Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A FRAMEWORK FOR SIMULATION VALIDATION & VERIFICATION METHOD SELECTION

Bill Roungas
Alexander Verbraeck

Department of Multi Actor Systems
Delft University of Technology
Jaffalaan 5
Delft, 2628BX, THE NETHERLANDS

Sebastiaan Meijer

Department of Health Systems Engineering
KTH Royal Institute of Technology
Hälsovägen 11C
Huddinge, 14152, SWEDEN

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 time and budget restrictions 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.

1 INTRODUCTION

Back in 1972, based on Forrester's work (1971), Meadows et al. (1972, 1974) 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 then increasing interest in environmental degradation due to human activities (Janssen and De Vries 1999). Even though the model gained support for being "of some use to decision makers" (Meadows et al. 1974) and generated the spark for many later global models, it had several shortcomings, for which it received a lot of criticism as well (Nordhaus 1973). 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 (Balci 1994).

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 (Schlesinger et al. 1979), or in layman terms validation address the question: did we build the "right" model (Balci 2003). On the other hand, verification is the phase of the study that ensures that the model and its implementation are correct (Sargent 2005), or in layman terms verification addresses the question: did we build the model "right" (Balci 2003). V&V has been a well-researched field with a significant amount of produced literature and commercial case studies. This wide range of research has returned numerous methods and techniques, which in turn create 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. 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 2 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 3, a case study is presented in order to illustrate how the proposed framework can be put in practice. Finally, in Section 4, the future potential extensions of the framework are presented and final remarks are made.

2 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.

2.1 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.

2.1.1 Step 1

V&V methods, as indicated by their definitions on Table 1, are different in many aspects; some methods are strictly mathematical whereas others can accommodate for the more qualitative aspects of simulations etc. Balci (1998) identified more than 70 V&V methods, which in turn categorized in four categories as follows: informal, static, dynamic, and formal. By all means, Balci (1998)'s list might not be exhaustive; yet, it is extensive and representative of the body of work that exist to date. This paper adopts the greatest part of this list but not the categorization; since this study actually proposes a new classification of V&V methods. 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 on Table 1.

2.1.2 Step 2

Simulations differ from one another in various ways; distinctions are made on whether they represent an existing system, or whether they simulate a system in a microscopic or macroscopic level, or whether they are intended for learning or decision making, and so forth. This study has identified 10 properties of simulations mainly based on literature but also on common sense (wherever there was a strong indication of the existence of a property but not a literature reference on it). Not all identified properties necessarily influence the selection of a V&V method, therefore this step is not only about identifying the properties but also determines which are the ones that really influence the effectiveness of a method.

The 10 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 (van Gunsteren and Mark 1998), since several methods require some sort of a check on a code level. Hence, this property is included in the analysis.

2. The simulation represents an existing real-system for which real data exist. 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 designed based on, like DEVS, DESS, System Dynamics etc. (Vangheluwe et al. 2002). There have been proposed several frameworks and methods on how to verify and validate DEVS (Byun et al. 2009, Saadawi and Wainer 2009), DESS (Di Benedetto et al. 2007, Jo et al. 2012), or system dynamics models (Barlas 1994, Forrester and Senge 1980), 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 (Overstreet and Nance 2004). While worldviews allow for more concise model descriptions by allowing a model specifier to take advantage of contextual information, there is not any evidence neither logically nor from a literature point of view that they have influence over the V&V method selection, hence they are excluded from the analysis.
5. The fidelity level of the simulation (Low, Medium, High) (Liu et al. 2008). While from a literature point of view there is no evidence to support the influence of the level of fidelity over 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 on the first property and is shown on Table 1, 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) (Morrison and Meliza 1999). This classification, which is adopted by the U.S. Department of Defense (1997), 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 in such a way so as to link the content presented during the session with reality (Fanning and Gaba 2007). It has been demonstrated that debriefing can in general facilitate validation (van den Hoogen et al. 2014, Lo et al. 2013), 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 (Zevin et al. 2012, Morgan et al. 2004), in decision making (Gass 1983), in concept testing (Nemani and Running 1989) 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 (Mavin and Maiden 2003) and CAS with multiple agents (Nilsson and Darley 2006, Louie and Carley 2008) 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 on the modeling process, since different approaches are required (Newtonian physics for object movement, Navier–Stokes equations for fluid behavior etc.) for modeling different systems (Landriscina 2013), literature, or more precisely 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 (Mylopoulos et al. 1999). Validating the simulation's requirements is indeed an important part of the V&V process (Balci 2004), since validation is always relative to the intended use (Pace 2004), 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.

2.1.3 Step 3

According to Sargent (2000), 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 (2000)'s characterization and aims at using it to classify the methods, in addition to the simulations' properties.

2.1.4 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 2.2, a methodology that combines all three steps aiming at the development of a framework for V&V method selection is proposed.

2.2 Methodology

As discussed in Section 2.1.2, dimensions 2, 3, 4, 5, 6, 7, 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 2.1.3, 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.
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.

Table 1: List of V&V methods & properties of simulations.

Method	**1**	**2**	**3**	**4**	**Source**
Acceptance Testing	No	No	Both	O. Val.	Schach (2011)
Alpha Testing	No	No	Both	O. Val.	Beizer (1990)
Assertion Checking	Yes	No	Hard	M. Ver.	Stucki (1977)
Audit	Yes	No	Soft	M. Ver.	Perry (2007)
Beta Testing	No	No	Both	O. Val.	Miller et al. (1993)
Bottom-Up Testing	Yes	No	Both	M. Ver.	Sommerville (2004)
Boundary Value Testing	Yes	No	Both	M. Ver.	Myers et al. (2011)
Cause-Effect Graphing	Yes	No	Hard	M. Ver.	Myers et al. (2011)
Comparison Testing	No	No	Both	C.M. Val.	Nayani et al. (1998)
Compliance Testing	No	No	Soft	O. Val.	Perry (2007)
Control Analysis	Yes	No	Hard	M. Ver.	Overstreet et al. (1985)
Data Analysis Techniques	Yes	No	Hard	D. Val. & M. Ver.	Perry (2007)
Data Interface Testing	No	No	Soft	D. Val.	Miller et al. (1993)
Debugging	Yes	No	Both	M. Ver.	Mihram (1972)
Desk Checking	Yes	No	Both	M. Ver.	Bunus et al. (2003)
Documentation Checking	Yes	No	Both	C.M. Val.	Balci (1998)
Equivalence Partitioning Testing	No	No	Hard	O. Val.	Juristo et al. (2012)
Execution Testing	No	No	Hard	C.M. Val.	Whitner et al. (1989)
Extreme Input Testing	No	No	Hard	O. Val.	Nayani et al. (1998)
Face Validation	No	Yes	Both	O. Val.	Hermann (1967)
Fault/Failure Analysis	No	No	Hard	C.M. Val.	Miller et al. (1993)
Fault/Failure Insertion Testing	No	No	Hard	C.M. Val.	Balci (1998)
Field Testing	No	Yes	Both	O. Val.	Shannon et al. (1976)
Functional (Black-Box) Testing	No	Yes	Hard	C.M. Val.	Myers et al. (2011)
Graphical Comparisons	No	Yes	Both	O. Val.	Cohen et al. (1961)
Induction	No	No	Both	C.M. Val.	Reynolds et al. (1976)
Inference	No	No	Both	C.M. Val.	Birta et al. (1996)
Inspections	No	No	Both	C.M. Val.	Ackerman et al. (1984)
Invalid Input Testing	No	No	Hard	O. Val.	Balci (1998)
Lambda Calculus	Yes	No	Hard	M. Ver.	Barendregt (1984)
Logical Deduction	No	No	Both	All	Whitner et al. (1989)
Model Interface Analysis	No	No	Soft	C.M. Val.	Balci (1998)
Model Interface Testing	No	No	Soft	C.M. Val.	Sommerville (2004)
Object-Flow Testing	No	No	Hard	O. Val.	Swisher et al. (2001)
Partition Testing	Yes	No	Hard	C.M. Val.	Howden (1976)
Predicate Calculus	Yes	No	Hard	M. Ver.	Backhouse (1986)
Predicate Transformations	No	Yes	Hard	M. Ver.	Dijkstra (1976)
Predictive Validation	No	Yes	Hard	O. Val.	Emshoff et al. (1970)
Product Testing	No	No	Both	O. Val.	Schach (2011)
Proof of Correctness	Yes	No	Hard	C.M. Val. & M. Ver.	Backhouse (1986)
Real-Time Input Testing	No	Yes	Hard	O. Val.	Balci (1998)
Regression Testing	Yes	No	Hard	M. Ver.	Whitner et al. (1989)

Reviews	No	No	Both	C.M. Val.	Perry (2007)
Self-Driven Input Testing	No	No	Hard	O. Val.	Law et al. (1991)
Semantic Analysis	Yes	No	Both	M. Ver.	Whitner et al. (1989)
Sensitivity Analysis	No	No	Hard	O. Val.	Van Horn (1971)
Stress Testing	No	No	Hard	O. Val.	Dunn (1987)
Structural (White-box) Testing	Yes	No	Both	C.M. Val.	Beizer (1990)
Structural Analysis	No	No	Hard	C.M. Val.	Whitner et al. (1989)
Submodel/Module Testing	No	No	Both	C.M. Val.	Balci (1990)
Symbolic Debugging	Yes	No	Hard	M. Ver.	Whitner et al. (1989)
Symbolic Evaluation	Yes	No	Hard	C.M. Val.	Ramamoorthy et al. (1976)
Syntax Analysis	Yes	No	Hard	M. Ver.	Beizer (1990)
Top-Down Testing	Yes	No	Both	C.M. Val.	Sommerville (2004)
Trace-Driven Input Testing	Yes	Yes	Both	D. Val. & C.M. Val.	Balci (1998)
Traceability Assessment	Yes	Yes	Both	C.M. Val.	Miller et al. (1993)
Turing Test	No	Yes	Both	O. Val.	Schruben (1980)
User Interface Analysis	No	No	Soft	O. Val.	Balci (1998)
User Interface Testing	No	No	Soft	O. Val.	Pressman (2015)
Visualization/Animation	No	Yes	Both	O. Val.	Sargent (2000)
Walkthroughs	No	No	Both	C.M. Val.	Myers et al. (2011)

3. For what type of requirements is the V&V method more suitable?
   Possible answers: Hard (Functional), or Soft (Non-Functional). 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.

2.3 Discussion

The intended use of Table 1 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 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 second property, i.e. the accessibility to the source code, and in contrary to the first 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 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

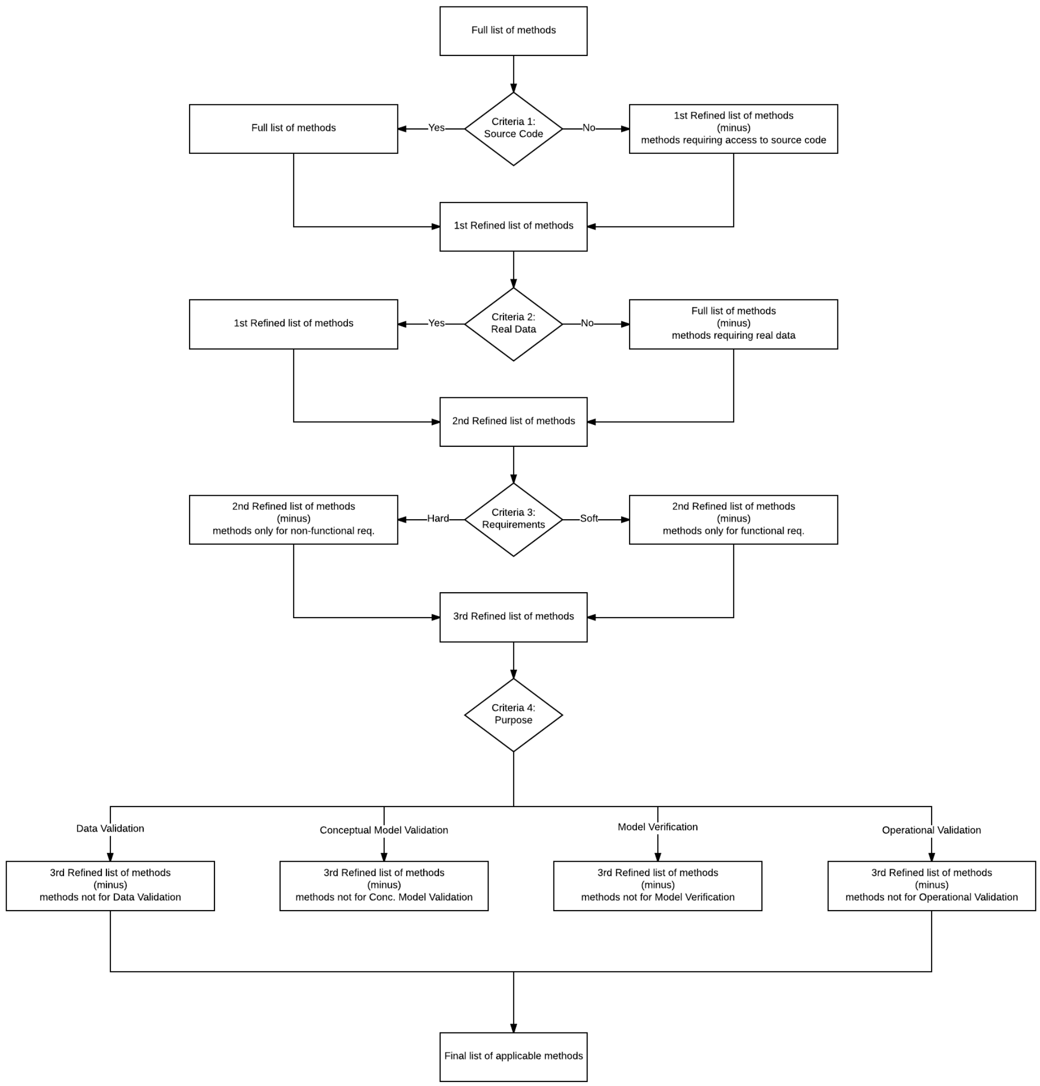

Figure 1: The proposed framework shown as a decision tree.

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 1 per se, but on the idea that the list of V&V methods can be narrowed down to a manageable level (Figure 1), thus making the V&V of a simulation faster, more accurate, and more cost effective.

There is a threat towards the validity of the content on Table 1. 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 4, future steps are proposed aiming at addressing and mitigating the above mentioned threat.

3 A CASE STUDY

In this section, a case study illustrates how the framework, through the use of Table 1, can be used. The case study is a computer simulation of a particular instantiation of the Dutch railway system. The authors

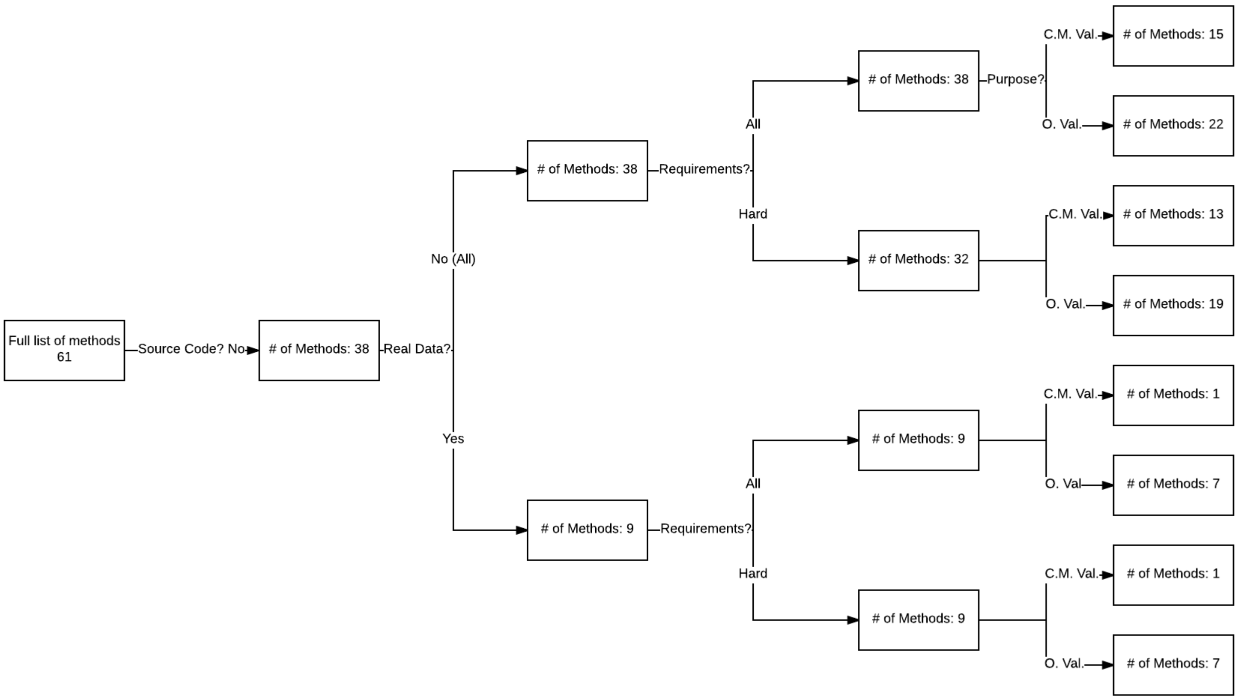

Figure 2: The refined list of V&V methods at each step.

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 1, consists of 61 methods. Then with every step, the list is narrowed down. For this particular study, the filtering of the methods is shown on Figure 2. 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 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 also partially 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.

4 FUTURE WORK & CONCLUSION

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.

With regards to future work, the list of V&V methods can be further enhanced by an extended literature review and perhaps also by conducting interviews with V&V experts. Equally, more simulation properties might exist that may influence the V&V method selection or some of the discarded properties, identified in Section 2.1.2, might be proven to be more influential than initially acknowledged. Finally, there is a need to further verify for each method the purpose for which they are more suitable.

Nevertheless, this paper paves the way for future research in the topic, and as discussed earlier, the novelty of the proposed framework does not lie on the results presented on Table 1, but on 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 would be of great importance for any future research to be focused on these relationships.

ACKNOWLEDGMENTS

This research is supported and funded by ProRail; the Dutch governmental task organization that takes care of maintenance and extensions of the national railway network infrastructure, of allocating rail capacity, and of traffic control.

References

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

Backhouse, R. C. 1986. *Program construction and verification*. Prentice-Hall International.

Balci, O. 1990. "Guidelines for successful simluation studies (tutorial session)". In *Proceedings of the 22nd Conference on Winter Simulations*, edited by O. Balci, 25–32. New Orleans, Louisiana, USA: IEEE Press.

Balci, O. 1994. "Validation, verification, and testing techniques throughout the life cycle of a simulation study". *Annals of Operations Research* 53 (1): 121–173.

Balci, O. 1998. "Verification, validation, and testing". In *Handbook of Simulation*, edited by J. Banks, 335–393. Engineering & Management Press.

Balci, O. 2003. "Verification, validation, and certification of modeling and simulation applications". In *Proceedings of the 35th Conference on Winter Simulation*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 150–158. New Orleans, Louisiana, USA: Winter Simulation Conference.

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

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

Barlas, Y. 1994. "Model validation in system dynamics". In *Proceedings of the 1994 International System Dynamics Conference*, 1–10. Sterling, Scotland.

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

Birta, L. G., and F. N. Özmizrak. 1996. "A knowledge-based approach for the validation of simulation models: The foundation". *ACM Transactions on Modeling and Computer Simulation* 6 (1): 76–98.

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

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

Cohen, K. J., and R. M. Cyert. 1961. "Computer models in dynamic economics". *The Quarterly Journal of Economics* 75 (1): 112–127.

Di Benedetto, M. D., S. Di Gennaro, and A. D'Innocenzo. 2007. "Diagnosability verification for hybrid automata". In *Hybrid Systems: Computation and Control*, edited by A. Bemporad, A. Bicchi, and G. Buttazzo, 684–687: Springer, Berlin, Heidelberg.

Dijkstra, E. W. 1976. "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*, 111–124. Springer, Berlin, Heidelberg.

Dunn, R. H. 1987. "The quest for software reliability". In *Handbook of Software Quality Assurance*. New York: Van Nostrand Reinhold.

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

Fanning, R., and D. Gaba. 2007. "The role of debriefing in simulation-based learning". *Simulation in Healthcare* 2 (2): 115–125.

Forrester, J. W. 1971. *World dynamics*. Wright-Allen Press.

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

Gass, S. I. 1983. "Decision-aiding models: Validation, assessment, and related issues for policy analysis". *Operations Research* 31 (4): 603–631.

Hermann, C. F. 1967. "Validation problems in games and simulations with special reference to models of international politics". *Behavioral Science* 12 (3): 216–231.

Howden, W. E. 1976, 9. "Reliability of the path analysis testing strategy". *IEEE Transactions on Software Engineering* SE-2 (3): 208–215.

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

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

Juristo, N., S. Vegas, M. Solari, and S. Abrahao. 2012. "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)*, 330–339: IEEE.

Landriscina, F. 2013. *Simulation and learning*. Springer.

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

Liu, D., N. D. Macchiarella, and D. A. Vincenzi. 2008. "Simulation fidelity". In *Human Factors in Simulation and Training*.

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

Louie, M. A., and K. M. Carley. 2008. "Balancing the criticisms: Validating multi-agent models of social systems". *Simulation Modelling Practice and Theory* 16 (2): 242–256.

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

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

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

Mihram, G. A. 1972. "Some practical aspects of the verification and validation of simulation models". *Operational Research Quarterly (1970-1977)* 23 (1): 17–29.

Miller, L. A., E. Groundwater, and S. M. Mirsky. 1993. "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)*.

Morgan, P. J., D. Cleave-Hogg, S. DeSousa, and J. Tarshis. 2004. "High-fidelity patient simulation: Validation of performance checklists". *British Journal of Anaesthesia* 92 (3): 388–392.

Morrison, John E. and Meliza, Lary L. 1999. "Foundations of the after action review process".

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

Mylopoulos, J., L. Chung, and E. Yu. 1999. "From object-oriented to goal-oriented requirements analysis". *Communications of the ACM* 42 (1): 31–37.

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

Nemani, R. R., and S. W. Running. 1989. "Testing a theoretical climate-soil-leaf area hydrologic equilibrium of forests using satellite data and ecosystem simulation". *Agricultural and Forest Meteorology* 44 (3-4): 245–260.

Nilsson, F., and V. Darley. 2006. "On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings". *International Journal of Operations & Production Management* 26 (12): 1351–1373.

Nordhaus, W. D. 1973. "World dynamics: Measurement without data". *The Economic Journal* 83 (332): 1156–1183.

Overstreet, M. C., and R. E. Nance. 1985. "A specification language to assist in analysis of discrete event simulation models". *Communications of the ACM* 28 (2): 190–201.

Overstreet, M. C., and R. E. Nance. 2004. *Characterizations and relationships of world views*. Washington Hilton and Towers, Washington, D.C., U.S.A.: ACM.

Pace, D. K. 2004. "Modeling and simulation verification and validation challenges". *Johns Hopkins APL Technical Digest* 25 (2): 163–172.

Perry, W. E. 2007. *Effective methods for software testing*. Wiley Publishing Inc.

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

Ramamoorthy, C. V., S.-B. F. Ho, and W.-T. Chen. 1976. "On the automated generation of program test data". *IEEE Transactions on Software Engineering* SE-2 (4): 293–300.

Reynolds, C., and R. T. Yeh. 1976. "Induction as the basis for program verification". *IEEE Transactions on Software Engineering* SE-2 (4): 244–252.

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

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

Sargent, R. G. 2005. “Verification and validation of simulation models”. In *Proceedings of the 37th Conference on Winter Simulation*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 130–143. Orlando, Florida, USA: Winter Simulation Conference.

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

Schlesinger, S., 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. 1979. “Terminology for model credibility”. *Simulation* 32 (3): 103–104.

Schruben, L. W. 1980. “Establishing the credibility of simulations”. *Simulation* 34 (3): 101–105.

Shannon, R., and J. D. Johannes. 1976. “Systems simulation: The art and science”. *IEEE Transactions on Systems, Man, and Cybernetics* 6 (10): 723–724.

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

Stucki, L. G. 1977. “New directions in automated tools for improving software quality”. *Current Trends in Programming Methodology* 2:80–111.

Swisher, J. R., S. H. Jacobson, J. B. Jun, and O. Balci. 2001. “Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation”. *Computers & Operations Research* 28 (2): 105–125.

U.S. Department of Defense 1997. “DoD Modeling and Simulation (M&S) Glossary”. Technical report.

van den Hoogen, J., J. Lo, and S. Meijer. 2014. “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*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 3505–3516. Savannah, Georgia, USA: IEEE Press.

van Gunsteren, W. F., and A. E. Mark. 1998. “Validation of molecular dynamics simulation”. *The Journal of Chemical Physics* 108 (15): 6109–6116.

Van Horn, R. L. 1971. “Validation of simulation results”. *Management Science* 17 (5): 247–258.

Vangheluwe, H., J. de Lara, and P. J. Mosterman. 2002. “An introduction to multi-paradigm modelling and simulation”. In *Proceedings of the AIS’2002 Conference (AI, Simulation and Planning in High Autonomy Systems)*, edited by F. Barros and N. Giambiasi, 9–20. Lisboa, Portugal.

Whitner, R. B., and O. Balci. 1989. “Guidelines for selecting and using simulation model verification techniques”. In *Proceedings of the 21st Conference on Winter Simulation*, edited by E. MacNair, K. Musselman, and P. Heidelberger, 559–568. Washington, D.C., USA: ACM.

Zevin, B., J. S. Levy, R. M. Satava, and T. P. Grantcharov. 2012. “A Consensus-based framework for design, validation, and implementation of simulation-based training curricula in surgery”. *Journal of the American College of Surgeons* 215 (4): 580–586.

AUTHOR BIOGRAPHIES

BILL ROUNGAS is a PhD researcher in the Department of Multi Actor Systems at Delft University of Technology. His research focuses on the design, validation, and knowledge management of simulations and serious games. His e-mail address is v.roungas@tudelft.nl.

SEBASTIAAN MEIJER is Professor at KTH Royal Institute of Technology, Department of Health Systems Engineering, and at Delft University of Technology, Faculty of Technology, Policy and Management. His email address is sebastiaan.meijer@sth.kth.se.

ALEXANDER VERBRAECK is Professor at Delft University of Technology, Department of Multi Actor Systems and part-time professor in supply chain management at the R.H. Smith School of Business of the University of Maryland. His email address is a.verbraeck@tudelft.nl.