Credibility of Modeling and Simulation via Triangulation

Mariusz Balaban
MYMIC
Portsmouth, VA
mariusz.balaban@mymic.net

ABSTRACT

Triangulation is a strategy for increasing credibility of research. It can be used to investigate the same phenomenon or system through comparison at different levels. The literature review covers state of the art of replication and comparison studies in Modeling and Simulation (M&S). The following discussion identifies main differences between the terms triangulation, replication and scenario. Finally, different dimensions indicating possible configurations of triangulation are explored providing initial guidance on how they could improve credibility of M&S research.

ABOUT THE AUTHORS

Mariusz A. Balaban has two Masters in Mechanical Engineering from Lublin Technical University, Poland. He is currently pursuing a Ph.D. in Modeling and Simulation (M&S) at Old Dominion University. Mr. Balaban is a M&S Scientist/Engineer at MYMIC LLC. His research interests are R&D governance, innovation, evaluation methods, global events modeling, social systems, enterprise decision support systems, maritime systems, applied and methodological aspects related to M&S; especially, the multi-method (hybrid) M&S approach. Mr. Balaban is a member of SCS, ACM SIG, and GK International. His email addresses are Mariusz.Balaban@mymic.net and mbala004@odu.edu.

Credibility of Modeling and Simulation via Triangulation

Mariusz Balaban
MYMIC
Portsmouth, VA
mariusz.balaban@mymic.net

INTRODUCTION

Balaban, Hester, and Diallo (2014) pointed at a pragmatic paradigm as the suitable philosophical stance for a M&S discipline that permits integration of both quantitative and qualitative methods. This generates both challenges and new opportunities. Padilla, Tolk, and Diallo (2013) described methodological challenges of M&S as a discipline. Among others, the M&S discipline should identify its cannons of research that are independent from its multiple application domains. They emphasized a need to increase trust in models, which relates to validity, replicability and tractability. M&S validation can be subjective because there is neither a perfect nor a one hundred percent accurate representation of a system. Harmon and Youngblood (2005) have defined validation as the process of generating information in the quest for truth. They have discussed the risk of the validation process as dependent on the quality of information, which is based on truthfulness as an essential measure of validation of information. They used objectivity, repeatability, timeliness, completeness, and accuracy as attributes of information quality. Considerations given to all these factors can make simulation study expensive, therefore verification and validation (V&V) efforts have to be often balanced with the cost versus benefits analysis.

Campbell and Fiske (1959) pointed out that validation is typically convergent. Triangulation is a strategy for increasing the validity of evaluation and research findings used to investigate the same phenomenon with an intent of convergence (Greene, 2007). Denzin (1970) specifies four types of triangulation: data triangulation, investigator triangulation, model/theory triangulation and method triangulation. Osgood (2007) anticipates an increase of the use of M&S tools that permit crossing a wide variety of levels of analysis and M&S method to facilitate convergence. This paper explores triangulation as a way to induce credibility within M&S research. It discusses a need for a better understanding of types of triangulation in M&S, and how to more systematically design and conduct triangulation studies by clearly specify the boundaries of triangulation.

LITERATURE REVIEW

Axelrod (2007) said that “Replication is one of the hallmarks of cumulative science. It is needed to confirm whether the claimed results of a given simulation are reliable in the sense that they can be reproduced by someone starting from scratch”(p. 11). He focused on the investigator dimension and pointed at challenges related to the lack of detailed reporting of M&S research advocating other forms of sharing information e.g. including a source code, a model, information on how to run it, and how to understand output files. Axelrod (2007) identified three types of equivalences: namely “numerical identity” with exact results, “distributional equivalency” where results cannot be statistically distinguished, and “relational equivalence” with the same internal relationship e.g. dynamics. Wilensky and Rand (2007) considered time, hardware, languages, toolkits, algorithms, and investigator(s) as main dimensions of replication, and proposed this particular order as increasing likelihood of generating different results. They also discussed challenges and proposed what should be included in a published replication work, and what should be included in replicated models. Merlone, Sonnessa, and Terna (2008) presented two strategies concerning implementation of agent-based simulations of a complex phenomenon. In the first approach they progressively implemented more complex models using different methods, which allowed for a deeper understanding of a model. For instance, System Dynamics method allowed for highlighting aspects for which close form solutions could not permit, while Agent Based Modeling (ABM) method allowed for heterogeneity of the agents, but was more difficult to interpret. In the second approach they implemented ABM using different modeling environments “…to determine whether the conceptual simulation model has been correctly translated into a computer program” (p. 11). The second approach was also divided into two options. In the first option, the model would provide a basis for successive implementations, which allowed assessing consistency of code and comparisons of different M&S environments. In the second option models would have independent implementations, which allow additionally for comparing and discussing implementation choices and assumptions. Moreover, they pointed at the importance of examining differences related to error accumulation e.g. floating point operation, different random number generators, or differences in algorithms (small differences can lead to significant variations of results). Miodownik, Cartrite, and Bhavnani (2010) contributed to the body of replication research. Based on their work they noted that the notion of exact replication is a difficult and perhaps unattainable standard, and that the design of replication should be based on required evaluation criteria of model-to-model results. Zhong and Kim (2010) reiterated that even if replication is successful, the opportunity to publish a paper is much smaller as compared to conducting original work. They concluded that replication can improve ABM reliability by forcing replicators to re-examine the conceptual model and assumptions in detail, engaging replicators in validation, and facilitating discussion between original authors and model replicators. Graca and Coelho (2012) conducted replication based on guidelines provided by Wilensky and Rand (2007). They encountered difficulties related to human language ambiguity within documentation, unfamiliarity with technologies used, model code bugs, and methodological limitations. In the light of their work they suggested the use of software engineering best practices to increase model transparency and improve research reporting practices to disclose all necessary data to improve results comparison. Seagren (2014) conducted a successful replication study and noted that this permitted them to additionally implement a richer and more complex version of an original model leading to insight about studied phenomena.

Borshchev and Filippov (2004) demonstrated representation of the same phenomenon based on different M&S methods. They showed and discussed how and when agent's states can be mapped to "stocks" of SD method. Similarly, this can be attempted for Discrete Event Simulation (DES), where an agent's states can be mapped to the process view of DES blocks. In practice, expansion can also take place because of methods' differences, which can mean two things: 1) a modeling error leading to unnecessarily inflated purpose, or 2) discovery of a desirable expansion unforeseen by the original purpose. Sansores and Pavón (2005) focused their efforts on specification of a conceptual model for ABM, which would permit elimination of ambiguities in model translation and execution within different simulation tools e.g. REPAST, MASON, and SDML. Their framework based on a Model Driven Architecture (MDA) pattern allows for code generation for MASON and REPAST providing modelers to contrast the results.

A different purpose of comparison is presented by Gutenschwager, Völker, Radtke, and Zeller (2012). They compared various simulation software tools to evaluate their shortest path algorithms by measuring execution times based on model size for different scenarios. This approach focuses on benefits of tools, where the purpose is not to replicate the results, but to investigate performance or to aid in a decision as to which M&S software should be used for a particular problem. It should be noted that this view does not fit into replication brackets as defined by Wilensky and Rand (2007). A single investigator could often conduct this type of study. Similarly, Agarwal and Juneja (2013) compared a stochastic mesh method and least square methods illustrating the trade-offs in selecting one over others based on improvement in mean square error of the estimator. Zhang, Huddleston, Brown, and Learmonth (2013) evaluated three methods: a close form probability based method, DES, and ABM within the problem of designing a police patrol district. Within a less complex scenario all methods sufficed, while only ABM permitted to represent more complex scenarios accurately. Generally, this type of study can also improve credibility of M&S study by supporting decision about which methods, tools, or algorithms are more adequate to tackle a given problem but itself do not focus on replication of compared results.

DISCUSSION

Wilensky and Rand (2007) defined replication "…as the implementation (replicated model) by one scientist or group of scientists (model replicators) of a conceptual model described and already implemented (original model) by a scientist or group of scientists at a previous time (model builders)" (p. 3). As defined above, replication involves an attempt to build a replica. This definition is quite constraining in the context of different replication types that can be conducted within the M&S discipline. The word "replication" can mean different things in M&S. A "model replication" is a different thing than an "experiment replication". A word replication may also pertain to running multiple scenarios for a simulation model, which is the exploration of simulation model space. Multiple simulation runs are experiment replications, which also have different meanings. For instance, in discrete stochastic simulation models, multiple simulation runs are used for achieving a desirable confidence level/steady state of a single scenario simulation output. In continuous deterministic simulation models, what if analysis are replications conducted by changing values of parameters as separate simulation runs, often called scenarios. In discrete stochastic simulation models exploration of multiple scenarios require additional multiple runs for each scenario to ensure statistically valid output values. As pointed out by Merlone et al. (2008) "…experiment replication can be easily achieved, reimplementation is a more rigorous process in which model is deeply reexamined"(p. 11). A word "reimplementation" could be synonymous to a "model replication" by the addition of the attribute "model". As shown above, there is a need to better specify terms to avoid their ambiguity.

Triangulation can involve different measures for the investigation of the same phenomenon or system with offsetting biases of different study dimensions e.g. investigators, methods, tools, data, and theory (Denzin, 1970; Greene, 2007). It should be pointed out that the purpose of study should specify phenomena and system in consideration in order to define the scope and context of triangulation. This particular view of triangulation, originated in social sciences, can be projected onto M&S (Balaban & Hester, 2013). The main idea in M&S triangulation is to induce credibility by comparison of M&S studies or their phases. This work introduces a term *pseudo-triangulation* as a type of triangulation conducted by the same individual who conducts the original research.

The terms replication, as defined by Wilensky and Rand (2007) and triangulation, as defined by Greene (2007) have similarity, but there are also some differences. Replication indicates induced credibility based on building a replica, which is expected to be similar to an original. Triangulation has less constrained general connotations related to convergence. Convergence can be achieved by focusing on replicability of a considered concept (theory) or differences related to methodological concerns and disambiguation of concepts (theories). Because the broader view is pursued in this paper, the term triangulation is employed.

Triangulation can be evaluated by its permitted variability and achieved convergence. It is anticipated that triangulation will increase or decrease the credibility of a study. Convergence as it pertains to a concept (theory) depends also on methodological concerns and evaluation of the triangulation type, and should be interpreted accordingly. In case of investigator triangulation that aims at confirmation and generates the same results, it confirms representation of phenomena generating more credible basis for insight or decisions. On the other hand, if this triangulation produces significant differences, this decreases credibility and often requires further investigation. The triangulation of internal methodological decisions related to different methods, simulation software tools, algorithms or even multiple different competing concepts of the same phenomenon should be interpreted quite the opposite. The different results between options should aid to choose a more suitable approach. Triangulation in this case when producing the same results can be seen as inconclusive because not any particular choice exhibits merit. On the other hand, if the results are different, they naturally point to a more suitable option, inducing credibility.

The process of creating compared items is a key methodological component influencing induced credibility and it is a base to delineate permitted variability. Reviewed literature identified multiple possible dimensions of triangulation study, which should be systematically analyzed when designing a M&S triangulation. Next, these dimensions are discussed in the context of designing a M&S triangulation.

DESIGN OF M&S TRIANGULATION

In a M&S triangulation based study, stakeholders and/or modelers would have at least a few decisions related to considered co-dependent dimensions: the level of triangulation as a starting point of a study, investigator(s), methods considered and preselected, the M&S environment, and data. These decisions set permitted levels of variability affecting comparability and closeness of results of compared studies and ultimately induced credibility. Please refer to Table 1during the following discussion.

Triangulation levels. Triangulation levels describe a starting point of a study. This work proposes triangulation levels based on a generic M&S research process:

- Purpose (P)
- Research question(s) (R)
- Concept (C) (theory) of phenomena and system(s)
- Simulation (S) model
- Verification and validation (V)
- Experimentation (E)
- Analysis (A)

When specifying a level of triangulation previous levels are asserted and available. The subsequent levels of triangulation decrease permitted variability and possible induced credibility because of limited convergence. For instance, assuming that we aimed at confirmation of results and compared levels would have generated the same results the P level would have permitted the highest level of variability, hence inducing the most credibility when compared to consecutive levels like R level, C level, and so on. During P triangulation only the purpose is given, and the following phases are conducted. R triangulation would have to provide P and R as a starting point. This pattern goes on until E level. It should be noted that not all phases are permitted, which depends on other dimensions. For instance, E level (given all phases leading to it) requires different investigators to analyze output data, describe insights, and possibly recommend decisions. Moreover, the proposed levels can be refined into different smaller levels. For instance, C level could be separated into high and low levels. A high-level could provide an overview of phenomena and system, and some dependencies between them using, e.g., a causal loop diagram. A low-level would operate on constructs of simulation, e.g., constructive methods like DES, ABM, or SD. Similarly, an S level pertains to the M&S environment, and could be separated into hardware and software. Software could be further divided into operating systems, programming languages, simulation software used, and different implementations of algorithms. Triangulation at different levels could provide benefits to investigate different research paths in the context of similarities and differences related to assessing repeatability.

Table 1. Dimensions of M&S triangulation study design

Method		Investigator		Level of triangulation or pseudo-triangulation							Induced Credibility
Same	Different	Same	Different	P	R	C	S	V	E	A	
X		X		No PseudoT	PseudoT Increase variability		No PseudoT (run replications, scenarios) Choice of concept				Between very low and low
	X Preselect (P,R,C)		X	Increase variability					Triangulation (T)	No T	Between medium and very high
X			X	Increase variability					Triangulation (T)	No T	Between medium and high
	X Preselect (P,R,C)	X		No PseudoT	PseudoT Increase variability		No PseudoT (replications, scenarios) Choice of concept				Between low and medium

Investigator. A triangulation study conducted by multiple investigators is scientifically a more adequate approach and has potential to induce more credibility than a pseudo-triangulation study. Pseudo-triangulation is problematic and it can be seen as unattainable as pure triangulation, situated between comparison and expansion. At the P level, a single investigator can propose additional research questions, which would expand the research scope. A triangulation conducted by the same individual within the established purpose does not facilitate comparability of results but exploration through additional questions. At the R level, multiple concepts of the same phenomenon and the following developments could be considered by an individual based on established research question(s). Concept divergence could be helpful to direct toward a theory convergence in case of subjective phenomena. The multiple concepts and their simulation models built by the same individual compared against real world data could produce refinements, and should ultimately lead toward a single unified concept. It should be emphasized that because the concepts are created by the same individual, they are not independent and objectivity of triangulated views is affected. On the other hand, pseudo-triangulation could help a single investigator to refine and support the choice of a concept within the study boundaries, increasing credibility of research. Even a pseudo-triangulation at R level using the same method could induce some credibility when separate simulation models are implemented, analyzed, documented and disseminated. It should be noticed that the starting point of pseudo-triangulation pertains only to the R and C levels, and above them pseudo-triangulation is not applicable because the same investigator can only explore the simulation model(s) in the context of given research question(s) and concepts (including implementation options and methodological concerns). Additional alternative designs of experiments or additional analysis would explore a simulation model, but not necessarily serve as a confirmation of the results, as in the case when conducted by separate modelers. As discussed above, a single modeler could to some degree benefit from pseudo-triangulation, but engaging different modelers would facilitate a more objective triangulation.

Methods. Different methods can be used during triangulation and pseudo-triangulation studies to further increase credibility of M&S research. Methods or a combination of methods used on separate instances should be able to realize the same concepts and possibly similar results for comparison, but through the lenses of different methods. Because the differences in methods could cause different results, the comparison of methods would be a part of explaining the differences in produced theories. In the situation when different methods continuously produce the same or very similar results based on the same situation, it may be said that to some degree the measures arising from different methods are suitable to triangulate the given situation (Ghrayeb, Damodaran, & Vohra, 2011). This way, one could approach confirmation of correctness of a triangulation study of given methods/measures. On the other hand, if a triangulation study produces significant differences, this indicates a need for further exploration, including the methods used. When triangulation or pseudo-triangulation aim at methodological choice, possible alternative methods can be used to implement simulation models based on an established conceptual model. The difference in results can provide more valuable indications of better methodological choices inducing credibility.

Methods preselecting. Another option that can influence triangulation research design pertains to preselecting methods. This means that the methods may be artificially imposed, which can influence the rest of the process. From the perspective of comparability of results the use of different methods can lead to more incommensurable simulation models. Method(s) could be preselected in order to lower variability of results by considering the same method(s) as the original M&S study at P, R, or C levels. It is important to notice that triangulation at a low concept level is a threshold point at which methods must be selected in order to implement a simulation model. Pass the C level methods would be given and would not be considered as a design option. The question is if preselecting method(s) at P and R levels is a justifiable practice when considering how this can limit possible variability. On the other hand, without the preselecting of methods incomparability of results is more likely.

Data. Depending on the purpose of the M&S study, data activities may involve both a search for secondary data and/or collection of primary data, which both can shape the triangulation study design. Especially when primary data collection is desirable and feasible this imposes additional considerations. Different M&S study design paths are possible. In the first case data collection and/or a data search could take place between P and C levels. This adds flexibility to examine phenomena but may constrain the M&S part of the study if the modeler waits for data before starting the model development. Another approach could be to consider developing the simulation model first and exploring which data should be collected based on the value of information that the data could facilitate. Yet another design could involve multiple phases that align both data and simulation layers according to larger intervention schema, as shown in Figure 1. When more data becomes available (e.g. qualitative insight changing theoretical view) additional M&S triangulation studies are conducted.

CONCLUSIONS

M&S domain should embrace complementarity (expansion) of methods and triangulation (comparison) as its research guiding principles (Balaban & Hester, 2013). While complementarity of methods offers enhancement to expansion of knowledge through flexibility, expressiveness, and uniqueness of representation, triangulation focuses on scientific evaluation to induce credibility of M&S research through comparison within different dimensions. This paper reviewed relevant M&S comparison studies, and proposed initial view of M&S triangulation study design dimensions. Future work requires in depth analysis of dimensions to eliminate possible ambiguities between their dependencies, and derivation of the scale for induced credibility in the context of expected and obtained results from M&S triangulation. This can be supported by multiple case studies that explore different possible configurations of triangulation.

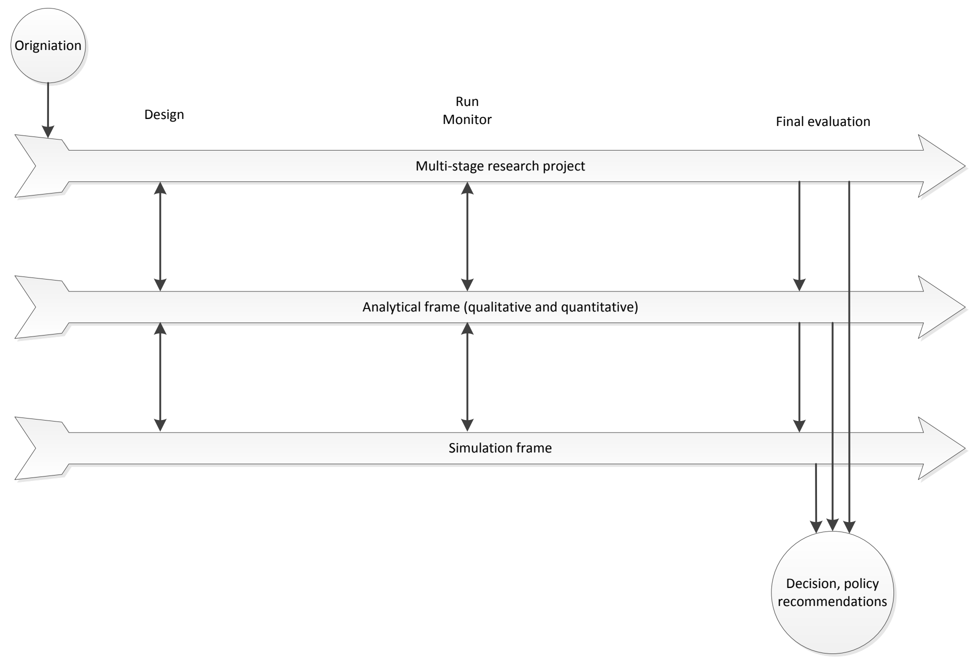

Figure 1. M&S activity within a larger multistage research project

REFERENCES

Agarwal, Ankush, & Juneja, Sandeep. (2013). *Comparing optimal convergence rate of stochastic mesh and least squares method for bermudan option pricing.* Paper presented at the Winter Simulation Conference.

Axelrod, R. (2007). Simulation in the social sciences. In J.-P. Rennard (Ed.), *Handbook of research on nature inspired computing for economy and management* (pp. 90-100). Hershey, PA: Idea Group Inc.

Balaban, M. A., & Hester, P. (2013). *Exploration Of Purpose For Multi-Method Simulation In The Context Of Social Phenomena Representation*. Paper presented at the Simulation Conference (WSC), Proceedings of the 2013 Winter, Washington D.C.

Balaban, M. A., Hester, P., & Diallo, S.Y. (2014). *"Towards a Theory of Multi-Method M&S Approach: Part I"*. Paper presented at the Proceedings of the 2014 Winter Simulation Conference (accepted), Savannah, GA.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*(2), 81.

Denzin, Norman K. (1970). The research act in sociology: A theoretical introduction to sociological methods.

Ghrayeb, Omar, Damodaran, Purushhothaman, & Vohra, Promod. (2011). Art of triangulation: an effective assessment validation strategy. *Global Journal of Engineering Education, 13*(3), 96-101.

Graca, Joao, & Coelho, Helder. (2012). *Replicating Hofstede's Cultured Negotiation: A First Update.* Paper presented at the Social Simulation (BWSS), 2012 Third Brazilian Workshop on.

Greene, J.C. (2007). *Mixed methods in social inquiry* (Vol. 9): Jossey-Bass.

Gutenschwager, Kai, Völker, Sven, Radtke, Axel, & Zeller, Georg. (2012). *The shortest path: Comparison of different approaches and implementations for the automatic routing of vehicles.* Paper presented at the Proceedings of the Winter Simulation Conference.

Harmon, SY, & Youngblood, Simone M. (2005). A proposed model for simulation validation process maturity. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 2*(4), 179-190.

Merlone, Ugo, Sonnessa, Michele, & Terna, Pietro. (2008). Horizontal and vertical multiple implementations in a model of industrial districts. *Journal of Artificial Societies and Social Simulation, 11*(2), 5.

Miodownik, Dan, Cartrite, Britt, & Bhavnani, Ravi. (2010). Between replication and docking:" Adaptive agents, political institutions, and civic traditions" revisited. *Journal of Artificial Societies and Social Simulation, 13*(3), 1.

Osgood, N. (2007). Using traditional and agent based toolsets for system dynamics: present tradeoffs and future evolution. *Proceedings of the 25th international conference of the system dynamics society*.

Padilla, Jose J, Tolk, Andreas, & Diallo, Saikou Y. (2013). *M&S methodological challenges.* Paper presented at the Proceedings of the Emerging M&S Applications in Industry & Academia/Modeling and Humanities Symposium.

Sansores, Candelaria, & Pavón, Juan. (2005). Agent-based simulation replication: A model driven architecture approach *MICAI 2005: Advances in Artificial Intelligence* (pp. 244-253): Springer.

Seagren, Chad W. (2014). A Replication and Analysis of Tiebout Competition Using an Agent-Based Computational Model. *Social Science Computer Review*. doi: 10.1177/0894439314534810

Wilensky, Uri, & Rand, William. (2007). Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation, 10*(4), 2.

Zhang, Yue, Huddleston, Samuel H, Brown, Donald E, & Learmonth, Gerard P. (2013). *A comparison of evaluation methods for police patrol district designs.* Paper presented at the Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World.

Zhong, Wei, & Kim, Yushim. (2010). Using model replication to improve the reliability of agent-based models *Advances in Social Computing* (pp. 118-127): Springer.