



Advancing reproducibility and replicability in simulation: Challenges and opportunities

Yilin Huang¹  and Deniz Cetinkaya²

Abstract

Computer simulation has become increasingly complex and widely applied across different domains. However, the reproducibility and replicability (R&R) of simulation models remain limited. Despite recent improvements, independent reproduction or replication of simulation experiments is rare in the literature. This paper provides an overview of the state of research on R&R in simulation, highlights recent developments, and discusses key concepts and future perspectives. It first examines how R&R has been viewed, approached, and evaluated and then outlines typical challenges and defining characteristics of R&R. Emerging opportunities are also discussed in light of community-driven practices, artificial intelligence, and quantum computing. Given the significant role of simulation in modern science, this paper argues that R&R studies of simulation are valuable research outputs and should be regarded as an integral and equally important part of scientific progress. R&R should be explicitly addressed and embedded into modelling and simulation practices, and supported by stronger community efforts. Researchers engaging in these efforts face substantial challenges, including those related to recognition and rewards, methodology, and scalability, many of which are under-researched.

Keywords

reproducibility, replicability, open simulation, methodology, reusability

1. Introduction

Many computer simulation models have been developed for science. It is widely recognised that simulation-based experiments, like other types of experiments, shall uphold reproducibility, a fundamental tenet of science.^{1–3} Although, in principle, research should be reproducible, there is a growing concern among scientists that only a limited fraction of published research can be reproduced, a situation sometimes called the “reproducibility crisis.”^{2–5} A nature survey⁶ revealed that more than 70% of responding researchers have tried and failed to reproduce another scientist’s experiments, and more than half of them have failed to reproduce their own experiments. Protzko et al.⁷ reported high replicability of novel social-behavioural findings, but the article itself was retracted due to methodological concerns. While many influential publications in different domains failed to be reproduced by independent researchers,^{8,9} non-reproducible publications seem to have been cited more often than reproducible ones, and this difference in citation does not change after the publication of the failure to reproduce or replicate.^{10,11} These findings raise alarms about the sustainability of science. Despite

improvements, many challenges in reproducible research are persistent.^{12–15}

In the field of modelling and simulation (M&S), reproducibility and replicability, abbreviated as R&R hereafter, is one grand challenge in both industry and academia.¹⁶ Compared with that of traditional (laboratory or field) experiments, the R&R of simulation models is presumably even worse.^{17,18} Computer simulation is increasingly popular in many scientific disciplines. However, computer models and experiments, especially dynamic stochastic simulations, are rarely reproduced or replicated by independent researchers.^{16,19,20}

¹Section Systems Engineering and Simulation, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands

²Bournemouth University, Poole, UK

Corresponding author:

Yilin Huang, Section Systems Engineering and Simulation, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, Delft 2628BX, The Netherlands.
Email: Y.Huang@tudelft.nl

A few R&R studies of simulation models exist. For example, Jalali et al.²¹ assessed 1613 articles that applied simulation modelling as a core method in health policy and epidemiology. They found that almost half of the articles did not report model details. A more in-depth evaluation of 100 of those articles showed that seven out of 26 evaluation criteria were satisfied by more than 80% of those articles. Only about 2% of these articles provided modelling code and had reproducibility discussions. Zhang and Robinson²² searched six prominent journals for articles focused on agent-based modelling (ABM). They found nine out of 348 resulting articles that aimed to replicate an existing model partially or entirely, indicating limited R&R studies. Riehl et al.¹⁸ assessed more than 11K simulation studies from 15 most renowned transportation journals. About 5% of the studies provided some form of repositories, and most offered “content of a rather mediocre level of usefulness.” Bajracharya and Duboz²³ and Donkin et al.²⁴ showed that different implementations of the same conceptual model on different modelling platforms can give significantly different results. The studies underline the necessity of R&R studies in simulation for producing more reliable results and understanding the potential pitfalls in simulation-based studies. In general, recent studies have called attention to R&R, which is a significant challenge in simulation-based research. Many published models and experiments lack sufficient documentation or accessibility for independent verification and validation.¹⁸ The challenges for R&R of simulation models appear to be more persistent than those for traditional experiments.^{17,25}

With the overarching goal of forming stronger community efforts for reproducible and replicable (R&R) simulation models, this paper aims to articulate the values, benefits, and challenges associated, as identified through the existing body of literature. The paper is partly based on Huang.²⁶ In this extended version, we identify the characteristics for R&R research, opportunities, and call for sustained and systematic efforts towards advancing R&R practices in the M&S field via promoting research and collaboration on this critical topic.

The rest of the paper starts with a brief review of the terminology used for R&R. Researchers in different domains do not have to align the use, but should be well-aware of the differences so that relevant work can be discovered in the literature across domains and disciplines. It is followed by different views and opinions on R&R, types of evaluation, and recent developments. We outline the challenges, characteristics, and opportunities presented by emerging technologies in relation to the R&R of simulation-based research. The paper argues that, given the current complexity and widespread use of diverse simulation models, the R&R of simulation needs to be explicitly addressed operationally in existing modelling practices by researchers who wish to engage in these

efforts, for which many social, conceptual, methodological, and scalability issues remain under-researched.

2. Reproducibility and replicability in simulation

There is no consensus in the scientific literature on what reproducibility is or should be.^{4,6} Closely related terms include replicability and repeatability, which have been used in different domains. They have long been used referring to the general concept of one experiment or study confirming the results of another by repeating the existing research in different ways.²⁷ However, within this general concept, the literature has not yet converged on a terminologically consistent conceptual framework.²⁷ Hence, the language of research reproducibility is non-standard.⁴ Different scientific disciplines and institutions may use the terms in inconsistent or sometimes even contradictory ways; some use them interchangeably.^{2,6,22,27–29}

Repeatability often implies that researchers can repeat the calculations of *their own* study and obtain the same results under the same setup.^{3,18,30} This paper will focus on R&R, namely the ability of *other researchers* to obtain consistent results when pursuing similar aims. (This “other researcher” may well be oneself in the future when the details of one’s own study become opaque without sufficient reporting and archiving.) In M&S, replications are also used referring to repeated runs of simulation models with different seeds of the random-number generators but otherwise the exact same model configurations – these runs are known as independent replications of simulation.³¹ For clarity, in this paper, reproducibility narrowly refers to the “computational reproducibility” of simulation models. We make the following distinction of R&R in M&S based on NAPress²⁷ and Hinsen.³²

- *Reproducibility* is obtaining identical results using the same input data, computational steps, methods, code, and conditions of simulation-based analysis.
- *Replicability* is obtaining consistent results across simulation-based studies aimed at answering the same scientific question, each of which has obtained its own data and/or uses different code.

According to this distinction, simulation studies that use the same core conceptual model but have different computational implementations – as those reported (Zhang and Robinson,²² Luijken et al.,²⁸ and Edmonds and Hales³³) are replication studies, regardless if the studies used the same input data.

Modellers commonly distinguish model conceptualisation from model implementation. The former, or a conceptual model, is often a textual, mathematical and/or diagrammatic description of model characterisation and

processes of interaction, based on a real system of modelling interest.¹⁷ A conceptual model is traditionally often not executable; thus, it may have ambiguities in how to compute model inputs to outputs.³⁴ The latter, or model operationalisation, or simply a simulation model, is a computational formalisation of a conceptual model into an executable computer programme where numerical output can be produced by executing (i.e. running) the implemented model.¹⁷

Reproducing or replicating simulation studies aims to demonstrate that a computer simulation experiment's results are repeatable and were not exceptional cases.^{19,34} Without verifying the claimed results through R&R simulation experiments, it is possible that published findings were incorrect due to, for example, programming errors, mistakes in the reporting or analysis of results, or misrepresentation of the simulation experiment.²⁰ Consistent results from model reproductions or replications can build confidence in the simulation mechanism used.³⁵

3. R&R research: why and why not

There are different views and opinions on the need for R&R of simulation studies in diverse research areas, and due to personal preferences, which can be equally interesting or at least open to debate. We briefly discuss those in this section.

3.1. Practicality and workload

Reproducibility can be hard or practically impossible to achieve for certain research areas, such as military and critical infrastructure, and for sensitive topics in medicine and other industries due to business interests, security, privacy, or ethical reasons, among others. These simulation applications often require non-disclosure agreements or are demanded by laws or regulations that limit or prevent the publication of information or software that is necessary to reproduce the simulation.^{16,18,36}

Making simulation (and results) more R&R demands dedicated, and sometimes extensive, time, and effort from the original researchers. According to a survey ($N = 87$) by Riehl et al.,¹⁸ major constraints include time, legal issues, lack of confidence to share material in the current state, lack of knowledge, and quality concerns. Many of those reasons also hold true for independent researchers or teams who aim to reproduce or replicate the work. As some put it, "publication is already a gruelling process, why would we increase our workload."³⁶

Currently, there is often no direct reward or consequence for researchers who exert effort for more R&R simulation. For example, the publication impact analysis by Riehl et al.¹⁸ showed that, while simulation studies received higher citations compared with studies that did not use simulations, when comparing simulation studies

with and without repositories, with good or less good repository quality, no significant citation difference can be found among those studies. There have been few or, at the very least, insufficient incentives to undertake what researchers view as additional work to make their simulation study more open and accessible.³⁶

3.2. Concerns and benefits

There are critical voices regarding R&R research. For example, Drummond³⁷ raises concerns about the strong influence the reproducible research movement is having on which papers get published; in addition, widening the responsibilities of peer reviews adds extra workload to reviewers and does not recognise the broad role the scientific community plays (at the post-publication stage) in determining the value of an idea. Drummond³⁸ states that reproducible research in some fields also requires open source code, which is a narrow interpretation of how science works; the effort necessary to meet the aim and the general attitude it engenders would not serve well any of the research disciplines. Fanelli³⁹ in his (Proceedings of the National Academy of Sciences) article argues that the rapidly growing scientific literature uncritically endorses a new "science is in crisis" or "reproducibility crisis" narrative, which is not only empirically unsupported and unreliable but also quite counterproductive and might foster cynicism and indifference in younger generations. Leonelli^{40,41} advocated for scientific pluralism and provided paradoxical real-life examples where the principles of Open Science (OS) and reproducible research clashed with responsible research measures and practices. The author further argued that unless relevant policies embrace a more sophisticated understanding of epistemic diversity, they may risk acting as a reactionary force that reinforces conservatism and increases inequity among researchers, given differences in power, resources, and visibility.

Regardless of the views or where the truth lies, clearly, not all models need or can be made R&R. It is a value and insofar often a voluntary community service being placed on good science by some researchers. For them, including the authors of this paper, reproducing and/or replicating a computational model can contribute to the scientific community in many different ways that cumulatively consolidate science. For example, commonly discussed benefits in the literature include developing shared understanding, obtaining an improved sense of accuracy, reliability, robustness, and range of plausibility of model results as well as empirical evidence to compare the models.^{4,22,28,34,42}

Besides those, many M&S researchers can readily relate to the frustrating scenario in which a new team member (whether an incoming PhD student or even one's future self) discovers that earlier simulation studies are not reproducible. Hence, R&R in simulation arguably matters first and foremost within research teams that seek continuity in

their own work. In addition, R&R studies are not only just a way to advance science but also a pedagogical tool, as they offer rich and complementary learning experiences for doctoral education.⁴³

3.3. Model reuse and interoperability

In addition, reproducing and/or replicating a computational model can be a good way to have a first assessment of the reusability of reported simulation models.^{30,36} Reusability generally refers to the degree to which research artefacts such as code, data, models, and documentation can be used again in different studies, contexts, or applications.^{44–46} It extends beyond reproducibility to emphasise the sustainability and adaptability of scientific outputs. Reusability is a dimension in the FAIR principles (Findable, Accessible, Interoperable, Reusable) of OS.⁴⁵ A reusable simulation model or data set is well-documented, openly accessible, and sufficiently modular to be integrated, modified, or extended for new research purposes.

Simulation studies often have two main types of audiences: methodological researchers and applied researchers.⁴⁷ The former reads a study to gain an overview of a method's uses, limitations, and potential improvements. The latter reviews a study with the main aim of applying the method or result to their own research problem. The reuse of simulation models can benefit both types of researchers.

The knowledge and data embodied by the simulation model are available to be utilised by model reusers as a tool to advance their own research agenda.³⁴ Model reuse also enables testing the broader parameter space of existing simulations.²⁴ It can further facilitate multi-modelling and hybrid simulation, that is, combining different models and modelling paradigms for the application of complex systems analysis.^{48–50} That said, reuse of research components and results, a concept closely related to R&R and OS, also requires a careful procedural approach, in the sense that it shall be treated with caution, as good practices can only emerge from diverse, context-dependent interpretations and responsible manners of implementation.⁴¹

An important factor enabling effective model reuse and collaboration across different research contexts is interoperability, which ensures that simulation models, data, and software can be integrated, exchanged, and executed across diverse platforms and tools. Interoperability is a critical enabler of R&R in simulation. The ability of different simulation platforms, software tools, and data formats to communicate seamlessly allows researchers to reproduce results across heterogeneous systems. Standardised data formats, modular software architectures, and open APIs facilitate the integration of components from diverse sources, reducing the risk of errors and inconsistencies. By promoting interoperability, researchers can more easily share models, validate results, and extend previous work,

thereby accelerating scientific progress and enhancing transparency.

4. Evaluation of R&R in simulation

For M&S, it is particularly beneficial to evaluate the R&R of a simulation model from two aspects of simulation: the method (i.e. the computational procedure of the simulation) and the simulated results. This distinction is inline with the “method reproducibility” and “result reproducibility” discussed in Goodman et al.⁴ The former means that the computer simulation is methodologically reproducible or replicable in theory and practice; the latter means that the simulated results can be quantitatively reproduced or replicated using the same computational method. Clearly, method R&R proceeds (and is necessary for) result R&R.

4.1. Method R&R

When the original simulation model (i.e. code and input data) is accessible to independent researchers, the method R&R can be directly evaluated. This typically means that independent researchers follow the simulation's computational procedure, as closely as possible – in the case of reproduction, using the same code, tools, and data – based on the workflows described in the original documentation and publication. If the original model is not accessible, the simulation has to be replicated, that is, assessed by newly implementing the simulation model based on the original publication and the model conceptualisation discussed within. Replicating a simulation, of course, can also be done independently when the original code and data are accessible. The actions needed to evaluate method R&R may appear straightforward; however, there are many associated challenges, which are briefly discussed later in the paper.

4.2. Result R&R

After method success, if the reproduced or replicated simulation generates outputs sufficiently similar to those of the original model, the reproduction or replication as a whole can be considered successful.¹⁹ The result R&R, that is, the quantitative measure of the similarity of results, however, is neither straightforward.⁵¹ For example, Muradchianian et al.⁵² reported difficulties in comparing multiple Frequentist and Bayesian measures because there is no established standard for the types of metrics used. Another complicating factor in the comparison is the different levels of publication bias.⁵² In relation to that, a broad categorisation of result R&R is the so-called “standards of equivalence” or “replication standards.” This is referenced in several studies^{17,19,22} and first appeared in the replication work of Axtell et al.⁴² The three general categories of model equivalence (from strict to loose) are

numerical, distributional, and relational equivalences, and they are summarised as follows.⁴²

1. Numerical equivalence (or identity) refers to the generation of exact reported results. It typically is not expected for stochastic simulations unless information on random seeds is specified.
2. Distributional equivalence is determined by showing that two studies produce distributions of results that cannot be distinguished statistically. This is determined by a statistical test of the null hypotheses.
3. Relational equivalence means that two models can be shown to have the same internal relationship among their results (i.e. inputs, parameters, and outputs). For example, two models show that a particular output variable is a quadratic function of time, or that a measure on a population decreases monotonically with the population size. This is the least demanding comparison, but for some theoretical purposes, it may suffice.

The separation of method and result R&R is particularly useful for simulation studies because R&R studies of complex models can be divided into two stages that are more manageable. The simulation method itself can be examined and tested or replicated first for methodological soundness, which is an important contribution of a simulation study. This stage also verifies the alignment of the conceptual model with the computational model and the experimental scenarios (and conditions). It tests, with the stated computational workflows and steps, whether the computational method can be executed as intended. In the second stage, the results are compared, which is typically performed in a traditional (non-simulated) replication study. For a simulation study of stochastic models, the numerical equivalence is expanded to distributional and relational equivalences, which are often more realistic and reasonable expectations depending on the particular goal of the individual simulation study.

5. Recent development and types of R&R studies

As research artefacts, simulation models should be R&R and potentially reusable by the broader scientific community. Establishing such properties not only enhances the credibility of individual studies but also supports the cumulative advancement of computational knowledge, including that of methodological adaptation and refinement. To this end, encouraging researchers to conduct R&R studies and to report both the successes and challenges plays a critical role in strengthening the evidence

base, making improvements, and fostering transparency in simulation research.²² Recently, there has been a marked increase in scholarly attention devoted to R&R across computational and simulation-based research. The alarming “reproducibility crisis” identified in several disciplines has motivated the development of domain-specific standards, methodological frameworks, and reproducible workflows.^{3,53–55}

The simulation community has conducted more studies now explicitly addressing the need for structured reporting and model sharing. For example, the RepliSims project²⁸ conducted replications of eight highly cited simulation studies and highlighted the key enablers of replicable statistical simulation. Monks et al.⁵⁶ introduced the STRESS documentation guidelines to strengthen the reporting of experimentation and results. Grimm et al.⁵⁷ updated the ODD (overview, design concepts, and details) protocol to promote consistent description of ABMs and enhance their reproducibility and comparability.

Notably, the computational biology and *in silico* medicine communities have been active in tackling R&R challenges.^{54,58,59} Relevant works include, for example, MIRIAM (minimum information in the annotation of models);^{60,61} MIASE (minimum information about a simulation experiment), minimum information about simulation experiments;⁶² SED-ML (simulation experiment description markup language),⁶³ COMBINE standards and formats,⁶⁴ as well as projects in VPH (virtual physiological human)⁶⁵ and *in silico* oncology.⁶⁶ Moreover, Knapp et al.⁶⁷ provided 10 simple rules for successful cross-disciplinary collaborations in computational biology. Ziemann et al.⁶⁸ introduced five pillars of computational reproducibility, which emphasised practices such as literate programming, code version control, and persistent data sharing to enhance reproducibility in bioinformatics and beyond.

Initiatives, such as Zhu et al.’s⁶⁹ integrated framework for assessing computational reproducibility, further elaborate on the importance of transparent workflows and code-data coupling, and the need for more holistic approaches to reproducible simulation-based science. Collectively, the increased efforts in R&R reflect a growing consensus that advancing reproducibility requires not only conceptual agreement but also concrete methods, workflows, and infrastructure tailored to the specific characteristics of simulation research. Within the growing body of work, three main types of research can be identified as shown in Table 1.

The first type, *empirical evaluation of R&R*, comprises studies that computationally reproduce or replicate published simulation models and associated results.^{28,70} Such a study typically evaluates one or a few published simulation-based studies, either by using original or new model code and/or data closely following the original

Table I. Three types of reproducibility and replicability (R&R) studies in simulation.

Type	Approach	Goal
I. Empirical evaluation of R&R	Computationally reproduce or replicate an original simulation-based study by using the original or new model code and/or data following the original experimental design and workflow.	To empirically evaluate the R&R of one or more published simulation-based studies
II. Assessment of R&R readiness	Typically assess the published simulation by indicators such as the openness of model code and data, the quality and coverage of reporting and documentation, and whether the work is cited or reused, etc. The original model is not computationally reproduced or replicated.	To assess the R&R potential of one or more published simulation-based studies
III. Research on or review of R&R studies	Theory building, development of methods, guidelines, tools; social and (research) culture studies, etc.	To improve the R&R of simulation practices and advance the field

experimental design and workflow. These studies are typically expensive to up-scale due to the time and effort required.

The second type, *assessment of R&R readiness*, refers to works that examine the published simulation by looking at indicators such as the openness of original model code and data, the clarity of workflows, the quality and coverage of original reporting and documentation, and whether the original publication was often cited or the model was reused by other publications, etc.^{18,21} This type of work, unlike Type I, does not directly evaluate R&R by executing the models and data. The assessment by indicators, however, can often cover a greater number of model publications, and the results are insightful to reveal the R&R potential of the original studies. It is also useful for scoping and prioritisation, enabling researchers to screen a large body of simulation-based studies and identify those that are promising for more detailed R&R investigations.

The third type of work encompasses diverse original research on simulation R&R or reviews of those works. They may be based on works of the two previous types. But instead of focusing on computationally reproducing or replicating individual original simulations or assessing their R&R readiness, the third type of work is aimed at generally improving the R&R of simulation practice.^{56,69,71,72} They can be intended for, for example, theory building, development of methods, guidelines, tools, as well as social and (research) culture studies, community engagement, policy analysis, and so on, to operationally and structurally improve R&R practices.

Even with recent developments and achievements, all three types of R&R studies are significantly lacking given the high number of computational simulation studies published in the literature and the increased importance of R&R of computational science in research. Such studies form the evidence base of R&R simulation-based research, provide theoretical and methodological foundations for more R&R practices, and have the potential to motivate

and engage more researchers in intellectual exchange and debates on this important topic.

6. Challenges in reproducing or replicating a simulation study

Different domains and disciplines are often of a distinct nature, thus resulting in models of a different nature. They can be, for example, with different levels of abstraction and details, and various stochastic characterisations representing uncertainty.¹⁷ Unlike in many physical sciences and engineering domains, systems that have less clear or agreed-upon “ground truth,” such as social systems or value systems, have many degrees of freedom and levels of uncertainty in model conceptualisation. In addition, technological advances and social development form a “seamless web,” via which the connected parts substantially shape and reshape one another.⁷³ This system perspective shapes how we develop models. See the models in the work by Shrestha et al.,⁷⁴ used for hardware development in the automotive industry and health care, versus the models in the work by Balkan et al.⁵⁰ for estimating infection risk, in the work by Wagenaar et al.⁷⁵ for neonatal care interventions, or in the work by Sundaram et al.⁷⁶ for incorporating justice considerations in energy transitions. The conceptualisation of the latter types of models is often subject to disparate or even inconsistent interpretations. This often means that simulation R&R studies of such models by independent researchers are particularly hard, even when the original code and data are available.

Many reported that communications (sometimes extensive personal interactions) between the researchers and the original author(s) of the work are helpful for simulation R&R studies.^{17,22,34} However, when such communications are not possible, for example, when the original researchers have left the field of work or are not reachable, a R&R study needs to rely on the original published materials. In

such cases, typical challenges reported in literature concern two categories of source materials: (1) the reporting of the original simulation studies and (2) the documentation, corresponding models, and associated data (if they are available). The challenges, briefly summarised in the following paragraphs, impede the process and the comparison of results when empirically reproducing or replicating a simulation study.^{20,22,33,77} We do not include contributing factors such as increased project complexity, publication biases, lack of incentives and funds, and legal and regulatory issues. These are systematic or structural and can be found in the works by Alston and Rick,² Baker,⁶ Begley et al.,⁷⁸ and Munafo et al.⁷⁹ We do not aim to give a detailed account of the challenges but to outline noteworthy ones in the literature.

Many reproducibility challenges are caused by incomplete or ambiguous reporting and documentation of simulation studies.^{17,22,56,80} It is not uncommon that the original work does not contain enough information regarding the conceptual model, computational model, and/or experimental conditions. Studying and understanding the conceptual model is often the most significant step²² where ambiguity in communicating a model and its experimental conditions can result in varying interpretations, including that of assumptions, mathematical processes, and mechanisms.^{17,22} When simulation models include stochastic processes, the experimental conditions, including the seeds, are often not explicitly mentioned or lack sufficient details, which makes numerical or distributional equivalence difficult to achieve, and the method's strengths and limitations poorly understood.^{22,81} Sometimes, the referred documents or online resources in the original publication cannot be retrieved.

The availability of the original study's source code and associated data is helpful for R&R.^{22,80} However, openness alone is often insufficient for ensuring reproducibility.³⁰ Many challenges can arise from the available model software and data, as well as the hardware, regarding their executability and comprehensibility.^{3,82} Software includes, for example, the required simulation environments (a.k.a. platforms), code libraries or toolkits, and the programming languages used for model implementation, as well as those for data management, processing, and analysis. The choices often have a strong influence on how models can be represented and interact, and they may yield different outcomes. Cross-platform and cross-language replications, as well as the portability and consistency of different algorithms, workflows, and their performance constraints, can be sources of significant variability between model results.^{3,24,30}

Sometimes, the descriptions of published reporting and documentation were explicit but had incorrect or inconsistent source code or data in relation to the descriptions.^{17,22} They can be caused by the translation from conceptual model to model implementation (or vice versa), or by the alignment of the report to the conceptual model or

implementation.¹⁷ This may also be due to versioning issues, such that the reporting was updated (e.g. during the reviewing process), but the code base or documentation (partially) was not.

Despite the ongoing progress towards more transparent reporting and documentation, significant challenges remain in achieving R&R in simulation research. Overcoming these challenges will require time and effort to adhere to reporting standards, as well as (research) culture, methodological, and infrastructural commitments to advance a truly R&R simulation science.

7. Characteristics of R&R simulation

Reproducing and replicating simulation studies form the methodological foundation for ensuring the credibility of simulation results. Although they share overlapping goals, each emphasises different aspects of transparency, verification, and validation in research practice.

On one hand, *reproducible* simulation research requires comprehensive documentation such that it enables an exact reconstruction of workflows. Achieving reproducibility depends on open and sustained access to all components of the computational process, including source code, data sets, software dependencies, and computational environments.⁸¹ Key reproducibility features include the following:

- *Source code availability*: All code used for simulations (including scripts, algorithms, and randomness control) should be archived and available in stable repositories, for example through public version-controlled platforms such as GitHub, GitLab, or institutional repositories with persistent identifiers.
- *Data set availability*: All data used for simulations should be available in stable repositories. Details about the metadata, variable definitions, distributions, and sampling procedures should also be explained to ensure the complete specification of the data set, for instance using standardised data documentation formats or data descriptors.
- *Environment set-up description*: Detailed explanation of the development and test environment should be provided, including software versions, hardware configurations, dependencies, and parameter settings. This can be facilitated by environment management tools such as Docker, Conda, Pip, or virtual machines, which allow the computational environment to be explicitly specified and reconstructed.
- *Workflow specification*: Each analytical step, from pre-processing to experimental steps, and post-processing, should be clearly explained and traceable either through literate programming tools (e.g. Jupyter Notebooks) or workflow manager tools.

Table 2. Reporting and documentation categories and features for reproducibility and replicability.

Category	Reproducibility	Replicability
Study design/Workflow	Hypothesis, variables, setup, detailed workflow	Methodology, population, protocols, analytical steps
Code/Simulation model	Full source code, versions, dependencies	Pseudo-code, model logic, assumptions
Data set	Metadata, all data, pre-processing routines	Metadata, procedures, sampling methods

On the contrary, *replicable* simulation research extends validation beyond the original implementation and data set. It tests the generalisability of the approach, that is, whether a result holds true under different implementation contexts, such as alternative programming languages, platforms, or modelling frameworks. Key characteristics for replicability include the following:

- *Experimental design*: Detailed explanation of the model and experiments should be provided, including the model logic, equations, assumptions, and conceptual structure. This may be supported by formal model descriptions (e.g. the ODD protocol for ABMs), diagrams, or mathematical specifications that allow independent reimplementations.
- *Data set information*: Metadata and variable definitions should be clearly documented. In addition, it is advisable to deposit the data in stable repositories to allow future teams to access and reuse the same data sets, or to generate equivalent data sets when replication relies on synthetic or simulated data.
- *Specification of stochastic processes*: Details about random-number seeds, distributions, and sampling procedures should be explained so that the structure of uncertainty is understood and can be independently reproduced using alternative implementations.
- *Limitations and constraints*: The domain context, limitations, and constraints should be explained to indicate where replication may fail due to computational or conceptual constraints, for instance when results depend on specific hardware architectures, restricted software, or unavailable real-world data.
- *Workflow transparency*: The overall workflow and analytical steps should be clearly described to ensure that others can understand, evaluate, and replicate the study, for example through high-level workflow diagrams or platform-independent protocols.

Table 2 summarises the essential reporting and documentation categories and features for R&R in simulation.

The study design of reproducible simulation shall be accurately described so that an independent researcher can understand the hypothesis, overall goal, variables, and setup and rerun the analysis using the same code and input data. For replicable research, researchers shall go beyond describing the original study setup and specify the

conditions under which a new team could independently recreate the study with potentially new data. The study design should, therefore, describe the population, methodology, data collection protocols, and criteria for generating equivalent or similar data sets.

The workflow of reproducible simulation should be accurately documented to enable exact repetition, with all code, scripts, and software versions fully disclosed. For replicable simulation, the workflow must outline the procedures for analysis, but the emphasis shifts towards generality, ensuring that a new implementation (with possibly different data or context) following the described protocol can address the same scientific question.

In addition, for both R&R, the research integrity and consistency between the reporting, documentation, code and data are essential. Clear and well-defined conceptual models, transparent methodological descriptions, appropriate use of theoretical frameworks, and thorough validation of results should be ensured to maintain rigour and reproducibility.

8. Emerging opportunities for R&R

Emerging opportunities are transforming R&R in simulation-based research, offering pathways to greater transparency, quality, and community impact. To fully realise these benefits amid increasing model complexity and diversity, R&R must be systematically embedded into simulation workflows and supported by stronger, collaborative community efforts.

8.1. Open science and community-driven practices

The advances in OS practices, such as the use of Free and Open Source Software (FOSS), open code repositories, and data sharing platforms, have lowered barriers to sharing and reusing simulation models.⁸³ Public repositories for code, models, and simulation data sets (e.g. GitHub, Zenodo and Open Science Framework) enable transparent access to the full modelling pipeline. Initiatives now promote not only the publication of model code alongside articles but also the packaging of workflows into user-friendly web applications, improving accessibility for researchers and practitioners beyond technical specialists. Containerisation and environment specification tools, such as Docker and Conda, enable the recreation of

computational environments, mitigating issues arising from software version incompatibilities. Community repositories and FAIR data principles are also increasingly standard, helping ensure that code, data, and documentation remain available for assessment and reuse.⁴⁵

Another key trend is the co-development of community-driven standards for documenting and sharing simulation studies.⁸⁴ This includes formal reporting guidelines for simulation design, experimental protocols, parameter definitions, and software environments.⁴⁶ Community platforms (e.g. streamlit-based apps) facilitate collective input and code reviews, encouraging reusable and robust model development. Tools that enable standardised archiving, provenance, persistent identifiers, and reproducibility checklists help embed good practices across the research lifecycle.

8.2. Utilising AI and software engineering best practices

Given the complexity and diversity of simulation models, R&R should not be an afterthought. Instead, workflows should be built for transparency from the onset using literate programming, containerised environments, version control, and modular code organisation. Research teams should adopt clear protocols for documenting model provenance, tracking changes, and linking code, data, and outputs.

As an emerging technology, artificial intelligence (AI) presents practical opportunities to enhance R&R in simulation. Frameworks for embedding R&R in simulation can employ AI to support or automate (or semi-automate) known best practices at each modelling stage. For example, AI-driven tools can assist in tasks such as workflow documentation, standardisation of naming conventions, and generation of metadata for models and outputs.⁸⁵ AI can also assist in benchmarking and testing reproducibility, generating synthetic data for replication, and supporting intelligent search and annotation of simulation resources.⁸⁶

Generative AI models can further assist in automating complex procedures or searching the literature for best practices and emerging standards and can be regarded as a specific and increasingly influential subclass within the broader landscape of machine learning (ML) and AI. Adaptive and generative AI methods combine self-optimising model behaviour with the ability to synthesise new data, models, or workflows. In simulation, adaptive techniques can continuously calibrate models based on feedback from simulation outputs, improving accuracy and robustness over time.

In the literature, recent studies have begun to explore specific uses of generative AI in simulation contexts.^{85,87,88} For example, Alshareef et al.⁸⁹ demonstrate how generative AI can assist in transforming activity and

flow-based diagrams into executable simulation models. Ghaffarzadegan et al.⁹⁰ introduce generative ABM, integrating large language models with agent-based simulations to enhance model construction. In addition, Jackson et al.⁸⁸ propose a framework that uses generative AI and natural language processing to automatically generate simulation models of logistics systems from verbal descriptions. These studies collectively illustrate the potential of AI to support and accelerate simulation model development.

However, the outputs of AI, especially generative models, are inherently unpredictable and may contain errors, inconsistencies, or biases. It is, therefore, imperative that researchers independently verify any AI-generated content or analyses before integrating them into simulation studies. Hence, the limits of AI with regard to R&R should be explicitly recognised, as AI can assist and augment reproducible workflows but cannot replace careful methodological design, documentation, or independent validation. ML and AI can also aid in parameter sensitivity analysis, uncertainty quantification, and anomaly detection, ensuring that model outputs are both robust and reproducible.

In addition, embedding model-driven engineering (MDE) into simulation modelling, especially in conjunction with AI, represents a major opportunity to systematise R&R within complex computational workflows. While MDE is not new, its integration with AI-assisted automation and OS frameworks can now redefine how simulation studies can be better designed and performed, and provide knowledge-based constraints on and verification for AI-based results. By defining explicit models and transformations, MDE facilitates traceability from conceptual designs to executable simulations, and its systematic application remains highly relevant for R&R.^{91,92} Figure 1 illustrates an AI-assisted simulation modelling workflow that embeds MDE and OS.

Integrating MDE within simulation workflows ensures that models are modular, parameterised, and version-controlled, enabling consistent replication across different computational environments. Moreover, MDE complements AI-assisted practices by providing well-defined model artefacts that AI tools can analyse, validate, or optimise. For example, AI algorithms can automatically check consistency between model layers, suggest optimisations, or detect discrepancies in model transformations, while MDE guarantees a structured representation of simulation logic and dependencies.

8.3. Quantum computing for simulation and the way forward

Quantum computing is another emerging topic offering a novel computational paradigm that has the potential to significantly expand the scale, complexity, and fidelity of simulation-based research, creating new opportunities for

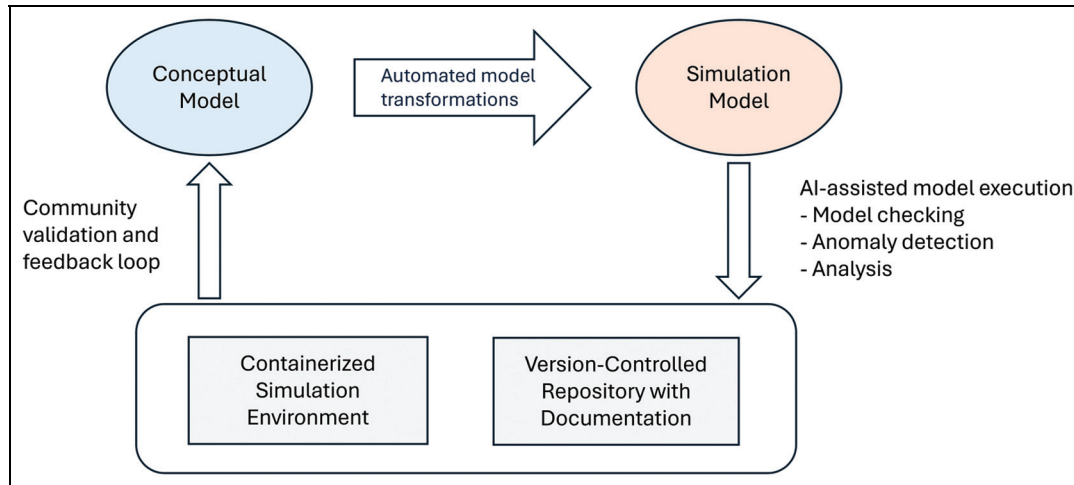


Figure 1. An AI-assisted simulation modelling workflow.

R&R. Quantum computing utilises the principles of quantum mechanics, including superposition and entanglement, to perform computations that would be infeasible for classical computers.⁹³

Unlike classical bits, which represent either 0 or 1, quantum bits (qubits) can encode multiple states simultaneously, enabling parallel exploration of large computational spaces.^{94–96} This capability makes quantum computing particularly promising for complex simulations in areas such as optimisation, molecular modelling, and stochastic processes, where traditional approaches face scalability limitations. This increased computational capacity enables researchers to explore larger parameter spaces and perform more exhaustive sensitivity analyses, which enhances the robustness and reproducibility of simulation outcomes.

The relationship between quantum computing and simulation is bidirectional. Quantum computing introduces new computational principles that may substantially extend the scope and efficiency of simulation methods, while simulation remains essential for the design, evaluation, and validation of quantum algorithms, architectures, and hardware systems.

However, while quantum computing offers opportunities for advancing simulation capabilities, it also introduces novel challenges to R&R. Quantum-enhanced simulations are inherently sensitive to hardware variability, noise, and stochastic quantum effects, making exact reproduction of results across different platforms difficult. Moreover, the current lack of standardised quantum software frameworks, coupled with limited access to high-fidelity quantum hardware, further complicates verification and replication efforts. Despite these challenges, by carefully designing hybrid quantum-classical workflows, employing modular and well-documented quantum subroutines, and integrating

classical verification methods, researchers can mitigate risks and foster replicable and transparent simulation studies in this rapidly evolving computational paradigm.

A significant area for future research lies at the intersection of quantum computing and agentic AI.^{97,98} This emerging research space introduces both conceptual and methodological complexities, particularly regarding the integration of quantum computational elements, such as superposition and entanglement, into simulation-based learning and decision-making frameworks.⁹⁹ Beyond purely technical integration, this line of inquiry also raises broader questions about model design, interpretability, and evaluation, as quantum-enhanced components may fundamentally alter the behaviour and outcomes of simulated agents.

Future research may, therefore, focus on the development of dedicated modelling and assessment frameworks to systematically investigate these hybrid systems. In addition, establishing reproducible and replicable protocols for such quantum–agentic systems will be essential to ensure that results can be independently verified, that integration strategies are robust across platforms, and that the benefits of these emerging technologies can be reliably evaluated and adopted in both research and applied contexts.

8.4. Practical implications and real-world applications

Emerging opportunities in R&R are particularly relevant to industry-focused applications of simulation-based research. In sectors such as aerospace, manufacturing, and health care, the ability to reliably reproduce simulation outcomes is vital for design verification, regulatory compliance, and risk assessment. By adopting reproducible workflows, organisations can reduce development costs,

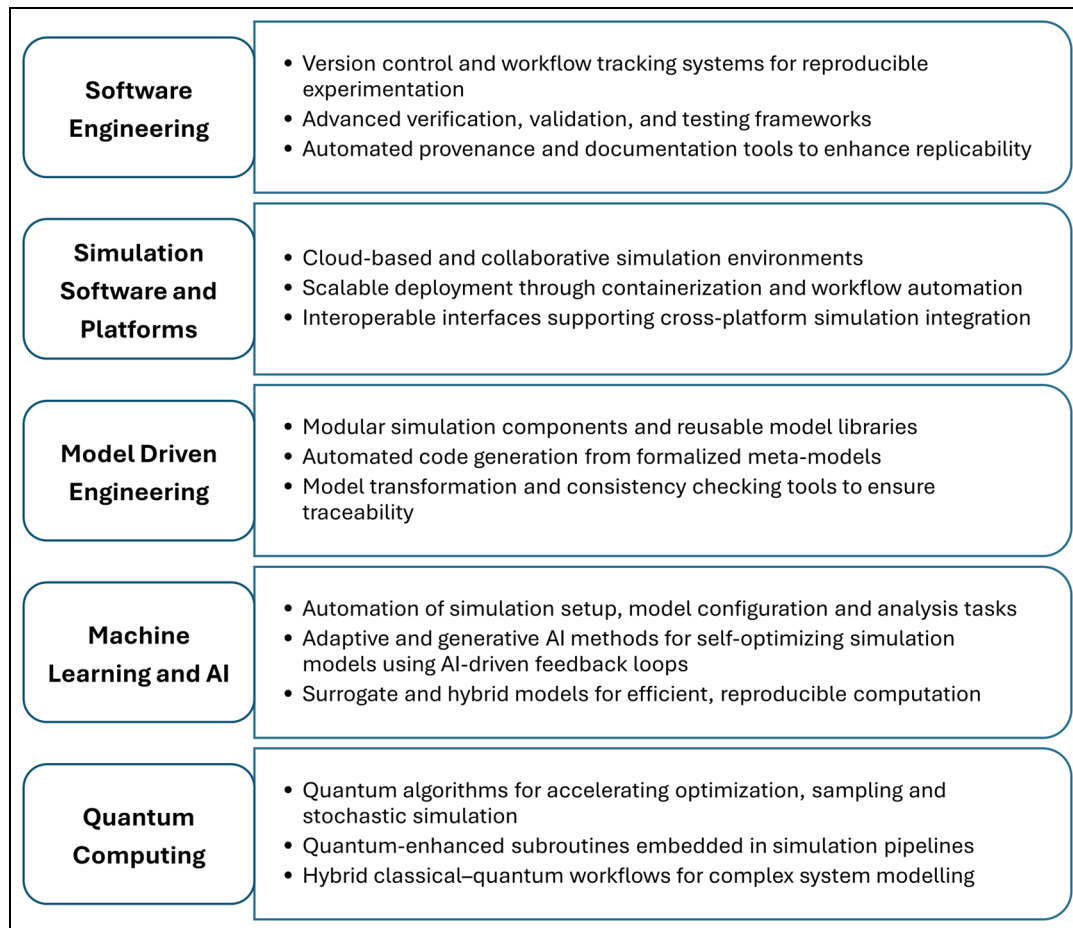


Figure 2. Emerging opportunities and technologies for reproducibility and replicability in simulation.

streamline testing procedures, and enhance confidence in simulation-driven decision-making. In addition, industry adoption of reproducible practices encourages collaboration with academia, facilitating knowledge transfer and the co-development of robust simulation models. Figure 2 lists some of the potential technologies that could advance the simulation field and specifically the R&R in simulation.

Once implementable methods and supportive tools are more widely available, the benefits of R&R in computer simulation are expected to extend across multiple domains. Enhanced R&R will not only facilitate knowledge transfer between methodological and applied researchers but also accelerate robust and reliable innovation in fields such as health care, manufacturing, transportation, energy systems, and environmental modelling. In particular, the integration of digital twin technologies offers opportunities for continuous validation and real-time simulation, bridging the gap between virtual models and real-world systems. There are many pilot projects as well as mature examples, from personalised medicine¹⁰⁰ to space missions such as digital twins of the James Webb Space Telescope¹⁰¹ and Mars missions,¹⁰² and those used for high-energy physics

experiments¹⁰³ and Earth systems.¹⁰⁴ Given their expected impact on society, the role of R&R in them is pivotal and self-evident.¹⁰⁵

9. Discussion and reflection

Simulation is a powerful method of inquiry enabled by computing. Unlike field experiments or laboratory investigations, simulation models generate results relying on computational routines. Thus, their computational R&R is a basis for meaningful analysis, corroboration, and further use of the results. Scientists and grant agencies have spent a large amount of time and funds on projects that develop new simulation models – while these efforts have been indispensable and fruitful, they often do not explicitly address computational R&R, despite the fact that many regard it as critical for scientific simulation.^{22,28,106}

At the same time, simulation models are becoming increasingly complex and more widely used. Recent developments in M&S also included more functionalities that use AI, particularly ML. The domains of application of simulation have expanded beyond the physical sciences

and engineering to include the social and behavioural sciences, among others. Those who use and develop scientific simulation models are well-trained in their respective domains, but not necessarily in the software aspects of computational methods. All of these make simulation R&R a more complex task to pursue and carry out. As a community of model users and developers, we need to recognise that R&R is extremely challenging and time-consuming in practice, and that the steps needed to tackle it are not going to be acceptable to everyone.^{36,38,106} Nonetheless, we need and have to talk (more often) about R&R as a community.⁵³ In particular, when simulation is used for decision support with real-world impact, for example, in safety-critical applications, the reliability of the computational results must be investigated.¹⁷

The recent growing emphasis on R&R has become particularly visible within the simulation community, where researchers are increasingly formalising methods and infrastructures to ensure transparent, verifiable, and shareable computational studies. Over the past decade, simulation-oriented conferences and journals have begun to recognise R&R not merely as an ethical imperative but as a technical and scholarly contribution in its own right. An example is the *ACM SIGSIM Conference on Principles of Advanced Discrete Simulation* (PADS), which has established an approach to reporting the reproducibility of simulation papers. These reports assess whether the original results can be independently reproduced using the authors' shared artefacts, code, and documentation. For instance, Rossi and Vabdin¹⁰⁷ reported on the reproducibility of the experiments by Piccione and Pellegrini.¹⁰⁸ Similarly, the journal *ACM Transactions on Modeling and Computer Simulation* (TOMACS) is aligned with ACM's Task Force on Data, Software, and Reproducibility and supports artefact review and badging policies.¹⁰⁹ These are signs of a cultural shift towards valuing R&R as a measurable research output. R&R should not be viewed solely as an ancillary best practice but as an integral component of scientific progress, enabling simulation studies to be independently verified, extended, and trusted across diverse application domains.

The OS initiatives taking place in many parts of the world have brought about positive changes, making science more transparent and accessible. However, for those who wish to engage in the R&R of simulation models, openness alone is insufficient. There is still a significant gap to overcome towards R&R and the potential reuse of simulation models. With the R&R research movement, many authors have also advocated for more structural and cultural change in institutions and research communities, for example, good institutional practice, appropriate incentive and evaluation systems, funding policies, journal guidelines and standards, training programmes, among others.^{36,78,106,110}

Besides those, another crucial question we need to ask ourselves, which is not yet sufficiently addressed, is how

to make R&R more attractive and operational, for both model builders (i.e. original researchers) and model reproducers or replicators (i.e. independent researchers)? They face related but distinct challenges. Therefore, they require different types of skills and support. For both, how can the "extra" time and effort be made more worthwhile and implementable, for researchers who primarily focus on original research and already have high workloads? There are no simple solutions. We discuss four reflexive thoughts that could fit well with the efforts in simulation modelling.

The first is to tie R&R in M&S more closely to model reuse. Model reusability is challenging in itself, while R&R is a promising first step towards model reuse. This can serve as an incentive for original and independent researchers, since reusing a computational model is a recognition of the original work, which also eases model reusers in their own research. It would be useful for researchers to have infrastructures and resources, such as registers of reproducible or to-be-reproduced simulation models, to promote domain-specific corroboration and potential model reuse. Currently, there is no lack of online code and model repositories and versioning systems, but how R&R is the simulation model indexed therein is highly unclear, and they generally lack peer evaluation. In this regard, more publication venues can consider publishing high-quality simulation R&R studies, as well as requesting and supporting reviews and evaluations of simulation artefacts.

The second is to develop processes, methods, and supporting elements, such as benchmarks and tooling, that can integrate well with existing simulation model development workflows and practices. While R&R is a socio-economic problem,³² it also has many methodological issues to overcome. Existing works of R&R in different disciplines often have their particular focus. For example, the *ACM Conference on Reproducibility and Replicability* (ACM REP, inaugurated in 2023) places a strong emphasis on computational issues (<https://acm-rep.github.io>). *ReScience C*,¹¹¹ an open-access journal, is dedicated to the publication of high-quality replications implemented with FOSS (<https://rescience.github.io>). While these initiatives and venues are beneficial to the M&S community, a simulation model is a special piece of software that focuses on the *imitation* of dynamic systems that change over time. We argue that general good practices often stem from software engineering, and they are helpful but often not sufficient for R&R simulation. For simulation, experimental workflows and steps, data trajectory of state changes, and results of different experiments, etc., should be captured and managed in an incremental M&S-process-based research cycle.³⁰ In addition, being able to identify the type and level of complexity of simulation experiments is a significant help in estimating the time and effort needed for R&R studies. Such methods and tools are generally not available but urgently needed in the field. Research

topics in this regard include experiment (and scenario) management, documentation, complexity evaluation, etc. The M&S community could organise regular R&R-specific sessions and workshops in conjunction with existing initiatives or at M&S-focused venues to engage more researchers and foster more discussion and collaboration within the community.

Moreover, because model conceptualisation entails simplification and abstraction based on many assumptions, given certain goals of a simulation study, modellers shall be facilitated in capturing such conceptual conditions in a more clarifying and methodological way.⁴⁸ There are efforts and recommendations towards this direction, for example, metadata structures, model description languages, and reporting guidelines.^{3,19,30,56,57,82} We also need a better understanding of researchers' and practitioners' needs and workflows, which are often highly contextual, to design (or avoid over-designing) methods and tools that can facilitate the wider adoption of more formal descriptions of conceptual models and assumptions, thereby enhancing model R&R.

Last but not least, R&R simulation entails a continual process and a cultural change towards more sustainable simulation research. Hence, it shall not be treated as an afterthought of a scientific study or as an end in itself.^{2,4,30} It calls for a way of working and thinking about how we develop and maintain our scientific simulations. Some might think that we need more openness, transparency, better methods and good practices to make simulation more R&R. While these elements are indeed essential, we argue that to effectively implement, adapt, and expand existing practices for more R&R simulation, we first need incentive building and engaging more researchers. How to effectively stimulate and sustain community engagement for R&R simulation is still largely unclear. Nonetheless, a good starting point is to explicitly consider and activate the important role of research institutions and graduate programmes, for example, in recognition and rewards, adapting research assessment criteria, and integrating R&R and OS elements in education and training.^{43,110} Advancing R&R in computer simulation demands not only guidelines, methods, and tools, but also structural, social, and (research) cultural investigations of simulation practices.

10. Conclusion


R&R in simulation face significant challenges. Yet, emerging opportunities, such as community-driven practices, AI, and quantum computing, offer promising paths for advancement. By ensuring that simulation results can be independently reproduced and replicated, researchers can mitigate uncertainties in the method and enhance the

credibility of simulation-based findings. General principles and guidelines, such as transparency, version controls, containerised environments, and adherence to community standards, remain essential for supporting these efforts. Simulation-specific methods and tooling, as well as ongoing community engagement, are also indispensable for improving R&R in simulation.

In addition, reviews of simulation studies will be insightful with more detailed R&R evaluation and assessment, to enable a clear mapping of current R&R evaluation and assessment practices, and identify the effectiveness of evaluation and assessment criteria and methods. Such efforts would also support the development of more robust methods to produce R&R simulation-based research. The work of Luijken et al.²⁸ and Heather et al.⁷⁰ presented such investigations of R&R. There remains a clear need to broaden the scope of such reviews and expand into other domains. To this regard, the development of more automated R&R evaluation and assessment methods and tools can help the simulation community implement such studies and achieve more scalable results.

The convergence of conceptual understanding, methodological innovation, supportive tools, and standardised practices aimed at enhancing R&R holds great promise for simulation-based research. By addressing R&R in a systematic and scalable manner, the simulation community can strengthen scientific rigour, foster collaboration and peer learning, and generate more credible insights into increasingly complex systems.

ORCID iD

Yilin Huang  <https://orcid.org/0000-0002-4396-1990>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Peng RD. Reproducible research in computational science. *Science* 2011; 334: 1226–1227.
2. Alston JM and Rick JA. A beginner's guide to conducting reproducible research. *Bull Ecol Soc Am* 2021; 102: e01801. <https://doi.org/10.1002/bes2.1801>
3. Antunes B and Hill DR. Reproducibility, replicability and repeatability: a survey of reproducible research with a focus on high performance computing. *Comput Sci Rev* 2024; 53: 100655. <https://doi.org/10.1016/j.cosrev.2024.100655>

4. Goodman SN, Fanelli D and Ioannidis JPA. What does research reproducibility mean? *Sci Translat Med* 2016; 8: 341ps12. <https://doi.org/10.1126/scitranslmed.aaf5027>
5. Fidler F and Wilcox J. Reproducibility of scientific results. In Zalta EN (ed.) *The Stanford encyclopedia of philosophy, summer 2021 Ed.* Metaphysics Research Lab, Stanford University, 2021.
6. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016; 533: 452–454.
7. Protzko J, Krosnick J, Nelson L, et al. RETRACTED ARTICLE: high replicability of newly discovered social-behavioural findings is achievable. *Nature Human Behav* 2024; 8: 311–319. <https://doi.org/10.1038/s41562-023-01749-9>
8. Rodgers P and Collings A. Reproducibility in cancer biology: what have we learned? *Elife* 2021; 10: 75830. <https://doi.org/10.7554/elife.75830>
9. Pavlov YG, Adamian N, Appelhoff S, et al. Eegmanylabs: investigating the replicability of influential EEG experiments. *Cortex* 2021; 144: 213–229. <https://doi.org/10.1016/j.cortex.2021.03.013>
10. Serra-Garcia M and Gneezy U. Nonreplicable publications are cited more than replicable ones. *Science Adv* 2021; 7: eabd1705. <https://doi.org/10.1126/sciadv.abd1705>
11. Hardwicke T, Szűcs D, Thibault R, et al. Citation patterns following a strongly contradictory replication result: four case studies from psychology. *Adv Method Pract Psychol Sci* 2021; 4: 40837. <https://doi.org/10.1177/25152459211040837>
12. Korbmacher M, Azevedo F, Pennington C, et al. The replication crisis has led to positive structural, procedural, and community changes. *Commun Psychol* 2023; 1: 3. <https://doi.org/10.1038/s44271-023-00003-2>
13. Nature SC. Challenges in irreproducible research, 2018, <https://www.nature.com/collections/prbfkwmwvz>
14. Derksen M, Meirmans S, Brenninkmeijer J, et al. Replication studies in the Netherlands: lessons learned and recommendations for funders, publishers and editors, and universities. *Accountabil Res* 2024; 32: 1285–1303. <https://doi.org/10.1080/08989621.2024.2383349>
15. Cobey KD, Ebrahimzadeh S, Page MJ, et al. Biomedical researchers' perspectives on the reproducibility of research. *PLoS Biol* 2024; 22: 1–15. <https://doi.org/10.1371/journal.pbio.3002870>
16. Taylor SJE, Khan A, Morse KL, et al. Grand challenges for modeling and simulation: simulation everywhere – from cyberinfrastructure to clouds to citizens. *SIMULATION* 2015; 91: 648–665. <https://doi.org/10.1177/0037549715590594>
17. Fitzpatrick BG. Issues in reproducible simulation research. *Bull Math Biol* 2019; 6: 1–6. <https://doi.org/10.1007/s11538-018-0496-1>
18. Riehl K, Kouvelas A and Makridis M. Revisiting reproducibility in transportation simulation studies. *Europ Transport Res Rev* 2025; 17: 22. <https://doi.org/10.1186/s12544-025-00718-9>
19. Yilmaz L and Ören T. Toward replicability-aware modeling and simulation: changing the conduct of M&S in the information age. In: Tolk A (ed.) *Ontology, Epistemology, and Teleology for Modeling and Simulation: Philosophical Foundations for Intelligent M&S Applications*. Springer, 2013, pp.207–226. https://doi.org/10.1007/978-3-642-31140-6_11
20. Axelrod R. Advancing the art of simulation in the social sciences. *Japan J Manag Inform Syst* 2003; 12(3): Special Issue on Agent-Based Modeling.
21. Jalali MS, DiGennaro C, Guita A, et al. Evolution and reproducibility of simulation modeling in epidemiology and health policy over half a century. *Epidemiol Review* 2021; 43: 166–175. <https://doi.org/10.1093/epirev/mxab006>
22. Zhang J and Robinson DT. Replication of an agent-based model using the replication standard. *Environ Model Softw* 2021; 139: 105016. <https://doi.org/10.1016/j.envsoft.2021.105016>
23. Bajracharya K and Duboz R. Comparison of three agent-based platforms on the basis of a simple epidemiological model. In: *Proceedings of the Symposium on Theory of Modeling & Simulation – DEVS Integrative M&S Symposium. DEVS'13*, San Diego, CA, 7–10 April 2013. Society for Computer Simulation International.
24. Donkin E, Dennis P, Ustalakov A, et al. Replicating complex agent based models, a formidable task. *Environ Model Softw* 2017; 92: 142–151. <https://doi.org/10.1016/j.envsoft.2017.01.020>
25. Fitzpatrick BG, Gorman DM and Trombatore C. Impact of redefining statistical significance on p-hacking and false positive rates: an agent-based model. *PLoS ONE* 2024; 19: 1–18. <https://doi.org/10.1371/journal.pone.0303262>
26. Huang Y. Reproducibility and replicability of simulation models. In: *2025 Annual Modeling and Simulation Conference (ANNSIM)*, 26–29 May 2025, pp.1–10. IEEE. <https://ieeexplore.ieee.org/document/11118683>
27. NASEM. *Reproducibility and Replicability in Science*. National Academies of Sciences, Engineering, and Medicine, The National Academies Press, 2019. <https://doi.org/10.17226/25303>
28. Luijken K, Lohmann A, Alter U, et al. Replicability of simulation studies for the investigation of statistical methods: the RepliSims project. *Royal Soc Open Sci* 2024; 11: 231003. <https://doi.org/10.1098/rsos.231003>
29. Gundersen OE. The fundamental principles of reproducibility. *Philos Trans Royal Soc A* 2021; 379: 2197.
30. Chen X, Dallmeier-Tiessen S, Dasler R, et al. Open is not enough. *Nature Phys* 2019; 15: 113–119. <https://doi.org/10.1038/s41567-018-0342-2>
31. Law AM. *Simulation Modeling and Analysis*. 4th ed. McGraw-Hill, 2007.
32. Hinsin K. Reproducibility and replicability of computer simulations. In: *ACM REP'24*, Rennes, France, 18–20 June 2024, <https://hal.science/hal-04621140/>
33. Edmonds B and Hales D. Replication, replication and replication: some hard lessons from model alignment. *JASSS – J Artif Societ Soc Simul* 2003; 6: U227–U253.
34. Wilensky U and Rand W. Making models match: Replicating an agent-based model. *J Artif Societ Soc Simul* 2007; 10: 2.
35. Edmonds B and Hales D. Computational simulation as theoretical experiment. *J Math Sociol* 2005; 29: 209–232. <https://doi.org/10.1080/00222500590921283>

36. Taylor SJE, Eldabi T, Monks T, et al. Crisis, what crisis – does reproducibility in modeling & simulation really matter? In: *2018 Winter Simulation Conference (WSC)*, Gothenburg, 9–12 December 2018, pp.749–762. <https://doi.org/10.1109/WSC.2018.8632232>
37. Drummond C. Reproducible research: a minority opinion. *J Exper Theoret Artif Intel* 2018; 30: 1–11. <https://doi.org/10.1080/0952813X.2017.1413140>
38. Drummond C. Is the drive for reproducible science having a detrimental effect on what is published? *Learn Publish* 2019; 32: 63–69. <https://doi.org/10.1002/leap.1224>
39. Fanelli D. Is science really facing a reproducibility crisis, and do we need it to? *Proceed Natl Acad Sci (PNAS)* 2018; 115: 2628–2631. <https://doi.org/10.1073/pnas.1708272114>
40. Leonelli S. Rethinking reproducibility as a criterion for research quality. In: *Research in the History of Economic Thought and Methodology: Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*. Emerald Publishing, 2018. <https://doi.org/10.1108/S0743-41542018000036B009>
41. Leonelli S. Open science and epistemic diversity: friends or foes? *Philos Sci* 2022; 89: 991–1001. <https://doi.org/10.1017/psa.2022.45>
42. Axtell R, Axelrod R, Epstein JM, et al. Aligning simulation models: a case study and results. *Comput Math Organiz Theory* 1996; 1: 123–141. <https://doi.org/10.1007/BF01299065>
43. Schwab A, Aguinis H, Bamberger P, et al. How replication studies can improve doctoral student education. *J Manag Sci Rep* 2023; 1: 18–41. <https://doi.org/10.1177/27550311231156880>
44. Robinson S, Nance RE, Paul RJ, et al. Simulation model reuse: definitions, benefits and obstacles. *Simul Model Pract Theory* 2004; 12: 479–494. <https://doi.org/10.1016/j.simpat.2003.11.006>
45. Wilkinson M, Dumontier M, Aalbersberg IJ, et al. The fair guiding principles for scientific data management and stewardship. *Scientific Data* 2016; 3. <https://doi.org/10.1038/sdata.2016.18>
46. Monks T, Harper A and Mustafee N. Towards sharing tools and artefacts for reusable simulations in healthcare. *J Simul* 2024; 19: 619–638. <https://doi.org/10.1080/17477778.2024.2347882>
47. Lohmann A, Astivia OLO, Morris TP, et al. It’s time! Ten reasons to start replicating simulation studies. *Front Epidemiol* 2022; 2: 973470. <https://doi.org/10.3389/fepep.2022.973470>
48. Huang Y and Nikolic I. Towards a multi-model infrastructure for integrated decision-making in energy transition. In: *International Multidisciplinary Modeling & Simulation Multiconference, I3M 2024, Fes, Morocco, 17–19 September 2025*. <https://doi.org/10.46354/i3m.2024.mas.002>
49. Brailsford SC, Eldabi T, Kunc M, et al. Hybrid simulation modelling in operational research: a state-of-the-art review. *European Journal of Operational Research* 2019; 278: 721–737. <https://doi.org/10.1016/j.ejor.2018.10.025>
50. Balkan BA, Sparnaaij M, Duives D, et al. Infection risk and economic activity trade-offs: decision-making in indoor venue operations for pandemic preparedness. *J Simul* 2025; 2025: 1–25. <https://doi.org/10.1080/17477778.2025.2546481>
51. Heyard R, Pawel S, Frese J, et al. A scoping review on metrics to quantify reproducibility: a multitude of questions leads to a multitude of metrics. *R Soc Open Sci* 2025; 12: 242076.
52. Muradchianian J, Hoekstra R, Kiers H, et al. How best to quantify replication success? A simulation study on the comparison of replication success metrics. *Royal Soc Open Sci* 2021; 8: 201697. <https://doi.org/10.1098/rsos.201697>
53. Crick T, Hall B and Ishtiaq S. Reproducibility in research: Systems, infrastructure, culture. *J Open Res Softw* 2017; 5: 73. <https://doi.org/10.5334/jors.73>
54. Tiwari K, Kananathan S, Roberts MG, et al. Reproducibility in systems biology modelling. *Mol Syst Biol* 2021; 17: e9982. <https://doi.org/10.15252/msb.20209982>
55. Troost C, Huber R, Bell AR, et al. How to keep it adequate: a protocol for ensuring validity in agent-based simulation. *Environ Model Softw* 2023; 159: 105559. <https://doi.org/10.1016/j.envsoft.2022.105559>
56. Monks T, Currie CSM, Onggo BS, et al. Strengthening the reporting of empirical simulation studies: introducing the stress guidelines. *J Simul* 2019; 13: 55–67. <https://doi.org/10.1080/17477778.2018.1442155>
57. Grimm V, Railsback SF, Vincenot CE, et al. The odd protocol for describing agent-based and other simulation models: a second update to improve clarity, replication, and structural realism. *J Artif Societ Social Simul* 2020; 23: 7. <https://doi.org/10.18564/jasss.4259>
58. Schaduagrath N, Lampa S, Simeon S, et al. Towards reproducible computational drug discovery. *J Cheminform* 2020; 12: 9. <https://doi.org/10.1186/s13321-020-0408-x>
59. Brunak S, Collin CB, Cathaoir KE, et al. Towards standardization guidelines for in silico approaches in personalized medicine. *J Integrat Bioinform* 2020; 17: 20200006. <https://doi.org/doi:10.1515/jib-2020-0006>
60. Le Novère N, Finney A, Hucka M, et al. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotech* 2005; 23: 1509–1515. <https://doi.org/10.1038/nbt1156>
61. Juty N, le Novère N and Laibe C. Miriam guidelines. In: Dubitzky W, Wolkenhauer O, Cho KH, et al. (eds.) *Encyclopedia of systems biology*. New York: Springer, 2013, pp.1367–1367. https://doi.org/10.1007/978-1-4419-9863-7_1176
62. Waltemath D, Adams R, Beard D, et al. Minimum information about a simulation experiment (MIASE). *PLoS Computational Biology* 2011; 7: 1001122. <https://doi.org/10.1371/journal.pcbi.1001122>
63. Smith LP, Bergmann FT, Garny A, et al. The simulation experiment description markup language (SED-ML): language specification for level 1 version 5. *J Integrat Bioinform* 2024; 21: 20240008. <https://doi.org/doi:10.1515/jib-2024-0008>
64. Golebiewski M, Bader G, Gleeson P, et al. Specifications of standards in systems and synthetic biology: status, developments, and tools in 2024. *J Integrat Bioinform* 2024; 21: 0015. <https://doi.org/10.1515/jib-2024-0015>
65. Viceconti M and Emili L (eds) *Toward Good Simulation Practice: Best Practices for the use of Computational Modelling and Simulation in the Regulatory Process of Biomedical Products. Synthesis Lectures on Biomedical Engineering (SLBE)*. Springer, 2024.

66. Stamatakos GS, Perez MA and Radhakrishnan R. Editorial: Multiscale cancer modeling, in silico oncology and digital (virtual) twins in the cancer domain. *Front Physiol* 2025; 16: 1614235. <https://doi.org/10.3389/fphys.2025.1614235>
67. Knapp B, Bardenet R, Bernabeu MO, et al. Ten simple rules for a successful cross-disciplinary collaboration. *PLoS Comput Biol* 2015; 11: e1004214.
68. Ziemann M, Poulain P and Bora A. The five pillars of computational reproducibility: bioinformatics and beyond. *Brief Bioinform* 2023; 24: bbad375. <https://doi.org/10.1093/bib/bbad375>
69. Zhu Z, Chen M, Ren G, et al. A framework for assessing the computational reproducibility of geo-simulation experiments. *Environ Model Softw* 2025; 186: 106323. <https://doi.org/10.1016/j.envsoft.2025.106323>
70. Heather A, Monks T, Harper A, et al. On the reproducibility of discrete-event simulation studies in health research: an empirical study using open models. *J Simul* 2025; 0: 1–25. <https://doi.org/10.1080/17477778.2025.2552177>
71. McDougal R, Bulanova A and Lytton W. Reproducibility in computational neuroscience models and simulations. *IEEE Trans Biomed Eng* 2016; 63: 2021–2035. <https://doi.org/10.1109/tbme.2016.2539602>
72. Samota EK and Davey RP. Knowledge and attitudes among life scientists toward reproducibility within journal articles: a research survey. *Front Res Metric Anal* 2021; 6: 678554. <https://doi.org/10.3389/frma.2021.678554>
73. Huang Y, Poderi G, Šćepanović S, et al. Embedding internet-of-things in large-scale socio-technical systems: a community-oriented design in future smart grids. In: Cicirelli F, Guerrieri A, Mastroianni C, et al. (eds.) *The Internet of Things for Smart Urban Ecosystems*. Springer, 2019, pp.125–150. https://doi.org/10.1007/978-3-319-96550-5_6
74. Shrestha SL, Chowdhury SA and Csallner C. Replicability study: Corpora for understanding simulink models & projects. In: *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, New Orleans, LA, 26–27 October 2023, pp.1–12. <https://doi.org/10.1109/ESEM56168.2023.10304867>
75. Wagenaar J, Dietz A, Huang Y, et al. Impact assessment of neonatal care interventions on regional neonatal care capacity: a simulation study based on clinical data in the Netherlands. *BMJ Open* 2025; 15: e104688. <https://doi.org/10.1136/bmjopen-2025-104688>
76. Sundaram A, Huang Y, Cuppen E, et al. Operationalizing justice in models used as decision-support tools in local and regional energy transition planning. In: *12th International Workshop on Simulation for Energy, Sustainable Development & Environment: SESDE*, Tenerife, Spain, 18–20 September 2024. <https://doi.org/10.46354/i3m.2024.sesde.002>
77. Niaz Arifin SM, Davis GJ and Zhou Y. Verification & validation by docking: a case study of agent-based models of anopheles gambiae. In: *Proceedings of the 2010 Summer Computer Simulation Conference, SCSC '10*, San Diego, CA, 11–14 July, pp.236–243. Society for Computer Simulation International.
78. Begley C, Buchan A and Dirnagl U. Robust research: Institutions must do their part for reproducibility. *Nature* 2015; 525: 25–27. <https://doi.org/10.1038/525025a>
79. Munafò M, Nosek B, Bishop D, et al. A manifesto for reproducible science. *Nature Human Behav* 2017; 1: 0021. <https://doi.org/10.1038/s41562-016-0021>
80. Navarro J, Deruyver A and Parrend P. A systematic survey on multi-step attack detection. *Comput Secur* 2018; 76: 214–249. <https://doi.org/10.1016/j.cose.2018.03.001>
81. Williams C, Yang Y, Lagisz M, et al. Transparent reporting items for simulation studies evaluating statistical methods: foundations for reproducibility and reliability. *Method Ecol Evolut* 2024; 15: 1926–1939. <https://doi.org/10.1111/2041-210X.14415>
82. Blinov ML, Gennari JH, Karr JR, et al. Practical resources for enhancing the reproducibility of mechanistic modeling in systems biology. *Curr Opin Syst Biol* 2021; 27: 100350. <https://doi.org/10.1016/j.coisb.2021.06.001>
83. Monks T and Harper A. Improving the usability of open health service delivery simulation models using python and web apps. *NIHR Open Research* 2023; 3: 48. <https://doi.org/10.3310/nihropenres.13467.1>
84. Zschaler S, Mustafee N, Harper A, et al. On simulation reuse in healthcare applications. *SIMULATION* 2025; 2025: 00375497251383912. <https://doi.org/10.1177/00375497251383912>
85. Monks T, Harper A and Heather A. Unlocking the potential of past research: using generative AI to reconstruct healthcare simulation models. *J Oper Res Soc* 2025; 2025: 1–24. <https://doi.org/10.1080/01605682.2025.2554751>
86. Gao C, Lan X, Li N, et al. Large language models empowered agent-based modeling and simulation: a survey and perspectives. *Human Soc Sci Commun* 2024; 11: 1259. <https://doi.org/10.1057/s41599-024-03611-3>
87. Giabbanelli PJ. Gpt-based models meet simulation: How to efficiently use large-scale pre-trained language models across simulation tasks. In: *Proceedings of the Winter Simulation Conference (WSC) WSC'23*, San Antonio, TX, 10–13 December 2023, pp.2920–2931. IEEE Press. <https://doi.org/10.1109/WSC60868.2023.10408017>
88. Jackson I, Saenz MJ and Ivanov D. From natural language to simulations: applying AI to automate simulation modelling of logistics systems. *Int J Prod Res* 2024; 62: 1434–1457. <https://doi.org/10.1080/00207543.2023.2276811>
89. Alshareef A, Keller N, Carbo P, et al. Generative AI with modeling and simulation of activity and flow-based diagrams. In: Guisado-lizar JL, Riscos-núñez A, Morón-fernández MJ, et al. (eds.) *Simulation Tools and Techniques*. Springer, pp.95–109.
90. Ghaffarzagdegan N, Majumdar A, Williams R, et al. Generative agent-based modeling: an introduction and tutorial. *Syst Dynam Review* 2024; 40: e1761. <https://doi.org/10.1002/sdr.1761>
91. Ledet JW, Teran-Somohano A, Butcher Z, et al. Toward model-driven engineering principles and practices for model replicability and experiment reproducibility. In: *DEVS '14: Proceedings of the Symposium on Theory of Modeling & Simulation – DEVS Integrative*, Tampa, FL, 13–16 April 2014.

92. Çetinkaya D, Verbraeck A and Seck MD. Model continuity in discrete event simulation: a framework for model-driven development of simulation models. *ACM Trans Model Comput Simul* 2015; 25: 2699714. <https://doi.org/10.1145/2699714>
93. Preskill J. Quantum computing in the NISQ era and beyond. *Quantum* 2018; 2: 79. <https://doi.org/10.22331/q-2018-08-06-79>
94. Nielsen MA and Chuang IL. *Quantum computation and quantum information*. 10th ed. Cambridge University Press, 2010.
95. Aaronson S. *Quantum computing since democritus*. Cambridge University Press, 2013.
96. Cetinkaya D and Abdelkader A. A review of quantum modeling and simulation approaches for lithium-ion batteries. In: *2025 Annual Modeling and Simulation Conference (ANNSIM)*, Madrid, Spain, 26–29 May 2025, pp.1–12.
97. Biamonte J, Wittek P, Pancotti N, et al. Quantum machine learning. *Nature* 2017; 549: 195–202. <https://doi.org/10.1038/nature23474>
98. Peral-García D, Cruz-Benito J and García-Pen~alvo FJ. Systematic literature review: quantum machine learning and its applications. *Comput Sci Rev* 2024; 51: 100619. <https://doi.org/10.1016/j.cosrev.2024.100619>
99. Pineda VG, Valencia-Arias A, Giraldo FEL, et al. Integrating artificial intelligence and quantum computing: a systematic literature review of features and applications. *Int J Cognit Comput Eng* 2026; 7: 26–39. <https://doi.org/10.1016/j.ijcce.2025.08.002>
100. De Domenico M, Allegri L, Caldarelli G, et al. Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. *NPJ Digital Med* 2025; 8: 37. <https://doi.org/10.1038/s41746-024-01402-3>
101. NASA. Why does the world (and NASA) need digital twins? 2025, <https://science.nasa.gov/biological-physical/why-does-the-world-and-nasa-need-digital-twins/>
102. Reimeir B, Leininger A, Edlinger R, et al. Digital twin for analog mars missions: investigating local positioning alternatives for GNSS-Denied environments. *Sensors (Basel)* 2025; 25: 4615. <https://doi.org/10.3390/s25154615>
103. CERN. Digital twins at CERN and beyond, 2023, <https://home.cern/news/news/knowledge-sharing/digital-twins-cern-and-beyond>
104. Hazeleger W, Aerts JPM, Bauer P, et al. Digital twins of the earth with and for humans. *Commun Earth Environ* 2024; 5: 463. <https://doi.org/10.1038/s43247-024-01626-x>
105. NASEM. *Foundational research gaps and future directions for digital twins*. National Academies of Sciences, Engineering, and Medicine, The National Academies Press, 2024. <https://doi.org/10.17226/26894>
106. Uhrmacher AM, Brailsford S, Liu J, et al. Panel – reproducible research in discrete event simulation – a must or rather a maybe? *2016 Winter Simulation Conference (WSC)*, Washington, DC, 11–14 December 2016, pp.1301–1315.
107. Rossi L and Vandin A. Reproducibility report for the paper: efficient non-blocking event management for speculative parallel discrete event simulation. In: *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. SIGSIM-PADS '24*, Atlanta, GA, 24–26 June 2024, pp.131–132, Association for Computing Machinery. <https://doi.org/10.1145/3615979.3665107>
108. Piccione A and Pellegrini A. Efficient non-blocking event management for speculative parallel discrete event simulation. In: *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. SIGSIM-PADS '24*, Atlanta, GA, 24–26 June 2024, pp.52–56. Association for Computing Machinery. <https://doi.org/10.1145/3615979.3656053>
109. ACM. Artifact review and badging policy. *Association for Computing Machinery Repository*, 2020, <https://www.acm.org/publications/policies/artifact-review-and-badging-current>
110. Kohrs FE, Auer S, Bannach-Brown A, et al. Eleven strategies for making reproducible research and open science training the norm at research institutions. *Elife* 2023; 12: e89736. <https://doi.org/10.7554/eLife.89736>
111. Rougier NP and Hinsenk K. Rescience C: a journal for reproducible replications in computational science. In: Kerautret B, Colom M, Lopresti D, et al. (eds) *Reproducible research in pattern recognition*. Cham: Springer, 2019, pp.150–156. https://doi.org/10.1007/978-3-030-23987-9_14

Author biographies

Yilin Huang is an assistant professor at the Faculty of Technology, Policy and Management at Delft University of Technology in the Netherlands. She received her PhD in Automated Simulation Model Generation from the same university. Her research focuses on modelling and simulation theory and methodology, with particular interests in data-driven methods, model conceptualisation, and the reproducibility, reusability, and interoperability of simulation models. She has worked on national and international simulation projects involving large-scale, complex socio-technical systems in various application domains, including transportation, logistics, smart energy systems, health care, and sustainability transitions.

Deniz Cetinkaya is a principal academic in Computing at Bournemouth University. She graduated from the Department of Computer Science and Engineering at Hacettepe University with honours in 2002. She received her PhD degree in Systems Engineering from the Delft University of Technology (Technische Universiteit Delft) in the Netherlands in 2013, and her MSc degree in Computer Engineering from the Middle East Technical University in 2005. Her research focuses on software engineering practice and methodology, model-driven engineering, quantum computing, modelling and simulation, and other related areas. She is a senior ACM member. Her email address is dcetinkaya@bournemouth.ac.uk.