Proceedings of the 1989 Winter Simulation Conference
E.A. MacNair, K.J. Musselman, P. Heidelberger (eds.)

TESTING, UNDERSTANDING, AND VALIDATING COMPLEX SIMULATION MODELS

Jerry Banks
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

ABSTRACT

This paper prescribes new methods for testing, understanding, and validating complex simulation models. The new methodology was developed by drawing analogies from systems in general, drawing inferences from specific systems, using statistical methods in novel ways, and by extending current validation methods.

1. INTRODUCTION

This presentation had its genesis in a visit by E.B. Vandiver, Director of the U.S. Army Concepts Analysis Agency (CAA) to Georgia Tech. During that visit, Mr. Vandiver expressed interest in developing new methods for verifying and validating (V/V) the extremely large simulation models developed and used by CAA. CAA models the highest levels of warfare; the division and above. These models are exercised to analyze various scenarios.

CAA models are extremely large, as much as 400,000 lines of SIMSCRIPT! Thus, ordinary techniques leading to a conclusive statement about the V/V of a model are not easily applied. Generally, it is impossible to conclude that a large, complex model of the sort at CAA is completely verified or completely validated. Numerous V/V techniques are applied, some of which are quite ingenious. Perhaps the closest that a CAA model achieves to being generally accepted as valid is to enjoy continued use for analysis and decision making over a long time span.

New methods were sought in four areas. Analogies from testing, understanding, and validating systems in general was one approach taken. In this approach, diverse systems such as banking, medical diagnosis, and the promotion and tenure system in a university were examined. Inferences were drawn from each of the systems. These inferences were then summarized into 64 constructs that could be employed in the V/V of complex simulation models.

A second area of investigation was specific large systems that use simulation in some way to increase their operational effectiveness. These large systems included the Strategic Defense Initiative, Marshall Space Flight Center, and the Plant Hatch nuclear reactor, to mention a few. Even though the purposes of these large systems is quite different than the purpose of CAA, numerous inferences were drawn to the validation of the large, complex models.

A third area of investigation was the use of statistical methods in new ways. One of the methods suggested is control charting. Examples were given concerning the application of control chart techniques to both the input and output data of one of the smaller models used by CAA. Several other statistical methods were described for CAA models and examples were shown concerning their application.

The last area of investigation was the extension of current validation methods. Most of the modeling agencies in the Army were queried to determine ingenious methods they were using for V/V. One of these methods is described in this paper.

2. ANALOGIES FROM TESTING, UNDERSTANDING, AND VALIDATING SYSTEMS IN GENERAL

The analogies are drawn from an examination of banking, medical examination, drug testing, restaurant reviewing, academic promotion and tenure, business, religion, election for political office, building construction, medical diagnosis, the Ph.D. program, food and nutrition, the airline industry, personal transportation, fire fighting, marketing, and others. All of these systems have characteristics that define them and procedures for maintaining high standards of performance, accountability, and reliability. These characteristics, procedures, and techniques are examined and analogies are drawn from them for applicability to the V/V of complex simulation models.

An example will clarify the manner in which the analogies are drawn. Consider the airline industry. The major airlines are continually validated by the service they provide. While there are countless variables over which they have no control, one area that can be influenced is the performance of the pilot/aircraft system.

The airlines' major source of pilots is the military. By the time an individual seeks employment with a carrier, he or she has served a minimum of seven years on active duty (including flight school) and accumulated 1500-3000 hours of "stick" time depending on the type of aircraft flown. Once hired by an airline, the new pilot spends about three months training to become a flight engineer. The pilot is then assigned to a crew and placed on probation for one year subject to dismissal for safety violations or unsatisfactory progress. Once the probation period has been completed, the pilot is required to have a flight physical and pass a check ride. This pattern of revalidation continues as the pilot progresses from flight engineer to co-pilot and upgrades to larger aircraft. Once promoted to the position of captain, the pilot is still required to have a physical each year, but must take a check-ride semi-annually.

There are two analogies from this description of the pilot. First, a model should be revalidated at projected intervals. Moreover, the revalidation effort should be conducted at different levels. The entire model, as well as separate modules, need to be reevaluated. Based on new data from field tests, system performance and/or expert opinion, the model or one of its components may be inadequate or inaccurate.

Second, models should be designed with diagnostic features. If the performance of a module or subprogram can be monitored separately, revalidation will be quicker and more comprehensive.

In the complete document, the aircraft was also examined. The analogy to the aircraft is the CAA analyst. The analogies indicated that the analysts should be revalidated periodically and that they should be proficient with more than one type of model.

After looking at sixteen different systems, and developing some 64 inferences by analogy, it can be seen that this technique is quite helpful in generating new methods for V/V. Lastly, it was observed that the analogies began to repeat themselves, so looking at additional areas may not be very useful.

3. INFERENCES DRAWN FROM LARGE SPECIFIC SYSTEMS

There are exceptionally large and complex systems in operation that demand precise specification of design, performance, and output. Oil refineries, nuclear reactors, the space shuttle, and an aircraft carrier are all examples of complex systems that work well on a daily basis. Examination of complex systems like these can identify V/V techniques used during their development, and their ongoing operation, to provide inferences for large simulations of complex systems.

We are particularly interested in simulation models of these types of large systems. We examined simulation models used in five large systems and developed 36 inferences. The systems examined include:

Plant E.I. Hatch
Georgia Power Company
(Training simulator for a nuclear power generation station)

TAC Thunder
CACI
(Interactive theater level combat simulation model)

Strategic Defense System
MITRE Corporation
(V/V of the simulation of the much discussed Starwars system)

National Airspace System
MITRE Corporation
(Validation of the simulation model for the FAA)

Marshall Spaceflight Center (MSFC)
NASA
(Training simulations for the space shuttle)

Consider the last of these, the MSFC located at Huntsville, Alabama. Conversations were held with numerous persons that had been involved in the development of simulation models for training astronauts to operate the space shuttle. Two of the simulations discussed were the space telescope and SPACELAB. These training simulations are far different than the complex models used by CAA. However, some seven inferences resulted from this investigation. One example is as follows:

Utilize extreme condition tests whenever possible to evaluate the performance of algorithms and modules. Simulation runs should be conducted with all parameters and variables initialized across a wide spectrum of permissible values (low, middle, and high). Utilize median input values for parameters and compare simulation response to predicted response. Utilize truth data (output response known to be true for a given set of initial conditions) as a comparison data base for simulation output.

4. STATISTICAL METHODS

Output analysis provides a specified degree of confidence on accuracy. There are two limitations when modeling large complex systems. There is a high cost in computer and elapsed time in performing replications. There also is a frequent lack of real world data from which to draw comparisons. Thus, it is usually impossible to perform a complete statistical analysis.

Several existing statistical methods can be modified to add to the credibility of complex simulation models. Control charts, acceptance sampling, fractional factorial analysis, cluster analysis, and time-series analysis techniques were examined for applicability. An example of these methods is control charting.

The $\overline{X}$ and R charts are used to control the mean and dispersion of samples, respectively. The $\overline{X}$ and R charts are used to control the mean and dispersion of individual values. It is assumed that the underlying distribution is normal. For the $\overline{X}$ and R charts, with sample sizes greater than or equal to 5, this assumption is not very critical, as the distribution of means approaches normality (according to the central limit theorem).

We propose the use of control charts for monitoring both input and output data. For example, While a model is running, different characteristics can be monitored. If the characteristic exceeds the control limits, a warning message is sent. (If the characteristic being monitored is the rate of travel, values related to movement can be given as output, and an analyst can determine if there is an error in the input, an error in the algorithm for determining movement, or that there is no error.)

The advantages from using control charting are that no replications are required and that monitoring model characteristics ($\overline{X}$ and R, or X and R) lead to greater understanding of the model.

5. EXTENSION OF VALIDATION METHODS

The validation of large, complex simulation models is normally a very difficult process. Quantitative analysis of the output often cannot be conducted because detailed and reliable data from the real world system we are modeling is generally not available. The usual compromise is to perform some subjective tests and evaluations in order to establish the validity of the model's output. The results achieved are commonly along the lines of being able to please some of the people some of the time, but never close to pleasing all of the people all of the time. Some of the more common validation techniques are as follows:

1. Face validation
2. Event validation
3. Model assumption validation
4. Comparison with historical data
5. Comparison to other models
6. Consistency checks
7. Extreme condition tests
8. Sensitivity analysis
9. Turing tests

These techniques are discussed by Banks and Carson (1984) and by Carson (1986).

During the course of this research, innovative applications of these techniques were found as they pertain to large scale complex models, particularly models of combat. One application uses a Turing test to compare data that is being electronically collected from near-real combat scenarios conducted during instrumented brigade-level exercises at the National Training Center (NTC) in Ft. Irwin, California to data generated by JANUS, a combat development model. JANUS is a graphics intensive, interactive model which allows the user to issue tactical orders and objectives through workstations.

TRAC-MTRY, an Army modeling agency located in Monterey, CA, surveyed leaders of combat units that have rotated through NTC and of units stationed at the NTC which play the opposing forces during exercises. The survey instrument gave a brief, but concise, tactical summary and battle losses experienced by both forces in a brigade or task force level battle generated by either an actual NTC battle or by the JANUS simulation.

In summary, those surveyed were unable to distinguish between data generated by the NTC or JANUS. The percentage of individuals correctly identifying actual NTC battle information ranged from 20% to 36% over five sets of battle data. A nearly identical range of 19% to 36% correctly identified JANUS generated data with over 50% answering "don't know" to the question of the source of the data. In fact, in two of the five data sets more respondents selected the JANUS replication as the actual NTC battle. This fact alone is valuable information in assessing a model's credibility.

6. SUMMARY

This presentation described the search for new V/V methods for large scale, complex simulation models. The search was made in four areas:

1. Analogies from testing, understanding, and validating systems in general.

2. Inferences drawn from large, specific system.

3. Use of statistical methods.

4. Extension of current validation methods.

Each of these generated many leads. The next step is to rank the ideas in their importance, for surely, all of them cannot be accomplished. The V/V budget is insufficient for the task, and, quite likely, some of the ideas are unworthy. After selecting ideas with potential, resources need to be allocated to implement those ideas on a pilot basis.

REFERENCES

Banks, Jerry, and J.S. Carson, II (1984) *Discrete-Event System Simulation*, Prentice-Hall, Englewood Cliffs.

Carson, John S. (1986) "Convincing Users of Model's Validity is Challenging Aspect of Modeler's Job," *Industrial Engineering*, Vol. 18, No. 6, pp. 74ff.

McQuie, Robert (1988) "Casualties in Wargames and Casualties in War (1941-1982)," *Phalanx*, Vol. 21, No. 4, pp. 19-20.

AUTHORS' BIOGRAPHY

JERRY BANKS is Associate Professor of Industrial and Systems Engineering. His area of interest is simulation languages and modeling. He is the author of numerous articles and books. His latest book is Getting Started with GPSS/H, co-authored with John Carson and John Sy, published by Wolverine Software Corporation. Recent books include Discrete-Event System Simulation, co-authored with John S. Carson, II, published by Prentice-Hall in 1984, Handbook of Tables and Graphs for the Industrial Engineer and Manager, co-authored with Russell G. Heikes, published by Reston Publishing Co., a division of Prentice-Hall also appearing in 1984, IBM PC Applications for the Industrial Engineer and Manager, co-authored with J. P. Spoerer and R. L. Collins, and published by Reston in 1986, Procurement and Inventory Systems Analysis, co-authored with W. J. Fabrycky, published by Prentice-Hall in 1987 and Principles of Quality Control, published by John Wiley & Sons in 1988. Dr. Banks received his Ph.D. in Engineering from the Oklahoma State University in 1966. He is an active member of many technical organizations. He serves as IIE's representative to the Board of Winter Simulation Conference and as Vice Chairman of the Board.

Jerry Banks
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332, U.S.A.
(404) 894-2312