www.palgrave-journals.com/jos/

Verification and validation of simulation models

RG Sargent*

Syracuse University, Syracuse, USA

Verification and validation of simulation models are discussed in this paper. Three approaches to deciding model validity are described, two paradigms that relate verification and validation to the model development process are presented, and various validation techniques are defined. Conceptual model validity, model verification, operational validity, and data validity are discussed. A way to document results is given, and a recommended procedure for model validation is presented.

Journal of Simulation (2013) **7,** 12–24. doi:10.1057/jos.2012.20; published online 7 December 2012

Keywords: simulation; verification; validation

1. Introduction

Simulation models are used for a variety of purposes such as in the design of systems, in the development of system operating policies, and in research to develop system understandings. The users of these models, the decision makers using information obtained from the results of these models, and the individuals affected by decisions based on these models are all rightly concerned with whether a model and its results are 'correct' for its use. This concern is addressed through model verification and validation, which is part of the model development process. Model verification is defined as 'ensuring that the computer program of the computerized model and its implementation are correct'. Model validation is defined as the 'substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model'. We discuss simulation model verification and validation in this paper, focusing primarily on simulation models that predict system behaviours such as system outputs. A related topic is model credibility and this is briefly discussed. Model credibility is concerned with developing in (potential) users the confidence they require in order to use a model and in the information derived from that model.

A model should be developed for a specific purpose (or application) and its validity determined with respect to that purpose. If the purpose of a model is to answer a variety of questions, the validity of the model needs to be determined with respect to each question. Numerous sets of experimental conditions are usually required to define the domain of a model's intended applicability. A model may be valid for one set of experimental conditions and invalid in another. A model is considered valid for a set of experimental conditions if the model's accuracy is within its *acceptable range of accuracy*, which is the accuracy required of the model for its intended purpose. This usually requires that the model's output variables of interest (ie, the model variables used in answering the questions that the model is being developed to answer) be identified and then their acceptable range of accuracy specified. A model's acceptable range of accuracy should be specified prior to starting the development of the model or very early in the model development process. If the variables of interest are random variables, then properties and functions of the random variables such as means and variances are usually what is of primary interest and are what is used in determining model validity. Several versions of a model are usually developed prior to obtaining a satisfactory valid model. The substantiation that a model is valid, that is, performing model verification and validation, is generally considered to be a process and is usually part of the (total) model development process.

It is often too costly and time-consuming to determine that a model is *absolutely* valid over the complete domain of its intended applicability. Instead, tests and evaluations are conducted until sufficient confidence is obtained that a model can be considered valid for its intended application (Sargent, 1982, 1984a). If a test determines that a model does not have sufficient accuracy for any one of the sets of experimental conditions, then the model is invalid. However, determining that a model has sufficient accuracy for numerous experimental conditions does *not guarantee* that a model is valid everywhere in its applicable domain. Figure 1 contains two relationship curves regarding confidence that a model is valid (Confidence in Model) over the range of 0–100% as they would occur in most cases (Anshoff and Hayes, 1973). The cost curve (and a similar relationship holds for the amount of time) of performing model validation shows that cost increases at an increasing rate as the confidence in the model increases. The value curve shows

**Correspondence: RG Sargent, Department of Electrical Engineering and Computer Science, L.C. Smith College of Engineering and Computer Science, Syracuse University, Syracuse, NY 13244, USA.*
E-mail: rsargent@syr.edu

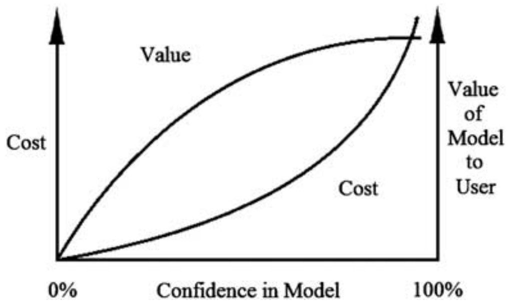


Figure 1 Confidence that model is valid.

that the value of a model to a user increases as the confidence in a model increases but at a decreasing rate. (In some cases, these curves may have a different shape for the lower confidence range but would usually be similar to what is shown in Figure 1 for the upper confidence range.) The cost of model validation is usually quite significant, especially when extremely high model confidence is required.

The remainder of this paper is organized as follows: Section 2 presents the three decision-making approaches used in deciding model validity, Section 3 describes two paradigms used in verification and validation, and Section 4 defines validation techniques. Sections 5, 6, 7, and 8 discuss data validity, conceptual model validity, computerized model verification, and operational validity, respectively. Section 9 describes a way of documenting results, Section 10 gives a recommended validation procedure, and Section 11 presents the summary.

2. Decision-making approaches

There are three basic decision-making approaches for deciding whether a simulation model is valid and each approach uses a different decision maker. All of the approaches require the model development team to conduct verification and validation as part of the model development process, which is discussed in Section 3. One decision-making approach, and a frequently used one, is for the model development team itself to make the decision as to whether a simulation model is valid. The decision is based on the results of the various tests and evaluations conducted as part of the model development process. It is usually better, however, to use one of the next two decision-making approaches, depending on which situation applies.

A better decision-making approach is to have the user(s) of a simulation model decide the validity of the model. In this approach, the users of the simulation model are heavily involved with the model development team when the team is conducting verification and validation of the model and the users determine whether the model is satisfactory in each phase of verification and validation. This approach is generally used with a model development team whose size is not large. In addition, this approach aids in model credibility.

Another decision-making approach, usually called 'independent verification and validation' (IV&V), uses a third party to decide whether the simulation model is valid. The third party (the IV&V team) is independent of both the simulation development team(s) and the model sponsor/ user(s). The IV&V approach is generally used with the development of large-scale simulation models, whose development usually involves several teams. The IV&V team needs to have a *thorough* understanding of the intended purpose(s) of the simulation model in order to conduct IV&V. There are two common ways in which the IV&V team conducts IV&V: (a) IV&V is conducted concurrently with the development of the simulation model and (b) IV&V is conducted after the simulation model has been developed.

In the concurrent way of conducting IV&V, the model development team(s) gives their model verification and validation test results to the IV&V team as the simulation model is being developed. The IV&V team evaluates these results and provides feedback to the model development team with regard to whether the model verification and validation is satisfying the model requirements and, when not, what the difficulties are. When conducting IV&V in this way, the development of a simulation model should not progress to the next stage of development until the model has satisfied the verification and validation requirements in its current stage. It is the author's opinion that this is the better of the two ways to conduct IV&V.

When IV&V is conducted after the simulation model has been completely developed, the evaluation performed by the IV&V team can range from simply evaluating the verification and validation conducted by the model development team to performing a separate thorough verification and validation effort themselves. Wood (1986) describes experiences over this range of evaluation by a third party on energy models. One conclusion that Wood makes is that performing a complete IV&V effort after the model has been completely developed is both extremely costly and time-consuming, especially for what is obtained. This author's view is that if IV&V is going to be conducted on a completed simulation model then it is usually best to *only* evaluate the verification and validation that has already been performed.

The IV&V approach is also useful for model credibility. When verification and validation is conducted by an independent (third) party and they conclude that the simulation model is valid, there is a much greater likelihood that others will accept the model as valid and results from the model as being 'correct'. Cases where this decision-making approach is helpful are: (i) when the problem associated with the model has a high cost or involves a high-risk situation and (ii) when public acceptance of results based on the model is desired.

3. Paradigms

In this section, we present and discuss two paradigms that relate verification and validation to the model development process. There are two common ways to view this relationship. One way uses a simple view and the other uses a complex view. Banks *et al* (1988) reviewed work using both of these ways and concluded that the simple way more clearly illuminates model verification and validation. We present one paradigm for each way developed by this author. The paradigm of the simple way is presented first, is this author's preferred paradigm, and is the paradigm used for much of the discussion in this paper.

Consider the simplified version of the model development process in Figure 2 (Sargent, 1981). The *problem entity* is the system (real or proposed), idea, situation, policy, or phenomena to be modelled; the *conceptual model* is the mathematical/logical/graphical representation (mimic) of the problem entity developed for a particular study; and the *computerized model* is the conceptual model implemented on a computer. The conceptual model is developed through an *analysis and modelling phase*, the computerized model is developed through a *computer programming and implementation phase*, and inferences about the problem entity are obtained by conducting computer experiments on the computerized model in the *experimentation phase*.

We now relate model verification and validation to this simplified version of the model development process (see Figure 2). *Conceptual model validation* is defined as determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is 'reasonable' for the intended purpose of the model. *Computerized model verification* is defined as assuring that the computer programming and implementation of the conceptual model are correct. *Operational validation* is defined as determining that the model's output behaviour has a satisfactory range of accuracy for the model's intended purpose over the domain of the model's intended applicability. *Data validity* is defined as ensuring that the data necessary for model building, model evaluation and testing, and conducting the model experiments to solve the problem are adequate and correct.

An iterative process is used to develop a valid simulation model (Sargent, 1984a). A conceptual model is developed followed by conceptual model validation. This process is repeated until the conceptual model is satisfactory. Next the computerized model is developed from the conceptual model followed by computerized model verification. This process is repeated until the computerized model is satisfactory. Next, operational validity is conducted on the computerized model. Model changes required by conducting operational validity can be in either the conceptual model or in the computerized model. Verification and validation must be performed again when any model change is made. This process is repeated until a valid simulation model is obtained. Several versions of a model are usually developed before obtaining a valid simulation model.

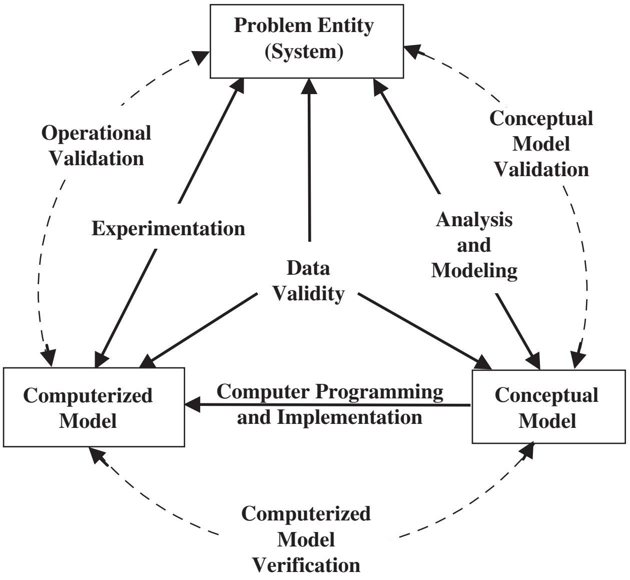


Figure 2 Simplified version of the model development process.

A detailed way of relating verification and validation to developing simulation models and system theories is shown in Figure 3 (Sargent, 2001b). This paradigm shows the processes of developing system theories and simulation models and relates verification and validation to both of these processes. (We note that Nance and Arthur (2006) use this paradigm as a Life-cycle Model.)

This paradigm shows a Real World and a Simulation World. We first discuss the Real World. There exists some *system or problem entity* in the real world, an understanding of which is desired. *System theories* describe the characteristics and the causal relationships of the system (or problem entity) and possibly its behaviour (including data). *System data and results* are obtained by conducting experiments (*experimenting*) on the system. System theories are developed by *abstracting* what has been observed from the system and by *hypothesizing* from the system data and results. If a simulation model exists of this system, then *hypothesizing* of system theories can also be done from simulation data and results. System theories are validated by performing *theory validation*. Theory validation involves the comparison of system theories against system data and results over the domain the theory is applicable for to determine whether there is agreement. The process of developing valid system theories usually requires numerous experiments to be conducted on the real system.

We now discuss the Simulation World, which shows a slightly more complex model development process than the previous paradigm shown in Figure 2. A simulation model should only be developed for a set of well-defined objectives. The *conceptual model* is the mathematical/logical/graphical

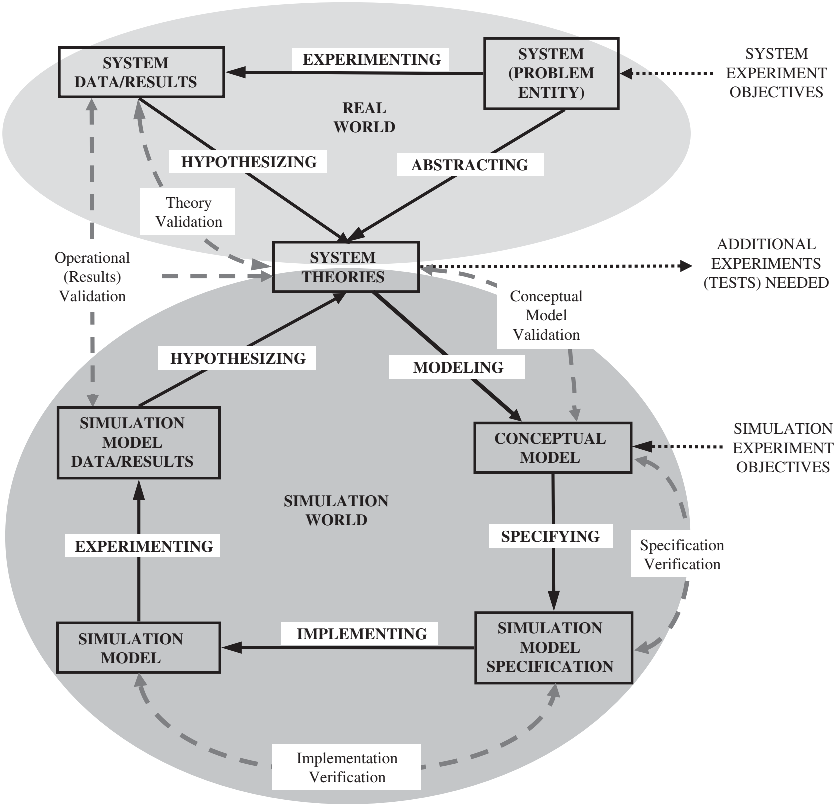

SYSTEM DATA/RESULTS
EXPERIMENTING
SYSTEM (PROBLEM ENTITY)
SYSTEM EXPERIMENT OBJECTIVES
REAL WORLD
HYPOTHESIZING
ABSTRACTING
Theory Validation
Operational (Results) Validation
SYSTEM THEORIES
ADDITIONAL EXPERIMENTS (TESTS) NEEDED
Conceptual Model Validation
HYPOTHESIZING
MODELING
SIMULATION MODEL DATA/RESULTS
CONCEPTUAL MODEL
SIMULATION EXPERIMENT OBJECTIVES
SIMULATION WORLD
EXPERIMENTING
SPECIFYING
Specification Verification
SIMULATION MODEL
IMPLEMENTING
SIMULATION MODEL SPECIFICATION
Implementation Verification

Figure 3 Real world and simulation world relationships with verification and validation.

representation (mimic) of the system developed for the objectives of a particular study. The *simulation model specification* is a written detailed description of the software design and specification for programming and implementing the conceptual model on a particular computer system. The *simulation model* is the conceptual model running on a computer system such that experiments can be conducted on the simulation model. The *simulation model data and results* are the data and results from experiments conducted (*experimenting*) on the simulation model. The conceptual model is developed by *modelling* the system for the objectives of the simulation study using the understanding of the system contained in the system theories. The simulation model specification is developed by *specifying* in writing the software design and the programming and implementing specifications of the conceptual model for the targeted computer system. The simulation model is obtained by *implementing* the model on the specified computer system, which includes programming the conceptual model whose specifications are contained in the simulation model specification. Inferences about the system are made from data obtained by conducting computer experiments (*experimenting*) on the simulation model.

Conceptual model validation is defined as determining that the theories and assumptions underlying the conceptual model are consistent with those in the system theories and that the model representation of the system is 'reasonable' for the intended purpose of the simulation model. *Specification verification* is defined as assuring that the software design and the specification for programming and implementing the conceptual model on the specified computer system is satisfactory. *Implementation verification* is defined as assuring that the simulation model has been implemented according to the simulation model specification. *Operational validation* is defined as determining that the model's output behaviour has a satisfactory range of accuracy for the model's intended purpose over the domain of the model's intended applicability.

This paradigm shows the relationships used in developing valid system theories and valid simulation models. Both of these are accomplished through iterative processes. To develop valid system theories, which are usually for a specific purpose, the system is first observed and then abstraction is performed from what has been observed to develop proposed system theories. These proposed theories are tested for correctness by conducting experiments on the system to obtain data and results to compare against the proposed system theories. New proposed system theories may be hypothesized from the system data and the comparisons made, and also possibly from abstractions performed on additional system observations. These new proposed theories will require new experiments to be conducted on the system to obtain data to evaluate the correctness of these proposed system theories. This process repeats itself until a satisfactory set of validated system theories has been obtained. Similarly, new proposed system theories can be hypothesized from simulation data and results if a simulation model exists of the system.

It is possible that proposed system theories cannot be tested for correctness, which is done by comparing them to system data, and therefore they cannot be validated. This occurs if experiments cannot be performed on an existing system or if it is too costly to do so. Proposed system theories may also be developed for non-existing systems such as in the design of a new system or of a modification to an existing system. These proposed system theories are developed from an understanding of how such a system will operate. They cannot be validated because the system does not exist to conduct experiments on. If propose system theories cannot be validated, they remain as *proposed system theories*.

To develop a valid simulation model, an iterative process similar to the one for the previous paradigm (given in Figure 2) is performed. The model development team develops the conceptual model using existing validated system theories, existing proposed system theories, new proposed system theories they develop, or some combination of the these three types of system theories. A simulation model specification is developed for the conceptual model that is used to implement a simulation model on the computer. Experiments are performed on the simulation model to generate data for use in operational validation. Several versions of a model are usually developed prior to obtaining a valid simulation model.

4. Validation techniques

This section describes validation techniques and tests commonly used in model verification and validation. Most of the techniques described here are found in the literature, although some may be described slightly differently. They can be used either subjectively or objectively. By 'objectively', we mean using some type of mathematical procedure or statistical test, for example hypothesis tests or confidence intervals. A combination of techniques is generally used. These techniques are used for verifying and validating the submodels and the overall model.

Animation: The model's operational behaviour is displayed graphically as the model moves through time. For example, the movements of parts through a factory during a simulation run are shown graphically. (Usually, only a relatively short time interval can be observed, which may result in not all behaviours being observed.)

Comparison to other models: Various results (eg, outputs) of the simulation model being validated are compared to results of other (valid) models. For example (1) simple cases of a simulation model are compared to known results of analytic models, and (2) the simulation model is compared to other validated simulation models.

Data relationship correctness: Data relationship correctness requires data to have the proper values regarding relationships that occur within a type of data, and between and among different types of data. For example, are the values of data collected on a system or model correct for some known relationship within some type of data such as an inventory balance relationship or a dollar relationship?

Degenerate tests: The degeneracy of the model's behaviour is tested by appropriate selection of values of the input and internal parameters. For example, does the average number in the queue of a single server continue to increase over time when the arrival rate is larger than the service rate?

Event validity: The 'events' of occurrences of the simulation model are compared to those of the real system to determine whether they are similar. For example, compare the number of fires in a fire department simulation to the actual number of fires.

Extreme condition test: The model structure and outputs should be plausible for any extreme and unlikely combination of levels of factors in the system. For example, if in-process inventories are zero, production output should usually be zero.

Face validity: Individuals knowledgeable about the system are asked whether the model and/or its behaviour are reasonable. For example, is the logic in the conceptual model correct and are the model's input–output relationships reasonable?

Historical data validation: If historical data exist (eg, data collected on a system specifically for building and testing a model), part of the data is used to build the model and the remaining data are used to determine (test) whether the model behaves as the system does.

Internal validity: Several replications (runs) of a stochastic model are made to determine the amount of (internal) stochastic variability in the model. A large amount of variability among the replications may cause the model's results to be questionable, and if typical of the problem entity may question the appropriateness of the policy or system being studied.

Multistage validation: Naylor and Finger (1967) proposed combining the philosophy of science methods of rationalism, empiricism, and positive economics into a multistage process of validation. This validation method consists of (1) developing the model on theory, observations, and general knowledge;

(2) validating the model's assumptions where possible by empirically testing them; and (3) comparing (testing) the input–output relationships of the model to the real system.

Operational graphics: Values of various performance measures, for example the number in queue and percentage of servers busy, are shown graphically as the model runs through time; that is, the dynamical behaviours of performance indicators are visually displayed as the simulation model runs through time to ensure that the performance measures and the model are behaving correctly.

Parameter variability–sensitivity analysis: This technique consists of changing the values of the input and internal parameters of a model to determine the effect upon the model's behaviour or output. The same relationships should occur in the model as in the real system. This technique can be used qualitatively—directions only of outputs—and quantitatively—both directions and (precise) magnitudes of outputs. Those parameters that are sensitive, that is, cause significant changes in the model's behaviour or output, should be made sufficiently accurate prior to using the model. (This may require iterations in model development.)

Philosophy of science methods: The three philosophy of science methods are *rationalism*, *empiricism*, and *positive economics*. Rationalism requires a model to be logically developed (correctly) from a set of clearly stated assumptions. Empiricism requires every model assumption and outcome to be empirically validated. Positive economics requires only that the model outcomes are correct and is not concerned with a model's assumptions or structure (causal relationships or mechanisms).

Predictive validation: The model is used to predict (forecast) the system's behaviour, and then comparisons are made between the system's behaviour and the model's forecast to determine whether they are the same. The system data may come from an operational system or be obtained by conducting experiments on the system, for example field tests.

Structured walkthrough: The entity under review is formally presented usually by the developer to a peer group to determine the entity's correctness. An example is a formal review of computer code by the code developer explaining the code line by line to a set of peers to determine the code's correctness.

Trace: The behaviour of a specific type of entity in a model is traced (followed) through the model to determine whether the model's logic is correct and if the necessary accuracy is obtained. (Most current simulation software provide for trace capability making the use of traces relatively simple.)

Turing test: Individuals who are knowledgeable about the operations of the system being modelled are asked whether they can discriminate between system and model outputs. (Schruben (1980) contains statistical tests for the Turing test.)

5. Data validity

We discuss data validity, even though it is often not considered to be part of model validation, because it is usually difficult, time-consuming, and costly to obtain appropriate, accurate, and sufficient data, and data problems are often the reason that attempts to validate a model fail. Data are primarily used for three purposes: for building the conceptual model, for validating the model, and for performing experiments with the validated model. In model validation, we are usually concerned only with data for the first two purposes.

Data needed on the problem entity for building the conceptual model include data for identifying and developing appropriate system theories, developing mathematical and logical relationships, estimating model parameter values, and developing and testing the model assumptions. Behavioural data are needed on the problem entity to be used in the operational validity step of comparing the problem entity's behaviour with the model's behaviour. (Usually, these data are system input/output data.) If problem entity behaviour data are not available, high model confidence usually cannot be obtained because sufficient operational validity cannot be achieved.

The concerns with data are that appropriate, accurate, and sufficient data are available, and all data transformations, such as data disaggregation, are made correctly. Unfortunately, there is not much that can be done to determine whether the data are correct. One should develop and use good procedures for (1) collecting and maintaining data, (2) testing the collected data using techniques such as data relationship correctness on known data relationships, and (3) screening the data for outliers *and* determining if the outliers are correct. (*Note*: Outliers should always be evaluated and, if correct, the reason for them occurring should be incorporated into the simulation model.) If the amount of data is large, a database of the data should be developed and maintained.

6. Conceptual model validation

Conceptual model validity determines that (1) the theories and assumptions underlying the conceptual model are correct and (2) the model's representation of the problem entity and the model's structure, logic, and mathematical and causal relationships are 'reasonable' for the intended purpose of the model. The theories and assumptions underlying the model should be tested using mathematical analysis and statistical methods on problem entity data. Examples of theories and assumptions are linearity, independence of data, and arrivals to the system follow a Poisson process. Examples of applicable statistical methods are fitting distributions to data, estimating parameter values from the data, and plotting data to determine whether the data are stationary. In addition, all theories used should be reviewed to ensure they were applied correctly. For example, if a Markov chain is used, does the system have the Markov property, and are the states and transition probabilities correct?

A conceptual model may be a single model or an overall model with submodels. Each model, whether a single model, an overall model, or a submodel, must be evaluated to determine whether it is reasonable and correct for the intended purpose of the conceptual model. This should include determining whether the appropriate detail and aggregate relationships have been used for the model's intended purpose, and also whether appropriate structure, logic, and mathematical and causal relationships have been used. The primary validation techniques used for these evaluations are face validation, structured walkthroughs, and traces. Face validation has experts on the problem entity evaluate the conceptual model, which may be a flowchart model, graphical model (Sargent, 1986), or a set of model equations, to determine whether it is correct and reasonable for its purpose. Structured walkthrough is having the conceptual model developer formally explain the model in detail to a set of peers for them to determine the conceptual model correctness. The use of traces is the tracking of entities through each submodel and the overall model to determine whether the logic is correct and whether the necessary accuracy is maintained. If errors are found in the conceptual model, it must be revised and conceptual model validation performed again.

7. Computerized model verification

Computerized model verification ensures that the computer programming and implementation of the conceptual model are correct. The major factor affecting verification is whether a simulation language or a higher-level programming language such as FORTRAN, C, or C++ is used. The use of a special-purpose simulation language generally will result in having fewer errors than if a general-purpose simulation language is used, and using a general-purpose simulation language will generally result in having fewer errors than if a general purpose higher-level programming language is used. (The use of a simulation language also usually increases the model execution times and reduces both the programming time required and the amount of flexibility.)

When a simulation language is used, verification is primarily concerned with ensuring that an error-free simulation language has been used, that the simulation language has been properly implemented on the computer, that a tested (for correctness) pseudo random number generator has been properly implemented, and that the model has been programmed correctly in the simulation language. The primary techniques used to determine that the model has been programmed correctly are structured walkthroughs and traces.

If a higher-level programming language has been used, then the computer program should have been designed, developed, and implemented using techniques found in software engineering. (These include such techniques as object-oriented design, structured programming, and program modularity.) In this case, verification is primarily concerned with determining that the simulation functions (eg, the time-flow mechanism, pseudo random number generator, and random variate generators) *and* the computerized (simulation) model have been programmed and implemented correctly.

There are two basic approaches for testing simulation software: static testing and dynamic testing (Fairley, 1976). In static testing, the computer program is analysed to determine whether it is correct by using such techniques as structured walkthroughs, correctness proofs, and examining the structure properties of the program. In dynamic testing, the computer program is executed under different conditions and the values obtained (including those generated *during* the execution) are used to determine whether the computer program and its implementations are correct. The techniques commonly used in dynamic testing are traces, investigations of input–output relationships using different validation techniques, data relationship correctness, and reprogramming critical components to determine whether the same results are obtained. If there are a large number of variables, one might aggregate the numerical values of some of the variables to reduce the number of tests needed or use certain types of design of experiments (Kleijnen, 2008).

It is necessary to be aware while checking the correctness of the computer program and its implementation that errors found may be caused by the data, the conceptual model, the computer program, or the computer implementation. (See Whitner and Balci (1989) for a detailed discussion on simulation model verification.)

8. Operational validity

Operational validation is determining whether the simulation model's output behaviour has the accuracy required for the model's intended purpose over the domain of the model's intended applicability. This is where much of the validation testing and evaluation take place. Since the simulation model is used in operational validation, any deficiencies found may be caused by what was developed in any of the earlier steps in developing the simulation model including developing the system's theories or having invalid data.

All of the validation techniques discussed in Section 4 are applicable to operational validity. Which techniques and whether to use them objectively or subjectively must be decided by the model development team and the other interested parties. The major attribute affecting operational validity is whether the problem entity (or system) is observable, where observable means it is possible to collect data on the operational behaviour of the problem entity. Table 1

Table 1 Operational validity classification

Decision approach	*Observable system*	*Non-observable system*
Subjective approach	● Comparison using graphical displays ● Explore model behaviour	● Explore model behaviour ● Comparison to other models
Objective approach	● Comparison using statistical tests and procedures	● Comparison to other models using statistical tests

gives a classification of the validation techniques used in operational validity based on the decision approach and system observable. 'Comparison' means comparing the simulation model output behaviour to either the system output behaviour or another model output behaviour using graphical displays or statistical tests and procedures. 'Explore model behaviour' means to examine the output behaviour of the simulation model using appropriate validation techniques, including parameter variability–sensitivity analysis. Various sets of experimental conditions from the domain of the model's intended applicability should be used for both comparison and exploring model behaviour.

To obtain a *high* degree of confidence in a simulation model and its results, comparisons of the model's and system's output behaviours for several different sets of experimental conditions are usually required. Thus, if a system is not observable, which is often the case, it is usually not possible to obtain an extremely high degree of confidence in a model of it. In this situation, the model output behaviour(s) should be explored as thoroughly as possible and comparisons made to other valid models whenever possible.

We note that there are two methods to provide simulation model inputs. One method is sampling from either empirical or theoretical input distributions and the other method is the use of system data traces, which is often referred to as trace-driven simulation (Law, 2007). In the latter method, data traces are collected on the system inputs and then these traces are used to 'drive' the simulation model instead of sampling from the input distributions. (Some specialized methods for validation of trace-driven simulation have been developed; see, eg, Kleijnen *et al* (2001) and Sargent (2010).)

8.1. Explore model behaviour

The simulation model output behaviour can be explored either qualitatively or quantitatively. In qualitative analysis, the directions of the output behaviours are examined and also possibly whether the magnitudes are 'reasonable'. In quantitative analysis, both the directions and the precise magnitudes of the output behaviours are examined. Experts on a system, often called subject matter experts, usually know the directions and often know the 'general values' of the magnitudes of the output behaviours. Many of the validation techniques given in Section 4 can be used for model exploration. Parameter variability–sensitivity analysis should usually be used. Graphs of the output data discussed in the subsection 'Graphical comparisons of data' below can be used to display the simulation model output behaviour. A variety of statistical approaches can be used in performing model exploration including metamodelling and design of experiments. (See Kleijnen (1999) for further discussion on the use of statistical approaches.) Numerous sets of experimental frames should be used in performing model exploration.

For non-observable systems, exploring the output behaviour of the models is the method primarily used to determine model validity. Experts on the system can make subjective decisions on whether the model outputs are reasonable. If other models exist of the system, the outputs of these models can be compared either subjectively or objectively to the outputs of the model being evaluated for validity. On the basis of all of the evaluations, a decision is made regarding the validity of the simulation model.

8.2. Comparisons of output behaviours

There are three basic approaches used in comparing the simulation model output behaviour to either the system output behaviour or another model output behaviour: (1) the use of hypothesis tests to make an objective decision, (2) the use of confidence intervals to make an objective decision, and (3) the use of graphs to make a subjective decision. It is preferable to use hypothesis tests or confidence intervals for the comparisons because these provide objective decisions. Unfortunately, it is often not possible in practice to use either one of these two approaches because (a) the statistical assumptions required cannot be satisfied or only with great difficulty (assumptions usually required are data independence and normality) and/or (b) there is an insufficient quantity of system data available, which causes the statistical results to be 'meaningless' (eg, the length of a confidence interval developed in the comparison of the system and simulation model means is too large for any practical usefulness). As a result, the use of graphs is the most commonly used approach for operational validity. Each of these three approaches is discussed below when using system data. These approaches are used in the same way when using data from another model instead of system data. (We note that there are advanced statistical methods that can be used for comparisons of output data such as the use of distribution-free statistical tests. These methods can also have difficulties in their use for validating simulation models such as requiring a large amount of system data to obtain useful results.)

Hypothesis tests. Hypothesis tests can be used in the comparison of means, variances, distributions, and time series of the output variables of a model and a system for each set of experimental conditions to determine whether the simulation model's output behaviour has a satisfactory range of accuracy for its intended application. The accuracy required of a model is usually specified as a *range* for the difference between the model variable and the corresponding system variable of interest (eg, the difference between the means of a model variable and the corresponding system variable).

The first step in hypothesis testing is to state the hypotheses to be tested:

$\boldsymbol{H_0}$: *Model is valid for the acceptable range of accuracy under the set of experimental conditions.*

$\boldsymbol{H_1}$: *Model is invalid for the acceptable range of accuracy under the set of experimental conditions.*

Two types of errors are possible in testing hypotheses. The first, or type I error, is rejecting the validity of a valid model and the second, or type II error, is accepting the validity of an invalid model. The probability of a type-I error, α, is called *model builder's risk*, and the probability of a type-II error, β, is called *model user's risk* (Balci and Sargent, 1981). In model validation, the model user's risk is extremely important and must be kept small. Thus, both type-I and type-II errors must be carefully considered when using hypothesis testing for model validation.Statistical hypothesis tests usually test for a single point. Since the acceptable range of accuracy for each model variable of interest is usually specified as a range, a hypothesis test that uses a range is desired. Recently, a new statistical procedure has been developed for comparisons of model and system outputs using hypothesis tests when the acceptable range of accuracy is specified as a range (Sargent, 2010). This new statistical procedure is applied at each experimental condition to determine whether the model is valid for that experimental condition. Both type-I and type-II errors are considered through the use of the operating characteristic curve (Hines *et al*, 2003; Johnson *et al*, 2010). Furthermore, the model builder's and the model user's risk curves can be developed using this new procedure. This procedure provides for (i) a trade-off to be made between the two risks for fixed sample sizes and (ii) trade-offs among the two risks and different sample sizes for variable sample sizes. See Sargent (2010) for details of performing this new procedure.

Confidence intervals. Confidence intervals (c.i.) and simultaneous confidence intervals (s.c.i.) can be obtained for the *differences* between means, variances, and distributions of different simulation models and system output variables for each set of experimental conditions. These c.i. and s.c.i. can be used as the model range of accuracy for model validation, where the model range of accuracy is the conference interval or region (for the s.c.i.) around the estimated difference between some function (eg, the mean) of the model and system output variable being evaluated.

To construct the model range of accuracy, a statistical procedure containing a statistical technique and a method of data collection must be developed for each set of experimental conditions and for each variable of interest. The statistical techniques used can be divided into two groups: (1) univariate statistical techniques and (2) multivariate statistical techniques. The univariate techniques can be used to develop c.i., and with the use of the Bonferroni inequality (Law, 2007) s.c.i. The multivariate statistical techniques can be used to develop s.c.i. Both parametric and non-parametric techniques can be used.

The method of data collection must satisfy the underlying assumptions of the statistical technique being used. The standard statistical techniques and data collection methods used in simulation output analysis (Law, 2007; Banks *et al*, 2010) can be used in developing the model range of accuracy, for example the methods of replication and (non-overlapping) batch means.

It is usually desirable to construct the model range of accuracy with the lengths of the c.i. and s.c.i. as short as possible. The shorter the lengths, the more useful and meaningful the model range of accuracy will usually be. The c.i. and s.c.i. lengths (1) are affected by the values of confidence levels, variances of the model and system output variables, and sample sizes, and (2) can be made shorter by decreasing the confidence levels or increasing the sample sizes. A trade-off needs to be made among the sample sizes, confidence levels, and estimates of the length of the model range of accuracy. Trade-off curves can be constructed to aid in the trade-off analysis.

Details on the use of c.i. and s.c.i. for operational validity, including a general methodology, are contained in Balci and Sargent (1984b).

Graphical comparisons of data. Data of the simulation model and system output variables are graphed for various sets of experimental conditions to determine whether the model's output behaviour has sufficient accuracy for the model's intended purpose. Three types of graphs are used: histograms, box (and whisker) plots, and behaviour graphs (see Figures 4, 5, and 6 for an example of each one). (Behaviour graphs use scatter plots to show relationships between two measures. Additional behaviour graphs with explanations are contained in Sargent, 1996a.) A variety of graphs can be developed that use different types of (1)

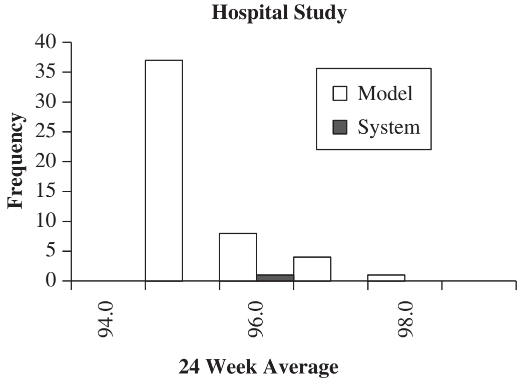


Figure 4 Histogram of hospital data.

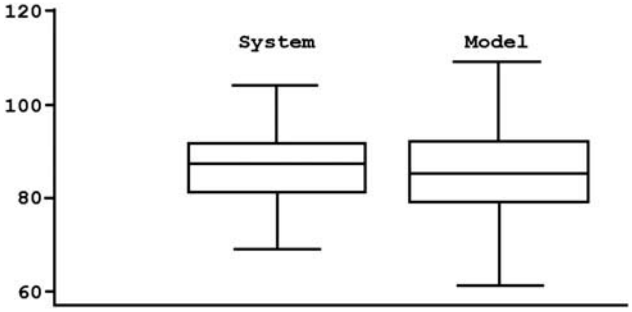


Figure 5 Box plot of hospital data.

measures such as the mean, variance, maximum, minimum, distribution, and times series of the variables, and (2) relationships between (a) two measures of a single variable (see, eg, Figure 6) and (b) measures of two variables. It is important that appropriate measures and relationships be selected with respect to the model's intended purpose to be used in developing graphs to validate a simulation model. (See Anderson and Sargent (1974) and Lowery (1996) for examples of sets of graphs used in the validation of two different simulation models.)

The simulation model data in these graphs are used as reference distributions (instead of theoretical distributions such a t or F distribution) to compare the system data against to determine if the model measure (eg the mean of an output) being evaluated has a satisfactory range of accuracy. We see in Figure 4 that the system data point lies within the model data, in Figure 5 that the model data has a slightly smaller median and more variability than the system data, and in Figure 6 we see that the relationship between the mean and standard deviation of reaction time for both the model and system is linear with the same slope. A subjective decision is made for each graph if the model accuracy is within the acceptable range of accuracy. We note that the same errors can be made when making these subjective decisions as in the hypothesis tests discussed in the subsection 'Hypothesis tests'.

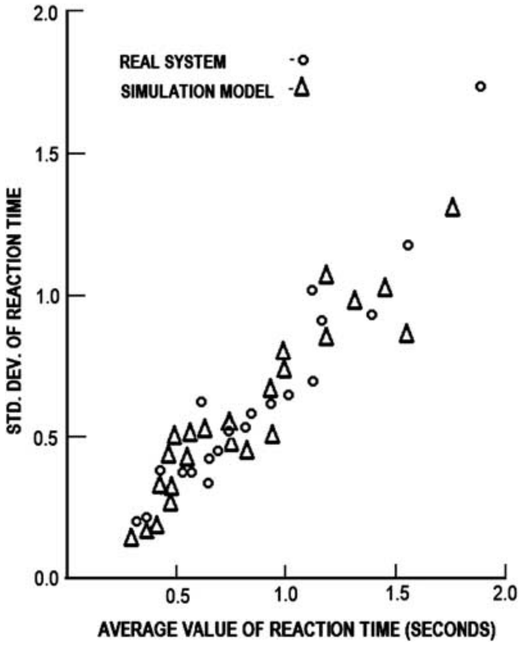


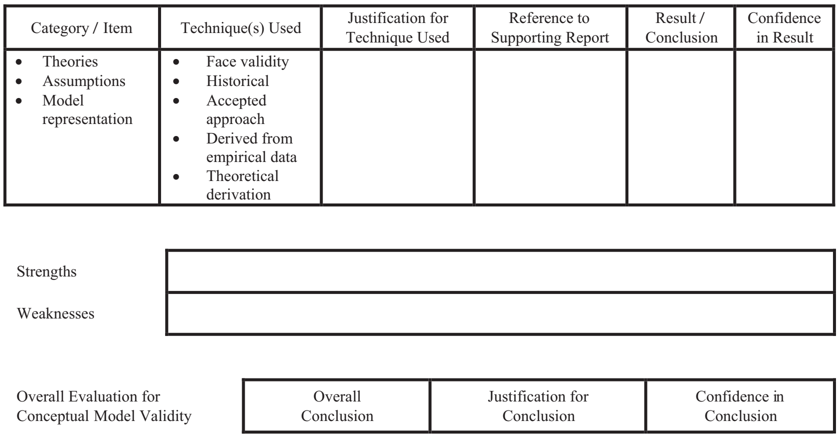
Figure 6 Behaviour graph of computer reaction time.

The only requirement on the data in these graphs is that they must be 'identically distributed'; the data can have any statistical distribution and be correlated. Sufficient simulation model data must be generated to be used as a reference distribution. The minimum number of data points needed depends on the amount of variability and correlation between data points—the larger each is the more model data that is required. (See Sargent (1996a, 2001a, b) for further discussion.)

These graphs can be used in model validation in different ways. First, the model development team (and also the users if they are deciding the validity of a simulation model) can use the graphs in the model development process to make a subjective judgement on whether a simulation model possesses sufficient accuracy for its intended purpose. Second, they can be used in the face validity technique where experts are asked to make subjective judgements on whether a simulation model possesses sufficient accuracy for its intended purpose. Third, the graphs can be used in Turing tests. Fourth, the graphs can be used in different ways in IV&V.

9. Documentation

Documentation on model verification and validation is usually critical in convincing users of the 'correctness' of a model and its results, and should be included in the simulation model documentation. (See Gass (1984) for a general discussion on documentation of computer-based models.) Both detailed and summary documentation are desired. The detailed documentation should include specifics on the tests, evaluations made, data, results, etc. The summary



Category / Item	Technique(s) Used	Justification for Technique Used	Reference to Supporting Report	Result / Conclusion	Confidence in Result
• Theories • Assumptions • Model representation	• Face validity • Historical • Accepted approach • Derived from empirical data • Theoretical derivation				

Strengths	
Weaknesses	

Overall Evaluation for Conceptual Model Validity	Overall Conclusion	Justification for Conclusion	Confidence in Conclusion

Figure 7 Evaluation table for conceptual model validity.

documentation should contain a separate evaluation table for data validity, conceptual model validity, computer model verification, operational validity, and an overall summary (see Figure 7 for an example of an evaluation table of conceptual model validity). (Examples of other evaluation tables are contained in Sargent, 1991, 1996b.) The entries of Figure 7 are self-explanatory except for the last column, which refers to the confidence the evaluators have in the results or conclusions. These are often expressed as low, medium, or high. (Data for this documentation are collected during the development of the simulation model.)

10. Recommended procedure

The author recommends that the following eight steps be performed in model verification and validation:

1. An agreement be made prior to developing the model between (a) the model development team and (b) the model sponsors and (if possible) the users that specifies the decision-making approach and a minimum set of specific validation techniques to be used in determining model validity.
2. Specify the acceptable range of accuracy required of the simulation model's output variables of interest for the model's intended application prior to starting the development of the model or very early in the model development process.
3. Test, wherever possible, the assumptions and theories underlying the simulation model.
4. In every model iteration, perform at least face validity on the conceptual model.
5. In every model iteration, at least explore the simulation model's behaviour using the computerized model.
6. In at least the last model iteration, make comparisons, if possible, between the simulation model and system behaviour (output) data for at least a few sets of experimental conditions, and preferably for several sets.
7. Prepare the verification and validation documentation for inclusion in the simulation model documentation.
8. If the simulation model is to be used over a period of time, develop a schedule for periodic review of the model's validity.

Some simulation models are developed for repeated use. A procedure for reviewing the validity of these models over their life cycles needs to be developed, as specified in Step 8. No general procedure can be given because each situation is different. For example, if no data were available on the system when a simulation model was initially developed and validated, then revalidation of the model should take place prior to each usage of the model if new data or system understanding has occurred since the last validation.

11. Summary

Model verification and validation are critical in the development of a simulation model. Unfortunately, there is no set of specific tests that can easily be applied to determine the 'correctness' of a model. Furthermore, no algorithm exists to determine what techniques or procedures to use. Every simulation project presents a new and unique challenge to the model development team. The verification and validation information presented in this paper should help with this challenge.

In this paper, we discussed 'practical approaches' to verification and validation of simulation models. For a discussion on the philosophy of model validation, see Kleindorfer and Ganeshan (1993).

There is considerable literature on model verification and validation (see, eg, Balci and Sargent, 1984a). Beyond the references already cited above, there are conference tutorials and papers (eg, Sargent, 1979, 1984b, 1990, 2000, 2005), journal articles (eg, Gass, 1983; Landry *et al*, 1983), discussions in textbooks (eg, Zeigler, 1976; Zeigler *et al*, 2000; Robinson, 2004; Law, 2007; Banks *et al*, 2010), USA Government Reports (eg, U.S. General Accounting Office, 1987; DoDI 5000.61, 2009), and a book by Knepell and Arangno (1993) that can be used to further your knowledge on model verification and validation.

Research continues on these topics. This includes such items as advisory systems (eg, Rao and Sargent, 1988; Balci, 2001), scoring models (eg, Gass and Joel, 1987; Balci, 1989; Gass, 1993), cost of validation (eg, Szabo and Teo, 2012) and new approaches, procedures, and techniques (eg, Balci *et al*, 2002; Ruess and de Moura, 2003; Sargent, 2010). We note that advisory systems have not come into use and that scoring models are rarely used as there are issues with them (Sargent, 2005). See Sargent *et al* (2000) for a discussion on research directions.

References

Anderson HA and Sargent RG (1974). An investigation into scheduling for an interactive computer system. *IBM Journal of Research and Development* **18**(2): 125–137.

Anshoff HI and Hayes RL (1973). Roles of models in corporate decision making. In: Ross M (ed). *Operations Research '72: Proceedings of the Sixth IFORS International Conference on Operational Research*. North Hollard: Amsterdam.

Balci O (1989). How to assess the acceptability and credibility of simulation results. In: MacNair EA, Musselman KJ and Heidelberger P (eds). *Proceedings of the 1989 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 62–71.

Balci O (2001). A methodology for certification of modeling and simulation applications. *ACM Transactions on Modeling and Computer Simulation* **11**(4): 352–377.

Balci O and Sargent RG (1981). A methodology for cost-risk analysis in the statistical validation of simulation models. *Communications of the ACM* **24**(6): 190–197.

Balci O and Sargent RG (1984a). A bibliography on the credibility assessment and validation of simulation and mathematical models. *Simuletter* **15**(3): 15–27.

Balci O and Sargent RG (1984b). Validation of simulation models via simultaneous confidence intervals. *American Journal of Mathematical and Management Science* **4**(3): 375–406.

Balci O, Nance RE, Arthur JD and Ormsby WF (2002). Expanding our horizons in verification, validation, and accreditation research and practice. In: Peters BA, Smith JS, Medeiros DJ and Rohrer MW (eds). *Proceedings of the 2002 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 653–663.

Banks J, Carson II JS, Nelson BL and Nicol D (2010). *Discrete-event System Simulation*, 5th edn. Prentice-Hall: Englewood Cliffs, NJ.

Banks J, Gerstein D and Searles SP (1988). Modeling processes, validation, and verification of complex simulations: A survey. In: *Methodology and Validation*, Simulation Series, Vol. 19, No. 1, The Society for Computer Simulation. Society for Modeling and Simulation International: San Diego, CA, pp 13–18.

DoDI 5000.61 (2009). DoD modeling and simulation verification, validation, and accreditation. 9 December.

Fairley RE (1976). Dynamic testing of simulation software. In: *Proceedings of the 1976 Summer Computer Simulation Conference*. Society for Modeling and Simulation International: San Diego, CA, pp 40–46.

Gass SI (1983). Decision-aiding models: Validation, assessment, and related issues for policy analysis. *Operations Research* **31**(4): 601–663.

Gass SI (1984). Documenting a computer-based model. *Interfaces* **14**(3): 84–93.

Gass SI (1993). Model accreditation: A rationale and process for determining a numerical rating. *European Journal of Operational Research* **66**(2): 250–258.

Gass SI and Joel L (1987). Concepts of model confidence. *Computers and Operations Research* **8**(4): 341–346.

Hines WW, Montgomery DC, Goldsman DM and Borror CM (2003). *Probability and Statistics for Engineers*, 4th edn. John Wiley: New York.

Johnson RA, Miller I and Freund J (2010). *Miller and Freund's Probability and Statistics for Engineers*, 8th edn. Prentice-Hall: Englewood Cliffs, NJ.

Kleijnen JPC (1999). Validation of models: Statistical techniques and data availability. In: Farrington PA, Black Nembhard H, Sturrock DT and Evans GW (eds). *Proceedings of the 1999 Simulation Conference*. IEEE: Piscataway, NJ, pp 647–654.

Kleijnen JPC (2008). *Design and Analysis of Simulation Experiments*. Springer-Verlag: Heidelberg.

Kleijnen JPC, Cheng RCH and Bettonvil BWM (2001). Validation of trace-driven simulation models: Bootstrapped tests. *Management Science* **47**(11): 1533–1538.

Kleindorfer GB and Ganeshan R (1993). The philosophy of science and validation in simulation. In: Evans GW, Mollaghasemi M, Russell EC and Biles WE (eds). *Proceedings of the 1993 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 50–57.

Knepell PL and Arangno DC (1993). *Simulation Validation: A Confidence Assessment Methodology*. IEEE Computer Society Press: Los Alamitos, CA.

Landry M, Malouin JL and Oral M (1983). Model validation in operations research. *European Journal of Operational Research* **14**(3): 207–220.

Law AM (2007). *Simulation Modeling and Analysis*, 4th edn. McGraw-Hill: New York.

Lowery J (1996). Design of hospital admissions scheduling system using simulation. In: Charnes JM, Morrice DJ, Brunner DT and Swain JJ (eds). *Proceedings of the 1996 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 1199–1204.

Nance RE and Arthur JD (2006). Software requirements engineering: Exploring the role in simulation model development. In: Roberson S, Taylor S, Brailsford S and Garnett J (eds). *Proceedings of the 2006 Operational Research Society Simulation Workshop*. Coventry: England, pp 117–127.

Naylor TH and Finger JM (1967). Verification of computer simulation models. *Management Science* **14**(2): B92–B101.

Rao MJ and Sargent RG (1988). An advisory system for operational validity. In: Hensen T (ed). *Artificial Intelligence and Simulation: The Diversity of Applications*. Society for Computer Simulation: San Diego, CA, pp 245–250.

Robinson S (2004). *Simulation: The Practice of Model Development and Use*. John Wiley: Chichester, West Sussex, UK.

Ruess H and de Moura L (2003). From simulation to verification (and back). In: Chick S, Sanchez PJ, Ferrin E and Morrice DJ (eds). *Proceedings of the 2003 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 888–896.

Sargent RG (1979). Validation of simulation models. In: Highland HJ, Spiegel MF and Shannon RE (eds). *Proceedings of the 1979 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 497–503.

Sargent RG (1981). *An assessment procedure and a set of criteria for use in the evaluation of computerized models and computer-based modeling tools*. Final Technical Report RADC-TR-80-409, U.S. Air Force.

Sargent RG (1982). Verification and validation of simulation models. Chapter IX. In: Cellier FE (ed). *Progress in Modelling and Simulation*. Academic Press: London, pp 159–169.

Sargent RG (1984a). Simulation model validation, Chapter 19. In: Oren TI, Zeigler BP and Elzas MS (eds). *Simulation and Model-based Methodologies: An Integrative View*. Springer-Verlag: Heidelberg, Germany, pp 537–555.

Sargent RG (1984b). A tutorial on verification and validation of simulation models. In: Sheppard S, Pooch UW and Pegden CD (eds). *Proceedings of the 1984 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 114–121.

Sargent RG (1986). The use of graphical models in model validation. In: Wilson JR, Henriksen JO and Roberts SD (eds). *Proceedings of the 1986 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 237–241.

Sargent RG (1990). Validation of mathematical models. In: *Proceedings of Geoval-90: Symposium on Validation of Geosphere Flow and Transport Models*. Stockholm, Sweden, pp 571–579.

Sargent RG (1991). Simulation model verification and validation. In: Nelson BL, Kelton WD and Clark GM (eds). *Proceedings of the 1991 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 37–47.

Sargent RG (1996a). Some subjective validation methods using graphical displays of data. In: Charnes JM, Morrice DJ, Brunner DT and Swain JJ (eds). *Proceedings of the 1996 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 345–351.

Sargent RG (1996b). Verifying and validating simulation models. In: Charnes JM, Morrice DJ, Brunner DT and Swain JJ (eds). *Proceedings of the 1996 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 55–64.

Sargent RG (2000). Verification, validation, and accreditation of simulation models. In: Joines JA, Barton RR, Kang K and Fishwick PA (eds). *Proceedings of the 2000 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 50–59.

Sargent RG (2001a). Graphical displays of simulation model data as statistical references. In: Ermakor SM, Kashtanov YN and Melas VB (eds). *Simulation 2001 (Proceedings of the 4th St. Petersburg Workshop on Simulation)*. Chemistry Research Institute of St. Petersburg University, St. Petersburg, Russia, pp 109–118.

Sargent RG (2001b). Some approaches and paradigms for verifying and validating simulation models. In: Peters BA, Smith JS, Medeiros DJ and Rohrer MW (eds). *Proceedings of the 2001 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 106–114.

Sargent RG (2005). Verification and validation of simulation models. In: Kuhl ME, Steiger NM, Armstrong FB and Joines JA (eds). *Proceedings of the 2005 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 130–143.

Sargent RG (2010). *A new statistical procedure for validation of simulation and stochastic models*. Technical Report SYR-EECS-2010-06, Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York.

Sargent RG, Glasow PA, Kleijnen JPC, Law AM, McGregor I and Youngblood S (2000). Strategic directions in verification, validation, and accreditation research. In: Joines JA, Barton RR, Kang K and Fishwick PA (eds). *Proceedings of the 2000 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 909–916.

Schruben LW (1980). Establishing the credibility of simulation models. *Simulation* **34**(3): 101–105.

Szabo C and Teo YM (2012). An analysis of the cost of validating semantic composability. *Journal of Simulation* **11**(3): 152–163.

U.S. General Accounting Office PEMD-88-3 (1987). DOD simulations: Improved assessment procedures would increase the credibility of results. U.S. General Accounting Office: Washington DC.

Whitner RG and Balci O (1989). Guideline for selecting and using simulation model verification techniques. In: MacNair EA, Musselman KJ and Heidelberger P (eds). *Proceedings of the 1989 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 559–568.

Wood DO (1986). MIT model analysis program: What we have learned about policy model review. In: Wilson JR, Henriksen JO and Roberts SD (eds). *Proceedings of the 1986 Winter Simulation Conference*. IEEE: Piscataway, NJ, pp 248–252.

Zeigler BP (1976). *Theory of Modelling and Simulation*. John Wiley and Sons, Inc: New York.

Zeigler BP, Praehofer H and Kim TG (2000). *Theory of Modelling and Simulation*, 2nd edn. Academic Press: London.

Received 6 October 2011;
accepted 16 October 2012 after two revisions