NOMF

EXPOSURE DRAFT

GUIDELINES FOR MODEL EVALUATION

84469

S. GENERAL ACCOUNTING OFFICE WASHINGTON, D.C. 20548
D-79-17 JANUARY 1979

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON, D.C. 20548

B-115369

This document is intended to guide the decisionmaker in the difficult task of determining how much confidence to place in a model's results. A model which is a representation of either a real or proposed process or system can be used as an aid to assist evaluators and analysts in supporting decisionmakers. Thus, models can allow analysts and decisionmakers to anticipate the implications of various policies on particular issues or systems when such implications are not readily susceptible to analysis with other tools.

While modeling is an extremely useful analytical approach, it is important to avoid the temptation to view a model as a magical "black box" which automatically gives reliable, valid answers. For various reasons, decisionmakers sometimes use a model's results without being fully aware of the theories, assumptions, approximations, and judgments that went into the development of the model. Also, it may be unclear how these factors affect the validity and reliability of the model's predictions. For those cases in which the model is at least partially implemented by means of a computer, the fact that a computer can perform immense numbers of calculations and produce large amounts of output very rapidly may give a false air of reality to the results. All too often, decisionmakers fail to ask, "How much confidence should be placed in the results provided by the model?" And, when they ask this question, there is all too often no sound basis for an answer.

This document is directed toward the person(s) who must answer this question of model confidence. More specifically, this document provides guidelines for the accumulation of evidence on which to base reasonable opinions, conclusions, judgments, and recommendations concerning the confidence which can be given to a model's results. It should be useful in planning a model evaluation effort. It provides a general overview of model evaluation and also identifies some concerns which should be considered before the results of a modeling effort are used by a decisionmaker.

The full-scale evaluation of a complex model can be an expensive, time-consuming effort requiring diverse talents and skills. Ideally, any model whose results will be used in the decisionmaking process should be subjected to such an evaluation. In reality, this will not always be possible because of constraints on time or resources. In these cases, if the use of the model plays a significant part in the decisionmaking process, consideration should be given to some level of evaluation. Time may permit no more than a quick but careful look by an expert in the field, or it may be possible to perform some, but not all, of the detailed analysis described in this document. The results of whatever evaluation is performed should be provided to the decisionmaker, accompanied by an assessment (insofar as possible) of the risks which may be involved in using the model without a more extensive evaluation.

We hope that this document will both increase and improve communication among decisionmakers, evaluators, analysts, and model developers. These guidelines are presented as a working document. They will need refinement in the light of experience gained from efforts to use them. Accordingly, we invite comments and suggestions for improvement from readers and particularly from those whose comments are based on the actual application of these guidelines. Such comments may be addressed to the Program Analysis Division of the General Accounting Office.

Elmer B. Staats

Comptroller General
of the United States

C O N T E N T S

		Page
CHAPTER		
1	INTRODUCTION	1
	Background	1
	Model Evaluation - Perspective and Role	2
2	THE MODELING PROCESS AND MODEL EVALUATION	4
	The Modeling Steps	4
	Model Evaluation	5
3	AN APPROACH TO MODEL EVALUATION	9
	Documentation	11
	Validity	12
	Computer Model Verification	20
	Maintainability	22
	Usability	23
	Evaluation Report	24
4	SUMMARY	25
APPENDIX	Model Documentation Checklist	27

CHAPTER 1

INTRODUCTION

BACKGROUND

The Congressional Budget Act of 1974 requires, among other things, that "the Comptroller General shall develop and recommend to the Congress methods for review and evaluation of government programs and activities carried on under existing law." A guidance document we issued earlier contained some fundamental requirements for evaluation and analysis to support decisionmaking. 1/ This document addresses one aspect of that support--modeling.

GAO has been reviewing models, particularly military models, for a number of years. More recently, in 1974, the Chairman of the Committee on Science and Technology, U.S. House of Representatives, in a letter to the General Accounting Office, noted that much of the information in the Federal Energy Administration's soon-to-be-available Project Independence Blueprint "was obtained by the use of computer simulation models." The Chairman requested GAO "to undertake a thorough review and analysis of the methodology used in the computer programs..." and cited several specific interests. Thus, GAO became engaged in the comprehensive evaluation of large scale models with this congressional request to evaluate the Project Independence Evaluation System (PIES), the formal name of the model used to support development of the Blueprint. The difficulty of this task soon became apparent. There simply did not exist any guidelines, much less standards, outlining how one might proceed.

In performing the analysis, GAO assembled and reviewed program material, the status of model evaluation, and interviewed numerous experts in the general field of modeling. The GAO findings, relative to PIES, were documented in a report to the Congress. 2/ An important side-effect of that document was that a foundation was laid for model evaluation by GAO. In particular, the PIES report said GAO believes emphasis should be placed on three areas: (1) model verifi-

---

1/ *Evaluation and Analysis to Support Decisionmaking*, PAD-76-9, U.S. GAO, Washington, D.C., September 1, 1976.

2/ *Review of the 1974 Project Independence Evaluation System*, OPA-76-20, U.S. GAO, Washington, D.C., April 21, 1976.

cation/validation, (2) sensitivity testing, and (3) model documentation. Moreover, each of the three was identified as being "essential in developing a computer model."

This effort was followed by a GAO initiated project in which the Transfer Income Model (TRIM), a large scale model used in welfare policy analysis, was reviewed and evaluated. This project also resulted in a report to the Congress. 1/ In the evaluation portion of the report, GAO further developed and refined the criteria used in the PIES report.

The need for a capability to evaluate models, data collection, analysis, and general agency activities in these areas was also acknowledged by the Energy Conservation and Production Act (Public Law 94-385, August 14, 1976). This act created a six member Professional Audit Review Team (PART), chaired by GAO, to perform these functions. PART's report to the President and the Congress 2/ contributed to model evaluation by describing actions needed to improve the credibility of energy models and data.

The models which were the bases of these reports are but two of a large number of similar models used by the Federal Government to assist policy analysts and decisionmakers in shaping their policy recommendations and decisions. These models are similar in the sense that they are large scale computerized models designed or used to help in making decisions about public programs which involve the lives of millions of Americans and large amounts of Federal funds. They are termed large scale in the sense that the system which they represent is large--in number of parts, in number of different types of parts, in number of functions performed, in number of inputs, and in absolute cost. There are more such models being developed all the time. Because such complex policy analysis models are costly to develop and require large staffs to maintain and exercise, they are the primary subjects of this document.

MODEL EVALUATION - PERSPECTIVE AND ROLE

There is a growing recognition of the need to assess large scale models. This task is unlike the situation in

---

1/An Evaluation of the Use of the Transfer Income Model--TRIM--To Analyze Welfare Programs, PAD-78-14, U.S. GAO, Washington, D.C., November 25, 1977.

2/Activities of the Office of Energy Information and Analysis, Professional Audit Review Team, Washington, D.C., December 5, 1977.

which GAO might be asked to audit financial accounts of an agency. In that case there exist standards and accepted procedures, while for model evaluation there are no generally accepted standards or methods. Hence, GAO perceives the need to expand upon the lessons learned in evaluating PIES and TRIM. Hopefully, this will stimulate a continuing discussion within the modeling community of the need for such guidelines and standards.

Model evaluation should not be a purely retrospective task. If there have been no foundations laid and no thought given to evaluation until the model is complete, or nearly so, then the task of the auditor or evaluator is made more difficult. The familiarization and understanding of a model that is required in an evaluation might take months to accomplish in the absence of appropriate documentation. Therefore, it is very important that the evaluative aspects of model building be considered at the start of the project and be carried out during model development as well as after the model is operational. Thus, this document also will be useful to those persons active in model development. Model builders should realize at the beginning that their products may be evaluated in terms of a set of criteria such as that discussed in this document. In this manner the model builders will know what their models may be measured against and this, hopefully, will encourage them to meet the evaluative criteria as a natural product of good model building, i.e., model evaluation must be an ongoing process and of continuing concern.

The guidance provided in this document is general in that it is neither restricted to the evaluation of some specific modeling methodology nor is it dependent upon the model having some particular theoretical foundation. Indeed, a portion of any evaluation involves assessing whether the theoretical foundations of the methodology developed or adopted during the formulation of the model are appropriate. The final report on a model evaluation study should provide an independent assessment of (among other things): (1) whether or not and under what conditions the potential user should use the model for the purpose described to the evaluators by that user; and (2) how the model should be used.

It is the intent of GAO that this document and its future refinements be useful to persons charged with the task of evaluating a model. To place the proposed evaluation criteria in their proper perspective, this document first discusses the modeling process itself.

CHAPTER 2

THE MODELING PROCESS AND MODEL EVALUATION

For purposes of this document a model is a representation of the underlying structure of a process or system. The system might be conceptual, ideal, or real. In general, a model has a simple and/or manipulatable structure relative to the system it represents. By making explicit the implications of alternative assumptions regarding key relationships of the issue or system under study, a model can provide a clearer understanding of these relationships. This definition is very general and can be applied to many different types of "representation," from a toy car (which represents an automobile) to a full-scale prototype of a supersonic aircraft; and from the game of Monopoly (which represents the real estate business in Atlantic City) to a set of mathematical equations that represent the behavior of the national economy. All of the above examples are relatively concrete or tangible. It should be noted, however, that models often are used to represent concepts or ideas or conditions of the future which do not exist in any tangible form.

The types of models considered as the primary basis for developing these guidelines have the following general characteristics: (1) they are models that are developed to assist the policy analyst or decisionmaker in selecting or evaluating various policies regarding governmental issues and programs (e.g., social, economic, political, or military programs); (2) they are mathematical models of a complex system and have been computerized; and (3) they are large scale models. Ideally, any model whose results will be used in the decisionmaking process should be subjected to some level of evaluation using these guidelines to the extent possible.

THE MODELING STEPS

Before introducing the evaluation criteria, it is pertinent to review the basic steps that are involved in any modeling effort. The steps listed below have been adapted from an earlier GAO report 1/ which describes a comprehensive five-phased approach for improving the management of computerized model development activities. It should be noted that these steps are interrelated and not always performed in the exact sequence listed.

--Describing the problem to be solved; defining the problem issues, study objectives, and assumptions.

1/Ways to Improve Management of Federally Funded Computerized Models, LCD-75-111, U.S. GAO, Washington, D.C., August 23, 1976.

--Isolating the system or process to be modeled; delineating the characteristics which can be modeled.

--Developing or adopting a supporting theory; developing a flow or logic diagram.

--Determining available data sources; formulating the mathematical model or set of models to be linked; analyzing data requirements and designing data collection procedures.

--Collecting data.

--Describing the program logic of the model including basic flow charts with input, processing, and output described; estimating parameters in the model; constructing and implementing the computer program(s).

--Verifying that the mathematical/logical description of the problem is correct and that the corresponding computer program(s) has been coded correctly; debugging the computer program(s).

--Developing alternative solutions and analyzing them using the model.

--Evaluating results and output obtained from the model.

--Presenting results with a plan for implementing recommendations.

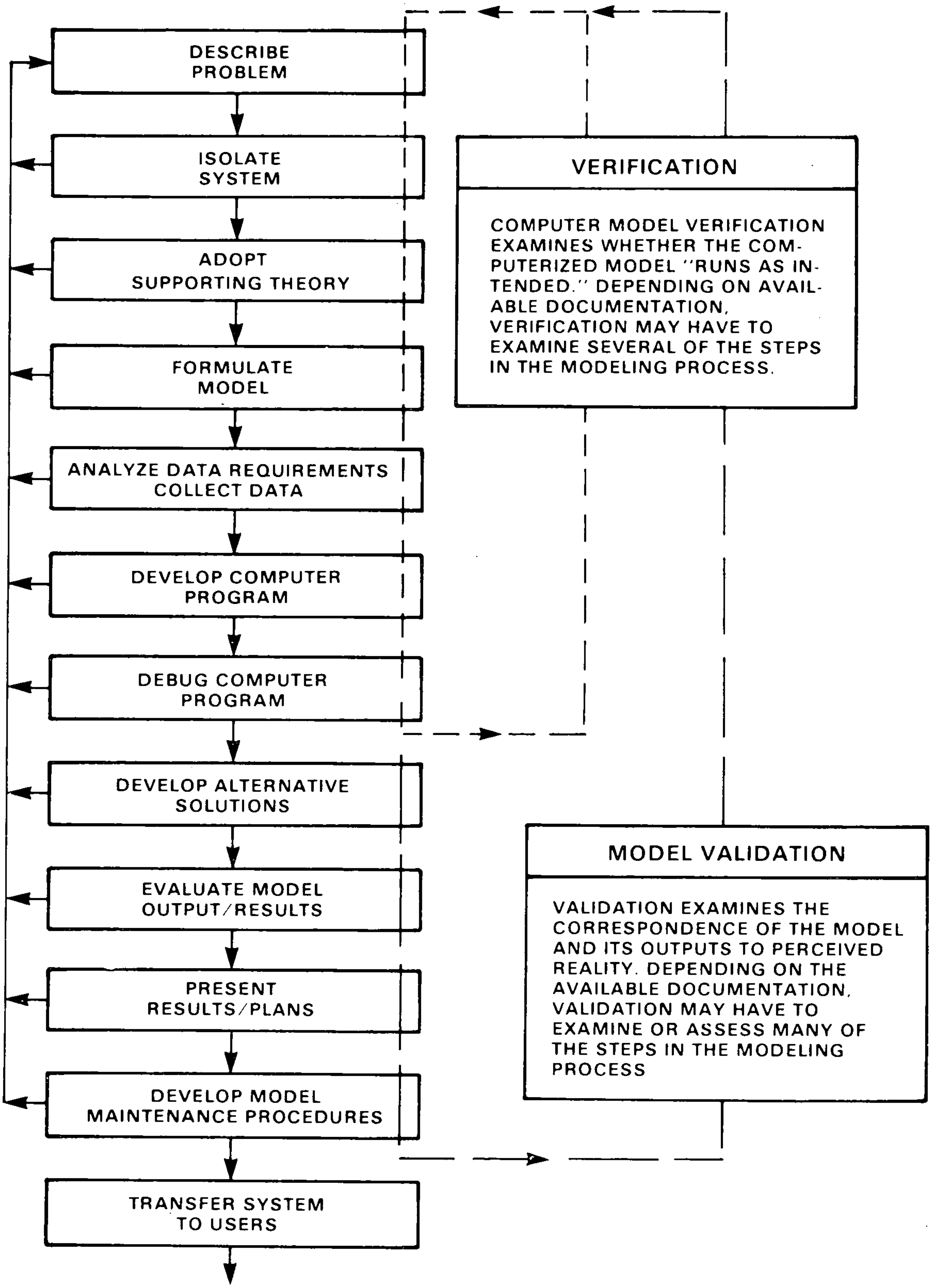
--Maintaining the model and data.

These steps in the modeling process, and their interrelationships are shown in figure 1. They are the responsibility of the model developers and sponsors, and if they are accomplished and documented in enough detail and scope, it should be less difficult for independent evaluators to review the appropriate documentation and to obtain the information needed to assess the model's validity. However, all too frequently adequate documentation is not available. This makes it more difficult for the evaluation team to understand the model and to conduct the evaluation.

MODEL EVALUATION

The use of complex, large scale models by many Government agencies is increasing due to better trained analysts and the development and refinement of analytical decisionmaking methodologies. At the upper levels of management, one cannot expect the diverse talents required by managers in senior positions to include a detailed understanding and apprecia-

FIGURE 1
BASIC STEPS IN THE MODELING PROCESS



tion of these methodologies. Thus, there is a need for procedures--or guidelines--by which independent investigators can evaluate models to assess the validity of a model's results so as to better guide the use and interpretation of the model by senior managers. Many models are of such size and complexity that evaluation by an individual working alone is effectively precluded. Thus, to evaluate large scale models in a reasonable amount of time, a multi-disciplinary evaluation team should be formed. The team should consist of personnel knowledgeable in the functional areas being modeled, the environment of the decisionmaker or other user of the model, mathematical modeling, computer science, and statistics.

Evaluation of a model, as the term is used here, does not mean second-guessing the intent and results of the model developers or sponsors. Rather, through evaluation, interested parties, involved or not involved in a model's origins, development, or use, can assess the model and its results by using an established set of criteria to accumulate evidence regarding the credibility and applicability of the model. Such a set is proposed and discussed at some length in the following chapter.

There are three primary concerns in advocating evaluation for complex models: (1) for many models, the ultimate decisionmaker is far removed from the modeling process (this is especially true in governmental areas) and a basis for accepting or rejecting the model's results by such a decisionmaker needs to be established; (2) users of a complex model developed for other purposes must be able to obtain a clear statement as to the applicability of the model to the new user's problem areas; and (3) for complex models, even if the decisionmakers and model developers have extensive interchange, it is difficult for the former to assess and to comprehend fully the results of carrying out the modeling steps (i.e., the interactions and impact of a model's assumptions, data availability, and other elements on the modeling process) without some formal, independent evaluation.

A final report on a model evaluation effort should include a statement summarizing the team's view of whether or not the model meets its design objectives and how the model should or should not be used. Such a report should include: statements concerning the model's <u>assumptions</u> and under what circumstances these assumptions hold, including remarks on the consistency of these assumptions and the completeness of the model; a review of the mathematical and <u>logical constructs</u> of the model; an analysis of the <u>data</u> utilized in terms of the data's original accuracy and appropriateness;

and a statement as to whether the total model environment (assumptions, data, computation, the model's assigned role in the decision process) is appropriate and accurate. Evaluation could, but need not always, include a detailed review of associated computer programs and their ability to perform. However, there is a need to state the steps which have (and have not) been made to assess computer program correctness and output reliability. Assumptions dealing with the programming specifications and program interfaces should also be assessed. Thus, the evaluation team certainly should avoid treating the model as a "black-box" that manipulates data; its concern is not with just what goes in and what comes out, but also includes data transformations and their rationale. The team should be able to replicate the model's results and analyze the sensitivity of results to changes in model assumptions and parameters.

It should be pointed out that very often there is a problem in the relevance of the evaluation to the "current" model, i.e., the model which exists at the time the reliability of its results is questioned. This can often be the case because there is a continuing effort to improve the model's performance. Moreover, the evaluation itself will generate certain suggested changes that may be adopted during the evaluation. The modifications which are made to improve a model can be changes in assumptions, data, etc. Thus, it is important for the evaluation team to specify the model version which is being evaluated and to communicate this to the model sponsor, if such is the arrangement.

The above discussion briefly indicated what model evaluation is. It is equally important to realize what it is not. It is not model certification. This issue is raised because model evaluation is sometimes mistakenly confused with model certification. Certification commonly refers to a guarantee that the model yields outputs or results that are suitably accurate for a particular application. This is an unattainable goal in dealing with the large scale models with which this document is concerned. Evaluation, on the other hand, acknowledges this limitation, and seeks to improve the model's usefulness by identifying its strengths, weaknesses, and appropriate uses as explicitly as possible. Finally, evaluation should, to the extent possible, recognize that these strengths, weaknesses, etc. need to be assessed in light of the alternative tools available to the potential user. Thus, the evaluation team should not ignore the possibility that a particular model may have important strengths and still be less useful than some other tool (another model, expert opinion, etc.). Conversely, a model may have significant limitations and still be the best analytical tool available for the immediate task at hand.

CHAPTER 3

AN APPROACH TO MODEL EVALUATION

The purpose of the present chapter is to provide a minimal set of criteria deemed necessary for model evaluation (see figure 2). These are based upon GAO's experience in program evaluation in general, upon GAO's evaluation of the PIES and TRIM models, and upon an extensive review of the available literature on model evaluation and program evaluation cited in the Bibliography (see in particular Gass, Evaluation of Complex Models, 1977, and Schellenberger, Criteria for Assessing Model Validity for Managerial Purposes, 1974).

FIGURE 2
CRITERIA FOR MODEL EVALUATION

DOCUMENTATION

VALIDITY

- Theoretical Validity
- Data Validity
- Operational Validity

COMPUTER MODEL VERIFICATION

MAINTAINABILITY

- Updating
- Review

USABILITY

It is very important to recognize that a model must not be judged only in the abstract against certain ideal goals. Careful consideration must be given also to its purposes, to the manner and the environment in which it is being or will be used, vis-a-vis other feasible approaches that might be used to solve the problem. What may be a relatively satisfactory operating model for one objective may be strikingly unsatisfactory for another.

Although the criteria to be discussed are based upon GAO's experience and reflect current thinking, subsequent experience probably will reveal that some or all of them are of greater or less importance than is now believed. They then would be subject to appropriate modifications or refinements. More generally, within the framework of the present discussion, this document provides guidelines for model evaluation which are themselves subject to evaluation and appropriate modification. This is both expected and welcomed. The result of this evolutionary process should improve the technology of model evaluation, and, ultimately, the usefulness of complex models in policy analysis and decisionmaking.

The proposed criteria are very general in nature, i.e., they do not depend upon either the subject matter or the modeling methodology. Therefore, two things are emphasized: (1) these criteria primarily reflect concerns any decisionmaker or model evaluation team would wish addressed before relying upon the results obtained from a model; and (2) the team will have to use a great deal of ingenuity, judgment, and experience when adapting the criteria to a specific model.

These criteria apply to any model evaluation effort. The extent to which they are applied in a particular case will depend not only on the judgment and experience of the evaluators but also on the needs of the model's users. Two important caveats are:

--Attention is focused on large scale computer models. While we feel that the criteria described in this chapter should apply to most model evaluation efforts, we do not attempt to specify the entire class of models to which they most directly apply.

--Since our interest is focused upon large scale computer models used to help analyze major programs, the model evaluation process is viewed here as a very extensive undertaking. To employ the same level of effort for all model evaluations would be wasteful indeed. We surely are not advocating this. Many times a quick but careful reading of available documentation on a model by an expert in the area (i.e., a "face validity" check) would enable a potential user to decide whether or not to further explore use of the model. There is an entire spectrum of possible levels of evaluation between a "face validity" check and the type of detailed analysis that is described in this document. In each case, the evaluators must place the proposed use of the model in proper context and help the appropriate decisionmakers decide what risks are acceptable. This should enable the decisionmakers to decide upon an appropriate level of evaluation.

DOCUMENTATION

Documentation, as the term is used here, is defined as written (or otherwise recorded) information concerning a model. This definition is purposely very general; it is intended not only to recognize but to highlight the fact that there are different levels of documentation designed to serve different purposes. It is convenient to distinguish here between two levels of documentation: descriptive documentation and technical documentation. The former consists of general information about the model such as its underlying theory, assumptions, limitations, constraints, relationships to other models to which it is linked, etc.; the latter consists of information that is sufficiently detailed to allow technical evaluation of the model, including details of the methodology used, mechanization, and running the model to permit the duplication and operation of the model.

As was emphasized in the discussion of the modeling steps, good documentation is an integral part of model development and use. Ideally, it should begin with the first step of model development and be kept current as each step of the modeling process is undertaken. Both descriptive and technical documentation are essential to achieve a proper understanding of a model and of its strengths and limitations.

Through documentation, people interested in a modeling effort--users, model developers, evaluators, et al.--can communicate about a model and its results. Clear, concise, and complete documentation is the foundation upon which such discussions should be based. Documentation is also important (1) to ensure that the model is thoroughly understood and can be operated and maintained in the present and the future, and (2) to facilitate independent evaluation of the model (i.e., by someone other than the model developer or initial user). It should be noted, however, that while good documentation is necessary from the viewpoint of the evaluator, it may not be from the viewpoint of the user. For example, a sponsor of a model development effort may not deem it important enough for his needs to provide adequate funds for documentation. The extent of a model's documentation is the responsibility of the developers and sponsors. The evaluation team must function within the confines of available documentation.

To summarize, the developer needs to document what has been done, why, and how. Documentation also should include claims for the model and evidence to substantiate those claims. Moreover, the record of the modeling process should be clear and intelligible to an informed, interested audience. It should be sufficient enough to permit the replication of the model's results by independent evaluators. The clarity, completeness, and conciseness of model documentation is also critical to the process of model evaluation. Since the quality of documentation cannot be assessed until the team has reviewed it in detail, the evaluation process begins and ends with documentation.

An annotated documentation checklist may be found in the appendix. This checklist does not depend upon the methodology employed or the subject matter. It is not intended that issues such as the need for different levels of documentation (depending on the purpose of the model or the intended audience) be addressed in this checklist. These issues are very important, of course, and need to be addressed in establishing model documentation guidelines and/or standards. The documentation standards for energy models currently being developed by the Professional Audit Review Team (see page 2) address such issues. Documentation is the responsibility of developers. Model sponsors should also recognize the need for this information and provide the developers with sufficient resources to produce adequate documentation.

VALIDITY

There is no reason to believe that a model is capable of approximating reality so well that its results can be accepted without reservation. This is the case even for those aspects of reality pertinent to the purpose the model was designed to serve. Its capability to do this from the perspective of the decisionmaker or analyst who is using the model is referred to as model validity, and the process involved is called validation. The definition and the determination of the degree of validity obtained is the major task of evaluation. It is important to realize that perceived reality is in the mind of the viewer, in this case the decisionmaker or the analyst. This makes an already difficult task even more difficult, and the process of validating a model requires interaction between the model developers and/or evaluation team and the users.

Frequently, the outputs of a complex model are predictions. It is then the task of the evaluation team to produce some statement concerning the accuracy of these predictions. The statement should include, where possible, comparison of a model's outputs against historical results or the results of field experimentation. A properly developed and documented model would include such analyses performed by the model developers. It would then be the task of the evaluators to assess the developers' predictions and supporting tests. However, the evaluation team may have to perform additional tests or at least replicate some documented tests.

Validation occasionally can be assisted by comparison to similar models, and by the top-down approach that uses control targets (e.g., known budget totals and subtotals) for checking on internal consistency. However, it is rarely possible to validate a decision-aiding model in this manner, since there is no real data about alternatives which are not implemented. Determining whether the model can be used to process historical input data and produce accurate historical output is probably the most basic and prevalent aspect of model validation. This validation aspect must be attempted for models of ongoing systems, but, of course, cannot be accomplished for models of new or proposed systems. For proposed models, the evaluation team must rely on the apparent credibility or reasonableness of the model as judged by those who are knowledgeable about the system being modeled. This group should include the decisionmakers and sponsors, as well as the model developers, and there must be evidence that they have reviewed the model in detail and agreed upon its structure.

A first-time model, a model of a proposed or conceptual system, or one that is based on assumptions about the future, are most difficult to evaluate. Here, the following considerations take on increased importance:

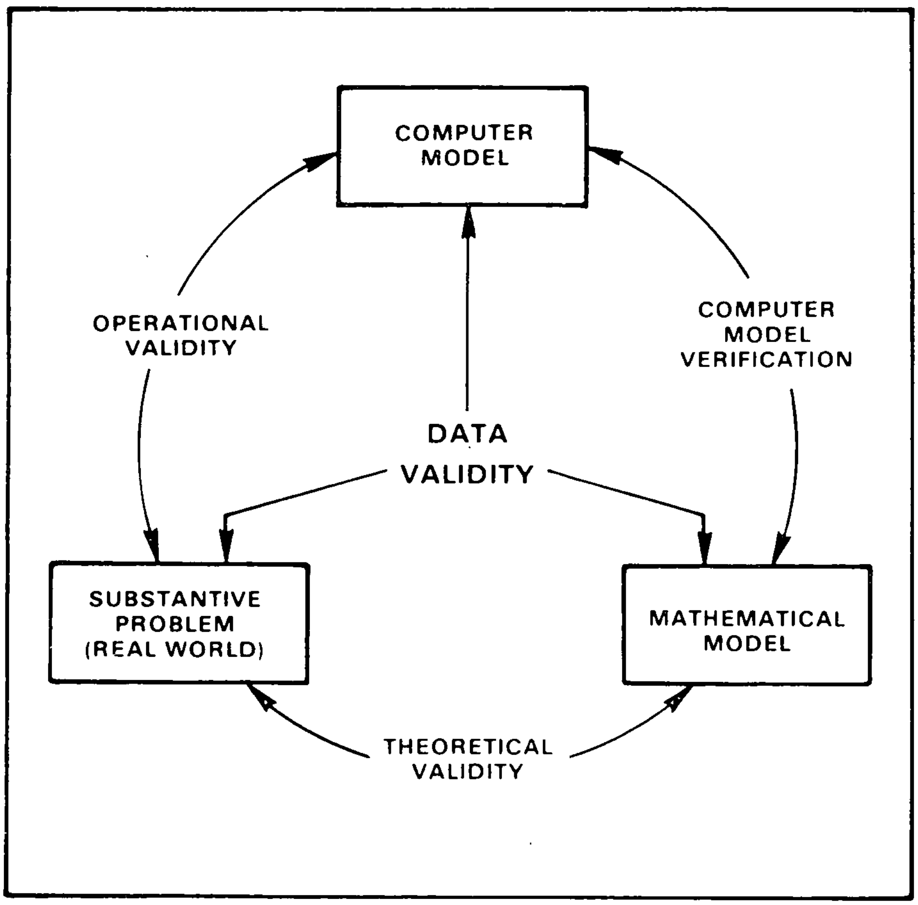
--Face validity which is a measure of general credibility (an initial expert opinion regarding the model's realism).

--Variable or parameter validity which is a measure of ability to interpret variations (a sensitivity analysis in which one or more input factors are changed to learn how they affect outputs).

In most complex modeling situations researchers have found that the decisionmakers often do not get involved in the details of either design or validation. The final model structure tends to be a product of the analysts and computer programmers. For a complex model of an existing situation, historical data frequently are very difficult to obtain (in terms of availability, completeness, cost, and analysis).

In sum, there is no validation procedure appropriate for all models; the tasks required for model validation must be adjusted on the basis of model structure, documentation, and other information that can be made available to the evaluators. Validity is viewed in this document as being comprised of three main subcategories--theoretical validity, operational validity and computer model verification. The relationships of these subcategories to each other and to computer model verification is depicted in figure 3 and will be discussed in detail in the following sections.

FIGURE 3
RELATIONSHIPS AMONG VALIDATION, VERIFICATION, AND MODELING PHASES



Theoretical Validity

Models are particularly useful tools and are frequently used by analysts and decisionmakers in the systematic investigation of questions or problems encountered by governmental planners or decisionmakers. The questions or problems are investigated using the model as a surrogate for the real world situation. The nature of the conclusions derived from the model, and the amount of credence and confidence that can be placed in it, depend in part on the results produced by the mathematical analysis of the model itself. It also depends significantly on the relationship between the problem and the model--what parts of the problem the model represents, and how well, and what parts of the problem the model distorts or fails to represent, and how badly.

Theoretical validation requires the evaluators to review the theories underlying the model and the major stated and implied assumptions which are embodied in that set of theories or which have been made to develop or adapt a theory to a problem. The applicability and restrictiveness of these assumptions in relation to the internal and external problem environments as viewed by decisionmakers must be examined, i.e., do they affect the model in such a way as to yield results for a problem that is different from the one originally stated? Have the underlying theories been adequately tested? Is it reasonable to assume that they are applicable to the problem at hand? The divergences uncovered by this review must be stated and discussed as to how they do or do not limit the validity of the model.

The evaluators must also verify that the transition from the theoretical model of reality (or perceived reality) to the mathematical model has been made correctly. This process will involve identifying and assessing the reasonableness of the most important assumptions made by the modelers in formulating the mathematical model. This is not as easy as it might seem, for assumptions come in many different forms. Explicitly stated assumptions are easy to identify. The difficulty lies in isolating the most important unstated or implied assumptions. Such implied assumptions may be present in the underlying theory or in the methodology chosen to apply the theory to the problem. Sometimes they are affected by the implicit and/or unintended biases of the model developers and computer programmers.

Many general methodologies have been devised for the construction of models. Some examples are regression analysis, linear programming, industrial input-output analysis, systems dynamics, and microsimulation. Each methodology is based upon special procedures and assumptions which may or may not be applicable to a specific situation. The evaluation team must assure itself that in the application of a particular theory and methodology sufficient care was taken to ensure that its assumptions were appropriate.

As an example, suppose an energy growth model is to be evaluated. Assume that the model relies on historical data and uses econometric techniques to estimate the parameters in the various structural equations. Further suppose that the model has one or more energy use and price variables among its dependent variables, and variables such as income, capital costs, imports, and government expenditures as the primary independent variables. Various criteria might be used to judge the underlying econometric methodology: Does the model capture the relevant energy policy issues? Are the assumptions realistic? For example, if the model includes an assumption about the responsiveness of energy use to price changes, has the assumption been empirically tested?

Analogous questions would have to be used for any methodology applied to a specific situation. For example, the first thing one should question in a linear programming model is whether the underlying assumption of linearity is appropriate.

A number of concerns have been raised in this section and their evaluation will, in general, be quite difficult. In addressing the concerns of theoretical validity, evaluators will have to address very broad, complex questions such as:

--What theories are considered to be relevant to the issue or problem to be modeled?

--What major assumptions have been used, either explicitly or implicitly, in fitting the theory to the problem?

--Have intangible issues such as political behavior, consensus maintenance and coalition-building, human values and attitudes, leadership and morale, and self-sacrifice been considered? Are these incorporated directly into the model? Indirectly? Not at all?

--How do these assumptions influence the modeling results, e.g., do they ignore certain interrelationships and, thus, not reflect the effect of significant data variations?

In addition, the evaluation team will need to examine the internal logic of the model. For a discussion of these aspects, see the section on computer model verification (p. 20).

Data Validity

Here the evaluation concern is two-fold:

--The accuracy, completeness, impartiality, and appropriateness of the original data.

--The manner in which the model deals with the transformation of the original data.

The distinction between and consideration of these two considerably different aspects of data validity is important. It is not sufficient merely to insure that the original data are accurate, complete, impartial, and appropriate. For example, a microsimulation model might require the specification of sources of non-wage income (such as interest, rent, and dividends) as data inputs. The only data source available for the model might provide information only on total non-wage income. While this original data may very well be accurate, complete, impartial, and appropriate for use by the model, the evaluator should determine whether the disaggregation of the original data is accomplished correctly.

To validate the data, the evaluator will have to answer such questions as:

--Do the data identify and measure the desired problem elements?

--Are the data sources clearly defined and the responsibilities for data collection established?

--Are the procedures for the collection and updating of data workable?

--Are the data obtainable within reasonable cost, time, and operational assumptions?

--Do the data collection procedures lead to impartiality with respect to the accurate recording of the data?

--Is the resulting data set representative?

--Are there audit procedures for the data collection activity; are they correct and do they aid in answering the above questions?

--Are aggregation or disaggregation procedures used in preparing data for use by the model?

--Have these procedures been documented so that their appropriateness may be determined?

--Are the data current?

Operational Validity

It is the inherent nature of models not to be able to reproduce exactly or to predict infallibly the real-world situation. Operational validity is concerned with assessing the importance upon the actual use of the model of these errors and divergences. This will require interaction between the evaluation team and the users. This is very important. It is the main check the team has to ensure that they comprehend the users' perception of reality.

The evaluators should be concerned with the divergence between the actual (real-world) and the outcomes predicted by the model. For some models, statistical tests can be utilized, e.g., comparing historical time series. The evaluators should determine if such tests can be applied and have been applied properly by the developers. As intermediate computational results (parameters) are usually used in the model to further the analysis, the team also needs to be concerned with whether there are errors in computed parameters, and if procedures to minimize any such errors were utilized. The evaluators must attempt to uncover any divergences and errors and their magnitudes and to assess their impact. The outcome of this operational validity review should include a listing of computed parameters, decision variables, and the extent to which errors and divergences in these computations can occur. To accomplish their task, the evaluators will have to answer such questions as:

--To what extent do the assumptions of the model and their divergences degrade the use of the results in the operational situation?

--Do the data requirements in terms of cost, time, accuracy and operations preclude gathering the necessary model inputs?

--Do the logic and numerical elements of the model as transformed into the computer program result in an invalid computational process?

--Are the predictive divergences of the model of a great enough magnitude to cause the model results to be unacceptable?

--Are the results of any trial solutions inconsistent with the expectations of the decisionmaker? If yes, how can the use of the model be justified? What changes have been or should be made?

--Are the expected cost savings of efficiency/effectiveness improvements attributed to the use of the model of a proper magnitude to justify the use of the model in its planned operational setting? Are the costs and benefits calculated correctly?

--Has the response of the model to changes in parameter values been determined? If yes, have complete tests been applied and results been presented to the user? If no, on what basis do the developers justify the use of a particular solution, with respect to parameter and input data values? Are sets of solutions presented to the user showing model outputs for the different possible ranges of data?

--What controls, if any, have been established to ensure that the final operational environment for the model is the same that was assumed in the original and modified model development plans? Are there model implementation plans? If so, are they realistic in terms of time, budget, and other resources and can they be accomplished? If no, how does the decisionmaker justify the use of the model?

--Are there confidence intervals on inputs and on the decision variables? If exogenous inputs are made to the model from external models such as the commercial econometric models, have these inputs been evaluated? How are errors propagated through the model?

COMPUTER MODEL VERIFICATION

To verify a model, the evaluators must ensure that it has the attributes which the developer imputed to it and that it behaves as intended. Basically, verification has been accomplished if it has been demonstrated that the computerized model "runs as intended." That is, the evaluators must also be concerned with validity in the translation of the mathematical model statement and formulation into a numerical computer process. This is a complex, multifaceted problem and requires that the evaluators explicitly identify and state in measurable terms, the model's intended purposes and determine whether there is sufficient evidence to establish that:

--the mathematical and logical relationships are internally consistent;

--the mathematical and numerical results are correct and accurate;

--the logical flow of data and intermediate results are correct;

--the important variables and relationships have been included;

--the computer program, as written, accurately describes the model as designed;

--the program is properly mechanized and debugged on the computer; and

--the program runs as expected.

Although these aspects are interrelated and not independent, it is important that they be verified separately because of the possibility that any one of them might not hold in a given case. Thus, just because the program has been written accurately and mechanized properly on the computer is no guarantee that it will run correctly. Assume that a model describes a problem of interest and the relevant programs have been computerized and debugged properly. Next, suppose that a decisionmaker wishes to determine the inputs required to obtain a desired output. The process of obtaining these inputs requires the application of a number of computational procedures. In spite of the fact that everything has been done correctly to this point, the program may still not run as expected, e.g., accumulated errors may become a large problem or an unexpected division by a very small number (almost zero), may lead to a meaningless result.

It is worth observing that once the computer model has been verified, the abstract or mathematical model recedes into the background. It is actually the computer model which will be evaluated according to the criteria outlined in this chapter.

Clearly, the evaluators will usually be constrained to perform the evaluation based on available documentation. This puts a heavy burden on the developers to prepare complete documentation that describes their verification process, e.g., test problems and results, or debugging procedures. However, whenever possible, the evaluators must attempt to clarify any detected deficiencies by discussing them with the developers. The evaluators should be able to replicate the tests conducted by the developer.

For any complex model, it will be difficult to state that the model has been completely verified. Where there are still concerns, the evaluator should state them along with an interpretation as to how these concerns (i.e., verification deficiencies) should be interpreted by any user. Some deficiencies will be minor and require minimum or no caution by the user, while others might be so major that a disclaimer on the model's verification should be promulgated. To establish confidence in the model's level of verification, the evaluator should comment on the use or lack of use of good design and coding techniques and aids, such as structured programming and systematic program change and updating procedures. When the above process has been completed, the outcome of verification is a summary statement describing any deficiencies, their impact on the ability to run the model, and whether the results of the model can be used explicitly or how they must be qualified.

Experience has shown that in the absence of computer model verification--at least main program flow, critical parameters, and program modules--the odds are that no one will really know what is going on. If the evaluators do not have sufficient evidence that the model has been properly verified, then they may decide to so report and to suspend their evaluation effort until the developer has satisfied this deficiency.

If some documentation is available and a more complete computer model verification is deemed necessary, the evaluators are referred to two other GAO publications. 1/

1/Audit Guide for Assessing Reliability of Computer Output, FGMSD-No.17-S/P, U.S. GAO, Washington, D.C., May 1978. Guide for Evaluating Automated Systems, Exposure Draft, U.S. GAO, Washington, D.C., March 1977.

These documents are intended to help an auditor assess the accuracy and reliability of a computer program and of its output (consistent with the auditor's intended use of the computer output). Included is a step-by-step approach and detailed audit procedures designed to lead to more uniform evaluations of internal controls.

MAINTAINABILITY

The next major evaluation criterion is model maintainability. This is concerned with how an acceptable model can be maintained during its life cycle so that it will continue to be an acceptable representation of the real system. Two aspects of maintainability are review and updating.

Review

Review is a preplanned and regularly scheduled program for reviewing the accuracy of the model over its life cycle. The evaluators must be assured that a review procedure has been established, and that it is functioning properly.

Some specific questions the evaluation team should ask include: Is there a formal procedure that requires the users, model developers, and/or current model maintainers and solution implementers to meet, discuss and decide what to do about divergences between the model predictions and the actual outcomes or proposed model and data changes; and to determine on a continuing basis whether the model is still valid, is not to be used any further unless specified changes are made, or is not to be used further under any conditions? Are change implementation procedures fail-safe (i.e., the current working system cannot be lost), do they encompass a proper testing methodology, and, very importantly, do they produce the necessary documentation?

Updating

The evaluators need to be satisfied that a procedure has been established to collect and to analyze information to determine if and when the model parameters or model structure should be changed, and that a process exists by which such changes are to be made.

Some specific questions relevant to updating include: Are there procedures for detecting when input data have changed? If yes, are the controls workable and are the changed data collected in a timely fashion so as to ensure that the model's calculations are not degraded or incorrect? Has someone been designated to be responsible for updating data sets and for analyzing the accuracy and the propriety of introducing updated data into the system? What procedures have been established to ensure that new data are entered without error? Has the computer program been written in a form that is readily modified?

Another related aspect that the evaluators must assess is adequacy of the training program associated with the model. The formalization of a training program is dependent on the model's application. However, training normally includes such items as formal lectures for computer systems personnel and other users, and briefings to decisionmakers on the model and how to interpret the model's output. It is important that: (1) revisions to the model are made known to systems personnel and decisionmakers; (2) the model results are presented to the user in a familiar and acceptable format; and (3) the user understands how the model should be used. As changes in the model are made, procedures are required to reflect such changes in the model documentation. The evaluators should determine if a proper process of updating and dissemination has been established.

USABILITY

In the final analysis, the usability of the model is a major concern of the potential model user. Thus, the evaluation team's report should contain a statement addressing this issue. Some factors which affect a model's usability include:

--Availability of data. Even if the data are known to exist, they might not be available for general use. For example, Bureau of the Census data which have been collected but have not yet been released might be necessary for a particular model; until such data are available, the model is not usable. How are privacy and freedom of information issues handled?

--The understandability of the model's output. Often the computer-produced output from a model is not in a form which is understandable, i.e., it may be a string of numbers with no explanatory text. If the results of the model are incomprehensible to the user, then, for all practical purposes, the model is not usable.

--The presentation format chosen. How representative are baselines, e.g., base year? Are the sensitivity data selected to show only one type of "finding?" What is the distribution of model results?

--The transferability of the model to another computer system.

--The accessability of the model, e.g., is it classified?

--The size of the model.

--The time of a typical run.

--The costs to set up and run the model in terms of both money and personnel, e.g., what is the efficiency of the computer model design in terms of the number of different runs needed to gain reliable insights?

The above list is neither exclusive nor exhaustive. It merely suggests some factors which can affect the usability of the model. Their relative importance will depend upon the problem at hand, and it will be up to the evaluators to determine this.

EVALUATION REPORT

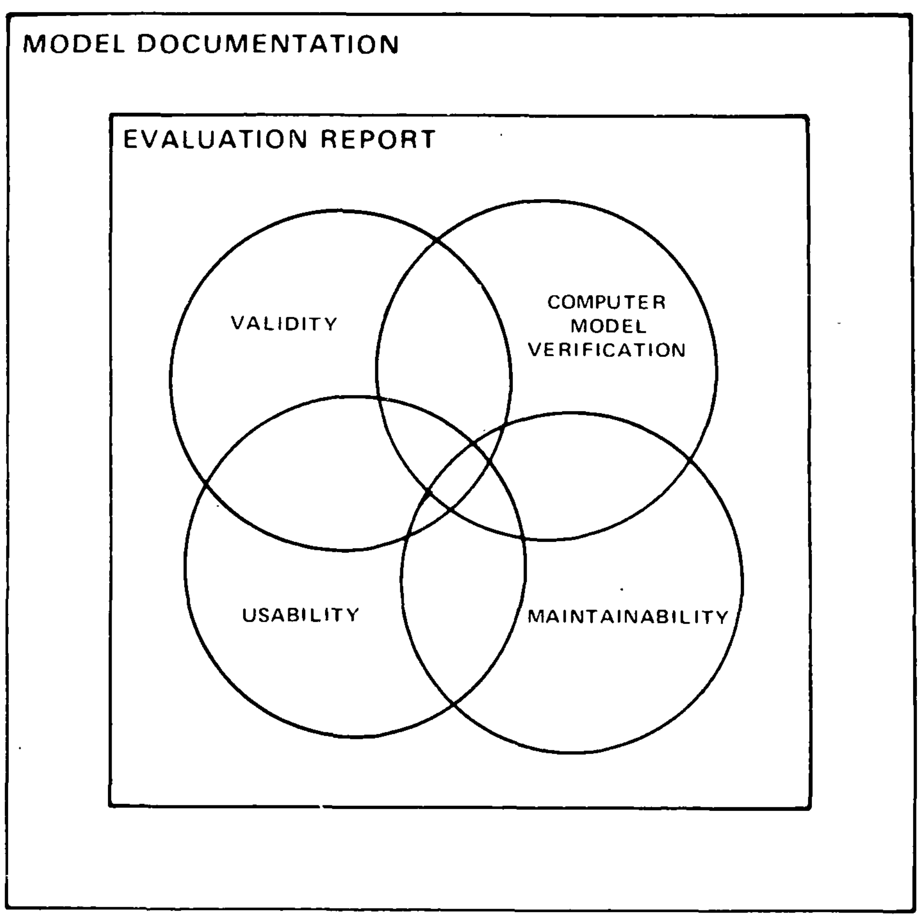
The evaluation report documents the evaluation team's view of when and how the model should be used. In other words, it delineates the team's view of the proper domain of the model's applicability. This judgment will be the result of a careful synthesis of the evidence which the team has accumulated during its evaluation effort. This statement should be comprehensive enough to enable the potential user to determine the different ways the model can be used, or whether or not the model should be used at all. In the context of this document the team's report will be one further element of model documentation. As such, it should provide a good vehicle for further communication concerning the model and its present or future uses among the developers, evaluators, potential users, and anyone else interested in the model.

CHAPTER 4

SUMMARY

Chapter 3 provides a list of criteria for model evaluation. As was the case in the list of modeling steps in Chapter 2, they are interrelated. They impact on one another and it makes little sense to consider them in isolation. These interrelationships are illustrated in figure 4. Also, some criteria will assume greater importance for the evaluation of a particular model. The exact mix or blend will depend upon such factors as the importance and complexity of the system the model was developed to approximate, the evaluation team's experience and perception of that system, and the arena in which the system

FIGURE 4
INTERRELATIONSHIPS AMONG EVALUATION CRITERIA



was modeled. For example, the model may have been evaluated by another interested party. This earlier evaluation might enable an evaluation effort to focus attention on previously identified weaknesses in the model. Or, a good evaluation of another model developed for the same purpose may have been completed; this might permit a comparison of the two models.

Obviously, a model evaluation which addressed every facet of each of these evaluation criteria would be a massive undertaking requiring a large commitment of staff, time, and money. Such an evaluation would probably be considered only for very large complex models that had, or could have, an impact on major programs.

It will be apparent to the reader from the discussion in this document that the evaluation of a model is not a routine, standardized process. Indeed, model evaluation is in its infancy and, at the present time, is more an art than an established methodology. It may seem reasonable to expect that the process will become progressively more systematic as experience accumulates.

APPENDIX

Model Documentation Checklist[1]

1. *Project information*
   - 1.1 *Project title*
     This should be the title of the overall project of which the modeling or simulation may be just a part.
   - 1.2 *Responsible organization*
     This is the name and address of the organization responsible for the overall project. If the project is supported by an external source (e.g., by a grant), this would be the organization responsible for the money and equipment furnished.
   - 1.3 *Contact*
     This is the person or persons to contact for further information. If the address is other than that given in 1.2, the full mailing address should be given. The organizational mail-stop or code and telephone number and extension should also be given.
   - 1.4 *Project objective*
     This is the objective of the overall project, which may be of greater scope than that of the modeling or simulation to be covered later.
   - 1.5 *Project duration*
     Give date that project was established and expected completion date.
   - 1.6 *Funding*
     - 1.6.1 *Source*
       Give name of funding organization.
     - 1.6.2 *Amount*
       Give total dollars. If equipment is contributed, list major items or estimate value.
     - 1.6.3 *Period*
       Give dates covered by support listed in 1.6.2

2. *Model development information*
   - 2.1 *Name of model*
     This might be the computer-callable name of the program. If an acronym, spell it out (e.g., WLDREC: WorLD RECycling model).
   - 2.2 *Name of modeler(s)*
   - 2.3 *Purpose for which model was developed*
     "The same" here will indicate the same as 1.4.
     - 2.3.1 *Specific*
       Give reason(s) for developing model (e.g., to test hypothesis that . . .).
     - 2.3.2 *General*
       Give other uses to which the model has been or might be put (e.g., to study other problems related to . . .).
   - 2.4 *Disciplines involved*
     These need not be fields of endeavor recognized as distinct disciplines (e.g., economics), but may be more descriptive of the work (e.g., land use).
     - 2.4.1 *Primary*
       Give the discipline(s) that the model was primarily developed to serve (e.g., international relations).
     - 2.4.2 *Supporting*
       List other disciplines required in the development of the model, preferably in descending order of importance.
   - 2.5 *Data required*
     Give kind of data (e.g., population) and source (e.g., census).
   - 2.6 *Method of development*
     Give method of development (e.g., theoretical, empirical, other).
   - 2.7 *Assumptions*
     List all assumptions concerning both data and causality that led to the model's being developed in the way it was.
   - 2.8 *Cost of development*
     Give actual or estimated total cost of the model and what the cost includes.

[1] McLeod, John, SIMULATION: FROM ART TO SCIENCE FOR SOCIETY, Simulation Today, No. 20, Dec 1973. A related checklist may be found in House and McLeod, Large-Scale Models for Policy Evaluation, pp.84-87.

2.9 *Availability*

2.9.1 *To developer*
Is the model operative? What will it take to make it operative?

2.9.2 *To others*
Is the model proprietary or classified? Can it be obtained by others? How? In what form (e.g., computer listing, deck, paper or magnetic tape, other)? What are the charges?

2.10 *Compatibility*

2.10.1 *Development of computer system*
On what equipment was the model developed?

2.10.2 *Other systems*
On what other computer systems has it been or might it be run with negligible change?

2.10.3 *Language(s)*
In what language(s) was the program written? Is it available in others?

2.11 *Extent of use*

2.11.1 *By developer*
What actual use has been made of the model by the developer? What use is planned?

2.11.2 *By others*
Has the model been used by others? By whom? To what extent?

3. *Model description*

3.1 *Model classification*
What kind of model is it? How is it run—batch or interactively? Locally or remotely? Is the computer time-shared?

3.1.1 *Focus*
Give the primary fields of interest that the model serves (e.g., political science, resource usage, etc.).

3.1.2 *Scope*
Give entity modeled (e.g., an industrial plant, a river basin, the U.S. Senate, etc.).

3.1.3 *Sophistication*
Where does the model fit in the "Fuzz to Fact" spectrum (e.g., preliminary studies, evaluating alternatives, predicting the future)?

3.2 *Block diagram of system modeled*
This should have a block for each component of the real-world system modeled, and show lines between the blocks indicating the causal relationships of the components as well as exogenous inputs and outputs.

3.3 *Program or wiring diagrams*
A program flow diagram should be shown in the case of digital models, a wiring diagram in the case of analog, and both in the case of a hybrid model.

3.4 *Notation*
A complete description of the notation used in 3.2 and 3.3, as well as any narrative description, should be included here. The notations and definitions must be carefully checked for consistency throughout the documentation.

3.5 *Validation*
Describe how the model was validated.

3.6 *Reference information*
This should be a computer listing of the program and the output of a standard check run for a digital model, a plot of a standard check run for an analog, or both in the case of a hybrid model. It should give all "numbers" used to set up the run, and be annotated in such a way that either the developer at a future date, or another user of the model, can make sure that if the model is to be rerun or used by someone else, he will be working with the model that he thinks he is working with.

3.7 *Distinctive features*
How does the model differ from related models? How is it better? What are its limitations? What are the possible pitfalls that might be encountered in its use?

3.8 *Model antecedents*
Have similar models been built before? If so, give proper credit. Is the current model a follow-on or a distinct "mutant"?

3.9 *Current relations*
Do other models exist that have the same or a closely related purpose? How does this model relate to them? Are they another attempt to solve the same problem, or can the results be expected to be complementary, i.e., to present two aspects of a larger problem? Are there possibilities of online interconnection and interaction between the models?

4. *Simulation(s)*

A simulation will be taken to mean an *experiment* performed on a model instead of the real-world simuland. Multiple runs using the same experimental design may be considered one simulation. However, if the design of the experiment or the procedure is changed, it should be considered another experiment and items 4.1 through 4.11 should be covered again.

4.1 *Title*
This should be descriptive of the simulation experiment, and may be made up of the model name plus a subtitle (e.g., "WLDREC: Effect of cost of recycling").

4.2 *Purpose*
This is the purpose of the individual experiment.

4.3 *Assumptions*
List assumptions made in the design of the experiment.

4.4 *Experimental design*
This should give the procedure to be followed in the experiment, step by step.

4.5 *Data requirements*
Give the data requirements for the individual simulation run(s) that differ from those for the model's reference run (item 3.6).

4.6 *Data used*
This might best be a computer listing of those lines of data that differ from the reference run.

4.7 *Run time*
This should be total as well as mainframe time.

4.8 *Cost per run*
This should be given for both mainframe and peripherals.

4.9 *Results*
This can be "raw data" (e.g., a computer printout) and plots, graphs, etc., prepared by hand or machine.

4.10 *Justification of assumptions*
This is most important, and should be done before any analysis of the results is attempted. Assumptions that influenced the development of the model as well as those related to the specific simulation experiment should be considered.

4.11 *Analysis*
Describe conclusions drawn from the results, and give reasoning where the conclusions are not obvious.

5. *Discussion*

5.1 *Comments*
Add any comments here that might further illuminate aspects of the project not covered elsewhere or, if preferred, give a brief narrative description for the benefit of the casual reader.

5.2 *Conclusions*
Relate development of the model and the simulation experiments to the overall project objective given in 1.4.

6. *Literature*

6.1 *Project reports*
List reports, presentations, and articles generated by the project.

6.2 *References*
List publications actually referred to in the documentation.

6.3 *Bibliography*
List publications which influenced the work documented or which are closely related to it.

BIBLIOGRAPHY

Apostel, L.: "Formal Study of Models," Chap. 1 in The Concept and the Role of the Model in Mathematics and Natural and Social Sciences, edited by H. Fredenthal, Gordon and Breach Publishers, N.Y., January, 1960.

Barker, W.G.: "The Use of Models in Urban Transportation," Department of Transportation, NTIS No. PB-222 893, April 1973.

Berman, M.B.: "Notes on Validating/Verifying Computer Simulation Models," Rand Report P-4891, The Rand Corp., Santa Monica, Calif., August 1972.

Brewer, G.: Politicians, Bureaucrats and the Consultant--A Critique of Urban Problem Solving, Basic Books, 1973.

Chaiken, J. et al: "Criminal Justice Models An Overview," R-1859-DOJ, The Rand Corporation, Santa Monica, Calif., October 1975.

Clark, J. and Cole, S.: Global Simulation Models: A Comparative Study, John Wiley and Sons, 1975.

deNeufville, R. and Stafford, J.H.: Systems Analysis for Engineers and Managers, McGraw-Hill Book Company, N.Y., 1971.

Deutsch, et al: Problems of World Modeling: Political and Social Implications, Ballinger Publishing, 1977.

Eilon, S.: "Mathematical Modeling for Management," Interfaces, Vol. 4, No. 2, February 1974.

Emshoff, J.R. and Sisson, R.L.: Design and Use of Computer Simulation Models, The Macmillan Co., N.Y., 1970.

Environmental Modeling and Decision Making, Holcomb Research Institute, Praeger Publishers, N.Y., 1976.

Fishman, G.S. and Kiviat, P.J.: "Digital Computer Simulation: Statistical Consideration," Rand Report RM-5387-PR, The Rand Corp., Santa Monica, Calif., 1967

Fromm, G., Hamilton, W.L. and Hamilton, D.E.: "Federally Supported Mathematical Models: Survey and Analysis," U.S. GPO. Stock No. 038-000-0021-0, Washington, D.C., 1975.

Gass, S.I.: "Evaluation of Complex Models," Computers and Operations Research, Vol. 4, pp. 27-35, March 1977.

Gass, S.I.: "A Procedure for the Evaluation of Complex Models," Proceedings First International Conference on Mathematical Modeling, University of Missouri, 1977.

Gass, S.I., and Sisson, R.S.: A Guide to Models in Governmental Planning and Operations, Sauger Books, Potomac, Md., 1975.

Greenberger, M., Crenson, M.A., and Crissey, B.L.: Models in the Policy Process, Russell Sage Foundation, N.Y., 1976.

Honig, J. et al: "Review of Selected Army Models," Department of the Army, Washington, D.C., May 1971.

Kleijnen, Jack P.C.: Statistical Techniques in Simulation, Parts I and II, Marcel Dekker, Inc., N.Y., 1974.

National Bureau of Standards: "Guidelines for Documentation of Computer Programs and Automated Data Systems," Federal Information Processing Standards Publication (FIPS) 38, Washington, D.C., February 15, 1976.

Operations Research Society of America: "Guidelines for the Practice of Operations Research," Operations Research, Vol. 19, No. 5, September 1971.

Pack, J.R.: "The Use of Models in Urban Policy Making," Fells Center of Government, University of Pennsylvania, 1974.

Pugh, R.E.: Evaluation of Policy Simulation Models, Information Resources Press, Washington, D.C., 1977.

Quade, E.S.: Analysis for Public Decisions, American Elsevier, 1975.

Sage, A.P.: Methodology for Large-Scale Systems, McGraw-Hill, 1977.

Schellenberger, R.E.: "Criteria for Assessing Model Validity for Managerial Purposes," Decision Sciences, Vol. 5, No. 5, 1974.

Shubik, M. and Brewer, G.D.: "Models, Simulations, and Games --A Survey," Rand Report, R-1060-APRA/RC, The Rand Corporation, Santa Monica, Calif., May 1972.

Strauch, Ralph E.: "A Critical Assessment of Quantitative Methodology as a Policy Analysis Tool," Rand Report, P-5282, The Rand Corporation, Santa Monica, Calif., August 1974.

U.S. General Accounting Office: An Evaluation of the Use of the Transfer Income Model--TRIM--To Analyze Welfare Programs, PAD-78-14, U.S. GAO, Washington, D.C., November 25, 1977.

U.S. General Accounting Office: Audit Guide for Assessing Reliability of Computer Output, FGMSD-No.17-S/P, U.S. GAO, Washington, D.C., May 1978.

U.S. General Accounting Office: Auditing a Computer Model: A Case Study, U.S. GAO, Washington, D.C., May 1973.

U.S. General Accounting Office: Advantages and Limitations of Computer Simulation in Decision Making, B-163074, U.S. GAO, Washington, D.C., May 3, 1973.

U.S. General Accounting Office: Computer Simulation, War Gaming, and Contract Studies, B-163074, U.S. GAO, Washington, D.C., February 23, 1971.

U.S. General Accounting Office: Improvement Needed in Documenting Computer Systems, B-115369, U.S. GAO, Washington, D.C., October 8, 1974.

U.S. General Accounting Office: Improvements Needed in Managing Automated Decision Making by Computer Throughout the Federal Government, FGMSD-76-5, U.S. GAO, Washington, D.C., April 21, 1976.

U.S. General Accounting Office: Review of the 1974 Project Independence Evaluation System, OPA-76-20, U.S. GAO, Washington, D.C., April 21, 1976.

U.S. General Accounting Office: Ways to Improve Management of Federally Funded Computerized Models, LCD-75-111, U.S. GAO, Washington, D.C., August 23, 1976.

Urban, G.L.: "Building Models for Decision Makers," Interfaces, Vol. 4, No. 3, May 1974.

Van Horn, R.: "Validation," Chapter in The Design of Computer Experiments, T.H. Naylor, Ed., Duke University Press, Durham, N.C., 1969.

Van Horn, R.: "Validation of Simulation Results," Management Science, Vol. 17, No. 5, January 1971.

Weiss, C.H.: Evaluation Research, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1972.

Zeigler, Bernard P.: Theory of Modeling and Simulation, John Wiley & Sons, 1976.

Single copies of GAO reports are available free of charge. Requests (except by Members of Congress) for additional quantities should be accompanied by payment of $1.00 per copy.

Requests for single copies (without charge) should be sent to:

U.S. General Accounting Office
Distribution Section, Room 1518
441 G Street, NW.
Washington, DC 20548

Requests for multiple copies should be sent with checks or money orders to:

U.S. General Accounting Office
Distribution Section
P.O. Box 1020
Washington, DC 20013

Checks or money orders should be made payable to the U.S. General Accounting Office. NOTE: Stamps or Superintendent of Documents coupons will not be accepted.

PLEASE DO NOT SEND CASH

To expedite filling your order, use the report number and date in the lower right corner of the front cover.

GAO reports are now available on microfiche. If such copies will meet your needs, be sure to specify that you want microfiche copies.

AN EQUAL OPPORTUNITY EMPLOYER
U. S. GENERAL ACCOUNTING OFFICE
WASHINGTON, D. C. 20548

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

U.S.MAIL

POSTAGE AND FEES PAID
GENERAL ACCOUNTING OFFICE

THIRD CLASS

Dr. Richard Barnes
Nat. Inst. Law Enforce Crim. Jus.
633 Indiana Avenue, NW
Washington, DC 20530