System Dynamics Review
System Dynamics Review vol 28, No 4 (October-December 2012): 396–411
Published online 11 October 2012 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/sdr.1481

NOTES AND INSIGHTS

Reporting guidelines for simulation-based research in social sciences

Hazhir Rahmandad[a*] and John D. Sterman[b]

Syst. Dyn. Rev. **28**, 396–411 (2012)

Supporting information may be found in the online version of this article.

Introduction and motivation

Reproducibility is central to the progress of science, and simulation-based research is no exception. Only when research results are independently reproducible can different scholars verify the results reported by others, build on each other's work, and convince the public of the reliability of their results (Laine *et al*., 2007). Given the widespread use of computational methods in different branches of science, many scientists have called for more transparency in documenting computational research to allow reproducibility (Schwab *et al*., 2000; Code, 2010; Peng, 2011). Simulation-based research in the social sciences has been on the rise over the last few decades (Gilbert and Troitzsch, 2005), yet a set of reporting guidelines that ensure reproducibility and more efficient and effective communication among researchers is lacking. As a result, many research reports lack the information required to reproduce the simulation models they discuss or the specific simulation experiments they report. In this paper we provide an initial set of reporting guidelines for simulation-based research (RGSR) in the social sciences, with a focus on common scenarios in system dynamics research. We discuss these guidelines separately for reporting models, reporting simulation experiments, and reporting optimization results. The guidelines are further divided into minimum and preferred requirements, distinguishing between factors that are indispensable for reproduction of research and those that enhance transparency. We also provide guidelines to improve visualization of research to reduce the costs of reproduction. Finally we offer suggestions to enhance the adoption of these guidelines.

To illustrate the challenge of documentation and reproducibility, we reviewed all the articles published in *System Dynamics Review* in the years 2010 and 2011. Of 34 research articles, 27 reported results from a simulation model. Of these 27, the majority (16; 59%) did not include model equations, two (7%) contained partial equations, and the rest reported the complete model, either in the text (3; 11%), in an online appendix

[a] Grado Department of Industrial and Systems Engineering, Virginia Tech, Falls Church, VA 22043, U.S.A.
[b] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
* Correspondence to: Hazhir Rahmandad, Grado Department of Industrial and Systems Engineering, Virginia Tech, Falls Church, VA 22043, U.S.A. E-mail: Hazhir@vt.edu

(5; 19%), or by referencing another publication (1; 4%). Similarly, only eight articles (30%) included the parameter values needed to replicate the base case. Only six (22%) included complete units for all the equations, with three offering partial coverage of units. Finally, the information needed to replicate a reported graph (e.g. scenario and parameter settings) was missing in eight studies and could not be verified without attempting full reproduction in another five. These findings are consistent with recent research that has examined model quality and documentation in *System Dynamics Review* and International System Dynamics Conference proceedings (Groesser and Tschupp, 2012). Despite a long tradition emphasizing model transparency, attention to modeling process, and reproducibility of results (Forrester, 1961; Sterman, 2000) the system dynamics literature is falling short of the goals for full reproducibility to which it aspires. Similar challenges to reproducibility are reported in a variety of disciplines and journals (Dewald *et al.*, 1986; Hubbard and Vetter, 1996; Ioannidis, 2005; McCullough *et al.*, 2006, 2008; Koenker and Zeileis, 2009).

In response, some fields developed guidelines for the reporting of models and simulation results, such as minimum information required in the annotation of biochemical models (MIRIAM) (Le Novere *et al.*, 2005), minimum information about a simulation experiment (MIASE) in systems biology (Waltemath *et al.*, 2011), IIE computational research reporting guidelines (Lee *et al.*, 1993), standards for describing agent-based models (Grimm *et al.*, 2006), and guidelines for mathematical programmers for reporting computational experiments (Jackson *et al.*, 1991). Others have called for reproducibility of all computational research (Code, 2010; Peng, 2011) and some go further, calling for provision of the full computational environment that produces published results (Donoho *et al.*, 2009; Morin *et al.*, 2012).

Here we propose standard reporting guidelines for simulation-based research in the social sciences to enhance reproducibility. We focus on common scenarios encountered in the field of system dynamics, but the guidelines should be informative for other modeling approaches as well. Table 1 defines the key concepts we use in this paper. The scope of these guidelines is limited to the reporting of the model and simulation results and does not attempt to specify best modeling or analysis practices: reasonable people can disagree about how a system should be modeled, but, we argue, all should document their work in such a way that it is fully reproducible by others.[1]

Simulation-based research reports results of simulation and optimization experiments on a model. Therefore the guidelines that follow are discussed separately for general visualization, reporting a model, reporting simulation experiments, and reporting optimization experiments. An optimization experiment often consists of many simulation runs, and as such will follow the requirements outlined for simulation experiments. However, optimization experiments also include additional reporting requirements that are discussed separately. For each type of information we identify minimum and preferred reporting requirements, where minimum requirements are essential for research reproducibility, and preferred requirements are recommended for enhanced communication and transparency.

A simple example

To provide a concrete example of these reporting requirements, we introduce a simple model, presented following the requirements we propose. The model is illustrative and

DOI: 10.1002/sdr

Table 1. Basic definitions for the concepts used in this paper

Model: a mathematical representation of a social system that can be simulated to generate numerical results.

Exogenous inputs: we distinguish between three types of exogenous model inputs. Model **parameters** are constant numerical values used in the model, including data inputs, parametric assumptions on functions used, and other numerical inputs to algorithms used in the model (e.g. the assumed time constant for inventory adjustment in a supply chain model). **Exogenous variables** are time-varying inputs that are fixed in advance and capture the dynamics of variables outside the boundary of the model (e.g., historical and projected population used in a macroeconomic model). **Pseudo-random number streams** (used in stochastic models) are generated through a random generation process and follow specified distributional assumptions specified by other parameters (e.g. random variations around the expected value in the incidence of new cases in an epidemiological model).

Simulation run: a single simulation consisting of computational operations on a model generating numerical results that represent some aspects of the system of interest given an instance of exogenous inputs. In comparing different simulation runs of a model we distinguish between **iterations**, simulation runs that use same parameter and exogenous variables values but differ in the realizations of pseudo-random number streams used in them, and **scenarios**, which differ in parameter or exogenous variable values.

Experimental setup: the design of simulation runs, in terms of the scenarios simulated and the number of iterations used, that inform simulation and optimization experiments.

Simulation experiment: a set of simulation runs that are conducted and some outputs of which are reported.

Optimization experiment: these experiments combine the results of a simulation model with a search algorithm to find values for a subset of model parameters that best match a desired outcome.

the numerical results reported here do not have any real-world significance. The model builds on the classical Bass diffusion model (Bass, 1969; as implemented in Sterman, 2000) and incorporates first-order autocorrelated noise (Sterman, 2000) in the adoption rate (AR) around the expected values.[1] Figure 1 provides a graphical representation of the model. Given that the model is small, we include the model parameters in the diagram; for example, we show that adoption from word of mouth (WOM) depends on three constants: the adoption fraction *i*, the contact rate *c* and total population *n*. We note, however, that there are circumstances in which authors may choose not to include all parameters in such diagrams. Specifically, for larger models and/or contexts where the purpose of the diagram is to communicate the overall feedback or stock and flow structure it is often appropriate to include in the diagram only those parameters that are important for the presentation and discussion, omitting others, for example, Figure 1 in Mojtahedzadeh (2012). While the full documentation of larger models, which will typically be presented in an appendix or online supplement, should show diagrams that correspond exactly to the full model, diagrams showing the structure of large models should generally be avoided in the main presentation—authors should avoid cluttering their papers with complex "gazinta diagrams" (after "goes into"; see Matthews and Matthews, 2008) indistinguishable from a plate of spaghetti. Table 2 specifies the model using the preferred

[1]All model variables and parameters include an abbreviation, at the end of the variable name in Figure 1, which is in capital letters for variables and in small letters for parameters. We use these abbreviations in presenting model equations within the text, but provide full variable names in the simulation models in the online Appendix.

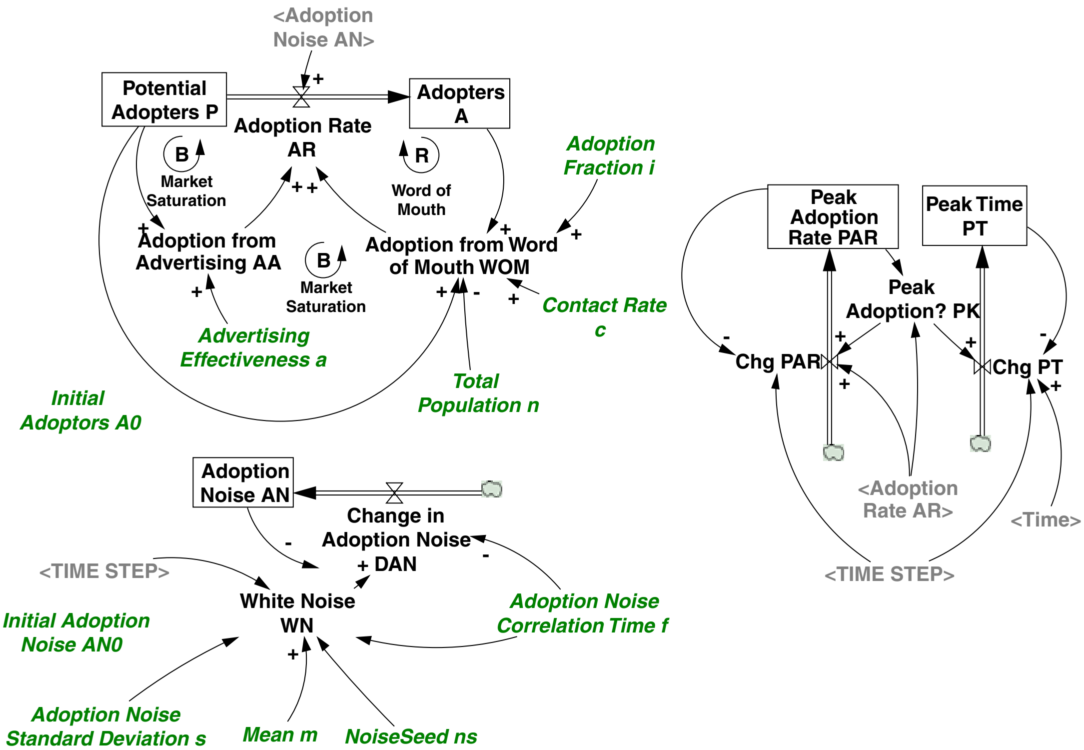

Figure 1. Graphical representation of the modified Bass diffusion model used as an example throughout this paper. Model parameters are identified by *italic* font.

requirements for model reporting. The complete model is also available in the online appendix for independent assessment and reproduction.

General visualization guidelines

Logically, reproducibility of research does not depend on how the model and the results are presented visually. However, if the goal is to facilitate independent replication of different studies, visualization matters. Poor visualization increases the costs of reproduction, leads to confusion and needless error, and therefore reduces the incidence and quality of reproducibility. We therefore provide brief recommendations for key visualization issues common in reporting models and simulation results in social sciences, noting that there is a large literature on visualization (Tufte, 2001; Bonneau *et al.*, 2006) and comprehensive coverage of this topic is beyond the scope of the current paper.[2]

[2]For a historical perspective on visualization in system dynamics see Lane (2008). *System Dynamics Review* has recently provided a set of useful guidelines, available at http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1727/homepage/ForAuthors.html (last accessed 20 August 2012).

- As mentioned above, avoid "gazinta diagrams" whose only function is to show readers that you have a complex model with lots of stuff in it. Such diagrams contribute nothing to the presentation of your model or insights from it. Unfortunately, examples of poor practice are plentiful (see, for example, Fig. 4 in Carter and Moizer, 2011).
- Avoid clutter in presenting causal relationships of a model. For example, use multiple views and avoid crossing causal links or overlapping variable names. Make sure polarity signs are visible.
- Use subsystem diagrams, model boundary charts, causal loop diagrams, simplified stock and flow maps, and other relevant summary visuals (Sterman, 2000) when, due to space limitations, presenting the causal diagrams for the full model is infeasible in the main body of the article.
- Use sans serif fonts such as Arial and Helvetica, and avoid serif fonts (e.g. Times) in visualizations.
- Choose a standard in naming and presenting variables and follow it consistently (e.g. friendly algebra with long variable names (Morecroft *et al.*, 1991) or Roman and Greek letter-based variable and parameter names).
- Limit the amount of information reported in graphs to the key items you want to discuss. Do not include default vertical and horizontal gridlines and other "chart junk" (Tufte, 2001).
- Make sure different lines portraying variables in graphs of simulation output are distinguishable whether viewed online or on paper, and in color or black and white (e.g., use both different colors and line thicknesses for different variables in the same graphs).

Model reporting requirements

Reproducibility requires standards for the presentation of the model structure and for the numerical results that are generated by the model. Model-reporting requirements should therefore be followed whenever a simulation model is discussed or any results reported.

Minimum model reporting requirement (MMRR)

Any model used to generate research results must be reported so that independent parties can recreate the model and simulate it in the base-case setting, on a computational platform of their choice, based on information provided in the reported research.[3] This requirement includes, but is not limited to:

- The computational operations the model is designed to perform shall be explained in plain text and provided within the paper or in an online appendix. Typically such documentation includes equations and algorithmic rules, all model parameters and initial values. The description should be sufficient to allow an independent party to implement and simulate the model.

[3]For our purposes, the research report includes both the main manuscript and any accompanying online supplement available to anybody without the need to contact the authors.

DOI: 10.1002/sdr

Table 2. Modified Bass diffusion model documentation following preferred model reporting requirements. MMRR compliance would be achieved by reporting the formulas and units

Formulations and comments	Units
$P(t) = P(0) + \int_0^t -AR(s)ds; P(0) = n - A(0)$	Person
The stock of potential adopters, *P*, declines as they adopt the innovation; the total adoption rate, *AR*, moves people from potential adopters to the stock of adopters, *A*. The initial number of potential adopters is given by the total population, *n*, less the initial number of active adopters, *A*(0).	
$A(t) = A(0) + \int_0^t AR(s)ds; A(0) = 1$	Person
The number of active adopters, *A*, accumulates the adoption rate, *AR*. There is no outflow from the adopter stock. The initial number of adopters is given by *A*(0), assumed to be a single person.	
$AR(t) = (AA(t) + WOM(t)) * Max(0, 1 + AN(t))$	Person/Year
The adoption rate, *AR*, is the rate at which potential adopters become active adopters. Adoption arises from advertising efforts, Adoption from Advertising, AA, and adoption from word of mouth, *WOM*. The actual adoption rate equals the expected value given by the sum of *AA* and *WOM*, modified by a random effect, adoption noise, *AN*, that captures stochastic variations in adoption arising from factors outside the boundary of the model. The *Max* function ensures that the adoption rate remains non-negative regardless of the realization of the random noise term *AN*.	
$AA(t) = a * P(t)$	Person/Year
Following the standard Bass model (Bass, 1969), adoption from advertising, *AA*, depends on the size of the pool of potential adopters, with the hazard rate of adoption from advertising given by advertising effectiveness, *a*, assumed, as in the Bass model, to be constant.	
$WOM(t) = c * i * P(t) * \frac{A(t)}{n}$	Person/Year
Adoption by word of mouth, *WOM*, is the product of the rate at which potential adopters have relevant contacts with other individuals, *c*, the probability that any such contact is with an adopter, given by the fraction of adopters, *A*, in the total population, *n*, and finally the probability of adoption given such a contact with an adopter, *i*. Assuming that the probability of contact with an adopter is given by the fraction of adopters in the population reflects the implicit assumption of the Bass model that adopters and non-adopters are well mixed and have the same behaviors with respect to their social contacts.	
$AN(t) = AN(0) + \int_0^t DAN(s)ds$	Dimensionless
Random variations in adoption are modeled by adoption noise, *AN*, assumed to be first-order autocorrelated noise, which is generated as an exponentially weighted average of white noise, *WN*, specified as identically and identically distributed (i.i.d.) noise, *IN*, assumed to be normally distributed. See Sterman (2000, Appendix B).	
$AN(0) = Normal(m,s) with\ Noise\ Seed = ns$	Dimensionless
The initial value of adoption noise *AN*(0) is set equal to the initial value drawn from the specified distribution of the pink noise process used to model the noise in adoption, given the specified mean and standard deviation for the pink noise ("adoption noise *AN*"). Initializing adoption noise from the steady-state distribution of pink noise stream avoids a spurious transient that would arise if Adoption noise were initialized at its expected value.	
$DAN(t) = \frac{WN(t) - AN(t)}{f}$	1/Year
The rate of change in or derivative of the adoption noise, *DAN*, is formulated as first-order exponential smoothing of a white noise stream, *WN*.	

(Continues)

DOI: 10.1002/sdr

Table 2. (Continued)

Formulations and comments	Units
$WN(t) = Normal\left(m, s\sqrt{\frac{2-\frac{dt}{f}}{\frac{dt}{f}}}\right)$, $with\,Noise\,Seed = ns$	Dimensionless
The white noise, *WN*, that drives the autocorrelated adoption noise is an independently and identically normally distributed random variable with a standard deviation scaled so that the pink adoption noise variable has the desired standard deviation, *s*.	
$PAR(t) = \int_0^t PK(s) * \frac{AR(s) - PAR(s)}{d} ds; PAR(0) = 0$	Person/Year
The peak adoption rate, *PAR*, is a summary measure of the diffusion dynamic. The *PAR* observed up to the current time in the simulation is calculated by adjusting the PAR from its current value to the current adoption rate, *AR*, whenever the current adoption rate is larger than the peak adoption rate observed to date, as determined by the peak indicator, *PK*.	
$PK(t) = if\ them\ else(AR(t) > PAR(t), 1, 0)$	Dimensionless
The adoption peak indicator variable, *PK*, tests is one if the current value of the adoption rate is the maximum observed so far and zero otherwise.	
$PT(t) = \int_0^t PK(s) * \frac{t - PT(s)}{d} ds; PT(0) = 0$	Year
The peak adoption time, *PT*, is a summary measure of the diffusion dynamic. The *PT* observed up to the current time in the simulation is calculated by adjusting the *PT* from its current value to the current time whenever the current adoption rate is larger than the peak adoption rate observed to date.	
$a = 0.01$	1/Year
Advertising results in adoption according the effectiveness of the advertising, *a*. The assumed hazard rate of adoption from exposure to advertising is 1% per year.	
$\mathrm{c} = 100$	1/Year
The rate at which active adopters come into contact with potential adopters, *c*. We assume 100 person-to-person contacts relevant to the focal innovation per year.	
$i = 0.015$	Dimensionless
The fraction of times a contact between an active adopter and a potential adopter results in adoption, *i*. We assume the probability of adoption conditional on a contact between a potential adopter and adopter is 1.5%.	
$n = 1E6$	Person
The population is assumed to be one million individuals.	
$m = 0$	Dimensionless
The mean value for the pink noise term influencing the adoption rate.	
$s = 0.1$	Dimensionless
The standard deviation of the pink noise in the adoption rate is assumed to be 10% of the expected adoption rate.	
$ns = 1$	Dimensionless
The noise seed, *ns*, specifies the particular pseudo-random number stream that affects adoption.	
$f = 2$	Year
The time constant for autocorrelation in the adoption noise.	
$d = 0.0078125$;	Year
The time step for the simulation, after sensitivity analysis on time step, was set to 0.0078125 (= 2^{-7}) years.	

DOI: 10.1002/sdr

- If a model extends a previously published and MMRR-compliant model in the publicly accessible literature, only the changes from the previously reported model need to be described.
- Units of measurement shall be reported for all variables and parameters.[4]

Preferred model reporting requirement (PMRR)

To increase the incidence and quality of model assessment and reproduction studies, modelers should provide information beyond the minimum requirements. Such information includes, but is not limited to:

- Sources of data (qualitative or quantitative) for different equations and algorithmic rules.
- Definition of all the variables used in the model and the logic behind their formulation.
- Source code in the original implementation platform, preferably in a format that can be freely accessed and simulated (e.g. for a Vensim model a .vpm or .mdl file that can be opened and executed by the freely available Vensim Model Reader).

Example of MMRR- and PMRR-compliant model descriptions

Table 2 follows the PMRR for the example model above. One could have achieved MMRR-compliant documentation by only including the formulations. The equations are represented using letters and short abbreviations for the variable names, as is standard and appropriate in, for example, scientific journals. Alternatively one could use longer, more explanatory variable names, so-called "friendly algebra" (Morecroft *et al.*, 1991), available in Figure 1, in explaining the equations in the text or in an appendix; the choice depends on the intended audience for the work. If only a subset of equations is discussed in the main text, full documentation, including all parameter values, must also be made available in an appendix or online supplement.

Simulation experiment reporting

A simulation experiment consists of setting up the model and conducting one or multiple simulation runs that generate numerical results. Simulation runs may differ in their parameter settings, i.e. belong to different scenarios, or in their driving random number streams, i.e., different realizations of the same scenario. The following reporting requirements apply to results reported from any simulation run(s).

Minimum simulation reporting requirements (MSRR)

Research should provide a detailed description of all the steps needed to repeat every reported simulation experiment and reproduce the results. Reproduced results shall be

[4]We acknowledge that some modeling disciplines put less emphasis on the units but we include providing units of measure for all variables and parameters in MMRR in light of their indispensable role in building and understanding system dynamics models.

DOI: 10.1002/sdr

consistent with the reported results within the computational error bounds expected from reproduction on different platforms and, in the case of stochastic models, differences arising from different realizations of pseudo-random variables. These requirements include, but are not limited to, reporting of:

- The software and hardware platform(s) used for the simulation.
- The simulation algorithm used, such as integration method and time step (for differential and difference equation models), meshing method (for spatial models), and event prioritization schemes (for discrete event and agent-based simulations).
- Any pre-processing (e.g. to generate exogenous inputs to the model) needed on the base-case model (described according to the requirement above) to enable reproduction of the reported experiments.
- Parameter settings required to reproduce any reported scenario, including parameter values for each scenario and, for Monte Carlo simulations, the joint distributions for the selection of parameters, including distributional forms, generating equations, and/or correlation matrixes, along with the sampling procedure used.
- The number of iterations per scenario.
- All post-processing steps (e.g. data reduction and aggregation computations, summary statistics, regressions on the simulation results) used to transform simulation outputs to reported results.

Preferred simulation reporting requirements (PSRR)

Reports of simulation experiments should include information that facilitates the assessment of the results beyond the minimum requirements. These include, but are not limited to, specifying:

- If any sensitivity analysis was conducted on robustness of the algorithmic parameters (e.g. sensitivity of results to time step or simulation method).
- Information on computational costs, including simulation time and processor information. This is especially important if computational costs are significant.
- The random number generation algorithm used and the noise seed (parameters specifying the exact stream of resulting pseudo-random numbers) for stochastic models.
- A measure of uncertainty (e.g., standard deviation, 95 percent confidence interval) in reported statistics in stochastic models and Monte Carlo analysis. The method used to calculate confidence intervals and other measures of uncertainty should be fully specified (e.g. empirical confidence interval versus one calculated assuming variations are normally distributed).
- In stochastic models, when differences between metrics across different scenarios are reported, the statistical significance of the difference and the significance testing method.
- The method for determining the number of significant digits presented in tables and graphs. When original code is provided, instructions for conducting the simulation experiment in the original platform.

Example of MSRR- and PSRR-compliant reports

Table 3 reports the results of a sensitivity analysis on the diffusion model described above. The analysis changes two of the model parameters over three values each (a full

DOI: 10.1002/sdr

Table 3. Results of a set of sensitivity analysis simulations for the example diffusion model. MSRR[1]- and PSRR[2]-compliant documentation are provided in table footnotes

Mean (SD) of *PAR* and *PT* with different *i* and *a*; based on 1000 iterations		Adoption fraction *i* 0.005	0.015	0.025
Advertising effectiveness *a*				
0.005	Peak adoption rate (*PAR*)	138125 (10428)*	395720 (34266)	649324 (58901)*
	Peak time (*PT*)	9.16 (0.85)*	3.82 (0.35)*	2.51 (0.23)*
0.01	Peak adoption rate (*PAR*)	140631 (10926)*	397741 (34733)	652055 (58804)*
	Peak time (*PT*)	7.71 (0.8)*	3.34 (0.32)	2.23 (0.21)*
0.015	Peak adoption rate (*PAR*)	143237 (11288)*	400188 (35612)	653820 (58936)*
	Peak time (*PT*)	6.85 (0.75)*	3.07 (0.31)*	2.06 (0.2)*

[1]The experiment includes nine different scenarios for model parameters in which the adoption fraction (*i*) and advertising effectiveness (*a*) are varied around base values of 0.015 and 0.01, respectively, as specified in the table. The table reports the mean and standard deviation for *PAR* and *PT* in ensembles of 1000 simulations for each scenario. Iterations in each scenario differ in the realizations of the adoption noise, *AN*, which varies the adoption rate around the expected value with a pink noise process with a standard deviation of 10%. The time horizon for each simulation is 25 years. Simulations were conducted using Vensim® software version 5.11 using Euler integration with a time step of 0.0078125 years. An asterisk indicates that the mean for a metric is statistically different, at $p \leq 0.01$, from the base case results (the center cell in the 3×3 table) based on the *t*-test for group means with unequal variances.
[2]Results were not sensitive to use of Runge–Kutta integration methods (RK2, RK4 and RK-Auto were tested) or smaller time steps. Iterations generated using "NoiseSeed ns" parameters [1–1000]. *PAR* numbers are rounded to the closest integer and *PT* numbers are rounded to two decimals. The pseudo-random number generator uses the rand1.c function from *Numerical Recipes in C* (Press, 1992). The 9000 iterations (nine scenarios with 1000 iterations in each) for this analysis, saving only *PAR* and *PT* once a year, took 2:20 minutes on a desktop computer with a Q9400 Intel Core 2 CPU at 2.66 GHz with a 64-bit Windows 7 operating system and 4 GB of RAM.

factorial analysis yielding a total of nine scenarios) and runs multiple iterations for each of these scenarios, calculating the sensitivity of two outcome variables of interest (peak adoption rate (PAR) and peak time (PT)) to these parameters. The table footnotes provide the MSRR and PSRR information.

Optimization experiment reporting

Optimization experiments can be applied to deterministic or stochastic models and are used for policy optimization, calibration (estimating parameters of interest by minimizing some function of the error between the simulation and data), dynamic programming, and finding equilibria in multi-player games, among others. The following information should be provided to enable reproduction of optimization experiments.

Minimum optimization reporting requirements (MORR)

Besides following the MSRR, the optimization objective function, search algorithm and search space underlying the optimization procedure shall be specified with enough detail to enable the reproduction of the optimization experiment by independent researchers. Exact numerical reproduction may not be feasible due to variations in pseudo-random number streams used in some optimization methods, or other platform-based differences

DOI: 10.1002/sdr

(e.g. in truncation or round-off error). However, the reproduced and reported results should, with sufficiently large samples, show no statistically significant differences. The minimum reporting requirements include, but are not limited to:

- The software environment in which the optimization has been implemented.
- The payoff function to be maximized (minimized) as a function of reported model variables. In game theoretic settings the payoff function of all the players involved shall be specified.
- The parameter space over which search for the best payoff value is conducted. If search parameters are not part of the model discussed above (e.g. feature-space definition and functional approximations used for approximate dynamic programming (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2007) the mapping of search parameters into model variables shall be reported.
- The search algorithm used shall be specified by reference to the original article introducing the algorithm and fully explaining any modifications or new search methods.
- If iterative methods are used, e.g. for finding game-theoretic equilibria (Kim and Kim, 1997; Sterman *et al.*, 2007; Rahmandad and Sibdari, 2012), the number of iterations needed for convergence shall be reported.
- The actual search that has led to the reported optimization results based on the algorithm used, e.g. the number of restarts of the search in the parameter space, the total number of scenarios simulated, and the number of iterations per scenario (for stochastic models).

Preferred optimization reporting requirements (PORR)

Reports of optimization experiments should go beyond the minimal requirements to include information that facilitates quick reproduction and the assessment of the results. These include, but are not limited to:

- Optimization implementation codes (e.g. Vensim payoff definition and optimization control files).
- Information on computational costs, including optimization time and processor information for each optimization experiment.
- For calibration/estimation results, a measure of uncertainty in the estimated parameters (e.g. 95 percent confidence interval).
- A measure of confidence in the generality of optimization results.[5] Examples include the number of unique local optima discovered divided by the number of restarts and, for stochastic models, the confidence level that the best local optimum is found (as different random number streams will find different local optima in the neighborhood).

Example of MORR- and PORR-compliant reports

We consider an optimization problem that builds on the diffusion model above to find the advertising policy that produces a desired peak adoption rate (PAR). Minimum reporting requirements follow, with preferred requirements provided in endnote ii and the Appendix.

[5]A global optimum cannot usually be guaranteed in simulation optimization, yet one can assess the probability of finding better peaks with additional search based on the expected return to further search.

DOI: 10.1002/sdr

In light of the stochastic nature of this model, we need to simulate the model multiple times and find an approximate value for the expected PAR. To do so we generate 1000 iterations of the model (using the subscript functionality in Vensim; see Appendix) and calculate the mean PAR in that sample. That value is then compared to a goal for PAR, which we set to 600,000 persons per year. We use Vensim's built-in optimization module, which uses a modified Powell conjugate search algorithm (taken from numerical recipes in C (Press, 1992) but modified to include additional constraints) to search values of "Advertising Effectiveness a" between 0 and 2 and find the value that minimizes the squared error between the PAR mean at the final time and the goal (of 600,000 people/year). We find the optimal value for Advertising Effectiveness $a = 0.361$ leads to the desired mean PAR. The optimal value was found after 706 simulations to accommodate 45 random restarts of the search in the parameter space.[2]

Discussion and recommendations

The guidelines developed in this article provide a starting point to standardize the reporting of simulation results in social science research. Such guidelines need to be updated based on feedback from the community of researchers who use them; therefore we encourage people to use the guidelines and suggest revisions and enhancements. A more challenging task is to promote the adoption of these guidelines within the research community. While each researcher benefits from using RGSR-compliant research by others, the additional work needed to develop such reports creates a collective action problem. We recommend the following steps to help address this challenge:

- The community can benefit from software programs that facilitate documentation of models. For example the SDM-DOC tool developed at Argonne National Laboratories offers documentation for models developed in Vensim® simulation environment as well as valuable model formulation diagnostics (Martinez-Moyano, 2012). The SDM-DOC tool is freely available through the System Dynamics Society website,[6] and we recommend it be used whenever possible. To illustrate, we include reports from the SDM-DOC tool for the example model used here in the e-companion.
- Developing good reporting habits should start early in the training of researchers. Advisors should require their students follow these standards in all internal reports and external papers. Such a requirement is likely to speed and improve communication between advisors and students and thus benefit the productivity of the research group overall, including that of the advisor.
- Senior faculty should promote these guidelines by highlighting RGSR compliance of research in their discussions with students and junior faculty, seminars, and other public encounters.
- Reproduction of published simulation research and assessment of its RGSR compliance can be a very useful training method for graduate students. It will also create pressure on other researchers to ensure RGSR compliance.
- Journals and conferences should ask authors to identify the compliance of their submitted work to the different components of RGSR. The additional transparency and

[6]At http://tools.systemdynamics.org/ (last accessed 20 August 2012).

DOI: 10.1002/sdr

reviewer trust that accompany such voluntary disclosure will provide an incentive for individual researchers to develop RGSR-compliant articles. Journals more heavily focused on simulation research should also require minimum RGSR compliance for all submissions. Several highly respected journals have already created such requirements (Editors, 2008, 2012).

- Journals and conferences should add additional questions to their review forms so that referees can report the degree of RGSR compliance of submitted articles.
- Recognizing that *System Dynamics Review* publishes a range of papers, including practitioner-oriented research, we recommend the journal require minimum RGSR compliance for all research-focused papers. Doing so will allow the journal to remain open to practice-based manuscripts in which the authors may not be able to reveal the full details of proprietary models, while setting appropriate standards for more academic publications.

Application of RGSR will allow researchers to better understand each other's work, engage in more and more effective collaborations, train their students more efficiently, identify errors in their work earlier, and gain the trust of the public. These benefits far outweigh any costs associated with developing RGSR compliant reports. It is also the right thing to do: reproducibility of research is at the core of science.

e-companion

All the Vensim files used for the example analysis, along with instructions for reproducing the reported results and full documentation using SDM-DOC, are available as an e-companion on the journal's web site.

Notes

1. The literature distinguishes between exact replicability and reproducibility. Replication in the context of simulation modeling would entail the ability for an independent party to generate precisely the same numerical results for a model, down to the last decimal. This is sometimes possible and, when it is, desirable: for example, the World3 model, as a deterministic simulation, should be (and is) replicable (Meadows, 1974). However, given the variations in simulation environments (software, computer hardware), random number generators, and other uncontrollable features of simulation-based research, exact replication is sometimes not possible (for example, when the realizations of the pseudo-random number generators used in stochastic simulations differ across software and hardware). Here we focus on reproducibility of the reported simulation experiments, meaning that substantive results can be reproduced (for example, when the results of reproducing a stochastic simulation yield the same confidence intervals and levels of statistical significance even though the realizations of the random variables in the two sets of simulations may differ).
2. *Additional notes for PORR compliance*: The additional model equations to formulate the optimization problem and optimization payoff and parameter setting files are available in the Appendix. All Vensim files are available in the e-companion. All 45

DOI: 10.1002/sdr

restarts found the same peak in the parameter space. Moreover, simulating the model with different noise seeds, using the optimum parameter value found above, leads to average *PAR*s that are within 1 percent of the goal, thus providing further confidence in the reliability of the results. The optimization, using a compiled version of the model, took 15:12 minutes on a desktop computer with Q9400 Intel Core 2 CPU at 2.66 GHz with 64-bit Windows 7 operating system and 4 GB of RAM.

Acknowledgements

We thank Mohammad Jalali for excellent research assistance, and David Lane, Rogelio Oliva, David Ford and participants in the 2012 International System Dynamics Conference for helpful comments and suggestions.

References

Bass F. 1969. A new product growth model for consumer durables. *Management Science* **15**(5): 215–227.

Bertsekas DP. 2007. *Dynamic Programming and Optimal Control*. Athena Scientific: Belmont, MA.

Bertsekas DP, Tsitsiklis JN. 1996. *Neuro-dynamic Programming*. Athena Scientific: Belmont, MA.

Bonneau G-P, Ertl T, Nielson GM. 2006. *Scientific Visualization: The Visual Extraction of Knowledge from Data*. Springer: Berlin.

Carter J, Moizer J. 2011. Simulating the impact of policy on patrol policing: introducing the emergency service incident model. *System Dynamics Review* **27**(4): 331–357.

Code YLSRD. 2010. Reproducible research addressing the need for data and code sharing in computational science. *Computing in Science and Engineering* **12**(5): 8–12.

Dewald WG, Thursby JG, Anderson RG. 1986. Replication in empirical economics: the Journal-of-Money-Credit-and-Banking Project. *American Economic Review* **76**(4): 587–603.

Donoho DL, Maleki A, Rahman AU, Shahram M, Stodden V. 2009. Reproducible research in computational harmonic analysis. *Computing in Science and Engineering* **11**(1): 8–18.

Forrester JW. 1961. *Industrial Dynamics*. MIT Press: Cambridge, MA.

Gilbert GN, Troitzsch KG. 2005. *Simulation for the Social Scientist*. Open University Press: Maidenhead, UK.

Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz S, Huse G, Huth A, Jepsen JU, Jørgensen C, Mooij WWM, Müller B, Pe'er G, Piou C, Railsback SF, Robbins AM, Robbins MM, Rossmanith E, Rüger N, Strand E, Souissi S, Stillman RA, Vabø R, Visser U, DeAngelis DL. 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modeling* **198**: 115–126.

Groesser S, Tschupp R. 2012. Quality of system dynamics research: evidence and recommendations for model-based studies. In *Proceedings of 2012 International System Dynamics Conference*. St Gallen, Switzerland: System Dynamics Society.

Hubbard R, Vetter DE. 1996. An empirical comparison of published replication research in accounting, economics, finance, management, and marketing. *Journal of Business Research* **35**(2): 153–164.

Ioannidis JPA. 2005. Why most published research findings are false. *Plos Medicine* **2**(8): 696–701.

Jackson RHF, Boggs PT, Nash SG, Powell S. 1991. Guidelines for reporting results of computational experiments: report of the adhoc committee. *Mathematical Programming* **49**(3): 413–425.

Kim DH, Kim DH. 1997. A system dynamics model for a mixed-strategy game between police and driver. *System Dynamics Review* **13**(1): 33–52.
Koenker R, Zeileis A. 2009. On reproducible econometric research. *Journal of Applied Econometrics* **24**(5): 833–847.
Laine C, Goodman SN, Griswold ME, Sox HC. 2007. Reproducible research: moving toward research the public can really trust. *Annals of Internal Medicine* **146**(6): 450–453.
Lane DC. 2008. The emergence and use of diagramming in system dynamics: a critical account. *Systems Research and Behavioral Science* **25**(1): 3–23.
Le Novere N, Finney A, Hucka M, *et al*. 2005. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology* **23**(12): 1509–1515.
Lee CY, Bard J, Pinedo M, Wilhelm WE. 1993. Guidelines for reporting computational results in IIE Transactions. *IIE Transactions* **25**(6): 121–123.
Martinez-Moyano I. 2012. Documentation for model transparency. *System Dynamics Review* **28**(2): 199–208.
Matthews JR, Matthews RW. 2008. *Successful Scientific Writing: A Step-by-Step Guide for the Biological and Medical Sciences*. Cambridge University Press: Cambridge, UK.
McCullough BD, McGeary KA, Harrison TD. 2006. Lessons from the JMCB archive. *Journal of Money Credit and Banking* **38**(4): 1093–1107.
McCullough BD, McGeary KA, Harrison TD. 2008. Do economics journal archives promote replicable research? *Canadian Journal of Economics—Revue Canadienne D'Economique* **41**(4): 1406–1420.
Meadows DL. 1974. *Dynamics of Growth in a Finite World*. Wright-Allen Press: Cambridge, MA.
Mojtahedzadeh M. 2012. Consistency in explaining model behavior based on its feedback structure. *System Dynamics Review* **27**(4): 358–373.
Morecroft JDW, Lane DC, Viita PS. 1991. Modeling growth strategy in a biotechnology startup firm. *System Dynamics Review* **7**(2): 93–116.
Morin A, Urban J, Adams PD, Foster I, Sali A, Baker D, Sliz P. 2012. Shining light into black boxes. *Science* **336**: 159–160.
Nussinov R, Bourne PE (eds). 2008. PLoS Computational Biology. Available: www.ploscompbiol.org/static/policies.action [16 September 2012].
Peng RD. 2011. Reproducible research in computational science. *Science* **334**: 1226–1227.
Press WH. 1992. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press: Cambridge, UK.
Rahmandad H, Sibdari S. 2012. Joint pricing and openness decisions in software markets with reinforcing loops. *System Dynamics Review* **28**(3): 209–229.
Schwab M, Karrenbach M, Claerbout J. 2000. Making scientific computations reproducible. *Computing in Science and Engineering* **2**(6): 61–67.
Sterman J. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill/Irwin: Boston, MA.
Sterman JD, Henderson R, Beinhocker ED, Newman LI. 2007. Getting big too fast: strategic dynamics with increasing returns and bounded rationality. *Management Science* **53**(4): 683–696.
Tufte ER. 2001. *The Visual Display of Quantitative Information*. Graphics Press: Cheshire, CT.
Verma IM (ed.). 2012. Proceedings of National Academies of Sciences, U.S.A.. Available: www.pnas.org/site/misc/iforc.shtml [16 September 2012].
Waltemath D, Adams R, Beard DA, *et al*. 2011. Minimum information about a simulation experiment (MIASE). *Plos Computational Biology* **7**(4): e1001122.

Appendix: The Vensim optimization details for section on optimization reporting guidelines

Additional equations for optimization section

All model equations reported in Table 2 remain valid; however, variables (but not model constants) are subscripted to be replicated for 1000 instances of subscript “Itr”. Additional equations for calculating the optimization payoff (PF) are added as follows:

$ltr : (i1 - i1000)$ Dimensionless

The subscript range Itr is defined to include 1000 members, i1 ... i1000.

$APAR = Sum(PAR[Itr!])/Elmcount(Itr)$ Person/Year

The average PAR value is calculated by summing the PAR over all Itr instances and dividing the result by the number of elements of Itr (1000 in this case).

$PF = ifthenelse(Time = FINALTIME, 1, 0) * (APAR - r)^2 Person2/Year2$

Payoff is calculated by comparing Mean PAR and Desired PAR at the end of simulation.

$r = 600000$ Person/Year

The desired Peak Adoption Rate is set to 600,000 people per year.

Text of optimization settings file (.vpd)

```
*P
Payoff PF/-1
```

Text of optimization settings file (.voc)

```
: OPTIMIZER = Powell
: SENSITIVITY = Off
: MULTIPLE_START = Random
: RANDOM_NUMER = Linear
: OUTPUT_LEVEL = On
: TRACE = Off
: MAX_ITERATIONS = 1000
: RESTART_MAX = 0
: PASS_LIMIT = 2
: FRACTIONAL_TOLERANCE = 3e-005
: TOLERANCE_MULTIPLIER = 21
: ABSOLUTE_TOLERANCE = 1
: SCALE_ABSOLUTE = 1
: VECTOR_POINTS = 25
: 0 < =Advertising Effectiveness a < =2
```


DOI: 10.1002/sdr