Proceedings of the 1992 Winter Simulation Conference
ed. J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson

COMPARISON OF MODELS: EX POST FACTO VALIDATION/ACCEPTANCE?

David A. Diener
Heston R. Hicks
Lawrence L. Long

Department of Graduate Logistics Management
School of Logistics and Acquisition Management
Air Force Institute of Technology
Wright-Patterson AFB OH 45433-6583

ABSTRACT

This research is a comprehensive investigation of model comparison using two air base operability (ABO) models. ABO model comparisons are common and, although such a methodology is generally not considered valid by simulation practitioners, there may be some real benefits to developing and documenting a suitable methodology. The process is not intended to replace model "validation," but we recognize that validation is often left undone or partially done. An appropriate ex post facto attempt at establishing model acceptability/validity via comparison could be most beneficial in contribution to the credibility of candidate models.

1 INTRODUCTION

Validation is recognized as a crucial step before using the results of a simulation study/model. Banks and Carson identify this step as perhaps the most important one in the process of carrying out a simulation study "because an invalid model is going to lead to erroneous results, which if implemented could be dangerous, costly, or both" (Banks and Carson, 1984). However, as important as validation is, it is also elusive in many cases. Many experts point out that complete or absolute validation can only be realized for simple models if at all (Law and Kelton, 1991; Williams and Sikora, 1991). Some classes of models are very complex and real world data nearly impossible to obtain. An example is the class of models used by the Department of Defense to study Air Base Operability (ABO). ABO models are developed as general purpose analysis tools to study the impact of wartime attacks on the ability of an air base to launch aircraft. The validation of these large, complex models is a difficult but important issue.

Simulation models should be assessed by their using organizations during the development phase of the model's life cycle. Many have pointed out the importance of user involvement in the model development stages, thereby enabling the user to simultaneously develop a sense of confidence in the structure and capabilities of the model (Balci, 1989; Law and Kelton, 1991; Banks and Carson, 1984). Unfortunately it is almost impossible for all the potential users of an ABO model to participate in model development. These models are developed over long periods of time, usually several years, and continue to evolve through on-going refinements. Developers provide these models to other users so that they can study particular, but wide-ranging, facets of ABO. These users then want to know whether these models are indeed "valid" and acceptable to use for their specific purpose. It would appear that there can be no blanket evaluation of a model that applies to every situation; however, some degree of "acceptability" is certainly desirable.

Simulation model predictive accuracy is, and should be, of key concern to serious model users. Simulation model acceptance, assumed based on user confidence of accuracy, appears to be directly influenced by verification, validation, credibility and accreditation. The simulation model user ultimately accepts or rejects a model as sufficiently accurate for use in supporting the decision making process.

The literature covers many different methods for assessing models using expert opinion, exhaustive analytical means, and real-system data. But in some instances, one or more of these alternatives is not available. Under these conditions there is a need for innovative assessment capabilities. We believe one alternative is the qualitative and quantitative comparison of similar models. We have found little documentary evidence that this type of assessment is frequently used as a formal tool; however we think that it does occur often in an informal sense. We believe, however, that a lack of real system data and the presence of a currently accepted model makes this a viable alternative for assessing the confidence decision-makers should place in unfamiliar models.

2 ACCREDITATION: A MORE MEANINGFUL CONCEPT

Williams and Sikora (1991) discuss the efforts of the Department of Defense to address how to answer that frequently asked question, "Has your model been validated?" They discuss some definitions that were developed by service and industry representatives that expand the idea of validation. They recognize that "validation is not an event, but a process" and that "a complex digital simulation can achieve degrees of validation, but true validation is a goal that can probably never be reached. Therefore it would be improper to refer to such a model as 'validated'."

Williams and Sikora (1991), recognizing the difficulty and elusiveness of the concept of validation, continue by defining a more meaningful term: accreditation. Accreditation is "an official determination that a model is acceptable for a specific purpose." Expounding further, they point out that accreditation is a "decision that a given level of validation insufficient for a model to be used in a particular application." Also to be considered is the importance of the decision being supported in determining the appropriate level of validation. Williams and Sikora continue to develop a framework for the concept of accreditation and state that work is continuing in this area.

Williams and Sikora (1991) recognize that many factors contribute to an accreditation decision, including the credibility of the candidate model. In turn, credibility is achieved through a number of factors which could be termed model reputation or performance history. Williams and Sikora list "the credibility of the model developers, the number of successful applications that have been made of the model, [and] the number of current users of the model" as some of the important credibility factors. One factor that contributes to establishing this credibility of a model is what we term "model comparison." In other words, does a model compare favorably in terms of output and features to an established, already accepted model?

3 MODEL COMPARISON - A REVIEW

Research has been done at the Air Force Institute of Technology concerning ABO model comparisons. This research recognizes that analysts considering the use of a new or unknown model often ask the question "How does the model compare to [a specific model they use or know about]?" rather than the question "Has the model been validated?" Models gain credibility and acceptance by familiarity and a favorable comparison to an "institutionalized" model leads to what we have termed "ex post facto validation." Although there is evidence to support the notion that this process takes place, there is little, if any, documentation of the formal process. Below is a short review of ABO model comparison in the literature.

Fossett et al (1991), in a discussion of validation of Department of Defense simulation models, state that there is evidence that model comparisons are used and may provide benefits: "We did, however, find evidence of an effort to validate the COMO III model by comparing its results to those from an Air Force model called SORTIE. The reasonable agreement of results when simulating similar conditions suggests that model-to-model validation can marginally strengthen credibility, especially when comparisons with real-world data are lacking."

A series of master's theses at AFIT have explored pairwise comparisons of various ABO models. In 1986, Noble compared the TSAR model to the Logistics Composite Model (LCOM). [Note: TSAR (Theater Simulation of Airbase Resources) and TSARINA (TSAR INputs Using AIDA) are ABO models developed by Rand Corporation for the U.S. Air Force.] Noble's stated purpose was "to begin the process of demonstrating the utility of the TSAR model to the Air Force manpower requirements analysts. In particular, the study aimed to show whether or not TSAR outputs could match those of the LCOM model, which has been a standard for manpower requirements forecasts for many years." He recognized that LCOM had many limitations and that "while TSAR includes many useful extensions that LCOM lacks, analysts are reluctant to use it exclusively because its ability to match the forecasting ability and suitability of LCOM, in those areas where the two models duplicate capability has yet to be proven." He summarizes the motivation for the comparison by stating "In short, maintaining both models results in a great deal of duplication, which is costly and inefficient. If users had proof that TSAR's outputs were comparable to LCOM's, TSAR could be used exclusively, in most applications."

Noble's study (1986) was inconclusive due to difficulties in aligning the databases of the two models. His effort, however, prompted two additional studies. As reported by Clark (1987), a follow-on study to Noble's thesis was conducted by Simulation Modeling Consultants (SMC). Their purpose was to manually create a TSAR data base from a simple LCOM data base and compare the manpower output. However, the SMC study also had difficulties because of database differences and the inability of the simulation to reach a steady-state.

Clark (1987) was the second follow-on to Noble's thesis (1986). He recognized the "institutionalization" of the LCOM model, suggesting that it was the standard against which similar models must be compared in order to gain acceptance. He states "If it can be substantiated that TSAR is as acceptable a predictor as LCOM of manpower requirements, the manpower analysts' ability to model wartime manpower requirements could be enhanced by TSAR's unique features." While improving significantly on the experimental design used by Noble, Clark found mixed results in differences between model outputs depending on the indicator of interest. He reemphasizes the importance of matching the chosen tool [model] to the particular problem at hand.

No further documentation of the TSAR-LCOM comparison are evident following Clark's work in 1987. Both models continue to be used by their respective user communities. In the 1988-1989 timeframe, a new ABO model was introduced called the Combat Base Assessment Model or CBAM. Users began to question whether they should use this new model. A report by Orlando Technology, Inc. (1989) addresses the issue of model comparison directly. Their stated purpose was to "evaluate the utility of the Combat Base Assessment Model (CBAM) .... This study was primarily a comparison to TSAR/TSARINA. The TSARINA attack simulation and the TSAR sortie generation simulation have, over time, become the de facto AF Standard ABO models." The Orlando Technology report goes on to recognize that the comparison of output trends of the two models would provide some insight into model differences and evaluation of how each contributes to ABO. They carefully and correctly avoid saying "validation," but in essence, their work certainly provided the ABO user community with insights into the "acceptability" of CBAM, especially those who accept the TSAR/TSARINA models. It is interesting to note that when TSAR/TSARINA were introduced, LCOM was "the standard of comparison," and now, through use and reputation, TSAR/TSARINA have emerged as the "de facto standards."

In 1989, yet another suite of ABO models emerged called the All Mobile Tactical Air Force or AMTAF. Leonhardt's thesis (1991), with support of an agency that was an LCOM user, addresses a comparison of AMTAF to LCOM. As he states, "This dilemma [Noble's assertion that LCOM and TSAR cannot be proven to duplicate results] has forced analysts to maintain both models resulting in duplication of effort. Since AMTAF has most (if not all) of the capabilities of TSAR, a case can be made that AMTAF could be used in place of the other two models as the single most reliable forecasting tool for aircraft resources, if it can be shown to be comparable to the forecasting abilities of LCOM." Leonhardt found, given limited variability in a factorial design, that "there is no significant difference between selected output variables produced by each model when the initial inputs to the models are as identical as possible given the inherent logic differences between the two models." he recognized that much more needs to be done to exclusively determine comparability of the two models.

A current project that will serve as a step towards evaluating model comparison as a validation/acceptance technique is underway at AFIT. We are doing a thorough comparison of the TSAR/TSARINA models to the comparable portion of the AMTAF suite of models in order to develop and test a model comparison methodology. The current focus of this work is on the base-level sortie generation process captured by TSAR and the SORGEN (sortie generation) module of AMTAF.

4 MODEL COMPARISON - A METHODOLOGY

The motivation for this research is the general acknowledgement of the use of model comparisons within the ABO community as a means of deciding whether or not to accept and use a new model. Our specific focus concerns the TSAR simulation model which is currently used by analysts while a newer model, the AMTAF simulation model, which the Air Force procured to improve its mission area planning capability and is purportedly easier to use, sits on the shelf. Both models possess the ability to simulate the capability of an airbase to generate and sustain sorties under wartime conditions. Estimates produced by the TSAR model are used and trusted while little is known about the capability of AMTAF to produce similar data. Since formal simulation development verification and validation activities have not produced sufficient user confidence to employ the AMTAF model, an alternative approach is indicated. Since the TSAR model is a used and accepted simulation of airbase operability and sortie generation estimation, a direct comparison of the common functional performance of AMTAF to TSAR provides a basis from which to assess equivalence. Our questions are: How can AMTAF and TSAR be compared to determine the extent to which they are equivalent? What constitutes a sound model comparison/assessment methodology?

4.1 Approach

Earlier efforts in model comparison have focused solely on the quantitative statistical equivalence of models. We believe a more robust comparison is possible in the quantitative arena and believe the addition of a qualitative comparison is essential to fully assess overall model equivalences. The foundation of our experimental framework is based on the construct found in two GAO reports shown in Figure 1 (GAO 1979, GAO 1987). Figure 2 is the pictorial representation of our experimental framework as adapted from Figure 1.

The common basis of the two diagrams is model documentation. The design, programmer, analyst, and user manuals are the key to the evaluation of any model. The proposed evaluation and research efforts rely on the depth and accuracy of the model documentation to support the comparison of AMTAF and TSAR. Since comparison of existing models is necessarily done after the model development phase has ended, it is extremely important that all available documentation be acquired and studied in depth to fully comprehend the intended use, features, and operation of the models under study. As shown in Figure 2, the evaluation/comparison process consists of three qualitative areas and one quantitative area.

4.2 Qualitative Comparison and Analysis

A qualitative evaluation of a simulation model is necessary to provide the potential user the requisite level of knowledge and understanding to competently employ the model and its capability. The potential user can establish this basic knowledge by conducting an in-depth review of the model's documentation. In Figure 2, three distinct areas comprise the qualitative comparison of SORGEN and TSAR:

1. Simulation Background and Documentation
2. Simulation Features and Input Database
3. Simulation Model Useability

4.2.1 Background and Documentation Comparison

A basic knowledge of each model's evolution is necessary to understanding what the models were originally intended to simulate. The researchers review the history of the models to determine when the models were originally developed, their original purposes for development and their evolution since then. We also research the documentation for both SORGEN and TSAR to learn the general classification of the models (e.g., theater, base, mission, etc.). Finally, we review the environment in which the models take place (i.e., the operational environment the models simulate), as well as the external environment (i.e., the computer systems on which the models operate). These findings are then documented and provide a general description of each model under study that may be referred to in future study or by other researchers or model users.

Next, each model's documentation is compared in five general areas: structure, level of detail, illustrations (diagrams and figures), examples, and ease of comprehension (clarity). This analysis is subjectively based on the combined perceptions of the researchers. The strengths and weaknesses of each set of documentation are recorded and the two sets of documentation are compared and contrasted to establish the degree of similarity. The overall intent of this analysis is to provide a measure of the suitability of the documentation and to assess its usefulness at the manager, analyst, and programmer levels.

4.2.2 Features and Input Database Comparison

The qualitative model comparison methodology continues with a comparison of features and databases. To compare the simulation models' specific features, capabilities, and characteristics, a catalog or listing of the features is assembled relying solely on the simulation models' written documentation to gain knowledge and insight into the models' individual capabilities. Comparison of the models' databases is accomplished by identifying common data requirements and mapping the data locations within each of the respective databases. Identifying the purpose and location of each data element provides the basis for the translation of a database for the quantitative comparison of the two simulation models.

A list of each model's capabilities is constructed and the two lists merged and sorted alphabetically to produce a combined listing of simulation features in a table. Assigned to each feature is a yes/no indication

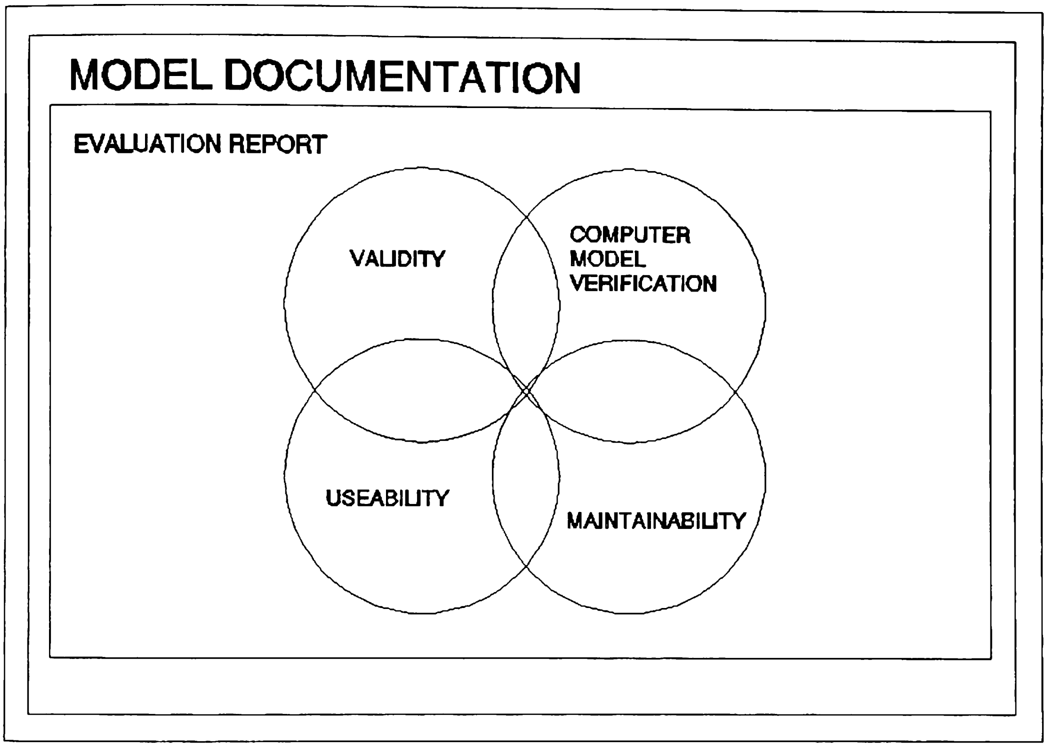

Figure1. Interrelationships Among Evaluation Criteria (GAO, 1979)

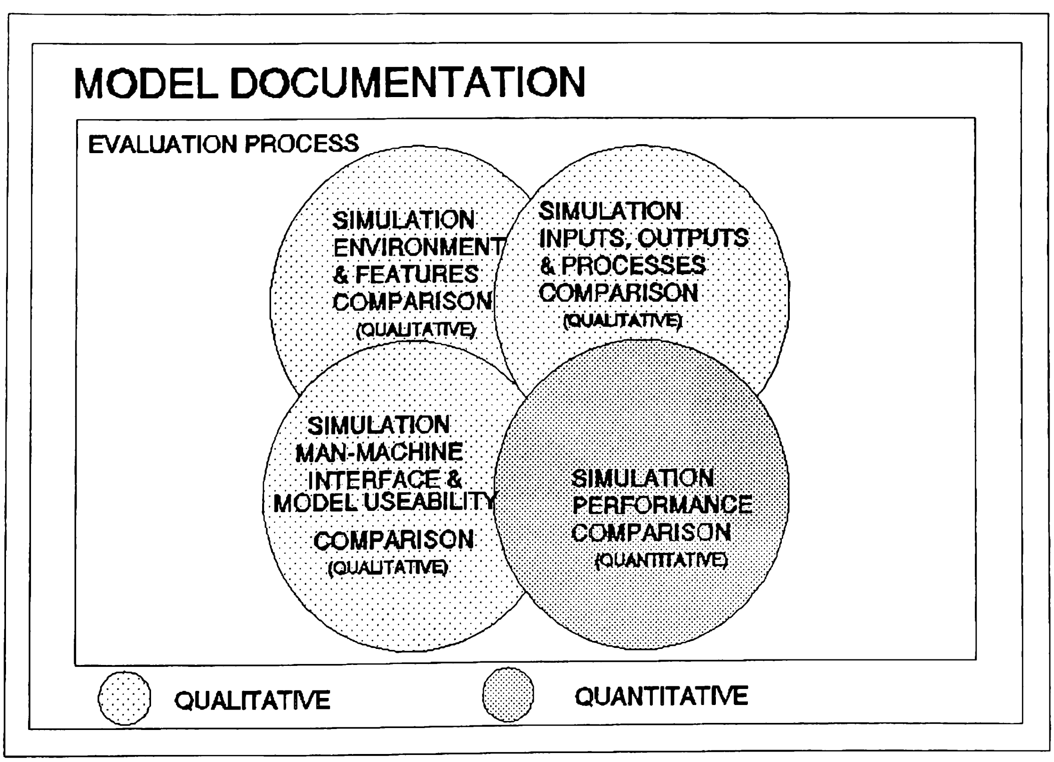

Figure 2. Construct for Model Comparison

of its presence or absence in each model. The table provides a tool from which to compare the capabilities of the two simulation models. In those areas where a simple yes or no is less than adequate in describing the presence or absence of a feature, exceptions are indicated and footnoted specifically within the context of the table.

The TSAR and AMTAF simulation models are driven by databases that provide information describing base facilities, personnel, aircraft, maintenance, supplies, policy and so on. Any similarity in database structure and content should indicate commonality between the simulation models. Since the goal of this research is to determine the extent to which the SORGEN module of AMTAF is equivalent to TSAR, it is necessary to construct databases that are equivalent to facilitate the quantitative comparison of the models. Since TSAR is the more widely accepted of the two simulation models, the representative TSAR database is translated to one useable by SORGEN. Identification of the required data elements and the determination of where they reside in the respective databases are critical to the translation process. To accommodate the database translation process and to expand the qualitative comparison of the models the two databases are mapped or cross-indexed.

4.2.3 Model Useability Comparison

An assessment of the useability, or ease with which a model is employed, may depend on several factors, among them the users' past experience in modeling, preferences for menu driven versus non-menu driven software, and the depth of study prior to the first modeling attempt. Since the determination of useability is different for individual users, it is appropriate to conduct a subjective evaluation of model useability over the course of the entire research project. The useability assessment is necessarily longitudinal in nature and the model user should be conscious of its importance from the beginning of the model comparison. Six categories of useability factors are subjectively assessed that are important to the success of the simulation effort. First, the differences that exist for each model in terms of the difficulty in learning to run the models. Second, the problems encountered during the production of databases for each model and also the difficulty experienced with manipulating the needed simulation data reports. Next, the availability of debugging tools with each model and the ease with which these are used to correct problems. Fourth, the problems encountered in implementing the experimental design, which includes the ease with which the model accommodates the manipulation of input data to achieve a successful experimental design, and whether multiple simulation runs may be submitted concurrently. Fifth, the time required to achieve each run, if batch submissions may be made, and whether the user's presence is necessary throughout a run. Finally, the adequacy of the simulations output, the content and format of system generated output, whether the user is provided with an ability to define output reports to suit specific needs, and whether the simulations provide the type of data necessary for the user to make informed decisions based on the results. Useability marks the beginning of the second category of the study, quantitative comparison.

4.3 Quantitative Comparison and Analysis

The quantitative comparison of simulation output provides the model user with another class of information on which to evaluate the equivalency of the models under study and answers the investigative question: Are the model outputs equivalent given equivalent inputs? Figure 3 provides a pictorial view of the quantitative comparison of the models.

The Measure of Merit. Conducting a quantitative comparison requires common numerical measures of overall performance. At the highest level of comparison, a single measure of the effects of all aspects of the simulation environment would be desirable. Since the TSAR and SORGEN models are designed to simulate the operation of a military air base, sorties generated (i.e., the number of combat missions flown during the simulation), provides an overall meaningful measure of air base operations. Assuming the simulation models are structured such that the interactive air base functions influence overall sortie production, sorties generated will be the measure of merit for the purpose of quantitatively comparing TSAR and SORGEN.

Database Translation and Equilibration. Quantitatively comparing TSAR and SORGEN requires a common input usable by both models. Much of the research conducted in the qualitative analysis contributes directly to this activity. The development of the table of features to permit comparing the models from a functional perspective and the preparation of the database mapping tools provides the mechanisms to support the development of equivalent input databases.

There are several databases available for TSAR that have been developed for the purpose of establishing policy, determining manning and resource levels for

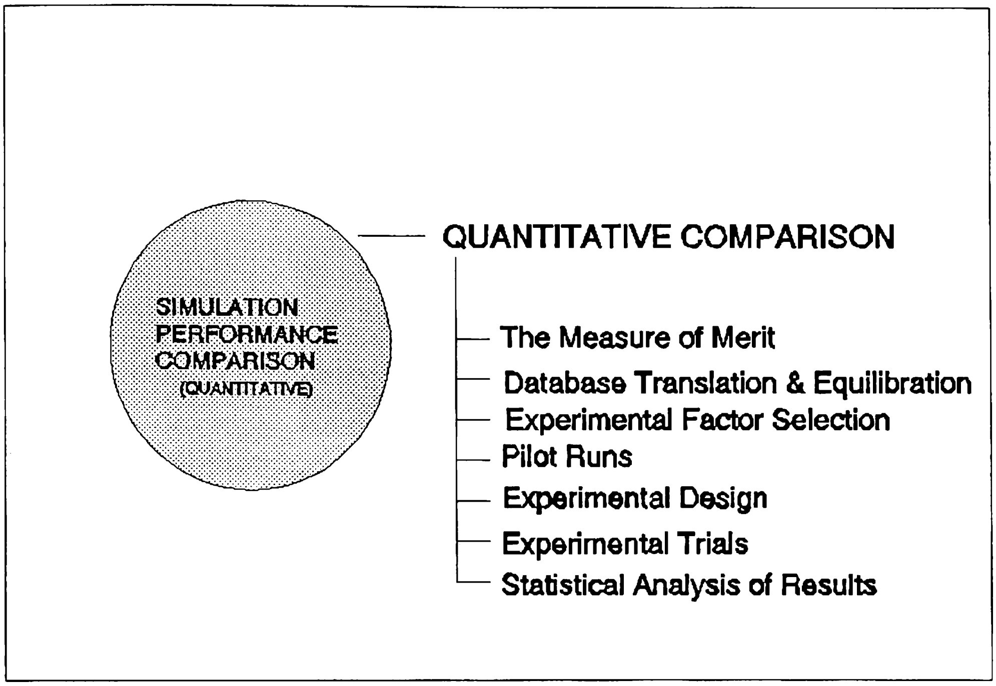

Figure 3. Quantitative Comparison

theater operations, and for studying the effect of enemy attacks on air bases. Some of these databases have been validated through actual operational tests of the modeled environment. The F-15 TSAR database has been thoroughly tested and used previously in extensive research and therefore, was chosen as the basis from which to develop equivalent databases.

Dissimilarities in the simulation models and their databases should be no surprise. When a feature is encountered that is not in both models or is not obviously replicated equally, the lesser capability becomes the standard and the model possessing extended capability in that particular function is constrained to a level that, as nearly as possible, equates it with the lesser capability. Where differences in fidelity (level of detail) are encountered, the data inputs are aggregated or disaggregated, where possible, to exercise the maximum number of common functions. This process is designed to exercise TSAR and SORGEN as thoroughly as possible.

Experimental Factor Selection. Several factors representative of air base operations must be chosen to build the experimental design for this research. These factors must be clearly defined in the input databases and capable of being varied to facilitate experimentation. In addition, the following criteria are used in the selection of factors:

1. Each factor must be present in each of the simulation models or capable of being implemented with a high degree of equality.

2. Each factor must have a direct impact on sortie generation and related to the logistics infrastructure.

3. Each factor must be directly influenced by operational demands (i.e., increased sortie rates) or use of more logistics resources.

4. The set of factors establishes a broad inference space (i.e., a significant portion of the logistics infrastructure is encompassed by the factors chosen).

Preliminary study by the researchers show that the two simulation models have a high degree of commonality in input data. Earlier research done on the TSAR model (Diener, 1989) identified factors which could be used in this research. These efforts narrowed our focus to eight factors which meet the above criteria:

1. Aircraft
2. POL (fuel)
3. Munitions
4. Missions
5. Personnel
6. Spares
7. AIS (avionics intermediate shops)
8. Support equipment

Experimental Design. A range of logistics scenarios is desired to assess comparability over a wide and realistic inference space. Thus a two-level factorial experiment is used to evaluate quantitative similarities of the two models. Therefore, levels for each of the eight factors must be selected.

For Factors 1, 5, 6, 7, and 8, the high levels are indicative of real world values based on the experience of the researchers. The low-level values are set at a 25% reduction of the high level values. Factors 2 (POL) and 3 (munitions) represent different quantities of resupply. The high-level resupply provides 90% of the POL and munitions required if 100% of the high-level mission demand were achieved. The low-level resupply provides 75% of the POL and munitions required if 100% of the high-level mission demand were achieved. Factor 4 (missions) reflects a sortie demand schedule. The value initially selected for the high level establishes a maximum sortie demand of 4.7 sorties per aircraft per day. The low level establishes a maximum sortie demand of 2.0 sorties per aircraft per day.

All of the factor levels are evaluated using a series of pilot runs designed to test the sensitivity of the models to each of the factor levels individually. Adjustments of the factor levels are made as necessary to insure they fall within the operational capability of both TSAR and SORGEN.

The selection of eight factors means that the size of a full factorial experiment would be 2^8 (256), given a two-level design. To reduce the number of treatments necessary to achieve a design that allows all two-way interactions of the factors to be examined, we use a 1/4 replication of the full factorial array (McLean & Anderson, 1984) which reduces the number of design points from 256 to 64. NOTE: The resolution of the fractional factorial to include estimation of all two-way interactions was chosen to foster parallel research on estimating metamodels from simulation data.

Experimental Trials. Each simulated treatment provides a cumulative average number of sorties generated over a 30-day period and the associated standard deviation. These figures are collected for each treatment for use in the quantitative comparison.

Another measure model users should be interested in when comparing models is that of run length, or time needed to accomplish individual simulation runs. This is important in terms of the amount of computer time logged, which could be costly if the organization leases time on a computer system. The times are collected and simple summary statistics are calculated followed by a paired difference test to assess whether a difference exists in the length of times both models run for the same treatments.

Statistical Analysis of Results. The statistical analysis of the simulation results concludes the experimental methodology. A paired difference test is used to evaluate whether a difference exists between the identical treatments run on TSAR and SORGEN. The paired difference test allows the evaluation of whether a significant difference in output exists between TSAR and SORGEN when comparing the difference of identical treatments.

4.4 Summary

This methodology embraces a more complete comparison of two models than attempted in previous research. The qualitative comparison is important because of its need to establish a level of confidence in AMTAF that has not been achieved earlier. The qualitative comparison provides the information necessary for potential users to assess the traits and characteristics of AMTAF and TSAR with respect to the environment they are intended to simulate. The comparison of input and output characteristics enables potential users to assess the suitability of the two models to his/her needs. Knowing what it takes to run the model and the outputs produced should prove useful in determining the suitability of the models to a particular purpose.

The quantitative comparison of AMTAF and TSAR establishes the level of equivalence between the two models in a specific functional area. This research does not compare all of the functions of either model. Our analysis is constrained to evaluation of critical logistics factors and their interactive influence on simulation results, and to those factors capable of being modeled similarly.

5 CONCLUSION

ABO model comparisons are common and, although such a methodology is generally not considered valid by simulation practitioners, there may be some real benefits to developing and documenting a suitable methodology. This research furthers that purpose as the most comprehensive investigation of model comparison. This is not intended to replace model "validation," but we recognize that validation is often left undone or partially done. An appropriate ex post facto attempt at establishing model acceptability/validity via comparison could be most beneficial in contributing to the credibility of candidate models.

REFERENCES

Balci, Osman. 1989. "How to Assess the Acceptability and Credibility of Simulation Results," *Proceedings of the 1989 Winter Simulation Conference*: 62-71.

Banks, Jerry and John S. Carson, II. 1984. *Discrete-Event System Simulation*, Prentice-Hall.

Clark, Capt Gregg A. 1987. *The Theater Simulation of Airbase Resources and Logistics Composite Models: A Comparison*. MS Thesis, AFIT/GLM/LSM/87S-15. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH.

Diener, David A. 1989. *Forecasting Air Base Operability in a Hostile Environment: Estimating Metamodels From Large-Scale Simulations*. PhD Dissertation. Purdue University, West Lafayette IN.

Fossett, Christine A. and others. 1991. "An Assessment Procedure for Simulation Models: A Case Study," *Operations Research*, 39:710-723.

General Accounting Office. 1979. *Guidelines for Model Evaluations*, PAD-79-17.

____. 1987. *DOD Simulations: Improved Assessment Procedures Would Increase the Credibility of Results*, GAO/PEMD-88-3.

Hicks, Heston R. and Lawrence L. Long. 1992. *A Comparison of the Theater Simulation of Airbase Resources and the All Mobile Tactical Air Force models*. MS Thesis, AFIT/GLM/LSM/92S-XX. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH.

Law, Averill and W. David Kelton. 1991. *Simulation Modeling and Analysis*, McGraw-Hill.

Leonhardt, Capt David P. 1991. *A Comparison of the All Mobile Tactical Air Force and the Logistics Composite Simulation Models*. MS Thesis, AFIT/GLM/LSM/91S-42. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH.

McLean, Robert a. and Virgil L. Anderson. 1984. *Applied Factorial and Fractional Designs*, Marcel Dekker, Inc.

Noble, Capt David R. 1986. *Comparison of the TSAR Model to the LCOM Model*. MS Thesis, AFIT/GLM/LSM/86S-54. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH.

Orlando Technology, Inc. 1989. "ABO Model Comparisons - CBAM/TSAR".

Williams, Marion L. and James Sikora. 1991. "SIMVAL Minisymposium - A Report," *Phalanx - The Bulletin of Military Operations Research*, 24:1-6.

AUTHOR BIOGRAPHIES

DAVID A. DIENER, Lieutenant Colonel, USAF, is the Head, Department of Graduate Logistics Management, School of Logistics and Acquisition Management, Air Force Institute of Technology (AFIT), Wright-Patterson AFB OH. He received a BS in Management and Economics from the USAF Academy in 1976, an MS in Logistics Management from AFIT in 1980, and a PhD in Operations Management from Purdue University in 1989. His research interests concern the estimation and use of metamodels derived from large-scale simulations and the accreditation/acceptance of large-scale military logistics models.

HESTON R. HICKS, Captain, USAF, is a candidate for the degree of Master of Science in Logistics Management at the School of Logistics and Acquisition Management, Air Force Institute of Technology (AFIT), Wright-Patterson AFB OH. His research interests include the use and accreditation of simulation models by the United States Air Force. He is a logistics officer whose experience is in both logistics operations and acquisition logistics.

LAWRENCE L. LONG is a candidate for the degree of Master of Science in Logistics Management at the School of Logistics and Acquisition Management, Air Force Institute of Technology (AFIT), Wright-Patterson AFB OH. He is a career logistician serving the Department of the Air Force. His modeling interests lie in the simulation and analysis of logistics processes related tot he support of combat forces. Mr. Long is a member of Sigma Iota Epsilon, the National Honorary and Professional Management Fraternity.