Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

VERIFICATION AND VALIDATION OF SIMULATION MODELS: AN ADVANCED TUTORIAL

Robert G. Sargent

Department of Electrical Engineering and Computer Science
College of Engineering and Computer Science
Syracuse University
Syracuse, NY 13244, USA

ABSTRACT

Verification and validation (V&V) of simulation models are discussed in this paper. Different approaches to deciding model validity are described and a graphical paradigm that relates V&V to the model development process is presented and explained. Conceptual model validity, model verification, operational validity, and data validity are discussed, documentation is briefly covered, and a recommended procedure for model validation is presented. References for further information are provided when the various aspects of conducting V&V of simulation models are discussed.

1 INTRODUCTION

This paper discusses verification and validation (V&V) of simulation models. V&V are concerned with determining whether a model and its results are "correct" for a specific use or purpose. Model verification is formally defined as "ensuring that the computer program of the computerized model and its implementation are correct" and model validation is defined as the "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model." Our discussion of V&V will focus primarily on simulation models that are used to predict system behaviors such as systems outputs. Two related topics are model credibility and model usability. Model credibility is concerned with developing in (potential) users the confidence they require in order to use a model and the information derived from that model. Model usability is determining if the model and its user instructions are easy to use. (See Sargent and Balci (2017) for a history on the development of V&V of simulation models.)

A model should be developed for a specific purpose and its validity determined with respect to that purpose. A developed model should usually be a parsimonious model, meaning the model is as simple as possible yet meets its purpose. Furthermore, the accuracy of a model (sometimes referred to as model fidelity) should usually be only what is needed to satisfy the model's use or 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. The developers and users of models, the decision makers using information obtained from the results of models, and the individuals affected by decisions based on models are all rightly concerned with whether a model and the model's results are "correct" for each question being addressed.

Numerous sets of experimental conditions are usually required to define the domain of a model's intended applicability. A set of experimental conditions contains a set of values for the set of variables that define the domain of 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 (i.e., 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, i.e., performing model V&V, 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; 1984). 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. The cost of model validation is usually quite significant, especially when extremely high model confidence is required. (See Sargent (2013, 2015a) for further discussion on the relationship of the cost of model validation and model confidence.)

The remainder of the paper is organized as follows: Section 2 presents the basic approaches used in deciding model validity; Section 3 describes a graphical paradigm used in verification and validation, and the model development process that emphasizes verification and validation; Sections 4, 5, 6, and 7 discuss data validity, conceptual model validity, computerized model verification, and operational validity, respectively; Section 8 covers documentation; Section 9 contains a recommended validation procedure consisting of eight steps; and Section 10 has the summary.

2 DECISION-MAKING APPROACHES

There are three basic decision-making approaches for deciding whether a simulation model is valid. Each of these three approaches uses a different decision-maker. All of the approaches require the model development team to conduct V&V as part of the model development process, which is discussed in Section 3. One decision-making approach, often used, 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 V&V of the model and the users determine if the model is satisfactory in each phase of V&V. This approach is generally used with a model development team whose size is not large. Also, 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 that 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 V&V 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 regarding whether the model V&V is satisfying the model requirements and when not, what the difficulties are. When conducting IV&V this way, the development of a simulation model should not progress to the next stage of development until the model has satisfied the V&V requirements in its current stage. It is this author's opinion the concurrent way 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 V&V conducted by the model development team to performing a separate thorough V&V effort themselves. Performing a complete IV&V effort after the model has been completely developed is usually both *extremely* costly and time consuming. 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 V&V that has already been performed.

When an IV&V team concludes that a 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 MODEL DEVELOPMENT PROCESS WITH VERIFICATION AND VALIDATION

In this section a graphical paradigm is presented in subsection 3.1 that relates model V&V to the model development process. Then in subsection 3.2 the model development process is described that includes V&V.

3.1 A Graphical Paradigm

There are two common ways to view how V&V relate to the model development process. One way uses a simple view and the other uses a complex view. A simple graphical paradigm is presented in Figure 1 that was developed by this author called the Simplified View of the Model Development Process (Sargent 1981; 1982; 1983; 2001b; 2013). A more-complex paradigm developed by this author that includes both the "Simulation World" and the "Real World" is contained in Sargent (2001b, 2013).

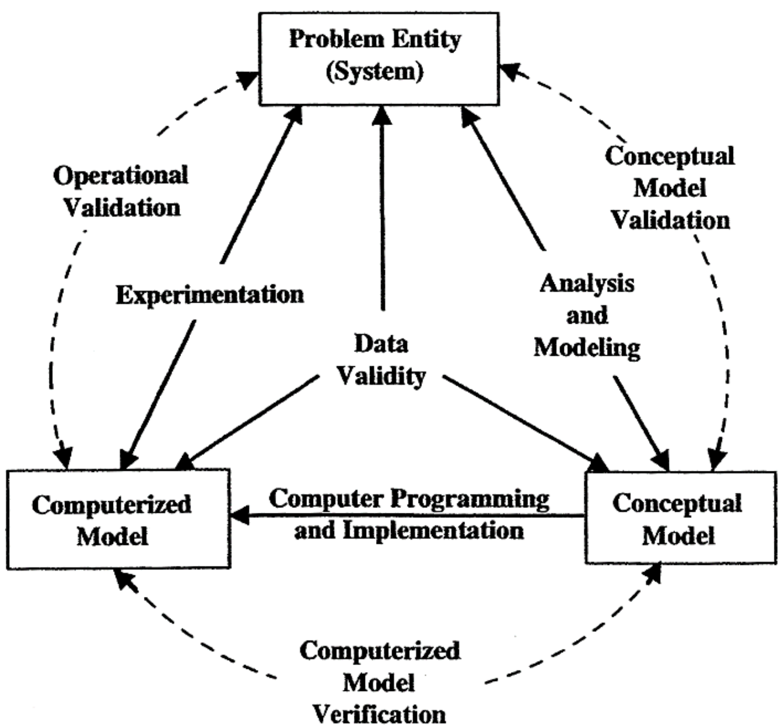

Figure 1: Simplified version of the model development process.

Consider the simplified version of the model development process in Figure 1. The *problem entity* is the system (real or proposed), idea, situation, policy, or phenomena to be modeled; 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 modeling 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 V&V to this simplified version of the model development process (Figure 1). *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 behavior 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. These items are discussed below.

3.2 Model Development Process

A model should be developed for a specific purpose or use. Additionally, a model should be a parsimonious model, which means that the model is as simple as possible yet meets its purpose. Also, the accuracy of a model need not be more than what is required for its purpose. A simulation model is a structural model meaning that the model contains logical and causal relationships that occur in the systems. Developing a valid simulation model is an iterative process where several versions of a model are developed prior to obtaining a valid model.

The model development process should include model V&V. Following the paradigm given in Figure 1, the iterative process shown in Figure 2 can be used to develop a valid simulation model (Sargent 1984). We first develop a conceptual model through analyzing the problem entity and then developing a model of the problem entity, remembering that a parsimonious model is desired. (See, e.g., Robinson (2017) for further information on conceptual modeling.) Then conceptual model validation is performed. This process is repeated until the conceptual model is satisfactory. Next a computerized model is developed of the (validated) conceptual model by developing a simulation model of the conceptual model and implementing it on a computer. Then computerized model verification is performed. This process is repeated until the computerized model is satisfactory. Lastly, operational validation is performed on the computerized model. Model changes required by conducting operational validity can be in either the conceptual model or in the computerized model. V&V must be performed again when any model change is made. This process is repeated until a valid simulation model is obtained. As stated above, several versions of a model are usually developed prior to obtaining a valid simulation model. There are numerous validation techniques that are used in conducting verification and validation. See, e.g., Balci (1994, 1998); Sargent (2013); Sargent and Balci (2017); Sargent et al. (2016); Whitner and Balci (1989); and references therein, for various techniques used in verifying and validating a simulation model.

4 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 needed 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.

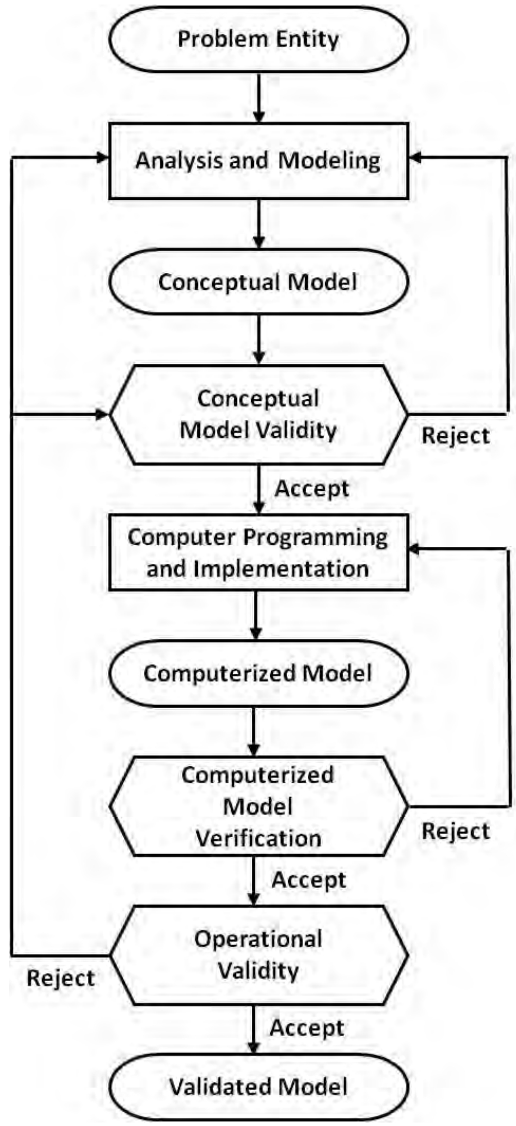

Figure 2: The model development iterative process.

To build a conceptual model we must have sufficient data on the problem entity to develop theories that can be used to build the model, to develop mathematical and logical relationships for use in the model that will allow the model to adequately represent the problem entity for its intended purpose, and to test the model's underlying assumptions. In addition, behavioral data are needed on the problem entity to be used in the operational validity step of comparing the problem entity's behavior with the model's behavior. (Usually, this data are system input/output data.) If behavior 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 ensure that the data are correct. One should develop good procedures for (1) collecting and maintaining data, (2) testing the collected data using techniques such as data relationship correctness (Sargent 2013), and (3) screening the data for outliers *and* determining if the outliers are correct. If the amount of data is large, a database of the data should be developed and maintained.

5 CONCEPTUAL MODEL VALIDATION

Conceptual model validity is determining that (1) the theories and assumptions underlying the conceptual model have *all* been identified, clearly stated, and determined to be 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 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 if 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?

Each submodel and the overall model must be evaluated to determine if they are reasonable and correct for the intended purpose of the model. This should include determining if the appropriate detail and aggregate relationships have been used for the model's intended purpose, and also if appropriate structure, logic, and mathematical and causal relationships have been used. The primary validation techniques used for these evaluations are face validation and traces. Face validation has experts on the problem entity evaluate the conceptual model to determine if it is correct and reasonable for its purpose. This usually requires either (1) the experts examining the flowchart or graphical model (Sargent 1986) and the set of model equations or (2) the experts receiving from the conceptual model developer a structured walkthrough, which is a formal detail explanation of the conceptual model. The use of traces is the tracking of entities through each submodel and the overall model to determine if the logic is correct and if the necessary accuracy is maintained. If errors are found in the conceptual model, it must be revised and conceptual model validation performed again. Furthermore, the theories and assumptions underlying each submodel and the overall model also need to be checked.

6 COMPUTERIZED MODEL VERIFICATION

Computerized model verification ensures that the computer programming and implementation of the conceptual model are correct and the model executes properly. The major factor affecting verification is what type of software is used for the simulation. The types of software used are (1) higher-level computer programming languages such as Fortran, C, C++, Java, R, or Python, which provide no simulation support, (2) "simple" simulation languages or packages that use a specific higher-level programing language and contains some or all of the simulation model execution functions such as the time-flow mechanism, event list management, and random variate generators, (3) "advance" simulation languages that use a specific computer language and contains both simulation model execution functions and modeling functions, which may include graphical capabilities, and (4) simulation software systems that provide complete support for model building, model execution, model analysis, etc. using some specific computer language. The advance simulation languages and simulation software systems are either general purpose or special purpose. Special purpose means they are designed for modeling and simulating specific types of problems such as risk problems or systems such as manufacturing systems or hospital systems. Using software that contain more simulation capabilities (increasing from 1 to 4 above and going to special purpose from general purpose) usually reduces the number of (initial) errors in developing a simulation, reduces the amount of time required to develop a valid model, reduces flexibility, and increases simulation execution times.

In performing computerized model verification, one must ensure that the conceptual model has been programmed and implemented correctly and executes properly in the computer language being used. This implies if an advance simulation language or a simulation software system is used that verification is primarily concerned with ensuring that the conceptual model has been properly developed and implemented and then executes correctly in the simulation language being used; and also that the software is error free, that the software has been properly implemented on the computer, and that a tested (for correctness) pseudo random number generator has been properly implemented. If a higher-level computer programming language or a simple simulation language is used, the computer program should be designed, developed, and implemented using techniques found in software engineering. (These include such techniques as object-oriented design, structured programming, and program modularity.) Verification is concerned in these two cases with ensuring that (1) the simulation model execution functions execute correctly *and* (2) the conceptual model has been properly programmed and implemented and executes correctly in the higher-level computer programming language being used.

There are two basic approaches for verifying (testing) simulation software and simulation models: static testing and dynamic testing (Fairley 1976). In static testing the computer program is analyzed to determine if it is correct by using such techniques as structured walkthroughs, correctness proofs, and examining the structure properties of the program (Law 2015; Sargent 2013). 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 if the simulation software and computerized models and their implementations are correct and models execute properly. The techniques commonly used in dynamic testing are traces, investigations of input-output relations using different validation techniques, data relationship correctness, reprogramming critical components to determine if the same results are obtained, and animation if it is being used. (Sargent 2013). If there are many variables, one might aggregate the numerical values of some of the variables to reduce the number of tests needed or use specific types of design of experiments (Kleijnen 1999; 2015).

It is necessary to be aware while conducting computerized model verification 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 model verification.)

7 OPERATIONAL VALIDITY

Operational validation is determining whether the simulation model's output behavior 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 steps that are involved in developing the simulation model including developing the system's theories or having invalid data. (See Sargent et al. (2016) for an in-depth discussion of operational validity of simulation with examples.)

Numerous validation techniques 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 behavior of the problem entity. Table 1 gives a classification of the validation techniques used in operational validity based on the decision approach and system observability. "Comparison" means comparing the simulation model output behavior to either the system output behavior or another model output behavior using graphical displays and/or statistical tests and procedures. "Explore model behavior" means to examine the output behavior of the simulation model using appropriate validation techniques. For both comparison and exploring model behavior (1) various sets of experimental conditions from the domain of the model's intended applicability should always be used and (2) parameter variability-sensitivity analysis (Sargent 2013) conducted. Furthermore, it is often desirable to use design of experiments (Kleijnen 2015) and also metamodels (Kleijnen and Sargent 2000) to aid in operational validity.

To obtain a *high* degree of confidence in a simulation model and its results, comparisons of the model's and system's output behaviors 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 a high degree of confidence in the model. In this situation the model output behavior(s) should be explored as thoroughly as possible and comparisons made to other valid models whenever possible.

Table 1: Operational validity classification.

Decision Approach	Observable System	Non-observable System
Subjective Approach	• Comparison Using Graphical Displays • Explore Model Behavior	• Explore Model Behavior • Comparison to Other Models
Objective Approach	• Comparison Using Statistical Tests and Procedures	• Comparison to Other Models Using Statistical Tests

7.1 Explore Model Behavior

The simulation model output behavior can be explored either qualitatively or quantitatively. In qualitative analysis the directions of the output behaviors are examined and also possibly whether the magnitudes are "reasonable." In quantitative analysis both the directions and the precise magnitudes of the output behaviors are examined. Experts on the system often know the directions and frequently know the "general values" of the magnitudes of the output behaviors. Many of the validation techniques can be used for model exploration. Parameter variability-sensitivity analysis should usually be used. Graphs of the output data discussed in Subsection 7.2.1 below can be used to display the simulation model output behavior. A variety of statistical approaches can be used in performing model exploration including metamodeling and design of experiments. (See Kleijnen (1999, 2015) for further discussion on the use of statistical approaches.) Numerous sets of experimental frames should be used in performing model exploration. Furthermore, simulation animation is useful to observe the behavior operation of a simulation model under various experimental frames to possibly see if any errors are occurring in the simulation model.

7.2 Comparisons of Output Behaviors

There are three basic approaches used in comparing the simulation model output behavior to either the system output behavior or another model output behavior: (1) the use of graphs to make a subjective decision, (2) the use of confidence intervals to make an objective decision, and (3) the use of hypothesis tests to make an objective decision. It is preferable to use confidence intervals or hypothesis tests for the comparisons because these allow for objective decisions. However, 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" (e.g., 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. Extreme care must be used in using this approach. Each of these three approaches is discussed below using system output data (Note: these same approaches can also be used with output data from a validated model instead of system output data when appropriate).

7.2.1 Graphical Comparisons of Data

The behavior data of the simulation model and the system are graphed for various sets of experimental conditions to determine if the model's output behavior has sufficient accuracy for the model's intended purpose. Three types of graphs are used: histograms, box (and whisker) plots, and behavior graphs using scatter plots. (See Sargent (1996a, 2001b) and Sargent et al. (2016) for a thorough discussion on the use of these for model validation.) Examples of a histogram and a box plot are given in Figures 3 and 4, respectively; both taken from Lowery (1996). Examples of behavior graphs, taken from Anderson and Sargent (1974), are given in Figures 5 and 6. A variety of graphs are required that use different types of (1) measures such as the mean, variance, maximum, distribution, and times series of the variables, and (2) relationships between (a) two measures of a single variable (see Figure 5) and (b) measures of two variables (see Figure 6). It is important that appropriate measures and relationships be used in validating a simulation model and that they be determined with respect to the model's intended purpose. See Anderson and Sargent (1974); Lowery (1996); Sargent et al. (2016) for other examples of sets of graphs used in the validation of simulation models.

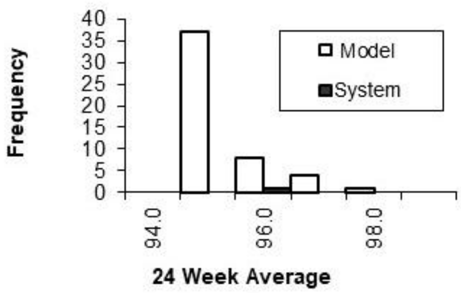

Figure 3: Histogram of hospital data.

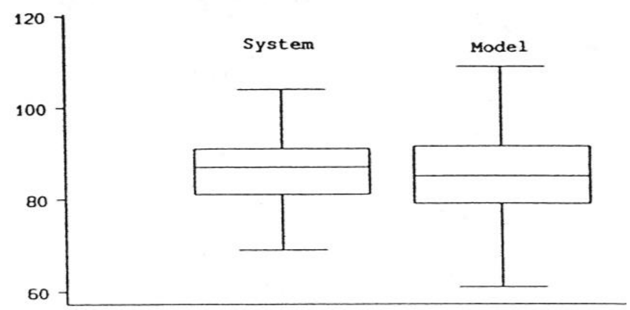

Figure 4: Box plot of hospital data.

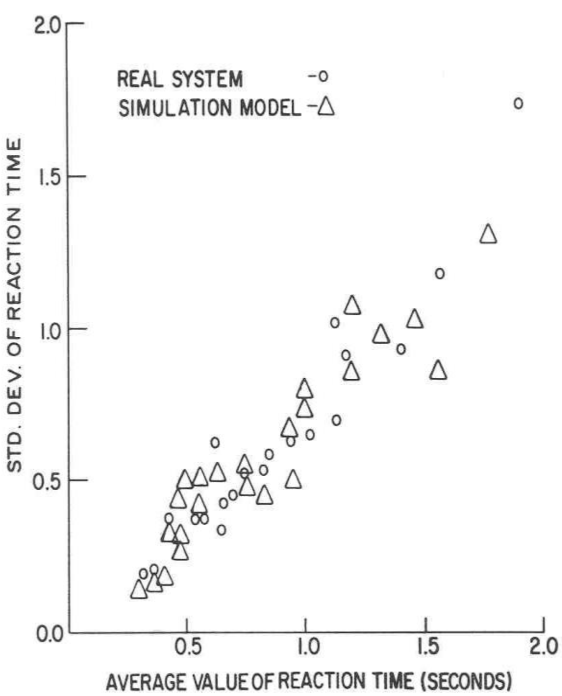

Figure 5: Computer reaction time.

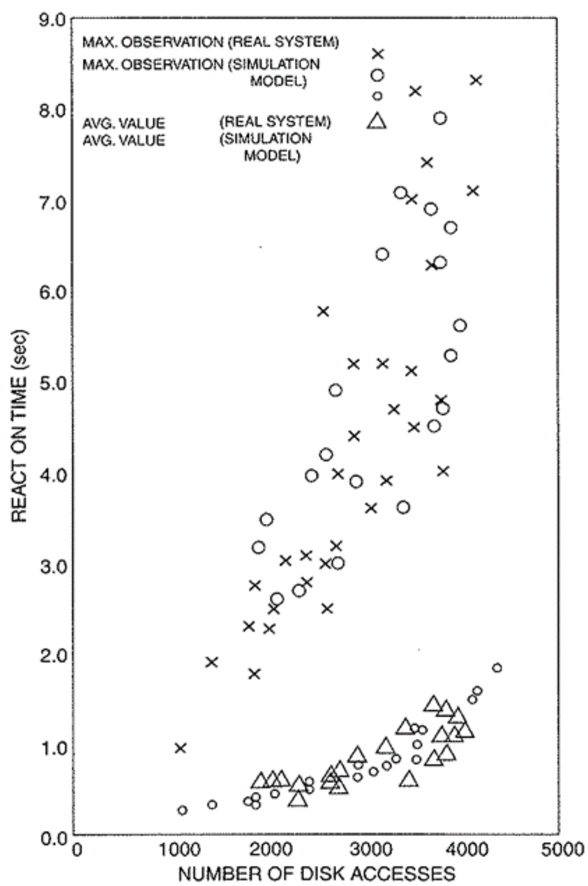

Figure 6: Computer disk behavior.

These graphs can be used in model validation in different ways. First, the model development team can use the graphs in the model development process to make a subjective judgment 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 judgments on whether a simulation model possesses sufficient accuracy for its intended purpose. Third, the graphs can be used in Turing tests (Schruben 1980). Fourth, the graphs can be used in different ways in IV&V. We note that the data in these graphs do not need to be independent nor satisfy any statistical distribution requirement such as normality of the data (Sargent 1996a; 2001a; 2001b).

7.2.2 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 if the simulation model's output behavior has an acceptable range of accuracy. An acceptable range of accuracy is the amount of accuracy that is required of a model to be valid for its intended purpose and is usually specified for each model variable of interest as a range for the difference between that model variable and the corresponding system variable.

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

- H_0 Model is valid for the acceptable range of accuracy under the set of experimental conditions.
- 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 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.

Classical 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, this author developed a new statistical procedure for comparisons of model and system outputs using hypothesis tests when the amount of model accuracy is specified as a *range* (Sargent 2015b); it is an Interval Hypothesis Test. This interval hypothesis test is applied at each experimental condition to determine if the model is valid for that experimental condition. Both type I and II errors are consider through the use of the operating characteristic curve (Hines et al. 2003; Johnson et al. 2018). Furthermore the model builder's and the model user's risk curves can be developed using a procedure associated with this interval hypothesis test. This procedure allows a trade-off to be made between the two risks for fixed sample sizes and for trade-offs among the two risks and variable sample sizes. See Sargent et al. (2015, 2016) for details and examples of the use of this new method for hypothesis testing the validity of models. Sargent et al. (2016) contains procedures for both the one-sample case when comparing a simulation model and an analytical model, and also the two-sample case when comparing the simulation model being validated against either another simulation model or the actual system.

7.2.3 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 (e.g., the mean) of the model and system output variable being evaluated. (Balci and Sargent (1984) contain details on the use of c.i. and s.c.i. for operational validity, including a general methodology.)

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 2015) s.c.i. The multivariate techniques can be used to develop an s.c.i. Both parametric and nonparametric statistical techniques can be used. The method of data collection used must satisfy the underlying assumptions of the statistical technique that is being used. One approach to developing a model range of accuracy is to use the standard statistical techniques and data collection methods used in simulation output analysis (Banks et al. 2010; Law 2015), e.g., using the methods of replication or (nonoverlapping) batch means.

8 DOCUMENTATION

Documentation on model V&V is usually critical in convincing users of the ‘correctness’ of a model and its results. This documentation should be included in the simulation model documentation and include both detailed and summary documentation. (Note: The simulation model documentation should include the model’s underlying assumptions and theories, along with other information.) The detailed documentation should include specifics on tests used, evaluations made, data, results, etc. The summary documentation should include a separate evaluation table for data validity, conceptual model validity, computer model verification, operational validity, and an overall summary table. The summary results should contain the confidence the evaluators have in the results and conclusions which are often expressed as low, medium, and high. Examples of different tables are contained in Sargent (1991, 1996b, 2013). (Also, see Gass (1984).)

9 RECOMMENDED PROCEDURE

This author recommends that the following eight steps be performed in model V&V:

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 each model iteration, perform at least face validity on the conceptual model.
5. In each model iteration, at least explore the simulation model’s behavior using the computerized model.
6. In at least the last model iteration, make comparisons, if possible, between the simulation model and system behavior (output) data for at least a few sets of experimental conditions, and preferably for several sets.
7. Prepare the V&V 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.

10 SUMMARY

Model V&V are critical in the development of a simulation model. Unfortunately, there is no set of specific tests that can be easily 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. Thus, sufficient knowledge on performing V&V of simulation models is required of every simulation team doing simulation studies.

This paper concentrated on the "fundamentals" and the "practical how to do" aspects of V&V of simulation models. There are other aspects of V&V. Perhaps the best place to see the various aspects with their references are in Sargent and Balci (2017)'s history paper on the V&V of simulation models. This includes, for example, the philosophy of model validation (see, e.g, Kleindorfer and Ganeshan (1993)), theories of simulation validation (see, e.g., Zeigler (1976), Zeigler et al. (2018)), and model accreditation (see, e.g., DoDI 5000.61 (2009), Sargent (2000)). Also, discussed in Sargent and Balci (2017) is the literature on model V&V with references that one can use to further their knowledge on simulation model V&V. This literature includes conference tutorials, journal articles, government reports, simulation textbooks, and books on model V&V (Knepell and Arangno (1993); Oberkampf and Roy (2010)).

Furthermore, the Sargent and Balci (2017) paper discussed the need to increase the awareness that "verification and validation should be adequately performed in all simulation studies" because currently not all simulation studies are doing this. This requires education that V&V should be adequately performed on *all* simulation models; and additionally, organizations requiring V&V to be adequately addressed in all simulation studies. (For example, the Department of Defense requires Verification, Validation, and Accreditation (VV&A) be performed and documented for all Models and Simulations per DoDI 5000.61 (2009).)

Also, there is the need to continue research in model V&V. An increased understanding is needed on how to verify and validate simulation models that are continuously increasing in complexity and for new uses. This increasing complexity is occurring because systems that are being modeled are becoming more complex such as "systems of systems" and systems that contain artificial intelligence and machine learning necessitating more complex models and simulations. These models include hierarchical models (Sargent 1993), hybrid simulation/analytic models (Shanthikumar and Sargent 1983), combined discrete-event/continuous simulation (Fahrland 1970) that uses mixed simulation models (originally called combined simulation but more recently being called hybrid simulation), and new modeling world views such as agent-based modeling for discrete-event simulation (see, e.g., Macal and North (2010)). Furthermore, simulation models are increasing being used in the operation of real time systems such as in the controlling of manufacturing systems and also for use in what is being called "digital twins" (Kulkarni et al. 2019).

REFERENCES

Anderson, H. A. and R. G. Sargent. 1974. "Investigation into Scheduling for an Interactive Computer System". *IBM Journal of Research and Development* 18(2):125-137.

Balci, O. 1994. "Validation, Verification, and Testing Techniques throughout the Life Cycle of a Simulation Study". *Annals of Operations Research* 53:121–174.

Balci, O. 1998. "Verification, Validation, and Testing". Chapter 10 in *The Handbook of Simulation*, edited by J. Banks, 335–393. New York, New York: John Wiley & Sons.

Balci, O. and R. G. Sargent. 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 R. G. Sargent. 1984. "Validation of Simulation Models via Simultaneous Confidence Intervals". *American Journal of Mathematical and Management Science* 4(3):375-406.

Banks, J., J. S. Carson II, B. L. Nelson, and D. Nicol. 2010. *Discrete-event System Simulation.* 5th ed. Englewood Cliffs, New Jersey: Prentice-Hall.

DoDI 5000.61. 2009. "DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)". December 9, 2009. U.S. Department of Defense, Washington. D.C.
Fahrland, D. A. 1970. "Combined Discrete Event Continuous Systems Simulation". *Simulation* February 1970, 61-72.
Fairley, R. E. 1976. "Dynamic Testing of Simulation Software". In *Proceedings of the 1976 Summer Computer Simulation Conf*erence, 40-46. San Diego, California: Society for Modeling and Simulation International.
Gass, S. I. 1984. "Documenting a Computer-based Model". *Interfaces* 14 (3): 84-93.
Hines, W. W., D. C. Montgomery, D. M. Goldsman, and C. M. Borror. 2003. *Probability and Statistics for Engineers*. 4th ed. New York: John Wiley.
Johnson, R. A., I. Miller, and J. E. Freund 2018. *Miller and Freund's Probability and Statistics for Engineers*. 9th ed. Englewood Cliffs, New Jeresy: Prentice-Hall.
Kleijnen, J. P. C. 1999. "Validation of Models: Statistical Techniques and Data Availability". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. Black Nembhard, D. T. Sturrock, and G. W. Evans, 647-654. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Kleijnen, J. P. C. 2015. *Design and Analysis of Simulation Experiments*. 2nd. ed. Heidelberg: Springer.
Kleijnen, J. P. C. and R. G. Sargent. 2000. "A Methodology for Fitting and Validating Metamodels in Simulation". *European Journal of Operational Research* 120:14–29.
Kleindorfer, G. B. and R. Ganeshan. 1993. "The Philosophy of Science and Validation in Simulation". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, 50-57. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Knepell, P. L. and D. C. Arangno. 1993. *Simulation Validation: A Confidence Assessment Methodology*. Los Alamitos, California: IEEE Computer Society Press.
Kulkarni, V., S. Barat, and T. Clark. 2019. "Towards Adaptive Enterprises Using Digital Twins". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 60-74. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Law, A. M. 2015. *Simulation Modeling and Analysis*. 5th ed. New York: McGraw-Hill.
Lowery, J. 1996. "Design of Hospital Admissions Scheduling System Using Simulation". In *Proceedings of the 1996 Winter Simulation Conference*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 1199-1204. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Macal, C. M. and M. J. North. 2010. "Tutorial on Agent-Based Modeling and Simulation". *Journal of Simulation* 4(3):151–162.
Oberkampf, W. L. and C. J. Roy. 2010. *Verification and Validation in Scientific Computing.* Cambridge: University Press.
Robinson, S. 2017. "A Tutorial on Simulation Conceptual Modeling". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zachareqicz, N. Mustafee, G. Wainer, and E. Page, 565-579. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Sargent, R. G. 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, R. G. 1982. "Verification and Validation of Simulation Models". Chapter IX in *Progress in Modelling and Simulation*, edited by F. E. Cellier, 159-169. London: Academic Press.
Sargent, R. G. 1983. "Validating Simulation Models". In *Proceedings of the 1983 Winter Simulation Conference*, edited by S. Roberts, J. Banks, and B. Schmeiser, 333-337. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Sargent, R. G. 1984. "Simulation Model Validation". Chapter 19 in *Simulation and Model-Based Methodologies: An Integrative View*, edited by T. I. Oren, B. P. Zeigler, and M. S. Elzas, 537-555. Heidelberg, Germany: Springer-Verlag.
Sargent, R. G. 1986. "The Use of Graphical Models in Model Validation". In *Proceedings of the 1986 Winter Simulation Conference*, edited by J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 237-241. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Sargent, R. G. 1991. "Simulation Model Verification and Validation". In *Proceedings of the 1991 Winter Simulation Conference*, edited by B. L. Nelson, W. D. Kelton, and G.M. Clark, 37-47. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Sargent, R. G. 1993. "Hierarchical Modeling for Discrete Event Simulation". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, 569-572. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
Sargent, R. G. 1996a. "Some Subjective Validation Methods Using Graphical Displays of Data". In *Proceedings of the 1996 Winter Simulation Conference*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 345-351. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Sargent, R. G. 1996b. "Verifying and Validating Simulation Models." In *Proceedings of the 1996 Winter Simulation Conference*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 55-64. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.
Sargent, R. G. 2000. "Verification, Validation, and Accreditation of Simulation Models". In *Proceedings of the 2000 Winter Simulation Conference*, edited by J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, 50-59. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2001a. "Graphical Displays of Simulation Model Data as Statistical References". In *Simulation 2001 (Proceedings of the 4th St. Petersburg Workshop on Simulation)*, edited by S. M. Ermakor, Yu. N. Kashtanov, and V. B. Melas, 109-118. Publisher: Chemistry Research Institute of St. Petersburg University.

Sargent, R. G. 2001b. "Some Approaches and Paradigms for Verifying and Validating Simulation Models". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J Medeiros, and M. W. Rohrer, 106-114. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2013. "Verification and Validation of Simulation Models". *Journal of Simulation* 7(1):12-24.

Sargent, R. G. 2015a. "An Introductory Tutorial on Verification and Validation Simulation Models". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1729-1740. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G. 2015b. "An Interval Statistical Procedure for Use in Validation of Simulation Models". *Journal of Simulation* 9(3): 232-237.

Sargent, R. G. and O. Balci. 2017. "History of Verification and Validation of Simulation Models". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zachareqicz, N. Mustafee, G. Wainer, and E. Page, 292-307. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G., D. M. Goldsman, and T. Yaacoub. 2015. "Use of the Interval Statistical Procedure for Simulation Model Validation". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 60-72. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Sargent, R. G., D. M. Goldsman, and T. Yaacoub. 2016. "A Tutorial on the Operational Validation of Simulation Models". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 163-177. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Schruben, L. W. 1980. "Establishing the Credibility of Simulation Models". *Simulation* 34 (3): 101-105.

Shanthikumar, J. G. and R. G. Sargent. 1983. "A Unifying View of Hybrid Simulation/Analytic Models and Modelling". *Operations Research* 31(6):1030-1052.

Whitner, R. B. and O. Balci. 1989. "Guidelines for Selecting and Using Simulation Model Verification Techniques". In *Proceedings of the 1989 Winter Simulation Conference*, edited by E. A. MacNair, K. J. Musselman, and P. Heidelberger, 559-568. Piscataway, New Jersey: Institute of Electrical and Electronic Engineers Inc.

Zeigler, B. P. 1976. *Theory of Modelling and Simulation*. 1st ed. New York: John Wiley and Sons, Inc.

Zeigler, B. P., A. Muzy, and E. Kofman. 2018. *Theory of Modelling and Simulation*: *Discrete Event & Iterative System Computational Foundations.* 3rd ed. London: Academic Press.

AUTHOR BIOGRAPHY

ROBERT G. SARGENT is a Professor Emeritus of Syracuse University. He received his education at The University of Michigan. Dr. Sargent has made numerous contributions to the simulation community in the forms of both scholarly publications and service work. His service work includes being the General Chair of the 1977 Winter Simulation Conference (WSC), serving on and chairing the WSC Board of Directors, being a Department Editor for the *Communications of the ACM*, holding the Presidency of what is now the INFORMS Simulation Society (I-SIM), serving as the Founding President of the WSC Foundation, and being the Founding of the Computer Simulation Archive. His many honors and awards include being a recipient of the 1988 Distinguished Service Award and the 2002 Lifetime Professional Achievement Award from I-SIM; the 2010 Board of Directors James R Wilson Award, the 40th Anniversary Landmark Paper Award, and the 50th Anniversary Titan Speaker from WSC; and the 2012 Distinguished Contributions Award from ACM/SIGSIM. Moreover, he was elected a Fellow of INFORMS in 2005, and received the University of Michigan College of Engineering Alumni Merit Award for the Department of Industrial and Operations Engineering in 2018. His email address is `rsargent@syr.edu`.