NBSIR 84-2963

A Survey of Techniques for Evaluating Emergency Planning Models and Data Bases

Robert E. Chapman, Robert G. Hendrickson and Stephen F. Weber

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Gaithersburg, MD 20899

November 1984

Sponsored by
Federal Emergency Management Agency
500 C Street, SW
Washington, DC 20472

NBSIR 84-2963

A SURVEY OF TECHNIQUES FOR EVALUATING EMERGENCY PLANNING MODELS AND DATA BASES

Robert E. Chapman, Robert G. Hendrickson and Stephen F. Weber

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Gaithersburg, MD 20899

November 1984

Sponsored by
Federal Emergency Management Agency
500 C Street, SW
Washington, DC 20472

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

PREFACE

The Operations Research Division (ORD) of the Center for Applied Mathematics, National Engineering Laboratory, National Bureau of Standards (NBS) is conducting a research effort under the sponsorship of the Federal Emergency Management Agency (FEMA) to develop a set of guidelines and methodology for applying them which promotes both a critical review and unbiased evaluation of FEMA models and data bases.

Large-scale models and data bases are key informational resources for FEMA. The type and nature of FEMA models and data bases, however, are perhaps unique within the Federal government in that they are concerned often times with severe changes or disruptions ranging from limited effects to extremes. Furthermore, in order to carry out its emergency missions, it is necessary for FEMA to determine which models, modelling techniques and data bases are appropriate for what purposes and which ones need modification, updating and maintenance. The development of evaluation guidelines is therefore of direct benefit to FEMA in discharging its emergency planning duties.

This report is the result of an in-depth review of the recent technical literature on model and data base evaluation. The focus of the report is on four areas which are of crucial importance to the development of a set of evaluation guidelines for FEMA; they are: (1) recommended practices for model and data base evaluation, (2) documentation - verification; (3) sensitivity analysis; and (4) specialized techniques for dealing with missing data or extreme events.

Special appreciation is extended to Dr. Saul Gass of the NBS ORD and of the University of Maryland, whose stimulating discussions have played an important role in all model and data base evaluation work being carried out at NBS. Dr. Gass also made available to the project staff his extensive collection of technical reports and articles on the subject of model and data base evaluation. These documents proved invaluable in the formative stages of the literature review. Special appreciation is also extended to Ms. Diane Cunningham, Library Division, NBS, for her thorough and efficient use of the numerous computerized information services. Without her assistance this survey of the literature would have been much more difficult and time consuming.

ABSTRACT

This report presents a state-of-the-art review of the technical literature on model and data base evaluation which relates to the modelling problems faced by the Federal government in discharging its emergency planning activities. Particular emphasis is placed on model/data base documentation and verification, an analysis of the structural properties of a model/data base, and the role of sensitivity analysis. Emphasis is also placed on four special topics (rare events, missing data, error propagation, and stability) which are involved in many emergency planning activities.

Key words: Assessment; data base; data quality; documentation; emergency planning; evaluation; mathematical model; model credibility; policy models; sensitivity analysis.

Table of Contents

PART I: INTRODUCTION AND OVERVIEW

1. BACKGROUND

During the last 15 years, many members of the operations research profession have focused on the development of large-scale, policy-oriented models. These efforts should be contrasted with the "golden age" of operations research between 1950 and 1970 when most models which were built were operational in nature (e.g., inventory management, queuing problems at a bank or manufacturing facility). The great successes of the operations research modeling methodologies of the 1950's and 1960's were seldom experienced in the policy analysis arena, however. The major source of the current lack of credibility associated with large-scale, policy-oriented models can be traced to the nature of the problem. In the case of an operational problem, the model's analytical results can be subjected to a face validity test, whereby experts who thoroughly understand the process being modeled (e.g., how to refine an incoming tanker load of crude oil so as to maximize profits) can judge whether or not the model's results are reasonable. The key point is that the recommended solutions can be tested in a working environment.

For planning and policy-making situations, there is seldom the well-defined, clear-cut set of objectives and constraints that typify an operational problem. The goals of a particular policy may be loosely stated and communication through the chain of command may be less than ideal. Some applications, especially the type for which models are used at the Federal Emergency Management Agency (FEMA), may be of a unique nature for which little or no useful information is available on how the system will respond given the occurrence of a particular event. The types of problems which confront the operations research analyst who provides policy-oriented information to the decision maker is summarized succinctly by Quade.

> A well-defined problem is one that can be given a clear-cut, well-defined formulation, amenable to rigorous analysis. A squishy problem is one with the property that any clear-cut, well-defined formulation of it will look like an unambiguous representation of the substantive problem only so long as we don't lean too hard on it, or question it too carefully or deeply...well-defined problems in policy analysis tend to be concerned exclusively with technological or physical questions.[1]

[1]Quade, E.S. Analysis for Public Decisions. New York: Elsevier, 1982.

A number of researchers, most notably Gass, have advocated that model evaluation should be an integral part of the building and use of any large-scale, policy-oriented model. Gass provides three reasons for advocating the evaluation of a policy model:[1]

(1) in many cases, the ultimate decision maker is far removed from the modeling process and needs a basis for deciding when to accept the model's results;

(2) users of a model developed for others must be able to obtain a clear statement of the applicability of the model to their problem area; and

(3) it is difficult to assess the impact of a model's assumptions, data availability, and other elements on the model's structure and results without a formal, independent evaluation.

In the context of this report, at least three topics are of critical importance to any comprehensive model evaluation effort. They are: (1) documentation-verification; (2) technical validity; and (3) sensitivity analysis. A brief discussion of each topic is presented here. These and additional topics, as well as their relationship to the various sections of the annotated bibliography, are discussed in section 3.

From a model user's point of view, documentation (the written description of the model) is essential if the model is to be usable, useful, and used. Since the abstract model is a mathematical representation, whereas the operational model appears as a computer code, it is necessary to ensure that the computer program:

(1) describes accurately the model as designed;

(2) is properly mechanized on the computer; and

(3) runs as intended.

The process of checking the relationship between the abstract and operational models is referred to as verification. Documentation is also essential if any degree of software portability is required (not just that the program can be run on a variety of machines, but that it produces the desired result). Documentation also serves to explain all relevant relationships between inputs, outputs and analysis which should facilitate its use. Complex models often involve subtle techniques which, in the absence of a buffer between the user and the model, could cause frustration and lead to a highly inefficient use of the model. Documentation and an executive code (e.g., extensive edit-checking and message generating capabilities) should serve to shield the user from unnecessary detail without withholding any information which is essential to confidently use the model.

[1] Gass, S.I. "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," Operations Research, Vol. 31, No. 4, 1983, pp. 603-631.

Technical validity requires the identification of all model assumptions, including those dealing with data requirements and sources. As a first step, one should identify all stated and implied assumptions, all decision variables, and any hypothesized relationships between variables. This step sheds light on the correspondence between the model and the real world phenomena it attempts to explain. Three types of assumptions may be readily defined. First, the mathematical assumptions include its functional form and the continuity of its relationships (e.g., fixed-proportion production function). Content assumptions, the second type, define all model terms and variables. They should also define the scope and limitations of the model. Causal assumptions, the third type, are concerned with the assumed or hypothesized relationships between terms and variables. Ideally, one would like to build a model which would produce true conclusions whenever all of the assumptions are true. To translate such abstract concepts into a form which is concrete and testable, it is necessary to:

(1) determine if the model's calculations are correct and accurate;

(2) analyze whether the logical flow of data and intermediate results are correct and consistent; and

(3) ensure that no significant variables and relationships have been omitted.

Sensitivity analysis is a way to compensate for uncertainty and to gain insight into how the model deals with specific types of input scenarios. Performing a sensitivity analysis requires the analyst to systematically vary the values of the model's parameters over some range of interest to determine if and how the recommended solution changes. This is done by running the model using values of the parameters selected according to an experimental design and comparing the (possibly new) solution set of decision variables and their values with the recommended solution. Once a sensitivity analysis has been performed and its results analyzed, the decision maker can then act with greater assurance that the implementation of any solution close to the recommended one will yield acceptable results.

2. APPROACH AND ORGANIZATION

A literature search was carried out at two levels. First, technical documents which were known to be relevant, based on past experience, were surveyed in order to develop a knowledge base. During the past five years, the Operations Research Division of the National Bureau of Standards has been very active in the area of model evaluation. Consequently, some of the most authoritative and comprehensive technical reports were either on hand or could be obtained easily from the authors. Conversations with the authors of a number of reports were also helpful in identifying their current thinking on the subject of model and data base evaluation. This information was then used to define the scope of the search and to focus on certain areas, such as sensitivity analysis, which appeared to be of particular interest to FEMA. Second, the information developed and targets of opportunity identified in the previous step were used to direct the on-line literature review.
The National Bureau of Standards Library has available a series of useful computerized literature searching services. Two of these services were utilized in the on-line review. They were: (1) ISI (Institute for Scientific Information); and (2) DIALOG. One file, ISTP&B, was searched via ISI; the remainder were searched via DIALOG.

ISTP&B (Index to Scientific and Technical Proceedings and Books) is a multidisciplinary index to the most relevant proceedings literature published internationally in journal issues, journal supplements, serials, and multi-authored books at the chapter level. ISTP&B is the on-line equivalent to Index to Scientific and Technical Proceedings with additional coverage of scientific multi-authored books. The file includes information on engineering and technology as well as the physical and applied sciences from 1978 to the present. During the course of the on-line review, nearly 300 titles were obtained.

NTIS is an extensive data base consisting of government-sponsored research, development, and engineering reports plus analyses, journal articles, and translations prepared by Federal agencies, their contractors or grantees. This data is produced and maintained by the National Technical Information Service of the U.S. Department of Commerce. It is the means through which unclassified, unlimited distribution reports are made available to the public. The NTIS data base includes material from both the physical and social sciences, including topics of immediate widespread interest, such as environmental pollution and control, energy conservation, technology transfer, health planning, society problems, and urban and regional development and planning. The file covers the period from 1960 to present. The on-line review of NTIS produced nearly 700 references.

MATHFILE provides essentially complete coverage of world-wide literature on pure mathematics and of those works in applied mathematics, computer science, physics, engineering, biology, operations research, and other fields that contain new and interesting mathematics. The data base is produced by the American Mathematical Society and corresponds to the printed publication Mathematical Reviews. Beginning in July 1979, almost all MATHFILE records contain the full text of the evaluative review or abstract printed in Mathematical Reviews. The data base also includes records for 1973 onward. The on-line review of MATHFILE produced more than 150 references.

The COMPENDEX data base is the machine-readable version of Engineering Index, which provides abstracted information from the world's significant engineering and technological literature. It provides world-wide coverage of approximately 3500 journals, publications of engineering societies and organizations, papers from the proceedings of conferences, and selected government reports and books. COMPENDEX covers citations from 1970 to the present. More than 100 references were produced during the on-line review of COMPENDEX.

The INSPEC data base is the largest English language data base in the fields of physics, electrotechnology, computers, and control. It corresponds to the printed Physics Abstract, Electrical and Electronics Abstracts, and Computer and Control Abstracts. Non-English language source material is also included but is indexed and abstracted in English. Journal papers, conference proceedings, technical reports, books, patents, and university theses are abstracted for inclusion in INSPEC. The total number of journals scanned is over 2000; 200 of these are abstracted completely. INSPEC covers citations from 1969 to the present. During the on-line review of INSPEC, nearby 100 references were produced.

COMPREHENSIVE DISSERTATION ABSTRACTS is a definitive subject, title, and author guide to virtually every American dissertation accepted at an accredited institution since 1861, when academic doctoral degrees were first granted in the United States. Approximately 99 percent of all American dissertations are cited in this file. In addition, it serves to disseminate citations for thousands of Canadian dissertations and an increasing number of papers accepted in institutions abroad. Professional (e.g., M.D., L.L.D.) and honorary degrees are not included. Individual, degree-granting institutions submit copies of dissertations or lists of dissertations completed to University Microfilms International (UMI). Citations for these dissertations are included in the data base and in the following UMI publications: Dissertation Abstracts International, American Doctoral Dissertations, and Comprehensive Dissertation Index. A complete listing of sources consulted in compiling this data base can be found in the preface to any volume of Comprehensive Dissertation Index. Approximately 20 citations were obtained from this data base.

ECONOMIC LITERATURE INDEX provides coverage of the worldwide literature on economics from over 200 major economic journals chosen for their usefulness to research workers and teachers of economics. Only articles in English or with English language summaries are included. The data base is produced by the American Economic Association and corresponds in part to the printed Index of Economic Articles. Among the many subjects covered are: general economics, theory, history, systems, economic growth, development, planning, fluctuations, domestic monetary theory, natural resources, industrial organization, urban and regional economics, and quantitative economic methods and data. Updates to the data base correspond to the "Subject Index of Articles in Current Periodicals" section of the Journal of Economic Literature. Each record includes basic bibliographic data plus descriptors and a descriptor code corresponding to the classification scheme used in the printed works. Although the ECONOMIC LITERATURE INDEX does not provide the actual text of abstracts on-line, it does indicate whether an abstract is available in the Journal of Economic Literature. The data base covers the period from 1969 to the present. The on-line review of ECONOMIC LITERATURE INDEX produced over 250 references.

The approach followed in all of the computer file searches involved the specification of a strategy in which key words are used in combination to retrieve citations relevant to the topic. The basic strategy used was to combine two groups of key words: (1) one group relating to the subject of either models or data bases; and (2) the other group relating to the analytical approach. Some of the key words used to qualify the analytical approach were: assessment; evaluation; verification; validation; confidence; documentation; and sensitivity analysis. Once the two groups of key words used to qualify the search were entered, the computer searched the file for the intersection of the groups. That is, all citations with any combination of at least one key word from each group would be counted. If the resulting count turned out to be too large, ways were sought to exclude possibly irrelevant combinations.

After analyzing the results of the literature search, a scheme was devised for categorizing the studies. There are five major categories: (1) model evaluation practices; (2) data base evaluation practices; (3) selected federal documentation guidelines; (4) sensitivity analysis; and (5) techniques for handling special cases. A brief description of each category is presented in the next section of the report. The annotated bibliography itself is presented in Part II.

3. OVERVIEW OF ANNOTATED BIBLIOGRAPHY

MODEL/DATA BASE EVALUATION PRACTICES

The material presented in chapters 1 and 2 focuses on a summary of recommended practices for model and data base evaluation. Due to the amount of information published on this subject, it is not surprising to find that authors often disagree on the specific issues to be addressed in a model/data base evaluation. There does, however, appear to be a general consensus on the major phases of such an evaluation. For example, Kuh and Wood[1] divide the model/data base evaluation effort into three interrelated phases:

(1) descriptive;

(2) analytical; and

(3) experimental.

The descriptive phase is essentially an overview of all model/data base documentation and an in-depth review of related literature on similar models or data bases. It focuses on an explicit statement of the model's objectives, structure, and principal results. During the course of the literature review, the analyst begins to formulate hypotheses for testing in the final phase. A preliminary evaluation of the appropriateness of the model's structure for dealing with the policy issues for which it was designed would also be made. In some cases, the plausibility of the results would be evaluated through comparisons with the results produced by other analysis efforts.

The analytical phase goes beyond the literature review because it includes an in-depth analysis of both the underlying technical documentation and the model's computer source code. The purpose of this exercise is to relate the technical documentation to the source code and to document any perceived departures. It is this process which was earlier referred to as verification. Unfortunately, the interpretation of a model's source code is often a very difficult task which may well demand an even higher level of programming skill than was required to build the model in the first place. Consequently, a large-scale, policy-oriented model will probably never be totally verified. A major objective of this phase is to produce information on:

(1) the empirical content of the model and associated data base, perhaps with comparison to other empirical studies of similar applications;

(2) the limitations on the model/data base's applicability due to its basic structure; and

(3) those issues in the model/data base's structure, empirical content, and applications which require further "experimental" analysis.

[1]Kuh, E., and D. Wood. Independent Assessment of Energy Policy Models. Palo Alto, CA: Electric Power Research Institute, EPRI EA-1071, 1979.

The experimental phase relies heavily on the analysis of experimentally generated data. These data are usually generated through direct, hands-on operation of the model or access to the data base. Direct operation makes it feasible to carry out complicated tests which provide insight into how the model operates. Another rationale for this approach is that the closer one gets to the operation of a model, the more likely one is to identify errors and discrepancies between implementation and documentation. The cornerstone of the experimental phase is the use of sensitivity analysis. Among the goals of such a sensitivity analysis are a critical evaluation of the local and/or global characteristics, such as uniqueness and stability, associated with various policy impacts. Determining the effects of uncertainty in the underlying data base is also a common goal. Sensitivity analysis is the subject of chapter 4 where these and other applications are discussed. Specialized techniques such as rare events, missing data, error propagation, and stability, are the subject of chapter 5.

SELECTED FEDERAL DOCUMENTATION GUIDELINES

The references cited in sections 3.1 and 3.2 have been selected as the principal representatives of documentation standards and guidelines intended for use by the Federal government and its contractors. They have been developed as guidelines for contractual obligations, based on a concensus of needed standards, and carry an authority based on a perceived need for universal uniformity in the presentation of information on model and software projects.

Federal Information Processing Standards (FIPS)

Documents Federal Information Processing Standards Publications (FIPS PUBS) are developed by the Institute for Computer Sciences and Technology (ICST) at the National Bureau of Standards (NBS) and are issued under the provisions of the Federal Property and Administrative Act of 1949, as amended; Public Law 89-306; Executive Order 11717; and Part 6 of Title 15 of the Code of Federal Regulations. The goals of the FIPS PUB series are to:

(1) improve the life-cycle efficiency and effectiveness of Federal information technology resources;

(2) facilitate the competitive and economic procurement of systems, components and services;

(3) improve the portability of data, software, and technical skills across systems;

(4) protect systems and networks against unauthorized access, manipulation, and abuse;

(5) reduce waste, errors, and unnecessary duplication in the application and use of systems; and

(6) increase the productivity of the Federal workforce.

Institute for Computer Science and Technology Documents

ICST is a center of technical expertise in information technology. ICST focuses primarily on helping the Federal government make effective use of computers and information technology. ICST products, services, and technical support are used by all levels of government and the private sector as well. ICST's major activities are:

(1) determining requirements for and participating in the development of national and international voluntary industry standards for computer products and services;

(2) developing guidelines, technology forecasts, and other products to aid in the effective management and use of computers;

(3) disseminating and exchanging information with Federal, State and local governments, industry, professional, and research organizations on computer use and standards needs;

(4) providing technical support for the development of government policies in information technology;

(5) providing direct technical assistance to Federal agencies on a cost reimbursable basis; and

(6) carrying out applied research and development.

SENSITIVITY ANALYSIS

The references in chapter 4 focus on how the use of sensitivity analysis permits decision makers to deal with uncertainty. The need for sensitivity analysis in a policy context is apparent when one recognizes that policy or decision makers usually want more than single numbers or simple sets of numbers as answers. They desire a credible assessment of the range of possible outputs and how these can vary with different actions. Can, for example, desirable end results be reached by altering some of the decision variables and what are extreme possible outcomes if circumstances do not evolve quite as originally perceived?

The two major types of sensitivity analysis, deterministic and probabilistic, are treated in a variety of articles, some of which use advanced techniques. In a deterministic sensitivity analysis, the policy maker wants to set values of model parameters at any of a number of alternative levels, all of which may be feasible choices, though only one set may become the basis for action.

The modeler must be able to test the effect of each such alternative. To do this an experimental design for selecting specific values of input elements for analysis is first required. The model results based on the use of

selected inputs would then be analyzed via statistical techniques to quantify the effect that an individual input or combinations of inputs (independent variables) have on the response (dependent variable).

Probabilistic sensitivity analysis involves the explicit recognition of uncertainty in a model. Such methods generally lead to a more comprehensive sensitivity analysis, including the manner in which the output reflects the input randomness. In a probabilistic sensitivity analysis, the same strategies for putting alternative scenarios through the model are used, but the process involves a comprehensive approach, including:

(1) determination of input probabilities and (particularly) joint probability distributions, as necessary;

(2) formulation of an experimental design for selecting specific values of input elements for analysis; and

(3) identification of a statistical format for analyzing and presenting the results.

Of the advanced sensitivity analysis techniques, model sampling and response surface methodology appear to have the greatest applicability to the types of modeling problems faced by FEMA. Model sampling is a procedure for sampling from a stochastic process (even a single random variable in the simplest case) to determine, through multiple trials (generally via Monte-Carlo methods) of a carefully constructed computer-based experimental design, the nature and effects of a probability distribution that would be difficult or impossible to determine by standard statistical arguments. Response surface methodology, on the other hand, is typically carried out under a specific statistical experimental design. The term "response surface" refers to a formal mathematical relationship expressing the anticipated value of the dependent variable in an experiment as a function of the independent variables of the experiment. The relationship is typically derived by statistical analysis of the results of designed factorial experiments. An effort of this type is often completed by sampling of the variables found important and thought to be random, and then running the model at their values.

TECHNIQUES FOR HANDLING SPECIAL CASES

Rare Events

Much of the planning which FEMA must carry out as a part of its mandated responsibilities is concerned with the way the locality/region/nation should respond to the occurrence of a rare event. The articles in this section of the annotated bibliography are for the most part of a Bayesian flavor since little information is usually available on how the system of interest responds to the occurrence of a rare event. Several approaches involve the pooling of expert opinion to tighten the prior distribution. Most articles focus on one of two subject areas. The first area is concerned with rare and/or violent weather conditions; it includes models which are more akin to traditional statistical approaches. The second area deals with reactor safety. Due to the complex technological nature of the problem and the general lack of data on system/component failures, analysis often relies heavily on fault trees.

Missing Data

Techniques for dealing with missing data are the subject of a rapidly growing literature in statistics. The techniques discussed in this section of the annotated bibliography range from complicated imputation methods, based on a pre-specified experimental design, to mathematical programming (linear and dynamic), and finally ad hoc, rule-of-thumb procedures. As shown in section 5.2, the technique actually used depends both on the way in which the data was collected and the intended use of the data.

Error Propagation

The theory of error propagaton is a well-established branch of statistics which may be used in conjunction with sensitivity analysis in evaluating a large-scale model. Techniques for studying error propagation also appear to have important applications for systems which deal with rare events. The articles in this bibliographic section focus on three basic subjects: (1) errors in model structure; (2) parameter errors; and (3) inherent variability. Model structure errors focus on the process of simplification which is usually necessary to make the model tractable. When major decision parameters are based on data which are aggregated, transformed, or synthesized, their values should not be assumed to be known with certainty. Fortunately, statistical techniques, based on the theory of error propagation, may be used to analyze the impacts on the overall system due to parameter errors. Inherent variability, such as non-homogeneous capital stock, also leads to errors if ignored in model development. For example, non-homogeneity of the capital stock in a particular sector of the economy can be captured in part through reference to a vintage model of capital.

Stability

Local and global stability considerations are of particular importance in any large-scale, dynamic-optimization model. Such models may be concerned with the identification of an optimal investment trajectory following a major disruption which affects a large portion of the nation/region's capital stock. One should note the close relationship to the types of problems studied under the topic of sensitivity analysis to those discussed here. In many applications, the lack of stability may be characterized by either discontinuities in the response or movement to an asymptote. For example, the value of the response may be dependent on the path of a set of variables through the solution space rather than the values of the variables. Such phenomena may be analyzed via catastrophe theory. The more classical applications of stability are based on the theory of differential equations. A set of theorems, due to Liapounov, provide the theoretical basis for much of the current work in this area.

PART II: ANNOTATED BIBLIOGRAPHY

1. MODEL EVALUATION PRACTICES

1.1 METHODOLOGY

Anderson, T., and T. Sawa. "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," Econometrica, Vol. 47, No. 1, 1979, pp. 163-182.

This paper evaluates numerically the cumulative function (cdf) of the two-stage least squares (TSLS) estimate of the coefficient of an endogenous variable in a single equation of a system of simultaneous equations. Tables of the cdf's are given for enough values of the standardized coefficient, the number of excluded exogenous variables, and the noncentrality parameter, so that the cdf can be obtained for any case by suitable interpolations and transformations. Other tables evaluate the accuracy of approximations to the cdf's based on the normal distribution and asymptotic expansions. Some properties of the TSLS estimate are inferred, and comparisons with other estimates made.

Baggi, D., and M. Shooman. Software Test Models and Implementation of Associated Test Drivers. New York: Polytechnic Institute of New York, POLY-EE-79-0057, 1980, 42pp.

In the past, more software tests were constructed by heuristics and by drawing upon experience with similar software. Recently, enough preliminary work has been done to propose an analytical construction of test cases. This report begins by defining five broad classes of software tests: Type 0, Type 1, Type 2, Type 3 and Type 4. In a Type 0 test, all instructions are exercised at least once. In Type 1 and 2 tests, all flowchart paths are exercised at least once. Type 1 is performed by forced traversal and Type 2 by natural execution. Types 3 and 4 correspond to an exhaustive interaction of all input and stored data. Clearly, Types 3 and 4 are infeasible and only a strategy lying between Types 1 and 2 can effectively be implemented. Since enumeration of all the paths in a given program is required for Type 1 and 2 tests, this report establishes the lower and upper bounds on the number of paths as a function of the number of deciders, describes a manual decomposition procedure to cut a graph into smaller subgraphs, and proposes an algorithm to machine-identify all paths. A complete Type 1.5 driver system for forced path traversal, implemented in PL/1, is then thoroughly described, together with suggestions on how to extend these techniques to other languages. A typical program is analyzed manually, tested with data and run through the system. Some evaluation of the usefulness of the system is eventually given in the light of the accumulated experience.

Butterfield, M., and T. Mules. "A Testing Routine for Evaluating Cell by Cell Accuracy in Short-Cut Regional Input-Output Tables," Journal of Regional Science, Vol. 20, No. 3, 1980, pp. 293-310.

In the development of so-called nonsurvey or short-cut methods of generating regional input-output tables, the evaluation of the methods often compares an estimated input-output matrix with a survey based matrix or similar benchmark matrix. The paper discusses the limitations involved in using conventional statistical techniques for such matrix comparisons and highlights the input-output characteristics that are overlooked by such tests. It goes on to suggest a package of complementary tests, which is designed so that the loss of power of any particular test is compensated by other tests in the package. The package is then applied to assessing three different methods of generating input-output matrices for the Australian State of Western Australia.

Chinnis, J. Forecasting International Affairs: An Empirical Test of a Markov Renewal Model. Woodland Hills, CA: Perceptronics, Inc., AD-A115 656/1, 1982, 75pp.

A major difficulty for designers of systems to forecast international affairs has been to allow the use of both observable data and subjective estimates. A novel approach has been developed based upon a Bayesian stochastic model. The present report discusses an experimental test of the approach in an actual intelligence setting. In this experiment, both direct subject probabilistic forecasts and those generated by computer models performed well. There was some suggestion that the direct assessments might be better for short-term forecasts and that the computer models might be most valuable in longer-term forecasting (30 days and longer) in dynamic situations.

Cohon, J., and D. Marks. "A Review and Evaluation of Multiobjective Programming Techniques," Water Resources Research, Vol. 11, No. 2, 1975, pp. 208-220.

Three criteria are established for the evaluation of the utility of multiobjective programming techniques for water resource planning. The criteria are computational efficiency, explicitness of trade offs among objectives, and the amount of information generated for decision making. The multiobjective approaches are classified into generating techniques, techniques which rely on the prior articulation of preferences, and techniques which foster iterative definition of preferences. The methods in the various classes are reviewed and evaluated in terms of the hypothesized criteria. The evaluations are then used in establishing conclusions about the applicability of the multiobjective approaches to water resource problems.

Core, J., J. Cooper, P. Hanrahan and W. Cox, "Particulate Dispersion Model Evaluation: A New Approach Using Receptor Models," Journal of the Air Pollution Control Association, Vol. 32, No. 11, 1982, pp. 1142-1147.

Recent advances in the development of receptor-oriented source apportionment techniques (models) have provided a new approach to evaluating the performance of particulate dispersion models. Rather than limiting performance evaluations to comparisons of particulate mass, receptor model estimates of source impacts can be used to open new opportunities for in-depth analysis of dispersion model performance. Recent experiences in the joint application of receptor and dispersion models have proven valuable in developing increased confidence in source impact projections used for control strategy development. Airshed studies that have followed this approach have identified major errors in emission inventory data bases and provided technical support for modeling assumptions. This paper focuses on the joint application of dispersion and receptor models to particulate source impact analysis and dispersion model performance and evaluation. The limitations and advantages of each form of modeling are reviewed and case studies are examined.

De Man, R. "Energy Models and the Policy Process," Simulation and Games, Vol. 14, No. 4, 1983, pp. 445-464.

During the seventies hundreds of energy models were developed, varying in modeling techniques, scope, complexity, and size. One class of models, energy macro models, was designed to describe demand and supply of energy for a large region or a country. These energy macro models were expected to contribute substantially to the quality of energy policy. The author indicates some basic problems inherent to these energy macro models and finds that these expectations were too optimistic. Two classes of problems are mentioned: scientific weaknesses in the models themselves, and the way models have been used in the policy process. This article focuses on the second class of problems and draws examples from the Dutch experience in using energy macro models.

Dhrymes, P., et al. "Criteria for Evaluation of Econometric Models," Annals of Economic and Social Measurement, Vol. 1, No. 3, 1972, pp. 291-324.

This multi-authored article develops a framework for systematically evaluating large scale econometric models. Reasonably self-contained aspects of model evaluation include parametric evaluation prior to the release of the model (model selection, parameter estimation, and pseudo-forecasts and structural stability tests) and evaluation after release of the model. Many operational procedures for parametric evaluation are noted; alternative, ad hoc procedures are necessary in some cases, given the present state-of-the-art. Non-parametric validation procedures are then outlined. These include single-variable measures, tracking measures, error decomposition, and cyclical and dynamic properties. A statistical appendix sketches some of the theoretical results used in the paper.

Draper, N., and A. Herzberg. An Investigation of First Order Designs for Extrapolation Outside a Sphere. Madison, WS: University of Wisconsin, Mathematics Research Center, MRC-TSR-1354, 1974, 10pp.

Methods for selecting a first order response surface design when both variance and bias error exist are applied to a situation which it is desired to extrapolate the fitted model in all directions outside of a sphere within which all the runs are to be made. The region of interest is a spherical shell.

Faldella, E., and G. Iuculano. "Automatic Evaluation of Regression Models," Alta Freq, Vol. 49, No. 1, 1980, pp. 31-34.

A generally applicable algorithm for the evaluation of regression models is described. The correctness of each contribution to the regression is verified according to static tests which provide a high level of significance. On the basis of the described algorithm a program has been developed for automatic identification of the functional relationships characterizing the regression models. The program has been fruitfully exploited for a large computing system performance evaluation.

Friedman, R., et al. Use of Models for Water Resources Management, Planning, and Policy. Washington, DC: Office of Technology Assessment, OTA-O-159, 1982, 242pp.

This assessment is intended as a guide for Congress on the potential to develop and use mathematical models to more effectively and efficiently analyze the Nation's water resource problems. Policy options are presented for improving the use of these tools within both Federal and State water resource agencies. The assessment reviews the current technical capabilities of models; the current and potential use of models by Federal and State agencies; and general impediments to effective model development, use, and dissemination. It draws on the results of surveys of water resource professionals in 27 Federal agencies and offices, from all 50 States, and from universities and the private sector. Mathematical models are new techniques that are increasingly being used for water analysis. The Office of Technology Assessment (OTA) focused on mathematical modeling as a means of assessing the broader issue of our ability to analyze and plan courses of action to deal with current and long-range water problems. OTA did not investigate the more traditional forms of water resource analysis that are being replaced by mathematical models, or the less commonly used physical models. The technical capabilities of models vary greatly among the water resource issues analyzed in this report; however, the assessment finds that models capable of analyzing many pressing water resource issues are currently available, and have significant potential for increasing the accuracy and effectiveness of information available to managers, decision makers, and scientists. Institutional constraints to model use, including lack of information about available models, lack of training in model use and interpretation, lack of communication between decision makers and modelers, and lack of general support services, are identified as major impediments to increased model use for water resource analysis.

Gass, S., editor. Validation and Asessment Issues of Energy Models. Washington, DC: National Bureau of Standards, Special Publication 569, 1980, 559pp.

This workshop proceedings contains 29 papers dealing with the theoretical and applied state-of-the-art of validation and assessment, with emphasis on energy mathematical modeling. The following areas are addressed: activities of the Department of Energy in assessment and validation, taxonomy and structure of assessment and validation, the relationship between model assessment and policy research, the Electric Power Research Institute's Energy Modeling Forum and projects, independent third-party model assessment, the Texas National Energy Modeling Project, management and improvement of the modeling process, complexity of model evaluation, definitions and structure of model assessment approaches, model access and documentation, energy and econometric models, and sensitivity analysis.

Gass, S., editor. Validation and Assessment of Energy Models. Washington, DC: National Bureau of Standards, Special Publication 616, 1981, 248 pp.

The Symposium on Validation and Assessment of Energy Models, held at the National Bureau of Standards (NBS), Gaithersburg, MD (May 19-21, 1980), was sponsored by the Energy Information Administration (EIA), of the Department of Energy (DOE), Washington, DC. The symposium was organized by the NBS Operations Research Division with a two-fold agenda: (1) to summarize the recent ideas and advances of model validation and assessment that have been applied to DOE energy models, and (2) to hold workshops on key open questions that are of concern to the validation and assessment research community. Speakers addressed current and future practices, the EIA model validation program, model structure and data, and model credibility. Full-day workshop sessions were held on the following topics: validating composite models, the measurement of models, the measurement of model confidence, model structure and assessment, sensitivity and statistical analysis of models, and model assessment methodologies. This volume documents the symposium proceedings and includes the formal papers presented, discussant comments, panel discussions and questions and answers, and summaries of the issues and conclusions reached in the workshops.

Gass, S. "Decision-Aiding Models: Validation, Assessment, and Related Issues for Policy Analysis," Operations Research, Vol. 31, No. 4, 1983, pp. 603-631.

The extension of operations research (OR) or decision-aiding models and OR methodology from operational and technological settings to the field of policy analysis has caused analysts and users to question whether the OR process can handle the requirements of this new area. This paper reviews the evolving nature of OR methodology, the difficulties that have arisen in applying it, and the attempts to improve the use of OR methodology in policy analysis. The analyst's role in providing information to the user to determine whether and how a specific OR model can be used as a decision aid is emphasized. Relevant information for the analyst to provide covers model validation, assessment, utility, confidence and documentation.

Granger, C., and P. Newbold. "Some Comments on the Evaluation of Economic Forecasts," Applied Economics, Vol. 5, 1975, pp. 35-47.

Although a number of articles have been written over the years on the subject of evaluating economic forecasts, the authors feel that current practice in this area remains deficient. In the first place, a number of the evaluation techniques presently employed can generate false conclusions. Perhaps more importantly, the standards which a set of economic forecasts have been required to meet are not sufficiently stringent, since the object of any evaluation exercise ought to be self-critical rather than self-laudatory. This paper examines the problem of evaluating from two viewpoints. First, one would like to know how good is a particular set of forecasts. Of particular importance in this respect is the concept of conditional efficiency. Secondly, a number of diagnostic checks which ought to be made on forecast adequacy are outlined. In a good deal of econometric work tests against serial correlation in the forecast errors should prove particularly useful in this context. The paper discusses the validity of various evaluation techniques which have been proposed in the past, and concludes that a number of these can lead to misleading interpretations. The difficulty appears to be in the fact that, for a 'good' set of forecasts, the time series and distributional properties of the predictor series generally differ from those of the predicted series.

Ginnsrud, G., E. Finnemore and H. Owen. Evaluation of Water Quality Models: A Management Guide for Planners. Washington, DC: U.S. Environmental Protection Agency, EPA-600/5-76-004, 1967, 176pp.

This report is designed as a handbook specifically oriented to water quality and water resource planners and managers. It presents a large amount of basic information concerning water quality modeling including procedures for: model evaluation, model selection, integration of modeling with planning activities, and contracting modeling projects. Planners without previous experience in water quality modeling may use the information and procedures included in the handbook to determine whether a water quality model could and should be used in a particular planning program, and which specific model would be cost effective. This includes a step-by-step procedure leading to the rejection or selection of models according to specific project needs. The handbook discusses the implications which accompany the decision to model, including the need for additional labor and specialized technical expertise. Methods and procedures for integrating the use and results of water quality models with other activities of the planning process are described as well as the respective merits of in-house and contracted modeling. The handbook also deals with the procedures for obtaining and using contractual services for water quality modeling. Step-by-step instructions are provided for the preparation of solicitations, evaluation of proposals and selection of contractors.

Greenberger, M., and R. Richels. "Assessing Energy Policy Models: Current State and Future Directions," Annual Review of Energy, Vol. 4, 1979, pp. 467-500.

A great deal has been learned about the nature, problems, and possibilities of model analysis during the course of the MIT assessment of the Regionalized Electricity Model. Additional information has been accumulated through participation in the Energy Modeling Forum. It seems likely that model analysis in some form will become a permanent and important component of policy studies. But the shape it will take is not yet clear. Some argue that modelers are in the best position to do assessments and that in the long run the recent development of model analysis as a separate discipline will give way to assessments performed by modelers themselves on their own work and the work of others. This is a possibility, although it raises questions of impartiality and it is not clear that modelers will ever wish to spend the very significant time inspecting the work of others that assessment requires. Others believe that model assessment and forum analysis will eventually grow to resemble each other more than they do today and ultimately merge into a single type of analysis. This is another possibility, and the current work of the Energy Modeling Forum is indeed moving in that direction. But each kind of model analysis offers its own set of advantages, and at the present time the division of labor is an effective one. Still another possibility is that the users of a model will gradually take more initiative themselves in performing careful assessments of the models they use or are considering. The authors' guess is that the future development of model analysis will follow several of these alternate paths simultaneously--including the main path of separate growth. They see model analysis gradually broadening in scope to include not only comparative model assessments, but policy study critiques as well, where the postmortem review is directed not at a model, but a total policy study within which a modeling effort may or may not be contained. The most interesting question to them is whether development along the main path will become well enough received, especially by policy makers and model users generally, to lead to institutionalization of model analysis in the form of a recognized professional discipline with laboratories and centers in many application areas and in many regions of the country.

Hoffman, A., and C. Deeter, editors. Proceedings of the Workshop on Validation Computer-Based Mathematical Models in Energy Related Research and Development. Fort Worth, TX: Texas Christian University, 1979, 214pp.

In June 1978 a workshop on validation of mathematical models in energy related research and development was held at the Dallas/Fort Worth Airport Marina Hotel. The workshop was sponsored by the National Science Foundation and was hosted by Texas Christian University. Invited participants included representatives of government agencies, public and private research organizations, industry, and consulting firms whose experience and expertise included a serious concern for the problem of model validation. They also represented a broad spectrum of energy technologies (solar, nuclear, wind, fossil fuels) and a wide range of mathematical modeling specialties (econometric models, policy models, system design and performance models, sensitivity analysis, probabilistic models, etc.). The objectives of the workshop were the following: (1) to promote mutual understanding through examples and case studies; (2) to provide an opportunity for exchange of technical information; (3) to identify areas where cooperation would be mutually beneficial; (4) to identify needed research areas; (5) to identify means of broadening communication channels; and (6) to produce a workshop report which would be widely circulated among model builders.

Judge, G., M. Bock and T. Yancey. "Post Data Model Evaluation," Review of Economics and Statistics, Vol. 56, No. 2, 1974, pp. 245-253.

Within the context of the general linear regression model, the squared error loss measure for gauging estimator performance, preliminary test, Stein rule, and least squares estimators are specified, evaluated, and compared. It is noted that the James-Stein positive part and the modified James-Stein estimators, not commonly seen in the econometric literature, are uniformly superior over the range of the parameter space to the conventional and preliminary test estimators, respectively.

Labys, W. "Measuring the Validity and Performance of Energy Models," Energy Economics, 1982, pp. 159-168.

This paper is concerned with approaches to validating and judging the performance of energy models. There has been a proliferation of models of all sorts. These models not only attempt to deal with many different aspects of industry and market behavior but also involve widely different modeling methodologies. Several attempts have been made to evaluate these methodological approaches. The major focus of this paper is on criteria for determining the validity of a model that are quantitative. Different validation criteria are suggested for econometric time-series models, input-output models, and mathematical programming models. Finally, an appeal is made for including validation measures in future energy modeling studies.

Melsa, J. "A Survey of Fixed Configuration Optimization in Estimation and Control," in Proceedings of the Hawaii International Conference on System Sciences, Dallas, TX: Southern Methodist University, AFOSR-70-0953TR, 1970, pp. 251-254.

Fixed-configuration optimization is the modern version of the classical parameter optimization problem. The concept is to fix the configuration of the system with the exception of certain parameters which are then selected so as to optimize the performance. Application of fixed-configuration optimization to deterministic control is often referred to as the specific optimal control (SOC) problem following the terminology used in an early paper by Agarwal and Sridhar in which several numerical techniques were examined for solving SOC problems. A later paper by Eisenberg and Sage extended the results of Agarwal and Sridhar; various aspects of the sensitivity problem in SOC and some possible solutions were examined in the papers by Burns and Kumar and Sims and Melsa. A natural extrapolation of the SOC work was the application to estimation; specific optimal estimation (SOE) was developed by Sims and Melsa by reformulating the problem as a deterministic equivalent problem. Recently the concepts of the SOC and SOE work have been combined by Hardaway and Melsa to form the specific optimal stochastic control (SOSC) problem. This paper surveys and evaluates these previous efforts.

McNees, S. "The 'Rationality' of Economic Forecasts," The American Economic Review, Vol. 68, No. 2, 1978, pp. 301-305.

Forecasts can be evaluated relative either to an alternative forecasting procedure (traditionally a naive method) or to a desired statistical property (such as unbiasedness, efficiency, or rationality). This paper employs two statistical tests to examine whether the ex ante forecasts of three prominent forecasters were "rational." The first is a test of unbiasedness, a common definition of rationality. Some investigators have taken acceptance of the null hypothesis of unbiasedness as sufficient for inferring rationality, even though unbiasedness need not imply most accurate or best. If rationality is understood in its strong form as incorporating all available information, this test is inadequate. A second test examines whether the forecasts could have been improved on the basis of information available at the time of the forecast. Rather than searching the universe of plausible explanatory information, this investigation is confined to past forecast errors as they were known to the forecaster at the time the forecast was made. The final section of the paper discusses the implication of these results as a test of part of the set of propositions known as the rational expectations hypothesis.

Mincer, J., and V. Zarnowitz. "The Evaluation of Economic Forecasts," in Economic Forecasts and Anticipations: Analyses of Forecasting Behavior and Performance, New York: National Bureau of Economic Research, 1969, pp. 3-46.

An economic forecast may be called "scientific" if it is formulated as a verifiable prediction by means of an explicitly stated method which can be reproduced and checked. Comparisons of such predictions and the realizations to which they pertain provide tests of the validity and predictive power of the economic model which produced the forecasts. Such empirical tests are an indispensable basis for further scientific progress. Conversely, as knowledge accumulates and the models improve, the reliability of forecasts, viewed as information about the future, is likely to improve. Forecasts of future economic magnitudes, unaccompanied by an explicit specification of a forecasting method, are not scientific in the above sense. The analysis of such forecasts, which we shall call "business forecasts," is nevertheless of interest. There are at least five reasons for this interest in business forecasts. First, to the extent that the predictions are accurate, they provide information about the future. Second, business forecasts are relatively informative if their accuracy is not inferior to the accuracy of forecasts arrived at scientifically, particularly if the latter are more costly to obtain. Third, the margin of inferiority (or superiority) of business forecasts relative to scientific forecasts serves as a yardstick of progress in the scientific area. Fourth, regardless of the predicitive performance ascertainable in the future, business forecasts represent a sample of the currently prevailing climate of opinion. They are, therefore, a datum of some importance in understanding current economic behavior. Finally, even though the methods which produce the forecasts are not specified by the forecasters, it is possible to gain some understanding of the genesis of forecasts by relating the predictions to other available data. This paper is concerned with the analysis of business forecasts for some of these purposes. Specifically, the authors are interested in methods of assessing the degree of accuracy of business forecasts both in an absolute and in a relative sense. In the Absolute Accuracy Analysis they measure the closeness with which predictions approximate their realizations. In the Relative Accuracy Analysis they assess the net contributions, if any, of business forecasts to the information about the future available from alternative, relatively quick and cheap methods.

Mitchell, T., and D. Wilson. Energy Model Validation: Initial Perceptions of the Process. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/CSD-50, 1979, 43pp.

The concept and practice of energy model validation are addressed. The objectives of this work are to give a general indication of what model validation means; to identify, in an organized way, a set of specific questions and activities that are pertinent to model validation; and to indicate ways in which mathematicians and statisticians can contribute in consulting and research roles. The approach includes a breakdown of the model development and modification process into various components: mental model formulation, mathematical model formulation, program design, coding and keypunching, computing process, comparison with data and expectation, real world information process, data generation process, model correction and adjustment, and the policy process. Some standards for the type of extensive model documentation necessary to permit a full-scale validation effort are suggested, and some specific questions to be examined are identified. The process of active validation and the role of mathematicians and statisticians are also discussed.

Mucciardi, A., and E. Gose. "An Automatic Clustering Algorithm and Its Properties in High-Dimensional Spaces," in IEEE Transactions on Systems, Man, and Cybernetics, Vol. smc-2, No. 2, 1972, pp. 247-254.

An efficient technique for approximating a joint N-dimensional probability density function has been described by Sebestyen and Edie. The algorithm searches for clusters of points and considers each cluster as one hyperellipsoidal cell in an N-dimensional histogram. Among the advantages of this scheme are: (1) the histogram cell descriptors--location, shape, and size--can be determined adaptively from sequentially introduced data samples of known classification, and (2) the number of cells required for a good fit can usually be held to a small number. No assumptions are required about the underlying statistical structure of the data. The paper presents heuristics which were developed to automate the selection of the control parameters. The properties of these parameters were studied as a function of dimension. Two of the control parameters were found to be linearly related to dimension. This provides a method for determining their value by extrapolation, thereby avoiding a great deal of computation.

Osborn, D., and F. Teal. "An Assessment and Comparison of Two NIESR Econometric Model Forecasts," National Institute Economic Review, Vol. 88, 1979, pp. 50-62.

This article presents a methodology for decomposing ex ante forecasting error into exogenous variable error, data revision error, model error, and judgment error. This method is applied to the forecasts made by the National Institute in February 1975 and February 1976. The first section describes the methodology including the NIESR forecasting procedure. Then the NIESR model (with some of its problems) is discussed together with the data used in the study. The method for decomposing the forecasting error for 1975 is applied and a similar analysis presented for 1976. Some conclusions and a summary complete the article.

Palm, F., and A. Zellner. "Large Sample Estimation and Testing Procedures for Dynamic Equation Systems," Journal of Econometrics, Vol. 12, No. 3, 1980, pp. 251-283.

Joint two-step estimation procedures, which have the same asymptotic properties as the maximum likelihood (ML) estimator, are developed for the final equation, transfer function, and structural form of a multivariate dynamic model with normally distributed vector-moving average errors. The ML estimator under fixed and known initial values is obtained by iterating the procedure until convergence. The asymptotic distribution of the two-step estimators is used to construct large sample testing procedures for the different forms of the model. [Two related articles, by McDonald and Darroch and by Palm and Zellner, are cited in the Appendix (A.1).]

Ramamoorthy, C. The Design Methodology of Distributed Computer Systems. Berkeley CA: University of California, Electronic Research Laboratory, AFOSR-TR-81-0558, 1980, 42pp.

This is a final report for research in distributed computer systems. Chapter 2 describes the top-down development approach. The development process is divided into four successive phases: (1) requirement and specification phase; (2) design phase; (3) implementation phase; and (4) evaluation and validation phase. Guidelines and automated tools for the first two phases are developed. A graphical method (using the max-flow min-cut algorithm and cut-tree concept) to decompose and partition a computer system into loosely coupled subsystems is proposed. The implementation and the evaluation and validation phases are outlined briefly only because they are very technology and architecture dependent. Chapter 3 examines the Petri net model for asynchronous concurrent systems. Procedures based on Petri net for predicting and verifying the system performance are presented. The computational complexities of these procedures are also shown. Chapter 4 examines the analysis techniques for deadlocks in asynchronous concurrent systems. In particular, the author studies in detail deadlocks caused by conflicts in mutually exclusive access to resources with the contraint that each resource type has only one member. Chapter 5 first classifies and then evaluates several existing software reliability models according to some proposed criteria. Then it develops a theory of software reliability based on the nature of the input domain of the program, i.e., the size of the errors and the number, complexity and continuity of equivalence classes formed in the input domain.

Saib, S., D. Andrews, J. Benson, N. Brooks and R. Melton. Advanced Software Quality Assurance. Santa Barbara, CA: General Research Corporation, CR-6-720, 1977, 265pp.

This is a report of the work performed by General Research Corporation during the Advanced Software Quality Assurance contract. Research was conducted in four major areas: (1) assertions, verification conditions and consistency proof; (2) design of an executable assertion language and the development of a preprocessor to implement this language; (3) development of a Software Quality Laboratory; and (4) evaluation of existing languages as applied to BMD software problems. This report gives a methodology for verifying software. The preprocessors which allow assertions to be placed in FORTRAN and PASCAL programs are described. The static analysis developed as part of the Software Quality Laboratory are also described and examples are given of their use. The Concurrent PASCAL programming language is applied to a generic model of a BMD software system.

Salkever, D. "The Use of Dummy Variables to Compute Predictions, Prediction Errors, and Confidence Intervals," Journal of Econometrics, Vol. 4, No. 4, 1976, pp. 393-397.

This paper presents a method for computing predictions, prediction error variances, and confidence intervals, which can be implemented with any regression program. It demonstrates that a regression estimated for an augmented data set, obtained by (1) combining n sample points with r forecast points and (2) including r dummy variables (each equalling one only for the corresponding forecast point), will yield r dummy variable coefficients and variances, which equal the corresponding prediction errors and prediction error variances. Since most programs lack special routines to calculate these magnitudes and manual computation is cumbersome, the proposed method is of considerable practical value.

Sargent, R. An Assessment Procedure and a Set of Criteria for Use in the Evaluation of "Computerized Models and Computer-Based Modelling Tools." Syracuse, NY: Syracuse University, Department of Industrial Engineering and Operations Research, WP-80-016, 1981, 25pp.

This paper contains a flexible and adaptive assessment procedure and a set of assessment criteria for use in having assessments conducted of general purpose computerized models and computer-based modeling tools. Basically, the procedure requires the use of a third party to make subjective judgments on a set of criteria specified by the assessment sponsor and to make an overall recommendation on the use of the model or modeling tool.

Tessmer, J. "Example of Software Validation Using a Factorial Design," in Proceedings of the Conference on the Design of Experiments in Army Research Development and Testing, Washington, DC: U.S. Department of Energy, 1983, pp. 363-385.

This paper reports on the efforts of the Department of Energy to test and evaluate, validate, a large computer model, which represents the world petroleum distribution system. The evaluation technique employed is a complete two to the fourth power factorial design. The main effects, as well as all second order effects are estimated. The technique provided critical insights into the nature of the software identifying errors and assisted in the development of a methodology for re-examining candidate crude mixes of oil stored by the government for use during possible petroleum interruptions.

Treyz, G. "An Econometric Procedure for Ex Post Policy Evaluation," International Economic Review, Vol. 13, No. 2, 1972, pp. 212-222.

The structural equation residual and the predetermined values for an econometric model can be observed ex post. If the former values are used to adjust the contemporaneous intercept terms in the structural equations, the model solution will trace out the observed values for the endogenous variables exactly. Under certain assumptions any changes from this solution that are caused by changes in predetermined variables can be attributed to the alternative policy under which these changes would have been made. By using this simulation method with the 1969 Wharton-EFU Model and experimenting with various personal tax rates, optimum tax policies for 1968 and 1969 are found for an arbitrary objective function based on an unemployment-inflation trade-off. Subject to the limits of this procedure, it is possible to determine whether a policy based on an unadjusted Wharton forecast would have been superior to the macroeconomic policy that was followed by the government. The simulation procedure suggested may be more appropriate than current practice for many ex post investigations with non-linear econometric models.

Weisbin, C., R. Peelle and A. Loebl. "An Approach to Evaluating Energy-Economy Models," Energy, Vol. 6, No. 10, 1981, pp. 999-1027.

The increasing importance of models that stress the energy portion of the economy has created the demand for evaluation, by other than the model developer, to determine the strengths and weaknesses of a model with respect to the needs of the intended user. The literature describing how evaluations should be performed shows agreement on several basic components of the evaluation process, but at the same time is confusing because, although the same words may be used to describe these components, they often have different meanings. Although the models of interest are quantitative and the corresponding computer programs yield numerical results, published guidelines on model evaluation have lacked operationally defined checkpoints that can be used to summarize concisely and quantitatively the status of knowledge at a given time. Information related to desirable endpoints such as user requirements in terms of model outputs and associated maximum uncertainties, data sensitivities, etc., have customarily been expressed in more qualitative terms that are difficult to measure. A set of guidelines is developed for evaluating energy-economy models; the components of a thorough model evaluation are defined in an operational manner. Quantitative measures based on sensitivity and uncertainty analysis of results important to the user are included. Examples of the various components are drawn from the literature and from evaluation by Oak Ridge National Laboratory of the Long-Term Energy Analysis Program (LEAP) used by the Energy Information Administration for making projections to year 2020 included in that agency's 1978 Annual Report to Congress. Possible paths for future development of evaluation techniques are indicated, particularly those for which evaluation components must be more strictly defined after further practical experience is gained.

Williams, W. and M. Goodman. "Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts," Journal of the American Statistical Association, Vol. 66, No. 336, 1971, pp. 752-754.

A simple method for the construction of empirical confidence intervals for time series forecasts is described. The procedure is to go through the series making a forecast from each point in time. The comparison of these forecasts with the known actual observations will yield an empirical distribution of forecasting errors. This distribution can then be used to set confidence intervals for subsequent forecasts. The technique appears to be particularly useful when the mechanism generating the series cannot be fully identified from the available data, or when limits based on more standard considerations are difficult to obtain.

Wood, D., and M. Mason. Recommendations Concerning Energy Information Model Documentation, Public Access, and Evaluation. Cambridge, MA: Massachusetts Institute of Technology, Cambridge Energy Laboratory, MIT-EL 81-016, 1979, 119pp.

This report reviews the Energy Information Administration's (EIA) response to Congressional and management concerns relating specifically to energy information system documentation, public access to EIA systems, and scientific/peer evaluation. The relevant organizational and policy responses of EIA are also discussed. An analysis of the model development process and approaches to, and organization of, model evaluation as well as a survey of model evaluation studies are also presented. A more detailed analysis of the origins of the legislated documentation and public access requirements is presented in Appendix A, and the results of an informal survey of other agency approaches to public access and evaluation are presented in Appendix B. Appendix C provides a survey of non-EIA activities relating to model documentation and evaluation. Twelve recommendations to improve EIA's procedures for energy information system documentation, evaluation activities, and public access are also presented.

Zarnowitz, V. "An Analysis of Annual and Multiperiod Quarterly Forecasts of Aggregate Income, Output, and the Price Level," Journal of Business, Vol. 52, No. 1, 1979, pp. 1-33.

This paper examines several sets of forecasts representing a variety of sources, models, and techniques. Predictions of percentage changes in GNP are relatively accurate, but partly because of negative correlations between errors in the corresponding forecasts of real growth and inflation. For the last two variables, the multiperiod quarterly forecasts for the 1970's have errors that show unusually rapid cumulations over longer spans. The accuracy and properties of the forecasts depend heavily on the economic characteristics of the periods covered, but only weakly and not systematically on the differences among the forecasts.

Zarnowitz, V. An Appraisal of Short-Term Economic Forecasts, Occasional Paper 104, New York: National Bureau of Economic Research, 1967.

This report is one of the products of the study of short-term economic forecasting, a project of the National Bureau of Economic Research. The primary purpose of the project is to assess the accuracy of short-term forecasts of aggregate economic activity in the United States. The materials compiled and analyzed are authentic (ex ante) forecasts of the nation's economic fortunes for the near future--the next year or two or several shorter periods. The forecasts cover comprehensive measures of the value or volume of output for the economy as a whole and its main sectors: gross national product (GNP), its major components, and industrial production. Eventually, other variables will be added, such as personal income, employment, unemployment, and the main price indexes.

Zarnowitz, V., editor. "Forecasting Economic Conditions: the Record and the Prospect," The Business Cycle Today: Fiftieth Anniversary Colloquium I, New York: National Bureau of Economic Research, 1972, pp. 183-239.

This article summarizes the studies conducted at the National Bureau of Economic Research in evaluating the accuracy of a number of macroeconomic forecasts. The materials analyzed in these studies consist of predictions by business economists, as well as by large-scale econometric models.

Zarnowitz, V. "On the Accuracy and Properties of Recent Macroeconomic Forecasts," American Economic Review, Vol. 68, No. 2, 1978, pp. 313-319.

This article reviews the record of four major forecasts of the Gross National Product (GNP) and the Implicit Price Deflator (IPD) from the fifties through the seventies. The end-of-year forecasts of annual percentage changes in nominal GNP earn good marks for overall accuracy. Moreover, they are found to have improved in the period since the early 1960's compared with the years after World War II. The real growth and inflation forecasts are less accurate. The former forecasts suffer from large turning point errors, the latter from large underestimation errors. The errors in predicting real growth are negatively correlated with the errors in predicting inflation, which helped to make the nominal GNP forecasts more accurate.

1.2 CASE STUDIES

Ahlers, D., and J. Lakonishok. "A Study of Economists' Consensus Forecasts," Management Science, Vol. 29, No. 10, 1983, pp. 1113-1125.

This study examines the ability of economists to forecast ten major economic series. The data for this study were provided by J. A. Livingston of the Philadelphia Inquirer, who since 1947 has collected forecasts for the upcoming 6 and 12 months. The results reveal that, in general, for the period from 1947 through 1978, the economists in Livingston's sample did not produce efficient forecasts and were not able to outperform simple statistical models. It should be noted, however, that a substantial and consistent improvement in forecasting performance by economists in Livingston's sample did occur over this same period. These results contain important information for managers who use macro-economic consensus forecasts.

Ballard, K., and R. Wendling. "The National-Regional Impact Evaluation System: A Spatial Model of U.S. Economic and Demographic Activity," Journal of Regional Science, Vol. 20, No. 2, 1980, pp. 143-158.

Discussed is a modeling methodology and an operational multiregional model. The model, called NRIES, comprises 51 state econometric models that are integrated into a bottom-up model of the U.S. economy. Each of the state stochastic equations are estimated econometrically using time-series data. There are 230 variables for each state. The specification of stochastic variables and the error performance of major aggregate variables are presented. A hypothetical policy application of the model, a federal program to reduce unemployment, is discussed.

Bradley, G., G. Brown and G. Graves. Review of Computational Aspects of Planning Models. Carmel, CA: Optimal Decisions, EPRI-EA-2255-V.1, 1982, 48pp.

This report represents the results of a review of the computational aspects of the CRA/EPRI Coal Market Analysis System developed by Charles River Associates. The computational analysis focuses on validation of the computer model, ease of solution by the user, and reduction of solution costs by use of alternative computational procedures. The analysis shows that the computer model faithfully implements the Coal Market Analysis System and that the computational procedures produce the results described in the model documentation. Analysis of the computational structure and solution of the model by two independent analyses of the embedded linear programming problems shows that significant computational savings could be realized by exploiting the special structure that is present in the model. The model is shown to be a network with gains that can be solved by specialized computer packages. The problems are also shown to have a significant generalized upper bound structure that can also be exploited using specialized computer packages.

Brookhaven National Laboratory. Dynamic Energy System Optimization Model. Upton, NY: Brookhaven National Laboratory, EPRI EA 1079, 1979, 24pp.

A set of models of the U.S. energy system has been developed at Brookhaven National Laboratory (BNL). The models are linear programs which minimize the cost of energy consumption, subject to supply, demand, and process constraints. These optimization models provide a consistent framework for studying the nation's energy system. One of the BNL models is the Dynamic Energy System Optimization Model (DESOM), which examines the national energy system over an extended time horizon. This report on the work done for the DESOM project is presented in four parts. First, the general formulation of the model is detailed, with particular emphasis on electric sector modeling and its validation. Next, the sample assessments are described. The fourth section comprises a user's guide to running the model with PDS/Gen software. The final section shows the reformulation required in order to employ a nested decomposition algorithm with this problem, and describes experience with this solution approach.

Brown, C., and T. Sheriff. "Approaches to Medium-Term Assessments," National Institute Economic Review, Vol. 86, 1978, pp. 55-64.

In 1965, the National Institute of Economic and Social Research published The British Economy in 1975 by W. Beckerman and associates. This article summarizes the results of a post-mortem on the book, which was carried out at the Institute, together with a general discussion of approaches to medium-term projections.

Chandru, V., N. Goldman, M. Manove, M. Mason and D. Wood. The ICF, Inc. Coal and Electric Utilities Model: An Analysis and Evalution. Cambridge, MA: Massachusetts Institute of Technology, Cambridge Energy Laboratory, MIT-EL-81-015-2, 1981, 147pp.

This volume presents an evaluation of the ICF, Inc. Coal and Electric Utilities Model (CEUM) documentation, and a verification of the model's implementation. Chapter 1 reviews the development history and previous application of the CEUM. Chapter 2 presents an evaluation of the CEUM documentation, and Chapter 3 extends the existing documentation by providing a detailed mathematical formulation of the LP portion of the CEUM. Chapter 4 reviews the program structure and operating characteristics. Finally, Chapter 5 presents the results of verifying the correspondence between documentation and computer implementation, the accuracy of implementation, and the effect of implementation errors upon model results.

Christy, P., and K. Horowitz. "An Evaluation of BLS Projections of 1975 Production and Employment," Monthly Labor Review, Vol. 2, No. 8, 1979, pp. 8-19.

An evaluation of BLS 1975 industry projections showed the macro projection of the U.S. economy to be the most important source of error. This overestimation of demand led to an overprojection of industry outputs, which combined with productivity factors to produce spurious accuracy in industry employment projections. Equations for investment and imports were particularly poor, while those for personal consumption expenditures and government purchases performed well.

Fair, R. "An Evaluation of a Short-Run Forecasting Model," International Economic Review, Vol. 15, No. 2, 1974, pp. 285-303.

In this paper the ex ante forecasting performance of the model described in A short-run forecasting model of the United States economy (1970) is examined for the period from the third quarter of 1970 through the second quarter of 1973. The ex ante forecasts have always been generated from the model in as mechanical a way as possible. The results indicate, contrary to the prevailing view in the literature, that reasonably accurate forecasts can be produced from a model that is not subjectively adjusted. The results also show, again contrary to results that have been obtained for other models, that the forecasting accuracy of the model is generally improved when the actual values of the exogenous variables are used in place of the ex ante forecasted values. The overall results appear encouraging enough to warrant the suggestion that a more scientific approach be taken to econometric forecasting.

Ford, W., J. Horwedel, J. McAdoo, R. Alsmiller and B. Toney. Operating Experience with LEAP from the Perspective of the Computing Applications Analyst. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/CSD/TM-141, 1981, 176pp.

In conjunction with the Energy Information Administration's interest in forecasting the long-range impact of new technologies, the Long-Term Energy Analysis Program (LEAP), which was used for the energy price-quantity projects in the 1978 Annual Report to Congress (ARC '78), was used in an ORNL research program to develop and demonstrate a procedure for evaluating energy-economic modeling computer codes and the important results derived therefrom. The LEAP system used in the ORNL research, the mechanics of executing LEAP, and the personnel skills required to execute the system are described. In addition, a LEAP sample problem, subroutine hierarchical flowcharts, and input tables for the ARC '78 energy-economic model are included. Results of a study to test the capability of the LEAP system used in the ORNL research to reproduce the ARC '78 results credited to LEAP are presented. This document is one report in a series of reports which describe the ORNL LEAP-related research program.

Fromm, G., and L. Klein. A Comparison of Eleven Econometric Models of the United States," American Economic Review, Vol. 63, No. 2, 1973, pp. 385-393.

Comparison of solution error and dynamic multipliers of major U.S. econometric models shows, despite some exceptions, that errors are within reasonable bounds and that there is relatively uniform agreement about the pattern and magnitude of fiscal policy impacts. However, the authors have just begun to obtain the standardized comparisons which are needed to assess the merits of different models and their usefulness for forecasting and structural policy analyses.

Fromm, G., and L. Klein. "The NBER/NSF Model Comparison Seminar: An Analysis of Results," Annals of Economic and Social Measurement, Vol. 5, 1976, pp. 1-27.

This paper reports comparisons of selected error characteristics and policy multipliers of 11 major econometric models of the U.S. economy. These results were generated by a cooperative effort of the leading model builders under the aegis of the Model Comparison Seminar of the NBER/NSF Conference on Econometrics and Mathematical Economics. Comparisons of turning point performance, error decomposition, alternative policy analyses, and other characteristics are now underway and will be presented in subsequent seminar symposia.

Gass, S., K. Hoffman, R. Jackson, L. Joel and P. Saunders. The NBS Energy Model Assessment Project: Summary and Overview. Washington, DC: National Bureau of Standards, NBSIR 80-2128, 1980, 43pp.

This report is a summary of the activities and technical reports for the Energy Model Validation Procedure Development project undertaken by the NBS Operations Research Division for the Department of Energy, using DOE's Midterm Oil and Gas Supply Modeling System (MOGSM) as a test vehicle. The reports cover: (1) assessment of the documentation of MOGSM; (2) analysis of (a) the model methodology, (b) characteristics of the input and other supporting data, (c) statistical procedures undergirding construction of the model, and (d) sensitivity of the outputs to variations in input; as well as (3) guidelines and recommendations for the role of the above topics in building models and in developing procedures for their evaluation.

Goldman, N., M. Mason and D. Wood. An Evaluation of the Coal and Electric Utilities Model Documentation. Cambridge, MA: Massachusetts Institute of Technology, Cambridge Energy Laboratory, MIT-EL-81-007, 1981, 241pp.

This report is the first in a series of the MIT Energy Model Analysis Program (EMAP) project. It is concerned with the development of procedures for planning and implementing effective documentation. It presents the results of an MIT analysis of policy model documentation as well as the Energy Information Administration's approach to model documentation.

Hansen, J., M. Becker and J. Trimble. The Economic Foundations of LEAP Model 22C. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL -5748, 1981, 115pp.

The Department of Energy's Energy Information Administration (DOE/EIA) annually publishes energy future projections in its Annual Report to Congress. These projections, which are divided into short, mid- and long-term, are based on DOE/EIA energy-economy models. The model for long-term projections is the DOE/EIA'S Long Term Energy Analysis Program (LEAP). This report identifies and presents an initial assessment of the major underlying economic assumptions of LEAP Model 22C, the version of LEAP used to prepare the 1978 Annual Report to Congress. The major assumptions of Model 22C are: competitive markets; constant input proportions in production processes; constant returns to scale in production processes; unlimited supplies of capital and variable inputs at fixed prices; continuous (infinitely divisible) units of input; investment based on perfect foresight; technical change limited to exogenous forecasts; minimal regional differences (except for coal production); unlimited supplies of foreign oil at exogenously specified prices; very little government regulation; and a final demand quantities which are essentially exogenously specified.

Henry, S., M. Sawyer and P. Smith. "Models of Inflation in the United Kingdom: An Evaluation," National Institute Economic Review, Vol. 77, 1976, pp. 60-71.

In this article re-estimates are made of some of the main econometric models of inflation that have been proposed for the U.K. Primarily these are wage inflation models, though reference is also made in the article to price equations. Estimates for each model are provided for a number of different postwar samples to investigate the reliability of each model and their adequacy in light of recent high rates of inflation. The results for most of the models the authors consider are generally negative, except for a wage inflation equation based on the hypothesis of a target rate of increase in real net earnings desired by workers, which shows a great deal of consistency over the different sample periods.

Kuh, E., and D. Wood. Independent Assessment of Energy Policy Models. Palo Alto, CA: Electric Power Research Institute, EPRIEA-1071, 1979, 262pp.

Energy policy models are playing an increasingly important and visible role in supporting both private and public energy policy research and decision making. As importance has increased so too has the need for model review and assessment to assist in establishing model credibility for users and those affected by model-based policy research. Toward this end EPRI has sponsored the MIT Cambridge Energy Laboratory project to assess two important energy system models, the Baughman-Joskow Regionalized Electricity Model and the Wharton Annual Energy Model, and to identify and analyze organization and procedural issues in the model assessment process.

Makridakis, S., and M. Hibon. "Accuracy of Forecasting: An Empirical Investigation," Journal of the Royal Statistical Society, Sec. A, Vol. 142, No. 2, 1979, pp. 97-145.

In this study, the authors used 111 time series to examine the accuracy of various forecasting methods, particularly time-series methods. The study shows, at least for time series, why some methods achieve greater accuracy than others for different types of data. The authors offer some explanation of the seemingly conflicting conclusions of past empirical research on the accuracy of forecasting. One novel contribution of the paper is the development of regression equations expressing accuracy as a function of factors such as randomness, seasonality, trend-cycle and the number of data points describing the series. Surprisingly, the study shows that for these 111 series simpler methods perform well in comparison to the more complex and statistically sophisticated ARMA models.

McFadden, D., et al. An Evaluation of the ORNL Residential Energy Use Model. Cambridge Energy Laboratory, MIT-EL 81-021, 1981, 145pp.

This report provides an evaluation of the architecture, empirical foundation, and applications of the Oak Ridge National Laboratory (ORNL) residential energy use model. A particular effort is made to identify the strengths and shortcomings of the model for alternative uses, and to identify areas where model structure and empirical support could be upgraded. Concrete suggestions are made for improvements in model logic, strengthening the empirical basis for behavioral and technical aggregation. The overall conclusion is that the model has the potential to provide adequate forecasts of the aggregate impacts at a regional or national level of policies whose effects on households are relatively homogeneous. There are a number of model changes which would be relatively easy to implement, and which should substantially improve forecasts of this sort. On the other hand, the aggregate architecture of the ORNL model makes it fundamentally unsuitable for applications to geographical areas smaller than DOE regions, or to policies which have a heterogeneous impact on households.

Personick, V., and R. Sylvester. "Evaluation of BLS 1970 Economic and Employment Projections," Monthly Labor Review, Vol. 99, No. 8, 1976, pp. 13-26.

In 1966, BLS published the results of its first attempt at projecting a detailed set of economic and employment variables. A comparison of these projections for the year 1970 with what actually occurred provides a framework for identifying sources of error and improving the projections methodology. Estimates of aggregate labor force, GNP, and employment were all within 4 percent of actual values, but some of the industry output and employment projections contained large errors. The largest source of error in the employment projections at the industry level was the productivity factor. Input-output coefficients were also significant contributors. Incorrect conditional assumptions, though affecting selected industries severely, were found to have relatively small impact on the final results.

Pfaff, P. "Evaluation of Some Money Stock Forecasting Models," Journal of Finance, Vol. 32, No. 5, 1977, pp. 1639-1646.

The predictive performance of a number of econometric models of the money stock is examined in this study. The econometric models whose parameters were estimated with quarterly 1947-60 data were used to forecast the 1961-70 money stock. The performance of the econometric models was mixed. The root-mean-squared-error (RMSE) of its forecasts, especially of those models without the lagged variable, is higher than that of a number of mechanistic models of the money stock. The econometric models, however, appear to be structurally more stable than the mechanistic models. These results reflect the strong time trend exhibited by the money stock data.

Saunders, P., editor. Selected Assessment Strategies Applied to Short-Term Energy Models. Washington, DC: National Bureau of Standards, NBSIR 83-2672, 1983, 153pp.

This report is one in a series focusing on the evaluation of complex mathematical models. The basic approach pursued in this document is patterned after an earlier analysis of the Department of Energy's Midterm Oil and Gas Supply Model (MOGSM). Several extensions of the earlier methodology are presented which assist the analyst in defining the degree to which certain evaluation activities are model dependent. The Department of Energy's Short Term Integrated Forecasting System (STIFS) was used as a vehicle for exercising the revised methodology. The technical content of the report is divided into three parts, reflecting three basic issues of model form, sensitivity and forecast performance. The first issue addressed relates to the structure of STIFS. It includes not only the mathematical assumptions implicit in the model but also data and software considerations. The approach to the second issue focuses on the measurement of climatological uncertainties and uses as its basis a Monte-Carlo experiment. The final issue deals with several techniques for evaluating the predictive performance of a model. Both classical statistical methods and an information theoretic approach are used to illustrate how such an analysis would be carried out in practice.

Singleton, F., S. Muthukrishman and R. Thompson. Verification/Validation of an Economic Process Model for Electric Power Generation in the United States. Palo Alto, CA: Electric Power Research Institute, EPRI RP 1055-1-3, 1979, 42pp.

The purpose of this paper is to report how a verification/validation effort was completed for an economic process model of electric power generation in the United States. Primary emphasis in effort was directed to replicating historical statistics in the 1965/1969 and 1970/1974 periods. This replication effort was accomplished by working toward meaningful replication first for the South Central and Middle Atlantic Regions and second for a 16-region delineation of the contiguous 48 states. Results of both phases show least-cost solutions of the model, as revised in the effort, replicate well statistics for important categories of fuel use and capacity investments. In 6 aggregated regions, the modeling solutions explained 97 percent of the variance in coal use in 1969 and 93 percent of the variance in coal use in 1974. Also, the capacity investment solutions of the model explained 70 percent and 84 percent of the variance in coal investments in 1969 and 1974, respectively. In addition, solutions of the model seem to match unfolding fuel use and capacity investment trends very well for the 1975/1985 period.

Su, V. "An Error Analysis of Econometric and Noneconometric Forecasts," American Economic Review, Vol. 68, No. 2, 1978, pp. 306-312.

In this study, the historical records of the Wharton quarterly forecast and the ASA-NBER survey forecast are analyzed. The complete record of the Wharton forecast includes at least two sets of forecasts for each quarter. Each forecast projects eight to ten quarters ahead. The premeeting forecast is usually made at the end of the first month of the quarter, two weeks before the Wharton quarterly meeting of its users. The post meeting forecast is made at the end of the second month of each quarter and is based on the forty-five day release data. The ASA-NBER survey forecast is a quarterly consensus forecast which was first released in December 1968. It consists of the forecasts of ten major economic variables for four or five quarters ahead made by forty to eighty economists that respond to the survey. The study finds that the ASA-NBER survey outperforms the Wharton model in predicting six (of the ten) variables in the first quarter. By the fourth quarter the ASA-NBER survey still forecasts better in four variables, but only significantly better in one variable. The conclusion is that the Wharton forecast becomes relatively better as the forecast period expands.

U.S. General Accounting Office. Review of the 1974 Project Independence Evaluation System. Washington, DC: U.S. GAO, OPA-76-20, 1976, 58pp.

The 1974 Project Independence Evaluation System is a complex evaluation, forecasting, and analysis system. It was developed by the Federal Energy Administration to use in preparing the Project Independence report of 1974. This initial version of the Project Independence Evaluation System provided an innovative framework for evaluating energy policy, even though it contained limitations. The Federal Energy Administration is aware of these limitations and is implementing a plan for improving the system. These changes are essential in order for the Project Independence Evaluation System to approach its full usefulness.

U.S. General Accounting Office. An Evaluation of the Use of the Transfer Income Model -- TRIM -- to Analyze Welfare Programs. Washington, DC: U.S. GAO, PAD-78-14, 1977, 107pp.

Models are used within the Government to perform program and policy analysis of complex issues in such areas as social welfare, food, energy, and transportation. Models allow analysts and decision makers to address issues which are not readily susceptible to other analytical techniques. Models can be extremely useful, and, in some cases, they are indispensable for dealing with analytical problems. However, before being used, models should be evaluated carefully in order to assure that they are used properly and that any uncertainties in the results are identified. The Transfer Income Model is used widely throughout the Government to analyze a broad range of welfare programs. GAO's assessment of the model demonstrated how changes in assumptions and modifications to the model can be very useful in certain circumscribed areas but should not be used for other types of analysis.

Weisbin, C., and R. Alsmiller. "Highlights of the ORNL-Model Evaluation Methodology and the Evaluation of LEAP," in Symposium on Energy Modeling III: Dealing with Energy Uncertainty. Washington, DC: U.S. Department of Energy, 1980, 20pp.

This paper describes progress made at the Oak Ridge National Laboratory (ORNL) toward development and demonstration of a methodology for evaluating energy-economic modeling codes and important results derived from these codes. The goal of this work is to supplement traditional evaluation methods with more quantitative procedures which will be of interest to the Office of Validation Analysis of EIA. ORNL is applying sensitivity theory as part of a comprehensive effort to quantify the importance of various data and model parameters to key results of interest. Two types of sensitivity-analysis approaches, screening and adjoint perturbation theory, are discussed. Other essential program elements described in some detail include computational experience, assessment of model structure and rationale, and evaluation of the model input data. Preliminary findings from application of this approach to the Long-Term Energy Analysis Program (LEAP) used in obtaining projections (Model 22C) for the 1978 EIA Annual Report to Congress are presented.

Wood, D., and M. Mason. Analysis Quality Report on the EIA Annual Report to Congress 1978, Volume III: Coal Supply. Cambridge, MA: Massachusetts Institute of Technology, Cambridge Energy Laboratory, MIT-EL 81-017, 1981, 112pp.

The Energy Information Administration (EIA) is charged by Congress to prepare an Annual Report to Congress (ARC) which includes projections of energy supplies, consumption and prices, as well as the relation of energy to other economic activity. As an aid to users of ARC, the EIA Office of Analysis Oversight and Access (OAOA) is preparing "Analysis Quality Reports" on particular components of the energy information analysis system used in developing the ARC-78 projections. This report focuses on the Coal Supply Module used for the midterm projections of the ARC-78. The Coal Supply Module is part of the EIA's National Coal Model. The review and analysis presented here is based upon the MIT Energy Model Analysis Program's (EMAP) evaluation of the documentation and implementation of the Coal Supply Module sponsored by OAOA, and an in-depth evaluation of a related model--the ICF Coal and Electric Utilities Model--which also employs the EIA Coal Supply Module.

2. DATA BASE EVALUATION PRACTICES

Burnett, R., and J. Thomas. Transposed-File Structures and Data-Manipulation Support for Statistical-Data Editing and Subset Selection. Richland, WA: Battelle Pacific Northwest Laboratories, PNL-SA-9907, 1981, 12pp.

Statistical analysis of large data sets often requires an initial data editing and preparation phase to check the validity of individual data items, check for consistency among related data, correct erroneous data, and supply (impute) values for missing data where possible. During this preparatory phase of analysis, it is often necessary to partition the data set into a number of subsets by logical selection and/or random-sampling techniques for purposes of hypothesis testing. This paper examines the data-management support required by these editing and subsetting operations in terms of lower-level data-manipulation functions and mappings between logical and physical data structures. Advantages of transposed data files for statistical applications are discussed in comparison with record-based structures. A specific self-describing transposed-file design is described in detail, with emphasis on representations of logical data structures commonly encountered in statistical data bases.

Easton, M., and B. Bennet. "Transient-Free Working-Set Statistics," Communications of the ACM, Vol. 20 No. 2, 1977, pp. 93-99.

After defining the terms used, the authors derive two useful difference equations with appropriate initial conditions. The practical difficulties which are encountered when these equations are used with large data sets are discussed. After discussing a stochastic model, a statistical method for evaluating probable complexity is derived. This is a valuable contribution to data-base analysis.

Ferber, R., et al. "Validation of Consumer Financial Characteristics: Common Stock," Journal of the American Statistical Association, Vol. 64, No. 326, 1969, pp. 415-432.

This study of the accuracy with which people report holdings of common stock was carried out by the Bureau of the Census with institutional cooperation as a sequel to the Federal Reserve Board Survey of Financial Characteristics (SFC) of 1963 using the same interviewers and identical field and data processing procedures. The results indicate substantial nonreporting of stock ownership, which represents a major source of bias. Nonresponse is positively related to economic status, and holdings reported by the respondents were less than those of the nonrespondents. The size distributions of stock holdings are biased. Perhaps most important, estimates of variances and of confidence intervals computed by the usual standard error formulas are invalid as applied to mean holdings of the total population.

Fry, J., and A. Merten. Toward the Development of a Data Translation Methodology and Selection of Target Database Structures. Ann Arbor, Mich: University of Michigan, Department of Industrial and Operations Engineering, AFOSR-TR-77-0792, 1977, 23pp.

Progress has been made toward development of a data translation methodology and selection of optimal target data bases at the University of Michigan. A model for implementing data translators has been formulated and verified through the development of a series of increasingly general data translators. Mechanisms have been developed for prescribing logical data transformations--Restructuring Specifications--and physical data descriptions--a Stored-Data Definition Language. Finally, results have been obtained on the evaluation, selection, and optimization of target data base structures.

Johnson, T., R. Montalvo and J. Naus. Data Validation Study. Rockville, MD: Geomet, Inc., GEOMET-7104-01, 1970, 308pp.

Research was conducted and methods were developed for detecting and locating errors in large data systems. Errors were assumed to originate from any source including recording, transcription, man-made, machine-made, etc. Mathematical formalism was developed to evaluate probabilities calculated from the results of tests applied to the data. The tests are based on redundancies and known relationships in the data, as well as past history of the data. The logical implication of sets of failed tests are evaluated for various models including dependency and independency of error assumptions. Methods were developed for the construction of dictionaries of tests and procedures given to check the dictionaries for consistency, completeness and redundancies. The effectiveness of tests to detect and locate errors was defined for single tests and sets of tests giving a way to evaluate the testing procedure.

Kipnis, V., and I. Pinsker. "Forecasting Short Time Series Based on the Randomization Principle," in Models, Algorithms, Decision Making, Moscow, USSR: "Nauka", 1979, pp. 38-61.

The authors consider the problem of forecasting time series. They note the difficulties that arise in the solution of practical problems, especially in processing short series. They give estimates relating the accuracy of the prediction to the quality of the approximation and extrapolation of applicable models. They propose algorithms that make it possible to refine prediction models successively until a best estimate is obtained of the quality of the extrapolation. (In Russian)

Loebl, A., and S. Cantor. Energy-Data Validation: Overview and Some Concepts. Oak Ridge, TN: Oak Ridge National Laboratory, CONF-810842-7, 1981, 18pp.

Energy data validation can be viewed operationally as a three-fold assessment process: (1) a determination of the quality of the data collected, i.e., an assessment of accuracy; (2) an analysis of the relevance and usefulness of the data so as to assess how closely the data collected is meeting the requirements of its users; and (3) an assessment of measures that can be taken to enhance the effectiveness of the data system under study. Assessment (1) is akin to the process of critical evaluation of data in the physical sciences. Assessments (2) and (3), the more distinctive features, underlie two important goals of validation. In brief, the analyst validates both the information and the requirements for the information, and as deficiencies in these two aspects are uncovered, the validation analyst formulates and evaluates the means for correcting these deficiencies. This paper focuses upon the use of an error model for systematizing the assessment of accuracy in a data-system validation study.

Mendelson, H., and U. Yechiali. "Optimal Policies for Data Base Reorganization," Operations Research, Vol. 29, No. 1, 1981, pp. 23-36.

One of the problems faced by a data base administrator is the determination of a file reorganization policy. The data base administrator has to balance the benefits of a reorganization, which restores processing efficiency and hence reduces operating costs, against the extra expenditures involved (unloading and reloading the files, system unavailability costs, etc.). In this paper the authors consider both policies of state-dependent reorganization and reorganization at fixed time intervals. They prove that the optimal state-dependent policy belongs to the class of control-limit rules, and show how the optimal control-limit can be evaluated. They then consider the policy of reorganization at fixed time intervals, and show that, when the record arrival process is Poisson, this policy is always inferior to the optimal control-limit rule (for the average-cost criterion). They conclude with an example where operating and reorganization costs increase linearly with system occupancy.

Pack, D. Preliminary Internal Data Screening - a Component of Quantitative Data Analysis in Data Validation. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/CSD-85, 1981, 35pp.

Internal data screening is validation of data by quantitative internal comparisons across records for a given variable, across variables for a given record, across time for a given variable and record, etc. Preliminary internal data screening has the objectives of producing: (1) a frozen, archived data base ready for efficient computer analysis; and (2) a general understanding of key variables, a rough cut at response meaningfulness, a qualification of the type of error present; and an assessment of the expected level of difficulty in the comprehensive validation. The relatively less important first objective is defined. It is seen to be achievable through simple computer methodology. The second objective is the paper's major focus. Initial methodological suggestions for accomplishing this objective centering upon the exploratory data analysis techniques of Tukey are made, recognizing the appropriateness of these techniques for a general examination of a large data base that may contain numerous errors.

Smith, C., and J. Browne. "Performance Engineering of Software Systems: A Case Study," AFIPS Conference Proceedings, Vol. 51, 1982, pp. 217-224.

This paper summarizes the concepts of performance engineering in large software systems and illustrates the application of performance engineering techniques to the early design phase of a large data base system. Performance engineering is a methodology for evaluating the performance of software systems throughout their life cycles. The case study given here demonstrates that it is possible to predict resource usage patterns of complex software systems even in early design phases of the system, although detailed predictions of resource usage are not likely to be validated. The results presented show the leverage of considering performance implications in the early design phases of a software project.

University of Maryland. Computer Science and Technology: Modeling and Implementation of Database Management Software. College Park, MD: University of Maryland, 1979, 250pp.

This report describes a research effort that was carried out over a period of several years to develop and demonstrate a methodology for evaluating proposed data base management system (DBMS) designs. Included in the report is a review of a performance prediction system with the following components: (1) a prototype implementation of a positional set processor DMBS, a measurement facility; (2) discussion of the model concepts and parameters; (3) a summary of the mathematical relationships embedded in the performance prediction model; (4) a review of the model evaluation problem and the status of ongoing verification, validation, and problem analysis activities; and (5) research results and suggested future research directions.

3. SELECTED FEDERAL DOCUMENTATION GUIDELINES

3.1 FEDERAL INFORMATION PROCESSING STANDARDS (FIPS) DOCUMENTS

U.S. Department of Commerce. Code for Information Interchange (ASCII). Washington, DC: National Bureau of Standards, FIPS PUB 1-1, 1980.

This report provides a standard code, character set, and collating sequence for use in Federal information processing systems, communications systems, and associated equipment. It adopts ANSI X3.4-1977.

U.S. Department of Commerce. Dictionary for Information Processing. Washington, DC: National Bureau of Standards, FIPS PUB 11-1, September 1977.

This report provides a common reference within the government for terms and definitions used in such information processing activities as the representation, communication, interpretation and processing of data by human or automatic means. The dictionary consists of a single alphabetic listing of over 4000 terms and their definitions. It adopts ANSI x3/TR-1-82.

U.S. Department of Commerce. Guidelines for Describing Information Interchange Formats. Washington, DC: National Bureau of Standards, FIPS PUB 20, 1972.

Characteristics of formatted information, which must be considered for interchange of such information, are identified and described. The objective is to clarify and improve documentation for formatted information transfer. The guidelines describe physical and logical characteristics. A glossary of terms is also attached.

U.S. Department of Commerce. COBOL. Washington, DC: National Bureau of Standards, FIPS PUB 21-1, 1975.

This report establishes the form for and the interpretation of programs expressed in FIPS COBOL. It adopts ANSI X3.23-1974, except for the report writer module.

U.S. Department of Commerce. Flowchart Symbols and their Usage in Information Processing. Washington, DC: National Bureau of Standards, FIPS PUB 24, 1975.

Standard flowchart symbols and their use are specified. The standard is also known as ANSI X3.5-1972.

U.S. Department of Commerce. Software Summary for Describing Computer Programs and Automated Data Systems. Washington, DC: National Bureau of Standards, FIPS PUB 30, 1974.

A standard software summary form is defined (SF-185), together with instructions for describing computer programs for identification, reference, and dissemination. The form is used to record summaries of programs developed or acquired by Federal agencies, and by GSA to register selected Government software.

U.S. Department of Commerce. Guidelines for Documentation of Computer Programs and Automated Data Systems. Washington, DC: National Bureau of Standards, FIPS PUB 38, 1976.

Guidance is provided for determining the content and extent of 10 document types used in program and systems development, covering the "development phase" of the software life cycle. The document types cover: Functional Requirements, Data Requirements, System Specification, Data Base Specification, Users, Operations and Program Maintenance Manuals, Test Plan, and Test Analysis Report.

U.S. Department of Commerce. COBOL Coding Form. Washington, DC: National Bureau of Standards, FIPS PUB 44, 1976.

A standard COBOL coding form is provided (SF-268), with an explanation of its use and physical specifications. The form is used in coding of source programs, or as an input document in transcription of COBOL source programs to a medium acceptable to computer systems.

U.S. Department of Commerce. Transmittal Form for Describing Computer Magnetic Tape File Properties. Washington, DC: National Bureau of Standards, FIPS PUB 53, 1978.

A standard magnetic tape transmittal form is provided (SF-277) with instructions for its use. Physical properties and other characteristics are recorded, to permit a receiving agency to process the tape.

U.S. Department of Commerce. Guidelines for Documentation of Computer Programs and Automated Data Systems for the Initiation Phase. Washington, DC: National Bureau of Standards, FIPS PUB 64, 1979.

Guidelines are provided for determining the content and extent of documentation during the "Initiation Phase" of the software life cycle. The following document types are covered: Project Request, Feasibility Study, and Cost Benefit Study.

U.S. Department of Commerce. Minimal BASIC. Washington, DC: National Bureau of Standards, FIPS PUB 68, 1980.

This report defines the syntax of the Minimal BASIC programming language and the semantics for its interpretation. It adopts ANSI X3.60-1978.

U.S. Department of Commerce. FORTRAN. Washington, DC: National Bureau of Standards, FIPS PUB 69, 1980.

This report specifies the form and establishes the interpretation of programs expressed in the FORTRAN programming language. The standard consists of a full language, FORTRAN, and a subset language, Subset FORTRAN. It adopts ANSI X3.9-1978.

U.S. Department of Commerce. Guideline for Planning and Using a Data Dictionary System. Washington, DC: National Bureau of Standards, FIPS PUB 76, 1980.

This report describes the capabilities of a data dictionary system (DDS). Selection considerations and guidance for pre-implementation planning, implementation, and operational use of a DDS are also provided.

U.S. Department of Commerce. Guideline for Planning and Management of Database Applications. Washington, DC: National Bureau of Standards, FIPS PUB 77, 1980.

This report summarizes a recommended discipline of application management for database systems. It also provides specific advice on applications planning and management, and on software selection.

U.S. Department of Commerce. Guide for the Implementation of Federal Information Processing Standards (FIPS) in the Acquisition and Design of Computer Products and Services. Washington, DC: National Bureau of Standards, FIPS PUB 80, 1980.

This report provides basic information about FIPS publications 0-67 and identifies the computer products and services to which they may apply. The official applicability statement is contained in each FIPS publication. This Guide is intended to serve as a supplemental reference in the use of FIPS publications 0-67 only and does not cover FIPS publications 68 and above.

U.S. Department of Commerce. Guideline on Integrity Assurance and Control in Database Administration. Washington, DC: National Bureau of Standards, FIPS PUB 88, 1981.

This report provides explicit advice on achieving database integrity and security control. It documents a step-by-step procedure for examining and verifying the accuracy and completeness of a database.

3.2 INSTITUTE FOR COMPUTER SCIENCE AND TECHNOLOGY DOCUMENTS

Adrion, W., M. Branstad and J. Cheiniavsky. Validation, Verification, and Testing of Computer Software. Washington, DC: National Bureau of Standards, Special Publication 500-75, 1981, 62pp.

Programming is an exercise in problem solving. As with any problem solving activity, determination of the validity of the solution is part of the process. This survey discusses testing and analysis techniques that can be used to validate software and to instill confidence in the programming product. Verification throughout the development process is stressed. Specific tools and techniques are described.

Branstad, M., J. Chernlavsky and W. Adrion. Argini Citation Validation Verification, and Testing for the Individual Programmer. Washington, DC: National Bureau of Standards, Special Publication 500-56, 1980, 26pp.

Guidelines are given for program testing and verification to ensure quality software for the programmer working alone in a computing environment with limited resources. The emphasis is on verification as an integral part of the software development. Guidance includes developing and planning testing as well as the application of other verification techniques at each life-cycle stage. Relying upon neither automated tools nor formal quality assurance support, the guidelines should be appropriate for applications programmers doing small development projects.

Argini, J., J. Bowden and M. Skall. NBS Minimal BASIC Test Programs - Version 2 - User's Manual. Washington, DC: National Bureau of Standards, Special Publication 500-70, 1980, 79pp.

This publication describes the set of programs developed by NBS for the purpose of testing conformance of implementations of the computer language BASIC to the American National Standard for Minimal BASIC, ANSI X3.60-1978. The Department of Commerce has adopted this ANSI standard as Federal Information Processing Standard 68. By submitting the programs to a candidate implementation, the user can test the various features which an implementation must support in order to conform to the standard. While some programs can determine whether or not a given feature is correctly implemented, others produce output which the user must then interpret to some degree. This manual describes how the programs should be used so as to interpret correctly the results of the tests. Such interpretation depends strongly on a solid understanding of the conformance rules laid down in the standard, and there is a brief discussion of these rules and how they relate to the test programs and to the various ways in which the language may be implemented.

Gass, S. Computer Model Documentation: A Review and an Approach. Washington, DC: National Bureau of Standards, Special Publication 500-39, 1979.

Recent studies and surveys have concluded that, in general, the documents produced to support the understanding and use of computer models are inadequate. This paper describes the issues and concerns of computer model documentation and proposes an approach for the development of adequate documentation. First, a number of documentation studies and reports are reviewed, including software documentation guidelines and model documentation procedures. Then, based on the relationship between the phases of the model life cycle and documentation information needs, a set of documents is proposed and described. The author takes a highly critical view of the past and present inadequate state of documentation procedures for computer models. The attention of computer model sponsors and developers must be directed to this area. Otherwise, the author feels, there will be an unfortunate decline in the use of decision models as aids in the analysis of important policy issues. The course of action recommended in this report is an extreme position as to the total information and number of documents required to produce adequate documentation. The author calls for the capturing of all information generated during a model's life cycle. Further research is needed to adapt this extreme position to the realities of cost, resources, model complexity, and model use.

Krasny, M. Documentation of Computer Programs and Automated Data Systems. Washington, DC: National Bureau of Standards, Special Publication 500-15, 1977, 66pp.

This document is the proceedings of a symposium which was organized to introduce the Government ADP-community to the concepts of when and how to apply the government-wide guidelines of FIPS PUB 38, Guidelines for Documentation of Computer Programs and Automated Data Systems in developing both agency standards and operational documentation. These proceedings of the symposium summarize the question and answer sessions presented in three parallel sessions for management, operations, and staff attendees.

Martin, R., and W. Osborne. Guidance on Software Maintenance. Washington D.C.: National Bureau of Standards, Special Publication, 500-106, 1983, 74pp.

This report addresses issues and problems of software maintenance and suggests actions and procedures which can help software maintenance organizations meet a working definition for software maintenance. It also presents an overview of current problems and issues in that area. Tools and techniques that may be used to improve the control of software maintenance activities and the productivity of a software maintenance organization are discussed. Emphasis is placed on the need for strong, effective technical management control of the software maintenance process.

Neumann, A. Management Guide for Software Documentation. Washington, DC: National Bureau of Standards, Special Publication 500-87, 1982.

This guide is to assist managers in the establishment of policies and procedures for effective preparation, distribution, control, and maintenance of documentation which will aid in re-use, transfer, conversion, correction and enhancement of computer programs. Such documentation, together with the computer programs themselves, will provide software product packages which can be transferred and used by people other than the originators of the programs. "Software" and "documentation" are defined, some documentation problems are discussed, and policies, procedures, and applicable standards are outlined. Appendices provide checklists in support of documentation policies and procedures, and references to relevant guidelines, standards, and the literature. A glossary of terms is included.

Neumann, A. NBS FIPS Software Documentation. Washington, DC: National Bureau of Standards, Special Publication 500-94, 1982, 294pp.

These proceedings provide a record of papers and discussions presented at a workshop held on March 3, 1982, at the National Bureau of Standards. The meeting was sponsored by the NBS Institute for Computer Sciences and Technology. In addition to papers presented, the record also provides remarks by discussants and other participants. The workshop covered a variety of topics pertaining to software documentation. Topical sessions included: case studies of and reports on application of existing standards, documentation for operation and maintenance, tools for improved documentation, proposals for new documentation standards, enhancing software sharing, improving human interfaces, and quality assurance of documentation. Sixty-three papers were presented in parallel sessions, and a summary session concluded the meeting; over 300 persons participated in the workshop.

Patrick, R. Performance Assurance and Data Integrity Practices. Washington, DC: National Bureau of Standards, Special Publication 500-24, 1978.

This report identifies the approaches and techniques now practiced for detecting, and when possible, correcting malperformance as it occurs in computer information systems. This report is addressed to two audiences: to the systems designer using stock commercial hardware and software who is creating a system which will tax the available hardware, software, or staff skills; and to the manager who wishes to chronicle the deficiencies in an existing system prior to improvement. It enumerates 67 items of current practice which prevent computer malperformance.

Powell, P. Planning for Software Validation, Verification, and Testing. Washington, DC: National Bureau of Standards, Special Publication 500-98, 1982, 89pp.

Today, providing computer software involves greater cost and risk than providing computer equipment. One major reason is that hardware is mass-produced by proven technology, while software is still produced primarily by the craft of individual computer programmers. The document is for those who direct and those who implement computer projects; it explains the selection and use of validation, verification, and testing (V,V&T) tools and techniques for software development. A primary benefit of practicing V,V&T is increasing confidence in the quality of the software. The document explains how to develop a plan to meet specific software V,V&T goals.

Sockut, G., and R. Goldberg. Data Base Reorganization--Principles and Practice, Washington, DC: National Bureau of Standards, Special Publication 500-47, 1979.

Data base reorganization can be defined as changing some aspect of the way in which a data base is arranged logically and/or physically. This paper contains tutorials and surveys. It introduces the basic concepts of reorganization, including why it is performed. Many examples of types of reorganization are described and are classified into logical/physical levels. The paper then covers pragmatic issues such as reorganization strategies, a survey of several commercial reorganization facilities, case studies, and data base administration considerations. Finally, several research efforts are surveyed.

Powell, P. Software Validation, Verification, and Testing Technique and Tool Reference Guide. Washington, DC: National Bureau of Standards, Special Publication 500-93, 1982, 138pp.

Thirty techniques and tools for validation, verification, and testing (V,V&T) are described. Each description includes the basic features of the technique or tool, the input, the output, an example, an assessment of the effectiveness and usability, applicability, an estimate of the learning time and training, an estimate of needed resources, and references.

U.S. Department of Commerce. Computer Model Documentation Guide, Washington, DC: National Bureau of Standards, Special Publication 500-73, 1981.

This document provides guidelines for preparing documentation for computer models. Recommended structures for four types of manuals providing model information for four different classes of audiences (managers, users, analysts and programmers) are presented. This document specifies the content of sections and subsections for each type of manual. Manuals prepared using these guidelines will enable persons interested in a model to understand the capabilities and limitations of that model.

4. SENSITIVITY ANALYSIS

4.1 METHODOLOGY

4.1.1 Standard Techniques

Bar-Shalom, Y. "Effect of Uncertainties on the Control Performance of Linear Systems with Unknown Parameters and Trajectory Confidence Tube," in Proceedings of the 1977 IEEE Conference on Decision and Control, New York: IEEE, 1977, pp. 892-902.

This paper deals with the quantitative assessment of the effects of two sources of uncertainty on the quadratic performance index used for controlling linear econometric models. The uncertainties that are most important are the errors in the parameters' estimates and the additive noise entering the dynamic equations. Errors in measurements are not considered. An algorithm based upon the open-loop feedback concept is presented that evaluates the increase in the cost due to parameter errors. The parameters are modelled as random variables with known statistics up to second order, assumed not to change during the control period. This is equivalent to assuming that during the control period the quality of the parameters' estimates will remain essentially the same. These results can be used to assess the potential usefulness of adaptive control for a problem.

Deeter, C. and A. Hoffman. "Energy Related Mathematical Models: Annotated Bibliography," Energy Conservation Vol. 18, No. 4, 1978 pp. 189-227.

The evaluation of various options for energy sources, energy conversion, and efficient use of energy has led to the use of computer-based mathematical models. To ensure that these models are reasonably accurate representations of the situation being studied, it is desirable to have available information concerning the analysis of error propagation in mathematical models when implemented on computers. As an aid to energy model builders who seek methods of analysis for their own work, as well as to provide a helpful guide for the research computer scientists working on the development of new methods for analysis of propagation or error, a classification scheme of error propagation methods is presented and an annotated bibliography of energy-related mathematical models classified by validation and error analysis methods is provided.

Faldella, E. and G. Iuculano. "Automatic Evaluation of Regression Models," Alta Freq, 49, No. 1, 1980, pp. 31-34.

A generally applicable algorithm for the evaluation of regression models is described. The correctness of each contribution to the regression is verified according to statistical tests which provide a high level of significance. On the basis of the described algorithm a program has been developed for the automatic identification of the functional relationships characterizing the regression models. The program has been fruitfully exploited for a large computing system performance evaluation.

Hendrickson, R. A Survey of Sensitivity Analysis Methodology. Washington, DC: National Bureau of Standards, NBSIR 84-2814, 1984, 81pp.

This survey on the methodology of sensitivity analysis presents a general statement of the several broad categories of this discipline for the purpose of pulling together the various approaches and theory, to show the extent and sophistication of new techniques, special applications and the relation of sensitivity analysis to model evaluation.

Paulson, A., M. Presser and E. Nicklin, Self-Critical and Robust Procedures for the Analysis of Univariate Complete Data. Troy, NY: Renselaer Polytechnical Institute, A-5, 1982, 38pp.

A statistical sensitivity analysis may be defined and performed in terms of the response of a vector of parameter estimates to variation in the way sample information is processed vis-a-vis the tentative underlying model. The mode of information processing is a generalization of likelihood and is indexed on a non-statistical parameter c. The case c = 0 corresponds to maximum likelihood. If the vector of parameter estimates is stable under moderate increase of the index c from 0, the tentative model and the data are internally consistent. A general procedure for the conduct of such sensitivity analyses is given along with several illustrations. For fixed, positive values of the index, one obtains a general robust estimation procedure.

4.1.2 Specialized Techniques

Alsmiller, R., J. Barish, and D. Bjornstad. Interim Report on Model Evaluation Methodology and the Evaluation of LEAP Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/TM-7245, 1980, 223pp.

This report describes progress made at ORNL toward development and demonstration of a methodology for evaluating energy-economic modeling codes. Important results derived from these codes are also discussed. To bolster traditional evaluation methods with more-quantitative procedures of interest to the Energy Information Administration, ORNL is applying sensitivity theory as part of a comprehensive effort to quantify the importance of various data and model parameters to the key results that are of interest. The Long-Term Energy Analysis Program (LEAP) was chosen as the initial focus for the research. LEAP is an energy-economy model which resides in the Long-Term Energy Analysis Division (LTEAD) of the Integrative Analysis Group in the Office of Applied Analysis, EIA. LTEAD developed Model 22C of LEAP for two reasons: (1) to prepare projections through the year 2020, which were needed for the 1978 EIA Annual Report to Congress and (2) to develop a base for analyses of specific options for Federal action.

Alsmiller, R., J. Barish, J. Drischler, J. Horwedel, J. Lucius, and J. McAdoo. Adjoint Sensitivity Theory and Its Application to LEAP Model 22C. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/TM-7789, 1981, 118pp.

Adjoint sensitivity theory can be used to determine the sensitivity of results of interest to each of the data elements that enter into the calculations. In this paper adjoint sensitivity theory is discussed, and its applicability to a large energy-economics model is demonstrated by applying it to a specific calculation carried out with the Long-Term Energy Analysis Program (LEAP). Numerical results for dR/dx, where R is the result of interest and x is any one of the data elements, are presented for all x for which dR/dx is appreciable and for several definitions of R. In a number of cases the accuracy of the dR/dx obtained by adjoint methods has been verified by direct calculations; these comparisons are also presented. The application of the theory as presented requires extensive development work in that a large amount of analytic differentiation must be carried out. In the course of the work, a method was developed that would allow all of the required derivatives to be obtained numerically using LEAP; this method is presented and discussed. Unfortunately, this numerical method is applicable only to codes with the very special modular structure of LEAP.

Harris, C. Issues in Sensitivity and Statistical Analysis of Large-Scale, Computer Based Models. National Bureau of Standards, Gaithersburg, MD: NBS GCR 84-466,1984, 40 pp.

An understanding of the uncertainty associated with the predictions of a large-scale model can improve their usefulness in policy analysis. But this has been done in relatively few computer-based modeling efforts to date. It is critical for users to have confidence in their model's outputs and therefore important that the model builders carefylly quantify potential variability in results. In this work, we examine the primary statistical methods for understanding possible randomness in large-scale-model prediction. We do not provide an exhaustive survey of mathematical methods for sensitivity analysis but rather focus on the statistical aspects of measuring uncertainty in model output and on the prediction of outcomes from unexpected changes in key model inputs.

Kramer, M., J. Calo and H. Rabitz. "An Improved Computational Method for Sensitivity Analysis: Green's Function Method with Aim," Applied Mathematical Modeling, Vol. 5, No. 6, 1981, pp. 432-441.

Sensitivity analysis of initial value models via the calculation of linear sensitivity coefficients is quite important for model evaluation and validation. Direct solution of the sensitivity equations for N-dimensional, M-parameter systems of ordinary differential equations requires the solution of M*N differential equations, which can become quite expensive for large-scale models. When M>N, the Green's Function Method (GFM), which requires solutions of N^2 differential equations with M*N subsequent numerical quadratures, is the most efficient computational technique for determining linear sensitivity coefficients. An algorithm, known as the analytical integrated magnus (AIM) modification of the GFM, is presented which dramatically reduces the computational effort required to determine linear sensitivity coefficients. The technique employs the piecewise magnus method for more efficient calculation of Green's Function Kernels, and treats the sensitivity integrals analytically. An application of this technique to a chemical kinetics system is presented in which the computational effort is reduced by an order of magnitude in comparison to the unmodified GFM.

Mucciardi, A., and E. Gose. "An Automatic Clustering Algorithm and Its Properties in High-Dimensional Spaces," in IEEE Transactions on Systems, Man, and Cybernetics, Vol. smc-2, No. 2, 1972, pp. 247-254.

An economical technique for approximating a joint N-dimensional probability density function has been described by Sebestyen and Edie. The algorithm searches for clusters of points and considers each cluster as one hyperelliposoidal cell in an N-dimensional histogram. Cell descriptors (location, shape, and size) can be determined adaptively from sequentially introduced data samples of known classification. The number of cells required for a good fit can usually be kept small. No assumptions are required about the underlying statistical structure of the data. The paper presents heuristics which were developed to automate the selection of the control parameters. The properties of these parameters were studied as a function of dimension. Two of the control parameters were found to be linearly related to dimension. This provides a method for determining their value by extrapolation, thereby avoiding a great deal of computation.

Porter, W. "Sensitivity Problems in Distributive Systems," International Journal of Control, Vol. 5, No. 5, 1967, pp. 393-412.

The effects of parameter variations on the external characteristics of distributive systems are considered. It is shown that this class of problems can be brought within the framework of modern system sensitivity analysis. The interaction between feedback and system sensitivity is analyzed.

4.2 CASE STUDIES

Boshier, J. and F. Schweppe. "Energy Model Validation: Systematic Sensitivity Analysis" in Proceedings of Lawrence Symposium on Systems and Decision Sciences, North Hollywood, CA: Western Periodicals Company, 1977, pp. 266-270.

This paper briefly reviews the formation of a "third-party" energy assessment laboratory. A discussion of the central role of sensitivity analysis validation is given. The evaluation of the regional electricity model, a large simulation model of the US electricity sector, is discussed.

Greenway, A., H. Ellis and R. Deland. Description, Evaluation, and Sensitivity Analyses of Principal US EPA Air Quality Prediction Models. Washington, DC: U.S. Department of Energy, ANL/EES-TM-111, 1980, 124pp.

Scientific validity of the principal assumptions used in the U.S. Environmental Protection Agency (EPA) air pollution prediction models was reviewed. The computational assumptions and equations used in the principal EPA models were reviewed, as was the recommended applicability of these models and their performance as reported in validation and comparison studies. In addition, a sensitivity analysis of model response to input parameter changes was conducted. The performance of the CRSTER, the Urban and Rural RAM Models, and by inference the MPTER Model of the UNAMAP series was reviewed and evaluated based on available studies of the performance of these models. It is concluded that the RAM (Urban) Model tends to overpredict the impact of sources with tall stacks, even in urban areas, due to the treatment of unstable cases. The RAM (Urban) Model implicitly accounts for building-effect downwash through enhanced plume spreading rates.

Weisbin, C., R. Peelle and R. Alsmiller. "Assessment of the Long-Term Energy Analysis Program Used For The EIA 1978 Report to Congress," Energy, Vol. 7, No. 2, 1982, pp. 155-170.

This paper evaluates the Long-Term Energy Analysis Program (LEAP), a computer model of the energy portion of the U.S. economy that was used by the Energy Information Administration (EIA) of the U.S. Department of Energy (DOE) for the 1995-2020 projections in its 1978 Annual Report to Congress. Essential considerations include; (1) the adequacy of its documentation; (2) the evaluator's experience in operating the model; (3) the adequacy of the numerical techniques used; (4) the soundness of the economic and technical foundations of the model equations; and (5) the degree to which the computer program has been verified. To show which parameters strongly influence the results and to approach the question of whether the model can project important results with sufficient accuracy to support qualitative conclusions, the numerical sensitivities of some important results to model input parameters were obtained using a novel adjoint technique. This technique enabled efficient determination of first derivatives of results of interest with respect to all input data (approximately 1600 non-zero sensitivities). Direct recalculation of the model with altered data sets was used to confirm the results and to assess the extent of non-linear effects. The input data are categorized and discussed, and uncertainties are given for some parameters as examples.

5. TECHNIQUES FOR HANDLING SPECIAL CASES

5.1 RARE EVENTS

Apostolakis, G. Mathematical Methods of Probabilistic Safety Analysis. Washington, DC: Research Applied to National Needs, NSF-RA-X-74-047, 1974, 317pp.

This report presents the mathematical methods useful in reliability and safety studies. The problems that these studies deal with concern: (1) the prediction of the probability that a specified function will be performed satisfactorily over a period of time or per demand; and (2) the identification of events and their probabilities, which may lead to unfavorable circumstances endangering the health of the public. This document includes summaries of illustrative applications as well as references to the literature. Chapter 2 describes the methods of handling problems involving one component or simple logical configurations--reliability theory. An introduction to the fundamentals of probability theory and statistics is given. The modeling of the failure of components by statistical distributions is discussed, followed by the mathematical description of various maintenance policies. Chapter 3 deals with analysis of complex systems. Fault-tree methodology is developed in detail, and its uses and limitations are investigated. Methods like failure modes and effects analysis are described. Special problems arising from software and human errors as well as the possibility of common mode failures are also discussed. Finally, the use of statistical techniques to handle major natural phenomena and methods of dealing with systems without exploiting their logical structure are investigated.

Apostolakis, G., and A. Mosleh. Use of Expert Opinion in the Evaluation of Probabilities of Rare Events. Los Angeles, California: UCLA, Department of Chemical, Nuclear, and Thermal Engineering, UCLA-ENG-7834, 1978, 88pp.

The evaluation of probabilities of rare events in the context of the subjectivistic theory of probability is the subject of this work. Several useful subsidiary methods for the achievement of coherent probabilities are discussed with particular emphasis on experts' opinions. As an application, a distribution for the frequency of reactor core melts is assessed using the available experience with power reactors. This distribution is subsequently modified, via Bayes' theorem, to include the estimate derived by WASH-1400, which is treated as an expert's opinion. A probabilistic model for the credibility of this opinion is presented, which includes the critics' points of view. The resulting (posterior) distribution is the assessed distribution of the frequency of reactor core melts based on a body of knowledge which includes the available experience and the WASH-1400 estimate.

Bryan, J., and T. Johnson. Statistical Prediction of Rare Events. Hartford, CN: Travelers Research Center, Inc., 7673-199, 1965, 65pp.

Research on statistical prediction of rare weather events was concentrated on the discovery and utilization of effective joint predictors, since extensive tests of additive models revealed a tendency for predictive accuracy to deteriorate for more severe weather as compared to that achieved in more common and clement weather. Procedures were sought to: (1) identify variables having predictive potential; (2) detect joint predictive information of a synergistic type, bound up in pairs of variables; (3) represent single and paired variables in forms suitable for the reconstruction of an unknown prediction function; and (4) concentrate their predictive information so that in subsequent predictor screening the risk of overlooking valuable contributors is mitigated. Appendix A is a test for joint effects of two variables, and Appendix B is a computer program to evaluate the test.

Lee, Y., and G. Apostolakis. Probability Intervals for the Top Event Unavailability of Fault Trees. Los Angeles, CA: UCLA, School of Engineering and Applied Science, UCLA-ENG-7663, 1976, 102pp.

The evaluation of probabilities of rare events is of major importance in the quantitative assessment of the risk from large technological systems. In particular, for nuclear power plants the complexity of the systems, their high reliability and the lack of significant statistical records have led to the extensive use of logic diagrams in the estimation of low probabilities. The estimation of probability intervals for the probability of existence of the top event of a fault tree is the subject of this work. Given the uncertainties of the primary input data, a method is described for the evaluation of the first four moments of the top event occurrence probability. These moments are then used to estimate confidence bounds by several approaches which are based on standard inequalities (e.g., Tchebycheff, Cantelli, etc.) or on empirical distributions (the Johnson family). Several examples indicate that the Johnson family of distributions yields results which are in good agreement with those produced by Monte Carlo simulation.

Martin, R. "Robust Methods for Time Series," in Applied Time Series Analysis, New York: Academic Press, 1981, pp. 683-759.

Outliers in time series can wreak havoc with conventional least-squares procedures, just as in the case of ordinary regression. This paper presents two time-series outlier models, points out their ordinary regression analogues and the corresponding outlier patterns, and presents robust alternatives to the least-squares method of fitting autoregressive moving average models. The main emphasis is on robust estimation in the presence of additive outliers. This results in the problem having an errors-in-variables aspect. While several methods of robust estimation for this problem are presented, the most attractive approach is an approximate non-Gaussian maximum-likelihood type method which involves the use of a robust nonlinear fitter/one-sided interpolator with data-dependent scaling. Robust smoothing/two-sided outlier interpolation, forecasting, model selection, and spectral analysis are briefly mentioned, as are the problems of estimating location and dealing with trends, seasonality, and missing data. Some examples of applying the methodology are given.

Morlat, G. Decision Theory Applied to Rare Events. Fontenay-aux-Roses, France: CEA Centre D'Etudes Nucleaires de Fontenay-aux-Roses, CONF-7606115-2, July 1976, 12pp.

The history of statistics reveals diverging viewpoints concerning the concrete interpretation of the mathematical subject termed "probability." Indeed, the frequency of rareness of the phenomena involved contribute strong elements of justification to the selection of a particular attitude. The theory of the decision intended by some of its promoters (Abraham Wald) to reinforce the bases for conventional statistics appears to have achieved an opposite result, at the same time quite strongly justifying the Bayes' methods (Leonard J. Savage). This can result in removing a strong reluctance to use probabilistic and "statistical" methods in the case of rare events. Nevertheless, a close examination of the concepts and axioms of the theory reveals requirements which are likely to be difficult to satisfy within the very definition of rare events. It appears that a collective effort for better satisfying these requirements would be extremely desirable. It could contribute to the improvement of the present situation by providing a better consensus concerning an estimate of the risks.

Parry, G., and P. Winter. Characterization and Evaluation of Uncertainty in Probabilistic Risk Analysis. Culcheth, England: UKAEA Risley Nuclear Power Development Establishment, Safety and Reliability Directorate, SRD-R-190, 1980, 35pp.

The sources of uncertainty in probabilistic risk analysis are discussed using the event/fault tree methodology as an example. The role of statistics in quantifying these uncertainties is investigated. A class of uncertainties is identified which is, at present, unquantifiable, using either classical or Bayesian statistics. It is argued that Bayesian statistics is the more appropriate vehicle for the probabilistic analysis of rare events and a short review is given with some discussion on the representation of ignorance.

Sampson, A., and R. Smith. Assessing Risks Through the Determination of Rare Event Probabilities. Pittsburgh, PA: Pittsburgh University, Department of Mathematics and Statistics, TR-80-9, 1980, 47pp.

The authors consider the problem in risk assessment of evaluating the probability of occurrence of rare, but potentially catastrophic, events. The lack of historical data due to the sheer novelty of the event makes conventional statistical approaches inappropriate. The problem is compounded by the complex multivariate dependencies that may exist across potential event sites. In order to evaluate the likelihood of one or more such catastrophic events occurring, they provide an information theoretic model for merging a decision maker's opinion with expert judgment. Also provided is a methodology for the reconciling of conflicting expert judgments. This merging approach is invariant to the decision maker's viewpoint in the limiting case of exceptionally rare events. These methods are applied to case studies in likelihood assessment of liquid natural gas tanker spills and seismic induced light water nuclear reactor meltdowns.

Thompson, W. Observations on Risk Analysis. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/CSD-45, 1979, 26pp.

This paper briefly describes WASH 1400 and the Lewis report. It attempts to define basic concepts such as risk and risk analysis, common mode failure, and rare event. Several probabilistic models which go beyond the WASH 1400 methodology are introduced; the common characteristic of these models is that they recognize explicitly that risk analysis is time dependent whereas WASH 1400 takes a per demand failure rate approach which obscures the important fact that accidents are time related. Further, the presentation of a realistic risk analysis should recognize that there are various risks which compete with one another for the lives of the individuals at risk. A way of doing this is suggested.

Uppuluri, V. Rare Events: A State of the Art. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/CSD-73, 1980, 51pp.

The study of rare events has become increasingly important in the context of nuclear safety. Some philosophical considerations, such as the framework for the definition of a rare event, rare events and science, rare events and trans-science, and rare events and public perception, are discussed. The technical work of the task force on problems of rare events in the reliability analysis of nuclear plants (1976-1978), sponsored by OECD, is reviewed. Some recent technical considerations are discussed, and conclusions are drawn. The appendix contains an essay written by A. Beachey, entitled: "A Study of Rare Events -- Problems and Promises."

Uppuluri, V. Risk Analysis and Reliability. Oak Ridge, TN: Union Carbide Corporation, Nuclear Division, CONF-791016-1, 1979, 12pp.

Mathematical foundations of risk analysis are addressed. The importance of having the same probability space in order to compare different experiments is pointed out. Then the following topics are discussed: consequences as random variables with infinite expectations; the phenomenon of rare events; series-parallel systems and different kinds of randomness that could be imposed on such systems; and the problem of consensus of estimates of expert opinion.

Uppuluri, V., and S. Patil. Inferences About Rare Events. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL/CSD-12, 1976, 21pp.

A rare event is defined in a qualitative manner as an event that can happen with a small probability. From the known relationship between the waiting time and the frequency of an event of interest, the "conjugacy" between the geometric distribution and the binomial distribution is deduced. This conjugacy is exploited to obtain the confidence limits on p, the probability of a rare event (success), when no successes are observed in n independent Bernoulli trials. In order to resolve the waiting time paradox that could arise in the case of this pair of conjugate distributions, the distribution of the interarrival time, relative to an aribtrary but fixed point in time, is obtained.

Zavarina, M., and V. Klingo. Estimating Limiting Values of the Probability of Rare Events. Wright-Patterson Air Force Base, Ohio, Foreign Technology Division, FTD-HT-23-1056-72, 1972, 8pp.

Confidence intervals of the probability of the appearance of rare events are estimated as a function of the scale of the sample by which the probability is determined. A table of confidence intervals for actually encountered frequencies at a confidence probability of 99 percent is given. It is noted from the table that true probability of the event (and the average period of its recurrence) may differ several times from that calculated from the sample of frequency of recurrence. For example, if the probability of some meteorological element, calculated over a 30 year period of observations, is 10 percent (average period of recurrence corresponding to ten years), the maximum true value of probability will be equal to 31 percent, and the minimum, to 1.1 percent. Consequently, the minimum possible average period for recurrence will be three years, and the maximum will be 100 years.

5.2 MISSING DATA

Andres, F., H. Ruderman and J. Sathaye. Dynamic Programming Approach to Estimating Gaps in Large Economic Data Bases, Berkeley, CA: Lawrence Berkeley Laboratory, LBL-4271, 1975, 20pp.

Many precise mathematical models of social systems have been developed. However, precise input data from the economic census is usually not available. Unfortunately, as the precision of the data increases so does the amount of missing data. First, and most commonly, known data is "withheld to avoid disclosing figures for individual companies." The number of withheld items becomes significant when the economy is divided into about three hundred or more industry groups for regions the size of states or smaller. This degree of disaggregation is often necessary to capture important interactions. Secondly, a major reason for missing data is that the same size for some categories is insufficient for reasonable confidence in the data. However, there is usually enough other information about the missing data to constrain it and to estimate a probability distribution for it within the constraints. A suboptimal dynamic programming algorithm is proposed to find a maximum likelihood estimator of the missing data items.

Bowling, S., and S. Lai. The Use of Linear Prediction for the Interpolation and Extrapolation of Missing Data and Data Gaps Prior to Spectral Analysis. Lexington, MA: Massachusetts Institute of Technology, Lincoln Laboratory, TN-1979-46, 1979, 25pp.

The spectral analysis of a series of equally spaced samples of a time-stationary process becomes difficult when samples are missing or sizable data gaps occur within the interval of interest. A linear prediction algorithm can be used to fill in the missing data with estimates that are spectrally consistent with the data that are observed. Simulated and practical radar examples demonstrate an improvement in resolution and a reduction of sidelobe interference levels. Computer programs are provided which accomplish the extrapolation and interpolation for complex data.

Brailovsky, V. "On the Influence of Missing Data in a Sample Set on the Quality of a Statistical Decision Rule." in Third International Conference on Collective Phenomena, New York: New York Academy of Science, 1980, pp. 124-138.

This paper deals with the case where an approximation function must be obtained with the help of a sample set lacking some values. The usual least mean-squares method (LMS) is compared with the author's so-called LMS-like method. The special case of a model of the two-category classifier is considered, and the effect of missing values in the sample set on the quality of a decision rule is investigated. The author also considers the situation in which one adds new samples to a sample set and thus improves the statistical estimations of the unknown parameters; at the same time, the quality of the decision rule becomes worse.

Burnett, R., and J. Thomas. Transposed-File Structures and Data-Manipulation Support for Statistical-Data Editing and Subset Selection, Richard, WA: Battelle Pacific Northwest Labs, PNL-SA-9907, 1981 12pp.

Statistical analysis of large data sets often requires an initial data editing and preparation phase to check the validity of individual data items, check for consistency among related data, correct erroneous data, and supply (impute) values for missing data where possible. During this preparatory phase of analysis, it is often necessary to partition the data set into a number of subsets by logical selection and/or random-sampling techniques for purposes of hypothesis testing. This paper examines the data-management support required by these editing and subsetting operations in terms of lower-level data-manipulation functions and mappings between logical and physical data structures. Advantages of transposed data files for statistical applications are discussed in comparison with record-based structures. A specific self-describing transposed-file design is described in detail, with emphasis on representations of logical data structures commonly encountered in statistical data bases.

Drummond, J., and F. Lawrence. A Least Squares Approach to Missing Meteorological Data. Las Cruces, NM: New Mexico State University, Physical Science Laboratory, ASL-CR-82-0008-1, 1982, 47pp.

In battlefield situations it is not uncommon to lose all or a portion of the meteorological data needed for artillery ballistic purposes due to malfunction of equipment or premature balloon bursting. This report describes several techniques to provide for these missing data. Further, a preliminary comparison of these techniques is made.

Frane, J. "Some Simple Procedures for Handling Missing Data in Multivariate Analysis," Psychometrika, Vol. 41, No. 3, 1976, pp. 409-415.

For analyses with missing data, some popular procedures delete cases with missing values, perform analysis with "missing value" correlation or covariance matrices, or estimate missing values by sample means. There are objections to each of these procedures. Several procedures are outlined here for replacing missing values by regression values obtained in various ways, and for adjusting coefficients (such as factor score coefficients) when data are missing. None of the procedures are complex or expensive.

Gourieroux, C., and A. Monfort. "On the Problem of Missing Data in Linear Models," Review of Economic Studies, Vol. 48, No. 4, 1981, pp. 579-586.

The problem of estimating the parameters of econometric models when some observations are missing has been treated along two different lines. In the first category of methods the missing data are considered as functions of additional parameters. In the second approach the missing data are considered as random variables on which some assumptions are made.

Horst, P. The Missing Data Matrix. Seattle, WA: Washington University, AD-655 916, 1967.

The report deals with a multiple regression approach to the estimation of missing elements in a data matrix. Three types of missing data matrices are discussed and methods for their analysis are presented. Computational equations together with mathematical proofs are included.

Kastenberg, W., T. McKone and D. Okrent. On Risk Assessment in the Absence of Complete Data. Washington, DC: Research Applied to National Needs, NSF/RA-760325, 1976, 41pp.

A major problem in assessing risks in advanced systems is that: (1) all of the potential consequences are not known nor anticipated; and (2) even when risks are anticipated there is uncertainty in the data. Several mathematical risk assessment theories are summarized in this report, and their deficiencies in the absence of complete data are discussed. The mathematical basis of risk assessment is considered, and the basic problems of moving from a discrete to a continuous consequence-frequency distribution are discussed. The application of these concepts to waste disposal and storage of high level radioactive waste, risk to ground populations due to airline crashes, and nuclear power plant risks are reviewed and discussed.

Jones, R. "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," Technometrics, Vol. 22, No. 3, 1980, pp. 389-395.

Akaikes' Markovian m-dimensional representation of an ARMA time series where m=max(p, q{ubar}) is used to formulate a collection of recursive Kalman estimation procedures that can be used to evaluate the associated likelihood function of the time series if it is also invertible and stationary. The approach can be used even if missing data or observation error are present. This approach is then applied to determine, by likelihood criteria, the "best fit" ARMA mode of the Palmer wetness index constructed from tree ring data.

Little, R., and D. Rubin. On Jointly Estimating Parameters and Missing Data Complete-Data Likelihood. Madison, WS: University of Wisconsin, Mathematics Research Center, MRC-TSR-2326, 1982, 10pp.

One approach to handling incomplete data occasionally encountered in the literature is to treat the missing data as parameters and to maximize the complete data likelihood over missing data and parameters. This paper points out that although this approach can be useful in particular problems, it is not a generally reliable approach to the analysis of incomplete data. In particular, it does not share the optimal properties of maximum likelihood estimation, except under the trivial asymptotics in which the proportion of missing data goes to zero as the sample size increases.

Morlat, G. On Decision Theory Applied to Rare Events. Washington, DC: Nuclear Regulatory Commission, NR-TR-003, 1976, 16pp.

Decision-making theory in the case of rare events is, by definition of the word rare, difficult, when approached from the viewpoint of conventional statistical methods: data analysis or conventional inferential statistics. Promoters of conventional statistical arguments seem, in fact, to have reinforced the argument for probabilistic approach using Bayes methods. A close examination of the concepts of various decision-making theories reveals that a combined effort could help to achieve a consensus among those favoring different methods of estimating risks.

Murray, C. Calculation of Power Spectrums from Digital Time Series with Missing Data Points. Greenbelt, MD: National Aeronautics and Space Administration, NASA-TM-82016, 1980, 66pp.

Two algorithms are developed for calculating power spectrums from the autocorrelation function when there are missing data points in the time series. Both methods use an average sampling interval to compute lagged products. One method, the correlation function power spectrum, takes the discrete Fourier transform of the lagged products directly to obtain the spectrum, while the other, the modified Blackman-Tukey power spectrum, takes the Fourier transform of the mean lagged products. Both techniques require fewer calculations than other procedures since only 50% to 80% of the maximum lags need be calculated. The algorithms are compared with the Fourier transform power spectrum and two least squares procedures (all for an arbitrary data spacing). Examples are given showing recovery of frequency components from simulated periodic data where portions of the time series are missing, and random noise has been added to both the time points and to values of the function. In addition the methods are compared using real data. All procedures performed equally well in detecting periodicities in the data.

Neave, H. Spectral Analysis of a Stationary Time-Series Using Initially Scarce Data. Madison, WS: Wisconsin University, Department of Statistics, M TR-152, 1968, 82pp.

The purpose of this paper is to introduce and study methods of estimating the spectrum of a stationary time series when given "initially scarce" data, i.e., data where the sampling period is shortened at some point during the period of observation. Two types of estimates are considered, the "basic" estimates, so-called because they derive from general theory which can essentially be applied to most missing data situations, and "alias-improved" estimates, whose use unfortunately cannot be extended to many other situations. The asymptotic properties of these estimates are studied in a forthcoming paper. The estimates are compared here in finite situations by Monte Carlo methods, and the alias-improved estimates turn out to be much superior.

Nuttall, A. Spectral Analysis of a Univariate Process with Bad Data Points, via Maximum Entropy and Linear Predictive Techniques. Newport, RI: Naval Underwater Systems Center, NUSC-TR-5303, 1976, 128pp.

A comparison of several methods for spectral estimation of a univariate process with equi-spaced samples, including maximum entropy, linear predictive, and autoregressive techniques, is made. The comparison is conducted via simulation for situations both with and without bad (or missing) data points. The case of bad data points required extensions of existing techniques in the literature and is documented fully here in the form of processing equations and FORTRAN programs. It is concluded that the maximum entropy (Burg) technique is as good as any of the methods considered, for the univariate case. The methods considered are particularly advantageous for short data segments. This report also reviews several available techniques for spectral analysis under different states of knowledge and presents the interrelationships of the various approaches in a consistent notation. Hopefully, this non-rigorous presentation will clarify this method of spectral analysis for readers who are nonexpert in the field.

Onukogu, I. "Another Alternative to Least Squares for Estimating Missing Values in Balanced Incomplete Block Designs (BIB)," Metron, Vol. 34, No. 1-2, 1978, pp. 181-186.

It is shown that the problem of estimating missing data in a BIB may be treated by changing coordinates in the estimation space from the (unknown) parameters of the model to the (known) observed values. The missing observation is subsequently expressed as unbiased estimates, i.e., linear combinations of the new sets of coordinates. The (pooled) average of such independent estimates gives the desired result. The primary advantage of the proposed method over the method of least squares is that such problems as (a) handling large matrices and (b) handling ill-conditioned (i.e., singular) matrices, which can occur in the latter, are avoided in the former.

Pezoldt, V. Proceedings of a Workshop on Rare Event/Accident Research Methodology. Washington, D.C.: National Bureau of Standards, Special Publication 482, 1977.

The volume contains the formal papers presented at a Workshop on Rare Event/Accident Research Methodology sponsored by the Human Factors Section of the Center for Consumer Product Technology, National Bureau of Standards held at NBS May 26-28, 1976. The topics addressed at the workshop and reflected in the papers in this volume include system safety engineering, hypothesis generation in accident research, epidemiological approaches to injury research, observational techniques for studying complex tasks, accident simulation, and methodological considerations being forced by the law.

Robbins, J. Ozone Data and Mission Sampling Analysis. Washington, DC: National Aeronautics and Space Administration, NASA-CR-159342, 1980, 156pp.

A metholodogy was developed to analyze discrete data obtained from the global distribution of ozone. Statistical analysis techniques were applied to describe the distribution of data variance in terms of empirical orthogonal functions and components of spherical harmonic models. The effects of uneven data distribution and missing data were considered. Data fill based on the autocorrelation structure of the data is described. Computer coding of the analysis techniques is included.

Strong, P., C. Swain and M. Swain. DOVE, A Rational Analysis of Sparse Data. Cambridge, MA: Massachusetts Institute of Technology, Department of Chemistry, TR-1, 1977, 23pp.

Realistic parameters are attainable in spite of missing data. DOVE can be useful, even when many or most data are missing, for generalized least squares fitting to evaluate a self-consistent set of all parameters in an expression for predicting all missing data, and without changing the predicted data, to transform the set of parameters obtained in phase 1 so that each final parameter has a simple, pure, realistic, physical meaning. Since predicted data are expressed as $a(i)x(j)+b(i)y(j)+...+c(i)$ with n product terms, phase 2 requires incorporation of $n^2 + n$ independent subsidiary conditions, of which 2n are arbitrary, i.e., merely fix zero reference points and scale unit sizes, but n^2-n are critical, i.e., must be relationships between particular parameters supported by other information. Both phases are illustrated by a two-mode application with 41 parameters, to fit the data plus 6 subsidiary conditions. Valid parameters are obtained although 30 of the 70 possible data are missing.

Titterington, D., and J. Jiang. Recursive Estimation Procedures for Missing-Data Problems. Madison, WS: Wisconsin University, Mathematics Research Center MRC-TSR-2427, 1982, 26pp.

Titterington in a previous paper proposed recursive methods for dealing with incomplete data. This present paper concentrates on versions of these for multiparameter problems involving missing data. Theorems are outlined from which asymptotic properties of the recursive procedures can be established and versions of the recursions are written down for problems in which the missing data are missing at random. After illustration with exponential family models, the case of multivariate Normal data is considered in detail. Numerical comparisons of the various methods are obtained using bivariate Normal data. Whereas the previous paper discussed incomplete data in general, the present one restricts attention to the problem of missing values. Typically, each experimental unit should have records of the values of several characteristics associated with it. Statistical analysis is made difficult if one or more of these values are missing on some units. To combat the heavy analysis required for a "proper" analysis of the data, comparatively simple recursive procedures are outlined in which the data are incorporated sequentially into the estimation scheme. Some comments are made about theoretical properties and special emphasis is laid on the case of the data from multivariate Normal distributions.

Titterington, D. "Analysis of Incomplete Multivariate Binary Data by the Kernal Method," Biometrika, Vol. 64, No. 3, 1977, pp. 455-460.

The procedure, due to J. Aitchison and C. G. G. Aitken, of analyzing multivariate binary data by kernel methods is extended to deal with missing data problems. The performance of the techniques is discussed for univariate, bivariate and for specific multivariate data. Comparisons are drawn with other approaches.

5.3 ERROR PROPAGATION

Goel, N. "Error Propagation and Catastrophes in Protein Synthesizing Machinery," in Nonlinear Systems and Applications, New York: Academic Press, 1977, pp. 125-145.

It should be pointed out that the word "catastrophes" in the title has no direct relation to the catastrophe theory of R. Thom. The subject of this paper is how the protein synthesizing machinery in cells may produce errors in the primary sequences of amino acids. Proteins with errors may lead to the production of other proteins with more errors. This process of increase of errors may continue and finally lead to "an error catastrophe" in protein synthesis and then death of the cell. In this paper various models of error propagation are discussed, as well as conditions under which errors may increase (leading to an error catastrophe) or decrease.

O'Neill, R., and R. Gardner. "Sources of Uncertainty in Ecological Models" in Symposium on Systems Methodology and Simulation, Rehonot, Israel, 1978.

Variance in model output is the result of a number of interacting causes. For purposes of the present review, the authors have grouped the sources under three headings: model structure, parameter error, and natural variability in ecological systems. Uncertainty associated with model structure results from constructing a simple mathematical model to describe a complex natural system. It is clear that the model can only represent limited aspects of the system behavior. Results are reviewed for error resulting from representing a two-component system by a single-variable model; for relative error propagation in alternative models for the same ecological system; and for error propagation in series of models of increasing structural complexity. Errors in the measurement of model parameters represent a second major source of uncertainty. Monte Carlo studies are reviewed on the sensitivity of variance on model outputs to variance on individual parameters, the effects of the distribution of parameter values on the distribution of output, and the propagation of parameter errors in a time-varying model and its implications for gathering validation data. Model construction and parameter measurement are errors that are introduced by the researcher. Ecological systems also contain uncertainty due to environmental (e.g., meteorological) variability, genetic variability within populations, and spatial heterogeneity. These sources of intrinsic variability also lead to errors if they are ignored in model development.

Serth, R., T. Hughes, R. Opferkuch and E. Eimutis. Source Assessment: Analysis of Uncertainty--Principles and Applications, Dayton, OH: Monsanto Research Corporation, MRC-DA-632, 1978, 181pp.

This report provides the results of a study that was conducted to analyze the uncertainties involved in the calculation of the decision parameters used in the Source Assessment Program and to determine the effect of these uncertainties on the decision-making procedure. A general procedure for performing an analysis of uncertainty is developed based on the principles of error propagation and statistical inference. It is shown that this simple and straightforward method represents an approximation to standard statistical techniques. Guidelines are established for precision in field sampling and analytical work, and for setting critical values of decision parameters.

Uppuluri, V. and W. Kuo, Survey of Error Propagation in Systems. Oak Ridge, TN: Oak Ridge National Laboratory, ORNL-CSD/TM-190, 1983, 22pp.

Error propagation analysis in reliable systems has been studied widely. Unlike classical sensitivity analysis which investigates the range of system performance, the distribution function of system performance studies error propagation analysis. Thus, error propagation analysis is essentially a statistical analysis. This paper reviews and classifies current research articles in error propagation. An overview is presented of error propagation applied to various systems and models. A standard analysis procedure is also given. Finally, several conclusions are drawn. Error propagation analysis is important for expensive or rare-event systems. This report can benefit those who analyze these systems.

5.4 STABILITY

Leadbetter, M. On Certain Results for Stationary Point Processes and Their Application. Chapel Hill, NC: North Carolina University, Department of Statistics, SER-641, 1969, 24pp.

The purpose of this report is to describe some viewpoints and certain results of point process theory, and their application. Section 2 contains series expressions for the distribution of the times between events both when measurement is made from a fixed time, and from an "arbitrary event" of the process. Relations between these series are also discussed. (These results were obtained in collaboration with R. J. Serfling.) Section 3 concerns applications, particularly to zero-crossings problems. In Section 4, the method of defining a stationary point process by means of a stationary sequence of non-negative random variables (due to Kaplan) is discussed and related to the previously described framework. Section 5 deals with the approximation of stationary point processes consisting of rare events by Poisson processes.

Paulre, B. System Dynamics and the Analysis of Change. New York: North-Holland Publishing Co., 1981, 387pp.

This document is a proceedings of the Sixth International Conference on System Dynamics held at the University of Paris IX, Paris, November 5-7, 1980. The document is divided into two parts: (1) current applications; and (2) methodology and techniques. Selected papers which appear in the proceedings are: R. J. Rahn, System dynamics and catastrophe theory; Maryse Martin, System dynamics and catastrophe theory; E. Renguet and D. M. Dubois, A stochastic approach of the catastrophe theory; and Patrick J. Starr and Jean-Pierre Pouplard, Identifying critical parameters in system dynamics models.

Tahir, S. Breakdown of Predictability: An Investigation on the Nature of Singularities, Trieste, Italy: International Centre for Theoretical Physics, IC-80/192, 1980, 18pp.

When relations are extrapolated beyond their premises of discovery, the operation sometimes results in an undefined object, i.e., one which cannot be identified within the given structure. The thesis is put forth that the occurrence of singularities is due to "incompleteness" in knowledge. An intuitive answer on how to deal with singularities (in, for instance, the real number system, space-time, quantum field theory) is presented first. Then a quasi-formalistic approach, e.g. non-standard models in higher-order languages and Lawvere's axiomatic formulation of categories, is set out. The independence of singularity with respect to other primitive notions of the universe of knowledge is noted.

Vanleeuwen, H. Methods for the Interpolation and the Extrapolation of the Results of Creep Rupture Tests. Amsterdam, Netherlands: National Aerospace Laboratory, NLR-MP-79007-U, 1979, 71pp.

A survey of extrapolation methods, graphic analysis, statistics and mathematical models relevant to creep rupture is presented. Minimal commitment and discrete function values, and creep diagrams are also discussed. (in Dutch)

APPENDIX A: ADDITIONAL REFERENCES OF POTENTIAL INTEREST

A.1 MODEL AND DATA BASE THEORY

Ben-Dov, Y. Optimal State Detection Policies for Coherent Systems. Berkley, CA: Operations Research Center, University of California, ORC-77-30, 1977, 103pp.

The problem of minimizing the expected cost of identifying the state of a coherent system (as functioning or failed) is considered. The system is composed of components, and only individual components can be tested. Efficient algorithms are presented for some special cases of coherent systems: series, parallel, parallel-series, series-parallel and k-out-of-n systems. The concept of the Importance of Components is used to develop a branch and bound algorithm which determines the optimal testing procedure for any general coherent system. However this algorithm is not always efficient in solving this problem. Some other closely related problems are discussed, such as how to identify the state of a system which is represented as a fault tree, or how to identify the failed components when the system is known to be failed.

Ben-Dov, Y. Optimal Testing Procedures for Coherent Systems. Berkeley, CA: Operations Research Center, University of California, ORC-77-23, 1977, 26pp.

This report deals with the problem of minimizing the expected cost of testing a coherent system. The concept of the Importance of Components is used to develop a branch and bound algorithm which determines the optimal testing policy for any coherent system.

Bracchi G., and G. M. Nijssen, Editors. Data Base Architecture. New York: North-Holland Publishing Company, 1979, 341pp.

These proceedings of the IFIP Working Conference on Data Base Architecture address the research and development problems related with the organization of data base systems and their interfaces. This collection of papers offers a comprehensive and penetrating analysis of the current fundamental issues in data base architecture. Among these are design and implementation aspects of advanced data base system organizations, design solutions for distributed data bases, access control, integrity, recovery techniques, data models, and high-level languages.

Brandon, C., R. Fritz and J. Xander. "Econometric Forecasts: Evaluation and Revision," Applied Economics, Vol. 15, No. 2, 1983, pp. 187-201.

Burmeister, E., and L. Klein, eds. Econometric Model Performance. Philadelphia: University of Pennsylvania, 1976.

Chow, G. "Evaluation of Macroeconomic Policies by Stochastic Control Techniques," International Economic Review, Vol. 19, No. 2, 1978, pp. 311-319.

Christ, C. "Judging the Performance of Econometric Models of the U.S. Economy," International Economic Review, Vol. 16, 1975, pp. 54-74.

Cicarelli J. "A New Method of Evaluating the Accuracy of Economic Forecasts," Journal of Macroeconomics, Vol. 4, 1982, pp. 469-475.

DBS Corporation, Development of Software for Computer Assisted Model Simplification. Washington, DC: U.S. Department of Energy, Energy Information Administration, DOE/EIA/10597-T1, 1980, 48pp.

This document is the final report of the DBS Corporation on the model project to develop computer-assisted model simplification. The contributions of DBS to this project were an initial overall project assessment, contributions to design principles and testing procedures, specific experimental designs, and initial test results. The main contributions of DBS to this project were in the area of LP matrix scaling, and particularly in the potential usefulness of shadow price information for model simplification. An algorithm for obtaining approximate shadow price information was developed and subjected to initial small-scale testing with promising results.

Dutta, J. "On Predictive Evaluation of Econometric Models," International Economic Review, Vol. 21, No. 2, 1980, pp. 379-390.

Fair, R. Estimating the Expected Predictive Accuracy of Econometric Models, Cowles Foundation Discussion Paper #480, Yale University, 1978.

Fujiwara, K. "On the Verification of Causal Systems-An Investigation of Simon's Method," Osaka Economic Papers, Vol. 1, 1980, pp. 89-102.

Galanc, T., and J. Mikus. "On Some Methods of Verification of Normative Forecasts," Ekonomicko-Matematicky Obzor, Vol. 12, No. 2, 1976, pp. 205-210.

Gallaire, H., Editor. Advances in Data Base Theory, New York: Plenum Press, 1981, 432pp.

Hatanaka, M. "A Simple Suggestion to Improve the Mincer-Zarnowitz Criterion for the Evaluation of Forecasts," Annals of Economic and Social Measurement, Vol. 3, No. 3, 1974, pp. 521-524.

Herbert, J. "Model Estimation and Verification-Some Recent Approaches," Empirical Economics, Vol. 4, No. 2, 1979, pp. 87-99.

Hymans, S. "Criteria for Evaluation of Econometric Models: A Correction," Annals of Economic and Social Measurement, Vol. 5, No. 1, 1976, pp. 161-162.

Hurley, K. "A Guide to Evaluating Forecasts," Business Economics, Vol. 11, No. 4, 1976, pp. 40-44.

Jayatissa, W. "Criteria for Evaluation of Econometric Models: A Correction," Annals of Economic and Social Measurement, Vol. 5, No. 1, 1976, pp. 161-162.

Lucas, R. "Econometric Policy Evaluation: A Critique," in K. Brunner and A. Meltzer, Carnegie-Rochester Conference Series on Public Policy, Vol. 1, New York: North Holland, 1976, pp. 19-46.

Martin, R. "The Illusion of Uncritical Evaluations of Econometrics," Business Economics, Vol. 16, No. 4, 1981, pp. 53-56.

McCarthy, P. "The Use of Balanced Half-Sample Replication in Cross-Validation Studies," Journal of the American Statistical Association, Vol. 71, No. 355, 1976, pp. 596-604.

McDonald, J. and J. Darroch. "Large Sample Estimation and Testing Procedures for Dynamic Equation Systems: Comment," Journal of Econometrics, Vol. 17, No. 1, 1981, pp. 127-130.

McNees, S. "An Evaluation of Economic Forecasts: Extension and Update," New England Economic Review, 1976, pp. 30-44.

Palm, F., and A. Zellner. "Large Estimation and Testing Procedures for Dynamic Equation Systems: Rejoinder," Journal of Econometrics, Vol. 17, No. 1, 1981, pp. 131-138.

Shapiro, H. "Is Verification Possible? The Evaluation of Large Econometric Models," American Journal of Agricultural Economics, Vol. 55, No. 2, 1973, pp. 250-258.

Stekler, H. "Evaluation of Econometric Inventory Forecasts," Review of Economics and Statistics, Vol. 51, No. 1, 1969, pp. 77-83.

Vogt, W. G., and M. H. Mickle, Editors. Modeling and Simulation. Research Triangle Park, NC: Instrument Society of America, 1981, pp. 1223-1644.

A.2 MODEL AND DATA BASE VERIFICATION

Benedict, B. "Effects of Equation Averaging on Model Verification," in Verification of Mathematical and Physical Models in Hydraulic Engineering, College Park, MD: University of Maryland, 1978.

Butterfield, M. "A Method of Model Verification for Large Perturbation Transients With Examples from Nuclear-Power," in International Conference on Control and its Applications, Coventry, England: University of Warwick, 1981, pp. 196-201.

Chrostowski, J., D. Evensen and T. Hasselman. "Model Verification of Mixed Dynamic-Systems," Journal of Mechanical Design: Transactions of the ASME, Vol. 100, No. 2, 1978, pp. 266-273.

Ebbesen, L., J. Perrault H. Sebesta and K. Lawrence, "Simulation, Modeling, and Verification for Large Systems with Random Disturbances," in Proceedings: 1979 Joint Automatic Control Conference, New York: American Institute of Chemical Engineers, 1979, pp. 559-563.

Hasselmann, K. "Construction and Verification of Stochastic Climate Models," in Climatic Variations and Variability: Facts and Theories, Hingham, MA: Kluwer Boston, 1981, pp. 481-497.

Kuo, C., and C. Blair. "Comparison of Verified Physical and Mathematical-Model," in Verification of Mathematical and Physical Models in Hydraulic Engineering, College Park, MD: University of Maryland, 1978.

Meegan, D., and D. Sprevak, "A Flexible Data Splitting Algorithm for Model Verification," in Compstat 1980; Proceedings in Computational Statistics, Vienna, Austria: Physica-Verlag Wien, 1980, pp. 297-303.

Sanden, B. "Verification of a Program Complex," BIT, Vol. 19, No. 2, 1979, pp. 244-255.

For verification of a complex of programs interacting in a complicated way, the programs can be described as processes in a state space. Properties to be verified can be expressed in a similar manner. In this paper such a method is applied to the restart facility of an online data base system.

Sargent, R. "Verification and Validation of Simulation-Models," in Progress in Modeling and Simulation, London, England: Academic Press, 1982, pp. 159-169.

Shannon, R. "Tests for the Verification and Validation of Computer-Simulation Models," in 1981 Winter Simulation Conference Proceedings, New York: IEEE, 1981, pp. 573-577.

Sheppard, S. "Verification of Simulation-Models," in Modeling and Simulation, Research Triangle Park, NC: Instrument Society of America, 1980, pp. 695-699.

A.3 MODEL AND DATA BASE VALIDATION

Angus, J., J. Schafer and A. Sukert. "Software Reliability Model Validation," in Proceedings Annual Reliability and Maintainability Symposium, 1980, pp. 191-199.

Balci, O., and R. Sargent. "Some Examples of Simulation-Model Validation Using Hypothesis-Testing," in 1982 Winter Simulation Conference Proceedings, New York: IEEE, 1982, pp. 620-629.

Balci, O., and R. Sargent. "A Methodology for Cost-Risk Analysis in the Statistical Validation of Simulation-Models," Communications of the ACM, Vol. 24, No.4, 1981, pp. 190-197.

Bell, C. "The Testing and Validation of Models," in Mathematics and Plant Physiology, London, England: Academic Press, 1981, pp. 299-309.

Bubenko, J. Third International Conference on Very Large Data Bases, New York: IEEE, 1977.

Colonna, G. "Hazard Assessment Computer-System (HACS) - Models Review and Validation," in 1982 Hazardous Material Spills Conference Proceedings, Rockville, MD: Government Institutes, 1982, pp. 414-421.

Cornelius, J., J. Ikerd and A. Nelson. "A Preliminary Evaluation of Price Forecasting Performance by Agricultural Economists," American Journal of Agricultural Economics, Vol. 63, No. 4, 1981, pp. 712-714.

Curtis, D.M. and S.G.T. Lafontaine, "Model Validation -Its Purpose, Value and Results," in Solar World Forum, Elmsford, NY: Pergamon Press, 1982, pp. 2659-2663.

Diwan, R. "Projections of World Demand for and Supply of Foodgrains: An Attempt at Methodological Evaluation," World Development, Vol. 5, Nos. 5-7, 1977, pp. 497-506.

Donigian, A.S., "Model Predictions vs Field Observations - The Model Validation Testing Process," in Fate of Chemicals in the Environment; Compartmental and Multimedia Models for Predictions, Washington, DC: American Chemical Society, 1983, pp. 151-171.

Ferber, R., et al. "Validation of a National Survey of Consumer Financial Characteristics: Savings Accounts," Review of Economics and Statistics, Vol. 51, No. 4, 1969, pp. 436-444.

Folley, L.J., and R.C. Williges, "Validation of User Models for Interactive Editing," in Proceedings of the Human Factors Society 26th Annual Meeting, Santa Monica, CA: Human Factors Society, 1982, pp. 616-620.

Gass. S.I. "Validation and Assessment Issues of Energy-Models," in Energy Policy Planning, New York: Plenum Publishing Corporation, 1981, pp. 421-441.
Ghali, M., M. Akiyama and J. Fujiwara. "Models of Regional Growth: An Empirical Evaluation," Regional Science and Urban Economics, Vol. 11, No. 2, 1981, pp. 175-190.

Griner, G.M., and J.A. Mango, "Validation of Simulation-Models Using Statistical Techniques," in Proceedings of the 1978 Summer Computer Simulation Conference, Arlington, VA:, AFIPS Press, 1978.

Keller, A.M., and G. Wiederhold, "Validation of Updates Against the Structural Data base Model," in Proceedings: Symposium on Reliability in Distributed Software and Database Systems, Piscataway, NJ: Computer Society Press, 1981, pp. 195-199.

Kenward, L. "Forecasting Quarterly Business Expenditure on Non-Residential Construction in Canada: An Assessment of Alternative Models," Canadian Journal of Economics, Vol. 9, No. 3, 1976, pp. 517-529.

Kost, W. "Model Validation and the Net Trade Model," Agricultural Economics Research, Vol. 32, No. 2, 1980, pp. 1-16.

Lalwani, C., and J. Brown, "Model Validation - View of Modeling in Social Sciences," in Proceedings of the 1978 Summer Computer Simulation Conference, Arlington, VA: AFIPS Press, 1978.

Leuthold, R. "On the Misuse of Theil's Inequality Coefficient: Model Validation and the Net Trade Model. The Food and Agricultural Policy Simulator," Agricultural Economics Research, Vol. 34, No. 4, 1982, p.39.

Ljung, L. "Model Validation," in Identification and System Parameter Estimation 1982, Elmsford, NY: Pergamon Press, 1983, pp. 73-75.

Ljung, L., and A. Vanoverbeek. "Validation of Approximate Models Obtained from Prediction Error Identification," in Link Between Science and Applications of Automatic Control, Elmsford, NY: Pergamon Press, 1979, pp. 1899-1906.

McNees, S. "An Assessment of the Council of Economic Advisors' Forecast of 1977," New England Economic Review, 1977, pp. 3-7.

McSorley, R., J. Ferris and V. Ferris, "Model Synthesis and Validation - Primary Consumers," in Nematicde in Soil Ecosystems, Austin, TX: University of Texas Press, 1982, pp. 141-156.

Obermayer, R., and S. Bowser, "Issues in Validation of Large-Scale Simulation Models," in Modeling and Simulation, Research Triangle Park, NC: Instrument Society of America, 1979, pp. 2061-2065.

Rodriguez, G. and D. Kunkel. "Model Validation and the Philippine Programming Model," Agricultural Economics Research, Vol. 32, No. 2, 1980, pp. 17-25.

Salaway, G., and K. Siler. "Validation Considerations During Parameter Development for Simulation Models," in Proceedings of the 1979 Summer Computer Simulation Conference, Arlington, VA: AFIPS Press, 1979, pp. 624-628.

Sargent, R. "Verification and Validation of Simulation Models," in Progress in Modeling and Simulation, London, England: Academic Press, 1982, pp. 159-169.

Sargent, R. "Model Validation -The Current State," in 10th IMACS World Congress, Montreal, Canada: International Association Mathematics Computers Simulation, 1982, pp. 112-114.

Schwartzkopf, S. "An Elementary Method for the Diagnostic Validation of Simulation-Models," in Proceedings of the 1980 Summer Computer Simulation Conference, Arlington, VA:, AFIPS Press, 1980, pp. 236-239.

Shannon, R. "Tests for the Verification of Computer-Simulation Models," in 1981 Winter Simulation Conference Proceedings, New York: IEEE, 1981, pp. 573-577.

Sommer, M. "Some Controversies on Predictive and Non-Predictive Validation Strategies for Dynamic, Socioeconomic Models," in Proceedings of the 1981 UKSC Conference on Computer Simulation, Woburn, MA: Ann Arbor Science Publishers, 1981, pp. 307-315.

Spall, J., and K. Wall. "On Hypothesis-Testing and Model Validation Employing State Estimation," in Proceedings: 1982 IEEE International Large Scale Systems Symposium, New York: IEEE, 1982, pp. 418-422.

Steinhorst, R. "Parameter Identifiability, Validation, and Sensitivity Analysis of Large System Models," in Systems Analysis of Ecosystems, Burtonsville, MD: International Co-operative Publisher House, 1979, pp. 33-58.

Tendolkar, N. "Application of Simulation in Validation of Algorithms of an Analytic Model of a Computer System," in Tenth Annual Simulation Symposium, Long Beach, CA: IEEE Computer Society Publications, 1977.

Thompson, J. "Example of Data-Poor Model Validation," in Decision Information, New York: Academic Press, 1979, pp. 405-408.

Thrall, R. "Cross-Validation of Models - How Can Industry, Government and Academics Check on Each Other," in Energy Modeling II: The Interface Between Model Building and Decision Maker, Chicago, IL: Institute for Gas Technology, 1980, pp. 281-288.

Wall, D, "Specification and Estimation of Econometric Models with Generalized Expectations," in Lecture Notes in Control and Information Science, Berlin: Springer, 1979, pp. 34-48.

Ward, R. "Approach to Validation Through Evolutionary Model-Building," in Proceedings of the 1977 Summer Computer Simulation Conference, La Jolla, CA: Simulation Council, 1977.

Wittenmark, B., and B. Barshalom, "Model Validation from Estimated Closed-Loop Performance," in Identification and System Parameter Estimations, Elmsford, NY: Pergamon Press, 1980, pp. 1103-1109.

A.4 DOCUMENTATION GUIDELINES

American Nuclear Society. Guidelines for the Documentation of Digital Computer Programs. Hinsdale, IL: American Nuclear Society, N431 (ANS-10.3), 1973.

Bevan, R. Using the Naval Surface Weapons Center (Dahlgren Laboratory) Computer Assisted Documentation (CAD) Technique. Dahlgren, VA: Naval Surface Weapons Center, 1977.

Brewer, G. "Documentation: An Overview and Design Strategy," Simulation and Games, Vol. 7, No. 3, 1976.

Enger, N.L., Documentation Standards for Computer Systems. Fairfax Station, VA: The Technology Press, 1976.

Gass, S., K. Hoffman, R. Jackson, L. Joel and P. Saunders, "Documentation for a Model: A Hierarchial Approach," Communications of the ACM, Vol. 24, No. 11, 1981, pp. 728-733.

Lleine, H., Software Design and Documentation Language. Pasadena, CA: Jet Propulsion Laboratory, California Institute of Technology, JPL 77-24, 1977.

London, K., Documentation Standards. New York: Petrocelli Books, 1974.

Nance, R., The Feasibility of and Methodology for Developing Federal Documentation Standards for Simulation Models. Blacksburg, VA: Department of Computer Science, VPI and State University, 1977.

National Aeronautics and Space Administration. Computer Program Documentation Guidelines, Washington, DC: NASA, NHB-2411.1, 1971.

Newton, O., and J. Weatherbee. "Guidelines for Documenting Computer Simulation Models," in Proceedings of the 1980 Winter Simulation Conference, New York: IEEE, 1980.

Sayani, H., et al. Automated Documentation System User's Manual. Ann Arbor, MI: Department of Industrial and Operations Engineering, University of Michigan, ISDOS Working Paper No. 123, 1975.

Schoeffel, W. An Air Force Guide to Software Documentation Requirements, Bedford, MA: Mitre Corporation, 1976.

Sibley, E., and H. Sayani. "Data Element Dictionaries for the Information Systems Interface," in Proceedings of the First National Symposium on Management of Data Elements in Information Processing, Washington, DC: National Bureau of Standards, 1974.

Ubico, R. Documentation Standards for the Strategic Environmental Assessment System (SEAS), Rockville, MD: Control Data Corporation, 1973.

U.S. Air Force. Computer Program Documentation Guidelines. Washington, DC: Assistant Chief of Staff, Studies and Analyses, Headquarters, USAF, 1976.
U.S. Air Force, Computer Program Documentation. Eglin Air Force Base, FL: Headquarters Armament Development Test Center, ADA 01-3, 1977.

U.S. Atomic Energy Commission. Management Information Systems--Systems Documentation Handbook. Washington, DC: Division of Management Information and Telecommunications Systems, U.S. AEC, 1972.

U.S. Department of Defense. Automated Data System Documentation Standards Manual. Washington, DC: U.S. DOD, 4120.17-M, 1972.

U.S. Department of Labor. ADP Documentation Guidelines. Washington, DC: U.S. DOL, DLMS-9, Data Processing Chapter 400, 1976.

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

Wright, M., Documentation of Computer Library Routines, Stanford, CA: Systems Optimization Laboratory, Department of Operations Research, 1976.

A.5 SENSITIVITY AND UNCERTAINTY ANALYSIS

Ahamad, B., and K. Scott. "A Note on Sensitivity Analysis in Manpower Forecasting," Journal of Royal Statistical Society, Series A, Part 3, Vol. 135, 1972, pp. 385-392.

Atherton, R. "Toward a Methodology for Model-Sensitivity Analysis," in 11th Asilomar Conference on Circuits, Systems and Computers, New York: IEEE, 1978.

Bronfenbrenner, M. "Sensitivity Analysis for Econometricians," Nebraska Journal of Economics and Business, Vol. 11, No. 4, 1972, pp. 57-66.

Cukier, R., H. Levine and K. Shuler. "Nonlinear Sensitivity Analysis of Multiparameter Model Systems," Journal of Physical Chemistry, Vol. 81, No. 25, 1977, pp. 2365-3266.

Iman, R. "Statistical-Methods for Uncertainty Analysis and Sensitivity Analysis Associated with Computer-Models," in Waste Isolation in the US and Elsewhere: Technical Programs and Public Communications, Grange, IL: American Nuclear Society, 1982, pp. 243-255.

Lee, M., and R. Ernst, "Criterion Performance-Model Considerations for Decision Systems-A Sensitivity Analysis," in Proceedings: International Conference on Cybernetics and Society, New York: IEEE, 1981, pp. 644-648.

Martin, D. "Functional Sensitivity Analysis of Mathematical-Models," in System Modelling and Optimization, New York: Springer-Verlag, 1982, pp. 663-670.

McKay, M. "Sensitivity Analysis and a National Energy-Model Example," in Proceedings of the 1979 DOE Statistical Symposium, Springfield, VA: NTIS, 1980, pp. 139-145.

Melendez, K., and A. Linder. "Sensitivity Analysis and Response Surface of a Simulation-Model of a Computer System," in Computer Science and Technology: Computer Performance Evaluation Users Group (CPEUG), Washington, DC: National Bureau of Standards, 1978.

Mihilov, V. "Sensitivity Analysis of Uncertainty in Seismic Sources Modeling on Seismic Hazard Mapped Parameters," in Proceedings of the 2nd U.S. National Conference on Earthquake Engineering, Berkeley, CA: Earthquake Engineering Research Institute, 1979, pp. 783-792.

Oblow, E., C. Weisbin, and R. Alsmiller. "Sensitivity Analysis for Energy-Economy Modeling," in Energy Modeling II: The Interface Between Model Builder and Decision Maker, Chicago, IL: Institute Gas Technology, 1980, pp. 425-435.

Schreiber, B. "A Different Approach to Sensitivity Analysis and Controllability of Dynamic-Models," in Proceedings of the 1980 Summer Computer Simulation Conference, Arlington, VA: AFIPS Press, 1980, pp. 116-121.

Srivastava, L. "Stability Analysis and Model Tests of Dam Foundations Subjected to Earthquake Loads," in Sixth Symposium on Earthquake Engineering, Nauchandi, India: Sarita Prakashan, 1978, pp. 175-179.

Steinhorst, R. "Parameter Identifiability, Validation, and Sensitivity Analysis of Large System Models," in Systems Analysis of Ecosystems, Burtonsville, MD: International Co-operative Publishing House, 1979, pp. 22-58.

A.6 EXTREME AND RARE EVENTS

Bendell, A., and W. Samson, "The Use of Rank Order Distributions for the Estimation of the Probabilities of Rare Events," in Proceedings: Third National Reliability Conference, Warrington, England: National Centre for Systems Reliability, 1981, pp. 43-53.

Bland, W. "TMI Rare Event - Lessons to be Learned," in 1982 Proceedings: Annual Reliability and Maintainability Symposium, New York: IEEE, 1982, pp. 17-20.

Busch, A., Colamosca, J. Hunter and N. Polhemus. "Assessing the Safety and Risk of Air-Traffic Control-Systems - Risk Estimation from Rare Events," in Aviation Forecasting and Systems Analyses, Washington, DC, National Academy of Sciences, ISBN 0-309-03112, 1980, pp. 24-28.

Lathrop, J. "Planning for Rare Events - Nuclear Accident Preparedness and Management - Overview," in Planning for Rare Events: Nuclear Accident Preparedness and Management, Elmsford, NY: Pergamon Press, 1981, pp. 3-12.

Rosbjerg, D. "Analysis of Extreme Events in Stationary Dependent Series," in Inputs for Risk Analysis in Water Systems, Littleton, Co.: Water Resources Publications, 1979, pp. 69-75.

Rowe, W. "Estimating the Probability of Rare Events -Approaches and Limitations," in Modeling and Simulation, Research Triangle Park, NC, 1980, pp. 517-528.

Watson, I. "The Rare Event Dilemma and Common Cause Failures," in 1982 Proceedings: Annual Reliability and Maintainability Symposium, New York: IEEE, 1982, pp. 5-10.

A.7 MISSING DATA

Anderson, O., J. Degooijer and K.D.C. Stoodley. "The Mean-Square Error of One-Step-Ahead Prediction for an Autoregressive Process When the Coefficients are Estimated with Missing Data," Time Series Analysis: Theory and Practice, New York: Elsvier North-Holland, 1982, pp. 53-63.

Brailovsky, V. "On Multivarate Linear-Regression with Missing Data Among the Independent Variables," in Fourth International Conference on Collective Phenomena, New York: New York Academy of Sciences, 1981, pp. 128-137.

Brailovsky, V. "On the Influence of Missing Data in a Sample Set on the Quality of a Statistical Decision Rule," in Third International Conference on Collective Phenomena, New York: New York Academy of Sciences, 1980, pp. 124-138.

Dahiya, R., and R. Korwar. "Maximum Likelihood Estimates for a Bivariate Normal Distribution with Missing Data," Annals of Statistics, Vol. 8, No. 3, 1980, pp. 687-692.

Elashoff, J., and R. Elashoff. "Two-sample Problems for a Dichotomous Variable with Missing Data," Journal of the Royal Statistical Society, Series C, Vol. 23, 1974, pp. 26-34.

Kleinbaum, D. "A Generalization of the Growth Curve Model which Allows Missing Data," Journal of Multivariate Analysis, Vol. 3, 1973, pp. 117-124.

Morrison, D. "A Test for Equality of Means of Correlated Variates with Missing Data on One Response," Biometrika, 1973, pp. 101-105.

Papaioannou, T., and T. Speevak. "Rank Correlation Inequalities with Missing Data," Communications in Statistics, A6, No. 1, 1977, pp. 67-72.

Pearce, S., and J. Jeffers. "Block Designs and Missing Data," in Journal of the Royal Statistical Society, Series B, Vol. 33, No. 1, 1971 pp. 131-136.

Rubin, D. "Inference and Missing Data," Biometrika, Vol. 63, No. 3, 1976, pp. 581-592.

Sargan, J., and E. Drettakis. "Missing Data in an Autoregressive Model," International Economic Review, Vol. 15, 1974, pp. 39-58.

Sun, F. "On Estimation with Missing Data," in Proceedings of the 21st IEEE Conference on Decision and Control, New York: IEEE, 1982, pp. 1197-1198.

Woolson, R., and W. Cole. "Comparing Means of Correlated Variates with Missing Data," Communications in Statistics, Vol. 3, No. 10, 1974, pp. 941-948.

NBS-114A (REV. 2-8C)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET (See instructions)	1. PUBLICATION OR REPORT NO. NBSIR 84-2963	2. Performing Organ. Report No.	3. Publication Date November 1984

4. TITLE AND SUBTITLE

A Survey of Techniques for Evaluating Emergency Planning Models and Data Bases

5. AUTHOR(S)

Robert E. Chapman, Robert G. Hendrickson and Stephen F. Weber

6. PERFORMING ORGANIZATION (If joint or other than NBS, see instructions)

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP)

Federal Emergency Management Agency
500 C Street S.W.
Washington, D.C. 20472

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)

This report presents a state-of-the-art review of the technical literature on model and data base evaluation which relates to the modeling problems faced by the Federal government in discharging its emergency planning activities. Particular emphasis is placed on model/data base documentation and verification, an analysis of the structural properties of a model/data base, and the role of sensitivity analysis. Emphasis is also placed on four special topics (rare events, missing data, error propagation and stability) which are involved in many emergency planning activities.

12. KEY WORDS (Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)

Assessment; data base; data quality; documentation; emergency planning; evaluation; mathematical model; model credability; policy models; sensitivity analysis

13. AVAILABILITY

☒ Unlimited

☐ For Official Distribution. Do Not Release to NTIS

☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.

☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES

15. Price