

What contributes to the quality of simulation results?

Jan Himmelspach

Department of Computer Science
University of Rostock
Rostock 18055, Germany

Adelinde M. Uhrmacher

Department of Computer Science
University of Rostock
Rostock 18055, Germany

ABSTRACT

Simulation as a method is widely used throughout industry and science. Thereby, the question arises how much trust can we put into the results achieved. The quality of the simulation results depend not only on the efforts put into the model and the simulation study, but also in designing the underlying simulation tools and as such in the quality of the simulation methods used: an aspect that is often neglected in the evaluation of simulation studies. Re-use, documentation, and repeatability of experiments have long been identified as core concepts in the quality of simulation studies, however, they likewise apply to the simulation software used. The many interdependencies in achieving high quality simulation results, and possible steps towards this goal will be illuminated by discussing general quality aspects and the role of emerging modeling and simulation frameworks in this process.

1 INTRODUCTION

Surveys like (Pawlikowski, Jeong, and Lee 2002) reveal that the quality of many simulation studies is amazingly poor: when the authors checked whether appropriate random number generators and simulation analysis procedures have been exploited in network simulation studies they found that the majority of the more than 2000 considered cases neglected these basic ingredients of a trustworthy simulation study. They concluded that the issue of valid simulation experiments in addition to valid models requires more attention if the crisis of (modeling and simulation) M&S credibility

shall not deepen further. Also (Verification, Validation and Accreditation) (VV&A) emphasize that the entire model life cycle, including the simulation experiments, has to be designed carefully to ensure the quality of M&S (Balci 1998, Robinson 2002, Troitzsch 2004, Law 2007, Sargent 2008, Leye et al. 2009). In the following we would like to take this appeal a step further, by illuminating yet another facet of M&S quality: what about the software that underlies modelling, simulation, and interpreting the simulation results?

Thereby, we broaden the scope of “products” that quality might refer to (Balzert 2008, p. 460) in M&S: the model, the simulation experiment, the predictions, the M&S methods, and the implementation. Also quality itself is not a homogeneous concept, e.g., (Balzert 2008) distinguishes quality according to being:

(a) Transcent: Quality is identifiable, absolute, unique and complete. It means that high standards and requirements for products have to exist and are consequently applied. Nevertheless quality can’t be absolutely defined and measured, about quality one can only judge with experience.

(b) Product centered: The quality of a product is exactly defined, it describes the product and can not only be measured, it can be used to show differences in quality and thus it allows to rank products. The needs of customers do not need to be completely reflected thereby.

(c) User centered: User centered quality means that a user decides whether a product has a certain quality or not, thus the decision can be very subjective.

(d) Process view: Quality is created by the correct creation of the product, which is ensured by well-defined processes (to make it “right the first time”).

(e) Cost/Use relation: Quality can be seen as a function of production costs and usability, and thus a product of a high quality is a product with a certain use and acceptable costs.

If we look at the typical VV&A processes, all of these qualities play a certain role. Verification emphasizes a rather production-centered view on quality, e.g., whether a product has been created in the right manner. Validation is associated with matching the results of modelling and simulation with other sources of knowledge, e.g., expectations, and data. Whereas face validation is a comparatively user-centered concept, statistical procedures in validation focus on the properties of the achieved product, the simulation results. One central goal in VV&A is establishing work-flows that shall ensure a high quality emphasizing the process view on quality. Accreditation assumes that quality of M&S can be identified in a transcendent manner. This is based on the confirmation by (external) experts, and associated measures, e.g., the curation of model bases. VV&A processes are expensive, thus how to reduce the costs of VV&A is always an issue. How are these different aspects of quality reflected in the different products?

2 Quality in M&S

2.1 Quality of models

To apply VV&A processes in model building to ensure the quality of models is widely acknowledged in M&S, as documented by regular tutorials at large conferences, e.g., Winter Simulation, and the vast amount of literature, e.g., (Robinson 2002, Balci 2003, Brade 2004, Troitzsch 2004, Law 2007, Sargent 2008, Wang and Lehmann 2008, Leye et al. 2009). However, there is still a lack of deeply integrated validation functionality in M&S products (Leye et al. 2009), and a surplus of models that have not been sufficiently validated. These observations might not be entirely independent. Whereas some reasons for insufficiently validated models might be located in the application area, e.g., the lack of data to validate a model against, other reasons might refer to more technical aspects of M&S. Among these how well is the validation process including the validation procedures suitable for the model supported in the tool, is the simulation engine sufficiently efficient to execute the required number of simulation runs in a reasonable time, and whether it works correctly or not (Edmonds and Hales 2003). Confidence in a model and thus its quality is coupled to costs: for the confidence to be higher, one needs to invest more (Sargent 2008) or one needs to lower the ante, e.g., to

facilitate VV&A processes technically by developing new methods and implementations.

2.2 Quality of M&S methods

When to use what method is a central but difficult question to answer, due to the plethora of modeling formalisms, simulation algorithms, and statistical procedures. Whether a particular model formalism is more suitable for a specific system and the simulation objective than another, is difficult to fasten on objective criteria. Whereas it is well known that numerical integration might introduce a significant bias into the achieved simulation results (if not carefully chosen) the idea that random number generators are crucial for discrete event simulation has only slowly taken hold in the M&S community and their users. Simulation runs produce multi-dimensional data that changes over time and provide ample challenges for state of the art statistical and visual analysis methods, as well.

Often answers when to use which method can only be found experimentally, e.g., by exploiting performance data bases and machine learning tools (Ewald, Uhrmacher, and Saha 2009). However, in this case methods can not clearly be separated from their implementation.

2.3 Quality of M&S software

The development of software can be supported by a variety of techniques ranging from general bottom-up to top-down design approaches, development workflows, development techniques like model driven development to development team techniques like extreme programming (Beck 1999). Each of the techniques is aimed at improving the software development process and at increasing reliability and maintainability of the software product, and thus the confidence in it. However, often these techniques imply an overhead, or the existence of teams of developers (e.g., at least two persons are needed for extreme programming), which are not always available in simulation projects. Testing is one of the most important steps in software development. As M&S tools are intrinsically complex, testing requires some effort. Diverse data structures and algorithms are combined, and bugs can be in any of them influencing simulation results and the derived predictions and decisions. In addition these data structures and algorithms are often highly specialized and expert knowledge is required to assess their correctness, e.g., are they implemented in a numerical safe manner, are they applicable at all, which numerical integration algorithm, which random number generator, which optimization method, which statistical analysis methods can be used, and how to best implement these.

One of the major problems in M&S is that many different fields are intertwined, and thus it is difficult to get a

complete overview and understanding of m&s techniques (Balci 1990). Many techniques require very good programming skills, especially if the performance of the algorithms shall be high, or if it comes to special execution environments like distributed computation. Thus, a background in M&S methodology, software engineering, programming skills, and the application domain is required to create high quality m&s products. Altogether a background that is not easy to find, and difficult to achieve in a few couple of years, that typically limit simulation projects.

That well tested, and documented M&S software will improve the overall quality of M&S and that reuse will play a key role in reducing costs and improving comparability, was stated already in the mid 90s (Minar, Burkhart, Langton, and Askenazi 1996). Consequently libraries and frameworks for M&S, which allow a broad reuse and which can be extended and tested by a variety of people are a key towards achieving high quality results in M&S.

3 The Role of Reuse in M&S

Reuse can be seen as a chance towards higher quality: a product developer needs not to be an expert in all areas, e.g., it can be sufficient to know that an event queue is a priority queue in order to use it implementation details are not required. This is the idea behind “component-based” approaches: the reuse of well-defined components reduces the costs, and projects become manageable which would take too long without reuse. However, “trust” is central for reuse – why should you believe that something which otherwise someone else has been created is sound, thus that it does what it shall do with a sufficient quality? Among the requirements are at least good documentation (of interface and internals, especially of restrictions, and dependencies), but in addition it should be clear which tests have been passed. Luckily reuse itself can help here as well: if tests can be reused for alternative implementations, and if these are constantly extended, such a test suite can be much better than a set of tests quickly created for testing on ones own.

3.1 Quality aspects in reuse

How are the different notions of quality reflected in the context of reuse?

Transcent: Components have to be curated by experts which can judge about their quality, an approach, e.g., undertaken for models in the curated branch of the BioModels database.

Product centered: For some M&S products quality measurements exist (i.e., benchmarks) which make M&S products or parts of them (partially) comparable. However, model quality is hard to measure in the end, and benchmarks can be misleading: what’s the worth of 10% percent speedup if development and maintenance costs are much higher?

User centered: In M&S users differ quite significantly, we have at least developers of M&S software and methods, users of M&S software developing models, executing, and analyzing simulation studies, and decision makers interested in the results of simulation. Within each the definition of quality varies as well, e.g., for developers efficient and sound implementations of methods defines a quality component for reuse, for modelers the ease of composing models, and for decision makers models and simulation experiments producing valid predictions in a wide range of settings. In addition, M&S is applied in many different areas of science, and with this area the respective science culture varies and with it what e.g., is expected of a validation process.

Process centered: This view can be applied to M&S software development as well as to modeling and simulation as such. Quality is the result of the correct creation of a product, something which might be easier if you can build on pre-defined bricks. Each required building process can be described in detail, and this can help beginners and experts to create M&S solutions of a higher quality.

Cost/Use: For M&S this means that the best precision with highest execution performance (M&S software), and a good, and sufficiently validated model for a certain price should be achieved. Consequently quality should increase the more we reuse, if reuse has lower costs than redevelopment.

Thus, reuse might be able to address many different aspects of quality when applied not only to modeling and simulation experiments but also to the software being used as well.

3.2 Reuse of models / model components

Model reuse is pretty common in some domains, where the potential building blocks are well known and standardized (e.g., electrical engineering or manufacturing), but is difficult to achieve in other domains such as sociology or systems biology. Currently several steps are undertaken to support the reuse of models, e.g., to provide unambiguous interface descriptions, to validate components carefully, to curate them using a 3rd party to check their quality, and to protocol simulation experiments executed.

Thereby, simulation experiments become a subject of reuse, the repeatability of simulation results increases our trust in the model and it can be used to increase our trust into M&S software as well (Jeschke and Ewald 2008).

3.3 M&S software

Many different M&S software products are available - ranging from application domain specific solutions, to “single model” tools, or just those being designed for a certain hardware infrastructure. If no existing tool can be reused and a tool has to be developed, one is likely to enter a complex, and error prone workflow. The development of good

and reliable software implies careful software design, implementation, and validation processes. In M&S it additionally implies deep knowledge of techniques required for M&S (Balci 1990). Thus, the quality of the software is dependent on the development process as such, and on the abilities of those working on the different steps. Reuse during development can significantly lower the efforts needed to create new software: libraries can provide functionality which can be easily integrated into a new application. If these are commonly used, and thus repeatedly tested, their reliability should grow, and thus help to support the reliability of simulation runs. Frameworks combine the functionality of a library with additional pre-defined processes (combining the functionality), and are intended to support the development of domain specific applications. Thus, subjects of reuse are functionalities as well as processes.

M&S frameworks are particularly suited to address the quality concerns not only at the level of M&S software and methods, but also at the level of modeling and simulation as well. For example, the flexible and scalable experimentation layer in JAMES II (Himmelspach, Ewald, and Uhrmacher 2008) makes use of a “plug’n simulate” concept (Himmelspach and Uhrmacher 2007) to execute experiments on different hardware infrastructures and to configure an efficient simulator on demand (Ewald, Himmelspach, and Uhrmacher 2008) or variants of simulation experiments can be executed with only marginal changes. Both “optimize” the cost / use relation with respect to execution time and the effort required in setting up and documenting experiments. Thus, the costs in validating models is reduced, which contributes to the model’s quality. These are the same means that can be exploited for testing new M&S implementations, thus adding to their quality. A seamless integration of validation means into tools can lower the barrier to do validation. Guiding users by well-defined workflows will improve quality as well.

4 Workflows in M&S

The process-oriented view on quality emphasizes the need to pursue established procedures in producing results and to execute each step of those procedures with care.

On first sight two workflows in M&S can be distinguished: modeling and simulation, i.e., the generation of models and the execution of simulation experiments. However, both are intertwined, as model creation requires simulation experiments. In validation these experiments become the focus of attention, however they might have an effect on the model, thus activating the model creation cycle. This is even more apparent when optimization or calibration is the objective of the simulation experiments.

However, another workflow exists that is typically not considered in M&S at all, although it depends on and influences the execution of simulation experiments: the workflow for designing and evaluating M&S software. This workflow which is tightly interwoven with the workflows above is dedicated to the “other products” of M&S, i.e., the development of methods and tools.

The identification of workflows, and the definition of minimal requirements of quality for each of these, in combination with a sufficient documentation of the overall experimentation process is mandatory (see also ISO 9001). Thus, M&S tools should support users on these tasks, e.g., by workflow controlled processes or by “auto-documentation” of models, experiments, and results thereof. This can improve the cost / use relation (by eliminating “orientation phases”) and the quality of the product from the user’s point of view, as the combination of guidance and reuse through the processes of model generation, designing simulation experiments, evaluating them, or evaluating M&S methods will increase the trust of the use in the results achieved.

5 CONCLUSION

M&S research produces a variety of products, including models, predictions, and tools. Quality applies to all of those. Quality itself is multi-faceted, it might be defined in product-centered, user-centered manner, it might take third parties into account, or might turn the emphasis on the process of generating the product. In VV&A many of these different facets are reflected. However, VV&A in M&S focuses on the quality of models and predictions, and takes the quality and reliability of M&S tools for granted. We argue that this presumption might lead to deepen the credibility crisis of simulation results. The generation and evaluation of M&S methods requires more attention, also to ensure the quality of models and predictions. To address quality concerns in a comprehensive manner we identified the concept of reuse as being central. M&S frameworks combine reuse of functionalities and processes together with an easy extension and thus could be the answer for different M&S users in a wide variety of application domains to address various quality concerns.

REFERENCES

- Balci, O. 1990. Guidelines for successful simulation studies (tutorial session). In *Proceedings of the 1990 Winter Simulation Conference*, ed. O. Balci, 25–32. Piscataway, NJ, USA: IEEE Press.
- Balci, O. 1998. Verification, validation, and accreditation. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 41–4. Los Alamitos, CA, USA: IEEE Computer Society Press.

- Balci, O. 2003. Verification, validation, and certification of modeling and simulation applications: verification, validation, and certification of modeling and simulation applications. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, 150–158: Winter Simulation Conference.
- Balzert, H. 2008. *Lehrbuch der Softwaretechnik - Softwaremanagement*. 2 ed. Spektrum.
- Beck, K. 1999, Oct. Embracing change with extreme programming. *Computer* 32 (10): 70–77.
- Brade, D. 2004. *A generalized process for the verification and validation of models and simulation results*. Ph. D. thesis, Universitt der Bundeswehr Mnchen.
- Edmonds, B., and D. Hales. 2003. Replication, replication and replication: Some hard lessons from model alignment. *J. Artificial Societies and Social Simulation* 6 (4).
- Ewald, R., J. Himmelspach, and A. M. Uhrmacher. 2008, June. An algorithm selection approach for simulation systems. In *Proceedings of the PADS*, 91–98. Rome, Italy: IEEE Computer Society.
- Ewald, R., A. Uhrmacher, and K. Saha. 2009. Data mining for simulation algorithm selection. In *Proceedings of the SIMUTools'09: 2nd International Conference on Simulation Tools and Techniques*.
- Himmelspach, J., R. Ewald, and A. M. Uhrmacher. 2008, December. A flexible and scalable experimentation layer for JAMES II. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. Mason, R. Hill, L. Moench, O. Rose, J. Fowler, and T. Jefferson, 827–835: IEEE Computer Society.
- Himmelspach, J., and A. M. Uhrmacher. 2007, March. Plug'n simulate. In *ANSS '07: Proceedings of the 40th Annual Simulation Symposium*, 137–143. Washington, DC, USA: IEEE Computer Society.
- Jeschke, M., and R. Ewald. 2008. Large-scale design space exploration of ssa. In *Computational Methods in Systems Biology, International Conference, CMSB 2008, Rostock, Germany, October 12-15, 2008, Proceedings*, Volume 5307 of *Lecture Notes in Computer Science*, 211–230. Berlin / Heidelberg: Springer.
- Law, A. M. 2007. *Simulation modeling and analysis*. 4 ed. McGraw-Hill International.
- Leye, S., J. Himmelspach, and A. M. Uhrmacher. 2009, March. A discussion on experimental model validation. In *Proceedings of the 11th International Conference on Computer Modeling and Simulation*, 161–167: IEEE Computer Society.
- Minar, N., R. Burkhart, C. Langton, and M. Askenazi. 1996, Juni. The SWARM simulation system: a toolkit for building multi-agent simulations. Technical report, Santa Fe Institute.
- Pawlikowski, K., H.-D. Jeong, and J.-S. Lee. 2002, Jan. On credibility of simulation studies of telecommunication networks. *Communications Magazine, IEEE* 40 (1): 132–139.
- Robinson, S. 2002. General concepts of quality for discrete-event simulation. *European Journal of Operational Research* 138 (1): 103 – 117.
- Sargent, R. G. 2008. Verification and validation of simulation models. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. Mason, R. Hill, L. Moench, O. Rose, J. Fowler, and T. Jefferson, 157–169.
- Troitzsch, K. G. 2004. Validating simulation models. In *18th European Simulation Multiconference. Networked Simulations and Simulation Networks*, ed. G. Horton, 265–270. The Society for Modeling and Simulation International: SCS Publishing House.
- Wang, Z., and A. Lehmann. 2008, October. Expanding the V-Modell XT for verification and validation of modelling and simulation applications. In *System Simulation and Scientific Computing*, 404–410. Asia Simulation Conference.

AUTHOR BIOGRAPHIES

JAN HIMMELSPACH is a post doc in the modeling and simulation group at the University of Rostock. His e-mail address is <jan.himmelspach@uni-rostock.de>.

ADELINDE M. UHRMACHER is a professor in the Computer Science Department at the University of Rostock. Her e-mail address is <adelinde.uhrmacher@uni-rostock.de>.