as inconclusive due to apparatus problems. Five years later, Michelson and Edward Morley (1838-1923) published the results of a newly run experiment on yet another new apparatus [19]. This had a negative effect for ether theory. There were criticisms of the experiment. A second article appeared in November, 1887 [20]. While there was much more controversy to come, the fate of ether had been sealed. Interestingly enough, the experimental plan of Michelson and Morley was never carried to completion. Only two complete sets of observations were taken out of a year-long protocol. The ether lingered on for almost thirty years; it did not die, but simply disappeared from main-stream physics.

What does this have to do with validation of large-scale simulations? The Michelson-Morley experiment represents perhaps the finest hour of the validation methodology. The experiment had been expected to confirm, not disconfirm, the ether theory. Because the story is so well-known and understood, we can illustrate how we might apply modern ideas of validation.

Section 10 reviews the state of the computational science and engineering (CSE) as I intend to use it. Section 11 develops the necessary machinery from philosophy of science to use logic in validation studies. Section 12 defines measures of uncertainty used in numerical analysis and basic concepts of sensitivity analysis. Section 13 develops the logic of the Michelson-Morley experiment. Section 14 applies our principles to the experiment.

10 Foundations

To understand the ins and outs of validation we must first understand what the scientists — and engineers — are trying to accomplish with their models. For science, a current view is that we seek to develop explanatory theories concerning the observational world. Engineers try to take those explanations and use them to solve problems. For each group, the question is, "How can I be certain that I have selected a correct causal relationship to effect the behavior I am interested in." James Bernoulli (1654–1705) started search for a solution to this question. This has led to two statistical poles: Bayesian and classical statistics. The relationship of the two schools of thought to science is controversial; see [14].

Regardless, our investigation revolves around two basic issues. The first issue is the question of derivation of the experiment itself. This has to do with the approximations used and the numerical effects caused by finite numbers.

The second issue deals with relative uncertainty. That is, given we know there are uncertainties in the observations, are these uncertainties enough to reverse the decision?

The basics of validation methodology were laid down by Aristotle in *Posterior Analytics.* My 1994 paper [25] discussed Aristotle's *Organon* (part of *Posterior Analytics*) [1] and Bacon's *Novum Organon* [2]. The two classical works are the foundation of what we now call the philosophy of science termed the "hypothesis-test" cycle. The computer adds a new concern. I proposed the following principles for the correctness of simulations:

P1 *Physical Exactness.* We strive to identify non-physical (mathematically convenient) assumptions and eliminate them.

P2 *Computability.* We must identify non-computable relationships. Most mathematical relationships in scientific computation turn out to be *approximate.*

P3 *Bounded Errors.* No formulation is acceptable without *a priori* error estimates or *a posteriori* error results.

P3 is clearly a numerical methods/machine arithmetic issue and **P2** is a problem for numerical analysis and computability theory. **P1** represents a new way of thinking for the scientist and engineer. **P1** leads us to consider confirmation as a problem addressed in the philosophy of science.

11 The Logical Machinery

A goal of many researchers has been to develop a system of *deductive* logic that can be used to describe *inductive* science. While there are many problems with this, a related question is reasoning in uncertainty. Another position is to develop a deductive system that bridges the gap between the observational world and the formal world. I describe the Carnap-Hempel system developed over several decades. Such a system would be a good introduction to formal methods into validation.

11.1 A Quick Review of Mathematical Logic

The formalization of reasoning starts with Aristotle *Organon* (part of *Posterior Analytics*) [1], although there was undoubtedly a long history of these investigations before he started. Aristotle's original goal was to have a mechanism for checking arguments during debates. After the end of the "classical period," logic languished until resurrected by, among others, Thomas Acquinas during the Scholastic era. Scholastic logic was used and improved through the 19th Century. Boole and Pierce greatly improved the notation if not the theory of logic. Pierce introduced the quantifiers, "for all" and "there exists" into common usage.

The modern version of logic can be found in any number of good texts; e. g., van Dahlen [30]. Logic is a formal language issue. In formal languages we describe the structure of the language and rules for manipulating the symbols of the language. In mathematical practice, there are fundamental statements that are primitive; these are the axioms. There can be many rules, but the most primitive beginning is to have *modus ponens*:

$$\{A, A \to B\} \Rightarrow B$$

or, in words, "From formulas 'A' and '$A \to B$' one can write 'B' alone." From these ideas alone we can formalize *sentential logical calculus*.

There is no controvery so far. However, when we get to the quantified formulas like $\forall x \exists y x < y$ that we get into trouble. For most situations, however, logic works the way we were taught in undergraduate mathematics.

11.2 Induction versus Deduction

Logic as we discussed above is known as *deductive* logic and this is the logic mathematics. In deductive logic, all the information needed for the proof is given in the hypotheses. In this very important sense, no mathematical proof ever discovers anything new.

In science, however, it is not the case that nothing new is discovered. In fact, in mathematics we talk about *proving theorems* but in science we talk about *establishing laws*. A good early science course should establish in us the following distinction:

1. In mathematics, a theorem is forever. Any datum meeting the restrictions of the hypotheses will meet the restrictions of the conclusion.
2. In science, a datum that meets the restrictions of the hypotheses may *not* meet the restrictions of the conclusion. Therefore, each experiment might fail. In fact inductive systems are contradictions of deductive systems.

These two concepts of mathematical logic and induction versus deduction were at the heart of some areas of philosophy from the mid-1930s to the mid-1970s. From the crisis in foundations in mathematics and logic from the late 1880s until Gödel's famous Incompleteness theorem, much of philosophy and mathematics was looking for *absolute* truth, called the *verificationist* movement. Gödel basically showed this is impossible. Even so, it is useful to have formal languages and formal rules — hence, the formal methods movement in computer science.

11.3 Axioms or Not?

One of the major differences between science and mathematics is the question of axioms. This actually is a question of what is truth — that is, episitemology. Mathematics answers this question in several ways, the most commonly held position is that of derivability. In the derivation scenario, the truth of a statement is shown by deriving an unbroken path from the axioms to the statement in question. This is the familiar axiomatic method from Greek mathematics.

On the other hand, science does not have this view, hewing instead to the Babylonian view. In science, computation is the key issue. The consequence of this is that the logical method developed for mathematics is not useful since there are no axioms. Feynman in [10] describes the different viewpoints and is *must* reading.

Any formal logic that would be useful for science would require the tools to connect the axiomatic world with the computational world.

11.4 The Confirmation Problem

The confirmation problem, as the validation problem is known in the philosophy of science literature, was the focus of the philosophy of science from the mid-1930s until the mid-1960s. This focus arose from events in 19th Century science. Carnap's 1936 and 1937 articles entitled "Testability and Meaning" [3] are the first in a lengthy development in the logical framework of confirmation. Carnap, together with Carl Hempel

(1905–1997) and others, developed "logical framework" for confirmation. In 1977 Frederick Suppe [28] edited the proceedings to the conference *The Structure of Scientific Theories: The Search for Philosophic Understanding of Scientific Theories*, bringing the field into focus. Karl Popper in *The Logic of Scientific Discovery* [23] said that logical theories were set up to fail; the so-called *falsification* theory. Thomas Kuhn's *Structure of Scientific Revolutions* [17] was among the more influential works on the *conduct* of science over confirmation. Kuhn tried to show that theories evolve due to confirmation/disconfirmation. Because many computer scientists and engineers have a formal methods view, I focus on understanding the logical formalisms of Carnap and Hempel.

The goal is to formulate the conduct of science in a first-order logical system, allowing us to investigate the issues of completeness and consistency. There are three basic categories of constructs to consider: (1) the objects and the measurements, (2) science-theoretic definitions and functions, and (3) the logical axioms and operators.

For physics and engineering, we have four objects: (1) idealizations, (2) inferred entities, (3) unobservable objects, and (4) theoretical postulates. These idealized objects are placed in space-time. Our functions are idealized operators over objects and space-time points. So too the relations. The question: how to relate observed objects to theoretical and how to understand idealized relations. That is the subject of this section.

11.5 Reduction Sentences

The basic issue is to relate three concepts: the pre-scientific meaning of a term and the ability to test or confirm the existence of the referent. Did the term *ether* mean anything? It would if we could confirm the existence of the referenced object. To *test* means that one has a procedure that answers a boolean question: if *yes* then we have a confirmatory instance and if *no* a disconfirmatory instance. Since *absolute* confirmation is impossible, we must speak in terms of *degrees of confirmation*. Therefore one confirmatory or disconfirmatory instance is not the whole answer.

We know that scientists and engineers do not use logic in the same way mathematicians and logicians use it. The distinction was drawn as the *analytic-synthetic* distinction[1]. Analytic statements are true or false based on the form of the constituent terms. Synthetic statements are true or false based on how accurately they describe some fragment of the empirical ("real") world. Thus, the scientific terms are synthetic and require a different approach known as a *counterfactuals*. Counterfactuals have several interpretations, but I prefer to think of them as primitive relations on the observations. The link between the observational world and the theoretical world is bridged by *correspondence rules* embodied in *reduction sentences*.

Our inferential structure must include statements that are synthetic as well as analytic. The concept of *consequence* must be defined for both synthetic or analytic consequence. In our first order language, call it L, derivations according to the logical rules are called *L-consequences* and those by scientific rules are called *P-consequences*. Because of this mixed mode of reasoning, we call two sentences S_1 and S_2 *equipollent* if S_1 and S_2 are (L or P) consequences of one another.

The distinction is over how "implies" is translated. Computer science, logic, and mathematics all translate "implies" into *material implication* which is false only when the hypotheses are true and consequent is false. Counterfactuals are *subjunctive implication*: the *false* case simply is uninteresting and does not occur because it is presupposed in the antecedent. Counterfactual inference is model-based.

Then why do we need counterfactuals? Carnap [3] considers the term *soluble in water*: "x is soluble in water" means "Whenever anything x is put into water, x dissolves." Suppose we have the following:

$Q_1(x,t)$	The body x is placed in water at time t.
$Q_2(x,t)$	The body x dissolves at time t.
$Q_3(x,t)$	The body x is soluble in water.

Our natural reaction would be to define Q_3 as

$$Q_3(x) \equiv (\forall t)[Q_1(x,t) \supset Q_2(x,t)]. \qquad (1)$$

[1]This is the major point disputed by philosophers. It is useful for us to accept the distinction.

But this does not capture what we want. Suppose we want $x = b$, a glass brick. Certainly, $Q_3(b)$ is false. But the right hand side of (1) is still true.

$$Q_3(b) \tag{2}$$

$$(\forall t)[Q_1(b,t) \supset Q_2(b,t)] \tag{3}$$

Looking at (3), $Q_1(b,t)$ must always be false and hence (3) must be true. The culprit is the form of (3). A way out of this problem is to introduce the *reduction sentence.*

A reduction sentence has the form

$$Q_1 \supset (Q_2 \supset Q_3)$$

In our example, we would have the form

$$(\forall x)(\forall t)[Q_1(x,t) \supset (Q_2(x,t) \supset Q_3(x,t)]$$

which we could read as "If x is put into water at time t, then, if x is soluble in water, x dissolves at time t."

11.6 Full Statement of the Language Requirement

Space restrictions preclude an in-depth development of the rules for developing languages that fit the Carnap/Hempel mold. See [28, pp. 50-52] for the rules and Hempel [12] for an example. I will only sketch developments here.

There are three types of statements: (1) logical, (2) observational, and (3) theoretical. The essence of the rules is that these three must be carefully woven together. There are two non-logical vocabularies: V_o for constants, functions, and relations for observational purposes and V_T for the theoretical terms of the theory. These yield languages L_o and L_T. The two languages are linked through *correspondence rules.* The correspondence rules can have two forms: (1) reduction sentences which link theoretical definitions to observation and (2) testing procedural statements. The semantics of L are taken from concrete observable events, objects, or times. Therefore, the variables in L_o statements must refer to variables or expressions in the observation language.

Some examples of correspondence rules that might come up in a discussion of Michelson-Morley come to mind. The definitions of `place` and `thing-at` given below are in the observational vocabulary and `distance` and `avg-vel-between` can be placed in the theoretical vocabulary. Clearly, *ether* is a theoretical term.

$$\begin{aligned}\texttt{place}(\bar{x}) \wedge \texttt{place}(\bar{y}) &\rightarrow \\ (\bar{y} - \bar{x})\ \texttt{finite} &\rightarrow \\ \texttt{distance}(\bar{x},\bar{y}) &= \|\bar{y} - \bar{x}\|\end{aligned}$$

$$\begin{aligned}\texttt{thing-at}(a,\bar{x},t) \wedge \texttt{thing-at}(a,\bar{y},t') &\rightarrow \\ (t - t' \neq 0 &\rightarrow \\ \texttt{avg-vel-between}(\bar{x},\bar{y}) &= (\bar{y} - \bar{x})/(t' - t).\end{aligned}$$

12 Uncertainty, Sensitivity, and Numerical Issues

12.1 Uncertainty and Sensitivity

Some views of validation seem to focus on quantification of, and reasoning about, uncertainty as the key issue in validation. This is not just classical statistics since the statistical community has two poles: Bayesian and classical statistics. The different types can lead to different conclusions. Fortunately, we only need to quantify the uncertainty.

One measure of uncertainty is that of relative error. Let $F(x)$ be an arbitrary, once continuous function. The *relative error* of F with respect to x is defined by

$$\text{rel. err.} = \frac{dF/dx}{F}. \tag{4}$$

Since F could be multivariate, dF is the total derivative.

12.2 Sensitivity Analysis

In any theory we would expect to see mathematical relationships among the objects of the theory. When subject to observations, observational errors are introduced. A natural question would be, "How large can these observational errors be and still not effect the decision?" In other words, how *sensitive* is our result to errors in the observations. This is formulated by asking how does change in the error of a parameter change the function's value. Using our arbitrary $F(x)$ again, we want to study how $F(x+\Delta x)$ changes with respect to the indvidual Δx_i's.

$$\text{sensitivity of } \Delta x_i = \left.\frac{\partial F}{\partial \Delta x_i}\right|_{\Delta x_j = 0} \qquad \text{where } j \neq i.$$

12.3 Error Analysis

We must insure that the principles (see Section 10) of computational science and engineering are enforced. There are two types of errors: 1) scientific errors, denoted δ_s and 2) numerical errors from approximations and finite arithmetic δ_a. By definition, δ_s is unknown. Let F be the function received from the science and suppose (as a gross simplification) F is linear.

Our measure of merit for δ_a is relative error.

$$\text{rel. err.} = \frac{|x - \hat{x}|}{x} = \frac{dF/dx}{F(x)}.$$

where x is the *true value* and $\hat{x}$ is the *computed value.*

For the *forward problem* where F and x are known and we want $y = F(x)$ we can determine the relative error in y easily:

$$\begin{aligned}
(F + \delta_a F)(x + \delta_a x) &= y + \delta_a y && (5)\\
F(x) + F(\delta_a x) + \delta_a F(x) + \delta_a F(\delta_a x) &= y + \delta_a y \\
\frac{F(\delta_a x)}{F(x)} + \frac{\delta_a F(x)}{F(x)} + \frac{\delta_a F(\delta_a x)}{F(x)} &= \frac{\delta_a y}{y} && (6)
\end{aligned}$$

The first term in Eq. 6 measures the *conditioning*, the second measures the *stability* of the method and the third term measures the arithmetic errors.

1. If the conditioning number is large, then the function is extremely difficult to compute. Therefore, a different formulation of the problem is needed.

2. If the conditioning number is small, the function is easy to compute. Stability now comes into play. Stability is the measure of how approximation errors are propagated. If the stability number is large, the method probably will not be satisfactory; find another method.

3. If both conditioning and stabilty are good, then the errors due to finite arithmetic determine how accurate the computed value is.

13 Illustrative Development of Æther

This section describes the basic tenets of the Fresnel theory. This section requires only an elementary understanding of fluids and the idea of interference patterns.

To understand Fresnel's theory, consider the æther as a fluid and light as a swimmer. If the swimmer swims back and forth between the banks (at right angles to the current), then she will be carried downstream on the two trips. When she swims with (or against) the current the speed of the river is added to (or subtracted from) to the swimmer's speed. Therefore, there is an apparent difference is speed. This speed difference would appear as an intereference pattern when this analogy is mapped over into light. The patterns can be measured quite accurately and the speed difference (the current) reflects the effect of the æther.

This section combines the developments presented in Michelson's and Morley's 1886 and 1887 papers.

13.1 The Formulas

Let the *index of refraction* n be the ratio of the speed of light in the external ether (v_e) to that internal (v) to the ether. This is further defined as the inverse ratio of the square root of the densities. Suppose the density of ether is 1 outside the box but $1+\Delta$ inside a small volume (because it is moving and compressing the ether):

$$n = \frac{v_e}{v} = \frac{\sqrt{1+\Delta}}{\sqrt{1}}. \tag{7}$$

We use these facts to develop the theory behind the experiment.

The question now is to determine what velocity of ether in the prism is needed to give the observed results. The speed of the ether must be

$$x = \frac{n^2-1}{n^2}. \tag{8}$$

Logical Development. In terms of our observational/theoretical distinction, this entire section is done in the theoretical language. In fact, we run into trouble as soon as we try to use our observational/theoretical model.

The value of x is not connected to any observable measurement: the observable v is eliminated. We cannot measure the density difference between air and ether. The ratio we can measure is given by $v = v/V$. This is the key to the refutation.

We now derive the equations to support the Michelson-Morley experiment. The definitions of the variables and the assumptions are given in Figure 5. We need to find the time difference to predict the fringe interference pattern difference. The experimental setup is shown in Figure 4.

The times for each leg are $T = D/(V-v)$ and $T_1 = D/(V+v)$. The round-trip time is clearly

$$T + T_1 = 2D\frac{V}{V^2-v^2}.$$

N. B. In the 1887 paper, we have a illustration of breaking principle **P1** given in Section 10. "... and the distance traveled in this time is $2DV^2/(V^2-v^2) = 2D(1+v^2/V^2)$, *neglecting terms of the fourth order*(italics mine)." Having now crossed the line, the report's authors must now continually refer to both the exact and approximate derivation.

The arm moving perpendicular to the path of the earth is actually undergoing a Lorenz contraction (we could also compute this from first principles). Therefore, the light on the other path must travel $2D\sqrt{1+v^2/V^2}$. To get the approximation compatible with the above, we expand this by the binomial expansion $(a+b)^n$, with $a = 1$ and $n = 1/2$. We obtain a first-order approximation for the length as

$$2D(1+v^2/2V^2).$$

Subtracting the two distances we get Dv^2/V^2. The distance divided by the wavelength of the light is equal to an integral number of wave length; that is, the bands occur at those intervals. The experiment was run using a sodium lamp. Sodium light emits at 5,900Å or $5.9 \times 10^{-7}m$.

One experiment instance is to be carried out by first setting up the apparatus one way, then rotating it 90 degrees. Therefore the total difference is twice the above, or (using the approximations), $2Dv^2/V^2$. The protocol called for one reading at noon and the other at night.

13.2 The Experimental Results

As reported in the 1887 paper, observations were taken July 8–12, 1887. Measurements were taken at Noon, then the apparatus was rotated clockwise for the evening measurements. Quoting from the report:

> "....It seems fair to conclude from the figure that if there is any displacement due to the relative motion of the earth and luminiferous ether, this cannot be much greater than 0.01 of the distance between fringes.
>
> Consider the motion of the earth in its orbit only, this displacement should be $2Dv^2/V^2 = 2D \times 10^{-8}$. The distance D was about 11 meters, or 2×10^7 wave-lengths of yellow [sodium]

> light; hence the displacement to be expected was 0.4 of a fringe. The actual displacement was certainly less than the twentieth part of this, and probably less than the fortieth part. But since the displacement is proportional to the square of the velocity, the relative velocity of the earth and the ether is probably less than one sixth the earth's orbital velocity, and certainly less than one-fourth.
>
> ...It appears ... if there be any relative motion between the earth and the luminiferous ether, it must be small; quite small enough entirely to refute Fresnel's explanation of aberration...." [29, p280–281]

The distances were to be measured using a graduated set screw which was reported to be able to measure 0.02 wavelengths.

14 Applying the Principles

14.1 Can We Compute Accurately?

We have already commented that the derivation makes gratuitous use of unwarranted approximations: the time difference can be computed without intermediate approximations. Without the approximations, the difference between the time along the orbit and the time perpendicular is

$$\frac{2D(\rho V^2 + \rho v^2 + V)}{-V^2 + v^2} \tag{9}$$

where $\rho = \sqrt{1 + (v/V)^2}$.

How well can we compute this? One measure is known as a *condition number*. The condition number is defined for a function $f(x)$ as

$$c = \frac{x}{f(x)} \frac{df}{dx}$$

If the condition number is "small" then the value of $f(x)$ is "easy" to calculate accurately. Since we have two variables, the condition number is a function of both. Using `Maple`, we find that (after a lot of work), that $c = 3 \times 10^{-9}$ using the values in Figure 5. That means we should be able to compute the time difference very accurately.

Using the definition of relative error, and again using `Maple`, we get a relative error of 3×10^{-13}. Therefore, the computed difference should differ from the real one in the 13th decimal place.

We can compute the values accurately enough to guarantee that observational errors are the dominate errors in the results.

14.2 The Logical Case On Michelson-Morley

For convenience, we copy Eq. 9 to the below and label the values as observational:

$$\begin{aligned} T_i &= \frac{2D_o(\rho V_o^2 + \rho v_o^2 + V_o)}{-V_o^2 + v_o^2} && (10) \\ \rho &= \sqrt{1 + (v_o/V_o)^2} && (11) \\ m_o\lambda_o &= T_i && (12) \end{aligned}$$

where T_i is the time difference with the auxiliary definition for ρ. Eq. 12 is the equation relating the spacing of the interference fringes to the time difference. m is actually what was calculated as 0.4 in the report on results. The report says that D is "approximately 11 meters" or 2×10^7 times the wavelength of yellow light. Using the values from Fig 5 and that definition we get that D is 11.8 meters. Solving Eq. 12 we find that the value of m_t should be 0.3971. Taking "certainly less than a fortieth" as 1/40, $m_o = m_t/40 = 0.00993$.

We can solve the *theoretical* T_i equation for v_t. If we use the observed values and the theoretical m value, we compute that

$$v_t = 29789.999\underline{99} \text{ with relative error } 0.3 \times 10^{-9}$$

(with the last two digits in error) but the observed speed in orbit would have to be

$$v_o = 4710 \text{ with relative error } 0.8 \times 10^0.$$

The zero exponent in the relative error says that the real and computed value of v_o do not even agree in magnitude.

We still cannot make a connection from the observational and the theoretical unless we accept a correspondence rule that allows

$$v_o = v_t.$$

How does this logically generate the contradiction to Fresnel's axiom? We have both v and V by measuring the speed of light in a vacuum. The value is given as $V = 2.997 \times 10^8$. If we allow such a rule then we can say that $v_o \neq v_t$ and therefore we have a disconfirming experiment.

14.3 Sensitivity and Uncertainty

In Eq. 9, there are no arbitrary constants and just three variables: D, V, and v. The sensitivity values are

Variable	Value
D	2.0000
V	3.1413×10^{-15}
v	-2.9157×10^{-11}

We can conclude that we are very *un*sensitive since the distance is the most sensitive but the easiest to measure: we should easily be able to measure D to 10^{-3}. The other values are many orders of magnitude smaller, especially at the reported errors.

The relative error is 1.517×10^{-16} as calculated in `Maple` with 20 digits of precision.

These numerical considerations indicate that there is low uncertainty about the value. The sensitivity coefficients indicate that the value is quite stable within the errors stated.

15 Lessons for Simulations

This section draws on the lessons learned above. Our task is to apply these lessons to simulation development.

15.1 Computer Simulation, Verification, and Validation

The pragmatic development of a computer simulation begins with the mathematical statements of the model. While most developments treat pencil-and-paper mathematics as numerical mathematics, one must usually rewrite the mathematics. In [26], I discuss a more comprehensive view of the various forms of mathematics. The short version is (1) scientists work in terms of the mathematical analysis of infinite precision, (2) numerical analysis works in terms of convergence on rational numbers, (3) numerical methods works in terms of an idealized floating point that may or may not exist, and (4) the machine manufacturers do what they can within broad guidelines.

The interplay among the various sources and levels of mathematics can cause many problems. Because of the possibilities of error in a simulation, three terms have evolved to denote activities used to guarantee correctness:

1. *Verification.* Verification is the process by which we ensure that the code as developed is in accordance with the programming specificiations; *i. e.*, the code correctly represents the specificiations (model).

2. *Validation.* Validation is the process whose goal by which we ensure that the model accurately reflects Nature; *i. e.*, the mathematics correctly represents the real system in the real world.

3. *Accreditation.* Accreditation is a process similar to validation. Validation is a never ending process due to the nature of science. Accreditation seeks to validate a simulation for a specific situation.

Several professional engineering societies have taken steps to make their members aware of these steps by developing standards. The American Institute of Aeronautics and Astronautics has worked especially hard in this area and has published "Guide to Verficiation and Validation of Computational Fluid Dynamics Simulations". This is well worth reading regardless of the focus of a simulation project. The Defense Modeling and Simulation Office of the U. S. Department of Defense has extensive resources relating to best practices in verification, validation, and accreditation (or by the commonnly used acronym VV&A).

15.2 Specific Comments on Validation and Verification

A predominant view of validation held by many workers in the U. S. is that validation is about dealing with intrinsic uncertainty. By *intrinsic uncertainty* I mean the uncertainty left over after verification is completed. Based on our simple example, what is the nature of this uncertainty?

1. Almost surely, the observations have statistical uncertainty when taken. But there could be systemic and instrument errors as well. For example, traffic counters can intermittently fail with no telltale traces of failure. This means that a simulation could be correct in the larger sense, but fail to validate since the validation data is corrupt.

2. The validation, verification, and development groups should work together throughout the life cycle. For example, the recent loss of the *Mars Polar Explorer* over failure to convert MKS to British units shows how difficult it is to ferret out mistakes.

3. Numerical errors can mask any information in the simulation. This is particularly possible for long running simulations where numerical cancellation can be a problem typified by physics codes that can run many days.

4. The scientific explanation may be inappropriate for the problem at hand. This happens when the regime of parameters is outside the "trusted" set during extrapolation. This can occur when simulations are used outside the validated parameter space. This problems is seen as particularly difficult in large engineering simulations.

5. The problem is so sensitive to initial conditions (like chaos) that we are unable to accurately simulate to the precision required to make a decision. Here, the simulation gives unwarranted credibility.

15.3 The Formal Methods Possibilities

The Carnap-Hempel logic provides an excellent opportunity for development of formal methods. The logical language itself is standard first order predicate calculus. The new element would be making it simple to write the counterfactual experssions and then integrating them into a whole. By the very nature of the process, it would seem quite feasible to use a logic programming language to develop a simulation writing environment.

15.4 Simulation Code Development

Development of generalized simulation writing environments requires a long lead time. This lead time might be justified for mission critical projects. But not all such projects may have the luxury to develop the logical system describing their project. This means that the simulation code developers must take up the slack.

In [26], I discussed many of the attributes needed to ensure quality simulations. Some very basic observations are:

1. Good practice matters. Unfortunately, academics don't teach it and organizations can't enforce it. Software engineering practice does not address computational science and engineering.

2. Speed must become subordinate to correctness.

3. Detailed specifications, quality assurance procedures, and formal testing are not enough. Uncertainty by using less-well-defined algorithms is several times worse than using formal mathematical definitions.

4. Double precision does not solve problems of unstable codes or ill-conditioned problems.
5. Paradigm shifts in language or formal methods do not appear to automatically solve the problem. But comprehensive and objective testing, formal methods, and multiple versions may be helpful.
6. Safe subsets for languages are very important.
7. Construction and testing of static code fault finders are needed to
   (a) Find formally undefined behaviors in languages and systems.
   (b) Help enforce known standards.
   (c) Screen out well-defined behaviors we know we should not use.
   (d) Help assess quality.
8. Documentation is not a panacea.
9. Software engineering metrics and processes, as currently practiced, measure nothing of interest except those in Chapter 8 of Fenton and Pfleeger [9].

These problems still exist and are ubiquiteous in systems. The point about languages goes beyond standard languages, reaching into the specialized language systems used for modeling and simulation.

15.5 The Control of Complexity in Simulations

The control of complexity of a model or simulation is an important property not seen in the Michelson-Morley experiment. The experiment was very simple, and as we have seen, not particularly prone to either observational or apparatus errors. Such is not the case with a large simulation running on several processors.

Charles Perrow, in *Normal Accidents* [22], advances a thesis that there are two causes for accidents such as Three Mile Island: mind-boggling component complexity and mind-boggling interconnectivity of components. We should expect there to be a point at which the complexity of the process overwhelms our cognitive ability to understand. I have called this *cognitive complexity* in [26]. The basic cure for cognitive complexity is tracability: the ability to trace model concepts into the the simulation. In a large simulation, tracability is not possible without help from language tools that tie documentation to code. A beginning step would be the requirement that simulations use literate programming techniques [18]. However, documentation is not a cure-all and much more research is needed in this area.

15.6 Sensitivity

The study of sensitivity has a huge literature base in its own right. Sensitivity can be easily approached from two standpoints: control theory and operations research. In fact, these two areas are also the foundations of many simulations studies. The problem is the automatic development of sensitivity studies for simulations. The problem is exacerbated by the long runtimes of many simulations in use in industry. It is not uncommon to have a simulation run several days *per output set.*

15.7 Uncertainty and Statistics

Finally, the defining context of validation is uncertainty. Uncertainty transcends simulation type, being just as important in discrete event simulations as it is for continuous simulations. Any simulation of a stochastic system already has a great deal of statistical processing associated with the data, but this should not fool one into thinking that the uncertainty is under control. Just how robust are those statistics? How sensitive are the simulations to the actual distribution?

Among statisticians and philosophers of both mathematics and science there is a healthy debate about whether classical statistics or Bayesian statistics is correct. Practitioners have a tendancy to ignore such arguments as “academic.” Unfortunately, these are not necessarily ignorable. The reason is that when all is said and done, validation is a judgement call and the judgement used is the judgement standard set by the entire intellectual community. For example, one might be roundly criticised for a judgement rendered outside *best practice*. Indeed, the Defense Modeling and Simulation Office of the U. S. Department of Defense has an information site at `www.dmso.mil` which points to the *VV&A Recommended Practices Guide (RPG).* A group choosing to ignore the RPG does so at its own peril.

The artificial intelligence community is the proximal source of a vast literature on reasoning with uncertainty. It is not clear exactly how this subject can be used in validation studies. It would appear to be another area ripe for formal methods.

16 Conclusions

We have presented a case study analysis of the MichelsonMorley experiment.

We have presented two concepts for the validation of scientific work. One concept, the reasoning aspect, has been formalized by 20th Century philosophers of science. The reasoning aspect requires accommodation to the reasoning modes of the scientists and the careful integration of the scientific reasoning into mathematical reasoning. We saw the interplay of the observational model and the theoretical model.

The second concept concerning the ability to compute values needed in the observational and theoretical world use the ideas of conditioning, *mathematical* error, and convergence. We did not explore the machine's effect through *round-off* error. We could do so by adding another term δ_r to Eq. 5.

The third and fourth concepts explore uncertainty and sensitivity of the final answers with respect to the errors in the problem. We found the computed values remarkably insensitive to the errors.

The task of finding workable validation techniques has hardly begun. While numerical analysis has worked out many procedures for numerical computations, most of the results are *a posteriori*. Understanding — indeed including — scientific reasoning in the mathematically intense theoretical model and how to logically link the observation and theoretical models is a wide open question. I present an expansion of these ideas in [26].

The question of how the lessons of the Michelson-Morley experiment can be used in simulations was partially addressed. Validation is such a direct concern that many standard ideas from areas like control theory and operations research have not been properly evaluated. I did suggest several promising approaches for possible formal methods.

Perhaps the most interesting lesson of all is that the physicists I talked look for *consistency* among many different values. Such scientists consider ranges of values more than a single, "correct" number. In fact, they may not be looking for a number at all: Some histories of science seem to imply that the physics world was looking for a way out of the æther for relativity; Michelson and Morley provided the exit.

References

[1] Aristotle. *Prior and Posterior Analytics.* Clarendon Press, Oxford, 1949.

[2] Francis Bacon. *Novum Organum.* P. F. Collier, NY, NY, 1902.

[3] R. Carnap. Testablity and meaning. *Philosophy of Science*, 3 and 4, 1936–1937.

[4] Rudolf Carnap. *Foundations of Logic and Mathematics.* Out of print, 1937.

[5] Rudolf Carnap. *Studies in Inductive Logic and Probability.* Out of print, 1971. 2 volumes.

[6] DMSO. Vv&a recommended best practices guide. US Department of Defense, November 1996.

[7] U. S. DOD. Dodd 5000.59: Modeling and simulation (m&s) management. DOD Directive, 1994.

[8] U. S. DOD. Instructions 5000.61: Instructions on vv&a. DOD Instructions, April 1996.

[9] N. E. Fenton and Shari Lawrence Pfleeger. *Software Metrics: A Rigorous and Practical Approach.* PWS Publishing Company, 2 edition, 1997.

[10] Richard P. Feynman. *The Character of Natural Law.* MIT Press, 1965. QC28.F5 1968.

[11] J. A. Goguen. Sheaf semantics for concurrent interacting objects. *Mathematical Structures in Computer Science*, 2:159–191, 1991.

[12] C. G. Hempel. A purely syntactical definition of confirmation. *J. of Symbolic Logic*, 8, 1943.

[13] Mary Hesse. *The Structure of Scientific Inference*. Macmillan, 1976.

[14] Colin Howson and Peter Urbach. *Scientific Reasoning: The Baysian Approach*. Open Court, La Salle, IL, 1989.

[15] Philip Kitcher and Wesley C. Salmon. *Scientific Explanation*, volume XIII of *Minnesota Studies in the Philosophy of Science*. University of Minnesota Press, 1989.

[16] George J. Klir. *Archtecture of Systems Problem Solving*. Plenum Press, 1985. ISBN 0-306-42867-3. QA295.K55 1985.

[17] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL, 2 edition, 1970.

[18] Silvio Levy and Donald E. Knuth. *The Cweb System of Structured Documentation/*. Addison-Wesley Pub Co, 1994.

[19] Albert Michelson and Edward Morley. Influence of motion of the medium on the velocity of light. *American Journal of Science*, 31:380, May 1886. 3d series.

[20] Albert Michelson and Edward Morley. On the relative motion of the earth and the luminiferous ether. *American Journal of Science*, 34:333–336, Nov 1887.

[21] OSD. Joint modeling and simulation system (jmass) joint initial requirements document (jird). U. S. DOD, February 1997.

[22] Charles Perrow. *Normal Accidents*. Basic Books, 1984.

[23] Karl Popper. *Logik der Forschung/The Logic of Scientific Discovery*. Basic Books, New York, 1935 (1959).

[24] Nicholas Rescher. *Scientific Explanation*. Free Press, New York, 1970. Q175.R394.

[25] D. E. Stevenson. Science, computational science, and computer science: At a crossroads. *Comm. ACM*, 37(12):85–96, 1994.

[26] D. E. Stevenson. A critical look at quality in large-scale simulations. *IEEE Computing in Science and Engineering*, 1(3):53–63, May/June 1999.

[27] D. E. Stevenson. A foundations of validation: The michelson-morley experiment. In *The Proceedings of the European Simulation Multiconference 1999, Warsaw, Poland, June 1-4, 1999*, pages 269–275, 1999.

[28] Frederick Suppe. Introduction and afterword. In Frederick Suppe, editor, *The Structure of Scientific Theories: The Search for Philosophic Understanding of Scientific Theories*, pages 3–244, 617–730, Urbana, IL, 1977. University of Illinois Press.

[29] Loyd S. Swenson, Jr. *The Ethereal Aether*. Univ. of Texas Press, Austin, TX, 1972.

[30] D. van Dalen. *Logic and structure*. Springer-Verlag, 1994.

[31] G. A. von Blaaw. Computer architecture. *Electoinishe Rechanlagen*, 14(4):154–159, 1972.

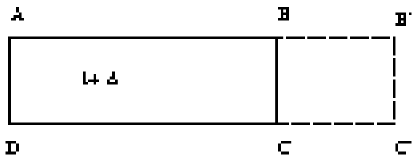

Figure 3: Diagram for Eisenlohr Development

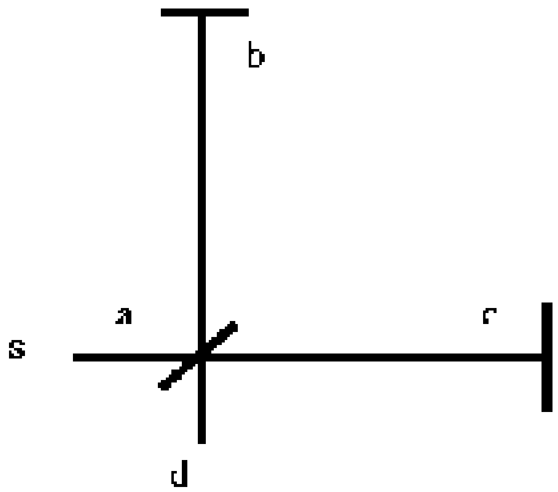

Figure 4: Schematic of Michelson-Morley Experiment

Variable Dictionary

Variable Name	Definition
v_e	The velocity of ether outside the prism
V	Velocity of light in the medium
v	Velocity of the earth with respect to the ether; i.e, in orbit.
D	Distance between the two points of the apparatus ab or ac
T	Time for light to go in direction a to c
T_1	Time for light to go in direction c to a

Note: See Figure 4 for diagram.

Value Correspondence Rules

$V = (299.8524 \pm 0.0790105478190518) \times 10^6 m/s$	NIST Mickelson Project
$v = 2.979 \times 10k/s = 2.979 \times 10^4 m/s$	NSSDC of NASA
$\lambda = 5.9 \times 10^{-7} m$	NIST
$v/V = 0.99 \times 10^{-4}$	
$(v/V)^2 = 0.99 \times 10^{-8}$	

Figure 5: Variable Definitions, Assumptions, and Values