Journal of Simulation

ISSN: 1747-7778 (Print) 1747-7786 (Online) Journal homepage: http://www.tandfonline.com/loi/tjsm20

Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines

Thomas Monks, Christine S. M. Currie, Bhakti Stephan Onggo, Stewart Robinson, Martin Kunc & Simon J. E. Taylor

To cite this article: Thomas Monks, Christine S. M. Currie, Bhakti Stephan Onggo, Stewart Robinson, Martin Kunc & Simon J. E. Taylor (2018): Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines, Journal of Simulation, DOI: 10.1080/17477778.2018.1442155

To link to this article: https://doi.org/10.1080/17477778.2018.1442155

JOURNAL OF SIMULATION, 2018
https://doi.org/10.1080/17477778.2018.1442155

THE OPERATIONAL RESEARCH SOCIETY
Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines

Thomas Monks[a], Christine S. M. Currie[b], Bhakti Stephan Onggo[c], Stewart Robinson[d], Martin Kunc[e] and Simon J. E. Taylor[f]

[a]NIHR CLAHRC Wessex, Faculty of Health Sciences, University of Southampton, Southampton, UK; [b]Mathematical Sciences, University of Southampton, Southampton, UK; [C]Trinity Business School, Trinity College Dublin, Dublin, Ireland; [d]School of Business and Economics, Loughborough University, Loughborough, UK; [e]Warwick Business School, University of Warwick, Coventry, UK; [f]Department of Computer Science, Brunel University, London, UK

ABSTRACT

This study develops a standardised checklist approach to improve the reporting of discrete-event simulation, system dynamics and agent-based simulation models within the field of Operational Research and Management Science. Incomplete or ambiguous reporting means that many simulation studies are not reproducible, leaving other modellers with an incomplete picture of what has been done and unable to judge the reliability of the results. Crucially, unclear reporting makes it difficult to reproduce or reuse findings. In this paper, we review the evidence on the quality of model reporting and consolidate previous work. We derive general good practice principles and three 20-item checklists aimed at Strengthening The Reporting of Empirical Simulation Studies (STRESS): STRESS-DES, STRESS-ABS and STRESS-SD for discrete-event simulation, agent-based simulation and system dynamics, respectively. Given the variety of simulation projects, we provide usage and troubleshooting advice to cover a wide range of situations.

ARTICLE HISTORY
Received 12 December 2016
Revised 24 October 2017
Accepted 14 February 2018

KEYWORDS
Simulation; reporting; reproducibility; discrete-event simulation; agent-based simulation; system dynamics

1. Introduction

The *reproducibility of* research findings from a study is at the centre of science. The simulation and the wider Operational Research and Management Science (ORMS) communities publish models and methods in order to advance knowledge and avoid reinventing the wheel. This issue is also of importance in industry, where models are built and maintained by a single person or a team of people and where studies using those models may need to be audited or repeated. Several authors have looked at the reproducibility of models within ORMS and found published peer-reviewed reports of models can be ambiguous, incomplete and hence difficult to reuse and extend (Boylan, Goodwin, Mohammadipour, & Syntetos, 2015; Dalle, 2012; Grimm et al., 2006; Kendall et al., 2016; Kurkowski, Camp, & Colagrosso, 2005; Rahmandad & Sterman, 2012). This is not unique to ORMS. In other model-based and empirical disciplines, there has been increasing calls to create guidelines to support authors in complete reporting of their models to maximise reproducibility (Grimm et al., 2006; Waltemath et al., 2011). However, there are still gaps in ORMS literature related to guidelines for reporting models. This article presents guidelines to support the reporting of models within Agent-Based Simulation (ABS), Discrete-Event Simulation (DES) and System Dynamics (SD). These three methods represent the most popular simulation methods within ORMS (Jahangirian, Eldabi, Naseer, Stergioulas, & Young, 2010). We describe these guidelines as the STRESS test (Strengthening the Reporting of Empirical Simulation Studies: STRESS). If followed, the STRESS guidelines provide authors with a way to maximise the chances of other researchers or practitioners reusing their work to either extend results or benefit society and give readers the ability to better judge the contribution of simulation studies. While the guidelines are focused on simulation models in ORMS, the principles they are based on could be applied to other modelling techniques.

The article is structured as follows. First, we review the reasons for publishing simulation studies to establish why reproducibility of models and results is critically important to simulation-based research. We follow this by reviewing the evidence examining the reproducibility of ORMS models, to illustrate the difficulties in reporting simulation models. To develop the STRESS guidelines, we review existing guidelines from other model-based

CONTACT Thomas Monks thomas.monks@soton.ac.uk
Supplemental data for this article can be accessed at https://doi.org/10.1080/17477778.2018.1442155.

disciplines; the complementary concept of design patterns for simulation development; and good practice papers across the three simulation fields. We then present an overview of the STRESS guidelines (complete checklists can be found in Supplementary material). Given the wide range of simulation studies that are carried out, we provide a detailed troubleshooting section on the practical use of the guidelines for reporting. Finally, we discuss the benefits of the approach and further work that could aid reporting.

2. Why do we publish simulation studies?

Research studies are published with the aim of extending existing knowledge and offering researchers and practitioners the opportunity to reuse and build upon others' work. When there is a paucity of detail, this becomes impossible and poorly reported simulation studies cannot be reproduced, extended or reused. It has been suggested that for models and results to be reproducible "modellers should be able to recreate the base case results of a simulation model and any simulation experiments even when using a different platform and software" (Rahmandad & Sterman, 2012). We repeat this argument, but acknowledge the difficulty in reproducing the exact results of a stochastic simulation across platforms and software. Within modern applications of computer simulation, study results should be reproducible on the basis of experimental lab conditions i.e., the model, software, code libraries and computer system specification need to be precisely reported. Within the context of simulation in ORMS, we argue that reproducibility has the following scientific, societal and practical benefits:

2.1. The advancement of operational knowledge

If a simulation model and its results can be reproduced, the model can be reused to investigate further hypotheses in the same application area or to test the generalisability of an efficient approach to managing operations in another context. For example, a group of authors may have developed an approach to increase the throughput of a car assembly line in a local factory and wish to test if their approach works in other lines, other factory operations, or even a completely different application.

2.2. To enable reuse of knowledge

It is well documented that the development of simulation models from scratch is expensive. Model reuse takes several forms (Robinson, Nance, Paul, Pidd, & Taylor, 2004). At one level this might be the reproduction and reuse of the full model to tackle a similar problem. Hospital accident and emergency models are often quoted as an area where such reuse might be possible (Fletcher, Halsall, Huxham, & Worthington, 2007; Fletcher & Worthington, 2009). At a lower level, it might be that smaller components within a model could be reproduced and reused within another model with a different purpose. For example, the chain of stocks and flows portraying the dynamics of workforce and its impact on satisfying demand has been widely employed in SD models related to workforce planning for different organisations and industries, (e.g., Brailsford & De Silva, 2015; Kunc, 2008).

2.3. To further conceptual modelling knowledge

Conceptual modelling is the process of deciding what to model and what not to model (Robinson, 2008). Often there are multiple levels of detail that could be employed to represent components within a model, or alternative ways to conceptualise a component. When reported accurately this offers an important resource to other researchers, with a range of modelling experience, who might be tackling similar problems.

2.4. To reuse data where none exists

In many applied problems, data are limited or missing. A report of a simulation model should include full details of distributions for a DES model, or constants and table functions within an SD model. Making such data publicly available will allow future modelling studies to use these values, improving the validity of this later work.

2.5. Testing of novel simulation methods

The validation of new output analysis methods, computational procedures or simulation optimisation methods requires full details of the method and the simulation case study to be reported in order to enable the reader to assess the quality of the proposed new method.

3. Is there a problem with reporting?

There is mounting evidence that the reproducibility of complex science is questionable. A recent survey in Nature suggests that around 50% of scientists believe there is a substantial reproducibility "crisis" (Baker, 2016). Within our own ORMS paradigm, several authors have investigated the reproducibility of research. Rahmandad and Sterman (2012), Boylan et al. (2015), Janssen (2017) and Kurkowski et al. (2005) investigate the reproducibility considering SD, DES, ABS and forecasting. Other authors have considered the issues of reproducibility with simulation models more generally (Dalle, 2012; Grimm et al., 2006).

Rahmandad and Sterman (2012) sampled a year's worth of articles from the academic journal System Dynamics Review. There were 27 papers that reported an SD model and scientific result. Out of 27 models, 16 (59%) included no equations at all while 2 (7%) reported

"some" equations. The set of equations that define the flow rates between stocks are pivotal to a quantitative SD model. Without these equations the model cannot be reproduced. If we consider another basic tenet of reproducibility – data – only 8 (30%) included the parameter values to reproduce the base case results. This result has to be contrasted with the initial reporting of SD models performed in Forrester's World Dynamics (Forrester, 1971), where the full model including parameters, equations and model logic are available.

We found no overarching review of DES reporting, but Kurkowski et al. (2005) review 114 DES models of Mobile Ad Hoc Networks (MANETS). They summarise the "common pitfalls" they find in the reporting of these models and conclude that the majority of studies are not reported completely and hence cannot be reproduced by other researchers. Some key findings were that 58% of the studies did not specify if a model was terminating or steady state; no studies detailed the pseudo random number generator; 93% of the studies did not include any comment on the need to deal with initialisation bias and the 7% that did failed to provide any documentation about the analysis procedure used to select a warm-up period; finally, 25% of the studies did not state the simulation software in which the model was implemented.

Boylan et al. (2015) investigate the reproducibility of forecasting models in a novel practical way. Two experienced teams of forecasters were tasked to reproduce the results of a famous forecasting paper (Miller & Williams, 2003). The teams were able to reproduce each other's results but not those of the paper. The authors of the original study were asked to clarify aspects of their paper, and responded positively. However, the teams still could not reproduce the results. The authors conclude that there is considerable scope for the improvement of the reporting of forecasting results and that it is uncommon for reviewers or editors to request sufficient details to reproduce results.

Janssen (2017) investigated the reproducibility of 2367 agent-based models returned from a search of ISI Web of Science. The study found that 50% of publications report complete or "some" equations. The authors' particular interest was in the provision of publically available model source code. Findings were that source code for the models was only available for 10% of the publications, although this appears to be slowly increasing. The authors note that the lack of transparency in how models work is slowing down knowledge creation and leads to duplication of effort in research.

At a more general level, Dalle (2012) outlines the case for reproducibility and the issues in achieving it within simulation. Insufficiently detailed publications are listed as a major obstacle. One facet of this is that there is "very little incentive to provide reproducible content"; journals do not ask reviewers to check for reproducibility and such a check is burdensome for reviewers. We note that no evidence is provided to support the author's propositions, but that it does fall in line with the comments of other authors from different simulation fields.

4. Developing the guidelines

To develop the guidelines we modified the approach of Moher, Schulz, Simera, and Altman (2010). Their approach is focused on healthcare but is sufficiently generic to be useful more widely. We adopted a pragmatic twofold approach. First we conducted literature searches to identify good practice articles from within ORMS and other model-based disciplines, as well as existing reporting guidelines for model-based research and empirical science. Existing guidelines are summarised in Table 1. The two lead authors (TM and CC) then converted the findings to an initial version of the guidelines. The second phase involved presenting the initial version of the guidelines to experts within the field at the 2016 OR Society Conference. A revised draft incorporating feedback from the conference was reviewed by four experts in DES, ABM, SD and large-scale simulation methods (co-authors: BSO, SR, MK and ST) who provided a detailed critique and revision.

4.1. Existing guidelines from model-based disciplines

Gass (1984) provides the earliest example of reporting guidelines for "computer based models". Although the documentation described is extensive, it offers readers little in the way of advice about the minimal elements that are necessary for reproducibility. The age of the guidelines also means that they lack the specialisation needed to report modern ABS, DES and SD models.

Rahmandad and Sterman (2012) develop the Minimum Model Reporting Requirements (MMRR); we note that these are published within a System Dynamics specialist journal and are most applicable to SD models. Guidelines are broken down into four areas: general visualisation, reporting of models (logic and algorithms), reporting of experiments and reporting of optimisation results. Each area is further broken down into a minimum and preferred level of reporting. Guidelines represent a "starting point" and "need to be updated on feedback from the community of researchers that use them". One weakness of the work is the authors do not include a simple checklist that authors and reviewers can follow.

In the context of Ecology, Grimm et al. (2006) (with minor updates in Grimm et al., 2010), propose a very structured protocol for documenting individual-based and agent-based models advocating a fixed structure for reporting. The authors break their protocol into sections on overview (i.e., the purpose of the model and general logic), design concepts (e.g., emergence and stochastic behaviour) and details (e.g., initialisation, data and agent

Table 1. Key model-based reporting guidelines published in peer-reviewed journals.

First author	Year	Guidelines	Paradigm	Key points
Gass	1984	GASS	Computer Simulation	• Splits documentation in four categories. Analyst, User, Programmer and Manager • Lack specificity for modern simulation approaches such as ABS
Grimm	2006/2010	ODD	Individual (agent)-based models	• Advocates a standard format for reporting • Highly relevant to modellers in ABS, but limited applicability elsewhere
Waltemath	2011	MIASE	Biological process simulation	• Simple, high-level and generalisable across simulation approaches. • Lack of detail means that authors and reviewers have limited specific guidance.
Rahmandad	2012	MMRR	System dynamics	• Minimum and preferred reporting guidance • Aimed to be general, but most applicable to SD • Lacks a simple checklist approach to reporting
Husereau	2013	CHEERS	Health economic evaluation	• Checklist approach • While some items are relevant the guidelines are specific to health economic modelling
Kendall	2016	GLP4OPT	Optimisation	• Checklist approach • Limited relevancy to the simulation community

attributes). The guidelines are named ODD (Overview, Design Concepts and Details). We note that the Grimm et al. (2006) guidelines are highly referenced within ecology. However, a recent study found that only 7% of 2367 agent-based model papers found in ISI Web of Science used the protocol (Janssen, 2017). No checklist is provided; however, muliple examples are provided in Supplementary material that are useful for illustration.

Waltemath et al. (2011) developed the Minimum Information about Simulation Experiment (MIASE) guidelines for biological process simulation. Here, models are created and simulated as testable hypotheses in order to determine whether or not they are compatible with experimental data or expected future observations. The MIASE guidelines are split across three areas: rules for documenting the model, rules to describe the simulation experiment and rules for dealing with model output. The MIASE guidelines are noticeably less detailed than other model-based guidelines we reviewed, such as Grimm et al. (2006) or Rahmandad and Sterman (2012). The advantage is that MIASE is quite general across simulation approaches; while the downside is that it does take more effort to adapt and apply them and it is more difficult to quickly assess that the guidelines have been followed.

Kendall et al. (2016) provide extensive reporting guidelines for optimisation research, making 54 recommendations in total. In general terms, the guidelines provide some advice to simulation modellers but the work shows some clear differences in focus between simulation and optimisation studies. The guidelines presented by Kendall will be very useful for simulation optimisation studies, where different algorithms are being compared and Pasupathy and Henderson (2006) complements this paper by introducing a testbed of simulation optimisation problems. This is beyond the scope of STRESS.

Husereau et al. (2013) synthesise 10 reporting guidelines for health economic evaluation and create a "user-friendly" checklist of 24 items. Given the context, the guidelines follow health economic terminology potentially unfamiliar to a more general simulation modelling community. Some items are health economic specific (e.g., discount rate and health outcomes); however, there are several that are transferable across disciplines. For example, comparators are equivalent to scenarios within simulation and details of input parameters are relevant across all modelling disciplines.

4.2. Good practice reporting papers

To illustrate good practice for reporting model logic in DES readers are directed to Günal and Pidd (2011) and the District General Hospital Performance Simulator (DGHPSim). The model is a complex, generic representation of a hospital (in the United Kingdom) and is split into sub-models that represent an emergency department, outpatient services, waiting list services and bed management. The four model "components" can be used separately (most notably the emergency department model in Gunal and Pidd (2009)) or combined to investigate the performance of a whole hospital. The workings of these individual models are reported in a manner that facilitates reuse. For example, the authors provide a high level overview diagram of how the models work together (as advocated by several of the existing guidelines for model-based research) along with more detailed diagrams of the sub-models. The authors describe the conceptualisation of each model, the level of detail included and the implementation as a computer simulation (for example which elements of the system are entities, activities and resources). The use of the

model is illustrated by the results of several scenarios. The only weakness of the report is that no "test" data are provided to allow researchers to recreate the results presented and verify that a model is working as expected.

For SD, readers are directed to Pierson and Sterman (2013) who report a model that explains the dynamics of airline earnings. The authors provide high and low level descriptions, using diagrams and text to explain the model. This includes details of all stock and flows, equations, simulation experiments, pre-processing of data and parameter values. We note that given the complexity of the model the authors make prudent use of the journal's online (peer-reviewed) supporting material policy. Their approach provides a good balance between keeping the main article at a reasonable length and the rigour needed to recreate the model independently. The authors developed the model in Vensim and, given System Dynamics Review policy, include the simulation model itself as supplementary material. The use of Vensim also allowed the authors to make use of SDM-Doc (Martinez-Moyano, 2012) a tool designed to automatically document the variables within a Vensim model.

For good practice in reporting ABS models, readers are directed to Yates, Ford, and Kuglics (2014) who report a detailed model in civil violence with Iran used as a test case. The paper illustrates one of the key benefits of unambiguous and complete reporting of models. That is, the research reuses and extends Epstein's (2006) model of civil violence. The authors cite the original work and describe the purpose and utility of their extension. This is followed by an overview of the logic of the original model such as the environment, agent states, state transitions and interactions between agents; readers are able to refer to the original work for more details. The main report details the extensions to the original model; for example, modifying the grid-based region with a continuous geographic region and transport network. The authors document model parameters and describe the model dynamics including equations where used and which elements of the model are stochastic. For the Iranian case study, the authors detail all experimentation elements, such as model run length and the number of replications used. Experimentation aims are incorporated as a 2^k factorial experimental design along with the range of parameter values used. The reporting of the model is framed as software independent; the authors also describe the software and programming tools used to implement the model.

5. An introduction to the guidelines

Before we detail the *Strengthening the Reporting of Empirical Simulation Studies* (STRESS) guidelines we encourage authors to take note of Table 2. We conducted a thematic analysis (Braun & Clarke, 2006) of the six model-based reporting guidelines that we reviewed. The six principles for reporting that we list are a simple and effective starting point for reporting simulation studies. These principles may also be applicable to other modelling disciplines falling within ORMS.

5.1. An overview of the guidelines

The idea of the STRESS guidelines is to support high quality reporting of simulation models in order to ensure that a model and its results are reproducible. The STRESS guidelines are split into six sections: objectives, model logic, data, experimentation, implementation and code access; STRESS includes 20 checklist items (Table 3). There are three specific instances of STRESS: STRESS-ABS, STRESS-DES and STRESS-SD, respectively. The three checklists can be found in Supplementary material. Here, we provide an overview of the general structure of the checklists and the key differences between the three versions.

5.2. Section 1: Objectives

This section contains three items that report clearly what the study is aiming to achieve. The first of these is the *purpose* and rationale for the project and includes the model's intended use or experimental frame (Pidd, 2006, p. 36). This helps other researchers and modellers understand the choices made in conceptualising the model. The second is the *model outputs* that the model will predict. The third item reports the *aims of experimentation* which provide more specific information about how the model is being used to achieve the stated purpose. For example, in modelling a simple queuing system such as a small shop, the purpose of the model may be to find the optimal number of servers to ensure good service; the model outputs might be average waiting time for service, the average utilisation of the servers and the cost of the system, while the aims of the experimentation would be to provide details of the input parameters that can be changed such as the number of servers or the structure of the queues and the objectives. In this case, there may be more than one objective, with the experiment finding a good trade-off between customer satisfaction (i.e., time in the queue) and the cost of the system. The remaining items should be followed with these objectives in mind.

5.3. Section 2: Logic

This section contains checklist items that ensure the logic of the base model and any differences in the logic of models implemented in different scenarios are reported clearly. This is the section where the checklists deviate most between the reporting of ABS, DES and SD. It is the most detailed subsection and includes five checklist items split between descriptions of the base model logic and the logic used in other scenarios. STRESS

recommends the use of a recognised simulation diagramming approach as an aid to communicate model design. Within the main text authors should limit diagrams to conceptual or simplified overviews but complex diagrams used to communicate complete model design should be included as Supplementary material.

The greatest differences across the three checklists are found in the model *components* section. Table 4 illustrates this difference across the three checklists. Components refer to the basic conceptual building blocks of the model. Hence, in the DES case, STRESS focuses on entities, activities, resources and queues, while for ABS models, STRESS focuses on the environment, agents, topology and interaction. In STRESS-SD, the focus is on stocks, flows and feedback loops. Authors are referred to the good practice papers in section 4.2 for detailed examples.

5.4. Section 3: Data

A model and its results cannot be reproduced without detailing the input parameters. The recommendations include listing details of data sources, input parameters for base runs of the model and scenario experiments, data pre-processing and assumptions. We illustrate the reporting of stochastic parameters in Table 5; deterministic parameters should be reported in a similar manner and readers are referred to Kunc and Kazakov (2013) for an example. The recommendations for reporting model data are common across the three modelling disciplines. We expect that, in most cases, following these recommendations will be unproblematic. However, there may be instances of modelling research where data are confidential or there are commercial reasons why data cannot be published. In these instances, reports should include hypothetical non-proprietary data so that researchers can still verify that a model has been reproduced accurately. Another factor that authors may have legitimate concerns about is the ethics of publishing data. In these circumstances, we encourage academic authors to consult their institution's research governance and ethics infrastructure and industry practitioners to consult their organisational data governance and data sharing agreements (ideally before collecting and using the data).

Table 2. The six principles of reporting simulation studies.

1. State the purpose of the study and the model's intended use
2. Provide enough detail to reproduce the results of the base run of the model and any simulation experiments conducted as part of the study
3. Ensure that descriptions of the model are software and hardware *independent*
4. Include data for verification and parameter values. Where proprietary or ethical issues prevent the inclusion of data, "hypothetical" test data should be included for verification purposes
5. Document all software and where necessary hardware-specific implementation
6. Provide additional visualisation of model logic or algorithms using a recognised diagramming approach

Table 3. General format of a STRESS checklist.

Section	Item no.	Checklist item
1. Objectives	1.1	Purpose of the model
	1.2	Model outputs
	1.3	Experimentation aims
2. Logic	2.1	Base model overview diagram
	2.2	Base model logic
	2.3	Scenario logic
	2.4	Algorithms
	2.5	Components
3. Data	3.1	Data sources
	3.2	Input parameters
	3.3.	Pre-processing
	3.4	Assumptions
4. Experimentation	4.1	Initialisation
	4.2	Run length
	4.3	Estimation approach
5. Implementation	5.1	Software or programming language
	5.2	Random sampling
	5.3	Model execution
	5.4	System specification
6. Code access	6.1	Computer model sharing statement

5.5. Section 4: Experimentation

This section has three items dealing with how the model is initialised, run length and the output estimation approach used. Reporting the *initialisation* of model experiments varies across the three checklists. For example, in DES, and where appropriate for stochastic ABS models, it is recommended that warm-up periods, warm-up analysis procedures (e.g., Welch's method or MSER-5; White and Robinson (2010)) and procedures for setting initial conditions for queues and activities are reported; SD would need to detail the initial values of stocks; and ABS needs to report an initial agent population size along with attribute values and environment set-up. In the case of initial conditions, one option is to tabulate this data within a Supplementary material.

The detail recommended for the *estimation approach* for model outputs reported depends predominantly on whether the model is deterministic or stochastic. More detail is required for stochastic models, as clarity is needed about how point estimates of outputs are produced. Within a stochastic model authors should state the approach that has been used to create independent samples of the output and how many samples have been taken, e.g., the number of replications. The use of variance reduction techniques such as common random numbers or antithetic variates should also be included. Table 6 illustrates some simple approaches to report this information.

5.6. Section 5: Implementation

We emphasise that the reporting of the design should be software independent; however, the reporting of software used may help clarify specific design choices or ambiguities. The final section of the STRESS guidelines recommends that authors report the specifics of the hardware and software used. The section is comprised of

Table 4. A comparison of checklist item 2.5 "model components" for DES, ABS and SD guidelines.

STRESS-DES		STRESS-ABS		STRESS-SD	
Item	*Recommendation*	*Item*	*Recommendation*	*Item*	*Recommendation*
Entities	Give details of all entities within the simulation including a description of their role in the model and a description of all their attributes	Environment	Describe the environment agents interact within, indicating its structure, and how it is generated. For example, are agents bound within a homogeneous grid, or do they have continuous movement through a detailed landscape incorporating geographic or environmental information?	Stocks /Levels	Give details of all stocks within the simulation including a description of their role in the model. Provide the stocks units and make sure that all inflows and outflows can easily be identified (note that in a large model a diagram is unlikely to sufficiently clear for other researchers to use. Please consider tables in addition to any diagrams)
Activities	Describe the activities that entities engage in within the model. Provide details of entity routing into and out of the activity	Agents	List all agents and agent groups within the simulation. Include a description of their role in the model, their possible states, state transitions and all their attributes	Flows/Rates	List all flows within the model along with units and equations. Describe the role of flows in the model e.g., if they act as a delay
Resources	List all the resources included within the model and which activities make use of them.		Describe all decision-making rules that agents follow in either algorithmic or equation form. Where relevant authors should report:• The data that agents access (i.e., internal attributes or external information from the environment) and how it is used • The objectives agents seek to achieve • The algorithms, optimisations, heuristics and rules that agents use to achieve objectives • How agents work together within a group along with any rules for changes in group membership • Predictions of future events and adaptive action	Constants/auxiliary variables	List all variables within the model and detail their equations (if applicable) including units.
Queues	Give details of the assumed queuing discipline used in the model (e.g., First in First Out, Last in First Out, prioritisation, etc.). Where one or more queues have a different discipline from the rest, provide a list of queues, indicating the queuing discipline used for each. If reneging, baulking or jockeying occur, provide details of the rules. Detail any delays or capacity constraints on the queues			Graphical functions / lookup tables	List and detail all graphical functions within the model and describe their data sources.
Entry / Exit Points	Give details of the model boundaries i.e., all arrival and exit points of entities. Detail the arrival mechanism (e.g., "thinning" to mimic a non-homogenous Poisson process or baulking)	Interaction topology	Describe how agents and agent groupings are connected with each other in the model. Define:with whom agents can interact, how recipients of interactions are selected what frequency interaction occurs How agents handle and assign priorities to concurrent events It is recommended that interactions are described using a combination of equations pseudo-code and logic diagrams	Sources and sinks	Give details of the model boundaries i.e., all infinite sources and sinks within the model
		Entry/Exit Points	Where relevant, define how agents are created and destroyed in the model		

four items: *software*, *random sampling*, *model execution* and *system specification*.

Software here refers to the commercial or open source software, simulation or general-purpose programming language or any other form of technology used to implement the model design covered by the previous items.

Relevant only for stochastic models the reporting of the algorithm for *random sampling* is important both for judging the validity of results and also for reproducing results. In some cases, the random sampling algorithm used may be documented within the simulation software; however, authors should not assume that other researchers have access to such software or its documentation. The implementation of variance reduction techniques should also be considered. For example, in the case of common random numbers authors should describe how streams or seeds are distributed across components within the model.

Model execution refers to how simulated time progresses within the model. For example, within SD this refers to the time-step interval and integration method, within DES this refers to the event processing mechanism (e.g., Three Phase), and in ABS this refers to the time-step and/or event processing. We note that in many commercial packages the exact details of the event processing mechanisms will be ambiguous or unpublished. In these instances, it is critical that author report the software version and build numbers.

For the final item, we note that the hardware and runtime recommendations are most relevant to large-scale models that may make use of cloud, grid or high-performance computing. Table 7 illustrates a straightforward approach to reporting this information.

5.7. Section 6: Code access

The final section of STRESS recommends that authors detail whether and how the computer model can be accessed by other researchers or other modellers within an industry team. There is only a single checklist item: model code sharing statement. STRESS does not specify how authors make the computer model available nor that they must. Industry modellers may wish to list a secure or local directory. Researchers who wish to share models may wish to include a statement such as "models are available on request" or provide a link to an open science repository that hosts the model code.

Table 5. Example reporting for stochastic parameters.

Activity	Distribution	Distribution parameters	Data source (sample size)
Service time a	Gamma	$\alpha = 4.5$; $\beta = 16.5$; min = 15	Observation (n = 125)
Service time b	Log normal	$\mu = 7$; $\sigma = 4$	Blogs et al. (2004) (n = 2000)
Service time c	Triangular	Min = 3, Mode = 8, Max = 15	Expert Opinion (n = 3)

Table 6. Examples for reporting of experimentation set-up.

- The model had a run length of 180 weeks. Based on a MSER-5 analysis, a warm-up period of 60 weeks was used. No initial conditions were included. All point estimates are based on the average of 50 replications of a model run
- The model had a run length of 30 days. The environment was initialised with a fixed size agent population (n = 10,000). All agents were in the potential adopter state initially and are connected in a random network generated using Watts–Strogatz's algorithm with mean degree 5 and rewiring probability 0.25. All results are based on an average of 1000 replications

6. STRESS usage troubleshooting

While there will be many circumstances where application of the guidelines is straightforward, we expect that there is a diverse set of circumstances where users may require some guidance to troubleshoot usage. Here we pose a series of likely questions and issues as well as suggested responses for how users of STRESS should address them.

Q: I have developed a hybrid simulation model. Do the STRESS guidelines apply?

Hybrid simulation models represent a mixed simulation approach. For example, the combination of DES and SD within a single model. The STRESS guidelines are applicable in these circumstances. In our DES and SD example, it is recommended to apply both the STRESS-DES and STRESS-SD guidelines to strengthen the reporting of the appropriate model components. However, some adaptation of the guidelines is needed to handle the interface between the DES and SD components of the models. It is recommended that authors report how the two (or more) methods communicate.

Q: Writing is a creative process. Can I structure my article in my own way?

The guidelines are not prescriptive in how to structure an article. The guidelines simply specify a minimum checklist to aid reproducibility of an author's model. Authors can structure their articles however they wish. We encourage authors to make use of Supplementary material and additional files (where possible). We also ask authors to remember that reviewers tend to want a manuscript that is easy to follow and appraise. Simple reporting that is clear and concise for scientific writing is good practice.

Q: Our model is a very large "mega-model" of an entire city. Do the STRESS guidelines apply?

The applicability of the guidelines depends on whether the contribution is the model and scientific result or the programming framework you have developed. For frameworks, it may be more appropriate to use standard documentation approaches from software engineering (e.g., Andrade et al., 2004; Insfrán, Pastor, & Wieringa, 2002; Rolland & Prakash, 2000). If it is a specific model with a specific result that you wish to

Table 7. Example reporting for implementation specifics.

- The DES model reported was implemented in the commercial software Anylogic 7.5.3 Researcher edition and made use of its Process Modelling Library version x.2. The pseudo-random number generator was provided by the Java class Random version x.y. The model was run on a Microsoft Surface Pro 4, with a 2.2Ghz Intel Core i7 processor and 16 GB of memory under Windows 10 (build 14,393). Model run time was 5 min per replication
- The SD model reported was implemented in iThink 10.0.3. The integration method was set to Euler's Method. The model had a run length of 15 years with a DT of 1 month. The model was run on an Apple Macbook Air, with a 1.7Ghz Intel Core i7 processor and 8 GB of memory under OS X El Capitan version 10.11.16. Model run time was under 1 min

disseminate (for example, how a transport policy affects overall city congestion) then it is a scientific paper that should follow the appropriate STRESS guidelines. We acknowledge that for very complex and large models some adaption may be required. The authors may wish to note this in a letter to the editor.

Q: Not all of the elements of STRESS are applicable in my case.

The STRESS guidelines are not rules they are *guidelines* to strengthen reporting of simulation models. There is no issue if authors find that certain sections of STRESS are not applicable. Reviewers may query omissions, and authors should be able to justify them.

Q. There are some unique features of my model that STRESS does not cover. Should I document them?

The spirit of the guidelines is reproducibility and we expect that the guidelines set the minimum information needed to report most models from ABS, SD and DES. If an author believes that their model requires additional detail in order to aid reproducibility, then they should include it for completeness of reporting.

Q. I have published my model code – why do I need to include the other details in STRESS?

We recommend that authors consider publishing code as an enhancement to reporting not the other way around. Some journals, such as System Dynamics Review, require authors to submit the model that they are reporting. Publication of open code will strengthen reproducibility (we note that not all authors may wish to share their code for commercial reasons). However, other researchers and practitioners may not have access to the commercial software used or may not have the right programming skills. Even when researchers are familiar with the programming languages used the code itself might be difficult to follow for a variety of reasons.

Q: My model is stochastic. What is a reproducible result?

There are several levels of reproducibility that might be achieved within the reporting of a model and its results: ordinality, precision and output uncertainty. Ordinality refers to the order of results, i.e., for a given model output, other researchers can confirm that scenario x is statistically better than scenario y and both x and y are better than scenario z. The precision of a result refers to the point estimate for a model output of a specific model configuration or the difference in output between two model configurations (scenarios). Within a stochastic model, for a given output and model configuration, the point estimate should be reproducible exactly or within the given confidence interval of reported results. It follows from precision that any sensitivity analysis conducted to assess uncertainty in model outputs due to uncertainty in model inputs should be reproduced in the recreated model.

Q: I am modelling using DEVS. Do the guidelines apply?

Discrete-event system specification (DEVS) has its own self-documenting formalisation (Zeigler, Praehofer, & Kim, 2000, p. 75). The STRESS guidelines are aimed at improving the completeness of reporting for DES, SD or ABS models that have been developed using less specialised approaches (i.e., simulation software, programming languages or general purpose programming languages). We see DEVS has the potential to form one part of the simulation reporting but these guidelines complement its use by incorporating other details needed for full reproducibility of the modelling study.

Q: Why are the extra details needed? If I want to recreate a model I can contact the authors of the work.

Yes that is fine and if necessary there is no reason not to contact the authors, but it assumes that the author(s) are still contactable, available, willing to respond and can remember. The published write-up is the permanent public record of the work. The reporting may not be perfect, but if the guidelines have been followed it will reduce the reliance on the authors and if necessary it will help an author answer the questions perhaps five years after publication (which might be seven years or more after the work was actually done). We refer the reader to the study by Boylan et al. (2015) where the authors of the work trying to be reproduced were contacted. It did not help.

Q: Will STRESS limit the write-up of "projects" i.e., the story of what happened in a simulation study

The reporting guidelines apply to the model itself and the model's results. The wider aspects of the modelling process/practice and its context are also of importance to the scientific and practical communities. This is a separate scientific area from STRESS and it is linked to Behavioural OR (BOR; Franco & Hämäläinen, 2016; Kunc, Malpass, & White, 2016). Such studies should follow the rigour of an appropriate BOR methodology and perhaps an appropriate quantitative or qualitative reporting guideline.

Q. My model and operations system is confidential and I cannot include all of the details recommended by STRESS. However, there are lessons from the work that are relevant to the simulation community.

In such cases, reviewers of the work wish to know that a rigorous approach to model development and analysis has been followed. One option to do this is to appoint an independent third party to quality assure the work whilst still maintaining confidentiality of proprietary information. It is recommended that a summary of the quality assurance is submitted along with the model. If the novel aspects of the work cannot be fully understood and verified without knowing the confidential information then it is better not to publish.

Q: I have concerns about other using my work without crediting the original authors of the work.

We recommend that authors publish their work under a creative commons licence. Authors are now licensers of their research and can choose a licence that suits their needs. A popular licence used by many open access publishers is the attribution licence CC BY (https://creativecommons.org/licences/). This means that anyone can reuse the work either in part or in whole for any purpose, for both commercial and non-commercial licence but must credit the original authors. Other licences mandate that licensees also make their published work available under the same licence terms as the original.

Q: I plan to publish the details of my model through a third party or academic institutions website and reference it in the manuscript. Do I still need to follow the STRESS guidelines?

We appreciate that many models in ORMS are large and complex. This has the potential to lengthen journal articles. We strongly recommend that, where available, authors make use of journal facilities for online supplementary material as opposed to third party websites. Third party websites may change, break or be taken down without an author's knowledge. This has the potential to affect review and reuse at a later date. Journal articles are a permanent public record of the model.

Q: Model verification and validation is not included in the guidelines. Should I include it in my write-up?

The short answer is Yes. It is recommended and preferred that details of model validation and verification (V&V) are included in both academic and industry reports of models. The STRESS test aims to increase the reproducibility of models and V&V is not a requirement to do so. Nonetheless, STRESS helps with reporting of V&V. It requires verification data to be reported, either the data used to produce the scientific results or hypothetical test verification data. Reporting V&V in full would require details of the tests, either statistical or based on expert judgement. It may also be necessary to provide more than one set of test data as a single set may not provide sufficient confidence to other researchers and practitioners who wish to reuse a model due to "not invented here syndrome" (Monks, Robinson, & Kotiadis, 2014; Robinson et al., 2004). So, V&V data and results can provide confidence that a model has been reproduced to sufficient accuracy. It is recommended that authors consult V&V literature (Onggo & Karatas, 2016; Sargent, 2013; Sterman, 2000; Windrum, Fagiolo, & Moneta, 2007).

Q: I have reused or adapted a published or publicly available model that has followed the STRESS guidelines. How do I report the model I am using in my paper?

It is only necessary to report the adaptations to the model, data or analysis that you have conducted. These modifications should follow the STRESS guidelines. Provide a reference to the original paper. The same applies to industry models, although industry models that adapt models from research may wish to include a copy of the academic paper in an appendix.

7. Discussion

In this article, we detail the development of the STRESS guidelines for ABS, DES and SD studies. We encourage authors, practitioners, editors and peer reviewers working across the three modelling disciplines to make use of the guidelines in their reporting and decision-making. If followed, the guidelines should increase the quality and completeness of model reporting and hence the likelihood of research being reused and extended. We believe that there are three main benefits for the academic simulation community. First, the guidelines help simulation model authors to write and *submit* better quality manuscripts to journals in the first instance. This offers the potential to reduce the quantity of rework requested by reviewers. Second, if peer reviewers make use of the guidelines then feedback on model documentation should be more structured and easier for authors to address. Third, a model that is reproducible is much more likely to be reused and in time to be cited by fellow researchers.

The guidelines also have tangible benefits for journal editors and peer reviewers. For those who review a study, the guidelines offer an additional structured approach to critiquing a manuscript and a standardised approach for assessing the quality of the research under review. This standardisation also provides more confidence to journal editors in relation to the quality of both reporting and review.

Looking forward, we have two expectations for STRESS. First, given the high volume of simulation studies published and changes in how simulation models are built and implemented, authors will inevitably

encounter instances where they believe the guidelines are insufficient for complete, unambiguous reporting of their models. As such, we expect that researchers will recommend amendments and refinements to the guidelines. The second is that the guidelines will be adopted by journals that publish high quality simulation research and will be used as a checklist for reviewers. We believe that this is a key way to overcome some of the adoption challenges articulated by other guidelines authors. However, we repeat the need for a grass roots movement to teach such skills to aspiring modellers (Rahmandad & Sterman, 2012). We can also see that the STRESS guidelines will be used outside of academia to provide a structure for recording information about simulation studies, as part of a knowledge management process.

An obvious extension to reporting guidelines is the development of automated documentation to reduce the effort required to report models. That is, software that can process a model and generate a software independent report of model logic, inputs and outputs. SDM-Doc (Martinez-Moyano, 2012) for Vensim is one example. The research challenge then is to support the wide range of simulation software available. A recent development that may facilitate such software is the Simulation Interoperability Standards Organisation's simulation reference mark-up language (https://www.sisostds.org/StandardsActivities/DevelopmentGroups/SRMLPDG-SimulationReferenceMarkupLanguage.aspx).

Another way to increase the efficiency of reporting is to conduct more research and development of design patterns for simulation models (North & Macal, 2014; Parker, Deadman, & Manson, 2008; Wolstenholme, 2003). Patterns are reusable solutions to a design problem that are both easy to communicate to other people and to understand. Patterns in themselves do not resolve reproducibility issues, but may be used within a reporting framework. Although implemented differently, pattern approaches have been introduced to both ABS and SD communities. Significant collaboration with the software engineering community could advance this area. There is also a large gap in pattern research within DES. The conceptual modelling literature may hold some insight here (e.g., Balci, Arthur, & Ormsby, 2011; Balci & Ormsby, 2007; Robinson, 2008; van der Zee, Holkenborg, & Robinson, 2012) .

8. Conclusions

Reporting guidelines provide a simple and powerful way to improve the quality and completeness of simulation model reporting. The STRESS guidelines are applicable across a wide range of ABS, DES and SD modelling studies. If a report of a simulation study can pass the "STRESS test" then we believe that the reusability of the simulation model and its results can be maximised. Over the next few years, we hope to see a range of ORMS journals and other journals that publish high quality simulation research adopt these guidelines. This paper has discussed STRESS applied to the modelling and simulation paradigms of ABS, DES and SD. Later work will consider how the guidelines are applicable to hybrid and distributed simulation.

Acknowledgements

One of the authors (TM) is funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) Wessex. The views expressed in this publication are those of the author and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) Wessex.

ORCID

Thomas Monks http://orcid.org/0000-0003-2631-4481
Christine S. M. Currie http://orcid.org/0000-0002-7016-3652
Bhakti Stephan Onggo http://orcid.org/0000-0001-5899-304X
Stewart Robinson http://orcid.org/0000-0002-6016-0167
Simon J. E. Taylor http://orcid.org/0000-0001-8252-0189

References

Andrade, J., Ares, J., Garc, R., #237, Pazos, J., Rodr, S., Silva, S.. (2004). A methodological framework for viewpoint-oriented conceptual modeling. *IEEE Transactions on Software Engineering, 30*(5), 282–294. doi:10.1109/tse.2004.1

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*, 452–454. doi:10.1038/533452a

Balci, O., Arthur, J. D., & Ormsby, W. F. (2011). Achieving reusability and composability with a simulation conceptual model. *Journal of Simulation, 5*(3), 157–165. doi:10.1057/jos.2011.7

Balci, O., & Ormsby, W. F. (2007). Conceptual modelling for designing large-scale simulations. *Journal of Simulation, 1*(3), 175–186. doi:10.1057/palgrave.jos.4250023

Boylan, J. E., Goodwin, P., Mohammadipour, M., & Syntetos, A. A. (2015). Reproducibility in forecasting research. *International Journal of Forecasting, 31*(1), 79–90. doi:10.1016/j.ijforecast.2014.05.008

Brailsford, S., & De Silva, D. (2015). How many dentists does Sri Lanka need? Modelling to inform policy decisions. *Journal of the Operational Research Society, 66*(9), 1566–1577. doi:10.1057/jors.2014.136

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77–101).

Dalle, O. (2012). *On reproducibility and traceability of simulations*. Paper presented at the Proceedings of the Winter Simulation Conference, Berlin, Germany.

Epstein, J. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton, NJ: Princeton University Press.

Fletcher, A., Halsall, D., Huxham, S., & Worthington, D. (2007). The DH accident and emergency department model: A national generic model used locally. *Journal of the Operational Research Society, 58*(12), 1554–1562.

Fletcher, A., & Worthington, D. (2009). What is a 'generic' hospital model?–a comparison of 'generic' and 'specific' hospital models of emergency patient flows. *Health Care Management Science, 12*(4), 374–391. doi:10.1007/s10729-009-9108-9

Forrester, J. W. (1971). *World dynamics*. Massachusetts: MIT Press.

Franco, L. A., & Hämäläinen, R. P. (2016). Behavioural operational research: Returning to the roots of the OR profession. *European Journal of Operational Research, 249*(3), 791–795. doi:10.1016/j.ejor.2015.10.034

Gass, S. I. (1984). Documenting a computer-based model. *Interfaces, 14*(3), 84–93. doi:10.1287/inte.14.3.84

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., & DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*(1–2), 115–126. doi:10.1016/j.ecolmodel.2006.04.023

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling, 221*(23), 2760–2768. doi:10.1016/j.ecolmodel.2010.08.019

Gunal, M. M., & Pidd, M. (2009). Understanding target-driven action in emergency department performance using simulation. *Emergency Medicine Journal, 26*(10), 724–727. doi:10.1136/emj.2008.066969

Günal, M. M., & Pidd, M. (2011). DGHPSIM: Generic simulation of hospital performance. *ACM Transactions on Modeling and Computer Simulation, 21*(4), 1–22. doi:10.1145/2000494.2000496

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., . . . Loder, E. (2013). Consolidated health economic evaluation reporting standards (CHEERS) statement. *BMJ: British Medical Journal, 346*. doi:10.1136/bmj.f1049

Insfrán, E., Pastor, O., & Wieringa, R. (2002). Requirements engineering-based conceptual modelling. *Requirements Engineering, 7*(2), 61–72. doi:10.1007/s007660200005

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L. K., & Young, T. (2010). Simulation in manufacturing and business: A review. *European Journal of Operational Research, 203*(1), 1–13. doi:10.1016/j.ejor.2009.06.004

Janssen, M. A. (2017). The practice of archiving model code of agent-based models. *Journal of Artificial Societies and Social Simulation, 20*(1), 2. doi:10.18564/jasss.3317

Kendall, G., Bai, R., Błazewicz, J., De Causmaecker, P., Gendreau, M., John, R., & Yee, A. (2016). Good laboratory practice for optimization research. *Journal of the Operational Research Society, 67*(4), 676–689. doi:10.1057/jors.2015.77

Kunc, M. (2008). Achieving a balanced organizational structure in professional services firms: Some lessons from a modeling project. *System Dynamics Review, 24*(2), 119–143. doi:10.1002/sdr.394

Kunc, M., & Kazakov, R. (2013). Competitive dynamics in pharmaceutical markets: A case study in the chronic cardiac disease market. *Journal of the Operational Research Society, 64*(12), 1790–1799. doi:10.1057/jors.2012.150

Kunc, M., Malpass, J., & White, L. (Eds.). (2016). *Behavioral operational research*. UK: Palgrave Macmillan.

Kurkowski, S., Camp, T., & Colagrosso, M. (2005). MANET simulation studies: The incredibles. *ACM SIGMOBILE Mobile Computing and Communications Review, 9*(4), 50–61. doi:10.1145/1096166.1096174

Martinez-Moyano, I. J. (2012). Documentation for model transparency. *System Dynamics Review, 28*(2), 199–208. doi:10.1002/sdr.1471

Miller, D. M., & Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting, 19*(4), 669–684. doi:10.1016/S0169-2070(02)00077-8

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Medicine, 7*(2), e1000217. doi:10.1371/journal.pmed.1000217

Monks, T., Robinson, S., & Kotiadis, K. (2014). Learning from discrete-event simulation: Exploring the high involvement hypothesis. *European Journal of Operational Research, 235*(1), 195–205. doi:10.1016/j.ejor.2013.10.003

North, M. J., & Macal, C. M. (2014). Product and process patterns for agent-based modelling and simulation. *Journal of Simulation, 8*(1), 25–36. doi:10.1057/jos.2013.4

Onggo, B. S., & Karatas, M. (2016). Test-driven simulation modelling: A case study using agent-based maritime search-operation simulation. *European Journal of Operational Research, 254*(2), 517–531. doi:10.1016/j.ejor.2016.03.050

Parker, D. C., D.G. B., Polhill, J. G., Deadman, P. J., & Manson, S. M. (2008). Illustrating a new conceptual design pattern for agent-based models of land use via five case studies - the MR POTATOHEAD framework. In A. Lopex Paredes & C. H. Iglesias (Eds.), *Agent-based modelling in natural resource management* (pp. 23–51). Valladolid: INSISOC.

Pasupathy, R., & Henderson, S. G. (2006, 3–6 December 2006). *A testbed of simulation-optimization problems.* Paper presented at the Proceedings of the 2006 Winter Simulation Conference. Monterey, CA, USA.

Pidd, M. (2006). *Computer simulation in management* (5 ed.). Chichester: John Wiley & Sons.

Pierson, K., & Sterman, J. D. (2013). Cyclical dynamics of airline industry earnings. *System Dynamics Review, 29*(3), 129–156. doi:10.1002/sdr.1501

Rahmandad, H., & Sterman, J. D. (2012). Reporting guidelines for simulation-based research in social sciences. *System Dynamics Review, 28*(4), 396–411. doi:10.1002/sdr.1481

Robinson, S. (2008). Conceptual modelling for simulation Part I: Definition and requirements. *Journal of the Operational Research Society, 59*(3), 278–290.

Robinson, S., Nance, R. E., Paul, R. J., Pidd, M., & Taylor, S. J. E. (2004). Simulation model reuse: Definitions, benefits and obstacles. *Simulation Modelling Practice and Theory, 12*(7-8), 479–494. doi:10.1016/j.simpat.2003.11.006

Rolland, C., & Prakash, N. (2000). From conceptual modelling to requirements engineering. *Annals of Software Engineering, 10*(1/4), 151. doi:10.1023/a:1018939700514

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation, 7*(1), 12–24. doi:10.1057/jos.2012.20

Sterman, J. D. (2000). *Business dynamics*. Boston, MA: McGraw-Hill Education.

Waltemath, D., Adams, R., Beard, D. A., Bergmann, F. T., Bhalla, U. S., Britten, R., & Le Novère, N. (2011). Minimum Information About a Simulation Experiment (MIASE). *PLoS Computational Biology, 7*(4), e1001122. doi:10.1371/journal.pcbi.1001122

White, K. P., & Robinson, S. (2010). The problem of the initial transient (again), or why MSER works. *Journal of Simulation, 4*(4), 268–272. doi:10.1057/jos.2010.19

Windrum, P., Fagiolo, G., & Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation, 10*(2), 8.

Wolstenholme, E. F. (2003). Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review, 19*(1), 7–26. doi:10.1002/sdr.259

Yates, J., Ford, A., & Kuglics, J. (2014). Identifying key parameters and trends in civil violence: A sub-regional, agent-based simulation approach using GIS. *Journal of Simulation, 8*(3), 179–194. doi:10.1057/jos.2013.26

van der Zee, D. J., Holkenborg, B., & Robinson, S. (2012). Conceptual modeling for simulation-based serious gaming. *Decision Support Systems, 54*(1), 33–45. doi:10.1016/j.dss.2012.03.006

Zeigler, B., Praehofer, H., & Kim, T. (2000). *Theory of modeling and simulation*. New York, NY: Academic Press.