A Framework for Simulation Validation & Verification Method Selection

Roungas, Bill; Meijer, Sebastiaan; Verbraeck, Alexander

Publication date
2017

Document Version
Final published version

Published in
Proceedings of the 9th International Conference on Advances in System Simulation, SIMUL 2017

Citation (APA)
Roungas, B., Meijer, S., & Verbraeck, A. (2017). A Framework for Simulation Validation & Verification Method Selection. In A. Ramezani, E. Williams, & M. Bauer (Eds.), Proceedings of the 9th International Conference on Advances in System Simulation, SIMUL 2017. (pp. 35-40)

A Framework for Simulation Validation & Verification Method Selection

Bill Roungas
Alexander Verbraeck

Department of Multi Actor Systems
Delft University of Technology
Delft, The Netherlands
Email: v.roungas@tudelft.nl,
a.verbraeck@tudelft.nl

Sebastiaan Meijer

Department of Health Systems Engineering
KTH Royal Institute of Technology
Huddinge, Sweden
Email: sebastiaan.meijer@sth.kth.se

***Abstract*—Thirty years of research on validation and verification (V&V) has returned a plethora of methods, statistical techniques, and reported case studies. It is that abundance of methods that poses a major challenge. Because of overlap between methods and time and budget constraints, it is impossible to apply all the available methods in a single study, so a careful selection of methods has to be made. This paper builds on two assumptions: a) that both simulations and V&V methods can be defined on the basis of different characteristics and b) that certain V&V methods are more suitable than others for different kinds of simulations. The present study aims at identifying the specific characteristics that make a V&V method more effective and more efficient than others, when confronting these with the simulations' different characteristics. The conclusion will advance a methodology for choosing the most appropriate method or methods for validating and/or verifying a simulation.**

Keywords–simulation; validation; verification; method selection.

I. INTRODUCTION

Back in 1972, based on Forrester's work [1], Meadows et al. [2], [3] introduced World 3, a simulation of the world for the years 1900-2100. The purpose of the simulation model was to project the dynamic behavior of population, capital, food, non-renewable resources, and pollution. The model's forecast was that the world would experience a major industrial collapse, which would be followed by a significant decrease in human population. The model became very popular especially because of the increasing interest in environmental degradation due to human activities [4]. Even though the model gained support for being "of some use to decision makers" [3] and generated the spark for many later global models, it had several shortcomings, for which it received a lot of criticism as well [5]. In turn, this criticism raised the question of whether, and to what extent, such simulation models are validated and verified. This is just one example of the notion that V&V is a fundamental part of a simulation study [6].

The term V&V is used to characterize two relatively different approaches, which almost always go hand in hand, validation and verification. Validation is this phase of a study that ensures that the simulation imitates the underline system, to a greater or lesser extent, and in any case to a satisfactory degree [7], or in layman terms validation address the question: did we build the "right" model [8]. On the other hand, verification is the phase of the study that ensures that the model and its implementation are correct [9], or in layman terms verification addresses the question: did we build the model "right" [8]. V&V has become a well-researched field with a significant amount of produced literature and commercial case studies. The large number of methods and techniques created by this wide range of research, is the greatest impediment to the designing of a V&V study.

The predetermined budget of a simulation study usually limits the amount of time and resources that can be spent on V&V. Additionally, the nature and the diverse characteristics of simulations limit the number of V&V methods that are applicable for each simulation. In other words, not all V&V methods are suitable for every simulation. To the best of our knowledge, a taxonomy for characterizing V&V methods and, subsequently, matching them with different simulations does not exist.

This paper aims at identifying the majority of the available V&V methods in order to classify them on the basis of their different characteristics and on whether they can validate or verify a simulation, and eventually match them with characteristics of simulation models.

Section II starts with a literature analysis on V&V methods, simulation properties, and simulation study phases, and then proceeds with introducing a methodology towards developing a framework for simulation V&V method selection. In Section III, a case study is presented to illustrate how the proposed framework can be put in practice. Finally, in Section IV, the future potential extensions of the framework are presented and final remarks are made.

II. THE FRAMEWORK

This section starts with a 3-step literature analysis and then proceeds with proposing a methodology for selecting one or more methods for a V&V study.

A. The 3-step Literature Analysis

The initial hypothesis of this study is that simulations exhibit certain properties that influence the effectiveness of a V&V method. Therefore, the 3 steps of the literature analysis are the following:

Step 1: Identification of V&V methods.
Step 2: Identification of simulations' properties.
Step 3: Identification of the phases of a simulation study.

1) Step 1: V&V methods, as indicated by their definitions on Table I, are different in many aspects; some methods are strictly mathematical whereas others accommodate the more qualitative aspects of simulations, etc. Balci [10] identified

 ISBN: 978-1-61208-594-4

more than 70 V&V methods, which he in turn categorized into four categories: informal, static, dynamic, and formal. Balci's [10] list is the most accurate representation of the body of work on V&V methods and, even to date, is considered as the most extensive one. This paper adopts the list in reference - but not the categorization - and goes as far as to propose a new classification of V&V methods. Further to the above, whilst the list is adopted in its entirety, some methods may occasionally appear to have been excluded. In effect, this occurs only when a particular method belongs to a group of methods, in which case if there are no significant differences between these methods, only the "parent" method is enlisted. Due to size restrictions, it is not possible to provide the definition of each method in this section. Nevertheless, references to detailed definitions can be found in Table I.

2) Step 2: Since simulations differ from one another in various ways, distinctions are made on whether they represent an existing system, or whether they simulate a system at a microscopic or macroscopic level, or whether they are intended for learning or decision making, and so forth. This is an indication that simulations can be characterized by various properties. Based on literature, this study has identified 10 properties of simulations. The rationale behind selecting those properties was to describe simulations with as much detail as possible. Hence, the properties span multiple levels. Not all identified properties necessarily influence the selection of V&V methods, therefore this step is not only about identifying the properties but also determining which are the ones that really influence the effectiveness of a method; in other words, this step serves as the rationale for choosing those properties of simulations that are applicable to specific V&V methods, and provides for the reasons behind this selection.

The 10 identified properties of simulations are the following:

1) Access to the source code of the simulation. Accessibility, or lack of it, influences the selection of a V&V method [11], since several methods require some sort of a check on the code level. Hence, this property is included in the analysis.
2) The simulation represents an existing real-system for which real data exist [12]. The existence of, or more importantly the lack of, real data heavily influences the selection process since several methods require real data and thus cannot be used when there are not any. Hence, this property is included in the analysis.
3) The formalism the simulation is based on, like Discrete Event System Specification (DEVS), Differential Equation Specified System (DESS), System Dynamics, etc. [13]. Several frameworks and methods have been proposed on how to verify and validate DEVS [14], [15], DESS [16], [17], or system dynamics models [18], [19], but they are either application specific or the same method can be used in more than one formalisms, making it independent of the actual formalism. Therefore, while formalisms are an important aspect of simulation modeling, their influence on the V&V method selection are minimal, ergo excluded from the analysis.
4) The simulation's worldviews: i) Process Interaction/Locality of Object, ii) Event Scheduling/Locality of Time, iii) Activity Scanning/Locality of State [20]. While worldviews allow for more concise model descriptions by allowing a model specifier to take advantage of contextual information, there is not any evidence from a literature point of view that they have an influence on the V&V method selection, hence, they are excluded from the analysis.
5) The fidelity level of the simulation (Low, Medium, High) [21]. While from a literature point of view there is no evidence to support the influence of the level of fidelity on the V&V method selection, common sense dictates that there must be some. Indeed, in order to characterize a simulation as of high fidelity, it must imitates an existing system and real-world data must exist, thus making the comparison and the final characterization possible. Therefore, as discussed in the second property and is shown in Table I, since the existence of data of the real system influences the V&V method selection, so does the level of fidelity, but since the correlation between real data and high fidelity is almost 1-to-1, the fidelity level is excluded from the analysis for reasons of simplification.
6) The type of the simulation (Constructive, Virtual, Live) [22]. This classification, which is adopted by the U.S. Department of Defense [23], should be seen more as a continuum rather than a discrete characterization. Once a simulation moves towards the Virtual or the Live side of the continuum, it can also be referred to as 'a game'. A game has the distinct characteristic that the game session is succeeded by debriefing, whereby the participants reflect upon the game session to link the content presented during the session with reality [24]. It has been demonstrated that debriefing can in general facilitate validation [25], [26], but except for two methods, i.e., *User Interface Analysis* and *User Interface Testing*, there is no evidence in literature on whether the type of simulation affects the V&V method selection. Hence, this classification is excluded from the analysis.
7) The purpose the simulation was built for (learning, decision making, etc.). Several case studies on V&V of simulations for different purposes have been reported; in training [27], [28], in decision making [29], in concept testing [30], etc., but there are no reports of specific V&V methods being more effective for a certain purpose. Hence, this property is excluded from the analysis.
8) The simulation imitates a strictly technical, a socio technical system (STS), or a complex adaptive system (CAS) with multiple agents. There are several studies on modeling and validating simulations for STS [31] and CAS with multiple agents [32], [33] but there are no indications that certain V&V methods are more effective for an STS or a CAS. Therefore, this property is excluded from the analysis.
9) The application domain of the simulation (logistics, business, physics, etc.). Although the application domain of the simulation plays a significant role in the modeling process, since different approaches are required (Newtonian physics for object movement, Navier–Stokes equations for fluid behavior, etc.) for modeling different systems [34], literature, or more precisely the lack of it, suggests that the V&V process

and thus the V&V method selection is not affected by the application domain. Hence, this property is excluded from the analysis.

10) The functional (hard goals) and non-functional (soft goals) requirements of the simulation [35]. Validating the simulation's requirements is indeed an important part of the V&V process [36], since validation is always relative to the intended use [37], in other words the use defined in the requirements. Hence, making a distinction between the hard and soft goals is paramount and as such this property is included in the analysis.

3) Step 3: According to Sargent [38], there are 4 distinct phases of V&V: *Data Validation*, *Conceptual Model Validation*, *Model Verification*, and *Operational Validation*. *Data Validation* is concerned with the accuracy of the raw data, as well as the accuracy of any transformation performed on this data. *Conceptual Model Validation* determines whether the theories and assumptions underlying the conceptual model are correct, and whether the model's structure, logic, and mathematical and causal relationships are "reasonable" for the intended purpose of the model. *Model Verification* ensures that the implementation of the conceptual model is correct. Finally, *Operational Validation* is concerned with determining that the model behaves accurately based on its intended purpose. This study adopts Sargent's [38] characterization and aims at using it to classify the methods, in addition to the simulations' properties.

4) Conclusion of the Literature Review: It is evident that selecting one method over another for a V&V study depends on several characteristics from both sides, i.e., the simulation and the methods, as well as the phase of the simulation study. In Section II-B, a methodology that combines all three steps aiming at the development of a framework for V&V method selection is proposed.

B. Methodology

As discussed in Section II-A2, dimensions 3, 4, 5, 6, 7, 8, and 9 are perceived to have little influence on the method selection, hence, there are excluded from the analysis. On the other hand, the purpose of the method selection, discussed in Section II-A3, seems to be crucial; in other words, it is important to differentiate on whether the selected method will be used for data validation, conceptual model validation, model verification, or operational validation. Therefore, the list of the dimensions is refined, and is expressed in questions, as follows:

1) Does the V&V method require access to the simulation model's source code?
Possible answers: Yes or No. A positive answer to this question means that this method can only be used when the person or persons performing the V&V have access to the simulation's source code, whereas a negative answer means that it can be used in any occasion regardless of the accessibility to the simulation model's source code. It should be noted that the current study - and consequently this dimension - is not concerned with the specific programming language the simulation is built on (Assembly, C++, NetLogo, etc.), but solely with whether the application of a V&V method depends upon having access to the source code.
2) Does the V&V method require data from the real system?
Possible answers: Yes or No. A positive answer to this question means that this method can only be used when data from the real system are available, whereas a negative answer means that it can be used in any occasion regardless of the availability of data from the real system. It should be noted that the current study - and consequently this dimension - is not concerned with the nature of the data in general (qualitative or quantitative), but solely with their existence and

TABLE I. LIST OF V&V METHODS & PROPERTIES OF SIMULATIONS.

Method	1	2	3	4	Source
Acceptance Testing	No	No	Both	O. Val.	[39]
Alpha Testing	No	No	Both	O. Val.	[40]
Assertion Checking	Yes	No	Hard	M. Ver.	[41]
Audit	Yes	No	Soft	M. Ver.	[42]
Beta Testing	No	No	Both	O. Val.	[43]
Bottom-Up Testing	Yes	No	Both	M. Ver.	[44]
Boundary Value Testing	Yes	No	Both	M. Ver.	[45]
Cause-Effect Graphing	Yes	No	Hard	M. Ver.	[45]
Comparison Testing	No	No	Both	C.M. Val.	[46]
Compliance Testing	No	No	Soft	O. Val.	[42]
Control Analysis	Yes	No	Hard	M. Ver.	[47]
Data Analysis Techniques	Yes	No	Hard	D. Val. & M. Ver.	[42]
Data Interface Testing	No	No	Soft	D. Val.	[43]
Debugging	Yes	No	Both	M. Ver.	[48]
Desk Checking	Yes	No	Both	M. Ver.	[49]
Documentation Checking	Yes	No	Both	C.M. Val.	[10]
Equivalence Partitioning Testing	No	No	Hard	O. Val.	[50]
Execution Testing	No	No	Hard	C.M. Val.	[51]
Extreme Input Testing	No	No	Hard	O. Val.	[46]
Face Validation	No	Yes	Both	O. Val.	[52]
Fault/Failure Analysis	No	No	Hard	C.M. Val.	[43]
Fault/Failure Insertion Testing	No	No	Hard	C.M. Val.	[10]
Field Testing	No	Yes	Both	O. Val.	[53]
Functional (Black-Box) Testing	No	Yes	Hard	C.M. Val.	[45]
Graphical Comparisons	No	Yes	Both	O. Val.	[54]
Induction	No	No	Both	C.M. Val.	[55]
Inference	No	No	Both	C.M. Val.	[56]
Inspections	No	No	Both	C.M. Val.	[57]
Invalid Input Testing	No	No	Hard	O. Val.	[10]
Lambda Calculus	Yes	No	Hard	M. Ver.	[58]
Logical Deduction	No	No	Both	All	[51]
Model Interface Analysis	No	No	Soft	C.M. Val.	[10]
Model Interface Testing	No	No	Soft	C.M. Val.	[44]
Object-Flow Testing	No	No	Hard	O. Val.	[59]
Partition Testing	Yes	No	Hard	C.M. Val.	[60]
Predicate Calculus	Yes	No	Hard	M. Ver.	[61]
Predicate Transformations	No	Yes	Hard	M. Ver.	[62]
Predictive Validation	No	Yes	Hard	O. Val.	[63]
Product Testing	No	No	Both	O. Val.	[39]
Proof of Correctness	Yes	No	Hard	C.M. Val. & M. Ver.	[61]
Real-Time Input Testing	No	Yes	Hard	O. Val.	[10]
Regression Testing	Yes	No	Hard	M. Ver.	[51]
Reviews	No	No	Both	C.M. Val.	[42]
Self-Driven Input Testing	No	No	Hard	O. Val.	[64]
Semantic Analysis	Yes	No	Both	M. Ver.	[51]
Sensitivity Analysis	No	No	Hard	O. Val.	[65]
Stress Testing	No	No	Hard	O. Val.	[66]
Structural (White-box) Testing	Yes	No	Both	C.M. Val.	[40]
Structural Analysis	No	No	Hard	C.M. Val.	[51]
Submodel/Module Testing	No	No	Both	C.M. Val.	[67]
Symbolic Debugging	Yes	No	Hard	M. Ver.	[51]
Symbolic Evaluation	Yes	No	Hard	C.M. Val.	[68]
Syntax Analysis	Yes	No	Hard	M. Ver.	[40]
Top-Down Testing	Yes	No	Both	C.M. Val.	[44]
Trace-Driven Input Testing	Yes	Yes	Both	D. Val. & C.M. Val.	[10]
Traceability Assessment	Yes	Yes	Both	C.M. Val.	[43]
Turing Test	No	Yes	Both	O. Val.	[69]
User Interface Analysis	No	No	Soft	O. Val.	[10]
User Interface Testing	No	No	Soft	O. Val.	[70]
Visualization/Animation	No	Yes	Both	O. Val.	[38]
Walkthroughs	No	No	Both	C.M. Val.	[45]

availability.

3) For what type of requirements is the V&V method more suitable?
Possible answers: Hard (Functional), or Soft (Non-Functional), or Both. A method might be focused on either the functional part or the non-functional part of the model or on both.
4) For which type of study is the V&V method more suitable?
Possible answers: Data Validation (D. Val.), Conceptual Model Validation (C.M. Val.), Model Verification (M. Ver.), or Operational Validation (O. Val.). A method might be suitable for one or more of the available categories.

Table I summarizes the results of the analysis. The last column, i.e., *Source*, indicates the origin of each method but it is also a source that justifies the choices in columns 2-5.

C. Discussion

The intended use of Table I is to act as a filtering mechanism. Whenever an individual or a team wants to verify and/or validate a simulation model, they can utilize this table to narrow down the applicable V&V methods according to the different properties of the simulation at hand.

With regards to the first property, i.e., the accessibility to the source code, and in contrary to the second property, access to the source code does not imply that the methods categorized under "Yes" are stronger. Usually, access to the source code is associated with verification and in some cases conceptual model validation.

With regards to the second property, i.e., the availability of data from the real system, by all means, methods categorized under "No" can be used whether real data exist or not. Nevertheless, the methods categorized under "Yes" are more powerful in the sense that, if used appropriately, they provide evidence or a data trace of how the simulation should work. Hence, whenever real data are available, the methods categorized under "Yes" should be preferred, unless an alternative method is definitely more suitable.

With regards to the third and fourth property, i.e., the type of requirements being tested and the purpose of the V&V study respectively, the answers are more or less self-explanatory. Some methods are more suitable for testing one type of requirement. As an example, regression testing is more appropriate for functional requirements (hard goals). Other V&V methods are better suited for one purpose, such as Structural (White-box) Testing, which is more appropriate for conceptual model validation, while others are more suitable for testing both types of requirements (e.g., Graphical comparisons), or for more than one purpose (e.g., Trace-Driven Input Testing).

The novelty of the proposed framework does not lie in the content of Table I per se, but on the idea that the list of V&V methods can be narrowed down to a manageable level, thus making the V&V of a simulation better grounded, faster, more accurate, and more cost effective.

There is a threat towards the validity of the content on Table I. The line between whether data from the real system are needed, or whether access to the source code is needed, or whether a specific requirement is definitely functional or non-functional, or whether the purpose is to validate the data, the conceptual model, the operational ability of the model, or to just verify the model, is not always clear and well defined. In Section IV, future steps are proposed aiming at addressing and mitigating the above mentioned threat.

III. A Case Study

In this section, a case study illustrates how the framework, through the use of Table I, can be used. The case study is a computer simulation of a particular instantiation of the Dutch railway system. The authors were assigned to validate the simulation model with regards to punctuality; the precision of the delays of trains in the model.

The initial list, as it is shown on Table I, consists of 61 methods. Then with every step, the list is narrowed down. For this particular study, the selection process for each property was as follows:

1) Access to the source code was not available; *Answer: No*. Using this criteria reduces the available methods to 38.
2) There were available data from the real system; *Answer: Yes*. Using this criteria eliminates 27 more methods totaling in 9 available methods. Although, all 38 methods can be used in this particular case.
3) The main focus was on the punctuality, ergo functional (hard) requirements, but comments were also expected on the non-functional (soft) requirements; *Answer: Both (but main focus on hard)*. If on the previous criteria *Yes* was chosen as an option, choosing either *Both* or *Hard* on this criteria leaves the list intact (Total 9 methods). The same applies if on the previous criteria *All* was chosen as an option and *Both* is chosen as an option on this one. On the contrary, if on the previous criteria *All* was chosen as an option and on this criteria *Hard* is chosen as an option, the list is further reduced by 6 methods to a total of 32 available methods.
4) The study was mainly concerned with the operational validity of the simulation, but to a degree also with the conceptual model validity; *Answer: C.M. Val & O. Val.*. Using this criteria and based on the selections on the previous criteria, the final number of available methods was reduced to between 1 and 15 for the conceptual model validation and between 7 and 22 for the operational validation.

TABLE II. REFINED LIST OF V&V METHODS OF THE CASE STUDY.

Method	1	2	3	4
Face Validation	No	Yes	Both	O. Val.
Field Testing	No	Yes	Both	O. Val.
Graphical Comparisons	No	Yes	Both	O. Val.
Predictive Validation	No	Yes	Hard	O. Val.
Real-Time Input Testing	No	Yes	Hard	O. Val.
Turing Test	No	Yes	Both	O. Val.
Visualization/Animation	No	Yes	Both	O. Val.

For the operational validation, which was the primary interest for the study, the final list of the seven methods is shown in Table II. From this list, in total four methods were used, namely the *Face Validation*, *Graphical Comparisons*, *Predictive Validation*, and *Turing Test*. *Predictive Validation* was first used to handle the initial datasets (simulation dataset & operational dataset) and to produce results for the different statistical tests. Then, a combination of the remaining three methods was used to ascertain the validity of the simulation.

In this section, the use of the proposed framework demonstrates clearly its effectiveness. As shown in Table II, the initial list of 61 methods was reduced in a matter of minutes to the manageable level of seven. By all means, the effectiveness of the framework is not only evident due to its time-saving nature but also due to the fact that it ensures that the chosen methods are appropriate for the simulation at hand and for the purpose of the V&V study.

IV. Conclusion & Future Work

In this paper, a framework for simulation validation and verification method selection was proposed. Various properties of simulations were taken into account and it was shown that indeed some of these properties, as well as the purpose of a V&V study, influence the method selection and thus, the result of the simulation study.

Moreover, the framework was applied on a case study, as a first step towards verifying its effectiveness. The case study showed that the framework is an effective time-saving tool, which also provides a safety net for choosing the method that best serves the intended purpose of the simulation and the V&V study.

With regards to future work, additional simulation properties may potentially influence the V&V method selection, or some of the discarded properties, identified in Section II-A2, might prove to be more influential than initially acknowledged. Moreover, there is a need to further verify the connection of each method to the simulation model's properties and the purpose for which they are more suitable; in other words, it should be verified that the answers on columns 2-5 in Table I are correct. Finally, more case studies, from the authors and more importantly from researchers unrelated to the authors, would further strengthen the validity and applicability of the framework.

Nevertheless, this paper paves the way for future research in the topic, and as discussed earlier, the main contribution of the framework does not lie in the results presented on Table I, but is related to the identification of the relationships between the V&V methods, and the simulation model's properties and purpose of the V&V study. Therefore, it is of utmost importance that any future research be focused on these relationships.

Acknowledgment

This research is funded through the Railway Gaming Suite 2 program, a joint project by ProRail and Delft University of Technology.

References

[1] J. W. Forrester, World dynamics. Wright-Allen Press, 1971.

[2] D. H. Meadows, D. L. Meadows, J. Randers, and W. W. Behrens III, The limits to growth. New York, U.S.A.: Universe Books, 1972.

[3] D. L. Meadows, W. W. Behrens III, D. H. Meadows, R. F. Naill, J. Randers, and E. Zahn, Dynamics of growth in a finite world. Cambridge, Massachusetts: Wright-Allen Press, 1974.

[4] M. Janssen and B. De Vries, "Global modelling: Managing uncertainty, complexity and incomplete information," in Validation of Simulation Models. SISWO, Amsterdam, The Netherlands, 1999, pp. 45–69.

[5] W. D. Nordhaus, "World dynamics: Measurement without data," The Economic Journal, vol. 83, no. 332, 1973, pp. 1156–1183.

[6] O. Balci, "Validation, verification, and testing techniques throughout the life cycle of a simulation study," Annals of Operations Research, vol. 53, no. 1, 1994, pp. 121–173.

[7] S. Schlesinger, R. E. Crosbie, R. E. Gagné, G. S. Innis, C. S. Lalwani, J. Loch, R. J. Sylvester, R. D. Wright, N. Kheir, and D. Bartos, "Terminology for model credibility," Simulation, vol. 32, no. 3, 1979, pp. 103–104.

[8] O. Balci, "Verification, validation, and certification of modeling and simulation applications," in Proceedings of the 35th Conference on Winter Simulation, S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, Eds. New Orleans, Louisiana, USA: Winter Simulation Conference, 2003, pp. 150–158.

[9] R. G. Sargent, "Verification and validation of simulation models," in Proceedings of the 37th Conference on Winter Simulation, S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, Eds. Orlando, Florida, USA: Winter Simulation Conference, 2005, pp. 130–143.

[10] O. Balci, "Verification, validation, and testing," in Handbook of Simulation, J. Banks, Ed. Engineering & Management Press, 1998, pp. 335–393.

[11] W. F. van Gunsteren and A. E. Mark, "Validation of molecular dynamics simulation," The Journal of Chemical Physics, vol. 108, no. 15, 1998, pp. 6109–6116.

[12] J. P. Kleijnen, "Verification and validation of simulation models," European Journal of Operational Research, vol. 82, no. 1, 1995, pp. 145–162.

[13] H. Vangheluwe, J. de Lara, and P. J. Mosterman, "An introduction to multi-paradigm modelling and simulation," in Proceedings of the AIS'2002 Conference (AI, Simulation and Planning in High Autonomy Systems), F. Barros and N. Giambiasi, Eds., Lisboa, Portugal, 2002, pp. 9–20.

[14] J. H. Byun, C. B. Choi, and T. G. Kim, "Verification of the DEVS model implementation using aspect embedded DEVS," in Proceedings of the 2009 Spring Simulation Multiconference. San Diego, CA, USA: Society for Computer Simulation International, 2009, p. 151.

[15] H. Saadawi and G. Wainer, "Verification of real-time DEVS models," in Proceedings of the 2009 Spring Simulation Multiconference. San Diego, CA, USA: Society for Computer Simulation International, 2009, p. 143.

[16] M. D. Di Benedetto, S. Di Gennaro, and A. D'Innocenzo, "Diagnosability verification for hybrid automata," in Hybrid Systems: Computation and Control, A. Bemporad, A. Bicchi, and G. Buttazzo, Eds. Springer, Berlin, Heidelberg, 2007, pp. 684–687.

[17] J. Jo, S. Yoon, J. Yoo, H. Y. Lee, and W.-T. Kim, "Case study: Verification of ECML model using SpaceEx," in Korea-Japan Joint Workshop on ICT, 2012, pp. 1–4.

[18] Y. Barlas, "Model validation in system dynamics," in Proceedings of the 1994 International System Dynamics Conference, Sterling, Scotland, 1994, pp. 1–10.

[19] J. W. Forrester and P. M. Senge, "Tests for building confidence in system dynamics models," in System Dynamics, TIMS Studies in Management Sciences, 14, 1980, pp. 209–228.

[20] M. C. Overstreet and R. E. Nance, Characterizations and relationships of world views, R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, Eds. Washington Hilton and Towers, Washington, D.C., U.S.A.: ACM, 2004.

[21] D. Liu, N. D. Macchiarella, and D. A. Vincenzi, "Simulation fidelity," in Human Factors in Simulation and Training, 2008.

[22] J. E. Morrison and L. L. Meliza, "Foundations of the after action review process," Alexandria, VA, p. 82, 1999.

[23] U.S. Department of Defense, "DoD Modeling and Simulation (M&S) Glossary," Tech. Rep., 1997.

[24] R. Fanning and D. Gaba, "The role of debriefing in simulation-based learning," Simulation in Healthcare, vol. 2, no. 2, 2007, pp. 115–125.

[25] J. van den Hoogen, J. Lo, and S. Meijer, "Debriefing in gaming simulation for research: Opening the black box of the non-trivial machine to assess validity and reliability," in Proceedings of the 2014 Winter Simulation Conference, A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, Eds. Savannah, Georgia, USA: IEEE Press, 2014, pp. 3505–3516.

[26] J. Lo, J. van den Hoogen, and S. Meijer, "Using gaming simulation experiments to test railway innovations: Implications for validity," in Proceedings of the 2013 Winter Simulation Conference, R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, Eds. Washington, D.C., USA: IEEE Press, 2013, pp. 1766–1777.

[27] B. Zevin, J. S. Levy, R. M. Satava, and T. P. Grantcharov, "A Consensus-based framework for design, validation, and implementa-

tion of simulation-based training curricula in surgery," Journal of the American College of Surgeons, vol. 215, no. 4, 2012, pp. 580–586.

[28] P. J. Morgan, D. Cleave-Hogg, S. DeSousa, and J. Tarshis, "High-fidelity patient simulation: Validation of performance checklists," British Journal of Anaesthesia, vol. 92, no. 3, 2004, pp. 388–392.

[29] S. I. Gass, "Decision-aiding models: Validation, assessment, and related issues for policy analysis," Operations Research, vol. 31, no. 4, 1983, pp. 603–631.

[30] R. R. Nemani and S. W. Running, "Testing a theoretical climate-soil-leaf area hydrologic equilibrium of forests using satellite data and ecosystem simulation," Agricultural and Forest Meteorology, vol. 44, no. 3-4, 1989, pp. 245–260.

[31] A. Mavin and N. Maiden, "Determining socio-technical systems requirements: Experiences with generating and walking through scenarios," in Proceedings of the 11th IEEE International Conference on Requirements Engineering. IEEE Comput. Soc, 2003, pp. 213–222.

[32] F. Nilsson and V. Darley, "On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings," International Journal of Operations & Production Management, vol. 26, no. 12, 2006, pp. 1351–1373.

[33] M. A. Louie and K. M. Carley, "Balancing the criticisms: Validating multi-agent models of social systems," Simulation Modelling Practice and Theory, vol. 16, no. 2, 2008, pp. 242–256.

[34] F. Landriscina, Simulation and learning. Springer, 2013.

[35] J. Mylopoulos, L. Chung, and E. Yu, "From object-oriented to goal-oriented requirements analysis," Communications of the ACM, vol. 42, no. 1, 1999, pp. 31–37.

[36] O. Balci, "Quality assessment, verification, and validation of modeling and simulation applications," in Proceedings of the 36th Conference on Winter simulation, R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, Eds. Washington, D.C., USA: Association for Computing Machinery, 2004, pp. 122–129.

[37] D. K. Pace, "Modeling and simulation verification and validation challenges," Johns Hopkins APL Technical Digest, vol. 25, no. 2, 2004, pp. 163–172.

[38] R. G. Sargent, "Verification, validation, and accreditation: Verification, validation, and accreditation of simulation models," in Proceedings of the 32nd Conference on Winter Simulation, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, Eds. Orlando, Florida, USA: Society for Computer Simulation International, 2000, pp. 50–59.

[39] S. R. Schach, Classical and object-oriented software engineering (8th edition). McGraw-Hill, 2011.

[40] B. Beizer, Software testing techniques (2nd edition). Van Nostrand Reinhold Company Limited, 1990.

[41] L. G. Stucki, "New directions in automated tools for improving software quality," Current Trends in Programming Methodology, vol. 2, 1977, pp. 80–111.

[42] W. E. Perry, Effective methods for software testing. Wiley Publishing Inc., 2007.

[43] L. A. Miller, E. Groundwater, and S. M. Mirsky, "Survey and assessment of conventional software verification and validation methods," in No. NUREG/CR–6018; EPRI-TR–102106; SAIC–91/6660. Nuclear Regulatory Commission, Washington, DC (United States). Div. of Systems Research; Science Applications International Corp., Reston, VA (United States), 1993.

[44] I. Sommerville, Software engineering (9th edition). Addison-Wesley, Reading, MA, 2004.

[45] G. J. Myers, T. Badgett, T. M. Thomas, and C. Sandler, The art of software testing. John Wiley & Sons, Inc., 2011.

[46] N. Nayani and M. Mollaghasemi, "Validation and verification of the simulation model of a photolithography process in semiconductor manufacturing," in Proceedings of the 30th Conference on Winter Simulation, D. Medeiros, E. Watson, J. Carson, and M. Manivannan, Eds., vol. 2. Washington, D.C., USA: IEEE, 1998, pp. 1017–1022.

[47] M. C. Overstreet and R. E. Nance, "A specification language to assist in analysis of discrete event simulation models," Communications of the ACM, vol. 28, no. 2, 1985, pp. 190–201.

[48] G. A. Mihram, "Some practical aspects of the verification and validation of simulation models," Operational Research Quarterly (1970-1977), vol. 23, no. 1, 1972, pp. 17–29.

[49] P. Bunus and P. Fritzson, "Semi-automatic fault localization and behavior verification for physical system simulation models," in Proceedings of the 18th IEEE International Conference on Automated Software Engineering. IEEE Comput. Soc, 2003, pp. 253–258.

[50] N. Juristo, S. Vegas, M. Solari, and S. Abrahao, "Comparing the effectiveness of equivalence partitioning, branch testing and code reading by stepwise abstraction applied by subjects," in IEEE Fifth International Conference on Software Testing, Verification and Validation (ICST). IEEE, 2012, pp. 330–339.

[51] R. B. Whitner and O. Balci, "Guidelines for selecting and using simulation model verification techniques," in Proceedings of the 21st Conference on Winter Simulation, E. MacNair, K. Musselman, and P. Heidelberger, Eds. Washington, D.C., USA: ACM, 1989, pp. 559–568.

[52] C. F. Hermann, "Validation problems in games and simulations with special reference to models of international politics," Behavioral Science, vol. 12, no. 3, 1967, pp. 216–231.

[53] R. Shannon and J. D. Johannes, "Systems simulation: The art and science," IEEE Transactions on Systems, Man, and Cybernetics, vol. 6, no. 10, 1976, pp. 723–724.

[54] K. J. Cohen and R. M. Cyert, "Computer models in dynamic economics," The Quarterly Journal of Economics, vol. 75, no. 1, 1961, pp. 112–127.

[55] C. Reynolds and R. T. Yeh, "Induction as the basis for program verification," IEEE Transactions on Software Engineering, vol. SE-2, no. 4, 1976, pp. 244–252.

[56] L. G. Birta and F. N. Özmizrak, "A knowledge-based approach for the validation of simulation models: The foundation," ACM Transactions on Modeling and Computer Simulation, vol. 6, no. 1, 1996, pp. 76–98.

[57] F. A. Ackerman, P. J. Fowler, and R. G. Ebenau, "Software inspections and the industrial production of software," in Proceedings of a Symposium on Software Validation: Inspection-Testing-Verification-Alternatives. North-Holland, 1984, pp. 13–40.

[58] H. P. Barendregt, "The Lambda-calculus: Its syntax and semantics," Studies in Logic and The Foundations of Mathematics, 1984.

[59] J. R. Swisher, S. H. Jacobson, J. B. Jun, and O. Balci, "Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation," Computers & Operations Research, vol. 28, no. 2, 2001, pp. 105–125.

[60] W. E. Howden, "Reliability of the path analysis testing strategy," IEEE Transactions on Software Engineering, vol. SE-2, no. 3, 9 1976, pp. 208–215.

[61] R. C. Backhouse, Program construction and verification. Prentice-Hall International, 1986.

[62] E. W. Dijkstra, "Guarded commands, non-determinacy and a calculus for the derivation of programs," in In: Bauer F.L. et al. (eds) Language Hierarchies and Interfaces. Lecture Notes in Computer Science, vol 46. Springer, Berlin, Heidelberg, 1976, pp. 111–124.

[63] J. R. Emshoff and R. L. Sisson, Design and use of computer simulation models. MacMillan, New York, 1970.

[64] A. M. Law and W. D. Kelton, Simulation modeling and analysis (3rd edition). McGraw-Hill, 1991.

[65] R. L. Van Horn, "Validation of simulation results," Management Science, vol. 17, no. 5, 1971, pp. 247–258.

[66] R. H. Dunn, "The quest for software reliability," in Handbook of Software Quality Assurance. New York: Van Nostrand Reinhold, 1987.

[67] O. Balci, "Guidelines for successful simulation studies (tutorial session)," in Proceedings of the 22nd Conference on Winter Simulations, O. Balci, Ed. New Orleans, Louisiana, USA: IEEE Press, 1990, pp. 25–32.

[68] C. V. Ramamoorthy, S.-B. F. Ho, and W.-T. Chen, "On the automated generation of program test data," IEEE Transactions on Software Engineering, vol. SE-2, no. 4, 1976, pp. 293–300.

[69] L. W. Schruben, "Establishing the credibility of simulations," Simulation, vol. 34, no. 3, 1980, pp. 101–105.

[70] R. S. Pressman, Software engineering: A practitioner's approach (8th edition). McGraw-Hill, New York, NY, 2015.