*Proceedings of the 2016 Winter Simulation Conference*
*T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.*

# PANEL – REPRODUCIBLE RESEARCH IN DISCRETE EVENT SIMULATION – A MUST OR RATHER A MAYBE ?

Adelinde M. Uhrmacher
School of Computer Science and Electrical Engineering
University of Rostock
18059 Rostock, GERMANY

Sally Brailsford
Business School
University of Southampton
Southampton, SO17 1BJ, UK

Jason Liu
School of Computing and Information Sciences
Florida International University
Miami, FL 33199, USA

Markus Rabe
Department of IT in Production and Logistics
Technical University Dortmund
44227 Dortmund, GERMANY

Andreas Tolk
Simulation Engineering Department
The MITRE Corporation
Hampton, VA 23666, USA

## ABSTRACT

Scientific research should be reproducible, and as such also simulation research. However, the question is – is this really the case? In some application areas of simulation, e.g., cell biology, simulation studies cannot be published without data, models, methods, including computer code being made available for evaluation. With the applications and methodological areas of modeling and simulation, how the problem of reproducibility is assessed and addressed differs. The diversity of answers to this question will be illuminated by looking into the area of network simulations, simulation in logistics, in military, and health. Making different scientific cultures, different challenges, and different solutions in discrete event simulation explicit is central to improving the reproducibility and thus quality of discrete event simulation research.

## 1 INTRODUCTION

Simulation is an experiment performed with a model and complements theory and real world experiments (Winsberg 2010). As an experimental science one would assume that repeatability, replicability, and reproducibility are deeply rooted within the scientific culture of simulation. However this is not the case, as observed by (Pawlikowski et al. 2002) and (Kurkowski et al. 2005) and recently (Sarkar and Gutiérrez 2014) in the scope of network simulation, and by (Merali 2010) and (Joppa et al. 2013) in general scientific computing: often the documentation is too incomplete to replicate or reproduce the research results by a third party. Requirements for reporting simulation experiments like *Minimum Information About a Simulation Experiment* (MIASE) (Köhn and Novère 2008), or *Minimum Model Reporting Requirements* (MMRR) (Rahmandad and Sterman 2012) have been proposed, requiring major documentation efforts. For example, according to MIASE the documentation should contain (in addition to the model) information about model composition and configuration, the simulation methods used (for instance, stochastic simulation algorithm), the tasks performed, and the complete collection of outputs. These requirements in documenting research results for an improved replicability or reproducibility by a third party are only fragmentally considered throughout the simulation community. Often due to too few replications or other mistakes in executing the simulation, even the repeatability of a simulation (repeating the same simulation by the same people achieving the same results) might not be granted.

Research findings of discrete event simulation are often based on or illuminated by executing simulation experiments. Thus a documentation of simulation experiments, e.g., by exploiting domain specific languages for describing, specifying and executing experiments (Waltemath et al. 2011, Ewald and Uhrmacher 2014, Schützel et al. 2014), appears central. Also the other artifacts that are produced in discrete event simulation which might include conceptual, formal, and implemented model, software, data, test suites, other digital or mathematical artifacts and which might be closely linked with the research findings should be easily accessible and reusable. This leads us to a central motivation of pushing reproducibility. It is less mistrust or the wish to adhere to a scientific norm, but rather helping that future researchers can more effectively build on previous research. Therefore, we will also abstract from a more technical point of view how to obtain the exact same simulation results (Dalle 2012) and concentrate on the current state of reproducible research, constraints that apply, and individual remedies to improve replicability and reproducibility of simulation research in the different areas.

## 2 REPRODUCIBILITY: A PROBLEM OF APPROACHES OR APPLICATIONS ?

The panel has been located in the agent-based track of the Winter Simulation Conference. In the area of agent-based modeling the problem to ensure replicability appears particularly challenging, as no standard for describing agent-based models does exist, many applications still resort to implementing models in a general purpose host language, and models tend to be complex. Initiatives like OpenABM https://www.openabm.org/ give evidence of the desire to more effectively share and utilize agent-based models. However, enabling reproducible simulation studies with agent-based models requires significant efforts. The ODD (Overview, Design concepts, and Details) protocol has been introduced to standardize the published descriptions of individual-based and agent-based models with the objective of facilitating understandable and complete model descriptions and to address the problem of a lack of replicability (Grimm et al. 2010). However, the standard comes, with its own challenges, e.g., redundant description of model parts, and how to describe the model's dynamics in an unambiguous manner, and shortcomings, e.g., a standardized description of the experiments done to validate the models has not been included so far (Grimm et al. 2010, Collins et al. 2015). However, the effort has undoubtedly led to a changed perception of the problem of reproducibility and a more rigorous approach in developing and documenting models. However, the degree of adopting the ODD protocol varies with the application area, e.g., demography, ecology, sociology, or healthcare. Whereas in some areas it is adopted as de facto standard, in other areas, even if an agent-oriented simulation approach is used, authors appear largely unaware about the ODD protocol. Therefore, to really understand the problem of reproducibility in modeling and simulation research one needs to consult the application areas and their constraints.

## 3 REPRODUCIBILITY OF HEALTH CARE SIMULATION BY SALLY BRAILSFORD

There can be little doubt that healthcare is a hugely popular application area for simulation modeling. Literature surveys describe many hundreds of examples (Fone et al. 2003, Katsaliaki and Mustafee 2011). However, one common feature of all these reviews and survey papers is the lack of reported implementation of model findings and recommendations. Indeed, Fone et al state that " . . . we were unable to reach any conclusions on the value of modelling in health care because the evidence of implementation was so scant." (Fone et al. 2003, p.333). All this raises the obvious question: how can we get simulation models used more widely in healthcare? One problem is that discrete-event simulation models are undeniably time-consuming to build. In healthcare applications, models are often very complex and detailed. Data collection is a lengthy (and painful) process which often requires Ethics Committee approval. Moreover, one of the key aspects of getting a model used involves convincing the client that the model has credibility. This can also be a time-consuming process. Questions frequently asked at the start of a new modeling project include "Who else has used this approach?" and "How do I know I can trust the model?" This section focuses on improving the reproducibility of published models as one possible step towards tackling some of these issues. This section is based largely on 25 years experience in the field and recently, since 2011, as Editor-in-Chief of the journal Health Systems. However, as an n = 1 case study, I have analysed a recent paper from the Journal of Simulation (Chen and Wang 2016) which describes a discrete-event simulation model of an Emergency Department (ED).

### 3.1 Current State of the Art

For quality assurance and academic integrity reasons, it is critically important that the peer-review process for published papers is rigorous and reliable. Therefore, in theory, any paper that describes a simulation model should be scrutinized with the same attention to detail as a mathematical proof. Arguably, the only way this can be done is either to examine the original model developed by the authors (and indeed, a small number of journals require this) or to attempt to rebuild the model from scratch. The latter would be a major undertaking. Nevertheless, it is required by the UK National Institute for Health and Care Excellence (NICE) whose role, similar to the US Food & Drugs Administration, is to assess the safety of new medications and (in the case of NICE) determine whether they should be funded by the National Health Service (NHS). As part of its Health Technology Assessment process, NICE uses cost-effectiveness models (often simulations) in order to determine cost-effectiveness over a long time period. NICE requires all such models produced by pharmaceutical companies to be independently audited by academics, and the models rebuilt from scratch. This can take several months, even by experts.

Clearly this is a gold standard approach and it would be impractical to require it for academic research papers. Moreover, in healthcare, reproducing any but the most artificial model published in the literature is tricky. The conditions imposed by Ethics Committees may prohibit the sharing of data, even anonymized, with other researchers. Furthermore, in many countries data may have to be purchased from the organizations that collect it, and is not given away free of charge. For example, in the UK NHS a vast amount of transactional data are collected for financial monitoring or performance evaluation purposes, and while summaries of some these data are in the public domain, these are usually highly aggregated and thus of little use for detailed modeling. More detailed data can be purchased by universities for research purposes (the more detailed, the more expensive), but of course can never be made publicly available.

### 3.2 Means to Improve Reproducibility

While it seems unlikely from a pragmatic standpoint that journals will ever require reviewers to rebuild whole models, there are steps that can be taken to increase model transparency and improve the quality of peer review. Journal editors have some responsibility here. As a reviewer and also as a journal editor myself, at the very least I would expect to see a high-level process flow diagram, for example the patient pathways through a hospital system or the different stages of a disease process. Most published papers in good quality journals include such diagrams, and also state the software used in the model, although there is rarely sufficient detail to allow the reader to reconstruct the model in its entirety. If you are lucky, the paper will contain details of (some of) the fitted probability distributions for the main activities, although obtaining the raw data to check the resulting parameters is usually impossible, for confidentiality reasons.

In the case of (Chen and Wang 2016), there is a simple diagram of patient flow in the ED, and a screen print (although of too low resolution to be readable) of the simulation model, which was developed in the software eM-Plant. There is an Appendix listing all the probability distributions and their parameters, which were fitted partly from administrative data obtained from the hospital, and partly using expert opinion. However, there are three different "Doctor" activities in the list and it is not clear what these are. Moreover two key things are missing, which mean that it would not be possible to reconstruct this model on the basis of the published paper. Firstly, we are not given full information about which resources (and how many) are required for each activity, and secondly and more seriously, we are not told the proportions of patients in each triage category. The latter is critical as without it, we cannot determine how many patients follow each path through the system. On the other hand, the simulation study appears to have been carried out rigorously and in close consultation with hospital stakeholders, and has been validated satisfactorily. This paper is by no means unusual in my experience.

There is a strong argument for a standardized approach to model description and documentation in published papers, which would allow models to be adapted more readily for reuse by other researchers. One further potential solution in healthcare might be the development of a massive synthetic patient dataset which could be used, rather like the MIPLIB library of test problems for integer programming algorithms, to allow researchers to compare their methods and models on a standardized dataset. This could be tailored to any specific demographic group, geographical location or healthcare delivery system. In the era of "big data" and open access publishing, such a concept is surely attainable.

### 3.3 Discussion

There is considerable interest in the use of generic models in healthcare (Günal and Pidd 2011). Rather than take a whole model off the shelf, Gunal and Pidd argue that the most promising approach seems to be developing a collection or library of modular building blocks which can be assembled and reconfigured before being populated with local data. However, the governance issues surrounding data will always mean that many real-world healthcare simulation models will never be placed in the public domain.

In my experience, building the model is the easy part of a healthcare simulation project. Data collection and stakeholder management jointly occupy at least 90% of the effort in any healthcare modeling project. Even if we are able to reuse or adapt an existing model, there is always the problem of the "not invented here" syndrome. For example, simulation models for Emergency Departments (EDs) abound in the literature, yet it is hard to believe that EDs differ so widely from each other that a model developed for Hospital *A* could not be applied, albeit with slight structural modification and obviously with different data, to Hospital *B*. Nevertheless, every ED manager thinks their department is unique. My personal experience has been that the process of stakeholder engagement, i.e. building the relationship between client and modeler and involving clinicians every step of the way, is a necessary and inescapable part of the model development process. This leads to a lot of wheel-reinventing, and often also - sadly - to the model not being used because it took so long to develop that the problem it was originally designed to address has changed. However, surely some of this wheel-reinventing could be avoided if modelers made more effort to ensure their published models were more easily reproducible.

## 4 REPRODUCIBILITY OF NETWORK SIMULATION BY JASON LIU

### 4.1 Current State of the Art

Simulation has been used and continues to be used quite extensively in studying computer networks and communication systems for evaluating new algorithms and protocols, for comparing design alternatives, and as fast prototyping tools. Although there exists a long history in network modeling and simulation, the first successful community-supported simulation framework can be attributed to ns-2, a project supported by DARPA in the mid 90s (Breslau et al. 2000). The result is a quite comprehensive suite of models, including a large collection of network protocols for both wired and wireless networks. They are widely used by the research community, even today. In fact, many simulators have been developed over the years to address different modeling and simulation needs, such as OPNET (http://www.opnet.com/), QualNet (http://scalable-networks.com/), OMNeT++ (https://omnetpp.org/) and ns-3 (http://www.nsnam.org/). Some have also been specially developed to model networks at scale, including SSFNet (Cowie et al. 1999), GTNeTS (Riley 2003), and ROSSNet (Yaun et al. 2003).

The issue about reproducibility has been frequently raised and seriously coped with in the past. Conceptually, reproducibility could be achieved with relative ease, especially when a common network simulator has been adopted by the community. In this case, researchers could publish the source code of the new module (for example, a new network protocol implementation in ns-2). In their papers, the authors could easily provide a reference to a public repository along with the necessary details (including the seed for the random numbers) for configuring the simulation experiments used to obtain the results appeared in the papers.

Unfortunately, the reality is far from satisfactory. Pawlikowski et al. (2002) claim that *"An opinion is spreading that one cannot rely on the majority of the published results on performance evaluation studies of telecommunication networks based on stochastic simulation, since they lack credibility."* The claim is based on a survey study of over 2200 papers in prominent network conferences, which indicates that a majority of the published results (approx. 77%) did not state the type of simulation used in the study, and only a small fraction of the papers (approx. 30%) specified the type of pseudo-random number generators used in the experiment (Pawlikowski 2003). Another study in 2005 (Kurkowski et al. 2005) zoomed in a top mobile ad hoc network (MANET) conference and found that, although the MANET research community relies heavily on the use of simulation (approx. 75% of the 151 papers appeared in the last five years included simulation results), the credibility of the simulation (mainly in terms of repeatability) has not improved at all. A more recent study by Sarkar and Gutiérrez (2014) reached the same conclusion.

There are additional problems associated with reproducibility. One obvious problem is that network simulators typically contain many hidden parameters (default configurations, unspecified protocol parameters). These hidden parameters may change between simulator versions, usually subject to the whims of the developers without proper documentation. As such, the simulator behavior may change unexpectedly; the results obtained from one version of the simulator may not be reproducible with other versions.

Another problem is associated with the "standard" simulation scenarios. In several communities, researchers conduct experiments by choosing from only a small set of well-established scenarios, which may not be designed originally to suit all situations. Using these standard scenarios can promote reproducibility, but may occasionally generate biased results, sometimes giving very wrong conclusions. For example, in early studies of TCP congestion control, researchers typically use "the dumbbell model" in their experiments, where all servers are put on one side of the dumbbell-shaped network and all clients on the other side. The servers and clients engage in pair-wise large data transfers in a synchronous fashion: all start from time zero and thus all experience similar end-to-end path delay. Unfortunately, all these experiments suffer from what is called "the phase effect", which *"is not relevant to the modern Internet", but "an artifact of unrealistic simulation scenarios."* (Floyd and Kohler 2003). Another example is the random waypoint user mobility model commonly in wireless simulations. Yoon et al. (2003) discovered that this model, if not parameterized correctly, may not generate steady-state results. It took the community many years, long after this discovery, to correct itself in the simulation experiments.

In today's network research, simulation still plays an indispensable role, however only as a part of the whole. A typical network study considers experiments using simulation, emulation, physical testbeds, or a combination of them. Li and Liu (2014) conducted a survey of five-year SIGCOMM papers (2007-2013) and found that, among papers involved with experimental studies, approx. 46% of the papers contain only experiments using physical testbeds, while approx. 20% papers contain only simulation studies. Most interestingly, nearly one third of the papers involve both simulation and physical testbeds. There is a trend to integrate simulation with the physical network testbeds, i.e., human, software, or hardware-in-the-loop models, e.g., Gu and Fujimoto (2007), Ahrenholz et al. (2008), Liu et al. (2009), Nicol et al. (2011), Tazaki et al. (2013), Erazo and Liu (2013). Obviously it is more difficult to achieve reproducibility from simulation that interoperates with the physical counterparts with non-repeatable behaviors. For this, we ought to first identify important characteristics of realistic and reproducible experiments. Handigol et al. (2012) suggest that they should include functional realism (the system must provide the same functionality), timing realism (exhibit similar timing behavior), traffic realism (experience similar network traffic), and easy replication (set up and run without difficulty).

### 4.2 Means to Improve Reproducibility

There is no known solution that can satisfy all characteristics. But partial solutions do exist. In particular, there are two technological trends that can make reproducibility easier. The first trend is that virtual machine solutions have become widely available. Simulation experiments, including both software code and data, can be packaged as "virtual appliances", with which one can run the experiments by instantiating the same execution environment on a virtual machine. This would avoid messy software/system dependency issues (software tool chain). A case in point is Handigol et al. (2012), where one can map network experiments to interconnected virtual machines (containers). One can create "runnable papers", where published results can be reproduced in a teaching environment with relative ease.

Another technological trend is that meta-clouds and cyber-infrastructure testbeds, such as CloudLab (https://www.cloudlab.us/), Chameleon (https://www.chameleoncloud.org/), and GENI (Berman et al. 2014), have become more accessible. They are shared resources and reconfigurable. They allow researchers to automate experiments, including resource allocation, model instantiation, execution, and data collection. An example is Apt (Ricci et al. 2015), allowing experiments to re-create the same execution environment on EmuLab (White et al. 2002) and GENI.

## 5 REPRODUCIBILITY OF SIMULATION IN LOGISTICS BY MARKUS RABE

In the following, the situation is analyzed for the logistics field with a view also on production enterprises, as they face a significant bunch of logistics problems. Within the company, the supply of parts and materials, the handling of finished parts and the disposal of waste are often characterized as production logistics. In the broader view, supply chain logistics covers the supply and distribution on long haul, regularly comprising suppliers or customers that are another legal entity or even multiple tiers of delivery and distribution. Transport logistics can be considered as part of the supply chain. Nevertheless, because of the detailed models required for transport simulation, these investigations are mostly conducted on a narrower cut of the full supply chain. Similar problems occur also in other sectors as e.g., heavy construction, civil engineering etc., and are, therefore, not excluded from this investigation. Most of these application fields can be seen as discrete systems with activities that do not follow general cycle times and are, therefore, modeled by discrete event simulation (DES). For special details, we find continuous models (energy models, chemical processes as part of the production, etc.) and for applications that show no major time dynamics Monte Carlo approaches.

### 5.1 Specific Issues in Logistics Experiments

Simulation models in the logistics field show characteristics that are not recognized in the same strength in other simulation application areas. These concern the type of the model, the source of the input data, and the closeness of detailed result data to the enterprises business. Logistics models, if detailed enough for simulation, are rarely built on mathematical relationships. In contrast, the simulation model is for the major part a description of local discrete decision rules that cover the control decisions and also business decisions. While the physical elements of the material flow are, with few exceptions, standard and not competitive, the way to use them for a most efficient logistics flow are considered precious know-how that may not be made public. On the other hand, there is no possibility to reproduce the results without exactly this knowledge on a very detailed level. Logistics models could, as all simulation models, operate with stochastic information based on theoretical distributions. However, it seems that users often don't trust such distributions and prefer to use real data. Just as examples, a real bill of material is used instead of a distribution that specifies how frequently a specific part is assembled, or real order data instead of a count per part and year combined with a distribution over the year. One of the reasons might be that there is significant correlation among parameters (e.g., specific parts are typically ordered in combination). Another potential reason could be the challenge to find a fitting distribution, as real data often show bi-mode or even multi-mode shapes that are found to be difficult to model as a mathematical distribution. As, consequently, real data are used in the model, it becomes obvious that most of these data will be strictly classified. Nevertheless, they can be used for a published study in some cases, but the data will then be given incomplete by (a) selecting data that are less sensitive; (b) changing real information like product names or customer names by artificial dummy information; (c) and excluding absolute figures from results, e.g., by omitting numbers on the ordinate of graphs or by giving percentages in result tables. On the other hand, there are parts of the simulation study that seem not to be influenced by nondisclosure or model structure problems, at least not in a major way. These concern parts of the credibility of the model, which heavily relies on the rigorous statistical evidence and a well-structured and systematic verification and validation (V&V) procedure. V&V is an essential part of a simulation study (Balci 1998), and there is sufficient research on the detailed V&V procedures along simulation studies in the logistics field (Rabe et al. 2009). For the statistical evidence, at least basic information like the number of replications conducted as well as the length of the transient phase (for non-terminating models) should be specified, and also argued why these parameters have been chosen in the given way. There are little arguments why this should not be published.

### 5.2 Current Situation

In order to get a clearer image of the situation, recent publications have been analyzed. With the aim to keep the task manageable but at the same time to have an acceptable level of representativeness, the proceedings of the last two issues of the ASIM Conference on Simulation in Production and Logistics have been analyzed (Dangelmaier et al. 2013, Rabe and Clausen 2015). This biannual conference is the only large event that is fully dedicated on this topic. Furthermore, the papers that have appeared either in the “Manufacturing Applications” or in the “Logistics, SCM, and Transportation” track of two Winter Simulation Conferences (Tolk et al. 2014, Yilmaz et al. 2015) were included in the analysis. In the logistics field, there are far more conference reports than journal papers. However, as journal papers might have a different attitude, the special issue in the Journal of Simulation on “Simulation in Production and Logistics” (Wenzel et al. 2016) has been included. From all these sources, papers that did not address a direct research issue, e.g. papers on field studies, kinematics problems, or guideline works, have been excluded from the analysis, leading to a total sample size of 193. In terms of application areas, 85 papers mainly address logistics, 84 production, 13 construction-related topics, and 11 other or general topics. In terms of the level of detail, 36 papers comprise supply chains, 49 address intralogistics and 24 describe transport-specific applications; further 9 focus on energy aspects and 75 papers address other topics, with a nice bunch of these on scheduling tasks for the production-oriented papers. Due to the approach of this research, most papers (139) report on DES, while a minority of 5 papers uses timesliced models, 10 papers give Monte Carlo applications, 8 continuous models, 7 hybrid approaches, and 24 papers use more than one or are not clear with respect to the applied technology. The hypothesis that many papers show real cases is proven by a number of 105 papers that explicitly address a real case (sometimes not naming the exact company). Even from the remaining 88 papers, many state that they use an adapted or simplified case based on a real system. As expected, models are not given in full detail. A minority of 10 papers disclose quite detailed information on the model (mostly those on artificial cases), while 41 give merely no information, 34 at least some indications, 57 an overview, and 46 explain parts of the model. Further, 5 give no details, but give references to other reports announcing that information could be found there. The situation for the input data is similar: 69 papers give merely no information, 67 just indications about the kind of input data used, 3 provide a reference to other sources, 49 partial information, and only 5 give apparently complete input (typically, for small toy samples). For the output data, the situation is much better, not fully surprising, as these are the results that should at least demonstrate the benefits of the investigation. Still, however, 40 papers do not give any substantial output data, 21 disclose only qualitative results, 130 give at least some quantitative data, and 2 provide broad and detailed result data. Not surprising but also not understandable, very few papers give clear indication about the statistical procedures. The vast majority of 129 papers does not discuss any issues coming from stochastic effects at all; 55 show at least some information (e.g., specify confidence intervals, number of replications, ) and only 9 papers give adequate information. Even worse, two thirds of the papers (129) do in no way discuss any validation and further, 14 mention validation with absolutely insufficient information. Only 9 papers (5%) give at least some understandable argumentation on how credibility of the model and its results have been achieved, and no single paper gives really convincing argumentation.

### 5.3 Discussion

Generally, it becomes obvious that publishing detailed models and data is mostly influenced by non-disclosure considerations. In cases where this does not apply, the detailed model and the applied input data are far too large to reproduce them in a paper, but they could be made available on side channels like Web sites. For general statistical parameters as well as for the exact V&V procedure there seems to be no barrier to publish it. The provision of information about the model and its results on a level of detail that might allow for reproducing the experiments is scarce. For the cases built on real-world data this seems to root in the fact that the logic of simulation models comprises a very detailed knowledge about the company's control strategy, which is regarded as precious and competitive. Also, input data (and often even output data) mirror information that would be very interesting to potential competitors and is, thus, not disclosed. This also leads to the effect that results are sometimes not given as absolute figures but as percentages, or even without quantification of the ordinate's data. Simulation studies that aim to solve specific problems for smaller sections of the process, however, might have a more principle approach and sometimes work on artificial or at least anonymized data. Generally said, the provision of simulation results is frequently not satisfying. Reasons are rarely given, but might root in the application of real-world model sections, just because they had been available to the investigator, and are consequently not open for publication. In other cases, models tend to be complex; specifying the full model would actually mean to publish the model itself. There would be an easy option to do this, obviously not within the specific papers, but through a public repository. Therefore, it might be worth it to encourage authors to publish input data and model, as long as they are not classified.

## 6 REPRODUCIBILITY AND MILITARY SIMULATIONS BY ANDREAS TOLK

### 6.1 Background on Military Simulation and the Approach on Reproducibility

Simulation and simulation–based experiments have a long standing tradition within the defense domain, as discussed with many examples in Loper and Turnitsa (2012). Although the predominantly perceived application domain is training and support of exercises, normally referred to as computer assisted exercises (CAX) as described among others in (Cayirci 2009), simulation–based experiments are conducted in support of analysis, procurement, support of operations, and several other domains. As such, the defense domain was among the leaders in supporting verification and validation efforts early on, as described in (Balci 1994). The *"The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology (JDMS)",* published quarterly by SAGE (eISSN: 1557380X, ISSN: 15485129) since January 2004 was created as a refereed archival journal devoted to advancing the practice, science, and art of modeling and simulation as it relates to the military and defense. The primary focus of the journal is to document, in a rigorous manner, technical lessons derived from practical experience. The journal will also publish work related to the advancement of defense systems modeling and simulation technology, methodology, and theory. It serves therefore as the main body of reference for the evaluation presented in this section. The insights and observations in this section are based on more than a decade of being a track chair, peer reviewer for various journals - including the JDMS - and some additional analysis of relevant publications, in particular in journals, as well as discussions with colleagues in similar positions.

### 6.2 Current State of the Art

Military simulation has three important constraints that are often cited as the reason for not supporting reproducibility of simulation-based experimentation results, as they are recommended in the introduction to this paper. All three issues are not limited to the military and security domain, but can be observed in other domains as well.

- *Classification and Security:* A lot of military information used within simulation-based experiments is classified. If information on the performance of a weapon system is secret, these data cannot be published. If a scenario is classified as not releasable to foreign nationals, it cannot be used for a publication in an international journal. While there are ways to obfuscate such data in collaborations, such as described in (McGowan and Raney 2008), no comparable ideas can be applied to support reproducibility without providing access to the classified information.
- *Intellectual Property:* The simulation systems used within the military and defense community are often the product of yearlong developments by industry. Millions of dollars went into the programming and testing of simulation algorithms that now can be used to conduct simulation–based experiments. However, it is in the professional and commercial interest of these organizations to protect their sophisticated algorithms, methods, and heuristics. This interest often beats the scientific wish to support reproducible results that require transparency.
- *Volume of Source Code and Data:* A related issue is the sheer volume of source code and data utilized for large scale military simulation applications. Many simulation system on the level of a military brigade still model each individual weapon systems and soldier, which results in more than 300 weapon systems and more than 15,000 infantry soldiers, supported by artillery, combat engineers, air defense, and many more troops. The US Army OneSAF baseline has approximately 2 million lines of code (Parsons, Surdu, and Franceschini 2005). Publishing all these code and the data is not really helpful in support of reproducibility.

As a result, the military domain is looking for other ways to increase the trust in the simulation results and increase the credibility of its simulation applications. In my professional activities of the recent years I have never had the privilege to review a submission with enough information to support reproducibility as envisioned in this paper.

### 6.3 Means to Improve Reproducibility

The NATO Code of Best Practice for Command and Control Assessment (COBP) (Alberts et al. 2002) was produced to facilitate high quality assessment in the area of C2 (Command and Control). The COBP offers broad guidance on the assessment of C2 for the purposes of supporting a wide variety of decision makers and C2 researchers. It addresses among other means of operations research also the use of simulation. One clear guidance is that no study shall be published without peer review to increase the overall credibility of the results. While reproducibility is a higher goal than credibility, it is a step into the right direction. In a similar way, the Guide for Understanding and Implementing Defense Experimentation developed by The Technical Cooperation Program (Labbé et al. 2006) emphasizes the need to apply scientific rigor and methods, including reproducible results.

Among the most referenced papers published in recent years in JDMS, the following three directly address means to improve reproducibility. Yilmaz proposes a need for contextualized introspective models to enforce transparency into simulation systems as well as their result (Yilmaz 2004). Harmon and Youngblood propose a model for simulation validation process maturity that includes transparency recommendation for design, implementation, and results (Harmon and Youngblood 2005). Kim et al. explain how the use of well understood formalisms, in their example DEVS, increases transparency and interoperation (Kim et al. 2010).

Another alternative worthwhile of discussion is the use of surrogate models that are neither classified nor expose the intellectual property protected elements of the original simulation system. If these surrogates are behavioral equivalent with the original model, they can be used to ensure the necessary transparency and reproducibility. The proof of structural and behavioral equivalence of two simulation system, however, is not trivial, as has been documented, among others, in (Yücesan and Schruben 1992, Pooley 2007, Bergen-Hill and Page 2010).

### 6.4 Discussion

Although the issues captured in section 6.2 often speak against full–scale reproducibility, not much speaks against publishing at least key source code and data that are in the focus of a journal article. To some degree, this is already applied. An example is given by Özhan et al. who use UML diagrams to describe in detail how the model field artillery in their experiment (Özhan et al. 2008). Classification, intellectual property, and volume of source code will continue to be the main challenges in the quest for reproducible simulation–based experiment results in the military and defense domain, but the need for transparency of algorithms and data to increase the credibility of these results has been recognized. When this insight will lead to recommendations or even requirements to publish source code and data with journal contributions in this domain remains to be seen. Alternative solutions, such as the use of equivalent surrogate models, is known to be challenging in themselves.

## 7 REPRODUCIBILITY BEYOND INDIVIDUAL APPLICATION AREAS

For the above four modeling and simulation areas, different approaches were pursued to help illuminate the current practice in ensuring reproducibility. In the area of network simulation, a series of surveys about reproducibility of published research exists which has been used as a basis. For the area of logistics, a short (informal, yet first) survey on reproducible research taking into account 193 recent publications has been conducted for this purpose. In health care and military, selective individual papers have been analyzed.

### 7.1 Reproducibility - Where are We ?

In all areas, the current state is non-satisfactory. Most of the work published cannot be reproduced. This applies to the area of network simulation, despite standard simulation tools and model libraries, as well as to modeling and simulation in health care, logistics, and military. However, what is missing for a reproducible modeling and simulation varies with the different application areas. Although in network simulation, incompletely documented experimental settings might threaten replicating research results, most problems might be taken care of by a more stringent documentation discipline in combination with exploiting recent technical developments. In contrast, the problems in the other three application areas appear more deeply rooted. In those areas, models and other artifacts of modeling and simulation research, like software or data, often constitute classified information and, as such, cannot be easily made accessible. Other arguments for the non-transparency of modeling and simulation research in these areas are the complexity of models and the problems induced due to non standardized modeling and simulation approaches. However, as revealed in the survey on logistics (but applicable also in the other areas), reasons for presenting non-reproducible research are rarely given and thus, left as guesswork in concrete cases. Interestingly in health care and military, the governmental institutions (as major players) appear well aware about the thread that non-reproducible modeling and simulation presents - even to the point to rendering modeling and simulation as a useless method. Therefore, specific measures shall support reproducible modeling and simulation results, which involve major efforts including a complete documentation of all validation and verification steps or rebuilding entire models from scratch. However, these measures are applied internally, and stand in stark contrast to the majority of research published.

### 7.2 Improving Reproducibility: Different Areas - the Same Means ?

In all four application areas, the role of easily accessible simulation tools that play the role of a de-facto standard in an area have been emphasized as one possibility to enhance reproducibility, and the lack of this as one major impediment of reproducible research. E.g. in network simulation, tools like ns2 with large libraries facilitate the accessibility of models as the prime artifacts of modeling and simulation and the reproduction of research results. However, even with standardized tools, their configuration, including possibly hidden parameters, influences simulation results and needs to be considered. The role of tools in reproducible research is also apparent in the area of agent-based research. Major agent-based modeling and simulation frameworks, e.g., Netlogo (Wilensky 1999) or Repast (North et al. 2006), encourage authors to publish their models online to support a reuse of models and a reproduction of research results. However, the information focuses on the model and does not expand to background information like input data, calibration, and validation processes the model was subjected to. Thus in addition, standardized guidelines for documenting simulation experiments are required.

Reproducible research shares many challenges with reusing models in general, e.g., developing model libraries that can easily be parameterized and specifying the limitations of a model's use in an unambiguous and compact manner (see section on network). In addition, the lack of a formal semantics of modeling approaches and of a compact description hampers the reproduction of research results. Thus, more efforts are required referring to formal domain specific modeling languages.

Other concerns, e.g., models being too complex or too many data being involved, can be addressed by technical means, like online supplementary material. Also runnable papers help in repeating experiments (as those occur in the same virtual environment) and as such increasing the confidence in the documented research findings. However, more information is needed to "reproduce" experiments, a process with which we typically associate scientists interpreting and evaluating the research and then reproducing the research results by same or similar methods.

Obviously, additional means also are required to overcome the problem of classified information. Synthetic data sets and exploiting adequate privacy and security methods have been identified as helpful across application fields. Pursuing the above strategies in combination with clearly differentiating between research results according to their degree of reproducibility (and thus according to their scientific rigor) could start changing the current practice to put more efforts into reproducible modeling and simulation research.

### 7.3 Journals, Conferences, and Workshops: is the Ball in their Court ?

Journals, conferences, and workshops are the outlets for research of scientific communities and, as such, reflect the culture of those communities. In natural sciences, it has long been a sine qua none to publish within the methods section any information required to replicate the research results. In addition, supplementary material was put to use to fill in the remaining gaps of information about the experiments done. Without this information papers could not be published and thus authors could not receive credit for the research done. Still this has not prevented non-reproducible research, as the replication study by Baggerly and Coombes illustrated (Baggerly and Coombes 2009).

Other disciplines have problems perceiving themselves as experimental sciences and as a result knowledge and procedures how to execute and document experiments leave much to wish for, as e.g. pointed out for the field of computer science by Walter Tichy (Tichy 1998), but likewise applicable to the other three application fields described above. Meanwhile, the computer science community has become aware that experiments are essential in computer science research and experiments need to be more carefully executed and documented (Mytkowicz et al. 2009). Non-reproducible research is considered a problem: "Repeatability and reproducibility are cornerstones of the scientific process, necessary for avoiding dissemination of flawed results." (Collberg and Proebsting 2016). In general, the calls for making main artifacts of computer science, e.g., software and data, available and accessible are getting louder.

An increasing number of conferences encourages authors of accepted papers whose research includes source code, mechanized proofs, data collection, test suites, models, or any other digital artifacts to submit these to an optional artifact evaluation. Therefore, conferences follow a rather similar pattern, see e.g. International Conference on Computer Aided Verification – CAV (http://i-cav.org/2015/evaluation/) or ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages – POPL (http://popl15-aec.cs.umass.edu/home/). Artifacts are reviewed whether they are 1) Easy to reuse. How easy is it to reuse the provided artifact? 2) Consistent. Does the artifact help to reproduce the results from the paper? 3) Complete. What is the percentage of the results that can be reproduced? 4) Well documented. Does the artifact describe and demonstrate how to apply the presented method to a new input? . The papers that successfully completed the artifact evaluation are marked by a corresponding badge. Other strategies to improve the reproducibility of research published at conferences are explored as well. For example, in 2016 the "International Conference for High Performance Computing, Networking, Storage and Analysis (SC)" (http://sc16.supercomputing.org/) exploits the student cluster competition to test the replicability of results presented in papers at previous SC conferences instead of running more traditional benchmark applications. Also two journals of the ACM, i.e., ACM Transactions on Mathematical Software (TOMS) (http://toms.acm.org) and ACM Transactions on Modeling and Computer Simulation (TOMACS) (http://tomacs.acm.org) offer a replication of computational results (RCR). TOMACS is the first simulation journal that pushes the reproducibility of modeling and simulation research. The first paper carrying the TOMACS RCR and Artifact Evaluation badge (Fig. 1) has just been published: all findings including figures but also the methods developed and implemented in (Feng et al. 2016) have been checked and findings replicated by an independent reviewer (Lück 2016). So far no simulation conference has established procedures to increase the reproducibility and replicability of research.


Figure 1: Badge of TOMACS

## REFERENCES

Ahrenholz, J., C. Danilov, T. R. Henderson, and J. H. Kim. 2008. "CORE: A real-time network emulator". In *MILCOM*, 1–7.

Alberts, D. S. et al. 2002. *NATO Code of Best Practice for Command and Control Assessment*. Washington, DC: CCRP Press.

Baggerly, K., and K. Coombes. 2009. "Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High-Throughput Biology". *Ann App Statist* 3 (4): 1309–34.

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. New York: John Wiley.

Bergen-Hill, T. A., and E. H. Page. 2010. "Out–of–order execution and structural equivalence of simulation models". In *Proc. of the Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 466–478: IEEE/ACM.

Berman, M., J. S. Chase, L. Landweber, A. Nakao, M. Ott, D. Raychaudhuri, R. Ricci, and I. Seskar. 2014. "GENI: A Federated Testbed for Innovative Network Experiments". *Comput. Netw.* 61:5–23.

Breslau, L., D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu. 2000. "Advances in network simulation". *IEEE Computer* 33 (5): 59–67.

Cayirci, E. 2009. "Multi-resolution Federations in Support of Operational and Higher Level combined/joint Computer Assisted Exercises". In *Proc. of the Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 1787–1797: IEEE/ACM.

Chen, T., and C. Wang. 2016, Feb. “Multi-objective simulation optimization”. *J Simulation* 10 (1): 50–68.

Collberg, C. S., and T. A. Proebsting. 2016. “Repeatability in computer systems research”. *Commun. ACM* 59 (3): 62–69.

Collins, A., M. D. Petty, D. Vernon-Bido, and S. J. Sherfey. 2015. “A Call to Arms: Standards for Agent-Based Modeling and Simulation.”. *J. Artificial Societies and Social Simulation* 18 (3).

Cowie, J., D. Nicol, and A. Ogielski. 1999, January. “Modeling the global Internet”. *Computing in Science and Engineering* 1 (1): 42–50.

Dalle, O. 2012. “On Reproducibility and Traceability of Simulations”. In *Proc. of the Winter Simulation Conference*, 244:1–244:12: IEEE/ACM.

Dangelmaier, W., C. Laroque, and A. Klaas. (Eds.) 2013. *Simulation in Production and Logistics 2013*. Paderborn: HNI-Verlagsschriftenreihe.

Erazo, M. A., and J. Liu. 2013. “Leveraging symbiotic relationship between simulation and emulation for scalable network experimentation”. In *Proc. of the 2013 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, 79–90. ACM.

Ewald, R., and A. M. Uhrmacher. 2014. “SESSL: A Domain-Specific Language for Simulation Experiments”. *ACM Transactions on Modeling and Computer Simulation* 24 (2).

Feng, C., J. Hillston, and V. Galpin. 2016, March. “Automatic Moment-Closure Approximation of Spatially Distributed Collective Adaptive Systems”. *ACM Transactions on Modeling and Computer Simulation* 26 (4): 26:1–26:22.

Floyd, S., and E. Kohler. 2003. “Internet Research Needs Better Models”. *CCR* 33 (1): 29–34.

Fone, D., S. Hollinghurst, M. Temple, A. Round, N. Lester, A. Weightman, K. Roberts, E. Coyle, G. Bevan, and S. Palmer. 2003. “Systematic review of the use and value of computer simulation modelling in population health and health care delivery”. *Journal of Public Health* 25(4):325 – 335. Publisher: Faculty of Public Health.

Grimm, V., U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. 2010. “The ODD protocol: A review and first update”. *Ecological Modelling* 221 (23): 2760–2768.

Gu, Y., and R. Fujimoto. 2007. “Applying parallel and distributed simulation to remote network emulation”. In *Proc. of the Winter Simulation Conference*: IEEE/ACM.

Günal, M. M., and M. Pidd. 2011. “DGHPSIM:: Generic Simulation of Hospital Performance”. *ACM Transactions on Modeling and Computer Simulation* 21 (4): 23:1–23:22.

Handigol, N., B. Heller, V. Jeyakumar, B. Lantz, and N. McKeown. 2012. “Reproducible Network Experiments Using Container-based Emulation”. In *Proc. of the 8th International Conference on Emerging Networking Experiments and Technologies (CoNEXT’12)*, 253–264.

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

Joppa, L. N., G. McInerny, R. Harper, L. Salido, K. Takeda, K. O’Hara, D. Gavaghan, and S. Emmott. 2013, May. “Troubling Trends in Scientific Software Use”. *Science* 340 (6134): 814–815.

Katsaliaki, K., and N. Mustafee. 2011. “Applications of simulation within the healthcare context”. *Journal of the Operational Research Society (JORS)* 62 (8): 1431 – 1451.

Kim, T. G., C. H. Sung, S.-Y. Hong, J. H. Hong, C. B. Choi, J. H. Kim, K. M. Seo, and J. W. Bae. 2010. “DEVSim++ toolset for defense modeling and simulation and interoperation”. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 8 (3): 129–142.

Köhn, D., and N. Novère. 2008. “SED-ML - an XML formate for the implementation of the MIASE guidelienes”. In *Proc. of the 6th International Conference on Computational Methods in Systems Biology*, edited by M. Heiner and A. Uhrmacher, 176–190: Springer.

Kurkowski, S., T. Camp, and M. Colagrosso. 2005. “MANET Simulation Studies: The Incredibles”. *ACM SIGMOBILE Mobile Computing and Communications Review* 9 (4): 50–61.

Labbé, P. et al. 2006. *Guide for Understanding and Implementing Defense Experimentation GUIDEx*. Norfolk, VA: The Technical Cooperation Program, NATO ACT.

Li, T., and J. Liu. 2014, November. "Cluster-Based Spatiotemporal Background Traffic Generation for Network Simulation". *ACM Transactions on Modeling and Computer Simulation* 25 (1): 4:1–4:25.

Liu, J., Y. Li, N. V. Vorst, S. Mann, and K. Hellman. 2009, March. "A real-time network simulation infrastructure based on OpenVPN". *Journal of Systems and Software* 82 (3): 473–485.

Loper, M. L., and C. D. Turnitsa. 2012. *History of Combat Modeling and Distributed Simulation*, 329–355. Hoboken, NJ: John Wiley & Sons, Inc.

Lück, A. 2016, March. "Replicated Computational Results (RCR) Report for &Ldquo;Automatic Moment-Closure Approximation of Spatially Distributed Collective Adaptive Systems&Rdquo;". *ACM Transactions on Modeling and Computer Simulation* 26 (4): 27:1–27:3.

McGowan, G., and C. Raney. 2008. "Integrating Multi-Level Security into the Joint Warfighter Training Environment". In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando*.

Merali, Z. 2010, October. "Computational science: ...Error". *Nature* (467): 775–777.

Mytkowicz, T., A. Diwan, M. Hauswirth, and P. F. Sweeney. 2009, March. "Producing Wrong Data Without Doing Anything Obviously Wrong!". *SIGPLAN Not.* 44 (3): 265–276.

Nicol, D., D. Jin, and Y. Zheng. 2011. "S3F: the scalable simulation framework revisited". In *Proc. of the Winter Simulation Conference*, 3288–3299: IEEE/ACM.

North, M. J., N. T. Collier, and J. R. Vos. 2006. "Experiences creating three implementations of the repast agent modeling toolkit". *ACM Trans. Model. Comput. Simul.* 16 (1): 1–25.

Özhan, G., H. Oğuztüzün, and P. Evrensel. 2008. "Modeling of field artillery tasks with live sequence charts". *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 5 (4): 219–252.

Parsons, D., J. Surdu, and D. Franceschini. 2005. "OneSAF: A Case Study in Agile Methods and Extreme Programming". In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando*.

Pawlikowski, K. 2003. "Do not trust all simulation studies of telecommunications networks". In *International Conference on Information Networking (ICOIN*, 899–908.

Pawlikowski, K., H. Jeong, and J. Lee. 2002. "On credibility of simulation studies of telecommunication networks". *IEEE Communications Magazine* 40 (1): 132–139.

Pooley, R. 2007. "Behavioural equivalence in simulation modelling". *Simulation Modelling Practice and Theory* 15 (1): 1–20.

Rabe, M., and U. Clausen. (Eds.) 2015. *ASIM Conference Proceedings - Simulation in Production and Logistics 2015*. Stuttgart: Fraunhofer Verlag.

Rabe, M., S. Spieckermann, and S. Wenzel. 2009. "Verification and Validation Activities within a New Procedure Model for V&V in Production and Logistics Simulation". In *Proc. of the Winter Simulation Conference*, edited by M. D. Rossetti, R. Hill, B. Johansson, A. Dunkin, and R. Ingalls, 2509–2519. Piscataway: IEEE.

Rahmandad, H., and J. D. Sterman. 2012. "Reporting guidelines for simulation-based research in social sciences". *System Dynamics Review* 28 (4): 396–411.

Ricci, R., G. Wong, L. Stoller, K. Webb, J. Duerig, K. Downie, and M. Hibler. 2015. "Apt: A Platform for Repeatable Research in Computer Science". *ACM SIGOPS Operating Systems Review* 49 (1).

Riley, G. 2003. "The Georgia Tech network simulator". In *Proc. of the ACM SIGCOMM Workshop on Models, Methods and Tools for Reproducible Network Research (MoMeTools'03)*, 5–12.

Sarkar, N. I., and J. A. Gutiérrez. 2014. "Revisiting the issue of the credibility of simulation studies in telecommunication networks: highlighting the results of a comprehensive survey of IEEE publications". *IEEE Communications Magazine* 52 (5): 218–224.

Schützel, J., D. Peng, A. M. Uhrmacher, and L. F. Perrone. 2014. "Perspectives on languages for specifying simulation experiments". In *Proc. of the Winter Simulation Conference*, edited by S. J. Buckley and J. A. Miller, 2836–2847: IEEE/ACM.

Tazaki, H., F. Uarbani, E. Mancini, M. Lacage, D. Camara, T. Turletti, and W. Dabbous. 2013. "Direct Code Execution: Revisiting Library OS Architecture for Reproducible Network Experiments". In *Proc. of the 9th ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT'13)*, 217–228.

Tichy, W. F. 1998. "Should Computer Scientists Experiment More?". *IEEE Computer* 31 (5): 32–40.

Tolk, A., S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller. (Eds.) 2014. *Proc. of the Winter Simulation Conference*. Piscataway: IEEE.

Waltemath, D., R. Adams, F. Bergmann, M. Hucka, F. Kolpakov, A. Miller, I. Moraru, D. Nickerson, S. Sahle, J. Snoep, and N. Le Novere. 2011. "Reproducible computational biology experiments with SED-ML - The Simulation Experiment Description Markup Language". *BMC Systems Biology* 5:198.

Wenzel, S., M. Rabe, and C. Laroque. (Eds.) 2016. *Special issue on simulation in production and logistics*, Volume 10 (2) of *Journal of Simlation (JoS)*.

White, B., J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar. 2002. "An Integrated Experimental Environment for Distributed Systems and Networks". In *Proc. of the 5th Symposium on Operating Systems Design and Implementation (OSDI'02)*, 255–270.

Wilensky, U. 1999. *NetLogo http://ccl.northwestern.edu/netlogo/*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Winsberg, E. 2010. *Science in the Age of Computer Simulation*. University of Chicago Press.

Yaun, G. R., D. Bauer, H. L. Bhutada, C. D. Carothers, M. Yuksel, and S. Kalyanaraman. 2003. "Large-scale network simulation techniques: examples of TCP and OSPF models". *ACM SIGCOMM Computer Communication Review* 33 (3): 27–41.

Yilmaz, L. 2004. "On the need for contextualized introspective models to improve reuse and composability of defense simulations". *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 1 (3): 141–151.

Yilmaz, L., W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti. (Eds.) 2015. *Proc. of the Winter Simulation Conference*. Piscataway: IEEE.

Yoon, J., M. Liu, and B. Noble. 2003, March. "Random waypoint considered harmful". In *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications (INFOCOM'03)*, Volume 2, 1312–1321 vol.2.

Yücesan, E., and L. Schruben. 1992. "Structural and behavioral equivalence of simulation models". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 2 (1): 82–103.

## AUTHOR BIOGRAPHIES

**Adelinde M. Uhrmacher** is Professor and Head of the Modeling and Simulation group within the School of Computer Science and Electrical Engineering at University of Rostock, Germany (adelinde.uhrmacher@uni-rostock.de).

**Sally Brailsford** is Professor of Management Science within Southampton Business School at the University of Southampton, UK (s.c.brailsford@soton.ac.uk)

**Jason Liu** is Associate Professor at the School of Computing and Information Sciences at Florida International University, USA (liux@cis.fiu.edu).

**Markus Rabe** is Professor and Head of the IT in Production und Logistics group within the School of Mechanical Engineering at the Technical University of Dortmund, Germany (markus.rabe@tu-dortmund.de).

**Andreas Tolk** is Computer Science Principal at MITRE Corporation, McLean, USA (andreas.tolk@gmail.com).