A historical review of the development of verification and validation theories for simulation models

Article *in* International Journal of Modeling Simulation and Scientific Computing · January 2017

DOI: 10.1142/S1793962317300011

Phillip J. Durst
Assured Information Security

Cindy L. Bethel
Mississippi State University

Derek Anderson
University of Missouri

A HISTORICAL REVIEW OF THE DEVELOPMENT OF VERIFICATION AND VALIDATION THEORIES FOR SIMULATION MODELS

PHILLIP J. DURST
U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Road
Vicksburg, MS 39180, USA
phillip.j.durst@erdc.dren.mil

CINDY L. BETHEL
Mississippi State University, Department of Computer Science and Engineering
Mississippi State, MS, 39762, USA
cbethel@cse.msstate.edu

DEREK T. ANDERSON
Mississippi State University, Department of Electrical and Computer Engineering
Mississippi State, MS, 39762, USA
anderson@ece.msstate.edu

Received (Day Month Year)
Accepted (Day Month Year)

Modeling and simulation (M&S) play a critical role in both engineering and basic research processes. Computer-based models have existed since the 1950s, and those early models have given way to the more complex computational and physics-based simulations used today. As such, a great deal of research has been done to establish what level of trust should be given to simulation outputs and how to verify and validate the models used in these simulations. This paper presents an overview of the theoretical work done to date defining formal definitions for, and methods of, verification and validation of computer models. Simulation models are divided into three broad categories: (1) analytical and simulation models, (2) computational and physics-based models, and (3) simulations of autonomous systems. This article presents the unique theories and methods developed to address verification and validation (V&V) of these models. Also presented are the current problems in the theoretical field of V&V for models as simulations move from single system models and simulations to more complex simulation tools. In particular, this article highlights the lack of agreed-upon methods for V&V of simulations of autonomous systems, such as an autonomous unmanned vehicle, and proposes some next steps needed to address this problem.

Keywords: Verification and Validation, Simulation Models, Computational Models, Physics-Based Simulations, Autonomous Systems

1. Introduction

Computer simulations of physical systems and processes became a critical part of science and engineering research in the mid to late 1950s as computer systems became smaller, faster, and more readily available. Early concepts and theories of simulations and their applications were first introduced in seminal literature[31]

[96] [52], and these concepts were further refined in MacDougall[50] and formalized in an article by McLeod[53], and later Rothenberg[78]. Early simulations were readily adopted into a wide range of research areas, including geology[37], human anatomy and biology[91] [3], and electronics[90]. These early studies and simulation applications laid the foundation for the prolific use of simulations observed today.

The use of simulations for both basic research and applied engineering increased exponentially as computers continued to become ever smaller, cheaper, and faster. In fact, modern simulations consist of both empirical/effects-based modeling to numerical, computational, and physics-based models, with early works in computational models focused on fluid flow problems[28] [17]. More attention is given to the computational fluid dynamics (CFD) simulations in Section 3.

For practical engineering problems, simulations are of value only if the results of these simulations can be trusted. Since the inception of M&S, the concept of verification and validation (V&V) of simulation models has been defined and formalized using theoretical frameworks and methods for V&V. These methods were developed primarily for empirical process modeling; however with the increased use of physics-based computational models, a qualitative, experimental-based method for V&V became the standard. If the simulation results more-or-less matched the experimental results, the simulation was considered reasonably validated and useful for more complex test cases, such as shown in Bonani et al.[21].

While a firm theory and methodology for the verification and validation (V&V) of the different types of models used today exists, there is one major area where these methods fall short: simulations of autonomous robots. Simulations of autonomous systems do not involve a single model or even a coupled system of two models (discussed in Section 3) but rather are simulation environments: software tools involving multiple models working in concert to simulate the outputs or behaviors of complex systems. In fact, no theory has been agreed upon to date for how V&V of simulation environments should be performed. While each individual model can be V&Ved, this does not necessarily guarantee that the overall simulation outputs will be accurate.

To that end, the goal of this paper is to provide a thorough background on the theory of V&V of simulations, including analytical and simulation models (Section 2), computational and physics-based models (Section 3), and simulations of autonomous systems (Section 4). Additionally, Section 4 discusses the limited V&V methods for simulating autonomous agents proposed to date and highlights the lack of research related to validation of simulations of autonomous robotic systems. Section 5 concludes this work by presenting the need for a new theoretical framework of V&V for simulation environments for autonomous robotic systems.

Several important terms are used throughout this paper, and it is important to define them. For the purposes of this paper, V&V "theories" are conceptual frameworks for how V&V should be performed as opposed to V&V "methods", which are the means by which V&V is implemented. In other words, V&V theories are the overarching concept of how V&V methods should be applied. As for model and simulation, the definitions provided by Engel[30] are used:

- Model: A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

- Simulation: A method for implementing a model over time.

Also for this paper, in order to logically present the V&V literature, models are broken down into three broad categories: (1) analytical and simulation models, (2) computational models and physics-based models, and (3) models of autonomous agents. By analytical and simulation models, this paper refers primarily to event-driven discrete probability-based models. Likewise, computational models and physics-based models refer to time-stepped continuous physics-based models with the key difference being computational models have a strict focus on solving the complex equations governing a system rather than simulating a general phenomenon.

In addition to the current work, several outstanding reviews of the V&V literature can be found. These include articles by Youngblood[104], Balci[12], and Petty[73]. The key difference between these and the current work is the focus on autonomous systems. All of these references provide an excellent review and details on current V&V techniques, they do not provide a history of the development of these techniques. Likewise, they do not address the lack of methods for simulations of autonomous entities, and robots in particular. Therefore, this paper seeks to fill in the gaps from these articles and point out the shortcomings in the current techniques for addressing simulations of autonomous robots.

2. Verification and Validation of Analytical and Simulation Models

A substantial amount of research has been performed to define methods for verifying and validating computer models. The bulk of the early research performed focuses on analytic/discrete event models and simulation models. These types of models are most often models of processes, such as economic forecasts or manufacturing outputs, versus physical phenomena, such as fluid flow.

2.1. *Early Efforts to Define Verification and Validation for Analytical Models*

The first models developed were exclusively analytical models which used computers to solve discrete event models, both equation and logic-based, as described in Schiber[88], which represent systems or problem entities. As such, the earliest adopters of computer simulations were not physical scientists and engineers but rather political/behavioral scientists[22] and economists[2, 33].

Coming from the political sciences, an early definition for what constitutes a simulation appears in Hermann[38]. Hermann defines a simulation as "the partial representation of some independent system" and he states that the primary use of simulations is to increase the "understanding of the system it intends to copy"[38].

The earliest work on defining verification of computer simulations was that of the economist Fritz Machlup[51]. Machlup provides the first formalized definition of the concept of verification as:

> "Verification in research and analysis may refer to many things, including the correctness of mathematical and logical arguments, the applicability of formulas and equations, ..., the reliability and exactness of observations, the reproducibility of experiments, the explanatory or predictive value of generalizations"[51].

Verification was further defined by Naylor and Finger[57]. They draw heavily from the previous work in model V&V done by Machlup et al.[51] in the field of economics and apply them to industrial models. They propose the idea of a validation score based on "the 'goodness of fit' of time series generated by computer models to observed historical series"[57].

Expanding upon this work, V&V were more formally defined by Fisher and Kiviat: "Verification ... determines whether a model actually behaves as an experimenter assumes it does; [and] validation ... tests whether the model reasonably approximates a real system"[32]. They further state that "Validation tests the agreement between the behavior of a simulation model and the observed behavior of a real system"[32]. Fisher and Kiviat's work is further developed by Horn[97], wherein he defines validation as "the act of increasing to an acceptable level the confidence that an inference about a simulated process is correct for the actual process"[97].

Mihram also sought to define V&V[56 55]. In[56], Mihram states that "the credibility of any modeling effort rests on the concrete demonstrations that the resulting model represents reality." Indeed, the idea of credibility is common throughout much of the early efforts to define validity.

To summarize, the earliest work in V&V of analytic models provided the following definitions for verification and validation. Verification of a model ensures that the model works as expected. Verification is a software-level process that does not necessarily require information about the models outputs; it is essentially a check that the model equations are implemented in software correctly. Validation is the testing of model outputs against experimental data to see if the model provides accurate outputs. Together, V&V of a model insures that the model works correctly and provides accurate and meaningful outputs. Furthermore, the ultimate goal of validation is to establish trust/credibility in a model such that the model can be used to predict problem entity behaviors in untested cases. These definitions are common amongst all future model and simulation types and will be used throughout the following sections.

2.2. *Verification and Validation of Simulation Models*

As computers became more powerful and prevalent, models moved from analytical models to simulation models and the simulation of a system rather than the solving of equations using computers. Many early simulation models were adopted within the field of operations research[47 26 27]. Analytical models prove largely inapplicable to the operational research field, which requires a simulation of a system or problem entity versus an analytical equation representing the system or problem entity. One key study on the transition from analytical to simulation models can be found in Shanthikumar and Sargent[89]. As this transition occurred, the ideas of V&V had to be redefined.

2.2.1. *Early Efforts and Definitions*

One of the first works that defined the transition from analytical to simulation models was that of Ignall et al.[42]. They argued that "the conditions assumed by solvable analytic models do not hold in the real world" and they go on to suggest the validation of analytical models against simulations. They do not, however, offer a solution for lthe validation of simulation models themselves.

One early method of validating simulation results is found in Kheir et al.[43], which offers a geometric scoring model based on the differences in simulated and live testing missile trajectories and target miss/hit distances. The idea of scoring metrics or fixed criteria is also found in Schellenberger[86]. While these works build upon earlier works (e.g., Naylor and Finger[57]) and attempt to provide a solid method of validating simulation models, they do not provide a firm theoretical basis for how validation should be performed.

One popular concept that arose out of simulation models was that of a simulation meta-model. A meta-model represents the generalization of a simulation model back into an analytical model. This generalization most often was in the form of a regression analysis of simulation results across "a number of situations (different parameters, variables, structure relationships) ... to arrive at an understanding of the system"[44].

The major benefit of meta-models at the time was that they were easily validated. A regression meta-model could be compared to real world results to check whether, statistically, the meta-model was in strong agreement with experimental data. In Friedman and Friedman[34], several approaches to meta-model validation were proposed.

2.2.2. *Formal Theories for Verification and Validation of Simulation Models*

The earliest work to establish a well-defined methodology for V&V of simulation models was performed by Naylor[58]. After some time, his work was followed by Schlesinger et al.[87]. Schlesinger defined validation as the "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model"[87]. For a time, it was the most commonly accepted definition for model validation.

The next formal effort in defining V&V was that of Őren and Tuncer, who defined the acceptability of simulation results. While focused more on acceptability of simulation studies in regard to not only result validity but cost and time concerns, Őren and Tuncer provided some insight into how to validate simulation results[67 68]. In these works, they defined validation as the "assessment of a model with respect to the real system"[68].

In later works, Sargent provided what is currently one of the most widely accepted definitions for V&V of simulation models. First proposed in Sargent[79] and further refined in Sargent[81], Sargent[80], and Sargent[82], he uses Schlesinger's definition for validation[87]. Sargent then goes on to propose, for the first time, actual formal methods for verifying and validating simulation models. He provides four methods for performing V&V[80 82]:

(1) The model development team decides if and when the model is valid for its application domain,
(2) The model development team works in close collaboration with the model's end users so that these users define the validation of the model,
(3) A separate third party assesses the validity of the model in independent verification and validation (IV&V), and
(4) The use of scoring models.

Sargent presented the relative strengths and weaknesses of each method. According to Sargent, the first approach, allowing the model developers to assess validity, is often too subjective and biased. The third method, the use of IV&V, provides the most credibility to the model, but is too prohibitively expensive to be practical[102]. The use of scoring models, such as those found in Balci[7] and Gass[35], is the least worthwhile method because "(1) a model may receive a passing score and yet have a defect that needs to be corrected, (2) the subjectiveness of this approach tends to be hidden resulting in this approach often appearing to be objective, (3) the passing scores may be decided in some (usually) subjective way, (4) the score(s) may cause over confidence in a model, and (5) the scores can be used to argue that one model is better than another"[80]. Sargent concluded by recommending method two (2) as the best method.

Sargent also presented a model for the model development process and the relationship between models and reality. The simple version of this model is shown in Figure 1 and the more detailed version is shown in Figure 2.

Since their original publication, these models have been widely adopted throughout the modeling and simulation community (see Balci[9], Robinson[76], and Kleindorfer et al.[45]). Sargent's model of V&V started from a set of system theories, or the generalization of the underlying problem entity. These system theories were then used to form analytical models which in turn formed the simulation models.

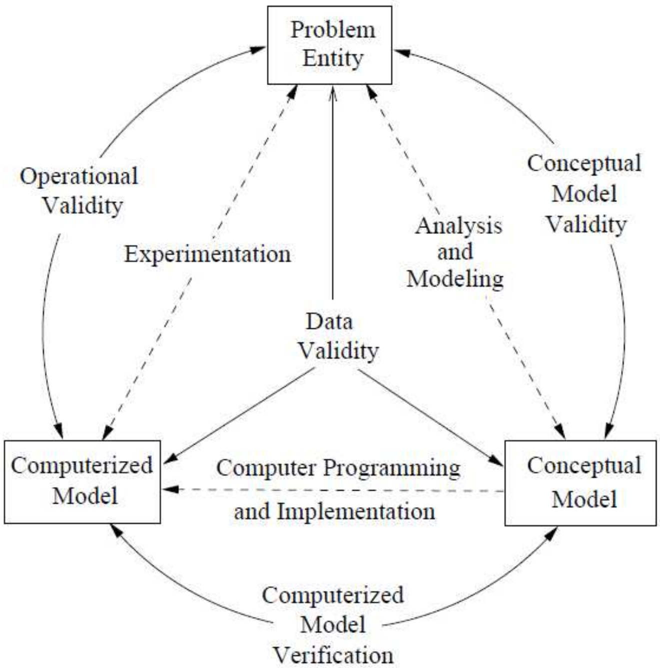

Fig. 1. The basic model development cycle proposed by Sargent[67].

The implementation of the simulation models was verified for proper coding and the simulation models themselves were validated through comparisons to real world experiments. If the comparison of real-world to simulation-world was "reasonable," the systems' theories and therefore the simulation model(s) were considered to be validated.

In light of this model, Sargent outlined several validation methods to apply throughout the model validation cycle. These validation methods included comparisons to other valid models, event validity, extreme condition testing, historical data validation, parameter variability sensitivity analysis, predictive validation, Turing tests, and others. Details on each of these methods are provided in Sargent[79].

For simulation models, Sargent's work is still the most widely accepted framework for how V&V should be performed. However, another leading expert in the field is Balci, who provided a more exhaustive taxonomy for the verification, validation, and accreditation (VV&A) process[10]. Starting from Oberkampf and Trucano[62], Balci defined a full taxonomy and dictionary of V&V processes and techniques. Another taxonomy for model trust and validation methods can be found in Őren and Tuncer[69]. Like Sargent, Balci stated that V&V should be a constant ongoing process throughout the model development cycle.

Balci championed the idea of statistical model validation through hypothesis

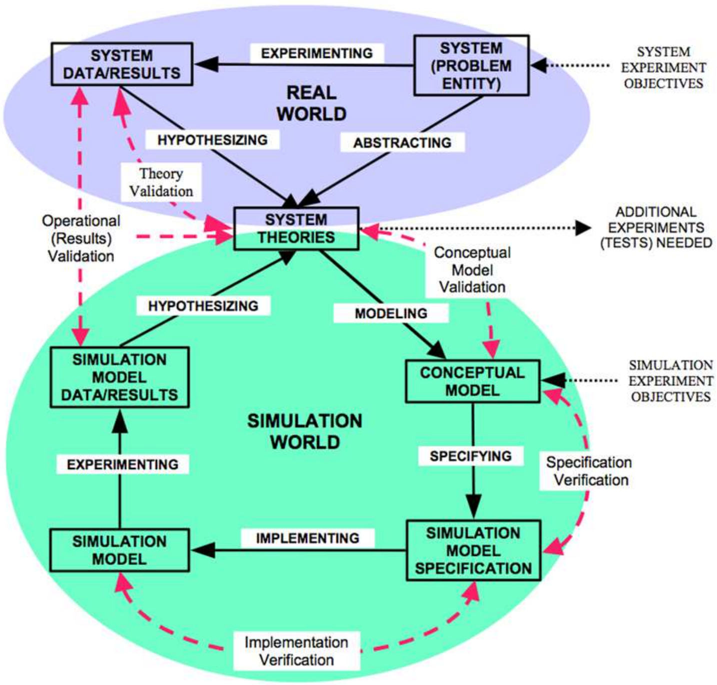

Fig. 2. The detailed model development cycle proposed by Sargent [82].

testing[14 15 16]. His recommended validation method was the comparison of graphical outputs from simulations (e.g. a meta-model) with experimental data and testing the degree of statistical agreement between the two. Balci also dealt with the concept of verification in Balci and Nance[13] and spoke to the necessity of verification in the formulation of simulation models.

2.2.3. *Later Efforts*

Sargent's method have been applied extensively throughout modeling and simulation research, and the concepts of model validation have evolved over the last 40 years. Sargent continued to work in this field to refine the applicability of his framework, as shown in Sargent[83], and his full body of work is summarized in Sargent[84]. Balci also continued his efforts in defining V&V, with later works found in Balci[8 11]. In these works, Balci proposed 15 principles for V&V. Taken from Balci[11], these principles are presented below:

(1) "V&V must be conducted throughout the entire M&S life cycle.
(2) The outcome of VV&A should not be considered as a binary variable where the model or simulation is absolutely correct or absolutely incorrect.
(3) A simulation model is built with respect to the M&S objectives and its credibility is judged with respect to those objectives.

(4) V&V requires independence to prevent developer's bias.
(5) VV&A is difficult and requires creativity and insight.
(6) Credibility can be claimed only for the prescribed conditions for which the model or simulation is verified, validated, and accredited.
(7) Complete simulation model testing is not possible.
(8) VV&A must be planned and documented.
(9) Type I, II, and III errors must be prevented.
(10) Errors should be detected as early as possible in the M&S life cycle.
(11) Multiple response problems must be recognized and resolved properly.
(12) Successfully testing each submodel (module) does not imply overall model credibility.
(13) Double validation problems must be recognized and resolved properly.
(14) Simulation model validity does not guarantee the credibility and acceptability of simulation results.
(15) A well-formulated problem is essential to the acceptability and accreditation of M&S results."

These principles contain the concepts of all of the works that come before Balci[11]. This article again claimed that V&V should be a constant process throughout model development (principle 1), and that V&V should involve a scoring mechanism (principle 2). Balci, like Sargent, also stated that a model was valid only for those cases for which the model has been V&Ved, or a model was applicable only to its particular domain (principle 6). The remaining principles expanded upon the methods of V&V from a qualitative perspective.

One notable contemporary of Sargent and Balci is Robinson. In Robinson[77], he defined verification as "the process of ensuring that the model design (conceptual model) has been transformed into a computer model with sufficient accuracy" and validation as "the process of ensuring that the model is sufficiently accurate for the purpose at hand"[77]. Both of these definitions are aligned with previous definitions. He also further echoed previous V&V theories by stating "A model is only validated with respect to its purpose. It cannot be assumed that a model that is valid for one purpose is also valid for another."[77] This stance also draws heavily from Sargent.

Despite all this work, the problem of V&V for these types of models continues to remain unsolved. In Sargent[85], he states that "there is no set of specific tests that can easily be applied to determine 'correctness' of a model." Taylor et al.[94] came to a similar conclusion, stating "From a technical system perspective, the verification and validation of models and of simulations in its own right are still challenges that have not been solved." However, despite their incompleteness, much of Sargent's and Balci's processes can be extended to computational and physics-based models. Many contemporaries working in the field of V&V for these types of models draw heavily from the processes and frameworks presented in the previous sections. The applicability of each of the methods presented and the models presented in Figure 1 and Figure 2 to computational and physics-based simulations, along with newly proposed methods, are discussed in the following section.

3. Verification and Validation of Computational and Physics-Based Models

With the advent of computational simulations, the concept of V&V for models and simulations shifted, and theoretical work into the nature of V&V sharply declined. Increased computational power enabled simulations to move from process models to simulations of physical phenomena, and this shift made simulations somewhat easier to validate. Most computational simulations focus on a fixed, small-scale system, with some of the earliest computational models, such as those found in Peyret and Taylor[74] and Nielsen et al.[59], focus on fluid flow and moisture transport.

Most early computational models can be thought of as numerical models (see Woodward and Collela[103], Oreskes et al.[70], and Calder et al.[23]). These models were validated by a one-to-one comparison with real world data using primarily the graphical methods described by Sargent, Balci, and others as presented in Section 2. These types of models generally fall into two categories: (1) computational models that seek to solve complex equations that model physical systems and (2) physics-based models that involve coupled systems of computational models to simulate a physical system.

In a sense, computational models can be thought of as analogous to analytical models, as both seek to determine the results of a given equation describing a system or problem entity, and physics-based models can be thought as analogous to simulation models, as both seek to recreate and/or predict the behaviors of a given physical system or process. The following section provides details on the V&V concepts and frameworks developed for computational and physics-based models and simulations.

3.1. *Verification and Validation of Computational Models*

The earliest work in defining V&V for computational models strictly involved CFD codes, and first attempts at defining V&V for computational fluid dynamics (CFD) codes are those described in Blottner and Larson[20]. In this work, Blottner defines verification as “solving the right equations” and validation as “solving the equations right.” Here, the definition of *verification* changed. Rather then ensuring the computer code compiles and executes correctly, now the code must be verified that it solves the correct equations. This verification is realized by comparing model outputs against known solutions for the given equations.

The first formal framework for V&V was proposed by Oberkampf[60] and later refined by Aeschliman et al.[1]. They recommended best practices for verification was “comparison to exact analytic solutions or results from previously verified codes and secondarily comparisons to experimental data.[1]” Similarly, the recommended method for validation “should be ... comparison to carefully designed and conducted experiments." It is worth noting that these methods are some of the same as those proposed by Sargent. These definitions were accepted and presented in Roache's canonical textbook " *Verification and Validation in Computational Science and Engineering*"[75].

Building off of this literature, the most often cited theories and practices in computational model V&V come from Oberkampf and Trucano and are presented in detail in [63]. They state that "verification is the assessment of the accuracy of the solution to a computational model by comparison with known solutions. Validation is the assessment of the accuracy of a computational simulation by comparison with experimental data ... Stated differently, verification is primarily a mathematics issue; validation is primarily a physics issue"[63].

Oberkampf and Barone also provided a formal model design process[61], as is shown in Figure 3. Here, the model's score is a function of its agreement to real-world experimentation.

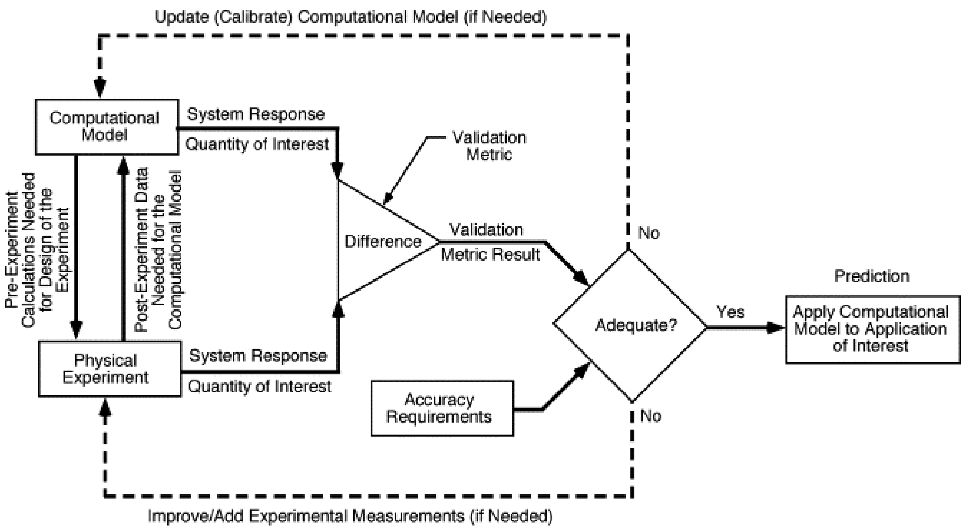

Fig. 3. A model development cycle for computational models[61].

In this approach, Sargent's "System Theories" were replaced with the underlying physics of the problem entity. Within computational models, the physics were generalized to form a parametric model of the equations to be solved - similar to how analytical models were formulated. The equations that can be implemented in code and the results of simulated computations were compared to physical experiments and then assigned a pass/fail score based on the scoring metric. If the model did not pass validation, the parameters were adjusted until the model was sufficiently accurate.

Comparison of simulation output to experimental results has become the standard V&V practice for computational models, and some theoretical work has been done to formalize this aspect of the V&V process. In Humphries and Gurney[41], they state that "if a computational model does not fit existing experimental data, then the model is probably inadequate." They extend this to provide four methods of data-fitting at four levels of accuracy:

> "(L1) Trends: matching directional changes in an experimental variable across experimental conditions; (L2) Means: matching mean values of an experimental variable across experimental conditions; (L3) Distribution: matching the distribution function of an experimental variable across experimental conditions; (L4) Exact values: matching experimental variable values across experimental conditions. Note that each level is a stronger validation of a model than its predecessors, and necessarily subsumes them: for example, if the model fits the mean data values, it must also have captured the trends in that data."[41]

Oberkampf et al.[66] and Oberkampf and Barone[61] provided a formal method to achieve this validation and offered numeric metrics for assessing a computational model's level of validation. In Oberkampf and Barone[61], V&V was again defined as "Verification is the assessment of the accuracy of the solution to a computational model ... validation is the assessment of the accuracy of a computational simulation by comparison with experimental data"[61]. Oberkampf and Barone[61] described several weaknesses in the graphical validation methodology given in Humphries and Gurney[41], including its non-statistical nature, i.e., comparing a fixed value computation to an inherently statistical physical process. Oberkampf and Barone took the position that a quantitative metric of comparison should be used to score the level of validation for a computational model. While this represented a scoring model, which Sargent argued against, it was a mathematical score and not subject to the degree of subjectiveness found in earlier scoring models. Specifically, for their work, Oberkampf and Barone referred to Balci's body of work, along with the more recent efforts found in Hills and Trucano[39], and used statistical confidence intervals as a means of scoring simulation validation in line with what was proposed in Sornette et al.[93], as discussed in more detail in the following section. These metrics, along with what a metric is and is not, and an example application can be found in Oberkampf and Barone[61]. Oberkampf and Trucano expanded on this effort and provided further benchmarking metrics[65].

3.2. *Verification and Validation of Physics-Based Models*

Physics-based models can be thought of as an extension of computational models. However, there are some differences. While computational models focus on using computing power to solve complex equations, physics-based models are designed to recreate and often predict the outcome of physical processes. To that end, most physics-based models are often part of multi-physics simulations wherein several physics-based models with some form of inter-model coupling are run simultaneously to recreate a complex physical phenomenon.

The current experiment matching approach to V&V has proven acceptable for single-system computational simulations, and in particular the solving of partial differential equations describing physical processes (e.g., heat transfer, fluid flow, etc). However, not as much work has been performed to determine whether this approach is valid for fully physics-based simulations for simulating complex systems or systems of systems. A good example of this is in the modeling of sensors. For example, a Light Detection and Ranging (LIDAR) sensor simulation or computational model can be validated against experimental data in a laboratory setting, but *does that mean that the same sensor model is validated for use in more complex environments with complicated sensor-environment interactions?* Furthermore, *does the model have to be re-validated for each possible environment using additional experimental data?*

The goal of physics-based models was to answer these questions. In theory, if the underlying physics of the system or problem entity can be modeled accurately, the model should provide accurate results regardless of the simulation environment. However, in practice this is often not the case. For example, in Hofmann[40], it is shown that a physics-based model will produce invalid results when the underlying assumptions are violated. The goal of physics-based simulations is the same as that of computational simulations: to capture emergent behaviors of the problem entity to test extreme cases not feasibly tested through experiments. This idea of physics-based simulations as a means of predicting the behaviors of complex physical systems is first found in Oberkampf and Trucano[29]. In this work, Easterling stated that "it is critically important, for the sake of credible computational predictions, that model-validation experiments be designed, conducted, and analyzed in ways that provide for measuring predictive capability"[29]. To that end, he provided a model for quantifying the uncertainty in model predictions, which is shown in Figure 4.

This framework started from the intended use of the computational tool and in a method similar to Sargent's original model, developed requirements from the intended use and calculated uncertainty measures coupled with real-world experiments to assess the model's validity. Based on performance, the model was adjusted until the requirements were met. Another similar framework for V&V of physics-based simulations was that of Bayarri et al.[18], which presented a six-step framework for V&V:

(1) "Specify model inputs and parameters with associated uncertainties or ranges - the Input/Uncertainty (I/U) map,
(2) Determine evaluation criteria,
(3) Collect data and design experiments,

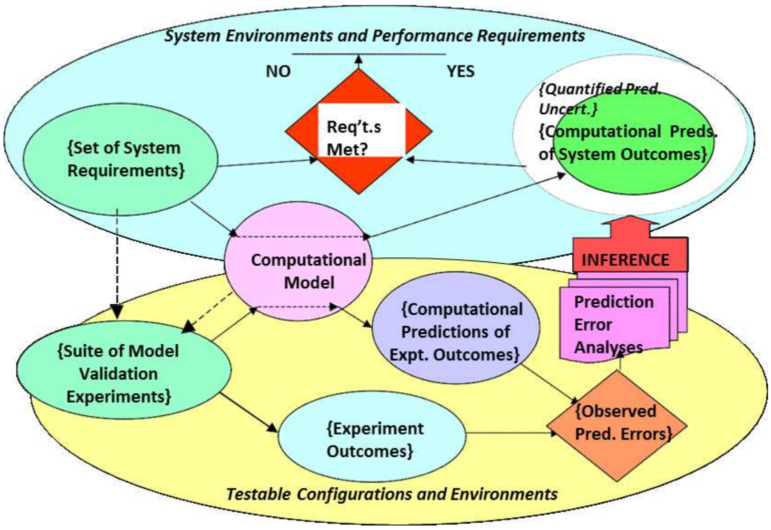

Fig. 4. The model for measuring uncertainty of model predictions developed by Easterling[29].

(4) Approximate computer model output,
(5) Analyze model output; compare computer model output with field data, and
(6) Feed information into current validation exercise and feed-forward information into future validation activities.

They further refine this framework in Bayarri et al.[19] to include uncertainty in the model inputs.

Building off of the concepts found in Babuska and Oden[5], Sornette and associates provided a canonical theory of validation for physics-based models and multi-physics simulations. Sornette et al. first proposed V&V concepts in [92] and with additional researchers further refined these concepts in Sornette et al.[93]. For their work, they defined *validation* "as the process of determining the degree to which a model is an accurate representation of the real world from the perspective of its intended uses"[93]. This definition is identical to Sargent's; a model must be validated to ensure some minimal degree of accuracy for a given system or problem entity. The only difference is that here the system/problem entity is a physical phenomenon in the real world instead of a generalized process.

They go on to argue that "a potentially far-reaching consequence of validation is to give the 'green light' for extrapolating a body of knowledge, which is firmly established only in some limited ranges of variables, parameters, and scales. Predictive capability is what enables us to go beyond this clearly defined domain into a more fuzzy area of unknown conditions and outcomes."[93] This argument echoed Easterling's idea that physics-based models should be used to predict behaviors in otherwise untestable conditions, which was the primary goal of computational models in general.

Unlike Oberkamf, Easterling and Sornette et al. deviated from a standard true/false statistical validation metric and proposed a more robust statistical measure of validation. Both works proposed a scoring model based on a level of trust given to model outputs based on prior knowledge of the system being modeled.

Another paper that discussed the V&V of multi-physics simulations is Avramova and Ivanov's 2010 paper[4]. Avramova and Ivanov proposed that the verification process now involved ensuring that the separate physics-based models communicated in an appropriate manner:

> "The extended verification procedure involves testing the functionality, the data exchange between different physics models, and their coupling, which is designed to model combined effects determined by the interaction of models The extended validation procedure compares the predictions from couples multi-physics code systems to available measured data and reference"[4].

Once verified, the V&V process tracked along with the framework and processes presented in previous sections: simulation outputs were matched to experimental outputs. This matching did not use Sargent's methods, but rather statistical metrics. This was due to the subjective nature of validation scoring methods, and many specific validation metrics have been proposed based on Sornett et al.'s arguments. An overview of the metrics used to date can be found in Oberkampf and Trucano[64], Liu et al.[49], and Ling and Mahadevan[48].

4. Verification and Validation of Simulations of Autonomous Systems

The closest related field to modeling and simulation of autonomous systems is the field of human behavior modeling. Human behavior models take the form of analytical models: the underlying problem entity is the expected behaviors or decisions of a human in a given situation or under a given set of circumstances, and these behaviors are modeled using equations. A good overview of human behavior modeling can be found in Papelis and Madhavan[72]. While numerous methods have been used to describe the human entity, such as finite state machines or fuzzy neural nets, in general, human behaviors can be modeled using equations to represent the mental state of the entity in question and models how different stimuli result in state transitions. These types of models are typically applied in the field of economics for use in predicting consumer behaviors and forecasting future spending Thaler[95], Camerer et al.[24], Padoa-Schioppa[71].

Within the field of robotics, many software simulation tools, such as described in Gerkey et al.[36], Koenig and Howard[46], Michel[54], and Wang et al.[98], have been developed to simulate autonomous ground robots. These simulation environments leverage computational and physics-based models of the robot's sensors and mobility platform to try to predict the autonomous behaviors of the robot. These simulations can be thought to mirror human behavior modeling: the robot has various internal states (position, velocity, world model, etc.) and transitions between states occur as a function of input sensor data. The goal of these simulators was the same as the previous models discussed in this paper: to be validated such that they could be used to predict robot behavior for those scenarios that could not be recreated in the field.

Simulations of autonomous robots is particularly challenging. These simulations involve multiple physics-based simulations working in concert. Each of these models need to be validated separately, but how to validate the overall simulation of these models combined remains an unanswered question. In fact, it has been shown in Weisel et al.[100] and Weisel[101] that separately validated models can produce invalid outputs when combined. Therefore, novel theories will be needed to provide the framework for what methods to use and how to use them to validate these complicated simulation tools.

Of the tools developed for simulating robots, only the Urban Search and Rescue Simulator (USARSim) developed by Wang et al.[98] has been validated to any degree. Several efforts have been undertaken to validate USARSim. Wang et al.[99] noted that "Validating USARSim ... presents a complex problem because the performance of the human-robot system is jointly determined by the robot, the environment, the automation, and the interface." In other words, USARSim is a complex simulation involving multiple physics-based models, and the interactions of these models must be verified. They went on to claim that a complete validation is impossible due to the complexity of the problem and the nature of autonomous systems.

In Carpin et al.[25], a qualitative validation was performed by comparing not the direct sensor model outputs, in this case simulated images to real images, but by comparing the outputs of image processing algorithms (e.g., edge detection or text recognition) using both real and simulated images. The results of this validation were favorable and provided some confidence in USARSim's camera model, but not a comprehensive validation of the model for other environments and algorithms. Finally, in Balaguer et al.[6], a somewhat formal approach to validating USARSim, and by extension any simulation of autonomous ground robots, is proposed as follows:

> "Whenever a new component (sensor, robot, etc.), is added to USARSim, one or more experiments should be designed in order to assess the accuracy of the simulation. These validation experiments will be performed twice, once in the real system and once in a simulation setting resembling as much as possible the real system used. After the two experiments, results should be quantitatively compared" [6].

So, as with the computational and physics-based model validations, the results of the simulation are compared between simulation and experimental testing. Unfortunately, no guidance is given as to how to set up these experiments or how to quantitatively compare the real-world and simulated outputs when the simulation involves a coupled system. New methods are needed to define these experiments and validation metrics. For example, the validation of a given sensor will also imply a validation of the simulation environment model, the range of environments and the algorithms that need to be tested to validate the sensor model are undefined. For M&S of autonomous robots to move forward, these questions need to be addressed and new methods for validation need to be developed.

5. Conclusions and Outstanding Problems

M&S has enjoyed a wealth of success in both basic research and engineering applications. As models and simulation methods have been developed, methods of V&V have been developed in kind, and these methods have been based on a solid theoretical background. Unfortunately, the development of theory and methods of V&V for simulations of autonomous systems has yet to be developed. Presented in this paper are the theories and formal methods of V&V developed to date. This review brings to light the lack of V&V methods for the complex simulation tools in use today.

The glaring problem in the field of V&V for models and simulations lies in the lack of a theoretical framework for V&V of simulations for predicting the behaviors of autonomous robotic systems. While a firm understanding of how to “trust” a given model, be it analytical, computational, or physics-based exists, currently no methodology for V&V of simulations seeking to predict autonomous robotic behaviors exists. To date, there is no firmly established formal methodology for obtaining trust in the outputs of these simulations or even for defining what “trust” means.

Despite this lack of validation, many simulation tools have been developed for the design, development, and testing of autonomous robots, several of which were presented in Section 4. However, very little trust has been placed in these simulation tools. Outside of basic research use, these tools are not widely used. Sargent’s call for increased validation and trust in the case of high risk applications is taken to heart, and simulation results simply are not trusted for the case of autonomous robots operating around humans. A solid theory and method for V&V of these simulation tools is necessary before M&S can be fully leveraged for autonomous robotics applications.

6. Acknowledgments

Permission to publish was granted by the Director of Geotechnical and Structures Laboratory.

References

1. DP Aeschliman, WL Oberkampf, and FG Blottner. A proposed methodology for computational fluid dynamics code verification, calibration, and validation. In *Instrumentation in Aerospace Simulation Facilities, 1995. ICIASF'95 Record., International Congress on*, pages 27–1. IEEE, 1995.
2. Jock R Anderson. Simulation: methodology and application in agricultural economics. *Review of Marketing and Agricultural Economics*, 42(1):3–55, 1974.
3. John F Andrews and Stephen P Graef. *Dynamic modeling and simulation of the anaerobic digestion process*. Environmental Systems Engineering Department, Clemson University, 1970.
4. Maria N Avramova and Kostadin N Ivanov. Verification, validation and uncertainty quantification in multi-physics modeling for nuclear reactor design and safety analysis. *Progress in Nuclear Energy*, 52(7):601–614, 2010.
5. Ivo Babuska and J Tinsley Oden. Verification and validation in computational engineering and science: basic concepts. *Computer Methods in Applied Mechanics and Engineering*, 193(36):4057–4066, 2004.
6. Benjamin Balaguer, Stephen Balakirsky, Stefano Carpin, Mike Lewis, and Christopher Scrapper. Usarsim: a validated simulator for research in robotics and automation. In *Workshop on Robot Simulators: Available Software, Scientific Applications, and Future Trends at IEEE/RSJ*, 2008.
7. Osman Balci. How to assess the acceptability and credibility of simulation results. In *Proceedings of the 21st conference on Winter simulation*, pages 62–71. ACM, 1989.
8. Osman Balci. Principles of simulation model validation, verification, and testing. 1994.
9. Osman Balci. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of operations research*, 53(1):121–173, 1994.
10. Osman Balci. Verification validation and accreditation of simulation models. In *Proceedings of the 29th conference on Winter simulation*, pages 135–141. IEEE Computer Society, 1997.
11. Osman Balci. Verification, validation, and accreditation. In *Proceedings of the 30th conference on Winter simulation*, pages 41–4. IEEE Computer Society Press, 1998.
12. Osman Balci. Verification, validation, and testing. *Handbook of simulation*, pages 335–393, 1998.
13. Osman Balci and Richard E Nance. Formulated problem verification as an explicit requirement of model credibility. *Simulation*, 45(2):76–86, 1985.
14. Osman Balci and Robert G Sargent. Some examples of simulation model validation using hypothesis testing. In *Proceedings of the 14th conference on Winter Simulation-Volume 2*, pages 621–629. Winter Simulation Conference, 1982.
15. Osman Balci and Robert G Sargent. Validation of multivariate response models using hotelling's two-sample t2 test. *Simulation*, 39(6):185–192, 1982.
16. Osman Balci and Robert G Sargent. Validation of simulation models via simultaneous confidence intervals. *American Journal of Mathematical and Management Sciences*, 4(3-4):375–406, 1984.
17. J Bardino, Joel H Ferziger, and William Craig Reynolds. Improved turbulence models based on large eddy simulation of homogeneous, incompressible turbulent flows. *Stanford Univ. Report*, 1, 1983.
18. Maria J Bayarri, James O Berger, Rui Paulo, Jerry Sacks, John A Cafeo, James Cavendish, Chin-Hsu Lin, and Jian Tu. A framework for validation of computer models. *Technometrics*, 49(2), 2007.
19. MJ Bayarri, JO Berger, John Cafeo, G Garcia-Donato, F Liu, J Palomo,

RJ Parthasarathy, R Paulo, Jerry Sacks, and D Walsh. Computer model validation with functional output. *The Annals of Statistics*, pages 1874–1906, 2007.

20. FG Blottner and DE Larson. Navier-stokes code ns3d for blunt bodies. part 1: Analysis, results and verification. *NASA STI/Recon Technical Report N*, 88:25855, 1988.
21. Fabrizio Bonani, S Donati Guerrieri, and Giovanni Ghione. Physics-based simulation techniques for small-and large-signal device noise analysis in rf applications. *Electron Devices, IEEE Transactions on*, 50(3):633–644, 2003.
22. Harold Borko. Computer applications in the behavioral sciences. 1962.
23. AC Calder, B Fryxell, T Plewa, R Rosner, LJ Dursi, VG Weirs, T Dupont, HF Robey, JO Kane, BA Remington, et al. On validating an astrophysical simulation code. *The Astrophysical Journal Supplement Series*, 143(1):201, 2002.
24. Colin F Camerer, George Loewenstein, and Matthew Rabin. *Advances in behavioral economics*. Princeton University Press, 2011.
25. Stefano Carpin, Todor Stoyanov, Yashodhan Nevatia, M Lewis, and J Wang. Quantitative assessments of usarsim accuracy. In *Proceedings of PerMIS*, volume 2006, 2006.
26. Hans G Daellenbach, John A George, and Donald C McNickle. *Introduction to operations research techniques*. Allyn and Bacon Boston, 1983.
27. Mark S Daskin. Logistics: an overview of the state of the art and perspectives on future research. *Transportation Research Part A: General*, 19(5):383–398, 1985.
28. Jim Douglas, Jr and Thomas F Russell. Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures. *SIAM Journal on Numerical Analysis*, 19(5):871–885, 1982.
29. Robert G Easterling. Measuring the predictive capability of computational methods: Principles and methods, issues and illustrations. Technical report, Technical report. Sandia National Laboratories, 2001.
30. J.H. Engel. A verification of lanchester's law. *Journal of the Operations Research Society of America, volume=2, number=3, pages=163–171, year=1954,*.
31. George William Evans, Graham F Wallace, and Georgia L Sutherland. Simulation using digital computers. 1967.
32. George S Fishman and Philip J Kiviat. Digital computer simulation: statistical considerations. Technical report, DTIC Document, 1967.
33. Nehemiah Friedland, Shlomo Maital, and Aryeh Rutenberg. A simulation study of income tax evasion. *Journal of Public Economics*, 10(1):107–116, 1978.
34. Linda Weiser Friedman and Hershey H Friedman. Validating the simulation metamodel: Some practical approaches. *Simulation*, 45(3):144–146, 1985.
35. Saul I Gass. Model accreditation: a rationale and process for determining a numerical rating. *European Journal of Operational Research*, 66(2):250–258, 1993.
36. Brian Gerkey, Richard T Vaughan, and Andrew Howard. The player/stage project: Tools for multi-robot and distributed sensor systems. In *Proceedings of the 11th international conference on advanced robotics*, volume 1, pages 317–323, 2003.
37. John W Harbaugh and Graeme Bonham-Carter. Computer simulation in geology. Technical report, DTIC Document, 1970.
38. Charles F Hermann. Validation problems in games and simulations with special reference to models of international politics. *Behavioral science*, 12(3):216–231, 1967.
39. Richard G Hills and Timothy G Trucano. Statistical validation of engineering and scientific models: Background. *Sandia National Laboratories, Albuquerque, NM, Report No. SAND99-1256*, 1999.
40. M Hofmann. Modeling assumptions: how they affect validation and interoperability.

In *Proceedings of the ACM European Simulation Interoperability Workshop*, 2005.
41. Mark D Humphries and K Gurney. A means to an end: validating models by fitting experimental data. *Neurocomputing*, 70(10):1892–1896, 2007.
42. Edward J Ignall, Peter Kolesar, and Warren E Walker. Using simulation to develop and validate analytic models: some case studies. *Operations Research*, 26(2):237–253, 1978.
43. Naim A Kheir and Willard M Holmes. On validating simulation models of missile systems. *Simulation*, 30(4):117–128, 1978.
44. Jack PC Kleijnen. Regression metamodels for generalizing simulation results. *IEEE transactions on systems, man, and cybernetics*, 9(2):93–96, 1979.
45. George B Kleindorfer, Liam O'Neill, and Ram Ganeshan. Validation in simulation: various positions in the philosophy of science. *Management Science*, 44(8):1087–1099, 1998.
46. Nathan Koenig and Andrew Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2149–2154. IEEE, 2004.
47. Richard C Larson and Amedeo R Odoni. *Urban operations research*. Number Monograph. 1981.
48. You Ling and Sankaran Mahadevan. Quantitative model validation techniques: New insights. *Reliability Engineering & System Safety*, 111:217–231, 2013.
49. Yu Liu, Wei Chen, Paul Arendt, and Hong-Zhong Huang. Toward a better understanding of model validation metrics. *Journal of Mechanical Design*, 133(7):071005, 2011.
50. MH MacDougall. Computer system simulation: An introduction. *ACM Computing Surveys (CSUR)*, 2(3):191–209, 1970.
51. Fritz Machlup. The problem of verification in economics. *Southern Economic Journal*, pages 1–21, 1955.
52. Francis F Martin. Computer modeling and simulation. 1968.
53. John McLeod. *Simulation: the dynamic modeling of ideas and systems with computers*. McGraw-Hill, 1968.
54. Olivier Michel. Webots: Symbiosis between virtual and real mobile robots. In *Virtual Worlds*, pages 254–263. Springer, 1998.
55. G Arthur Mihram. *Simulation*. Elsevier, 1972.
56. G Arthur Mihram. Some practical aspects of the verification and validation of simulation models. *Operational research quarterly*, pages 17–29, 1972.
57. Thomas H Naylor and Joseph Michael Finger. Verification of computer simulation models. *Management Science*, 14(2):B–92, 1967.
58. Thomas H Naylor and Joseph Michael Finger. Verification of computer simulation models. *Management Science*, 14(2):B–92, 1967.
59. DR Nielsen, JW Biggar, et al. Water flow and solute transport processes in the unsaturated zone. *Water Resources Research*, 22(9S):89S–108S, 1986.
60. William L Oberkampf. A proposed framework for computational fluid dynamics code calibration/validation. *AIAA Paper*, (94-2540), 1994.
61. William L Oberkampf and Matthew F Barone. Measures of agreement between computation and experiment: validation metrics. *Journal of Computational Physics*, 217(1):5–36, 2006.
62. William L Oberkampf and Timothy G Trucano. Verification, validation, and accreditation recommended practices guide. Technical report, Defense Modeling and Simulation Ofce, Ofce of the Director of Defense Research and Engineering, 1996.

63. William L Oberkampf and Timothy G Trucano. Verification and validation in computational fluid dynamics. *Progress in Aerospace Sciences*, 38(3):209–272, 2002.
64. William L Oberkampf and Timothy G Trucano. Design of and comparison with verification and validation benchmarks. *Sandia National Laboratories, Albuquerque, NM, Technical Report Sand No*, 2006.
65. William L Oberkampf and Timothy G Trucano. Verification and validation benchmarks. 2008.
66. William L Oberkampf, Timothy G Trucano, and Charles Hirsch. Verification, validation, and predictive capability in computational engineering and physics. *Applied Mechanics Reviews*, 57(5):345–384, 2004.
67. Tuncer I Ören. Concepts and criteria to assess acceptability of simulation studies: a frame of reference. *Communications of the ACM*, 24(4):180–189, 1981.
68. Tuncer I Ören. Quality assurance in modelling and simulation: a taxonomy. In *Simulation and Model-Based Methodologies: An Integrative View*, pages 477–517. Springer, 1984.
69. Tuncer I Ören. Quality assurance in modelling and simulation: a taxonomy. In *Simulation and Model-Based Methodologies: An Integrative View*, pages 477–517. Springer, 1984.
70. Naomi Oreskes, Kristin Shrader-Frechette, and Kenneth Belitz. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147):641–646, 1994.
71. Camillo Padoa-Schioppa. Neurobiology of economic choice: a good-based model. *Annual review of neuroscience*, 34:333, 2011.
72. Yiannis Papelis and Poornima Madhavan. Human behavior. *Modeling and simulation fundamentals*, page 271, 2010.
73. Mikel D Petty. Verification and validation. *Principles of modeling and simulation: A multidisciplinary approach*, pages 121–149, 2009.
74. Roger Peyret and Thomas Darwin Taylor. Computational methods for fluid flow. *New York, Springer-Verlag, 1985, 368 p.*, 1, 1985.
75. Patrick J Roache. *Verification and validation in computational science and engineering*. Hermosa, 1998.
76. Stewart Robinson. Simulation model verification and validation: increasing the users' confidence. In *Proceedings of the 29th conference on Winter simulation*, pages 53–59. IEEE Computer Society, 1997.
77. Stewart Robinson. Simulation model verification and validation: increasing the users' confidence. In *Proceedings of the 29th conference on Winter simulation*, pages 53–59. IEEE Computer Society, 1997.
78. Jeff Rothenberg. *The nature of modeling*, volume 3027. Rand, 1989.
79. Robert G Sargent. Validation of simulation models. In *Proceedings of the 11th conference on Winter simulation-Volume 2*, pages 497–503. IEEE Press, 1979.
80. Robert G Sargent. A tutorial on verification and validation of simulation models. In *Proceedings of the 16th conference on Winter simulation*, pages 114–121. IEEE Press, 1984.
81. Robert G Sargent. An expository on verification and validation of simulation models. In *Proceedings of the 17th conference on Winter simulation*, pages 15–22. ACM, 1985.
82. Robert G Sargent. An overview of verification and validation of simulation models. In *Proceedings of the 19th conference on Winter simulation*, pages 33–39. ACM, 1987.
83. Robert G Sargent. Verification and validation of simulation models. In *Proceedings of the 37th conference on Winter simulation*, pages 130–143. Winter Simulation Conference, 2005.

84. Robert G Sargent. Verification and validation of simulation models. *Journal of simulation*, 7(1):12–24, 2013.
85. Robert G Sargent. Verification and validation of simulation models. *Journal of simulation*, 7(1):12–24, 2013.
86. Robert E Schellenberger. Criteria for assessing model validity for managerial purposes*. *Decision Sciences*, 5(4):644–653, 1974.
87. Stewart Schlesinger, Roy E Crosbie, Roland E Gagné, George S Innis, CS Lalwani, Joseph Loch, Richard J Sylvester, Richard D Wright, Naim Kheir, and Dale Bartos. Terminology for model credibility. *Simulation*, 32(3):103–104, 1979.
88. Thomas J Schriber. Introduction to simulation. In *Proceedings of the 9th conference on Winter simulation-Volume 1*, page 23. Winter Simulation Conference, 1977.
89. JG Shanthikumar and RG Sargent. A unifying view of hybrid simulation/analytic models and modeling. *Operations research*, 31(6):1030–1052, 1983.
90. Harold Shichman and D Hodges. Modeling and simulation of insulated-gate field-effect transistor switching circuits. *Solid-State Circuits, IEEE Journal of*, 3(3):285–289, 1968.
91. MF Snyder, Vincent C Rideout, and RJ Hillestad. Computer modeling of the human systemic arterial tree. *Journal of Biomechanics*, 1(4):341–353, 1968.
92. D Sornette, AB Davis, K Ide, KR Vixie, V Pisarenko, and JR Kamm. Algorithm for model validation: Theory and applications. *Proceedings of the National Academy of Sciences*, 104(16):6562–6567, 2007.
93. Didier Sornette, Anthony B Davis, James R Kamm, and Kayo Ide. A general strategy for physics-based model validation illustrated with earthquake phenomenology, atmospheric radiative transfer, and computational fluid dynamics. In *Computational Methods in Transport: Verification and Validation*, pages 19–73. Springer, 2008.
94. Simon JE Taylor, Azam Khan, Katherine L Morse, Andreas Tolk, Levent Yilmaz, and Justyna Zander. Grand challenges on the theory of modeling and simulation. In *Proceedings of the Symposium on Theory of Modeling & Simulation-DEVS Integrative M&S Symposium*, page 34. Society for Computer Simulation International, 2013.
95. Richard Thaler. Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1):39–60, 1980.
96. Leonard Merrick Uhr. *Pattern recognition: Theory, experiment, computer simulations, and dynamic models of form perception and discovery.* John Wiley & Sons Inc, 1966.
97. Richard L Van Horn. Validation of simulation results. *Management Science*, 17(5):247–258, 1971.
98. Jijun Wang, Michael Lewis, and Jeffrey Gennari. A game engine based simulation of the nist urban search and rescue arenas. In *Simulation Conference, 2003. Proceedings of the 2003 Winter*, volume 1, pages 1039–1045. IEEE, 2003.
99. Jijun Wang, Michael Lewis, Stephen Hughes, Mary Koes, and Stefano Carpin. Validating usarsim for use in hri research. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 457–461. SAGE Publications, 2005.
100. Eric W Weisel, Mikel D Petty, and Roland R Mielke. Validity of models and classes of models in semantic composability. In *Proceedings of the Fall 2003 Simulation Interoperability Workshop*, volume 9, page 68, 2003.
101. Eric Werner Weisel. Models, composability, and validity. 2004.
102. David O Wood. Mit model analysis program: what we have learned about policy model review. In *Proceedings of the 18th conference on Winter simulation*, pages

248–252. ACM, 1986.
103. Paul Woodward and Phillip Colella. The numerical simulation of two-dimensional fluid flow with strong shocks. *Journal of Computational Physics*, 54(1):115–173, 1984.
104. Simone M Youngblood. Literature review and commentary on the verification, validation and accreditation of models and simulations. *Johns Hopkins University, Laurel, MD*, 1993.