NBS SPECIAL PUBLICATION 534

U.S. DEPARTMENT OF COMMERCE / National Bureau of Standards

Utility and Use of Large-Scale Mathematical Models

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau's technical work is performed by the National Measurement Laboratory, the National Engineering Laboratory, and the Institute for Computer Sciences and Technology.

THE NATIONAL MEASUREMENT LABORATORY provides the national system of physical and chemical and materials measurement; coordinates the system with measurement systems of other nations and furnishes essential services leading to accurate and uniform physical and chemical measurement throughout the Nation's scientific community, industry, and commerce; conducts materials research leading to improved methods of measurement, standards, and data on the properties of materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government Agencies; develops, produces, and distributes Standard Reference Materials; and provides calibration services. The Laboratory consists of the following centers:

Absolute Physical Quantities[2] — Radiation Research — Thermodynamics and Molecular Science — Analytical Chemistry — Materials Science.

THE NATIONAL ENGINEERING LABORATORY provides technology and technical services to users in the public and private sectors to address national needs and to solve national problems in the public interest; conducts research in engineering and applied science in support of objectives in these efforts; builds and maintains competence in the necessary disciplines required to carry out this research and technical service; develops engineering data and measurement capabilities; provides engineering measurement traceability services; develops test methods and proposes engineering standards and code changes; develops and proposes new engineering practices; and develops and improves mechanisms to transfer results of its research to the utlimate user. The Laboratory consists of the following centers:

Applied Mathematics — Electronics and Electrical Engineering[2] — Mechanical Engineering and Process Technology[2] — Building Technology — Fire Research — Consumer Product Technology — Field Methods.

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides scientific and technical services to aid Federal Agencies in the selection, acquisition, application, and use of computer technology to improve effectiveness and economy in Government operations in accordance with Public Law 89-306 (40 U.S.C. 759), relevant Executive Orders, and other directives; carries out this mission by managing the Federal Information Processing Standards Program, developing Federal ADP standards guidelines, and managing Federal participation in ADP voluntary standardization activities; provides scientific and technological advisory services and assistance to Federal Agencies; and provides the technical foundation for computer-related policies of the Federal Government. The Institute consists of the following divisions:

Systems and Software — Computer Systems Engineering — Information Technology.

[1]Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington,D.C. 20234.
[2]Some divisions within the center are located at Boulder, Colorado, 80303.

The National Bureau of Standards was reorganized, effective April 9, 1978.

Utility and Use of Large-Scale Mathematical Models

Proceedings of a Workshop
Held at the National Bureau of Standards
Gaithersburg, Maryland

April 28-29, 1977

Edited by:

Saul I. Gass

National Engineering Laboratory
National Bureau of Standards
Washington, D.C. 20234

Sponsored by:

Operations Research Division
Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards
Washington, D.C. 20234

U.S. DEPARTMENT OF COMMERCE, Juanita M. Kreps, Secretary

Jordan J. Baruch, Assistant Secretary for Science and Technology

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

Issued May 1979

Library of Congress Catalog Card Number: 79-600045

National Bureau of Standards Special Publication 534
Nat. Bur. Stand. (U.S.), Spec. Publ. 534, 217 pages (May 1979)
CODEN: XNBSAV

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON: 1979

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $4.25

Stock Number 003-003-02060-5

(Add 25 percent additional for other than U.S. mailing).

ABSTRACT

The Workshop on the Utility and Use of Large-Scale Mathematical Models held at the National Bureau of Standards, Gaithersburg, Maryland (April 28-29, 1977), was a "first" for its purpose to examine the problem of how to improve the use and utility of large-scale mathematical models in the Federal Government. The Workshop speakers addressed specific problem areas, including: the present status of model use in DOD and non-DOD applications, issues facing developers, problems of model implementation, transfer and development in the energy field, model assessment and evaluation, use in policy analysis, comparison of models, management of the modeling process, model software and documentation, and guidelines, standards and management improvement activities. This Proceedings volume presents the papers and much of the discussion that took place at the Workshop, along with a summary of directions for needed research.

KEYWORDS: Documentation; energy; evaluation; guidelines; implementation; large-scale; management; mathematical models; policy analysis; software; standards; transfer.

CONTENTS

WELCOMING REMARKS

A. J. Goldman

Good morning. Happiness may be a warm puppy for some people, but for me, today, it consists in welcoming you most warmly at the beginning of this Workshop. That welcome is offered on behalf of the National Bureau of Standards (NBS), and more specifically on behalf of our Applied Mathematics Division and its Operations Research Section, which is hosting the meeting.

I should like to devote a few minutes to explaining why we take such great pleasure in the convening of this Workshop. In the sixteen years since the formal birth of our Operations Research Section (and during several prior years of its existence as an informal embryo), we have undertaken a great variety of mathematical modeling activities for a great many Federal agencies. Though mainly model developers, we have also served as methodology contributors, model users, monitors of modeling efforts, evaluators of models, on occasion pall-bearers to models -- the whole gamut of roles. This long and sometimes painful history has taught us a number of disconcerting truths:

- that a model can be conceptually sound, but algorithmically inefficient or inaccurate,
- or algorithmically nifty, but conceptually or empirically dubious,
- or technically excellent in every sense, but not useful
- or both excellent and useful, but go unused,
- or, whether excellent or not, be misused and so on (you can readily add to the litany).

These observations have led us to three conclusions, reinforced by many conversations with colleagues and strongly corroborated by recent studies and events:

1. The very large Federal investment, in the development of decision-aiding mathematical models, has not "paid off" as it can and should. Some of the more obvious contributing causes involve the absence of articulated procedural guidelines and professional standards.

2. The attendant disappointments can delay and diminish use of the great (perhaps, indispensable) potential of modeling to illuminate major public issues and to improve Government decisions and operations.

3. The mainstream elements of the professional modeling community should exercise leadership in identifying and diagnosing the underlying problems and in moving toward their correction. Otherwise, less palatable prescriptions may be forthcoming from quarters less sensitive to

some of the realities of the modeling process, in particular its creative/innovative elements and their need for flexibility.

Accordingly, for the past several years we have been proposing a program of research, experimentation and development aimed at better understanding of the issues involved, and at technical aids, guidelines and protocols helpful in improving the planning and execution of model-development projects, the documentation and evaluation of models, and their subsequent maintenance and application by users. In this effort we have been joined by colleagues in the NBS Institute for Computer Science and Technology, though their interests are centered more in the "functional fidelity" of real-time decision systems than in the policy and planning-aid models emphasized here. From the legislative branch, the General Accounting Office is pressing NBS for greater activity in this area of responsibility. Thus, an intensification of effort in the fairly near future seems quite likely.

We are delighted to provide, through the Workshop, a forum in which the modeling community can sharpen its perception and articulation of this delicate topic. We are anxious to learn your views on the principal deficiencies, priorities, and opportunities for corrective action in this field. A mundane incidental, the absence of a registration-fee, symbolizes our understanding that we will be among the main beneficiaries. We look forward to listening to the Workshop's presentations, and participating in its deliberations. Thank you for coming. I want especially to thank Saul Gass, who has taken the time -- from his professorial and chairman's duties at the University of Maryland and his presidential duties for the Operations Research Society of America -- to work with us in the modeling area and in particular to organize this meeting.

THE WORKSHOP ISSUES

Saul I. Gass

To our knowledge, this Workshop is the first of its kind -- a pioneering effort to propose and discuss approaches to improving the utility and use of mathematical models in the Federal government. In his opening remarks, Alan Goldman described some of the reasons that compel us to seek such improvements, and the NBS long-term interest and involvement in this area. It is his view that "The mainstream elements of the professional modeling community should exercise leadership in identifying and diagnosing the underlying problems and in moving toward their correction." In organizing this Workshop, we have kept this point in mind. The speakers and attendees were all invited on the basis of their professional stature, their demonstrated concerns in these matters, and their commitment, as modeling professionals, to seek and to work for improvements in the Federal government's use of models.

The phrase "large-scale mathematical models" implies complexity in terms of model structure and data requirements, computational procedures, and interpretation and use of outputs and results. I believe the modeling community feels that they have demonstrated or could readily demonstrate the power of such models for many governmental decision areas. But, based on recent surveys (to be discussed next by their principal investigators), the general impression is that many models have been little used nor long remembered. Contrasting exceptions do exist. The FEA's Project Independence Evaluation System (PIES) has been used by both the Ford and Carter Administrations to evaluate alternative energy initiatives, and is a viable and ongoing complex model. The EPA's Strategic Environmental Assessment System (SEAS) is a complex model for evaluating the long-range impact of activities and policies on the environment at national and region levels. It was used by EPA for developing its 1975 report to Congress on the cost of clean air and water, and by the Council for Environmental Quality to project pollution in their 1974 report to the President. The SEAS model has lost its major supporters in EPA and OMB and its future utility is in question. The costs of both PIES and SEAS are in the multimillion dollar range.

There is a pressing need to close the gap between what model developers can actually do with their models and the understanding of such applications by the designated user. Thus, a major purpose of the Workshop is to determine how we can improve the utility and use of large-scale models.

A few definitions are in order. Utility implies usefulness and usability. A model can be considered useful if it can be shown to have attained its stated objectives. A model can be considered usable if it is understandable and plausible to both technicians and policymakers, economic to run on a computer, and accessible to those who wish to use it. If a model is useful and usable, it stands a good chance of being used, i.e., has high utility, especially if the potential users receive proper training in it use and interpretation [1].

The surveys [2, 3] indicate that about 75-80% of non-DoD models are developed by contractors and grantees, with 59% being done by universities and 20% by profit and not-for-profit organizations. In the DoD area, 55% of the models are developed externally, with 5% by universities and 50% by profit and not-for-profit groups; see Table 1.

This data indicates that most government models are developed by someone other than the ultimate user. The involvement of user groups in the developmental process is, of course, implied, but the final products can be assumed to be based on the concepts and analyses of the contractors or grantees. Thus, model improvement activities, be they guidelines, standards or whatever, must take into consideration the business and technical interests and capabilities of nongovernment model developers. My concern is that, unless this experienced and important class of model innovators and developers is an active party to any Federal model improvement program, we might find government agencies setting guidelines and standards without developer review. And, as Alan Goldman noted, this might reduce the modeler's ability to innovate, be creative and have the flexibility necessary to produce the best state-of-the-art and beyond model.

To my mind, this is the major reason for the Workshop. The total modeling community -- of which the Workshop participants are key members -- needs to address the model improvement problem to ensure that any proposals deemed necessary will be wanted, accepted, and viable by developers, as well as users.

A basic list of issues concerning guidelines, standards and management improvement that is open for discussion is given in Figure 1. During the course of the next two days we will have presentations that encompass many of these issues and describe specific models and assessment activites. We want this to be a Workshop in the true sense of the term and want to encourage discussion during and after the formal presentations.

On Friday, we will sum up the views of the participants as to what research directions should be pursued, and what approaches will or will not work to improve model utility. At that time, we will open for discussion the GAO report [4], the evaluation questionnaire [5, 6], and the individual issues raised and discussed.

Based on tape recordings, notes and speaker handouts, I will attempt to develop a Workshop proceedings and a summary of our views. I want to thank all of you for attending, and for your cooperation in helping me to arrange the Workshop.

Table 1

MODELS BY DEVELOPING INSTITUTION*

MODEL DEVELOPER

MODEL TYPE	Government Agency	University	For Profit	Not For Profit	Total
Non-DOD	36 21%	104 59%	20 11%	 9%	175
DOD	59 45%	7 5%	37 28%	29 22%	132

*Sources: References [2], [3]. The report [4] shows that for 57 models, 75% were developed under contract or grant.

ISSUES CONCERNING MODEL GUIDELINES, STANDARDS AND MANAGEMENT IMPROVEMENT

1. Feasibility of model guidelines or standards,
2. Model management proposals,
3. Model documentation standards,
4. Model evaluation and assessment procedures,
5. Relation to programming standards and documentation,
6. Voluntary or mandatory guidelines,
7. Contract and grant conditions,
8. Role of government contract technical monitor,
9. RFP requirements and statement of work,
10. Model dissemination requirements,
11. Government model testing and verification center,
12. Computer model clearinghouse,
13. User training requirements,
14. Use of financial and milestone (PERT) review techniques,
15. Financial penalties for not meeting agreed project objectives,
16. Model review boards during the life of the project,
17. Definition of large-scale computer-based models to which guidelines apply,
18. Ways of measuring improvement in modeling process,
19. Experimental or other approach to initiating any guidelines,
20. Approach for determining final set of guidelines, and
21. Process for monitoring and changing guidelines.

Figure 1

REFERENCES

[1] Anon, COMPUTER SIMULATION METHODS TO AID NATIONAL GROWTH POLICY, prepared for the Subcommittee on Fisheries and Widelife Conservation and the Environment, U.S. Government Printing Office, Washington, D.C., 1975.

[2] G. Fromm, W.L. Hamilton and D.E. Hamilton, FEDERALLY SUPPORTED MATHEMATICAL MODELS: SURVEY AND ANALYSIS, U.S. Government Printing Office, Stock No. 038-000-00221-0, Washington, D.C., 1975.

[3] M. Shubik and G.D. Brewer, MODELS, SIMULATIONS AND GAMES - A SURVEY, Rand Report, R-1060-ARPA/RC, May 1972.

[4] Anon, WAYS TO IMPROVE MANAGEMENT OF FEDERALLY FUNDED COMPUTERIZED MODELS, U.S. General Accounting Office, LCS-75-111, Washington, D.C., August 22, 1976.

[5] S.I. Gass, "Evaluation of Complex Models," COMPUTERS AND OPERATION RESEARCH, Vol. 4, No. 1, March 1977.

[6] S.I. Gass, "A Procedure for the Evaluation of Complex Models," PROCEEDINGS OF FIRST INTERNATIONAL CONFERENCE IN MATHEMATICAL MODELS, University of Missouri, Rolla, 1977.

REVIEW OF THE DOD MODELING EFFORT AND MODELING AS A PROFESSION

Garry D. Brewer

I think many of you in the room are aware of the fact that my colleague Martin Shubik and I some years ago produced a questionnaire that was administered by a unit of the General Accounting Office. The questionnaire was intended to get a fairly interesting sample of all the operational models, simulations and games in the Department of Defense's active inventory as of about 1971 or 1972. That work was published in a RAND Corporation document titled "Models, Simulations and Games: A Survey." The survey instrument contained numerical as well as descriptive information on what we saw at that particular point in time. I would like to remark that that probably was a unique experience and trying to duplicate it might even be impossible. The General Accounting Office group had been tasked independently and at roughly the same time by Mr. Mahan, of the House Appropriations Committee, to go find out "what those guys over in the Pentagon are doing with war games." They had gone off on their own without technical help, but with the kind of entree the GAO has when they come representing the House Appropriations Committee. We accidentally happended on this group of auditors in a meeting much like this and served as unofficial collaborators, I guess that is a good word for it, in training their people and in developing a lengthy questionnaire. Without that kind of collaborative experience, we would have had zero hope of anyone providing an answer. Because the GAO was the administering agent of the questionnaire, which went some 70 pages, we got better than a 90% response rate for those particular models, simulations, and games that we and they identified as being interesting. I'm not going to talk about the survey. The document was published in 1974, it is available under the title, plus the RAND designation number R-1060-ARPA/RC. You are probably familiar with how to get in touch with the RAND Publications. The Corporation itself put in some extra money because it felt that it had some professional, let me stress the word "professional," obligations to get this particular study out and available to a wider constituency. At last report, something on the order of a little less than 3,000 copies of the survey had been distributed by various means by RAND. For RAND it was a best seller.

I tried to sit down and structure my presentation as a loose discussion. It just didn't work -- there is too much to talk about. So I prepared a paper on communication issues. This approach was taken because it is a nice, short way of trying to summarize what Martin and I found to be some of the more glaring problems in professional development. In fact, there is not very good communication among individuals who are either buying, building or using models of any size, but particularly the large-scale ones. Largely stimulated by the interest that the survey generated and some other work that Shubik and I did (by way or reviewing literature and publishing very critical results of some of that literature) we were encouraged by RAND corporate management to go ahead and produce a regular book-length statement of the state-of-the-art. That was the basic task. It is finished and is being published in the fall of 1978 by Harvard University Press under the title, "THE WAR GAME: A Critique of Military Problem Solving."

The topic I outlined for myself is professionalism -- the communication issue. You all know dimensions of the question of professionalism (or the lack of it) that could obviously be discussed. However, communication is critical to the group assembled here.

The level of professional communication concerning models, simulations, and games is dangerously low. There is a great need for better coordination, documentation and communication of how models, simulations, and games are used at the operational, research and bureaucratic interfaces. Merely completing a study or analysis according to contract specifications is insufficient. What becomes of the study, and how the study gets used are far more important considerations, and they are not well communicated in the present system. It is essential that a more rational expenditure of resources be established to ensure that evaluations of previous studies and analyses are done and recorded widely; this is of far greater value than additional studies and analyses run without benefit of such inquiries. We just continue doing the nth study and never try to accumulate a track record or determine who is doing a good job and why, and who is using these models and to what effect. That kind of information doesn't exist and it should.

Weapons evaluation studies, for instance, that are either unused or misused may be worse than no studies at all. All you have to do is look at the current "debate" with respect to strategic arms to get some sense of the abuse of analytic power, and it's our fault. It's not really the fault of politicians who take and misuse the numbers generated by our models. For example, ill-conceived procedures of stewardship (e.g., military activities with high rates of turnover and personnel discontinuities which produce short memories), coupled with highly uneven documentation standards and procedures, account for much low and/or ineffective model use. If no one remembers why an existing model was built, for whom it was intended, or what its peculiar operational characteristics are, it is likely that the model will be used incorrectly or a new one may have to be built from scratch. This kind of wasteful activity can be directly attributed to poor or nonexistent documentation of one sort or another. That statement has to be modified, as the sources of waste and abuse are really a lack of attention and resources -- PROFESSIONAL attention and resources -- being paid to the documentation question, we will get to that in a moment.

The sum total of professional experience is currently unnecessarily fragmented. Groups of professionals are often not aware of the existence of others doing fundamentally the same work, but in another place. There are two basic dimensions to the problem: one is the need to create information about the collective experience and subsequently, to retain and transmit this information to others involved in the decision process responsible for the production, construction and use of military models. The second is the absence of institutional memory and furthermore, even if such existed, it would not matter because there is no means of transmitting that information to other individuals in the field. The lack of communications even among builders in this business is astounding. This group is not that large and they still don't know very well what each other is doing. Little pockets of ignorance -- that, I think, is a summary description of the business.

The first dimension (creating information) calls attention to the pressing need for documentation, library efforts, and a host of management controls that would together produce much information needed by current processes and practices. The second dimension (developing an institutional memory) signals the need to understand the variety of impediments to communications that currently exist and the variety of related design requisites to overcome these impediments. And 1gd$b "rm om"!†v*L it's not really the problem of politicians and it's not the problem of users -- it's a professional problem and it's one that we've got to be a lot more serious about than I think we have been up until now. No one individual or no institution has a complete map of the whole system at any level of detail or comprehension. Lacking such a map of the whole, the system merely drifts. Who keeps tabs on the individual conditions, standards, and industry norms? Nobody. Who is evaluating the effects of the deficient documentation practices on the aggregate enterprise, and the impact of the high turnover of military users and producers in the quality and effectiveness of model use? Nobody. Who is studying the implications of the apparent trend toward increased in-house capabilities and willingness to build a new model? Our survey shows a decided tendency over the last ten or fifteen years for the military, as a reaction to the McNamara "whiz kid" days, to train and use their own analysts and to rely less heavily on outside practitioners for work. This results in less demand for professional standards and scrutiny, even less than before, and less demand to document. There is a clear tendency to adhere in these studies to uniform military standards rather than to external professional ones that we all recognize as being more appropriate. All you have to do is look at certain trends in the analysis business, where dollars are drying up for the CNA's, IDA's and the RAC's of the world, to realize that this is a real trend, and one that no one has talked much about.

Who is responsible for sensing cues from the overall system that would signal needed reseach that is likely to have payoffs, not only for the military but for the profession as a whole? For example, I would like to cite the recent revelation by Paul Bracken and some others at the Hudson Institute that many of the military models used to study warfare over the land masses of northern and central Europe, fail to account for the existence of cities. A "sensational" discovery. Crazy. [See Paul Bracken, "Urban Sprawl and NATO Defense," SURVIVAL 18:6 (Nov./Dec., 1976):254-260.]

Who initiates transfers of knowledge from one operational setting to others? The answers here, also, is no one. I will recall, for example, the heroic efforts by one academic to set-up a laboratory in Santa Barbara -- trying to build a straightforward time-sharing network and creating the whole system from scratch. He simply did not know that this technology had been in place for years on the military side. He didn't know about it or who was responsible, because there was no information available outside the military community. So he had to rediscover it for himself. This is but one example; there must be hundreds of others like it.

Detailed investigations of the impediments to the creation and diffusion of knowledge is needed in terms of the effects of proprietary motivations on the accurate representation of model production and use; the extent to which entrepreneurial incentives and impulses override scientific incentives to produce effective analyses; the results of personalized desires to "advance the state-of-the-art" rather than to solve the client's problems; the extent to which classification is invoked to obscure work of questionable value. The results of such investigations would go a long way toward resolving the broader issue of who has power in military analysis systems. Shubik and I sat down and tried to figure out from the very beginning of the average model's "life" to the end who was responsible for its various aspects. And we were horrified after that analysis, which is recorded in the book noted earlier, to conclude that usually no one person is responsible from start to finish, and hardly anybody is responsible for doing the evaluation of usage of models. No one is interested.

An acute area of interest is documentation; trying to get the standards, trying to get some realization on the part of users and funders that documentation experience is technically important, but particularly for large-scale, potentially high-use models it is absolutely essential. I don't know what would be a reasonable rule of thumb. It is an empirical question as to the amount of resources that should be devoted to the documentation effort. I do know that differences exist between successful and unsuccessful software houses. In my experience, in and around Los Angeles, it had to do with the proportion of the resources set aside for documentation. Software houses that stay in business set aside about half of the available dollars to document. Software houses that go out of business set aside less than half or they don't do it at all. That might be a rough first approximation of the magnitude of the resources needed to document properly -- maybe half.

Besides creating some standards, we need to create a body of technical expertise in this area that just doesn't exist. The practice of documentation is clearly related to the building and use of models, but it goes beyond that. I think there is a clear need, if the resources and expectations of demand have been created, for a generation of technical skills and job descriptions that don't currently exist. They range from the ability to write programs and actually run and understand and use the model as well, at one end of the technical spectrum, all the way to simple library efforts. Just to keep track of who's using the model, what they cost, and so on. We haven't done a very good job here. We haven't thought very much about it. Part of the reason is that the people charged with the analytic responsibility are under incredible pressures and deadlines to build these models. They do not appreciate the importance of documentation because they are handling things as discrete events rather than as a part of a larger professional process of development and improvement. The second thing, is that it is not their job. But then, whose job is it? We have not answered that.

We asked questions in our survey about ways of improving deficient professional communication. Let me quickly run through some of our findings and results. We asked about clearing-houses, regional centers, and external professional review boards. We might cite that over half of the models in our survey have never been exposed or reviewed outside of the group who built and used them. More than half of the sample has never had any sort of professional review. Scandalous is the only word to decribe the condition. Scandalous. We have no one to blame but ourselves.

With respect to the establishment of a clearing-house that records data about all modeling activity in DOD (my earlier point that no one knows, even at a rough, crude level what is generally going on), respondents were quite favorably disposed. We found that something on the order of 70% of the people that we interviewed in the survey thought it would be a good idea as long as it didn't interfere with their business. They would be willing to make reports and willing to go ahead and contribute a gross level of documentation about the model and model use and cost to this kind of clearing-house or regional center.

While they are willing to provide gross descriptive information about what the model does and costs and who is responsible, and so on, everyone digs in their heels and says no when it comes to standardization, excluding professional review. Documentation is okay if somebody else does it but certainly not me. It creates more bureaucracy, brings more headaches, and besides that, who wants to do it? That is basically the answer. And that's the next point. It is okay to ask about gross descriptive information in a clearing-house, but when it comes to the nitty gritty, it is unacceptable. Well, I'm not pleased about this situation and I don't think any of us in this room should be. That's really the issue. Are we willing to expose our work to professional scrutiny, comment, and criticism? Are we willing, as funders, to spend money necessary to get excellent professional review? The only answer can be yes, and it's a question of getting people to realize how and why the answer should be yes.

In thinking about what kinds of strategies might be developed to improve communications, several come to mind. A multiple attack on several fronts needs to be mounted. It's not just a simple minded thing of saying document the hell out of everything, review everything. That isn't going to get it for you. Those are just two elements in what has got to be wholesale jerking up our own professional boot straps. Other things come to mind when one stops to think about what other professionals do and why they are professionals instead of just hobby groups, which is the way in which I would describe much of what goes on here -- a hobby. For instance, we need to develop means to create more than one modeling perspective of any given problem. It's not inconceivable that for-hire institutions could be funded by the Congress to begin this task. I think the partial experience through the Congressional Budget Office and budget process is an indication that it doesn't have to be a destructive enterprise. Why not let multiple contracts on any given problem instead of having one contract and solution? That is seldom considered.

Clearly, in areas where models have very little data, e.g., strategic studies, we should proceed in phases -- we had better have alternative models. Redundancy is essential if one has no or limited data. If one had data, you could point out contradictions. But if you don't, you need alternative models. If the facts correspond with different models and they came out with similar insights, then one should be comfortable that the insights probably make sense. But if one relies on a single model, confidence in the results must diminish. Nonetheless, there are too many one-model studies that claim a spurious validity; none of these models is valid, and we know that for a fact.

Journals represent the next point that I want to bring out. I think the OPERATIONS RESEARCH JOURNAL has got more of a responsibility than it has carried out in past years -- to publicize military analyses more than it has. They show up occasionally; they show up in bits and pieces or as technical notes at one point or another. I think there is room in the JOURNAL and indeed a need, given the proportion of operations research people and resources devoted to this field, for it to be publishing more about military analyses.

Catalogs. There are various catalogs put out in various formats. I think that the catalog activity and the clearing-house activity are closely related. The notion of improving documentation goes beyond technical standards to include concerns about utility. Model use should be made a part of catalogs.

REVIEW OF THE NON-DOD MODELING EFFORT

Gary Fromm

(Dr. Fromm's remarks were based on the following material which is Chapter 1 from [1]).

The growing complexity of modern society and the rising demand for response to social problems has led to needs for better analyses of the structure of our system, better methods of anticipating future difficulties, and better means of predicting the effects of alternative actions. Models, which might be termed representations of processes, have a role to play in responding to all of these needs. They are useful in developing our understanding of physical, economic, social and other phenomena; they can be used for forecasting; and they can be employed to simulate the impacts of different structural and policy scenarios.

For these reasons, models are increasingly being employed by governments and the private sector. However, knowledge has been limited about the extent to which and the ways in which Federal agencies use models. Little information has been available concerning the types of models constructed, the level of support (money and manpower) provided for the development of models, the difficulties involved in model development and use, and the ways in which results have been applied in administrative and political decisions. Neither has extensive information been compiled on the course of various modeling efforts - how projects are initiated, by what criteria potential efforts are judged, how work is monitored and documented, what validation and evaluation tests are conducted, and how results are disseminated.

SURVEY PROCEDURES AND THE MODELING UNIVERSE

The need for greater knowledge of government modeling efforts has been recognized by the Federal Council on Science and Technology and the National Science Foundation. This recognition led to their sponsorship of a survey by Data Resources, Inc. and Abt Associates Inc. of non-defense Federal agency endeavors in this field.

Sources such as the National Technical Information Service, the Smithsonian Information Exchange computerized abstract files, and agency records on grants were used to identify over 650 models involving some aspect of social decision making. A mail survey then obtained detailed information from over 230 project directors and 80 Federal agency project monitors on the uses and characteristics of currently extant models. While the lists of projects compiled were not exhaustive, both the sample and the universe are felt to be representative of the nature and scope of Federally-supported, non-defense modeling activity in the social-human, decision making area.

Although the survey found predominant application to subjects involving economics, the models were directed to a broad range of other problems, from simulations of agricultural production to analyses of the criminal justice system. Topics treated included the prediction of economic activity at various levels (national, regional, industry), population dynamics, transportation networks, route scheduling for refuse collection, etc.

Irrespective of topic, over 90 percent of the models were computer-based. However, there was a great diversity of other structural characteristics. Median size was 25 equations, but the range in scale was great. About a quarter of the models had less than 10 equations. 30 percent reported more than 30 equations, and six included over 1,000 equations. Stochastic models (those estimated stochastically or including random error terms) were somewhat smaller, on the average, than those whose parameters were obtained using other techniques.

There is a rough correlation between scale and the time needed for development. The average development time for the models surveyed was about 17 months, with some of the larger systems requiring several years between initiation and operational status. The model's life spans are difficult to estimate, partially because of the recency of modeling activity; over 90 percent of projects began development after 1966, and over half after 1969. However, project directors reported a median two-year period (which largely seems to be independent of topic area) between operational status and a need for recalibration or reestimation. An estimated five years is required before major structural change (redevelopment and respecification) must occur.

Federal agencies developed about 20 percent of these models internally, with the rest being "extramural" projects. The majority of models (60 percent) were developed by researchers at universities, normally with grant funding. Private for-profit and non-profit research institutions developed the remaining 20 percent, usually under contract. The practice of supporting modeling work at different types of institutions varied considerably across agencies. For instance, the Department of Commerce and the independent financial agencies (such as the Federal Reserve Board, Federal Deposit Insurance Corporation, and Federal Home Loan Bank Board) most frequently developed models internally, while the Department of Housing and Urban Development and the Environmental Protection Agency placed the highest proportion of model development projects with private research organizations.

Variations in size, complexity, developing institution, and funding arrangements were reflected in development costs of the models. Although the majority of models required less than $50,000 for development, prices ranged to over $3 million, for an average development cost of $140,000. Taken together, the 222 models which responded to this question represent a total cost of more than $31 million. Extrapolating to the universe from which the sample was taken, the cost would approach $100 million. Federal funding accounted for an average of 75 percent of the cost of extramural projects, with the remainder most commonly contributed by the institutions at which the models were developed. While the "quality" of models is an elusive concept, those characteristics which normally are taken to be indicators or correlates of quality did show improvement with cost, as did policy use.

PURPOSE, CHARACTERISTICS, AND USE

Most project directors cited multiple purposes for their models. Over 70 percent named at least one policy-related purpose, such as selection among policies or programs, evaluation of policy/program effectiveness, or development of policy/program concepts. Those models for which a policy purpose was not mentioned generally were intended to advance the state of knowledge in a particular field, or were developed to serve general educational goals (including training the modeler).

On an overall basis, models seem to be used much less frequently than their designers or sponsors intend. Project directors indicated that actual use of their models fell significantly short of intended use for all but one category (the exception was general education). Moreover, notwithstanding the great degree of policy intent, actual policy application appears to have the highest shortfall of use.

Use is difficult to measure precisely, and different indicators yield different apparent levels of use. Nonetheless, it would appear that at least one-third and perhaps as many as two-thirds of the models failed to achieve their avowed purposes in the form of direct application to policy problems. Some models, of course, make indirect contributions to policy by improving knowledge in a field or by adding to the state-of-the-art of policy simulation. However, this is of small comfort, given the significant costs of modeling (in terms of both expenditures and the use of highly skilled personnel) and the missed opportunities to achieve improved policy analysis and decisions.

The results of the survey suggest several reasons why higher rates of direct application to policy purposes have not been achieved. One difficulty is the often specialized and detailed nature of policy issues in contrast to the more general focus of models.

In part, the lack of detail of models is caused by the absence or high cost of obtaining or processing "fine-grained," specialized information. Too often, data from prior studies or standard statistical references are outdated, inappropriately structured, or too highly aggregated for policy analyses. Data for the surveyed models were most frequently (in 76 percent of the cases) drawn from published sources, which rarely are fine-grained. Some new data were collected in nearly half the projects, but both project directors and agency monitors still indicated that data availability was the greatest constraint to development and application of models. Models with policy purposes required special data collection activities more often than models not oriented to policy use.

The survey provided no information on how many developers of models may have chosen to use available data to avoid the costs of collecting and compiling new statistics. However, since models are costly to begin with, budgetary constraints might often deter modelers and sponsors from seeking otherwise highly useful specialized data.

While data availability and cost pose serious problems for making models more useful for policy purposes, the primary cause of low policy utilization rates for model probably is attributable to the "distance" between model builders and potential policy makers. The specific needs of policy makers must logically be communicated to developers of models in order for the resulting systems to be most useful for the examination of policy alternatives. Under current modes of operation, a number of procedural and institutional factors limit the interactions of policy makers and modelers, and thus increase the likelihood of imperfect communication.

Although results from the survey in this area are not always statistically significant, the following patterns are evident in the sample and support this conclusion:

o The survey found that most models originated independently with their designers (78 percent of all cases) as compared with funding agencies (11 percent) or users (4 percent). Shortfalls on policy use were highest with designer-originated models. In addition, when an idea for an extramural project did originate inside a Federal agency, it most often came from a research unit rather than a unit with program or policy responsibilities. Policy use suffered accordingly.

o Most modeling is conducted outside the sponsoring and potential user agencies, and, in more than a quarter of the cases, a third institution (for example, a State or local government agency) is an intended user. The shortfall in actual as opposed to intended use was largest for such third-party user agencies, next largest for funding agencies (for extramural projects), and smallest in cases where the developing institution was the same as the user institution.

o Most extramural projects are supported through grants, with very little specification by funding agencies of desired detail and characteristics of final products. The rate of policy use was highest for models funded with greater specification of performance requirements (which generally was true under contract rather than grant arrangements).

o Real-time interaction between developers and users was low. In over 50 percent of the cases, findings were presented through the comparatively impersonal, inflexible, and infrequent media of written reports, articles, and books, rather than through direct briefings (19 percent) or runs of models and analysis of results by user agencies (34 percent).

Closely related to the issue of distance between users and developers is the problem of policy makers' capabilities to use models after they are constructed. There are two dimensions to this problem: the knowledge and skills of policy officials, and the operational ease of using the models. Both developers and funding agency personnel commented that policy makers often lack the training which would enable and enhance appropriate use of models. Both project directors and agency monitors rated "ease of use by non-technicians" as the second most important constraint (after data availability) limiting the utility of models.

The flip side of the coin is the documentation problem. If a model is developed externally and the intent is for non-developers users to run it and directly analyze results, adequate documentation is a logical prerequisite to policy use. While costs of transferring models to Federal agencies appear to be low (nearly 90 percent of project directors indicated a relocation cost of less than $5,000), documentation was considered inadequate to enable other than project personnel to set up and run the models in about 75 percent of the cases.

Moreover, most documentation took the form of reports and articles dealing with the structure and characteristics of the models and seldom included user manuals, operating instructions, or computer programs. Use rates were higher in the presence of any form of documentation, and highest when user manuals had been published. Such manuals were more often produced when funding agencies specified the desired characteristics of models and when funding was carried out under contracts rather than grants.

Finally, it is important to note some factors not generally related to policy use. No particular subject areas or structural characteristics of models were found to lead to consistently greater or lesser use. Models sponsored by some agencies received greater use than others, but this mainly reflected different support arrangements (for example, internal vs. external development) and degrees of contract or great specificity. Models developed at private research organizations were more used than those developed at universities, again reflecting the grant/contract and specification patterns.

Model size was not significantly related to the rate of policy application. Models with a large number of equations were more often intended for policy use. However, among these models, policy use occurred at similar rates in all size categories. Within the sample, there was a consistent pattern for higher-cost models to be used more, but the relationship between cost and use was not highly significant from a statistical standpoint. There was not support for the hypothesis that smaller models, because they can ostensibly be oriented to a specific type of decision, are more useful.

IMPLICATIONS FOR FEDERAL MODELING POLICIES

No comprehensive Federal policy on modeling currently exists, but a number of agencies have established or are considering policy actions. The Department of Health, Education and Welfare has established a committee to review all proposed modeling efforts in the sub-agencies within its purview. The General Accounting Office is studying ways to evaluate models, and the Environmental Protection Agency has initiated an experimental effort to increase communications between developers and potential users of policy models. In most of the agencies surveyed, requirements for specific review, validation or dissemination procedures have been placed on individual modeling efforts.

The policies which have been established or discussed can be generally divided into four groups. The first concerns the broad purposes for which models should be supported. The second group relates to the relative funding by types of models, developing institutions, funding arrangements, and so forth. The third set of policies involves establishment of regulations or requirements for the way models are developed, or for the form of final products. Finally, some policies would amount to internal initiatives within Federal agencies to enhance the development and utility of models.

PURPOSES OF FUNDINGS. Whether models should be funded in response to particular decision requirements, in a general attempt to expand knowledge about certain subjects, or in an effort to advance the methodology of modeling and related techniques, is a question far broader in scope than this study. The survey focused principally on decision applications, and did not examine desired priorities among funding purposes.

Little research has been devoted to the question of what kinds of models apply to what decisions. Comments from agency monitors suggested that models are most advantageous where alternative policies are compared in terms of predicted outcomes. But much more information is required for policy consideration in this area.

EMPHASES IN FUNDING. Agencies now carry out practices which amount to substantially differing funding policies. Some emphasize internal model development, while some mainly fund work at universities or research organizations. Some are quite specific about what type of model should result from funded projects and others are not.

This study provides no conclusive evidence that any one of these emphases is "best." Where application of models to policy decisions is intended, the survey suggests that the probability of utilization is greater for models developed internally or with considerable specification of requirements to the external developer. Under current patterns, this specification is greatest under contract funding.

There is no indication from the survey that any particular types of models in terms of subject areas, structural characteristics, size, and so forth are more likely to be used in policy decisions than others.

REQUIREMENTS IN DEVELOPMENT. Most proposed requirements for developers of models concern either documentation or validation. This study offers some evidence in support of the need for and desirability of greater documentation: there was a consistent pattern of higher utilization rates when the models were reported to be better documented. There is no information which argues for particular types or amounts of documentation, but gains from the provision of user manuals seem sizable.

The study addressed validation issues only by asking modelers and sponsors for their opinions on standards or requirements for review of models. While respondents conceded benefits such as increased credibility, their overall reaction to the imposition of such standards was negative, based on fears of red tape and stifled innovation.

INTERNAL INITIATIVES. The idea of a Federal clearinghouse for models has often been suggested, and respondents reacted very favorably to the proposal. Further, attempts to develop a "universe" of models for the survey made it clear that current sources of such information are far from comprehensive. There is a question of priorities, however: the problem of utilization appears more severe than that of development, and the clearing-house as usually defined seems more an aid to developers than to users.

The possibility of Federal efforts to develop standardized computer routines or technical procedures was also presented in the survey. The overall reaction was positive, but many respondents felt such an effort would duplicate existing private work.

Some agencies are discussing or undertaking efforts to increase the ability of potential users to understand and apply models to their decision problems. Survey respondents suggested ideas along these lines, including scheduled briefings and conferences throughout development of models, review panels composed of potential users, and straightforward training efforts.

Lack of data was noted as the most severe constraint on the modeling efforts surveyed, and several respondents argued for a Federal effort to make more integrated socio-economic data available to modelers.

CONCLUSION

Judging from responses and opinions expressed in the survey by model builders and Federal agency personnel, and by independent sources, modeling and other rigorous analytical techniques can make significant contributions to the examination of policy alternatives and the alleviation of social problems. However, in order to realize these opportunities and to raise the low policy application return on most current modeling and analytical research expenditures, improvements must be made in the availability of data, in procedures used to fund and monitor modeling and analytical research, and in information flows between analysts, model builders, and policy makers. In general, guidelines and strategies for the conduct of research within and sponsored by Federal agencies are now lacking, and should be considered by appropriate authorities within or across agencies at an early date.

REFERENCE

[1.] "Federally Supported Mathematical Models," G. Fromm, W. L. Hamilton, and D.E. Hamilton, Stock No. 038-000-00221-0, U.S. GPO, Washington, D.C. 20402.

ISSUES FACING MODEL DEVELOPERS - I

Seth Bonder

When Saul asked me to do this, he didn't exactly tell me what I was supposed to do. I'm sorry to say I read the GAO report only yesterday, and doing so made me wish my prepared remarks were more directly responsive to it. I was both pleased and appalled at what I read -- really, both simultaneously. What I have prepared is a group of slides to quickly give you a sense of what some of the developer problems are in the defense community, and I'm talking about a specific type of model: large-scale defense models. These are basically what I would call general purpose force models; that is, large-scale mixes of land and air forces in Europe, that kind of modeling activity. As shown on the outline (slide 1), I have prepared some slides about some of the defense issues addressed by the models merely to give you an idea of the spectrum of questions and issues that some of these models are intended to illuminate when they are used. I want to separate the model from its use and, in fact, the user from the decision maker. They're really different activities. We may build a model for a technical agency in the Pentagon who will use it and then present the results to the decision maker which the GAO report refers to as "management." I am not sure how they use that word. So there really are three populations of players. You will see many of my biases; because I not only build models, I use them, and I teach them in universities. I also manage organizations that do this. So I have a sense of all the areas. I'll talk briefly about model types, perhaps carrying coal to Newcastle, to let you know the kinds of different models that are developed in the defense community. Then I'll spend, hopefully, most of the time on developer considerations, because you have a mixed set of pressures as a developer as to what kinds of models to build for users. Next, I'll present a summary of development trends, concluding with a statement vis-a-vis the GAO report about what I think is a myopic point of view. I am going to do all of this rapidly, somewhat as a subliminal presentation.

Slide 2 lists a set of weapon characteristics. Our clients would like models to tell them (1) is it worth doing R&D to improve those kinds of things or (2) should I write specifications for systems that do those things better... really technical kinds of questions. Slide 3 lists illustrative system choice questions; the problem is which one to buy between comparable systems. Do I buy a new tank, or keep the old tank? Do I buy a new close air support aircraft, A-10, or do I use the F4? Do I buy RPV's or a Mohawk intelligence collection system? These are questions of choice between systems, and there is a lot of money involved in these choices. Moving up the line to higher kinds of policy oriented issues -- not quite policy yet -- slide 4 lists the next questions about material mix. This is not only the choice between System A and B, but how many of which type, and usually not comparable systems. Shown here are attack helicopters vs. A-10's, a very crucial issue now between the Air Force and the Army. Which one of those systems should we buy? And how should they be mixed together? TOW vs. CLGP; one branch of the Army vs. another branch of the Army; air defense artillery, that is, the Army's ability

(SLIDE 1)

OUTLINE

- DEFENSE ISSUES ADDRESSED BY MODELS
- MODEL TYPES
- DEVELOPMENT CONSIDERATIONS
- SUMMARY OF DEVELOPMENT TRENDS

(SLIDE 2)

WEAPON CHARACTERISTICS QUESTIONS

- TANK FIRING RATES
- TANK ROUND FLIGHT TIMES
- ANTI-TANK WEAPON HIT AND KILL PROBABILITIES
- ARTILLERY RANGE CAPABILITIES
- ARTILLERY TARGET LOCATION ERRORS
- CAS AIRCRAFT RANGE AND SPEED
- AIR DEFENSE WEAPON FIRE CONTROL CAPABILITY
- ATTACK HELICOPTER ORDNANCE LOAD CAPABILITY

.
.
.

(SLIDE 3)

SYSTEM CHOICE QUESTIONS

(CHOICE BETWEEN COMPARABLE WEAPONS OR OTHER SYSTEMS)

- XM1 VERSUS M60A3
- DRAGON VERSUS MILAN
- A-10 VERSUS F-4 IN CAS ROLE
- RPV VERSUS MOHAWK (QUICK LOOK) INTELLIGENCE COLLECTION

.
.
.
.

(SLIDE 4)

WEAPONS MIX QUESTIONS

(SELECTION OF TYPES AND NUMBERS OF WEAPONS)

- ATTACK HELICOPTERS VERSUS CAS AIRCRAFT
- TOW VERSUS CLGP
- ADA VERSUS AIR INTERCEPTORS

•
•
•

to shoot down airplanes, vs. interceptors, the Air Force's ability to shoot down aircraft; how do we determine an appropriate mix of those kinds of things?

Slide 5 presents larger-scale force structure questions, concerning the amounts or relative proportions of combat arms within the Army. That is, how many and which types of divisions to have, how much of field artillery and how much air defense artillery. Within the Air Force, there are questions of how to mix the amount of close air support aircraft with the amount of airborne interdiction, with the amount of counter-air capability. How much of which things to buy? These are large-scale, organizational, force structure questions. The problem clearly exists between the Army and the Air Force of how many divisions vs. how many wings. Now you get to very high level OSD types of questions, and then you get international kinds of questions tied in, in part, to the arms control question. How many of which U.S. Forces vis-a-vis how many of the West German forces vis-a-vis how many British forces? When you get into arms control like the MBFR (Mutual Balanced Force Reduction), what do we trade with the Soviets and the Pact nations? These are current questions; I don't want to say models help solve them, but people think about using models to address them. As indicated on slide 6, there are other kinds of questions on force employment issues, how to use the systems when we buy them, fronting tactics, size of initial force vs. mobilization, what should we put on the ground in Europe now vs. what should we bring over in 10 or 15 or 30 days, and many more of those kinds of issues.

The next slide (slide 7) considers the study of individual processes, that is, the impact of reducing communications time; is it worthwhile to improve logistics, command-control, movement? Slide 8 indicates the need to study trade-offs of one particular process for another. How much intelligence information is required vs. the increased congestion in the communication system you get? The military want to make those kinds of trades, because they buy systems, either intelligence systems or communications systems -- intelligence gathering vs. target acquisition capabilities -- a whole bunch of trade-offs like that. And then finally, the next slide (slide 9) presents net-assessment questions, comparative NATO-Pact issues. We use very effective, very sophisticated airplanes. Those of the Warsaw Pact nations are not quite as sophisticated, and they buy more of them. Is that better or worse? Should we employ that kind of policy? We use large reserve forces; they use an echelon concept. It is a comparative kind of thing. We use a moderate amount of artillery; they use lots of artillery systems. Is that a better way? And if we trade in arms control, how should we trade?

Those are the kinds of questions that models address, and therefore, you require a whole spectrum of models. One model doesn't address all questions; so there are different levels of models, both technical as well as force structure models.

I've used the word models somewhat generically. Let me talk about three or four types that are used. And I think these are mentioned in some of your reports. War games (slide 10) -- that is the name of your book, but you are using it in a broader sense than I am.

(SLIDE 5)

FORCE STRUCTURE QUESTIONS

(AMOUNTS AND RELATIVE PROPORTIONS OF COMBAT ARMS)

WITHIN ARMY

- NUMBER OF ARMORED DIVISIONS
- NUMBER OF INFANTRY DIVISIONS
- NUMBER OF MECHANIZED INFANTRY DIVISIONS
- AMOUNT OF FIELD ARTILLERY
- AMOUNT OF AIR DEFENSE ARTILLERY

•
•
•

WITHIN AIR FORCE

- AMOUNT OF CLOSE AIR SUPPORT CAPABILITY
- AMOUNT OF AIRBORNE INTERDICTION CAPABILITY
- AMOUNT OF COUNTERAIR CAPABILITY

•
•
•

BETWEEN ARMY AND AIR FORCE

- BALANCE OF FIELD ARTILLERY, ATTACK HELICOPTERS, AND CLOSE AIR SUPPORT
- BALANCE OF AIR DEFENSE ARTILLERY AND AIRBORNE INTERDICTION

•
•
•

AMONG NATO FORCES

(SLIDE 6)

QUESTIONS OF TACTICS AND DOCTRINE

(FORCE EMPLOYMENT ISSUES)

- ALTERNATIVE FRONTING TACTICS
- SIZE OF INITIAL FORCE VERSUS MOBILIZATION CAPABILITY
- LOCATION OF INITIAL DEFENSIVE POSITIONS
- DETERMINATION OF CONDITIONS FOR DELAY, DEFENSE, AND COUNTERATTACK
- SIZE OF LOCAL SUPPLY STOCKPILES VERSUS CAPABILITY TO RESUPPLY
- FIRE SUPPORT ALLOCATION STRATEGIES
- MAINTENANCE OF AIR ALERT VERSUS GROUND ALERT FOR CLOSE AIR SUPPORT AND INTERDICTION AIRCRAFT
- EMPLOYMENT OF ATTACK HELICOPTER AS A FIRE SUPPORT RESOURCE OR USE IN A SCREENING ROLE
- SIZE AND LOCATION OF RESERVE FORCES

•
•
•

(SLIDE 7)

EXAMINATION OF VALUE OF

INDIVIDUAL PROCESSES

COMMUNICATIONS: IMPACT OF REDUCTION IN MESSAGE PROCESSING TIMES

LOGISTICS: BENEFIT OF GREATER SUPPLY LEVELS (BY AMMUNITION TYPE OR POL TYPE)

COMMAND AND CONTROL: EFFECTS OF REDUCED DECISION LAGS WHICH DECREASE

- RESERVE COMMITMENT TIMES
- FIRE SUPPORT DELIVERY TIMES
- SUPPLY DELIVERY TIMES

.
.
.

MOVEMENT: EFFECTS OF GREATER MOVEMENT SPEEDS WHICH DECREASE

- RESERVE COMMITMENT TIMES
- SUPPLY DELIVERY TIMES

.
.
.

INTELLIGENCE: EFFECTS OF BETTER ESTIMATES OF

- ENEMY STRENGTHS
- ENEMY LOCATIONS
- WEATHER

.
.
.

(SLIDE 8)

EXAMINATION OF TRADE-OFFS AMONG PROCESSES

- AMOUNT OF INTELLIGENCE INFORMATION REQUIRED VERSUS INCREASED CONGESTION OF COMMUNICATIONS SYSTEM

- TGT ACQ/INTELL NEEDS VERSUS FIRE SUPPORT REQUIREMENTS, E.G., RECON AIRCRAFT VERSUS ATTACK AIRCRAFT

- INTELLIGENCE GATHERING CAPABILITY VERSUS TARGET ACQUISITION CAPABILITY

.
.
.

(SLIDE 9)

NET ASSESSMENT QUESTIONS

(COMPARATIVE NATO/PACT ISSUES)

- COMPLEX, HIGHLY-EFFECTIVE AIRCRAFT VERSUS SIMPLER, LESS EFFECTIVE AIRCRAFT WHICH ARE MORE RELIABLE AND REQUIRE LESS SUPPORT

- USE OF LARGE RESERVE FORCE VERSUS ECHELONING CONCEPT WITH VERY SMALL RESERVE FORCE

- MODERATE VERSUS LARGE RELATIVE PROPORTION OF FIELD ARTILLERY

- COMMAND CONTROL STRUCTURE

- INTELLIGENCE STRUCTURE

.
.
.

(SLIDE 10)

WAR GAME

- PLAYERS REPRESENT COMMANDERS AND STAFF
- COMPUTER-ASSISTED ASSESSMENTS
- EXPENSIVE TO DEVELOP
- HIGH OUTPUT VARIANCE
- INAPPROPRIATE FOR EXAMINATION OF MANY SYSTEM ALTERNATIVES
- GOOD DIAGNOSTIC TOOL

I am talking about war games in which you have players that do the decision making and represent behavior. That is, they do the force allocations; they decide who moves where, what aircraft attack which kinds of bases. There are players that do all the behavioral activities. And that's very useful because we don't know how to model those activities very well. In fact, nobody does. However, there are some techniques that have been developed that try to do that in an automatic sense. We have developed a lot of computer-assisted war games. They are very expensive to develop -- expensive in development time and money. I've seen them take five, six, seven, eight years to develop, spending multi-multi-multi-millions of dollars to develop war games. One that started in '67 was not really completed and useful until '72 or '73. It goes through a lot of iterations. They have high output variations, clearly. Change the players, you get different results. The behavioral impact on model results are very, very significant. Model results are very sensitive to that. If people choose a different alternative in using their resources, then they get different outputs, clearly. You can win the war, or lose the war, in different ways. I think they (war games) are very inappropriate for examining many alternative ways of doing things. Why? Because they take too long. In '71 it used to take six months to play ten hours of combat. Now it takes about two weeks to play ten hours of combat and one situation. If you want to vary the things we talked about in the questions, you can't do it. They are very nice diagnostic tools. They tell you where the problems are, not how to fix them, because you are observing what is taking place. That's one type of model.

Next, (slide 11) is the simulation model, and I mean this very precisely. There aren't any players, so you somehow simulate the behavioral actions, usually by what are called the rules of engagement or decision logic of some kind. The model development processes are what I am going to use, to define what I mean by simulation vis-a-vis analytic models. What you do is decompose the process, i.e., you try to figure out what happens in the world and lay out a sequential structure of the process. In effect, you normally sequence the events and activities, how you think they may occur. Clearly, it doesn't make sense to do this in Ann Arbor; you have to find out how the process operates and interact with the people who live in the process -- that's what we try to do. And then once you do that in the simulation, by my definition, you act out the process to solve it. That is, you literally go through and lay out that process. The simulation can either be deterministic or stochastic -- I think everybody recognizes that. Most people, for example, simulate transportation networks in a deterministic fashion. They can be stochastic, where if you use probability distributions as inputs, you get as output sample probability distributions. Because you sample when solving simulation models by Monte Carlo sampling procedures, they are called, in fact, Monte Carlo simulations. You produce sample probability distributions as output. Some comments on the simulations. They are much more abstract than war games because they don't have any players. They are less expensive to develop and use than war games, but simulations also are fairly expensive. You've got to take two, three, four years to build some reasonably decent sized ones. I know of a battalion level simulation that may take an hour per replication (a Monte Carlo simulation) on a 370/168 third generation computer. If you are nice, you do five replications so you get five data points for some distributions.

(SLIDE 11)

SIMULATION MODEL

- NO PLAYERS
- MODEL DEVELOPMENT
    - DECOMPOSE PROCESS
    - SEQUENCE EVENTS AND ACTIVITIES
    - SOLVE BY "ACTING OUT" PROCESS
    - DETERMINISTIC OR STOCHASTIC
- COMMENTS
    - MORE ABSTRACT THAN WAR GAME
    - LESS EXPENSIVE THAN WAR GAMES TO DEVELOP AND USE BUT STILL HIGH
    - DIFFICULT TO INTERPRET RESULTS

If you want to vary anything, you talk about hundreds of days and almost years to do a study. So we're not very efficient doing the studies. And also, their results are very difficult to interpret. When people build simulations, when I say they act out the process, they really put everything in they can think of. They throw every possible variable one might think of into a simulation, and in great detail, rather than trying to do some aggregations.

Next, (slide 12). The third generic type of model category is what I would call analytic -- models which I think most of us tend to try to build. There are no players again, and the model development process is very like the simulation. You decompose it to try to understand how it operates. You may even sequence the thing; but rather than acting out the process to solve it, what you try to do is build analytic structures for a lot of the events and activities or aggregations of them. Maybe you build a system dynamics model. You assume it could be described by a linear-programming structure, or maybe it looks like a large-scale set of differential equations could represent these dynamics. Then you literally build little analytic structures of pieces, and you stand back and make one big assumption: "I think it all goes together by this formula -- this integrating mathematical structure." You solve it by mathematical operations if you're lucky. Most often you can't do that, and you use numerical procedures. That is, you may use numerical integration techniques if it is an integral equation structure. You may use Runge-Kutta solution techniques, or a multitude of others like it.

I'm sure everybody here knows, but most of the practicing community does not, that analytic models can either be analytic or stochastic. That is, you can, in fact, use probability distributions as input and get as output probability distributions analytically, mathematically. As a simplistic example, if I want to know the sum of two random variables, $X + Y$, for most random variables I can get the distribution for Z, mathematically. Most of the community doesn't understand that for some reason. These models are appreciably more abstract than simulations, and less expensive to use. You don't have to replicate, for example, the stochastic models. Analytic models are usually much quicker to run. They run quickly compared to large-scale simulations. I think their results are easier to interpret; if for no other reason than because you can look at the equations and say, "I think I understand what is happening in the equations." Now the equations could be wrong, but at least you can interpret them to see why the results are occurring.

If the user is analytic (which he should be), he should be able to interpret them too. My impression is that many users as well as developers in our community, are not technically capable of doing (or understanding) the mathematics at the level required for model development or use. Developers and users. I base that, Saul, as you know, by looking at the complaints in the ORSA JOURNAL, for example. They can't read the journals because they can't do college level mathematics, let alone graduate level problems in random processes and other sorts of mathematical logic. They just don't do that. The problem doesn't arise because the models are ill-structured, but because generally users are ill-prepared to understand them.

(SLIDE 12)

ANALYTIC MODEL

- NO PLAYERS
- MODEL DEVELOPMENT
  - DECOMPOSE PROCESS
  - DEVELOP ANALYTIC DESCRIPTIONS OF EVENTS, ACTIVITIES, OR AGGREGATES OF THEM
  - INTEGRATIVE MATHEMATICAL STRUCTURE
  - SOLUTION

    MATHEMATICAL OPERATIONS

    NUMERICAL PROCEDURES
  - DETERMINISTIC OR STOCHASTIC
- COMMENTS
  - APPRECIABLY MORE ABSTRACT THAN SIMULATIONS
  - LESS EXPENSIVE TO USE
  - FACILITATE SENSITIVITY ANALYSES
  - EASIER TO INTERPRET RESULTS

Next, (slide 13). We have been developing another form of model at VRI -- the hybrid-analytic simulation models. There are no players, again, but a mixture of simulation and analytic techniques. Where you know a lot about the process, you should attempt to describe it analytically. It's only when you know a little bit about it that you generally simulate it. The degree to which you mix analytic and simulation techniques depends on the level; that is, it varies with the model and the kind of development area. For the small unit model in the defense area, generally we can analytically model attrition and acquisition. We have lots of data on those subprocesses; and I've also shown that target allocation, line of sight, and terrain characteristics can be modeled analytically. For example, target allocation may be described by a set of differential game kinds of concepts. We can do nice analytics in attrition and acquisition, and you will notice that those are physical processes, not much behavioral stuff there. We generally simulate movement, more often than not the environmental characteristics, and force and target allocations, behavioral activities; also communications which is basically behavioral, although there is some physical communication. We attempt to simulate those processes. On the large unit, we try to go more analytic because we want them to run faster, and there are lots of systems. We tend to model analytically attrition, acquisition, and some of the behavioral activities. We still simulate movement and command control. We don't know very well why people move and how they make decisions in command control. We model these processes by what are called tactical decision rules, and some of the newer techniques allow the user to vary these behavioral activities very readily just like other model inputs.

There are two types of hybrid analytic/simulation models: what I call free standing, or independent, and fitted parameter. The first one just runs by itself. The other requires you to run a simulation to estimate parameters for the hybrid model. So you may run a Monte Carlo simulation to estimate attrition rates for use in an analytical model of combat.

All of this has been somewhat background material. I want to talk about the problems in developing the models, about what I call conflicting considerations and implications in model building (slide 14). First item, the combat and military processes are very complex. The military for 200 years has said "Our processes are complex." I'd like to say in an analytic way: there are literally tens of processes and thousands of variables describing them that can, in fact, influence the output significantly. Sometimes small variations in parameters (even in different types of models for the same process) produce significantly different outputs. It's a complex process which suggests, as a builder, you have to put everything into it. So you are led to build simulations and war games. On the other hand, there is also very little data to build these models -- that is to understand the process by which to build them, much less to use the models. Therefore I suggest that the models should not be used for what I call evaluations. They're not, in fact, "verified." (The reports have used "validated." I always reverse those two words; validate to me means mathematical consistency, verify means to produce what the real world's going to do.) With very little data, and since you can't verify those models very well (although you can verify some of the pieces), they should not be used as point estimates for what is going to happen in the

(SLIDE 13)

HYBRID ANALYTIC/SIMULATION MODEL

- NO PLAYERS

- MIXTURE OF ANALYTIC AND SIMULATION TECHNIQUES:
  VARIES WITH MODEL LEVEL

SMALL UNIT:

ANALYTIC:	ATTRITION, ACQUISITION, ... , (TARGET ALLOCATION, LOS)
SIMULATION:	MOVEMENT, LOS, FORCE AND TARGET ALLOCATION, COMMUNICATIONS, . . . ,

LARGE UNIT:

ANALYTIC:	ATTRITION, ACQUISITION, TARGET ALLOCATION, LOS, COMMUNICATIONS, INTELLIGENCE, . . .
SIMULATION:	MOVEMENT, COMMAND CONTROL, ... (COMMUNICATIONS)

TYPES

FREESTANDING OR INDEPENDENT

FITTED PARAMETER

(SLIDE 14)

CONFLICTING CONSIDERATIONS AND IMPLICATIONS IN DEVELOPMENT OF MODELS FOR DOD PLANNING

1. COMPLEXITY OF COMBAT PROCESS ⇒ SIMULATIONS / WAR GAMES

2. ABSENCE OF DATA TO VERIFY COMBAT MODELS ⇒ ANALYSIS, NOT EVALUATIONS ⇒ ANALYTIC MODELS

3. REQUIREMENT FOR EVALUATIVE STUDIES ⇒ SIMULATIONS

4. RESOURCE CONSTRAINTS ⇒ ANALYTIC MODELS AND/OR ⇒ SIMULATIONS WAR GAMES

5. REQUIREMENT FOR HIGHER ECHELON EVALUATIONS ⇒ HYBRID ANALYTIC/ SIMULATIONS

6. USER UNDERSTANDING ⇒ WAR GAMES / SIMULATIONS

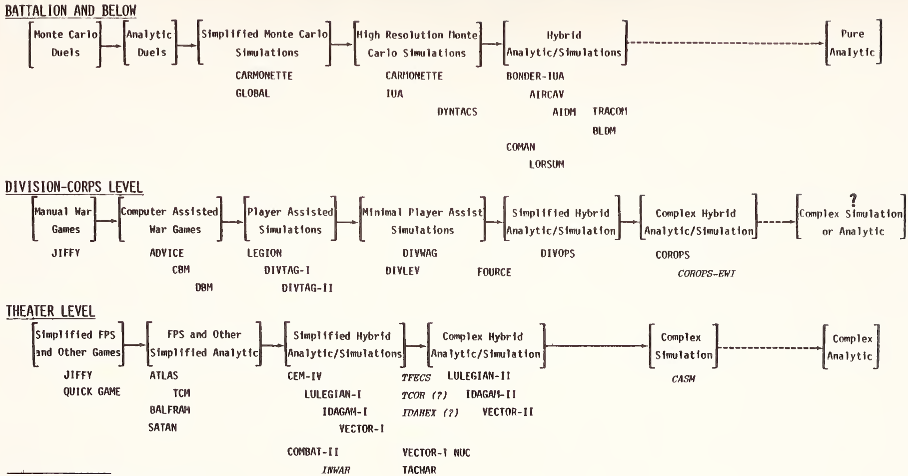
future. "You ought to buy nine airplanes because that will win the war in Europe." They ought to be used for what I call analysis -- to get some insights of where the rough trade-offs are, where the high marginal returns are. You do that by lots of sensitivity analysis; and if you want to do a lot of sensitivity analysis, you don't use simulations or war games, you go to more analytical or hybrid-analytical structures. However decision makers want evaluation studies; they want numbers. That drives you back toward the simulation mode, which means long running times, difficulty in using them; but they want numbers. However there are resource constraints, in both building and using them and, in fact, on technical capability, not only on numbers of people and time and money, but also on ability to develop them. Analytic models are developed quicker, and you see them quicker; but a different level of intellect is required to build them, i.e., mathematical capability if you like, and some ability to integrate lots of data intelligently. There isn't much of that around. There really isn't. I'm not trying to sound egotistical; but you know I teach modeling, I do studies, I observe it, I critique it; and there just isn't much around.

There is a requirement for higher-echelon evaluations, not only in a small unit. There is an interesting phenomenon that occurs. We model the small unit better than we do a larger one, because the processes tend to be more physical. When we are into the physical processes of ballistics, of destruction, physical destruction, we can run experiments. We do very poorly on the behavioral activities which are what drive a lot of the higher echelon operations...command control, movements. So I put a question mark there. I have since filled that in, and I now call that hybrid-analytic. And the last one says look, models will not be used unless the guy who is going to use them understands them. There is very high correlation between the two. The more you build analytic models the less they tend to get used, until you get a bright new user community. That is, unless they see their horses running down the battlefield, they don't like to use them. He's laughing -- I'm serious. If you write a set of differential equations, the user doesn't see his horses and therefore he says he won't use the models. It's taken literally ten years from the day that John Honig's shop sponsored some analytic work I did, until people started to use the models now -- after ten years of comparing simulations and analytic model results. They use the analytic and hybrid-analytic simulations now, but they check back to simulation often.

Let me give you a summary and then make a comment on the GAO survey. What I'm trying to show you is some trends, really, and not requirements on developing models; and I believe it has an impact on what I think is the myopic view of the GAO report. In this slide (slide 15) I show you three levels; battalion, division and theater. And there is really a set of trends that evolve through all of those. Harvey Wagner recently told me it's also evolving in the inventory area. Let me show you how we go about building models in an area over a long period of time, an interesting phenomenon. We start out with very simple models. For example, in the battalion area, we started with very simple Monte Carlo, one on one, lots of random numbers. Clint Anker came along and said he could do that mathematically, and he built the theory of stochastic duels. This was back in the '50's. We started to move (we recognized that didn't quite solve the larger problem of battalions

(slide 15)

SUMMARY OF TRENDS IN US



[1]Models in italics are not as yet developed.

and forces above that), and we moved to simplified Monte Carlo simulation. CARMONETTE and GLOBAL are some examples. We went to very, very detailed high resolution simulations -- CARMONETTE, IVA, and then DYNTACS, probably the most sophisticated, complex one. Notice the differences. A lot of them are complex but not sophisticated, Monte Carlo simulations at the battalion level. DYNTACS is the one that takes about an hour for a replication to run. And then we started to learn by observing what took place there, how to come back out of it. We went from a little bit of detail, down to a lot of detail and back out toward analytic descriptions until roughly we were here now with a lot of hybrid-analytic simulation structures, developed over the last roughly five to seven years; based on observing what we thought was the real world in those detailed simulations. I think there we are moving out to really pure analytic models. That is, I think there is a step over the last year or two, that I haven't shown here, where we can pretty much replicate the results in the higher resolution simulations by fairly straightforward mathematical structures that we have now learned how to build. So we've gone from simple analytic to higher resolution simulation back out. We've done the same thing roughly in the corps division/corps level; I won't go into that. I want to make some comments here, though, on the theater level.

There was a very simple analytic model back in the '50's, something called JIFFY. JIFFY uses a firepower-score concept, which two years ago I gave the name of the "phlogiston theory of combat". That is, it really is idiotic, and it's been in great disrepute. We started with a simple analytic, a little more complex analytic, and moved to analytic simulation techniques. We are getting into very high resolution simulation techniques at the theater level, not stochastic, but deterministic ones. The next generation will be something called CASM, which is supposed to be a very high resolution simulation. This period of time from year to year is 25 years, and we are just learning.

The point I want to make on the GAO report is that it is very myopic. We should separate development of the models from the studies to begin with. If you think about it, developing models is, in a sense, developing descriptive theories about the processes, and this is distinct from decision issues about the process. We should try to understand the processes! We should be very careful about trying to legislate standards for getting validated, verified models. I can't imagine government intervention would have sped up the process from Plato to Kepler to Einstein in any way to get verified theories in physics. I think this is a long process to try and understand via experiments. A lot of the models have verified submodels of helicopter activities based on experiments, of artillery based on experiments, etc. We never run large scale wars to check the whole thing out. I think we are talking multiple, multiple years and to try to legislate the creation of verified, validated models in a period of two or three years, is nonsense. I think it's clear that we ought to control the redundancy -- there ought to be redundancy; but it ought to be monitored and controlled in an effective, scientific way. We ought to have the models used only for intellectual purposes until we can get some good verified ones, and continue to build data bases and new model structures that seem to predict better in the real world.

ISSUES FACING MODEL DEVELOPERS - II

Dennis Meadows

Today we must deal with the fact that there is an extraordinary diversity of modeling methods in use for policy assessment. This diversity affects two aspects of model building. One is inculcating a set of professional competences and standards in the individual model builders. The other is tailoring a model to the needs and resources of the large corporate and public bureaucracies who are typically their clients. Today, I will evaluate the GAO proposals as they relate to the first aspect. (See Appendix for the GAO proposal.)

Since the early 1970s, I have called for improved standards in the modeling profession, particularly at the interface between the model builder and the model client. No one denies that modeling efforts are highly variable in their quality, and that much money invested in the construction and analysis of models is completely wasted. However, now that I see a concrete proposal by GAO to improve the quality of modeling, I begin to anticipate the difficulties these particular standards might engender. While I do not agree with the GAO standards, I share many of the concerns that led to them. I believe that GAO has misconstrued the nature of the problem and has put forth proposals that will simply stifle the symptoms rather than solve the underlying difficulties.

I will put the GAO proposal in perspective by listing the several images of the modeling process that are implicit in the GAO report. If you agree that the images are incorrect, then we should move to find appropriate substitutes for the GAO recommendations. But it is important that we provide some concrete alternatives, for the current state of the field is quite unsatisfactory.

Of course, my own views are substantially influenced by the context within which I practice my art. I carry out modeling in an academic milieu; thus my work differs from that of the rest of you in two ways. First, there is the possibility of doing model-based research in a somewhat more leisurely fashion at Dartmouth than would typically be the case. Our group does not have a fast response capability, because a large segment of our productive capacity is composed of students who are locked into a rhythm of course work and thesis research. Students do not just sit around ready to be called forth like a troop of workers as soon as a new client walks in the door with a contract under his arm. Thus, I have had to seek out those programs which offer the prospect of long-term funding. As a consequence, my views probably differ from those who key their modeling to the short-term demands of policy makers.

QUESTION: Isn't it frustrating if the student isn't around anymore when the insight comes along?

MEADOWS: No, because though our modeling often goes through cycles of five or six years devoted to one set of closely related topics, each student is actually engaged in a specific modeling effort that does have an identifiable client and addresses a specific set of questions. Though our group has not worked on a large variety of issues over the last eight years, we have produced about forty discrete models and around 150 model-related reports.

QUESTION: Is there usually one process area and one general subject area?

MEADOWS: All our work deals with the general topic of population-resource interactions. The effort includes studies of land use, zoning, energy supply and global modeling.

QUESTION: My experience is that learning about a particular process does take time. I'm surprised that you have been supporting 40 separate kinds of structures with ten people, and can develop the necessary depth of understanding of their context.

MEADOWS: There is a great deal of overlap between our models, and each one builds on the work that has gone before. It would certainly be impossible for our group of about 15 people to construct 40 useful models in 40 completely unrelated policy fields.

To exert control over a process one must understand its properties. Let me list several properties that the GAO report seems to imply characterize the field of modeling.

There seems, first of all, to be implicit in the report the notion that there is a one-to-one relationship between the modeler and decision maker. Of course that is typically false, at least for longer-term issues. Often the person who is the source of money for a modeling effort is not a decision maker at all. This is particularly true in the public sector. For example, NSF actually has nothing to do with the decision maker likely to be effected by the modeling work they support. The Foundation's staff may even have been forbidden to talk to the decision makers, much less to make decisions themselves. In the agencies like DOT and DOE, the people who support and monitor models, provide funds, and who would presumably be responsible for implementing the GAO standards, are typically not decision makers. At best they are analysts who may conceivably have some input to decision makers but many analysts in Federal agencies do not have any input to the decision making process at all. Even when they do, their decision maker client is typically not the only one responsible for responding to a specific problem. And even if he were, his decision would be based on many other considerations in addition to the output of a computer model.

Thus, GAO standards should not be formulated as if they apply to a single analyst who is developing a model for a single decision maker who will rely solely on the model-based recommendations.

Before pointing to another conception of the modeling process which seems implicit in these standards, I should say that I have enormous respect for the GAO staff. The problem they are addressing is an important one, and their past work in this field has typically been among the best available. Their organization is one of the first to which I turn when locating positions for any of my students who are seeking an internship in Washington. Though I am critical of their draft report, I readily admit that I do not have a comprehensive alternative set of standards to propose. I only hope my comments will be helpful in stimulating all of us to think about what a better set might be.

The second image of modeling which I consider inappropriate is the idea implicit in the GAO standards that a modeling project can be started and stopped on short notice and with no negative long-term consequences. To abide by GAO proposals, I would have to complete phase I of a modeling project, then I would hand in my documentation and pause while some group deliberates to see whether I pass the various tests. If I do not pass, I must start over or shift to other work. If I do pass, then I get a little money and authorization to start activities within phase II. The assessment of any large model is likely to be a lengthy process. Thus significiant time would typically elapse between completion of phase I an initiation of phase II. During that time, there would be no certainty about the prospects for follow-up support. With that uncertainty about some funding, I would certainly lose the best of my staff. I nurture on outside money, an infrastructure including xerox machines, secretaries and programmers. I must have continuity. Nobody gives me program money to support staff and idle time. I have to sell every hour so that it can be supported from contracts. I would even find the GAO proposal ethically untenable. I have students who are dependent on me for support. When a student applies for admission to my graduate program, I must commit two years of support. A funding process that can eliminate projects without substantial advance warning could leave me unable to satisfy obligations I have incurred to support staff and students. Perhaps the total program rather than specific modeling projects should be the focus of monitoring and control.

Gary Fromm made the excellent point that the GAO report implies a model is a static entity. Do it once and it is finished forever. Nothing further is required. I have never seen a model that matches that description. I know now things that I could do to improve every single model I have built in the past. Our models are in a constant process of evolution. Indeed, we have trouble freezing the modeling process long enough to capture some one specific version of the model comprehensively on paper. This notion of a model as a dynamic entity has to be better incorporated in the GAO standards.

A fourth image implicit in GAO's draft report is that model deficiencies arise from errors which are foisted off on the Federal bureaucracy by those actually doing the work. In fact, many modeling problems are attributable much more directly to members of the Federal bureaucracy. It might be more useful to define a set of standards that would have to be satisfied by any government bureaucrat before he was given money to fund model development and use.

There is general appeal in being a program manager with six to ten million dollars to give out for modeling research. To be the source of support for a large, computer-based analysis effort accords status, raises GS rating, and secures warm, personal attention from potential contract recipients. For many Federal program managers, supporting modeling is an end in itself. I suggest that would-be managers should have to pass a certification test. I have dealt with some program managers who actually have had a personal understanding of modeling and who showed greater sophistication in assessing a model's strengths and weaknesses. I have also dealt with program managers who simply have had no clue whatsoever about the modeling process and its application. By setting standards, imposing a specific problem focus and ruling on the boundary of the modeling effort, such program managers seriously compromise modeling research. I suspect that the professional skills and standards of model builders today are much higher than those of the typical model buyers.

Another image promulgated by the GAO text is that there are no professional modeling standards today. That is simply not true. There are professional standards. They are not uniform, nor are they universally shared, but the best econometricians know what constitutes good econometric research. The preeminent system dynamists recognize the work of other leaders in their field and perceive when others are doing shoddy work. The same goes for input-output modelers and for practitioners of other techniques. Unfortunately the ability to transfer modeling standards across methods is very low, but the standards do exist. The problem is that they are seldom implemented and enforced on those carrying out the work, but we do have some basis on which to build.

Another idea is that modeling is analogous to solving a mathematical problem, that there is one right solution to the problem. I suggest the analogy of painting a picture, is very much more appropriate. A model is a portrait of reality. There are many different ways to paint the same landscape. Model validation is a bit like the effort to bring Rembrandt, Van Gogh, and Miro to agree on one style of painting. More allowance should be left in any modeling standards for the fact that tastes in models can vary.

Another assumption implicit in the GAO suggestions is that documentation is carried out only at the end of the model project, in order to let others know the mode and results of the analysis. To the contrary, the most important part of documentation is the standards employed in reporting progess on the analysis throughout the project. It should not only be possible for the client to find out how a finished model works, he should be able to go back into the records of the project and find out why it did not work in midstream. That kind of documentation is seldom mandated, but its implementation could have an extremely important impact on the quality of work. Their effect would come not so much because documentation standards let others find out about mistakes in completed models, but because the threat of potential discovery could automatically cause analysts to upgrade their analysis and exert more care throughout the design of the model.

Next is the notion that model is designed always to have some impact on a decision. As we all know, that is typically false. Many models result from an implicit partnership between someone who has to spend money this year in order to justify his next budget request, and someone who likes to build models as a rather easy way of earning his salary. It is a liaison between these two, each of whom satisfies the other's needs, which may call forth most models that are actually built by the Federal government today. Other modeling projects are simply initiated to justify some previously derived decision. These may be legitimate and useful purposes. They should be acknowledged in the design of any modeling standard review procedure.

Even where implementation is the goal, it does not take place as implied by the GAO text; the model is not transferred in isolation to the client's computer for his use. Let me cite just one aspect of the implementation process we've employed at Dartmouth. Every summer we conduct a two-week seminar designed to teach policy makers and other clients something about the underlying methodology of model development and use. The conference does not make its participants into competent modelers, but it certainly makes them into rather well-informed model users. Each seminar draws on models that we have completed through our past work, so we are still implementing models that were built and paid for several years ago. Indeed it may be this latter stage of use which is our most important accomplishment, one that occurs long after the GAO standards would cease to apply.

QUESTION: Do you mean by a modeler one of those whose place and status has been enhanced by handing out money?

MEADOWS: No, I am talking about people who are in a position to be influenced by the results of a model, the day-to-day decision makers. In the case of our seminars, we typically would get planning officials from industry, staff people from the relevant Congressional committees, program managers from ERDA, staff people from GAO.

QUESTION: Were some under the impression that they were taking a course which carried status with it?

MEADOWS: Certainly they were. And some gained little more ego satisfaction. But in many instances, there actually was significant learning. The purpose of the course was to create. The client often implements the model, not by acquiring it and using it personally, but by hiring one of our students to join the agency responsible for decision making.

QUESTION: I don't know how long you have been doing this but it is a matter of some feedback. I started doing that in 1965 with people who were generally middle level managers. That has a tremendous impact about six or eight years later when they rise to levels of Assistant Secretaries or comparable. They are really knowledgeable on what to do and what not to do.

MEADOWS: And that is the point of the exercise. The process of conducting a useful modeling effort does not cease at the point where the GAO standards would stop. Indeed, the major impact of a good modeling effort may only start at that point.

Because I think the GAO analysis errs in its conceptualization of the modeling process, I do not believe that implementation of the GAO recommendations will much improve the real quality of models. It may simply tie up the good modelers in generating a great deal more paperwork. There is nothing about the GAO's 25 steps that can magically call forth increased ethics. Quite the contrary; it may diffuse the efforts of those working hard to create a first rate product. If we really want to improve the quality of models, we must first of all recognize that the problem lies on both sides, and that program managers themselves must adhere to certain standards. Then we might consider certifying modelers rather than models. This would be analogous to the practice in other professions, such as law and medicine. We can certainly identify areas of knowledge which any literate modeler should have mastered. These topics should be brought systemically into educational programs which should come to be certified much as law and engineering curricula are today. This will be a tedious process which impacts on the quality of models only gradually, but I think it is far superior to the implementation of comprehensive third party assessment. For one thing, it takes a relatively skilled modeler to make a perceptive analysis of another modeling effort. Good modelers are already in short supply. We should not tie up half of them in the analysis of work being done by the other fifty percent. This is not to imply that nothing can be done through independent assessment. Mechanical evaluation by a third party simply to certify that certain kinds of information are available about the model to an interested party could be easily accomplished and would be very useful. I therefore suggest turning our attention to the nature of documentation standards while working to identify the character of enhanced modeling education programs.

ISSUES FACING MODEL DEVELOPERS - III

Dan Maxim

Reading a book called the "Trenton Pickle Ordinance" is a delightful way to spend an evening. The book is a collection of laws that have been enacted at various times throughout the U.S. One of my favorites is a law to the effect that it is illegal to wake a sleeping polar bear for the purpose of taking his picture in Alaska. The interesting implication is that there are people that you need to tell that it is inappropriate to wake this sleeping polar bear. I hope you won't find my comments equally banal or unnecessary.

Let me offer one or two basic points. The first is directed to the people who are contracting for models: you can help improve the process of model development and implementation by direct participation in the process. By this I don't mean refusing to give out money unless there is a project plan with appropriate milestones and briefings. I mean being an intellectual architect and direct participant. I personally believe very, very strongly that that's one of the ways in which you will get significantly better products.

QUESTION: Do you see that in a public agency?

MAXIM: We have two or three fair-sized contracts right now where that's taking place and they are all in tremendous shape. In fact, if I had to single out any one variable that I think is probably the most important to project success I'd name direct client participation. This participation should be as a worker, not as a monitor or someone to go out to lunch with or to listen to briefings or whatever, but someone who actually is expected to produce major work elements. One of the best people in our firm, Frank Cook, was employed by a major aerospace firm years ago -- I won't tell you the name because what I'll say might be to some degree unflattering -- but in any event they made airplanes. They literally made the airframes but they would subcontract out the avionics. The engineers who were designing the airframes really knew their business, were pleasant and easy to deal with, and they were fully competent professionals. The engineers who were in the avionics end of things started out being no more or less competent than their structures counterparts, but then they spent a lot of time going to lunch, reading catalogs, subcontractor reports and so forth. Five years later you could detect the onset of incompetence and paranoia. Ten years out the paranoia was fully developed and there was just no point whatsoever in having them, I think. The same phenomenon is likely to hold true in the models business: contract monitors who participate in the work are likely to be better people for that participation.

There's a corollary to the effect that you shouldn't let senior people in your own firm be administrators all the time. There is tremendous pressure in a consulting firm for people that are articulate and competent to be the 'front men.' Then after a time they lose touch and don't know what they're selling any more. Proposals are well written but less realistic.

At MATHTECH we have had a policy, a very deliberate policy, to the effect that if you're a senior person you work. This policy may have some negative implications in terms of the short-term rate of growth, but it also has, we think, significant positive implications in terms of the quality of the product.

However it is accomplished, I think direct participation is important. It increases the realism of the product. In our case, I should add, we're closer to decision makers than you at NBS or universities might be. The people that are our clients are users. They're also close to people who make policy and so important policy questions are reflected in the analyses. Participation also increases the clients understanding of the model and its limitations. It increases their ego involvement in it. It's not only that they get to give away X millions of dollars, but they're in part an architect of the work and as a consequence, are more likely to implement solutions wisely.

The second set of comments that I have relate to an activity that everybody calls by different words. Norm Agin from our corporation calls it "intellectual post processing," which is, "after you've completed the initial analysis, then what?" What are all the checks you put it through, the validity and plausibility checks, sensitivity analyses, a fortiori analysis, break even analysis and related ideas. There's a lemma ascribed to the economist Will Baumol which goes, "All budgets are big at the beginning." Regardless of the amount of the contract, it's large when you start out. The tendency is to start things off by saying you're really going to do this one right. You spend a lot of time spinning your wheels, and then as due dates get closer and closer, the intensity of the work picks up, more and more assumptions get made, and very often, what gets eliminated or hurried in the process is this whole topic of "intellectual post processing." These activities somehow don't get done because after all, there isn't much paperwork associated with them anyway, and there is the need of getting the deliverable product.

I think it might be helpful if we could better discipline overselves as consultants or contractors. I have tried many, many times to discipline myself with mixed success. Perhaps a way in which that problem could be made a little bit simpler is by planning projects which are done in two phases, or at least have significant milestones in them, that force you to do certain things by certain times and allow you sufficient budget so that you will get to do this "intellectual post processing" which is perhaps the most important thing.

There's a book by Townsend called "Up the Organization." It's a very interesting book in many ways. Like many works, it contains lots of mutually exclusive propositions that are asserted with equal vigor. It's very entertaining, and much of its contents are hard earned wisdom. One of the things he observes in dealings with accountants in particular, is that they are asked to prepare various financial statements under extreme time pressure (while board meetings are in progress, for example). Townsend suggests that the analysts get a stamp which reads "Prepared under pressure and not fully understood."

It seems to me that if we all were prepared to be a little more candid and stamp some of our preliminary reports in this way we would get improved quality of decision making.

I'd like to underscore the "documentation now" point that has been made by Dennis Meadows. Document as you go along as opposed to at the end, if only because it increases the probability that you will have the time to do the "intellectual post processing" that you need to do if you don't feel under pressure to write this report. If you have a good collection of working papers, writing a report is often simply a matter of piecing these together. Therefore, you can have some time to think about what has been done and how it can be made more useful.

Well, I could go on with things that I think in the main are obvious, but if you just took some of those points to heart I think it would be an improvement. I know Saul Gass is concerned about keeping to schedule so I will be mercifully brief.

QUESTION: It seems to me that the guideline about being involved with a client is a good one. It stands out every time anyone sits down and listens to ways to be effective. I would point out, though (I'll make the statement to see if anybody can do some verification of their experience with somebody else's), it's virtually impossible to do that with ERDA, it's virtually impossible to do it with NSF. Those are two major sources of funding for pretty important kinds of models, and it's not always an option open to you to do that.

Now you can run through a contract in a way which involves you with some decision maker, although it becomes difficult to find anybody that assumes they have any responsibility in the area of energy, but it isn't possible in ERDA. There's no programmer of energy there, who can sit down and work with you on the model. You're really lucky perhaps if you can get a guy up for a day once a month.

MAXIM: Yes. Our experience with ERDA contracts has been similar although we have gotten participation, but oddly enough not from ERDA, even though it's been ERDA funded. ERDA supplies personnel from Argonne and Oak Ridge and various other such institutions to help monitor and participate in the work. This is one approach that has been successful. In general, I'm not entirely sure it's something that we as consultants can do a great deal about. But we can strongly recommend this policy. Many people in the audience are or will be involved in disbursing funds in one way or another. I think that if they took this suggestion seriously, they're likely to wind up with significantly better products. I say this fully aware of the other pressures that you face. As one more demand on your time, it must be evaluated in the context of other priorities. But consider this, if you don't think it's important enough to assign someone to, maybe the project isn't that important in the first place. It appears the real resource system constraint is not so much the financial resource, but the availability of people within the government to participate in these things.

QUESTION: I think, at least in DOD which has had a lot of experience with models, and now a lot of analysis organizations with similar technical people you expect a lot more participation. You get a C.T.O.R. whose job is to sit back and say "I gotcha" occasionally, but never really stick his nose in. Some of them are very good technical people. We have got to come up with a mechanism that requires C.T.O.R.'s to be technically capable and to participate at a level which you describe.

MAXIM: I don't think you can require anyone to be technically capable. You can surely require them to participate. I think that if they do participate and they are the sorts of people you want to hire, they'll be embarrassed into maintaining technical competence.

QUESTION: You have to be careful. I can think of one sponsor, it's a little side story, who I asked when doing a project with, "Do you know anything about this business?" He said, "Oh yes, I took a course in O.R." I said, "That's good, what did you take courses in? Did you have anything in mathematical programming?" He said, "I know all about that FORTRAN stuff."

MAXIM: We've heard that one before.

QUESTION: One area of concern with that -- I can see several areas of concern about delegating. There are some obvious ones. This may be the one sharp technical guy in the shop, and you hate to lose him. If this guy turns out to be the one who understands about the big model you've contracted for, he may be stuck within the organization on promotion lists. Finally, there is the sort of suspicion that where the modeling firm says, "Lend us one of your good people." What is in fact happening is that you are being co-opted from subsequently being critical of the product that's being used. So there are all these considerations which don't reverse any points of what you said but have to be weighed with the consequences.

MAXIM: I agree -- but perhaps if they do participate there will be less to be critical of -- so the co-option issue is less important.

QUESTION: I just want to make some comments. I agree with Dan, and say that the issue is that there should be a staff change in ERDA rather than to say we'll keep building models the way it is currently being done. That is one of the key issues in the whole discussion here, in the Workshop, and that is why they aren't being used more. And one of the reasons that they are not being used is this business of involvement and we should have a definition of success or failure set by whether they are used or not used.

MAXIM: Perhaps this should be reflected in the GAO guidelines.

QUESTION: One other thing which I think is just opposite to something that John said. Sometimes government agencies are structured in such a way that there's a difference between the user and the decision maker. In some cases, the users may participate. They may be part of one agency which contacts our R&D office which responds to users within the whole agency. The R&D people may participate, but that still doesn't mean that it is going to be used up the line within the government agency. So that there are some problems with organizations within the government agencies.

QUESTION: I have one small comment to what Dan said, if one is permitted to mention commercial types in this context. Just reviewing the few jobs I would consider unqualified successes, every one of them had a client personally committed to the organization throughout the job. The one which was most successful committed a vice president for a full year to the job. It made implementation much easier.

QUESTION: I want to say something optimistic. I said it earlier and I'm surprised that I've been put in the position of saying anything in opposition. Whatever the words, I think it made a difference. I think it improved the debate and the question I ask is what would have happened if they didn't do that piece of work. That is an interesting question to raise. I don't know how you would answer it, but this is a subjective opinion on my part.

Li
ch
at
a
No
a
he
Th
ex

on
th
wa
ci
Th
Th
we
In
cl
wa
ke
th
ge
un
so
de

cha
the
to
sta
way

fro
a p
nat
tha
mod
not
the

ISSUES FACING MODEL DEVELOPERS - IV

Alexander Pugh III

I resisted the temptation to comment on the last several speakers but I like what they're saying and I could take it one step further and that is to change the perspective from looking at a model as an end, to rather looking at a model as a means to an end. In each one of these instances somebody has a decision that he wants to make and he has chosen to use modeling as a support. Not in every instance obviously, for some model for model's sake, but I think a good model is characterized by a situation where a guy asks a question and he wants an answer and he believes that the model is a means to a good answer. This is not one of the characteristics of the GAO report. It looks almost exclusively at the model as being an end rather than part of a process.

I'd like to underscore this by talking about a process my own firm used on several occasions, in which we've gone through the modeling process up to the point of building the model and stopped. We clearly did not know if this was going to be of any use. We had a situation where there was need of participation. The people were not buying a model, they were buying a process. They were sitting down with us as we could help them deal with their problems. They recognized something was wrong and they wanted to get a grip on it, so we helped them articulate their problem, helped them articulate the structure. In that sense, we got the model started as far as a flow diagram. But it was clearly stated at the beginning that if we ever carried it a step further, it was to be introduced as a social model. Here's a case where the structure was key, but the utility of the mathematical model perhaps was zero. Nevertheless, the process was highly useful. They participated from day one. There were generally more of them than of us throughout the process, and they came away understanding the problem area perhaps for the first time and understanding some of the forces that were operative so that they could make some intelligent descisions about it.

QUESTION: How would you define success of a model, or failure?

PANELIST: I think in terms of the ability to influence decisions and change. This broadens things out somewhat. Dennis mentioned that sometimes the influence is not to the first point but rather to a later point. I'd like to target on the first point. In our case the objective, from a practical standpoint, is simply that resources are allocated in a different way then they would have been otherwise.

PANELIST: I think you can go maybe a step farther than that, at least from my point of view. I think it's been successful if you shed some light on a particular problem. Now whether the decision maker wants to use your information and allocate resources differently, that's kind of a separate part of that process. He may say yes, but for other reasons not considered in the model structure or some other reason, for whatever reason I have, I choose not to use that. I think if the models we develop produce some insight into the problem areas, and learning about the dynamics of it, I think it will be be successful. Whether or not he uses it, -- we'd like him to, clearly. But decisions are made; in fact I would point out that the whole premise of this process is that in the Federal Government there is a rational decision making process -- perhaps one model upon which the Federal Government operates. It is not at all clear that it's the correct model, that there is any decision maker or two or three. I think that if models produce some insights for people to look at, it will be for the Government, and be successful.

QUESTION: How do you measure that insight?

PANELIST: Participating in using the model. If you learn some things about the process and you can communicate those to people who make the decisions, that's how you judge it, it's a subjective thing. It's not a quantitative algorithm where you score it. In fact, somewhere I noticed in the report it had scales from 1 to 10 in how to measure certain things. I don't think you do that.

QUESTION: It's not time then?

PANELIST: Oh, no. That's a purely subjective kind of thing. But it's hard to establish that a model has been used, it tends to be very easy to establish that it has not been used. And so long as we're dealing in a field where most models in fact aren't used, then that's a good question to ask for most of them. Let me give you an example. The stuff we did way back when we were talking about the old battle tank kind of program, we did some analysis to show the fact that things ought not to have been done the way they eventually were done. And the reason they were done is because there were political agreements between the Republic of Germany and the United States about which company they would get to do the job. That was a separate issue from who should build it from -- the structure of their system, and how they compared. It was a political agreement. Okay. And new models don't account for the political agreements. They assess the candidates' systems.

PANELIST: One extension on what you said, I think, is there are many instances, at least quite a few instances where the output of a model does not provide insight but are definitely used incorrectly. There you really can't fault the modeler because he has done the best he could, but they are frequently used out of context incorrectly and do not provide insight. That's -- you can't fault him for not being successful.

PANELIST: I think that comes back to the issue of how close is the model to where the client wants to get it. I can think of some clients that did what I considered silly things but he would instantly agree with me, that in terms of our mutual understanding that was silly. He had some other reasons, political, for example, for doing it that way, but he knew he was flying in the face of our mutual understanding of what was going on.

PANELIST: But you see, one department can develop a model and fully understand it and somebody else -- let's take the participation of the decision maker. The closer they get together the higher the quality of the result, in terms of the success we are speaking of here. The further they are apart and in the military situation where a model gets started by one group and others, I assume it was a hundred percent change in personalities by the time the model gets implemented, --

QUESTION: You mean delivered --

PANELIST: You can use the word delivered. But our rotation scheme is going to assure that. To bridge that large a period of time, puts a hell of a burden on the model for instance, for success of that model in production, they must be that large --

PANELIST: Speaking to the last point about insight being lost by this transfer of personnel in an organization, or whatever, that can have implications with respect to documentation. There is a question as to whether or not one needs sort of defensive or preventive documentation as well as problem documentation, and so you need to say not only what the model does, but what is doesn't do that some fool might think it does. And sometimes perhaps some modeling alternatives that were considered but dismissed, and why.

e[illegible]
an
se
sc
ar
th
un
of
a
pa
pi

th
ar
De
ti
va
sc
de
ve
ne
pe
ba
po
cu
an
me
ri
ex
sh
ti
ho
ge
ge
th
tw
So
ex
de

wo
to
wi

MODEL IMPLEMENTATION

Richard Larson

The topic that I'll be talking about will focus on some implementation experiences. I've been working with the urban emergency services, police and ambulance in particular. Some of my students have been working on other urban services doing similar work. I'd like to extrapolate from a few case examples some properties that I think are required of all OR studies and models that are meant to be implemented. Then I'd like to talk about one particular model that I've been involved with for the last five years or so that seems to be undergoing an ongoing implementation effort, and what I feel indicates some of the complexities of the issue. The implementation process is, I believe, a very slow, long-drawn-out process, and requires a lot of commitment on the part of the model builders, and the user community. Hopefully we will get a picture for some of that process.

First of all, let's discuss three experiences that I've had working with the New York City Rand Institute when it was in its prime in the late 1960's and the early '70's (Figure 1). This was with the New York City Police Department. Very quickly, just to give you an idea of a few of the complexities, one situation occurred in lower Manhattan where it was suggested that various of the police precincts of lower Manhattan should have different scheduling rules for police officers, since the temporal distribution of demands for services were distinctly different, for instance Wall Street versus East Village. So one could gain some efficiency and some effectiveness by switching people around from precinct to precinct over a 24-hour period. The situation lent itself to a queuing analysis, requiring only back of envelope results, and its suggested perhaps switching one or two police officers or patrol units from one precinct to another at each particular time of day. From a system-wide perspective, the Police Commissioner and others in planning research were quite happy with the idea. They implemented this program for a while and found out that the model predictions were right on target. Perhaps the end results were even a little bit better than expected. So the top manager was quite happy. The idea was eventually shelved though, and it was called a "failure" because it was found to be politically infeasible. Why? The Precinct Commanders who were on duty during hours when their precincts had been depleted felt that they were left dangerously uncovered and were not in a position to react on receiving an emergency call. Their own objective function was like a minimax objective, where they want to minimize the chance of the worst possible thing happening, like two planes crashing above their precinct and debris falling from the skies. So here's a case where a model worked, the model was implemented in a trial experiment, the decision makers were happy, but we found that there were other decision makers whose own self-interests were seriously violated.

There's a second example, very briefly, which occurred when I had been working with Rand for about three or four weeks, and my assignment was how to reallocate their 16,000 man patrol force. We had our first briefing with the Commissioner and his staff, and he deliberately sat next to me at

THREE PERSONAL EXPERIENCES

1. DON'T TAKE MY MEN AWAY!

2. WHY SPEND TIME AND MONEY MAKING DECISIONS?

3. WHAT, DESIGN A NONPERFECTLY WORKING SYSTEM?

FIGURE 1

the briefing table, and his remark was, "I hope you have an answer for us now that you've been working on this problem," and I said, "I'm sorry, I didn't. There are 16,000 men doing very complex work in a very complex city, and I'm just getting going on the project." And he said, "Well, I don't understand because every year I assign a limited duty sergeant half-time for two weeks, and after that time he gives me the numbers." By giving me the numbers means that he gave the answers on what to do with the 16,000 men. That indicates the level of effort that they put in to that kind of decision making problem. The 16,000 men had huge dollar consequences per year and yet one person week of effort was put into allocating.

A third example is what happens when you have probabilistic system operation and you have to design the system to satisfy probabilistic performance criteria. An example here, again with the New York City Police Department, is the 911 system, the police emergency number. It was not operating properly. How did they find this out? Letters to the Editor of the New York Times saying, "I called up last Saturday night and waited for 29 minutes with a ringing telephone, hung up an tried again and waited 28 minutes for somebody to answer the phone." The Commissioner had been told, well, everything is working fine, because the interlevel management didn't want to tell him some of the scheduling problems they were having.

So again in a briefing with the Commissioner, I said, "Commissioner, in order to reschedule your personnel we need some performance criteria. Would you like five percent of the calls to wait for 15 seconds, or one percent of the calls not to wait at all, what kind of performance criteria would you like"? He refused to give any performance criteria other than no calls are to have any delay whatsoever. In other words, it was politically infeasible for him to accept an imperfectly working system. We had to change our design and accept certain constraints and reschedule personnel without any explicit policy statements from this decision maker. He was unfamiliar with the consequences of probabilistic operations of the system, and the only way he could get the zero chance of delays was to take everybody out of the police cars and all dispatchers out of the dispatcher room and put them all in this huge telephone switching center, and obviously infeasible alternative.

QUESTION: One of the points here is you should never really try to solve the problem via a third party. If you've got somebody specifying the problem for the Precinct Commander, that is the Commissioner, and the Precinct Commander is the one who is worried about you taking his men away, then you've got a third party you're talking to solve the problem with the guy down here.

LARSON: The Precinct Commander does indeed have authority over those men. They are "his" men AND they are the "Commissioner's" men. So who is the decision maker in this case?

QUESTION: If you want to take the Precinct Commander, you never really get to that problem, because he is interested in the allocation of men to his precinct. He never gets to address the question of whether some other group might --

QUESTION: -- because the guy down there that you are working with is the guy that is close to the problems he has in this precinct and in other precincts, whereas the Commissioner wouldn't say it was his problem because he was governing the thing, and enforcing --

LARSON: But he did say it was his problem. From the system point of view you have 77 independently operating mobile service units and if you could do something to help them cooperate a little bit and make them less independent, you would increase the total system efficiency.

QUESTION: There are other sciences that do address this sort of problem. Organization development people, who work on the control of the economy and who get at the sources of manpower data and do know how to get at the sources of manpower data and do know how to get the Commander of the whole system to perform. They could come in if they could demonstrate the power that I've seen they have. It is a specialized science.

LARSON: That point and others are also demonstrated in other projects we have worked on at MIT in a course called, "Analysis of Urban Service Systems" (Figure 2). One of these was a school busing case in a locality near Boston, having nothing to do with racial balance. In particular, the problem was reducing the school bus budget as it was going up about 20 percent per year. The class used some heuristic algorithm techniques and some back of envelope calculations and they showed how they could reduce the number of buses for one year and save the city about $130,000 a year. The town was so happy with this that they hired the students as consultants for $10,000, and had them implement their ideas city wide, and it worked.

One of the kinds of things that the students found out, though, was that they had to slightly redesign the district line between the north high school and the south high school in the town. The analysis indicated that it would be very efficient to do this. The found out that by doing this they switched the star half-back from the north high school to the south high school in his senior year. This is something that is clearly politically infeasible. That is not the type of thing that you would evaluate a computer model on. No computer model should be expected to "know" these kinds of things, and in the decision making process these kinds of things should be allowed to occur and readjust themselves naturally. There was a little gerrymandering with the district line so that the star half-back was included in his former high school.

This is an example, I think, where a decision maker has to be in charge. The model, I think, should be evaluated more for its assistance to decision making. We've had two or three speakers this morning who have emphasized that in their presentations. I think the word "optimization" is a poor word to use, at least in the public sector applications that I've seen, whether it be emergency services, school buses, other kinds of services, because of the inability for us to define mathematically the objective functions and constraints. We don't know what they are until all of a sudden we start finding things out and these things fall out of the woodwork. So the decision maker, who has an intimate knowledge of the city, combined with the computational

STUDENTS' EXPERIENCES

1. SCHOOL BUSING

2. CLUMPING

3. BULK MAIL

4. MUNICIPAL COURTS

5. AMBULANCES

6. QUEUING IN THE FUEL CRISIS

FIGURE 2

power of the computer or whatever other model he is using (even the back of an envelope), makes a fairly good team.

There are all kinds of reasons for implementation difficulities in public sector problems (Figure 3). I have just been talking about the problems, objectives and constraints. We don't know how to talk about productivity. You have internal resistance to innovation, people who have been in the organization for 25 years or 30 years are now in decision making positions. A lot of these organizations, at least in municipal services, tend to be somewhat insular, perhaps fraternal, and therefore they distrust outside technical experts and sometimes with good reason, given some of their experiences with them. Also, some of these systems are operationally complex and whatever models we have in some of these areas are still in their infancy. They are first cuts -- back of envelope kinds of things. We still don't know a lot about the operation of some of these systems.

There's another part to it, too. Suppose we develop complex probabilistic models that include major operational complexities. The user, whether a police sergeant or hospital administrator, has to understand the system intuitively in order to become an intelligent user. So that makes the learning curve rather steep and requires an ability to conceptualize. It's different kind of thinking and these people aren't used to thinking in terms of their own systems. They tend to think of their system in terms of big events that occurred in the past -- occurred in the extreme, e.g., what happened on October 2, 1952. This kind of thing is not conducive to planning for the average situation.

If we accuse decision makers of not being as intelligent as us, or whatever words we choose, we have already defined a big gap, between "us" and "them." I think we have to be brought closer together somehow, maybe to be forced to live as assistants for half a year or something before we even undertake any kind of modeling analysis (Figures 4 and 5).

Jan Chaiken, who is going to speak next, has been doing some research funded, I guess, primarily by HUD, on the lack of impact of all our types of models used in the public sector (Figure 6). He has found that one of the key model attributes which is related to success or failure of implementation is the data base requirement for the model, and this is true of my work too. I found that even in models which require what I think are relatively modest data bases, which might require partitioning the city up into census blocks, or something like getting data for each block, is a mind-blowing kind of thing for a lot of cities and municipalities. These municipalities can't even get their local computer systems to give them results within four months or six months. I know a lot of cities with police departments that are keeping hand tallies of crime rates and arrests, because the computer wouldn't give them that information for six months or so. A key agency characteristic deals with the allocation process. Like I explained before with the Police Commissioner, who said, "Well, every year I assign a half-duty sergeant two weeks to solve this problem for 16,000 men and allocate them." That decision making process was considered adequate for allocating

REASONS FOR IMPLEMENTATION DIFFICULTIES

1. ILL-DEFINED OBJECTIVES AND CONTRAINTS.

2. LACK OF PRODUCTIVITY MEASURES.

3. INTERNAL RESISTANCE TO INNOVATION.

4. RESISTANCE TO OUTSIDE TECHNICAL.

5. OPERATIONAL COMPLEXITY.

FIGURE 3

SAVAS' LIMITATIONS OF URBAN ANALYSIS

1. PROBLEM SOLVING VS. INCREMENTAL AMELIORATION.

2. DIFFUSE DECISION MAKING.

3. WHO CLAIMS SUCCESS?

4. ANALYSIS IS POLITICAL.

5. ANALYST AS CHANGE AGENT.

6. "SYSTEM" VS. "SUBSYSTEM".

7. "IF WE CAN GET TO THE MOON, WHY CAN'T WE...".

FIGURE 4

SAVAS, AGAIN

8. NO FEASIBLE SOLUTIONS.

9. THE IRRELEVANCE OF TECHNICAL ELEGANCE.

10. UNDEREMPLOYED MODELS: THE MODEL IS PRODUCT.

11. THE PROBLEM-FINDING ELITE.

12. FEEDFORWARD VS. FEEDBACK.

13. THE HOLISTIC ANALYST.

FIGURE 5

THE CHAIKEN, ET AL.
STUDY (1976)

LACK OF IMPACT OF MODELS

KEY MODEL ATTRIBUTE: DATA BASE REQUIREMENT

KEY AGENCY CHARACTERISTICS:

1. MODEL TO REPLACE PROCESS NOW CONSIDERED ADEQUATE
2. VANISHING ADVOCATE
3. LACK OF STAFF PROFESSIONALISM

FIGURE 6

16,000 men, so why spend $100,000 for models to do that. This is a real problem.

Another problem alluded to this morning by Saul and others, is what Jan calls the vanishing advocate problem. If you have a modeling effort which takes a year or more to do, and it is a modeling effort for a local agency, you may even be employed there, you're interacting with them, then you probably have a very innovative person as your contact person. The fact that the person is innovative indicates that he just might be of considerable value to the agency; he has a higher probability of being promoted, transferred, or finding a more lucrative job in the short term. Therefore, it is very likely he is going to be gone by the time you finish. That has happened to me several times and I'm sure Saul can talk about this having happened to him several times.

And then we also have what we call lack of staff professionalism (Figure 6). We didn't really mean a deficiency in administrative capability, but rather a mismatch with the prior training and education of people who are able administrators but not quite at the same level in the technical and quantitative areas. Then sometimes they have but a one or two weeks summer course, and come back oversold. They think a model will solve everything, and they're gong to be very disappointed when they're burned on their first modeling contract.

A colleague of mine at MIT and his graduate student have done some studies in the police area in Boston, St. Louis and Los Angeles (Figure 7). The Boston case was the one I was involved in and that was a simulation model developed for the Boston Police Department. Our contact person there was a former MIT graduate. He got a job offer for twice as much money outside, and he took it. By the time the model was ready to be implemented, a lieutenant was in his place and the model was never used to make decisions. In St. Louis there was a sergeant who was very innovative in the middle and late 60's, early 70's. He was promoted out of planning research and had no more dealings with planning research. So that model, which was good enough at the time for IBM to take and market as a resource allocations package in other cities, St. Louis stopped using because its advocate was promoted out of that position. In Los Angeles, the IBM model which was taken from St. Louis, was applied. The Los Angeles system unfortunately works differently than the St. Louis system, so therefore there was a conflict between the operational approaches to the model. In Los Angeles the model was too embedded in concrete to allow change easily.

QUESTION: Not only can the advocates of the model disappear but the analyst who is associated with it can disappear.

LARSON: Absolutely. This is a particular problem if the analyst doesn't have the luxury of staying with it year after year, because of marketing considerations, staying in business, that sort of thing.

QUESTION: Following the vanishing advocate, what normally happens is he's the innovator and all the guys around him think he's crazy. So when he leaves, one of those gets into the slot, do what I just did a month

COLTON-HEBERT STUDIES

1. BOSTON SIMULATION AND FOLLOW-ON WORK.
(VANISHING ADVOCATE AND MORE.)

2. ST. LOUIS QUEUING MODEL.
(VANISHING ADVOCATE.)

3. LOS ANGELES QUEUING MODEL.
(THE SYSTEM CHANGED.)

FIGURE 7

ago, cut the contract. They don't ever take the chance of trying to destroy it themself. They're the ones who didn't want it to begin with and their boss or their advocate told them, "I want you to do this." Boy, when he leaves you better leave to!

LARSON: The model, no matter how good it is, it may be the greatest model in the universe, it has now a political tinge to it. It is associated with that other guy in the former regime and therefore it is doomed to failure.

Well, quite briefly, one thing that we started about five years ago was given in the fancy name of hypercube queuing model. It is basically used to address spatial deployment and dispatching questions like the following situation. There's a precinct in New York City which happenes to be a precinct north of Kennedy International Airport (Figure 8). It's partitioned into a number of police beats, and these guys are dispatched across beat boundries or within their own beats. There's a guy who center was up in the north central part, but he is dispatched all over the place during an eight hour tour. The idea was to use this spatially distributed queuing model to assist the decision maker in making deployment decisions in this kind of situation.

So we created this model. We've been trying to explain its use and utility to the IACP, International Association of Chiefs of Police, right across the way here, and some others. This implementation effort goes on. The model has now been in the public sector about two and one half years. It was sponsored jointly by NSF to MIT and by HUD to RAND.

One of the problems with a model like this is, it turns out that the spatial distributed queuing model had certain non-linearities, certain complexities involved with it. It had probabilistic reasoning and one of the key things we found that even back-of-envelope reasoning explained to the user community is mind-blowing and is very hard to understand. And to show you how elementary it is, we've had problems explaining these kinds of contexts to the user community. The users are police departments, ambulance services, fire departments. The kinds of reasoning you see in the literature are things like this; doubling the number of patrol doubles the amount of preventive patrol (Figures 9-12). Always a linear kind of reasoning. Preventive patrol is what police do when they are not doing anything else; driving around. Well, you can say, its wrong as it more than doubles the amount of preventive patrol, because there's a fixed time spent on calls for serivce and there's a simple example that indicates that in certain areas doubling your amount of officers triples the amount of preventive patrol. You can even raise it by a factor of 5 or a factor of 10 -- a very simple idea to us but a very novel idea to -- let's say a police patrol planner or sombody designing an ambulance service. The same kind of linear reasoning is applied to travel time where it would be as you double the number of patrol units, you halve the travel time. Isn't it obvious? Well no, there's a square root law involved there which we derived. It would bring your reduction down 30% not 50%. You would have to quadruple your patrol to halve your travel time. These kinds of back-of-envelope things are not automatically understood by the user of a computer model. No matter

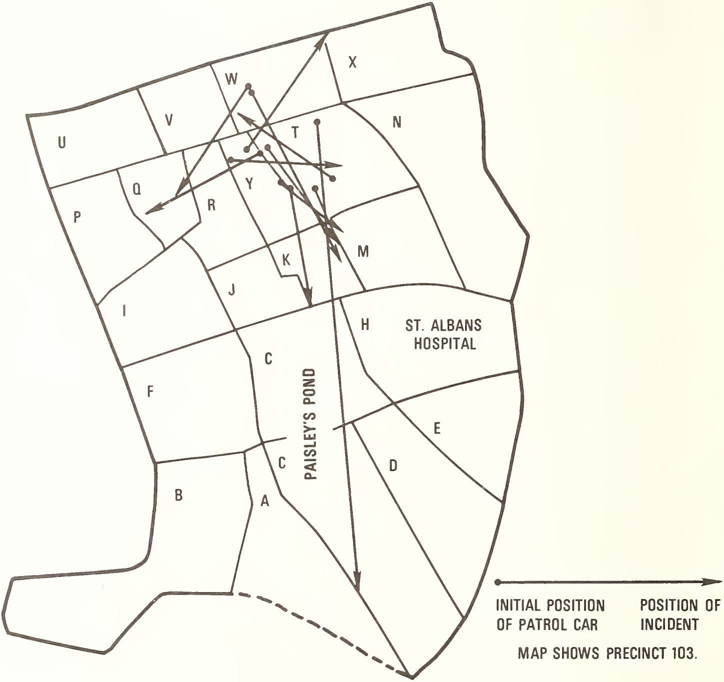


Figure 8 The Radio Runs of the Patrol Car Assigned to the Three Sectors, K, Y, and T

STATEMENT	RIGHT OR WRONG?
DOUBLING THE NUMBER OF PATROL UNITS DOUBLE THE AMOUNT OF PREVENTIVE PATROL.	WRONG; IT MORE THAN DOUBLES THE AMOUNT OF PREVENTIVE PATROL.

EXAMPLE: A) ONE UNIT PATROLS 4 HOURS AND ANSWERS CALLS FOR FOR SERVICE 4 HOURS.

AMOUNT OF PREVENTIVE PATROL = 4 HOURS.

B) A SECOND UNIT IS ADDED. NOW WE HAVE 4 HOURS OF CALL-FOR-SERVICE TIME AND AN

AMOUNT OF PREVENTIVE PATROL = 12 HOURS

(A TRIPLING OR PREVENTIVE PATROL EFFORT).

FIGURE 9

STATEMENT	RIGHT OR WRONG?
DOUBLING THE NUMBER OF PATROL UNITS HALVES THE AVERAGE TRAVEL TIME.	WRONG; IT TYPICALLY REDUCES TRAVEL TIME BY ABOUT 30% (NOT 50%).

FIGURE 10

STATEMENT	RIGHT OR WRONG?
THE FRACTION OF DISPATCH ASSIGMENTS THAT ARE INTER-BEAT DISPATCHES IS USUALLY SMALL ENOUGH TO IGNORE IN MOST CASES.	WRONG: THIS FRACTION AT LEAST EQUALS THE AVERAGE UTILIZATION FACTOR AND IT MAY BE CONSIDERABLY LARGER.

EXAMPLE: SUPPOSE UNITS EACH WORK ON CALLS FOR SERVICE AN AVERAGE OF 45 PERCENT OF THE TIME. THEN 45 PERCENT OF DISPATCHES WILL BE INTERBEAT DISPATCHES.

FIGURE 11

STATEMENT	RIGHT OR WRONG?
WORKLOADS OF UNITS WILL BE BALANCED IF WORKLOADS OF THEIR RESPECTIVE BEATS ARE ALL BALANCED.	WRONG: INTERBEAT DISPATCHES AND THE "BURDEN OF CENTRAL LOCATION" REQUIRE THAT BEAT WORKLOADS BE UNBALANCED IN ORDER FOR UNIT WORKLOADS TO BE BALANCED.

FIGURE 12

how good the computer model is, no matter how complex it is, no matter how good the interface is, the user is just not going to understand it, be able to interpret it and he will not use it for making decisions. I have a couple of other examples of this, but time is running short so I won't present them.

QUESTION: Did they understand these back-of-envelope things when you got finished explaining it?

LARSON: Yes, but I had to take about ten minutes on each of them.

QUESTION: That's not so bad.

LARSON: I had two examples which I presented at an IACP seminar. The hypercube model which I have talked about has been used for police sector or beat design. There are a lot of different objectives here and I don't have time to explain all (Figure 13). The fact is that any particular user has different trade-off values for each of these kinds of things. Some are system wide objectives for system wide efficiencies and others are neighborhood objectives to minimize inequities among neighborhoods. So we have to build the model to allow each type of user to utilize his preferences that way. Again, a reader reasons that optimization can be applied to a situation like this. The kinds of rules of thumb which the user has to be familiar with are back of envelope reasoning associated with these kinds of things: compactness of sector, sector area, travel time square root law, the burden of central location i.e., if a response unit is centrally located in the precinct he is more likely to get a heavier work-load in the precinct than a unit that is not so, because he is going to get out-of-sector dispatches (Figure 14). So all these properties are global properties of a microscopic model, and if you don't understand them, you're not going to be an intelligent user of that computer base model. So the model that has been developed, which Jan will talk about in his presentation, basically is a model to be used as follows: client proposes the particular design; the computer calculates and develops performance measures; the client scratches his head and says do I want to accept this or do I want to propose a new design. And iteratively go through the process (Figures 15-16). My own personal feeling is that at least in the municipal area, where performance measures and constraints are so problematic, except for garbage collection and mail delivery, that you need this kind of iterative feedback to the decision maker as an integral part of the whole process. There are all kinds of outputs, that are measured, that are calculated (Figure 17). This is an analytic model having many equations. One has to solve it numerically. You don't get equations out, you get numbers for all these kinds of things for any particular proposed objective. I have been personally involved in implementation experiences in the modeling community.

In Quincy, Massachusetts, two years ago, we used a police model on the MIT computer system. For a unique set of reasons in each case, the results of the analyses were implemented, implying a change sector design.

In Wilmington, Delaware, this model was used to design an experiment which ran through the entire last year and which has just been evaluated. In St. Louis, Missouri, we're using the model to evaluate a new technology, the

SECTOR DESIGN OBJECTIVES

- WORKLOAD BALANCING
- AVERAGE COMMAND-WIDE TRAVEL TIME
- POLICE ACCESSIBILITY TO NEIGHBORHOODS
- CROSS-SECTOR DISPATCHES
- PATROL FREQUENCIES
- NEIGHBORHOOD INTEGRITY
- NEARNESS TO CURRENT SECTOR PLAN
- USE OF MAJOR STREETS AS BOUNDARIES
- OTHERS??

FIGURE 13

RULES OF THUMB IN SECTOR DESIGN

1. SECTOR AREA VS. SECTOR TRAVEL TIME.
2. COMPACTNESS OF SECTORS.
3. EFFECT OF DIFFERING TRAVEL SPEEDS.
4. CROSS-SECTOR DISPATCHES AND WORKLOADS.
5. A PATROL UNIT'S WORKLOAD IS NOT EQUAL TO ITS SECTOR'S WORKLOAD.
6. THE BURDEN OF CENTRAL LOCATION.

FIGURE 14

USE OF THE MODEL: ITERATIVE IN NATURE

- FIRST PLANNER PROPOSES A PARTICULAR DESIGN OF SECTORS
- THEN COMPUTER CALCULATES THE RESULTING VALUES OF THE PERFORMANCE MEASURES.
- THEN PLANNER WEIGHS THIS EVIDENCE IN WITH REMAINDER OF KNOWLEDGE OF THE AREA, AND DECIDES WHETHER TO ACCEPT THE PROPOSED SECTOR PLAN OR TO DEVISE AN ALTERNATIVE.

FIGURE 15

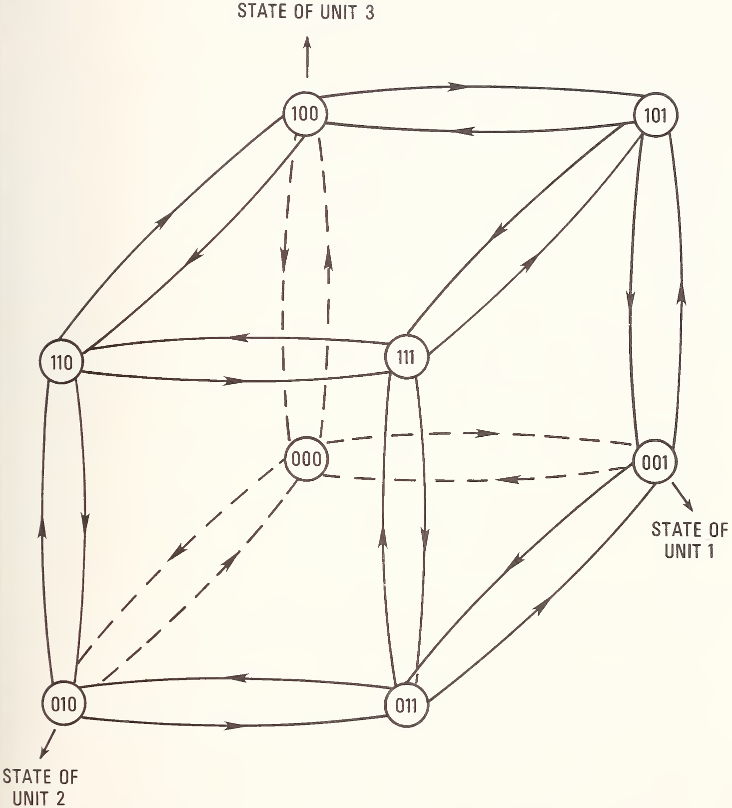


Figure 16 State Space for a 3-unit Problem

HYPERCUBE MODEL

OUTPUTS:

AVERAGE REGION-WIDE TRAVEL TIME;

WORKLOAD OF EACH UNIT (MEASURED IN FRACTION OF TIME BUSY SERVICING CALLS);

WORKLOAD IMBALANCE;

REGION-WIDE FRACTION OF DISPATCHES THAT ARE INTERSECTOR;

FRACTION OF DISPATCHES TO EACH UNIT THAT ARE OUT-OF-SECTOR;

FRACTION OF DISPATCHES IN EACH SECTOR THAT REQUIRE OUT-OF-SECTOR UNITS;

FRACTION OF CALLS DELAYED IN QUEUE;

AVERAGE TRAVEL TIME TO CALLS IN EACH SECTOR;

AVERAGE TRAVEL TIME FOR EACH UNIT;

AVERAGE TRAVEL TIME TO CALLS FROM EACH GEOGRAPHICAL CELL;

FRACTION OF CALLS FROM EACH CELL THAT ARE HANDLED BY EACH OF THE UNITS.

FIGURE 17

St. Louis police car automatic monitoring system. In this case, we had a model which started with an unsolicited grant, there was no real narrowly defined user community. What we're trying to do now is get a broad base user community which will be composed of consultants, people from universities, patrol projects and a number of different types of individuals (Figure 18).

Well, from this work and student experiences, we tried to categorize factors influencing success in the degree of implementation in three different areas (Figure 19). The three areas are technical, the model user interface, an the political concerns in working with municipal government (Figures 19-21). In the technical area, it's easy to deal with the accuracy of the model and is it better of than alternatives. One thing that I would like to disagree with, in spite of what has been mentioned before, is that a model should be evaluated on its predictive accuracy. I think a model should be evaluated on its ability to improve decision making. If you categorize something to the extreme, you could have a model whose outputs are uniformly a factor of two off. But if there is a factor of two off or every alternative you consider, then it would still rank order the alternatives appropriately and even give you a relative comparison betwen them, a correct one. It's really not the predictive power of the model. Rather it's the comparison of decision making alternatives which exist or can be constructed. We're talking about performance measures and all these kinds of things. Turnaround time is a key consideration. Cost to collect data to operate is usually always underestimated. It is my experience that about 90 percent of the time it has been a key bottleneck to implementation in the technical area.

QUESTION: Is turnaround time time for model implementation, development and use of the model?

LARSON: This turnaround time is time for model development and implementation, which sometimes is 18 months, sometimes is two years, in which case the guy who commissioned the model is probably no longer there.

QUESTION: To reapply this model is 18 months to two years?

LARSON: No, I'm talking about models in general. If the model is commissioned and no model like this exists, you have to start from scratch. That is the time until it's available for implementation.

QUESTION: I'd like to suggest that at some point in the conference we probably want to talk about Seth's remarks about the gradual evolution of the the military, Dennis's remarks about some of the cycles in what he's doing, your remarks and my own experience about the two-year period is all the state allows you. We ought to say something about some of these.

LARSON: Yes, of course. The second class of considerations I think, is the model user interface. Quite often a model will be produced, a computer package will be dumped on the user's doorstep, whatever documentation we have is there, and then the model creator goes away and there is no attention paid to the interface. The user though is the key part of the whole decision making process. If you evaluate models on their ability to aid decision making,

HYPERCUBE MODEL

QUINCY, MASS.

ARLINGTON, MASS.

WILMINGTON, DELAWARE

ST. LOUIS, MISSOURI

ROTTERDAM, THE NETHERLANDS

STOCKHOLM, SWEDEN

CALIFORNIA CIG CITIES

FIGURE 18

FACTORS INFLUENCING SUCCESS OR DEGREE OF IMPLEMENTATION:

TECHNICAL

MODEL ACCURATE?

BETTER THAN ALTERNATIVES?

MEANINGFUL PERFORMANCE MEASURES?

ADAPTABLE TO A PARTICULAR COMMUNITY'S NEEDS?

FLEXIBLE IN ON-GOING IMPLEMENTATION?

COST TO COLLECT DATA AND TO OPERATE?

TURN-AROUND TIME?

FIGURE 19

MODEL-USER INTERFACE

UNDERSTANDABLE OUTPUTS?

DATA EASILY CHANGED?

ALTERNATIVE POLICIES EASILY EXPLORED?

TURN-AROUND TIME?

TIME AND OTHER INVESTMENTS TO CONQUER LEARNING CURVE?

EASILY EXPLAINABLE TO VARIOUS TYPES OF AGENCY PERSONNEL?

EFFORT REQUIRED TO CHANGE MODEL?

FIGURE 20

POLITICAL

POSITION AND POWER OF INDIVIDUAL USING MODEL.

SHORT AND LONG-RANGE GOALS OF KEY DECISION-MAKERS IN USER'S AGENCY.

NECESSITY FOR TECHNICALLY "PROVING" AN ALREADY SELECTED POLICY.

EXTENT OF IMPROVED PUBLIC IMAGE DUE TO HAVING TECHNICAL SUPPORT.

NATURAL TIME CONSTANTS AND CONTRAINTS OF AGENCY.

FIGURE 21

then you have these kinds of things involved too. In the municipal sector, the effort to change a model, the effort to not be put in a straightjacket -- sometimes the models that I have seen, will put a straightjacket on the spatial configuration of resources throughout a city because whoever was commissioned to put the package in, programmed the computer in a way which is impossible to change. So sometimes computerization actually places more limits on decision making than appeared before. All these kinds of things have to be considered. Here the turnaround time refers to the user sitting down at a terminal. Is it on-line, is it overnight, is it batch processing?

QUESTION: For our information, what is the extent of documentation?

LARSON: The hypercube model? The hypercube model has an executive summary, it has a technical piece which talks about the equations; actually there are about three or four papers on that. It has a user's manual, with actual programs; it has a card file with a lot of cards in it, comment cards, and there are about four or five case studies which are written up.

QUESTION: They are expensive?

LARSON: Well, this has occurred over a period now of about a four or five years.

QUESTION: When a municipal client gets into this, do they do it on their own funds or do they get grants for this?

LARSON: When a municipal client gets involved, in my experience more than not, they usually get their funding from outside. Probably the LEAA funding or something like this.

QUESTION: Have you been involved in actually helping them get the funding? Say for St. Louis or --

LARSON: Sometimes. The last thing I'd like to talk about now that my time is over is, I think, the most difficult points, at least in the municipal sector, that I've seen, and that is the political attributes of the model and the position that the model builders and users are in. I know this is one of Garry's favorite topics. He can say a lot more than I can about the position or power of the person that's using the model, how it is perceived by others in the agency. You have to consider the goals of the decision makers in the agencies, both in short term and long-range goals, particularly with respect to promotion or salary increase, or the relative advantage over other people and how does this all fit in. Sometimes, and this was mentioned this morning and Garry refers to it in his book on urban problem solving, we have the necessity of providing something which has already chosen to be the policy that is going to be implemented. Sometimes the model builders are required for this purpose. I think we should all be very sensitive to be included in that type of situation. Sometimes it is a good public image associated with having bright outside technical support and so the model builder can be a good PR gimmick, if you like. And then there is the time constant -- I like to think of the time constant of an agency as that time required to get a 50% turnover of personnel. If we're talking, let's say, municipal service, if they have a 30-year retirement plan and things are in equilibrium, it takes about 15 years for you to get a 50% turnover of personnel and sometimes a minimum of 10 years. And 10 or 15 years is a natural time constant for systems. If you have GAO for NSF or any other organization going in six months after the 18-months modeling effort was launched, to check your chance of success or failure in that modeling effort, that evaluation totally ignores the time constant of the system you've been working with. The constraints of the system, speaking of constraints a little bit differently, constraints include the tenure of the top person. The police chief is rarely there for more than two years as head of a large municipal police department. You have the next election for Mayor, you have the power or non-power of the city manager when he is up for reappointment; so these are all constraints, too. I think the realities of attrition played a key role in some of the workings of the New York City Rand Institute. Certainly we know that the Mayoral elections in New York City played a key role in what happened to the Institute.

Usually, from my experience, modeling efforts are launched at times that are independent of, and not sensitive to, the natural time constants and constraints in the agency. Somehow the model's developers have to be aware of this and try to fit it in with their work.

QUESTION: I want to return to documentation. Could you say, since you do have an amount of documentation which is substantial, how the planning and financing of this documentation was accomplished?

LARSON: Well, a lot of the documentation was supported by NSF through grants to MIT and our final report is coming out this year, a four-volume book. One of the volumes focuses almost solely on the technical model and we encouraged the user agencies to be cooperative in the chapters that are case studies. So in a sense they became very excited about this because they could publish some work, which they ordinarily do not do, and so that gave rise to case studies. HUD funded grants to do the follow-up work on implementation of these models and that gave rise to a case study in New Haven on the hypercube model, in a number of other cities, on some of the deployment models. I think that HUD and NSF have been unusually interested in follow-up at least in the short term on some of this implementation which therefore gave rise to case studies. Both HUD and NSF were very interested in a tiered hierarchial level of documentation for the model, for different kinds of audiences.

QUESTION: Could you fill us in on the personalities involved. I don't know HUD all that well. My impression is that it is not ordinarily passionate in support of research and documentation.

LARSON: I'd rather talk about --

QUESTION: Dick, why not LEAA.

LARSON: Why isn't LEAA involved with this? They are involved with it to an extent in St. Louis and Wilmington. St. Louis is evaluating an automatic monitoring system for police cars. Wilmington is designing and evaluating a police patrol experiment. But LEAA usually views its role as underwriter and implementer of things. I don't see them necessarily as creators of new tools. They were not interested in this four or five years ago because it was a new unproven too, but NSF and HUD were.

IMPLEMENTATION OF EMERGENCY SERVICE DEPLOYMENT MODELS IN OPERATING AGENCIES

Jan M. Chaiken

BACKGROUND

A growing body of literature is suggesting the virtues and benefits of using modeling techniques of operations research to resolve governmental problems [2, 12, 16, 23, 30]. Yet all careful studies of the actual use of models by decision makers have drawn sobering conclusions about the chances that such models will actually be applied as intended. Even in the Department of Defense, which has been sponsoring modeling activities for many years, Shubik and Brewer [36] found, in a written survey conducted in 1970-71, that under half of the models, simulations and games had produced results worthy of presentation to policy makers in a briefing, and a smaller fraction had an impact on policies. Fromm, et al. [13] obtained similar findings in a 1973 survey of Federally funded non-defense models. They stated that "at least one-third and perhaps as many as two-thirds of the models failed to achieve their avowed purposes in the form of direct application to policy problems."

Many of the models examined in these two surveys were built for use by the Federal agency that funded the work. The implementation history of computer-based models intended for use by agencies of local government has been, in general, even less promising. In 1974, Michael Lawless conducted an interview survey of 39 recipients of models intended to be used by criminal justice agencies such as police departments, courts and correction agencies [27]. He found that only 18 percent of the recipients had used or were using the model. Attributes of the model itself (its programming language, data requirements, or conceptual complexity) were obstacles to implementation in only 28 percent of the instances of nonuse. The primary obstacles to implementation for these criminal justice models were unrelated to the model's characteristics. For example, it was very often that the case that a single advocate in the potential user agency saw the need for a model, conducted a search for the appropriate one, sponsored his choice before agency administrators, and pursued implementation. If the advocate became discouraged, or lacked political skills, or was promoted to a better position because of his skills, the model was not used. Chaiken, et al. [9] labelled this problem "the vanishing advocate." Other types of problems that led to nonuse were disputes between analysts and policy makers having no relationship to the virtues or lack of virtues of the model, and acquisition of models for a purpose that did not arise. In 12 percent of the cases studied, the reasons for nonuse could not be determined.

Lawless also made several observations about the characteristics of practitioner agencies and model builders that impede implementation of models. These appear to be sufficiently general that we may hypothesize their applicability to agencies of local government other than the criminal justice agencies that were surveyed. The first characteristic of practitioner agencies mentioned by Lawless is that models are often introduced to improve operations that are already generally considered satisfactory. In other words, the potential user agency does not perceive a need for change or a dissatisfaction with the status quo. Most observers of organizational behavior consider innovation to be unlikely to occur in such circumstances (see, for example, references 20 and 42).

Second, the search for a model tended to cease when the first possible suitable model was found. Thus, consideration was infrequently given to selecting the best model for the purposes at hand, and many model recipients found that the model did not actually meet their needs.

Third, lack of professionalization among local agency planners is an obstacle to implementation of models. In many instances, planners do not have advanced training, a tradition of using analysis to make decisions, or a world view that extends beyond their immediate organization. Using models would have been an activity very different from their usual style of work.

Finally, many local government agencies lack the technical resources to use computer-based models. They may not have access to any computer system, and, if they do, the system may not be able to compile the high-level languages ordinarily used for models. Collection of the data needed as input for a model may also be a major obstacle in such agencies. It is easy to envision that a governmental agency which has not yet introduced elementary data processing procedures would find a model to be too technologically advanced for its purposes.

In regard to the model builders themselves, Lawless pointed out that many of them have little interest in the implementation process and no incentive to become involved. Moreover, their special capabilities as researchers might be wasted if they spent time implementing models, and these capabilities certainly do not qualify them as able implementers. Thus, in parallel with the problem of the advocate noted above, one often observes a problem of the "vanishing model builder."

HUD-FUNDED DEPLOYMENT MODELS

This paper describes the implementation record of six models that were sponsored by the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development (HUD) with the specific intent of overcoming the commonly experienced obstacles to implementation of models. All six were designed for use by local policy, fire or emergency medical service agencies for analysis of deployment policies (how many emergency units to have on duty, where they should be located, what their response areas should be, and how they should be dispatched). The work was conducted at Rand during 1973-75 and has been summarized by Walker [39].

To assure that the models met actual needs of local decision makers (rathe than possibly imaginary needs invented by the model designers), HUD required that the models must be field tested in several cities. This experience also permitted validating the models, that is, checking that their output matched reality. Moreover, after the models had been tested, their characteristics were modified to meet the needs of users. For example, their capabilities, output formats, and mode of use (interactive or batch) were changed in some instances. Most of the field tests were described in written case studies, which have also been summarized by Walker [39].

To enhance the likelihood that the models could subsequently be transferred to other agencies of local government with little or no assistance from the model designers, HUD required that the models not be written in unnecessarily obscure programming languages and that they be completely documented. The documentation for all the models included the following [39].

° An EXECUTIVE SUMMARY, containing a nontechnical introduction to the model, information to assist an administrator in deciding whether to use the model, and details about how the computer program can be obtained.

° A TECHNICAL DESCRIPTION, designed to provide an analyst with an understanding of the theoretical underpinnings of the model.

° A USER'S MANUAL, describing step-by-step how the model is operated once it is installed on a computer system.

° A DESCRIPTION of the computer program. This document was written for data processing personnel and provides sufficient information to permit installation of the model, construction of the required data base, and modification of the model, if desired.

After the models were completed and documented, HUD awarded a small contract to Rand for maintenance of the models. This work included responding to user inquiries, fixing bugs in the programs or errors in the documentation as they were brought to our attention, and collecting information about the uses (if any) of the models. Direct, onsite assistance to users was not provided under this contract. Rather, the purpose was specifically to determine the extent to which the models would be used with only the most limited types of dissemination activities. A survey of the recipients of the models, conducted under this contract, provided information about the extent and nature of their use.

It should be noted that the surveys by Fromm, et al. and Lawless mentioned above, included emergency service deployment models but were not specifically focused on such models. None of the models described here had been completely documented at the time of the earlier surveys, and three of them had not yet been built.

The six models of the study were the following:

PARAMETRIC ALLOCATION MODEL (PAM): This model was designed by Rider [32, 33, 34]. It is intended to be used by fire departments or ambulance agencies for for rough analysis of the number of firehouses or garages needed in each of several large subregions of the jurisdiction served by the agency.

FIREHOUSE SITE EVALUATION MODEL (FHSEM): This model was designed by Dormont, Hausner and Walker [11, 38]. It is intended to be used by fire departments to evaluate specific proposed locations of firehouses.

SIMULATION OF MODEL OF FIRE DEPARTMENT OPERATIONS (FIRESIM): Written by Carter [4, 5], this model is a detailed simulation written in SIMSCRIPT I.5. It can be used to evaluate practically any deployment policy, including firehouse locations and dispatching practices.

PATROL CAR ALLOCATION MODEL (PCAM): Designed by Chaiken and Dormont [8], this model is intended for use by police departments and is similar in purpose to the PAM.

HYPERCUBE QUEUING MODEL: This model was designed by Larson [7, 24, 25, 26], partially under HUD funding and partially under NSF funding to MIT. It is intended for use by police and ambulance agencies for design and evaluation of fixed sites for their units and/or response areas for the units.

SIMULATION MODEL OF POLICE PATROL OPERATIONS (PATROLSIM): Written in SIMSCRIPT II.5 by Kolesar and Walker [21, 22], this model is similar to FIRESIM in its design, data requirements, and applications.

PATTERNS OF USE AND NONUSE

Despite the fact that all six model were carefully tested before they were released, errors in the programs and/or the user's manuals were found for each of the models, except the PAM. None of these problems was successfully resolved by the users. Rather, they were referred to Rand for appropriate action. The difficulties were usually easy for the model designer to repair, but would probably have been extremely difficult or impossible for another person, no matter how well trained. In one case, the diagnosis was not apparent even to the model designers.

This experience points to the necessity for agencies that support the design and documentation of models to support at least modest maintenance activities subsequently. No model is "perfect," and it seems likely if errors are found -- even if they are in infrequently used options, as was the case with the FHSEM, the PATROLISM, and the Hypercube Model -- the model will rapidly fall into disuse unless appropriate corrections are made. All known errors in the computer programs were reported within the first year after release of the models. In other words, no new bugs have been reported during the last six months. However, it remains to be seen whether a twelve-to-fifteen-month maintenance period would have been adequate.

Aside from difficulties with bugs and errors in the user's manuals, over half of the users had to change the program in some way before operating them. Some of the changes were very minor and were already anticipated in the user's manuals as possibly desirable. Others were more substantial but routine; they involved changes to make the program compatible with the user's compiler. In some cases, hundreds of lines of programs were changed for this purpose, but the user apparently did not consider this activity a major obstacle.

A totally unexpected development was the complete rewriting of two programs into other languages. There are now at least four versions of the Parametric Allocation Model, two of which were written in FORTRAN by users who had no contact with the model designer. Similarly, the COBOL version of the Hypercube Model was written by a user who was not associated in any way with the model designer. Since the new versions meet some user's needs better than the original programs, and the programmers are not in a position to disseminate the models, the task of documenting the new versions and making copies available of future users has been assumed by Rand. This task was not anticipated as part of the maintenace activity.

Most emergency service agencies operate the model on a computer that does not belong to them. Typically, the computer is owned by a university or a commercial service bureau. While the designers and their colleagues almost invariably provided an opportunity for interactive use of the models by the agencies they assist, other users (i.e., those without such assistance) more frequently use the models in batch mode. Although there are some instances of interactive use by agencies that have no outside technical assistance, it appears that the virtues of interactive use as perceived by the designer are irrelevant to most users. Indeed, Nelson Heller (private communication) reports that the additional cost of interactive operation is actually an obstacle to use of the Hypercube Model by some agencies. (In interactive mode, the user pays for the time he is connected to the computer system as well as for operation of the program.)

In analyzing the survey responses to determine the conditions that are conducive to use of models, it is difficult to sort out certain temporal effects. In particular, the earliest users of the models are the ones who are most likely to have had time to make operational changes based on the output from the models. These users also differ from later users in other respects as well -- for example, nearly all of them have had a personal contact with the model designer or one of his colleagues. Thus, the data show that personal contact with the designer or a colleague occurs in most instances where operational changes have occurred, but this observation may not be relevant for anticipating future events.

Some patterns in the survey responses appear to be independent of the passage of time. First, nearly all of the users who did not have direct Rand assistance obtained their data from computerized information systems. In several of the cities where field tests of fire deployment models were conducted by Rand staff, the researchers set up procedures for keypunching data from previously collected manual records. These procedures, which were continued by the fire departments, were considered to be a side benefit from having conducted the study. However, it appears that for an agency without outside technical assistance the absence of computer-readable data may frequently be an insuperable obstacle to using models. An alternative explanation of this observation is that agencies whose use of computers has not yet evolved to the stage of routine data processing will not have personnel who are sufficiently skilled to use deployment models.

Second, in every instance where an agency has made operational changes based on the output from one of the six models reviewed here, an attempt had been made to validate at least part of the model's output by comparing it with real data. Validation is a practice that model builders invariably recommend, but the forces compelling validation by users appear to be stronger than mere recommendations. Namely, the local analyst realizes that the fire chief, police chief, or city council will not be persuaded to take action based on estimated performance measures that have not been shown to be trustworthy. Paul Scheuer, a systems analyst in the Toledo Division of Police, reported "we compared PCAM's estimated travel time to actual travel time and found that it varied by approximately SIX SECONDS from actual results. Close enough!" (Emphasis in the original.) Similar encouraging experiences were reported by other uses who went on to implement changes in operations.

In instances where validation efforts revealed disparities between the output of the model and actual data, progress invariably stalled or terminated. The usual next step was either to abandon the model or to search for improvements that could be made in the data or the computer program. Concerning an application of the Hypercube Model in Anchorage, Thomas McEwen of PRC Public Management Services, Inc., reports "The program seems to work best in small, compact areas. Some of the dispatching rules in the program seem to be invalid in large areas when compared to actual practice." This application did not result in operational changes. The Edmonton Police Department reported that PCAM's assumptions do not match the department's operations. "Output for queue delays and probabilities of encountering a queue do not appear realistic. We found we were unable to use PCAM in its present form and plan future use of the model after we change the queuing equations to reflect our operations." The Yonkers Fire Department, which is not continuing to use the Parametric Allocation Model, reports that its travel-time estimates do not appear realistic because "topography and geography are not fully taken into account."

In summary, then, the fairly widespread use of PAM, FHSEM, PCAM and the Hypercube Model indicates that the models frequently survive validity checks, but it does not indicate that they are universally applicable. On the other hand, FIRESIM and PATROLSIM, which can be adjusted to be valid in nearly any city, have not been used at all after their initial tests because of their complexity. Therefore, there is a trade-off between validity and usefulness. A model that can be used by many local governmental agencies is likely to incorporate simplifying approximations that make it invalid for some applications, so each new user must repeat the activity validation.

OBSTACLES TO ACQUISITION OF MODELS

An additional survey was conducted to determine why individuals who have been instructed in the use of emergency service models might choose not to acquire copies of the models. The recipients of the survey instrument were students in a course presented in the summer session at the Massachusetts Institute of Technology by Richard Larson and Amadeo Odoni. The course, which was entitled "Analysis of Urban Service Systems," provided an opportunity for the students to hear lectures by Larson on the Hypercube Model and by me on PCAM and to operate the models in interactive mode using demonstration data base.

The predominant response from students who had not acquired either model was that they do not work for an emergency service agency and, therefore, have no use for the models. Among students who could potentially use the models but did not acquire them, the main explanations were as follows:

- They could not persuade their superiors to use the models or to budget funds for such purposes.

- They have requested funds in next year's budget.

- They lack appropriate computer support.

FACTORS INFLUENCING IMPLEMENTATION

During a period of approximatley two years that followed field tests in five cities, around ten individuals or agencies received copies of the Parametric Allocation Model and the Firehouse Site Evaluation Model. All but one recipient used the models, and most users have recommended operational changes based on the output. Over 35 agencies received copies of the Patrol Car Allocation Model, and a similar number received the Hypercube Queuing Model. Over 80 percent of recipients have used these models or have taken concrete steps to use them. In the case of PCAM, all users who obtained realistic output have made operational changes based on the results, but eight Hypercube Model users found they did not want to make changes. After initial field tests no one has used either FIRESIM or PATROLSIM. On balance, this experience is encouraging but not unique. For example, Chaiken, et al. [9] reported in 1975 that the JUSSIM model [3] had been acquired by 35 agencies or individuals, and Jack Barry [1] reported in 1977 that 52 cities and countries had used the Fire Station Location Package (FSLP) designed by Public Technology, Inc. [29].

Based on our experience with the six emergency service deployment models described in this paper, some observations can be made about important factors in the implementation process. These observations are presented as opinions or hypotheses, since the available information is not adequate to support firm conclusions. In particular, it is very difficult to determine the reasons for nonuse of a model. In most instances, we deduce nonuse from the fact that we have never heard from the recipients, and they did not respond to our surveys.

DOCUMENTATION

The documentation of a model plays many roles. Of course, the existence of a user's manual is an absolute prerequisite for dissemination of a model to recipients who do not have technical assistance from the model designer. Thus, none of the models described here could have spread beyond the Rand- or MIT-assisted test cities in the absence of a user's manual. On the other hand, the availability of a user's manual does not guarantee dissemination. The user's manuals for PATROLISM and FIRESIM were written according to the same format and with the clarity as the user's manuals for the other four models, and their availability was announced in the same media. Yet these models have not experienced dissemination.

Indeed, it is reasonable to believe that the documentation of the simulation models, especially their executive summaries, have actually discouraged dissemination. The executive summaries specifically warn their readers about the cost and difficulty of operating the model and about the necessity of having a SIMSCRIPT compiler. If the executive summaries had engaged in "salesmanship," glossing over the difficulties, perhaps these models would have been ordered by larger numbers of people, who only later would realize they could not use the model.

Documentation also serves the purpose of alerting potential users to the availability of a model. In this context, a clear, brief, inexpensive executive summary is probably more important than a user's manual. While the primary means by which recipient learned about a model was through knowing or meeting the model designer or one of his colleagues, substantial numbers of recipients first learned of the model through its documentation.

A curiously important aspect of documentation is an annotated program listing. While few users ever inspect this part of the documentation carefully, its presence is evidently reassuring in several years. It suggests to the reader that the program in "finished" and not subject to repeated modifications, even though this may not be true. Moreover, it clearly indicates that the program is not proprietary and is provided without any restrictions on changes to be made by the user. Most important, it demonstrates that the model designer has enough confidence in his or her product to expose it to the critical eye of other model builders.

PERCEPTION OF IMPACT

Not only must a model address a problem that the potential user considers worth analyzing carefully, but also the nature of the likely impact from using the model must be readily perceivable. There is a tradeoff here, because simulation models are powerful tools for addressing a variety of important issues, but precisely because of their flexibility, it is difficult for the potential user to imagine exactly what he or she will do with the model. Possibly this is a partial explanation for the lack of dissemination of FIRESIM and PATROLSIM. By contrast, the other four models have more limited uses, but their potential applications can be easily understood.

The Hypercube Model provides an informative example of the importance of an understandable impact. This model actually has a variety of possible uses, but it has come to be known as a tool for designing patrol beats in police departments. Where it has been used by people who are not associated with Larson, the purpose has been primarily beat design.

Do deployment models address important problems? In truth, considering the issues to which fire chiefs, police chiefs, city managers, and mayors devote their attention, questions related to the temporal and geographical allocation of response units must be judged to have relatively low priority. Questions related to the total resources that will be devoted to the police patrol function or the fire suppression function are more important to municipal administrators, but the extent to which the output from models can actually influence these decisions is still open to question. For example, in Yonkers and New York City, a budget reduction forced a decrease in the number of fire companies, and the models were then used to determine how the cuts should be made. In Denver, the study revealed that a smaller number of fire companies could provide about the same level of service [17], but interest in this possibility preceded the study.

As a general matter, deployment models appear to be used for decisions that are not very important but must be made. The users are people who are charged with at least partial responsibility for the decision, and would prefer to make a good decision rather than simply a satisfactory one.

DATA REQUIREMENTS

One might guess that if one model requires lesser amounts of data than another model, or more readily available data, it is more likely to be acquired and used. However, this is not necessarily so, since the model requiring less data will also ordinarily be less accurate and have lesser capabilities. Even when both models can be used to answer the same policy question, the simpler model may not be the one chosen. For example, in Fresno, the Hypercube Model was used to allocate patrol cars to geographical commands, a function which can be performed more easily with PCAM.

As mentioned earlier, an agency that is considering the possibility of using a model does not usually view itself as making a choice among alternative models. Therefore, we might expect the number of people who acquire a particular model to be affected more by the techniques used to disseminate it than by its comparative advantages in relation to other models. In Fresno's case, the police department was offered a opportunity to participate in a field test of the Hypercube Model; the possibility of choosing PCAM instead did not arise.

After a model is acquired, data requirements do appear to have an influence on whether or not the model is actually used until completion of a study. Evidently most users do not really come to grips with the problem of collecting data until the program is in hand. Several instances of aborted uses of the Hypercube Model occurred because of the difficulty of obtaining data, and probably some of the presumed nonusers of PCAM (those who did not respond to the survey) were unable to collect necessary data. The one known nonuser of PCAM (Kansas City) did not have a data problem. Kansas City actually wanted to redesign its patrol beats, a function that can be performed by the Hypercube Model, but not by PCAM.

PROGRAMMING LANGUAGE

Many emergency service agencies have difficulty finding a computer system that they can use and that will compile a program written in BASIC (the Parametric Allocation Model), PL/1 (the Hypercube Model), or especially SIMSCRIPT (the two simulation models). Occasionally, the agency cannot compile FORTRAN, but this problem occurs less frequently. Nearly all agencies can compile a COBOL program.

To use one of the simulation models, an agency must be able to modify the program as well as compile it. The absence of SIMSCRIPT programmers in municipal government, therefore, presents a severe restriction on the prospects for disseminating those models, probably the most important restriction.

ROLE OF THE ADVOCATE

Again in this study we usually found a single person in the agency who was an advocate for the model, pushing its implementation to a successful conclusion. Dedicated and politically skillful advocates have played an important role in all the examples of application that have been led to changed operations. The degree of dedication they possess is illustrated by the examples, mentioned earlier, where models are being used despite the absence of funding or authorization to do so. The advocates trust, in these cases, that they will be able to persuade their superiors to use the model, once it is running.

INTEREST OF THE MODEL BUILDER

The fact that PCAM and the Hypercube Model have been disseminated more widely than PAM and FHSEM, is partially explained by the continued interest of the designers of the first two models in patrol allocation research. The designers of PAM, FHSEM, and the two simulation models subsequently went on to other fields of research. They are examples of "vanishing model builders."

FEDERAL FUNDINGS

Field tests of PAM and FHSEM by Rand in Jersey City, Tacoma and Wilmington were funded by HUD, and the National Science Foundation funded field tests of the Hypercube Model conducted by MIT and the Institute for Public Program Analysis. Moreover, the earliest PCAM users all had funding from the Law Enforcement Assistance Administration (LEAA) for resource allocation. Thus, Federal funding has been involved in a sizeable fraction of the cases where models have been used.

More recent recipients of the models have not had Federal funding, and over half of PCAM users are expending local funds. Thus, it does not appear that the availability of LEAA funds to police departments, and the absence of a similar source of Federal funding to fire departments, accounts for the large number of PCAM users and compared to, say, FHSEM users. However, LEAA's continuing investment in improving the planning capabilties of criminal justice agencies may now be influencing the interest in PCAM and the Hypercube Model.

VERIFICATION AND VALIDATION

I have already mentioned the importance of verified models -- that is, computer programs that are debugged and work as the model designer intended -- and of the validation process. Continued dissemination of these models would not be possible, or even ethical, if a large number of users found they do not work as claimed.

PARTITIONER-TO-PRACTITIONER TRANSFER

Many model users reported in their survey responses that they have recommended the model to other agencies. I, therefore, expected to find that many of the most recent recipients of models first heard about them from other satisfied users. This, however, was not the case. Emergency service agencies still become aware of the models by having their personnel attend training courses or by discovering the documentation of the model.

Nonetheless, I believe that satisfied users are playing a major role in the dissemination process. Their influence is not felt at the stage where an agency first becomes aware of the model, but later, when the agency is deciding whether and how to use the model. At this point, the potential user often makes telephone calls or site visits to determine what has happened with the model in other cities. Only if the news is generally encouraging will the potential user turn into an advocate for the model in his or her own agency.

ACKNOWLEDGEMENTS

I am grateful to the Office of Policy Development and Research, U.S. Department of Housing and Urban Development, for supporting the activity of following up on the use of these models. Nelson Heller and Richard Larson were very helpful in providing me with lists of recipients of models, copies of completed questionnaires they had distributed, and related information. I especially wish to thank the many people who responded to the questionnaires we mailed to them.

This paper is based on a more detailed report of the same name: Rand Corporation Report P-5870, May 1977.

REFERENCES

[1] Barrett, Jack, "The PTI Experience," presentation at the Workshop on Utility and Use of Large-Scale Mathematical Models, National Bureau of Standards, Gaithersburg, Maryland, April 1977.

[2] Beltrami, Edward, MODELS FOR PUBLIC SYSTEMS ANALYSIS, Academic Press, 1977.

[3] Blumstein, Alfred, "A Model to Aid in Planning for the Total Criminal Justice System," in Leonard Oberlander (ed.), QUANTITIVE TOOLS FOR CRIMINAL JUSTICE PLANNING, U.S. Department of Justice, Government Printing Office, 1975. See also Chaiken et al. [9] for a complete list of earlier JUSSIM references.

[4] Carter, Grace, SIMULATION MODEL OF FIRE DEPARTMENT OPERATIONS: PROGRAM DESCRIPTION, The Rand Corporation, R-1188/2-HUD, December 1974.

[5] Carter, Grace, Jan Chaiken, and Edward Ignall, SIMULATION MODEL OF FIRE DEPARTMENT OPERATIONS: EXECUTIVE SUMMARY, The Rand Corporation R-1188/1-HUD, December 1974.

[6] Carter, Grace, Edward Ignall, and Warren Walker, A SIMULATION MODEL OF THE NEW YORK CITY FIRE DEPARTMENT: ITS USE IN DEPLOYMENT ANALYSIS, The Rand Corporation, P-5110-1, Jul7 1975.

[7] Chaiken, Jan, HYPERCUBE QUEUING MODEL: EXECUTIVE SUMMARY, The Rand Corporation, R-1688/1-HUD, July 1975.

[8] Chaiken, Jan and Peter Dormont, PATROL CAR ALLOCATION MODEL: EXECUTIVE SUMMARY (R-1786/1-HUD/DOJ); USER'S MANUAL (R-1786/2-HUD/DOJ); PROGRAM DESCRIPTION (R-1786/3-HUD/DOJ), The Rand Corporation, October 1975.

[9] Chaiken, Jan, Thomas Crabill, Leo Holliday, David Jaquette, Michael Lawless and Edward Quade, CRIMINAL JUSTICE MODELS: AN OVERVIEW, The Rand Corporation, R-1859-DOJ, October 1975. Also published by the Government Printing Office, April 1976.

[10] Dormont, Peter, "The Parametric Allocation Model: Batch Program Description and User's Handbook," unpublished paper (available from Chaiken), September 1975.

[11] Dormont, Peter, Jack Hausner and Warren Walker, FIREHOUSE SITE EVALUATION MODEL: DESCRIPTION AND USER'S MANUAL, The Rand Corporation R-1618/2-HUD, June 1975.

[12] Drake, Alvin, Ralph Keeney and Philip Morse (eds.), ANALYSIS OF PUBLIC SYSTEMS, MIT Press, Cambridge, Mass., 1972.

[13] Fromm, Gary, et al., FEDERALLY SUPPORTED MATHEMATICAL MODELS: SURVEY AND ANALYSIS, prepared for the National Science Foundation by Data Resources, Inc., and Abt Associates, Inc., June 1974.

[14] Hausner, Jack and Warren Walker, AN ANALYSIS OF THE DEPLOYMENT OF FIRE-FIGHTING RESOURCES IN TRENTON, NEW JERSEY, The Rand Corporation, R-1566/1-TRNTN, February 1975.

[15] Hausner, Jack, Warren Walker and Arthur Swersey, AN ANALYSIS OF THE DEPLOYMENT OF FIRE-FIGHTING RESOURCES IN YONKERS, NEW YORK, The Rand Corporation, R1566/2-HUD/CY, October 1974.

[16] Helly, Walter, URBAN SYSTEMS MODELS, Academic Press, 1975.

[17] Hendrick, Thomas, Donald Plane, et al., AN ANALYSIS OF THE DEPLOYMENT OF FIRE-FIGHTING RESOURCES IN DENVER, COLORADO, The Rand Corporation, R-1566/3-HUD, May 1975.

[18] Ignall, Edward, Peter Kolesar, Arthur Swersey, Warren Walker, Edward Blum, Grace Carter and Homer Bishop, "Improving the Deployment of New York City Fire Companies," INTERFACES, Vol. 5, pp. 48-61, 1975.

[19] Ignall, Edward, Peter Kolesar and Warren Walker, USING SIMULATION TO DEVELOP AND VALIDATE ANALYTICAL EMERGENCY SERVICE MODELS, The Rand Corporation, P-5463, August 1975.

[20] Knight, K., A DESCRIPTIVE MODEL OF THE INTRA-FIRM INNOVATIVE PROCESS, Journal of Business, Vol. 40, pp. 478-496, 1967.

[21] Kolesar, Peter and Warren Walker, A SIMULATION MODEL OF POLICE PATROL OPERATIONS: EXECUTIVE SUMMARY, The Rand Corporation, R-1625/1-HUD, March 1975.

[22] -------, A SIMULATION MODEL OF POLICE PATROL OPERATIONS: PROGRAM DESCRIPTION, The Rand Corporation, R-1625/2-HUD, February 1975.

[23] Larson, Richard, URBAN POLICE PATROL ANALYSIS, MIT Press, Cambridge, Mass., 1972.

[24] -------, A HYPERCUBE QUEUING MODEL FOR FACILITY LOCATION AND REDISTRICTING IN URBAN EMERGENCY SERVICES, The Rand Corporation, R-1238-HUD, March 1973. Also in COMPUTERS AND OPERATIONS RESEARCH, Vol. 1, pp. 67-95, 1974.

[25] -------, URBAN EMERGENCY SERVICE SYSTEMS: AN INTERACTIVE PROCEDURE FOR APPROXIMATING PERFORMANCE CHARACTERISTICS, The Rand Corporation, R-1493-HUD, January 1974. Revised version in OPERATIONS RESEARCH, Vol. 23, pp. 845-868, 1975.

[26] -------, HYPERCUBE QUEUING MODEL: USER'S MANUAL (R-1688/2-HUD): PROGRAM DESCRIPTION (R-1688/3-HUD), The Rand Corporation, July 1975.

[27] Lawless, Michael, "Implementation of Models," Chapter 7 in Chaiken, et al. [9].

[28] Merideth, Jack and Anthony Shershin, EMS AND FIRE ACTIVITIES IN THE SOUTH FLORIDA REGION, School of Business and Organizational Sciences, Flordia International Unversity, Miami, Working Paper 75-4, November 1975.

[29] Public Technology, Incorporated, FIRE STATION LOCATION PACKAGE: CHIEF EXECUTIVE'S REPORT, FIRE CHIEF'S REPORT, PROJECT LEADER'S GUIDE, PROJECT OPERATIONS GUIDE, Washington, D.C., 1974

[30] Quade, Edward, ANALYSIS FOR PUBLIC DECISIONS, American Elsevier, New York, 1975.

[31] Rand Fire Project, FIRE DEPARTMENT DEPLOYMENT ANALYSIS, Chapter 14, in preparation.

[32] Rider, Kenneth, A PARAMETRIC MODEL FOR THE ALLOCATION OF FIRE COMPANIES: EXECUTIVE SUMMARY, The Rand Corporation, R-1646/1-HUD, August 1975.

[33] -------, A PARAMETRIC MODEL FOR THE ALLOCATION OF FIRE COMPANIES: USER'S MANUAL, The Rand Corporation, R-1646/2-HUD, August 1975.

[34] -----, A PARAMETRIC MODEL FOR THE ALLOCATION OF FIRE COMPANIES, The Rand Corporation, R-1615-NYC/HUD, April 1975.

[35] Rider, Kenneth, Jack Hausner, et al. AN ANALYSIS OF THE DEPLOYMENT OF FIRE-FIGHTING RESOURCES IN JERSEY CITY, NEW JERSEY, The Rand Corporation, R-1566/4-HUD, August 1975.

[36] Shubik, Marting and Garry Brewer, MODELS SIMULATIONS, AND GAMES - A SURVEY, The Rand Corporation, R-1060-ARPA/RC, May 1972. Also to appear in THE WAR GAMES: A CRITIQUE OF MILITARY PROBLEM-SOLVING, in preparation.

[37] Singleton, David, Fire-Fighting Productivity in Wilmington, A Case History," PUBLIC PRODUCTIVITY REVIEW, Vol. 1, pp. 19-29, 1975.

[38] Walker, Warren, FIREHOUSE SITE EVALUATION MODEL: EXECUTIVE SUMMARY, The Rand Corporation, R-1618/1-HUD, June 1975.

[39] -------, THE DEPLOYMENT OF EMERGENCY SERVICES: A GUIDE TO SELECTED METHODS AND MODELS, The Rand Corporation, R-1867-HUD, September 1975.

[40] Walker, Warren, David Singleton and Bruce Smith, AN ANALYSIS OF THE DEPLOYMENT OF FIRE-FIGHTING RESOURCES IN WILMINGTON, DELAWARE, The Rand Corporation, R-1566/5-HUD, July 1975.

[41] Weissberg, Richard, USING THE INTERACTIVE HYPERCUBE MODEL, Massachusetts Institute of Technology, Cambridge, June 75.

[42] Zaltman, G., R. Duncan, and J. Holbek, INNOVATIONS AND ORGANIZATIONS, Wiley, New York, 1973.

THE PTI EXPERIENCE

Jack Barrett

Public Technology Inc., is a non-profit, tax exempt, public interest corporation established in 1971. Our primary mission is to facilitate the transfer of technology among State and local governments.

One of our main program areas is computer-based analytical models to support decision making. We have four operational systems. The first is a fire station location model; its objective is to help State and local governments locate fire stations. We have a park and recreational facility location system, to help locate parks and recreation facilities. We have an ambulance location system and finally we have a land use forecasting methodology. These systems have been used by 82 cities and counties around the country, ranging from Anchorage, Alaska to St. Petersburg, Florida, and in size from Dallas, Texas to Hope, Arkansas.

Our first, oldest and as a consequence most understood system is our fire station location system. I say "understood" because we learned very early that we couldn't understand how a system like a fire station model really worked until we saw how it fit within the decision making environment which it was designed to fill. It is necessary, perhaps even critical, to understand the political and decision making environments in which a model is placed before we can understand how the model can best be used to support decisions. Frankly, we didn't learn how the fire station location model as a technology worked, how it facilitated decisions, until it had been used in about twenty cities.

Our current version of the fire station model has a very simple concept. The basic assumption is that travel time to fire hazards is an important factor to consider in locating fire stations. The basic structure of the model includes three data bases. The demand data base, consisting of the disaggregation of the city into zones; each zone is assigned travel time requirements. A supply-data base which indicates potential sites for stations. Finally, a network data base, which links supply and demand together. Fire station plans are designed by local staffs, and analyzed to see how well each plan, a subset of potential sites, succeeds in meeting the response-time criteria.

A very simple model, very susceptible to technical comprehension and it's now been very widely used. Seventy three cities and counties have used this. Some cities have reduced the number of stations they had, and others have added stations and still others have stayed pat.

The fire station model used to be much different, however. When it was first developed it was an optimization system, designed to establish the minimum number of stations needed to satisfy local travel time requirements. Instead of a simple evaluation structure, it had a complex minimization/optimization structure. We quickly found out two things. One, the Fire Chief didn't understand what the model did. As a consequence he didn't understand how to interpret what the model said. The consequence of that was he didn't have confidence in the results. And two, he didn't want the computer to tell him where to build his station in the first place. He wanted to pick the station locations. All he wanted the computer to do was to show him what the advantages and disadvantages of his selections were, or, at least as frequently, to provide analytical justification to top management of his intuitive judgements.

Out of this we learned a very important lesson. The user should be involved in designing the model that is designed to help him make decisions. This should be obvious but it is something that we didn't truly comprehend until we had been through three cities. However, through installing our fire station model we began to understand the system. As we began to understand it, we modified it. Now we think we have a system that does what the user wants it to do.

Ever since our initial experience with the fire station model, we have made great efforts to get potential users to articulate how they want each new system to help them. Then we have tried very hard to give them what they want. This is most clear in the case of our newest, most ambitious, most experimental, and, as a consequence, least understood system, i.e, our land use forecasting model.

Before we started development work on this system, we convened what we called a User Requirements Committee. In this case, we used local staff from all around the country that we thought were potential users for this methodology. We got them to provide detailed product specifications. At that meeting we didn't just ask them, "What do you want?" but we provided them with a number of specific questions to which they gave us specific answers. What they wanted boiled down to this. They wanted an analytical set of steps, a proces that local staff could work through, to produce short-range, five or ten years, forecasts that were analytically based. They told us that they wanted a system that brought uncertainty about the future to the surface and did not hide it through assumptions or through mathematical averaging techniques. They wanted the uncertainty brought to the attention of the human decision maker to be dealt with by someone who understood the local environment. They did not want a long-range, complex, land use forecasting number cruncher. They wanted to figure out where the city would grow, and they wanted to have confidence in those decisions, through use of basic and analytical methodology.

Well, we produced a methodology along these lines that is now being tested experimentally, with high success thus far, in Eugene, Oregon. The basic concept again is very simple. The theory of the model is that there are some factors which are critical to the development of certain land uses. And there are other factors, like arterial access, which increase the likelihood of a zone being developed if it meets all the critical needs of a land use. The system therefore consists of three steps. Zones are screened to determine if they have the minimum requirements of a land use. Zones that pass the critical test are then ranked according to how high they score with respect to certain quantifiable factors, like arterial access. Then planners subjectively interpret the listing and either accept the ranks or overrule them through knowledge of other factors the system did not consider. To do this, the planners have to understand the system. They have to understand what the system is and what the system is not.

For the sake of clarity, we even introduced, on purpose, technical inefficiency in the system design. We created eleven computer programs instead of three, just so the computer process would be very clear to the user.

The land use forecasting model is again, a very simple system, and aimed at helping to support decisions, not at producing tne answer for the decision maker. Our land use forecasting methodology seeks to help the decision maker find the answer himself.

We think there are several characteristics that our land use forecasting and fire station location, and all of our other models have in common. These characteristics are not things that we thought of; these are characteristics that have resulted because of our dealings with cities and counties and through an active effort to find out what the decision maker wants. Out models do not try to tell the decision maker what the answer is. They support decision making and are not prescriptive. They are conceptually simple and thus, understandable and therefore, the results are believable. They try very hard to fit within the existing decision making environment. They seek to capitalize on the skills of the local user and to build upon his knowledge and then to incorporate the user's experience in the design of the models themselves.

Finally, our models are not frozen. We have constantly modified our systems as we have learned more and more about them, trying to get them more and more attuned to what the decision maker wants.

If I had to summarize PTI's experience with one single recommendation, it would be this: try to find out what the user wants, try to build what he wants, and then, as you begin to understand the system, modify what you have built so that it meets his needs even better. My view is that it is both the simplest and the surest path to developing models that the decision makers will use and find helpful.

QUESTION: For your oldest models, do you still have to provide personal or onsite assistance to users?

BARRETT: We think that training users is very important. Yes, we do. Every time we have had a new client who wants to use the system, we go and explain the system. When we transfer the fire station location model to a new jurisdiction, I go to the manager and tell him what the system does, spend time explaining all the concepts in the system, all the data bases, and how they are linked to each other. Then I go down to the lower level, go down to the management department heads and spend an hour with them. It's a simple model, but it is very important for them to understand. If they understand it, they will have confidence in it and use it.

QUESTION: And if they change management, do you have to go in and do it all over again?

BARRETT: Well, that occasionally has happened, yes. But usually management doesn't change, and when the management changes the new manger may have a different set of priorities and may not want to do the project at all.

QUESTION: What documentation standards does PTI have on these models and how do they relate to the fact that you continually revise or update the program?

BARRETT: Well, we figure that we don't understand the models as we try them in the cities. Our initial documentation is Xerox copies. That is, we have documentation typed and Xeroxed. After we begin to understand the system, then we do the best job we can on documentation. We get professional artists to illustrate the documents. Then we go to very pretty type-set documents so that someone will enjoy looking through them and get the general idea through the illustrations. We didn't do that for the fire station model until we had two years of experience with it.

QUESTION: But this documentation is basically user manuals. What about the technical documentation of the program?

BARRETT: We either place technical documentation in appendices of user manuals or deal strictly with Xerox reproductions.

QUESTION: Do you collect and make or analyze the data bases from the different cities in order to abstract general rules and do some research?

BARRETT: Well, we are doing some of that now. But we have not reached any final conclusions.

THE FEA PROJECT INDEPENDENCE EXPERIENCE

Harvey Greenberg[1/]

I want to review some items that strike me as being the key things that come out so far: the issue of documentation; reviewing of models, usually by people other than modelers; the impact of a model on the decision process; a notion of standards, the absence of which seems to be a bad thing; the concept of separation of modeling, data acquisition and analysis, which I guess means the same thing as the structure; and finally, the idea of training, getting talented people into the field of modeling.

The underlying factor that has been omitted, in my opinion, is the dependence on the environment under which these things take place, and some of the tradeoffs that are sometimes elusive. With regard to documentation, and particularly the environmental factor under which modeling activities happen, I have a few comments to make.

The environment I'm in tends to be in crash mode always, either virtual or real, depending on whether you can meet deadlines, and so on. But in any case there is certainly a perceived crash mode in everything we do. My experience at FEA with long-term modeling is that it's something that is going to take about two months, and more typically the turnaround in doing something is more like a couple of days, if you have the luxury of Saturdays and Sundays in between. You might have as little as a few hours to get certain things to happen.

The most recent project has been like that and has really highlighted a lot of these points. That project was providing the analytic support evaluation of the President's program, which, in its initial form, was considerably different than the form that was presented before a joint session of Congress and the American people. The impact we had through analysis and through honest, objective modeling was perceived by the people that were involved in working with the White House staff, with modeling on the fly as what had to be done, it would have been really impossible to adopt a puristic view of documentation. Now this isn't minimizing the importance of documentation and trying to understand some guidelines. It does, however, highlight that if we

---

1/ This discussion occurred when Dr. Greenberg was at the Federal Energy Administration and used PIES to provide analytic support to the analysts formulating the National Energy Plan. Since the formation of the Department of Energy, Dr. Greenberg's work environment has changed. Since the time of this discussion, several factors have contributed to major environmental changes, namely: DOE has been formed, putting PIES, and Dr. Greenberg, into the Energy Information Administration (EIA); much of the automated documentation and data tracking to which Dr. Greenberg alluded is now a reality; more of the key PIES people have left, and some new people have been hired; comprehensive reviews of PIES are scheduled for 1978.

had some sort of a planned operating environment, an environment where we can plan what we are going to be doing, with a reasonable horizon and with a reasonable collection of resources by which we can make realistic estimates of each thing, then we can work out how to minimize the total time to make everything happen. This includes modeling and documentation, and this is documentation of not only of the model itself, but also of the data that the model was using -- our data being a very controversial thing itself. If we tried to minimize that, it would put us beyond the deadline, say something like April 20th. And so a tradeoff might be either to have a better and well-documented model which doesn't get used or to have a poorly documented model which is used, and where the modeler is a part of the analytical team, so that we can make the appropriate offline adjustments and provide the appropriate inputs to the advice that has to take place in the decision process.

MEADOWS: You give the impression that you are creating a new model every time a new assignment is given to you. But the PIES system is now four years old. The PIES model structure is certainly not revamped every time you get an assignment. Why is it not possible, given your enormous staff, to remove several staff members from responsibility for the crises? Assign them instead of documentation of your model.

GREENBERG: There are several parts to that question. I want to answer each part. In the first place, the group of people who are really close enough to the model to really know what's going on and be able really to to anything with it is not as large as you might imagine. In the second place it is not true that PIES now is anything like the PIES of four years ago. There's been major changes. It has always been necessary to make modeling adjustments in the evolution of PIES. Sometimes this can be anticpated, so the model can be made more flexible. However, some necessary changes for analysis hit us by surprise. For example, we never thought that the Federal Government would regulate the price of natural gas to intrastate pipeline companies, which is part of the plan.

Now there's a number of things that come up that require policy analyses that weren't thought of before and require some adjustments to what PIES represents. So that doing some of the modeling on the fly means that there some adjustments that have to be made in PIES, and there's a very small number of people who really know enough about it to make these adjustments. That's what I mean by modeling on the fly. I don't mean to imply that PIES should be rebuilt every time something comes up. I do mean to imply that PIES is constantly expanding and changing to deal with the kind of refinements which are necessary, and that during the analysis of the plan, the crash mode intensifies that effort.

MEADOWS: Eighty percent of the current PIES structure must have existed a year ago.

GREENBERG: I couldn't think of it on a percentage basis, but I would say 90 percent of the data has changed, and much of that has come from changes in offline analysis with just trivial changes in the raw data.

MEADOWS: In the environment just described, and in the absence of substantial documentation, how do you establish for the President or for critical outsiders that the work that you have done this rapidly is in fact correct? How will you ever be able to establish validity? In fact on what grounds is even the staff itself sufficiently satisified with the model?

GREENBERG: All right. This brings us to the next thing I was going to say in the discussion of documentation. What we do now as far as data documentation is a sort of archeology. We try to fish through the thing and remember where the numbers came from. In the case of the short time frame for complete analysis, such as we've been involved in the President's program, there is much difficulty because there is more attention paid to writing memos for the record and the like, to try and keep track of things. There is more of that done now than there was, but it's still not perfect and is certainly subject to the forgetfulness and so on. The main documentation, consistent documentation, is done in the post-modeling effort by contracts.

MEADOWS: How many members of this six person staff were present during, for example, the first two years of the model construction?

GREENBERG: One.

MEADOWS: So one person from the original group has been there two years.

GREENBERG: No, I guess two. What we worry about somestimes is this; given demands with one hour turnaround, two hour turnaroud, a day turnaround, a week turnaround; one dare not sit and wait for the question and only then run to the model and do an analysis. One had better be using the model continually in a somewhat anticipatory fashion. One builds a storehouse of information about the problem, the areas, so that one can respond without actually running the model sometimes.

MEADOWS: Do you anticipate anyone on the Hill will challenge your numbers?

GREENBERG: I anticipate everybody on the Hill challenging our numbers.

MEADOWS: Based upon your description right now it would seem you will find it very difficult to defend your studies before Congress.

GREENBERG: I don't think so, because we've already done some things about documenting the numbers and not all the numbers necessarily come from the Federal Energy Administration. We're working with the White House Staff, and we've got some other sources, that will then have responsibility for documenting some of the data that is used. Our responsibility for data documentation is now being taken very seriously. It is being worked on by some key people, but it still requires a very intensive effort. We're anticipating a need to present documentation, full documentation, of what it is that we've done. We are in the process of doing that.

But the problem is that the very first time around, I guess when they were still looking around for models they might use, PIES was one of several that might help with the analysis. We delivered some of the preliminary results based on what their idea of what the program was at that time, and articulated the right kinds of question; this is perhaps the most important use of modelers/analysts--surfacing the ambiguous. Our rule of thumb is that if there is an ambiguity stated in the program that raises a question that we need answered in order to know how to model it, then the answer is necessary to write the law. So we asked the right kinds of questions, got off the ground and were able to produce preliminary results. I think that being able to do that built up a certain relationship that got us into doing the bulk of the analysis throughout that period. Suppose we hadn't done that, we'd said, "Well, wait a minute now, let's go back to our drawing boards and figure out a work scope, plan the documentation, plan where we're going to get our data, etc." Within a few days we will have been one of the many people that were busy with that, and they would not have come to us for the analysis, such as it is.

MEADOWS: If the documentation just describes what you have done and not the structure of the model, how does Congress determine the validity of what you provided?

GREENBERG: Well, I'm going to get to that. I think the subject of validity is separate from the subject of documentation.

MEADOWS: Harvey, can I ask you what you do when some of your people quit?

GREENBERG: That's a very serious problem. When Bill Hogan left, many thought PIES might fall apart, but somehow it didn't. Somehow we picked up the slack and carried on. I just got here at the tail end of other things that were happening, and in retrospect and piecing together my personal experiences with what I heard took place before -- it seems there were a lot of people that could have done a lot but weren't really given a chance. Somehow I think that if the environment is such (and I want to get to that in the issue of impacts) that people are turned on by the whole thing, are electrified by being involved in this, then it brings them out.

MEADOWS: Of course every analyst in this room has documentation and validation problems. I would like to register, however, a note of strong disagreement, with the image presented here of how a group comes to understand a model. I have had the frustration of working intensively over a period of months to understand the full range of behavior implicit in a 300 equation model that I personally constructed. I found even at the end of the period that I would not always predict nor interpret the model's behavior accurately on the first attempt. I know everyone here has been surprised by the behavior of a model he has built, even the simplest one. PIES involves thousands of equations and numerous different models devloped by different people at different times, with different interfaces and inputs. I simply will not accept the suggestion that a group of people can ignore formal documentation procedures, merely become "electrified" and fully comprehend that model.

GREENBERG: So you are disagreeing with something I just said.

MEADOWS: Those of you who have used models will appreciate that you can rationalize any output as being correct. You look at a model run, and say, "Of course, I see why that occurred." When I was a graduate student I remember doing that. I rationalized one astounding result. Then my advisor said, "Hey, dopey, look at those fractions, they're upside down!" And I was then able to turn around and rationalize the opposite. Just like that.

COMMENTS: We did exactly the same thing, with a model which must be at least three orders of magnitude easier to understand than a PIES model, so I know that you are doing that because everybody does it. And the danger is even greater when the thing isn't documented.

MEADOWS: Do you feel that your user really felt that he was getting completely honest and valid results?

GREENBERG: I have no idea. I don't know what he did. I don't know what he thinks he did.

I think a partial resolve of some of these problems, or at least one avenue we're exploring, is developing some software concepts that may go pretty far towards alleviation at least as far as data documentation is concerned. A simpler example that comes to mind is the Brookhaven data base which is in, and for itself, documented. Now this data base is smaller and less diverse than we use at DOE, so that our problems are a lot more complicated in trying to use something like that. My own background in software gives me the feeling that with an appropriate amount of attention, data base concepts can really go a long way in providing self-documenting data bases.

QUESTION: Why do you need a data base that is self-documented?

GREENBERG: Well, they have recorded along with each number some basic source information about that number. I think that some development along those lines may be useful to pursue in solving some of these problems, particularly if you perceive a crash mode environment as being rather perennial. I think that there are some distinctions between that and a 25-year development of a model.

Let's talk about reviews. Now I don't know too much about what sort of reviews other models have been subjected to. Bill Hogan mentioned about six reviews that the PIES has been subjected to. Three of these have been coordinated by NSF, and one was conducted by GAO shortly after the PIB. Then there was a second round of reviews that took place. The interesting thing, I think, is the fact that in most cases FEA initiated or instigated the review process. That is, FEA didn't review themselves but they, for example, gave NSF the money to set up and conduct, or have others develop, a review process. So this has been done several times, and I suspect it will be done again, sometime over the next few months, that an outside agency or a university or some organization of people outside of DOE will once again review PIES.

I think that it reflects our basic attitude; that those of us who are involved with PIES really prefer to have the exposure; that is, we're really quite aware of many of the deficiencies in the PIES model, and we solicit people to provide us with alternatives that are better. We do not have a marital relationship with PIES. If there are better ways to provide decision making advice on energy, we would adopt them, and so we even tend to go out of our way to try and get people to suggest better ways of modeling PIES or what we use PIES for.

QUESTION: How big an effort were these reviews?

GREENBERG: I wasn't there. Bill, how big of an effort was in the reviews that took place?

HOGAN: I don't remember the exact funds but I would say that in the MIT review there must have been a half a dozen people who participated in this over a period of a month or so for part-time, and one or two people who spent a longer period of time doing it themselves. They were able in that time frame to identify a sufficient set of problems to make for a very interesting discussion. I don't know how much Battelle spent on it. The last one done by RFF with several different groups, each one taking a part, and I would say that each component had like a man-month or so devoted to it in trying to analyze it, and there were half a dozen or so of those --

GREENBERG: In terms of impact, I think the factors that are often noticed in talking about a model's impact, or trying to predict what impact it would have, have to do with issues of technical quality, ease of future understanding and the salesmanship of promoters. I think another factor, at least in the case of PIES, is the fact that it resided at the Federal Energy Administration, a regulatory agency (and now at DOE), because if the same exact model had been built by the same people at a university, I don't think the impact would have been the same. I think being a DOE model certainly is a factor in what impact it would have. In particular, you can't ignore the fact that DOE may be asked to give testimony in Congress, even though other agencies and the executive branch that need to evaluate things may ignore it, and that's probably going to happen. So almost by mandate DOE would attempt to be involved in any of these kinds of things. So that I think that another factor on impact is the way the model is built.

QUESTION: How do you define impact?

GREENBERG: I wouldn't attempt to give general definitions. The reason that I am saying that PIES had impact is because of my recent experience analyzing the program as it's being written down by White House staff, and then over time and on a daily basis having people meeting with the White House staff, and then coming back and meeting daily, during the day or the night, to decide what the next model changes should be. Then we make our PIES runs and do the analysis, and so on, and have that whole processing problem to work out in two months. We can see the kind of program that goes before the nation, and we can see this whole thing evolve and that the prior analysis we did was useful.

QUESTION: Can you say how much prior analysis is effective?

GREENBERG: I can, yes, certainly for myself. I wouldn't attempt to try to quantify it or defend it to you, but I certainly have perceived --

QUESTION: In my line of work there are two supreme tests of any analytical model. One is for the policy decision level at the White House. In the White House, if you present all the assumptions and the model you are using, you can see that impact of the model, step by step. Those guys at the White House are not dummies, but very smart; they're smarter than you are. They'll ask you a lot of questions. You've got to be prepared to provide answers. You can see a policy being formulated based on some models.

As far as lobbying is concerned, Federal lobbying, these guys don't give a damn what model they use. You can explain how fast my model is, of the million dollars I spent for the model; he doesn't give a damn about it. All that he cares about is that you did, beyond a reasonable doubt, the best you can do to prove your case. Now if you can do that through your models, you're home free. If you cannot do that, you are sunk.

This is the final use of our models. It doesn't mean that any work -- one equation or ten thousand equations -- that's the way they do it.

GREENBERG: I think the notion of a perfect model is irrelevant. I think there's always going to be an imperfection. The issue is that, under limited time, you have to make decisions about various kinds of programs that say in effect, what we should do with our resources; you can either choose to say, "Since there's no model I'll procrastinate," or you can choose to say, "Since there's no model that's perfect, I shall simply look at my data and seek out what I think the effects will be and use that." Or you can go to the depths of available models and, imperfect as they are, use them to try to draw some inferences about what's likely to happen if you do this, that or the other thing. I think the last is what is essential as a course of action.

Moving right along to standards, my comment is there is a danger of being monolithic. The idea of generating standards without paying attention to environmental effects, such as the kind of environment I have described myself to be in, I think would be a mistake. And I think for example that you -- assuming that the outcome of this meeting were to be sufficiently influential that all of a sudden Congress passed a law adopting whatever standards we came up with, and that all government work had to obey those standards -- would tie some hands, and that would not be very useful. And so I think one has to be careful of creating standards that don't allow for environmental factors that are important.

QUESTION: There are other possibilities. One, professional standards and standards in practice to serve as environmental protection -- I've got here for example, "A professional shall not attempt model adaptations which will affect the future of the nation in a significant way with less than a 24-hour turnaround." That could be a protection of you people --

GREENBERG: Would that include military decision assistance during wartime?

QUESTION: There are levels of review on this thing. I know Congress is going to review it. No, it's better than that. Or maybe I'm naive, but I can understand the pressures that the profesionals are going to be under or have been, were, or will be again, and maybe the answer is that the turnaround is too fast, and they ought to be given a couple of weeks after the whole thing is over, not only to document what happened during that period but also to reflect on it. To put body and soul together, but afterwards to reflect on what has happened and decide whether they believe what they said.

GREENBERG: There's a related issue here, of substitutability, to try to make up standards, and try to apply some of them to all people. One of the things that is implicit in some of the discussions surrounding standards is a common mistake in government actions, particularly in trying to lay out personnel requirements. It's the idea that one analyst is completely substitutable for another, or one economist is completely substitutable for another. This is totally fallacious. It's not any crew of people that could have put certain things together.

If standards were to be developed on a basis of "it takes this much time to do this" or "it takes this mixture of people to do this," where the people are strictly defined in terms of functional characteristics, like he's an or analyst or he's an economist, I think they would be bad standards. So another problem that I see with standards is the issue of substitutability.

QUESTION: So you're happy with the first situation.

GREENBERG: I didn't say that. What I said was there's some other factors which haven't been mentioned yet that need attention. I don't pretend to have all the answers, and I don't pretend that the method I know about solves all of them.

MEADOWS: Once we know about those other factors, will we be able to develop better standards? Or do you think we would decide that the standards which are currently implicit in the field are about as good as we can do?

GREENBERG: I don't know. I haven't given it as much thought as you apparently have, so I really wouldn't venture a guess.

The concept of separation is one I disagree with. In the environment I'm in at least, there is a tremendous importance and premium to have a modeler and analyst being the same person, rather than having modelers and analysts separated. I guess I view modelers and analysts this way, and I don't view things as being totally sequential since there's a lot of feedback and interaction that takes place in the environment now. It certainly seems that there might be other environments, particularly when modeling is set up for a more generaly purpose, where once you can outline that it is going to be sequential, then the separation may not be a bad idea. But in the environment I'm in, that separation would not be a good idea.

MEADOWS: You misinterpreted what I was saying. I said it was not necessary to wait until the decision makers pose a problem before building a model. We can start to model the structures of various processes involved in a set of potential future problems. Let us build some good model structures and take the time to do that well. You may say put together a model in a few months, and I would suggest the value was correlated with the time it took to put it together.

GREENBERG: Well, we changed a model that already existed during the period of those few months.

MEADOWS: You can make better interpretation of studies with models, the better you know the model.

GREENBERG: Exactly, I agree.

MEADOWS: I would think the modeling process in the areas of importance to this country, like energy, transportation, and so on, should go on continuously and be considerably enriched with data, learning from studies where there is support for the models.

GREENBERG: I want to comment on the other part, having to do with keeping them overlapping. After the model is run, there's several kinds of the outcomes. After you've seen a counterintuitive answer, one of the things you can discover is that you've made a mistake and turned a fraction upside down. You can count that as part of the debugging and shakedown process and reduce error or aid diagnostic analysis.

Let's talk about the time after the model is somewhat stabilized, and you are conducting various kinds of applications. I can go back to some of the things we did, say, before the crash mode. We were doing more leisurely kinds of studies such as some of the things we did CONAES.

What happens is that there is a certain percentage of the time the model is in some sense "right," and what you get out of delving into it is some new insights. For example, there was a scenario we ran, which we called the "dirty screnario," which allowed old coal to be burned without scrubbing. Intuition is that if you remove the scrubbing requirements, then more coal would be burned, because coal is substantially cheaper. We obtained the counterintuitive answer that the model preferred consuming a little less coal. But after the model delving into that, and you discover that the heat rates are different, making more efficient use of the coal. Thus, you can go through what I would not consider a rationalization, but a perfectly legitimate explanation of what happens. So we've gained some new insights in the process of using the model, in this case as a learning system.

Another possibility is that the model is in some sense wrong, but the reason for it being wrong it somewhat subtle. And in the process of discovering why it went wrong, you gained some new insights and of course the model gets corrected. For example, you might have to deal with a situation we ran dealing with coal conversion in utilities. The fact that part of the model is linear causes peculiar phenomena. The coal conversion that we modeled produced an answer that was subtly wrong. The analysis as to why led to an an offline disaggregate analysis.

Then a third possibility is the model is wrong but not for the suspected reason. For example, we ran a trial run to gain some insights as to what would happen if we supposed under this new gas policy that the total curtailment of the nation is going to be something like one quadrillion Btu across the nation. But we didn't anticipate in advance exactly where this curtailment would occur. Sort of like the model, we equilibrate and decide where this curtailment should occur. Prior intuition suggested at least two incorrect guesses. One of the guesses was that Chicago would be curtailed. A second guess, but also wrong, is that those furthest away, like New England, from where the gas supply is (Texas and Louisiana), are going to be the ones who are going to suffer curtailment --- because of transportation costs. In some cases, the anomalous results were resolved by finding where the model was wrong. In other cases, where we understood why the model did what it did, then we believe the results and changed our intuition. Stability may be measured by the frequency of model error (case 1) relative to model precision (case 2).

The training issues are becoming increasingly important, at least as far as PIES is concerned. Right now it's not easy for a senior analyst to learn PIES in less than several months to the point where he (or she) can contribute to modeling or analysis. I'm talking about smart Ph.D.'s new to our staff; it's taking them on the order of five or six months to really understand what's going on. So it's gotten quite out of hand. There are some problems, both on our end in trying to do our housekeeping and taking the time that is necessary to do some revamping to make learning easier; also, the issue of trying to get people with some background in modeling and whatever universities and other such places could do.

Let me just summarize what I think are some of the things that we might want to do now. First, I think these kinds of dicussions provide a forum which is tremendously useful. So the idea of continuing such discussions is good. I think that we can evaluate approaches to education. Maybe look at the Harvard Business School case studies approach and what all, and to take a serious look at what it would take to attract and maintain high quality professionals in the modeling field. And I think this maybe needs a deeper scrutiny than we've conducted, a more scientific approach to the evaluation of approaches to education. I think we need to go deep into analyzing the use of models. I think that the kind of thing like the survey that Dr. Fromm pointed to earlier today is a step in a direction that I would agree with, but I think that a lot more is needed -- a lot deeper kind of survey with a more subsequent analysis taking up questions such as if the failure rate depended on the model size, does HEW have a higher failure rate for this, that and some other type of thing. I think a deep study conducted by leading professionals who are very savvy in modeling and measuring the model impact in some form, going deep into the question, actually going through all the gory details it might take in analyzing the use of models -- I think this would be very useful and would probably be my favorite priority in terms of what should be done.

QUESTION: You might mention the legislative requirements for documentation and access.

GREENBERG: Right. I've mentioned the requirements, but not the legislative --

QUESTION: When Congress passed the FEA renewal legislation, it required FEA to submit to Congress all the PIES documentation, programs and parameters, and make them available to any one who wanted to use it.

HONIG: If the program were documented, do you think the outside reviewers should run the program? If so, how should it be run?

GREENBERG: I don't know. That's a subject for study, John. We haven't come to a conclusion on that, and I don't know an off-the-cuff answer to that.

MEADOWS: What would it take to validate a model, especially after you have six different people to evaluate it?

GREENBERG: I really don't know how to answer that either. I think that some of the techniques that have been classified as standard for trying to validate any sort of forecasting model, probably in the largest sense are inappropriate for PIES for a variety of reasons, not the least of which is discontinuities in our energy outlook such as embargo and the like. So I think there are serious problems here, I think, some things we're sensitive to, but not things that I personally have been involved about.

MEADOWS: If standard techniques are not appropriate for validating PIES, then others must be developed. The PIES model has had more influence on national energy policy than any other model; it is important and it is visible. The problems you have in validating it arise in large measure because it is what we call a goulash model, one that combines many different types of models: a linear program, input-output matrix, econometric model, and others that differ in their underlying paradigm.

QUESTION: Could you call it eclectic rather than goulash?

MEADOWS: Eclectic, fine. We surveyed agricultural models that have been developed, several at a cost exceeding $1 million. One observation borne out by our analysis is that eclectic models performed less well on a number of important dimensions than those models which were elaborated within one paradigm. The reason is clear: no model exists in isolation, there always has to be a professional at the interface between the model and the real system. He must constantly monitor it to make sure the clients do not ask questions that lie outside the legitimate scope of the model. He can also supply that intuitive judgment necessary to take the model results and interpret their relevance for policy. Where you have a pure method, econometrics or whatever, the professional's relationship is relatively easy to establish and maintain. But when you link different kinds of models together, as in PIES, there is no longer any single professional who really can perform the overall monitoring function. The model becomes a black box. And many results of analysis with the model are really more speculative than scientific.

Questions concerning standards of documentation and validity are more difficult to answer when you leave the confines of one well worked out discipline to create a conglomerate model which is composed of several different kinds of submodels pasted together. We have the impression from our work on eclectic agricultural models that there never was one person in any of the modeling teams who fully understood the whole model system. At best each person in the team had some confidence about his own submodel, but he was not able to monitor its relation to the real world because his model mainly interfaced with the rest of the submodels. Professional control was lost.

GREENBERG: Are you suggesting that modeling should be limited in such a way that if one person can't fully understand it, then it shouldn't be done?

MEADOWS: I didn't suggest anything of the sort. I was merely summarizing empirical results not, recommendations. I said that the severity of many problems concerning us here today is influenced by the extent to which a professional can stand at the boundary between a mathematical formalism and real life. The ability of the professional to monitor his analysis with wisdom and insight declines precipitously, as soon as you start putting together a bunch of methods in one operating system.

GREENBERG: Are you suggesting that there is some new mechanism we need?

MEADOWS: I have seen some good professional standards evolving to guide design and ue of econometric models and system dynamics simulation models. Linear programming and dynamic programming models are also generally worked out within the context of rather thoroughly discussed and widely known standards. I have never seen a single methodological treatment of guidelines relevant to the use of eclectic models. There are no texts on the subject, and the analysts engaged in analysis of the PIES system seem not to have any generally accepted rules to guide their own work.

HOGAN: As you can see I have an emotional reaction. I disagree with everything you (Meadows) said. I think you're wrong at all points and I think that although there are problems, and you can do it wrong, you can also do it right and there are ways to get around all your objections. I'd be happy to have an evening session to discuss this, but I just want to go on record.

THE EPRI/NBER ENERGY MODEL ASSESSMENT PROJECT

David Kresge

This project I am reporting on is on the other side of the fence from most participants in this workshop. It happens that I'm on the side of virtue since I'm doing model assessment, rather than model building. Needless to say all of us involved in the model assessment project are also modelers. We spent much of the first part of the project, in fact even as we were proposing the project, trying to deal with the question that Jan brought up earlier. Namely, all our friends came up and said, "What are you doing with your lives? Why are you toiling over other people's ashes rather than building you own models for greater glory?" And I must say that until today I hadn't been able to come up with a very good answer. We would say to other people, "Well, it seemed like a good idea at the time. It seemed like someone ought to decide whether these models are valid. After all, there is always the danger, it doesn't very often come to pass, but there is a danger that someone might actually pay attention to one of these models. And in that case it would be nice to know just how bad or good it is. It's rarely a question of right or wrong, but just how adequate or inadequate it is in the uses for which it is being considered."

So we made a proposal to EPRI, the Electric Power Research Institute which obtains its funding from the privately owned electric utilities. EPRI sponsors hardware research primarily, but it does do a little methodological and socio-economic research, and of course, that's where we're involved.

Now the other nice thing in all this is this delightful quote on the board which came from the Greenberger et al. book. It's not entirely coincidential, I think, that the main recommendation in that book is that there should be third-party assessment. The first author of the book, Martin Greenberger, is also the head of the EPRI program which is funding our study, so it seems that he was in the position of putting his own recommendations into practice.

So what we are involved in is third-party energy model assessment. The idea, though, is not so much to assess a particular model or a couple of models, but rather to develop some sort of methodology by which you can carry out model assessment, put that methodology into practice, and set up a laboratory facility where you can assess models on an on-going, continuous basis and can continue to develop the methodology. That is the task we're involved in, though we're only about six weeks into the project. It now has a one-year time horizon, though we expect, unless we fall flat on our faces, that it will indeed turn into a laboratory facility that will have a longer life than that.

But right now we're in the midst of trying to deal with an operational thing, with the kinds of questions that have come up in a much more general way in the discussions this morning and earlier this afternoon. Namely, what is it we want to do in order to assess models, what kinds of criteria can we apply, because we certainly feel very strongly that the criteria are not at all obvious. They become even less obvious when we get into the eclectic models. I think it was correctly pointed out that these are not impossible to assess, it's just that it is extremely difficult to figure out how to do it. And that is especially so because the models which are of primary interest are models which look well into the future, certainly this is true in the energy models areas. You cannot wait until the year 1985 or the year 2000 to find out whether the model predicted right or not. Yet it is very important to know how much confidence you can have in a model or how much confidence you can have in the model relative to other analytical approaches. We are treading in a virgin territory but we are putting on our combat boots and tromping in there nonetheless, because it seems that someone has to do it and it's kind of exciting to try to do it, especially since we don't have to try to assess our own models, but we can pick on somebody else's. Quite frankly, I would not want to have someone look at any model that I would have built in the same detail we're looking at other people's models. As another example, I would dearly love to get at the guts of PIES and give Bill a hard time on what's going on in there. Of course, that would be a massive undertaking.

Let me now try to use this diagram, Figure 1, to give you a feel for the kinds of things we're looking at and the kinds of general approaches we are taking. What I have on the top of this diagram is a very, very terse, grossly oversimplified version of the modeling process.

QUESTION: Are you confining yourself to econometric models or is there any constraint on the type of models you are looking at?

KRESGE: No, though, in practice, we are going to look at two models in the prototype phase. But in principle we are not restricting overselves. The models that we're looking at involve econometric, engineering, input-output, and even some process analysis. So that even with the two models we have selected we're not particularly restricted.

One way we have chosen to deal with this issue of how to analyze eclectic models operationally, is by undertaking the model assessment project with a team. It's the same sort of team we would put together to build the models. We have systems programmers; we have electrical engineers since we're dealing with a model of the electric utility industry; we have financial regulatory people; we have general macroeconomic, IO systems modelers; and we have very high-powered computer support because we are at the National Bureau's Computer Research Center. We are tearing these models apart; perhaps much more than one would want to on a general basis. Because we are trying to develop methodology, we are ripping into the things to the extent of actually reprogramming the entire model. We are starting at the gross methodological level and working all the way down the line by line coding.

We're working on this fairly intensively because we don't want it to drag on too long, even though it is a prototype. We figure that we should be able to do an assessment in a matter of, say, three to six months even for a very complex model, but we're trying to do even our prototype assessment on roughly

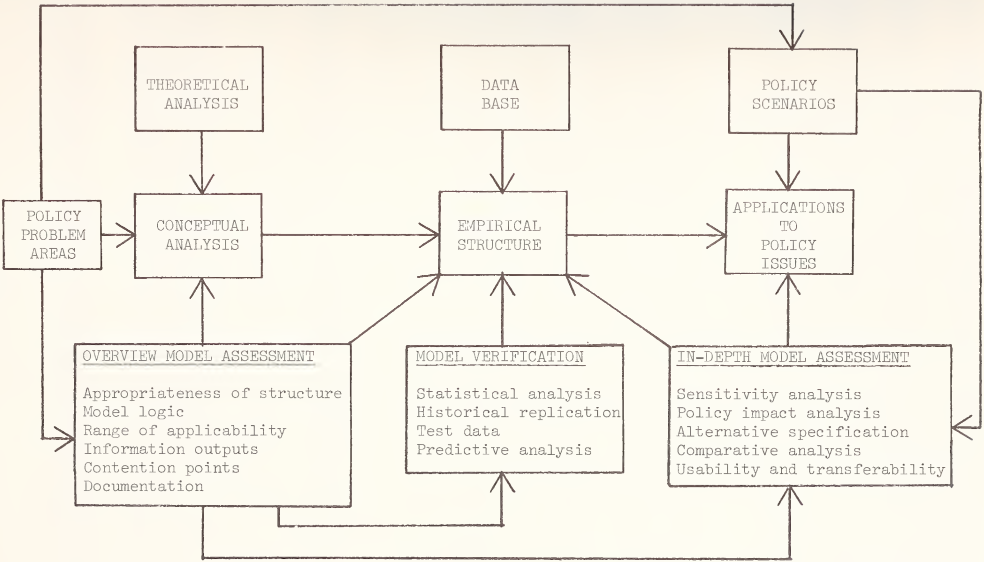


Figure 1 Energy Model Verification and Assessment

that same time schedule. We expect to have our first go-round done by the end of the summer.

Now let me try to explain the methodology we are using, and I again want to stress this is methodological. We are using an operational test but we are interested in developing general model assessment techniques. What is shown in the diagram is, as I said, grossly oversimplified version of the modeling process, showing the stages in model development. Something which has come up time and again in our discussions today is that the models we are talking about are designed to deal with decision problems. They are not academic exercises only. It is very important to begin with some sort of recognition of the types of policy problems that you want the model to deal with. It's both important in the development of the model itself and in the assessment process.

Now it is also important to recognize that the original client may not be the only client. Furthermore, the reason the model was originally developed is not the only area you might want to look at. You want to have some feel for the kinds of areas that a reasonable man might use this model to look at, so you want to define the policy applications as broadly as reasonable. Given the policy problem you want to deal with, you next begin the theoretical analysis to put together a general conceptual framework for use in the model. You then develop a data base, use that data base to implement this conceptual framework, and you have a quantitative or empirically implemented modeling structure. Typically there is a lot of back and forth in this process as you find out that you need to revise what you thought was the appropriate conceptual framework, you re-implement it, and then move back and forth until you finally get a set which is both consistent and in line with what it is you want to accomplish.

QUESTION: Does that imply, Dave, that all models are perfectly structured --

KRESGE: No. By empirically implemented I mean that you're attaching numerical values to the parameters of the model. They may come from going out and talking to the engineer who knows how this process works.

QUESTION: There are sets of equations devised somewhere else and developed either by process analysis or other means?

KRESGE: Yes. You see, the distinction I made between these two stages is that the first stage is purely conceptual, there are no numbers involved there, while the second stage has been quantified, by whatever means. I tried a couple of different words to summarize the second stage and I finally decided "empirical" was as general as I could come up with -- "quantitative" might have been a better word.

QUESTION: In the box marked Conceptual Analysis, is the sense that it has functional use represented merely by f( ), this is a functional box, or has it non-functional --

KRESGE: Probably, I don't regard it as terribly essential for what I am doing -- but I probably would just have f( ). You might also be making decisions about whether to use input-output analysis, if you are dealing with an economic model, or whether to use an income determination model. Or, whether to use process analysis in the technical analysis.

QUESTION: Somehow I get the feeling that you're doing the same thing that the guy did who orginally built the model.

KRESGE: As a matter of fact, what I'm talking about now is how I view the model building process. I haven't started talking about the assessment process yet.

Now let me finish outlining the final step in building and applying the models, so we can get to the assessment part which is, after all, the point of this project. At the final stage we again bring in the policy problems we want to deal with. I have again tried to be as general as I could by saying that somehow you have to convert your problem into specific sets of policy actions or decisions or input parameters or whatever. I just describe those as policy scenarios.

QUESTION: I would like to suggest that for some purposes one could find it useful to put another box between Empirical Structure and Applications - the actual choice of alternative numerical methods --

KRESGE: Yes, that might be a useful way to emphasize another thing we're looking at in the assessment. To give an example of how numerical methods can be important we found that in one of the models the solution algorithm has very, very poor convergence criteria. If you just change your policy a little bit, the fact that on one pass you may have converged at a high point, above the true solution, on the next pass you may converge on a low point, the difference between those two passes may be larger than the policy impact you're analyzing. In this case, a pure numerical computational problem can totally destroy the value of any policy impact analysis that you're doing. That's a nitty gritty problem but there's no way that a reasonable policy maker or model user could be expected to identify that kind of a problem in a model. It illustrates one of the reasons why, unhappily, it taks a very detailed analysis to know what you've got.

Now, we can get to the lower row of boxes in the diagram, which are the ones that deal with the model assessment process. We started out by saying that there were two distinct approaches to model assessment, and we were to do an example of each in the current model assessment project. The first is an "overview" assessment, and the other is an "in-depth" model assessment. Happily, Saul Gass sent his papers in early so that I was able to use his terminology, to put in a middle box called "model verification."

That process is involved in both the in-depth and the overview model assessment, though it's involved at different levels and uses slightly different procedures.

Nonetheless, we still have two basic levels, the overview and in-depth, at which to approach the assessment problem. The key operational distinction is that in the in-depth model assessment, we feel it is essential that, as third-party evaluators, we operate the model in a "hands-on" fashion. We want to have no intermediaries between us and the model. That means that it is not even sufficient to go to the model builder and say "make the following eight runs of the model and give us the output." That is not what we call hands-on operation. In an overview, we did not intend to make any runs at all with the model. These two approaches tend to shade into each other when we do go to the model developer and ask him to make some runs of the model. We now have the feeling that this approach could be one of the most difficult in which to tell exactly what we've got. Even an honest modeler, though of course all modelers are honest, tends to make adjustment when you ask him to run the model. If the model produces garbage, he's not going to send you that garbage. He will instead twiddle the dial here and change a parameter there because he knows that the model blew up just because of a quirk. He will make adjustments before you get the output. Often it's such an obvious quirk to him that he may even forget to tell you that he corrected it. It is essential to a true in-depth model assessment that the assessors run the model themselves.

In an overview, we focus on the conceptual framework first, concentrating on the functional forms rather than the quantitative parameters. An evaluation of the appropriateness of a specific functional form relative to the policy problem is a very important step. We also look at the model logic, and this may involve a fairly detailed look at the program. Even for the overview process, we feel very strongly we have to have access to the full computer program used in this model.

We next try to come up with some notion of the range of applicability of the model. This is most easily done in a negative way, we find. As a matter of fact, one of the things that is occurring to us very quickly and unfortunately very powerfully, is it is very hard to make positive statements about a model. You can say, "It can't do this, it's got an error here," or "It's got a weakness there." But, it's very hard to say, "The model is clearly appropriate for this problem." Because you know as soon as you say that, someone is going to turn around and show a reason why it isn't.

This is unfortunate because we think we are dealing with models that are quite good. We deliberately tried to pick models that we were confident that we were not going to completely discredit. We think the main value of the model assessment will come out of analyzing a fairly good model and saying, it's weak in these areas, it's strong in these areas and it can be improved in the following ways. We see the model assessment process as a positive and constructive type of activity, not as an activity that is trying to prove that a model is worthless.

QUESTION: You too, are going to have some sort of criterion for measuring this. Let me ask about price. I'm wondering if you would be willing to estimate what this procedure would cost if we put it on the DRI model or something of that sort.

KRESGE: The quarterly DRI model?

QUESTION: Yes.

KRESGE: We were looking at something similar.

QUESTION: The Wharton model?

KRESGE: The Wharton is one of the ones we're doing.

QUESTION: How do you estimate the cost of --

KRESGE: Overview or in-depth?

QUESTION: In-depth. A typical ballpark figure.

KRESGE: Between $100,000 and $200,000, though I'm pulling the number right out of the air.

QUESTION: Do you have staff experts in energy processing, in addition to experts in modeling?

KRESGE: Yes, that's right. We have people from the M.I.T. electrical engineering department, since this is being done as a joint project with the M.I.T. Energy Lab. Also, we have someone who is on loan from a power company, and we have three people from the electrical engineering department.

QUESTION: How can you assess the reasonableness of a particular assumption? I think that the rates will go up quadratically unless something else happens -- and I think of some point which changes some numbers -- how do you assess the reasonableness of the various assumptions?

KRESGE: I wonder if I could defer that question until I've gone through the rest of the steps. Because the statistical analysis and the historical replication of test data are procedures designed to answer that question.

QUESTION: On the one hand, you second guess the models. On the other hand, as I look down the list, you're looking at the model structure with its logic -- you are sucked into its conceptual frame of reference. As you look back on first principles, you say was this a sensible way to go about a modeling effort. There's a boundary line between second guessing a guy from an ab initio basis, which means practically doing everything up to model layout yourself --

KRESGE: That's what you don't want to do. We started out by looking at the policy problems the model is trying to deal with, and then asked what kind of a structure would be appropriate to deal with those problems. We did that before we looked at the model.

QUESTION: Okay, distinguish that from the appropriateness of the structure of the particular model.

KRESGE: As knowledgeable people in the field, we try to outline what we would like to have there if we were building a model, look to see what the model has, and if those who things don't match then try to spell out our reason why we think --

QUESTION: If you decide that another structure is better, you have to start building a model from scratch.

KRESGE: You would have to in order to improve the deficiencies in this one, but our task is to assess the range of applicability of this model. If we can identify structures that on theoretical grounds seem essential to deal with a particular problem and if those structures are not there in the existing model, we can say that's a limitation on the range of applicability. That's one of the key elements of an overview assessment. We try to tell the user where this model cannot be appropriately used. Again, we regard that as a positive thing. We're not just saying it's an error in the model. We're trying to flag users by saying, "Don't use this model for that problem. It doesn't have the mechanisms there that will allow you to analyze that."

QUESTION: Can you suggest a methodology for evaluating procedure models? And if so, how do you do that without ranking them, and if you rank them you're not sure they're going to be all positive?

KRESGE: If you mean evaluating and ranking, I would guess that's almost like doing anything else that involves a complex objective. I would be very reluctant to rank models in the sense of saying this one is a better model than that one. On the other hand, if you define a very, very specific decision problem for me, I might be able to do it in that context, but I doubt that that's an appropriate use of an assessment laboratory, except on some sort of contract basis. I think it's more useful to say that the model has this set of strengths and this set of weaknesses. It cannot be applied to these sets of problems because it does not have the appropriate structure; on the other hand, it does seem to us to be adequate or quite strong in the following areas. But then the most important thing is to give the reasons for why you are saying that.

Again, I wonder if we could try and muddle through some of these steps, because I keep getting ahead of my story on this. I have been trying to answer your questions without telling you what we're really in the process of doing. It's the implementation, I think, that really counts here. What are we doing with these model assessments.

In the first part of this overview assessment, the information output, of course, is very closely related to the range of applicability. A particularly damning point would be if there are certain information outputs that we feel give the impression that the model is applicable to a problem it is not applicable to. That would be regarded as a very, very poor characteristic of a model. And of course, it's not all that impossible.

A key output of overview assessment is to identify points of the model which seem to us to be particularly critical. The points of the model that are essential to the analysis of the problem we're dealing with, and that have to be looked at very carefully in order to tell whether the model is okay or not. They are not points where we say that the model is completely inappropriate because it just doesn't have the right structure. Rather they are elements that are there, but we don't know how well they are done.

We feel that one of the points of the overview assessment is to tell you whether or not we need to do an in-depth assessment. If on the basis of the the overview we can tell you that the model is not appropriate to your problem, there's no need to do an in-depth. It is also possible, in principle, that we could tell you on the basis of the overview that the model is perfectly adequate; then too we wouldn't need to do an in-depth. But that outcome is quite unlikely. In the more general case, the overview assessment would end up with a bunch of contention points where we don't know whether the model is adequate or not until we look at the empirical implications, and chances are that we would have to look at that in detail. In other words, the overview assessment might conclude the model is adequate structurally but whether it is adequate in practice or not would depend on the precise parameters and on the precise dynamic properties of the model. So one of the points of the overview is to identify the issues that have to be looked at in more detail.

Another key criterion in the assessment is "documentation." We tend to get very emotional if there is not adequate documentation, since it means our life is that much more difficult. To give you a horror story on that, one of the models we're looking at and the one that we are going to conduct an in-depth assessment on, looked to us like it had really excellent documentation, and in fact by current modeling standards it does. It gave good documentation on every single subroutine in the model with the exception of one which had the ominous name of "MAIN." Our first assumption was that MAIN, was just a call-up routine, all it did was call up subroutines. Wrong! It had lots of substantive elements, it had lots of integrated structure, and we had zero documentation on it. Our programmers were able to look at the code and unscramble it, but that's a very painful route to go, particularly when it's a key program. So the documentation turned out to be above average, but still far from adequate.

QUESTION: I notice that there is no feedback in the diagram, that flow diagram on the board. It all feeds forward. Isn't it possible that in that model evaluation you will get to a step where you want to go back to the top again just to relook at your assumptions as you go down --

KRESGE: Change the model?

QUESTION: No, not to change the model, you use that for the structure of your evaluation, I think?

KRESGE: Yes, right. I am not sure what the feedback would do.

QUESTION: Usually in problem solving, at some point or other, whatever problem it is you're solving, there's a point at which you go back, you have a way of going back to review your own assumptions and your initial ways of looking at things. It seems to me as you get down to a model you might find something that you might want to go back to look at.

KRESGE: In the overview?

QUESTION: Sure.

KRESGE: We all know that that's the way it works in practice. I think there is an even more important feedback which is also not in the diagram. If the assessment is done on a continuing basis, there will be feedbacks to the evolution of the model and that again stresses the constructive aspect of model assessment.

Saul has told me that I have something less than five minutes left and I still haven't gotten to the good stuff, which is the in-depth model assessment or model verification.

By verification we mean testing the operation of the model against existing data. Verification is a necessary, though not a sufficient condition, for model adequacy. What we're doing with the in-depth assessment in the area of verification may be a matter of overkill. We are planning to replicate all of the statistical analysis. We are then going to run the model against the historical data to see if we can replicate the historical time pattern.

QUESTION: Re-estimate by the same statistical criteria --

KRESGE: We're just trying to replicate the results.

QUESTION: The same outlook?

KRESGE: Yes. We really have equations that have developed by some kind of regression analysis; in most you have F's in conceptual analysis; you have normative --

QUESTION: In the model we have precise functions because we have someone's model.

KRESGE: I understand. Re-estimate their functions, as they specify, using their data. The best way for coefficients. To see if we know exactly how they got their coefficients.

QUESTION: That's not testing the output.

KRESGE: I'm not talking about testing the output; I'm talking about verifying the model to first determine where it came from, and then go on to see if it can replicate history (assuming that the model is supposed to do so). In this instance, the authors have done some verification, but we also know they fudged. For instance, they show that their model tracks history very well, the simulated and actual data points lie practically on top of each other. One of the key variables in the model is the price of gas; they use what they call a shadow price. The shadow price is determined by jiggling the gas price around until they get gas consumption, as estimated by the model, equal to the level it actually was. That kind of process is not really historical replication but is a twisting of the dials to force the model to track what you want to see. This process is more accurately described as calibrating or fudging - depending on whether it's your model or someone else's model. If it's your model it's calibrating but if someone else's it's fudging. It's not necessarily illegitimate but we want to know precisely what they did. And what would the model have done if they hadn't made the calibration adjustments.

On some of the engineering components of the model we can put in actual test data. We can put in a test signal of one form or another. We can also do predictive analysis where we have data beyond what they use for explanation. In the in-depth assessment, we are completely reprogramming the model and we are going to run it on our computer in a strictly hands-on fashion.

QUESTION: What do you mean by reprogramming?

KRESGE: Reprogramming the logic of the model in a different computer language.

QUESTION: What was their language, and what was your reason for that?

KRESGE: There are two reasons for doing it. One is that if you just accept their code, you keep the glitches and bugs in it. In reprogramming, our first requirement is that we have to be able to replicate their results exactly. We've already found that when there are programming errors in the model, we have to bring them along. Because the only way we can tell that we have the same model as they do is if we replicate their results exactly. If we find an error in their program we've got to put it in our program. Of course, we know how to take it back out and correct it later. The point is that if you just look at their code and leave it alone you won't pick up all their errors. No matter how good a programmer you are working with, you will find errors by reprogramming that you won't find by reading.

The other reason for reprogramming is that we're putting the model in a language which is much more suitable for sensitivity analysis and we can get a much better output. We're putting it in a highly interactive, high level language that the Bureau has developed called TROLL. In this language it's easy to change parameters, re-estimate using alternative specifications, or experiment with new functional forms. Those are all tests that we have to use in large numbers and if we are going to do that we've got to do it efficiently. The reprogramming of an existing model to make alternative specifications is very difficult, because you make mistakes and then you're not doing a fair assessment. You forget that if you change this equation you have to change it six subroutines later as well. If it's your program, even though it's not your model, you're less likely to make that error.

QUESTION: But you will introduce your own mistakes, your own glitches.

KRESGE: You have to be able to replicate the modelers' results.

QUESTION: After you reprogram, you're putting in your own errors.

KRESGE: If you have their model and the results from that model, then after you reprogram you have to be able to reproduce their results exactly.

QUESTION: The point is, what if you don't reproduce their results exactly?

KRESGE: Then you cannot do the assessment.

QUESTION: Then you do not know if you have errors in their program which you have not detected, or whether you have reached a set of errors in processing your own program.

KRESGE: Precisely.

QUESTION: By reprogramming it in TROLL, your computer language, you probably have effectively rendered your reprogramming immune to rebuttal analysis by the people who did the original modeling.

KRESGE: There is a way of translating back and forth, but your point is well taken. If we cannot replicate their results then we cannot do the assessment by this method. Because we don't know if we have simply made a programming error or whether we are in fact evaluating the model.

QUESTION: Are you going to replicate their results including their errors?

KRESGE: Yes.

QUESTION: In other words, they have a program, they have made an error, you translate that into TROLL, including that error, and replicate?

KRESGE: Correct

QUESTION: I guess the point I've missed is why you are going into TROLL.

KRESGE: In order to do the sensitivity analysis --

QUESTION: It's easier to do the sensitivity in TROLL than if you had done it in the original language?

KRESGE: Correct, and using the original language it would have been difficult to discover the same level of coding errors. We would have had to come up with an alternative procedure for doing that. For instance, in PIES it's out of the question to reprogram, it's too large. I also said that we were not conducting this procedure as a general methodology, this may be overkill. But, because we don't know how we can say on a priori grounds where we're most likely to find errors, it may turn out not to be overkill.

QUESTION: It sounds to me like two kinds of arguments. It seems to me if in fact you are so blessed as being the knowledgeable group in the area, why don't you just build the model? Why don't you just build a model for the community to use? Let somebody check your model.

GASS: But the basic idea is that of trying to develop an assessment methodology, not really trying to develop new models.

QUESTION: But it sounds like the assessment is twice as difficult as building a model in the first place.

KRESGE: But let me point this out. Even though we are setting this up as a prototype, starting from scratch and trying to develop a methodology and using what may be regarded as overkill methods, we are still expending a small fraction of what it costs to develop the model.

QUESTION: If they're the group that -- somebody asked this question before -- they have all the process modeling and all the model building power, and people who are the experts, why don't they just build the models?

QUESTION: They are not asked to pursue final analysis. They may turn out some kind of theory, it may be incorrect. They have not discovered it, you know, after spending half their budget. And they may have to do a major rejob. And this thing only asks whether they continue to decide whether they want bad theory. If the answer is bad theory then they finish the report on the first segment --

KRESGE: This study gives in-depth opportunity to learn about errors and defects in large-scale models, and that is a very valuable addition to knowledge quite independent of the goal of producing a better model.

issu
mode
part
like
last
noth
acce

abou
sign
of S
of s
deta
form

talk
doing
code,
do, I
of mo
the i
a ref
which
numbe
is to
alway
havin
which
answe
proce
find
times
decis
think
models
after
lems,
can ex

A
Modeli
throug
by sel

THE ENERGY MODELING FORUM

Wiliam Hogan

My plan today is to describe my work, but first I want to focus on two issues. First, for the record, everyone here would agree that we build large models for a purpose, not just to make a modeler's life interesting. Modeling, particularly large modeling, should not have its own imperatives. It's more like President Carter's view of nuclear power: something to go to only as a last resort, but it happens that this last resort is often needed. But there's nothing per se that is valuable about large models; small models would be quite acceptable if they answered the questions.

The second issue, which keeps returning in our discussions, is the debate about standards vs. modeling as an art form. I noticed "Standards" is on a sign at the gate here. Standards are obviously something the National Bureau of Standards ought to be worried about, but we must approach the establishment of standards gradually, to understand modeling well enough to measure the details without killing the valuable contributions that are more of an art form.

What I want to talk about is in a different direction from what we've been talking about most of the day. Instead of going down, as Dave Kresge was doing, into the guts of the models and trying to understand every module, the code, and the nitty gritty detail, which, of course, is a necessary thing to do, I want to go in the other direction and talk about the usefulness and use of models, and efforts to try to deal with models that way and get back to the idea of simplicity and smallness. For those people who haven't seen it, I refer you to a paper by Art Geoffrion in a recent issue of INTERFACES, in which he states that the purpose of mathematical programming is insight, not numbers. We can say the same thing about large-scale modeling. The purpose is to make things better understood, to give insight into the problems. I always summarized this as advice to decision makers, which I can report never having had any trouble having them accept: if the model produces an answer which is counterintuitive, your optimal decision rule is to assume that the answer is wrong. If you use that decision rule, but continue with the modeling process, sometimes it changes your intuition, because upon investigation you find that the model is right and the intuition was wrong. But then, too, many times you find out that the model is wrong. So if you're forced to make a decision, and your intuition and model don't agree, pick your intuition. I think this approach gets you around a lot of troubles about the black box models, which are complicated and obscure. If you can't explain an answer, after the fact, as Harvey Greenberg was doing very well on some of his problems, then you probably ought not to believe in it until such time as you can explain it.

And that brings me to the subject that I want to talk about -- the Energy Modeling Forum and trying to make models useful. I will try to go quickly through the preliminaries, because they have been gone over several times today. My self-image is that my profession is modeling and my hobby is vegetable gardening, but the empirical results that we have been referring to as "Fromm evidence" might not confirm my view. I might just be part of the modeling hobby show, and I worry about the fact that two out of three models, from the Fromm studies, are never used. I view this a prima facie evidence of a scandal. In a lot of these studies of modeling, the need is suggested to improve information flow between model developers and policy makers. We've been talking about documentation and standards and the quality of the models, etc., but a perfect model, perfectly documented, if complex and not understood by decision makers, is not going to be used. Everybody recognizes that, but things are getting worse, not better. We must make models more useful. It's a problem that a lot of people are concerned about and familiar with, so EPRI proposed the creating of this project, called "the Energy Modeling Forum (EMF)." I will try to explain what the EMF is doing, working with Martin Greenberger through Stanford University (Fig. 1).

We held a workshop last summer, with about a hundred people, who were interested in energy policy modeling, asking questions about how we can go about trying to improve communication, the comparative study of models, and so forth. There were a lot of suggestions, many of them similar to the kind of discussion we're having today. We mulled over several objectives and summarized it all as "working to improve the usefulness and use of energy policy models" (Fig. 2). We're trying to provide a communications link between model users and developers. We're doing so by conducting comparative studies of several energy models, and we're doing this by focusing on a specific energy issue. We can get people to pay attention to decision making, but they are not going to pay attention to abstract discussions of the models. They are not interested in models. They are interested in issues. If we take an issue, that disciplines the conversation. Take several models and try and apply them to that issue, and in the process you learn something about the models and how to use them. Hopefully, you communicate this process to the decision makers, and there also is some reverse flow to the modelers. We illustrate the strengths and the weaknesses of the existing energy models. We also get feedback on requirements for successful application and interpretation, and we identify research areas in modeling.

This is my model of how that process takes place (Fig. 3). We divide the world into model users and model developers. (Of course, some people actually change roles over time.) The heart of the operation is the working group in the center. The working groups are composed of people that probably are called model users, sophisticated model users, and model developers. Examples of a good representative, if we view the Office of Technology Assessment in the model user category, are the energy group in the Office of Technology Assessment that does work for the Congress; Bruce Pasternak, when he was with the FEA, etc. These are fairly senior staff people. We've had only one person that I would classify as a real, live decision maker participating to date, and that's Gordon Corey, who is Vice Chairman of Commonwealth Edison in Chicago. He is interested in participating. Of course, model developers are very willing to participate. At the start of the process, we pick some subject, then organize a working group, and give them some set of models, structure tests for the models, try to understand why we get the results that we get, and explain these results --hopefully in some fairly simple and intuitive way.

emf energy modeling forum

WHAT IS THE STATUS OF POLICY MODELING?

-- DECISION MAKERS AND ANALYSTS ARE FRUSTRATED BY THE DIFFICULTY OF REALIZING THE EARLY PROMISE OF MODELING.

-- STUDIES SHOW THAT 2 OUT OF 3 MODELS ARE NEVER USED.

-- THESE STUDIES AGREE ON THE NEED TO IMPROVE INFORMATION FLOW BETWEEN MODEL DEVELOPERS AND POLICY MAKERS.

-- EPRI PROPOSED THE CREATION OF AN ENERGY MODELING FORUM.

FIGURE 2

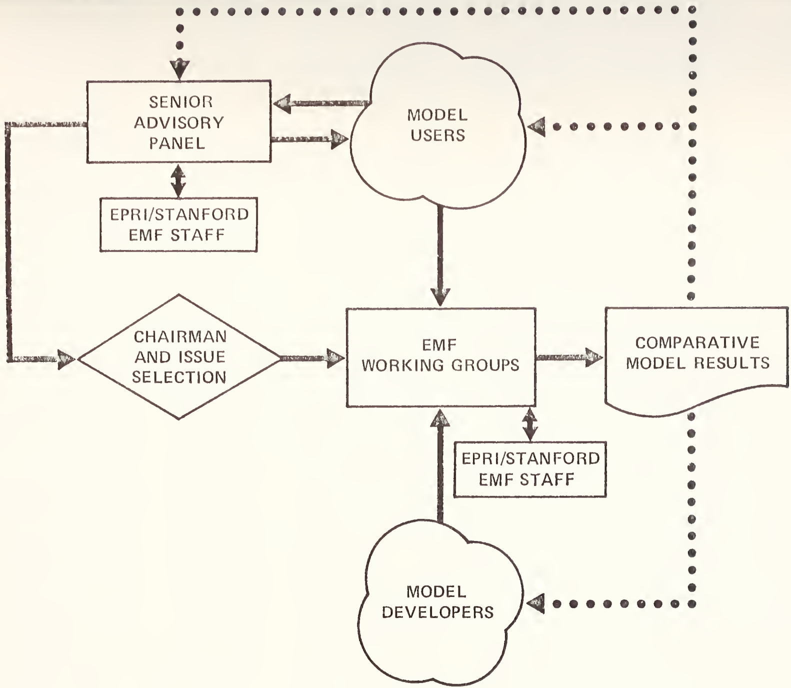

SENIOR ADVISORY PANEL
EPRI/STANFORD EMF STAFF
MODEL USERS
CHAIRMAN AND ISSUE SELECTION
EMF WORKING GROUPS
COMPARATIVE MODEL RESULTS
EPRI/STANFORD EMF STAFF
MODEL DEVELOPERS

Figure 3

That's what we call the comparative model result -- there is a feedback, and so forth. We initiated the process to experiment by actually taking an issue and constructing a working group in an ad hoc manner. We tested it, starting out in September, and we're just about finished now. I will show you quickly some of the results. (I'm speaking only for myself because I couldn't quite get everybody to sign the final draft, but we're almost done.)

We are about to organize an advisers' panel, which is a very crucial element in this operation. We didn't want to organize an advisory panel before we could demonstrate the process, however, so we waited until just recently. And now we do have such a group. We had our first meeting in Washington last week. We have a Senior Advisory Panel steering us, chief executive officers and others at similar levels in a number of private firms, and Congressmen, Senators, etc. It's a fairly senior group of people, and they are going to help us by trying to pick the issues that we ought to be working on. I am pleased to report that we had a very strong level of interest. We've been talking to a lot of people about this, and there's an amazingly strong interest in the kind of activity we're discussing here. Decision makers really are worried about the role of modeling; they're aware of the role that models play; they don't understand them but they would like to; and there's tremendous feedback from the group. I was surprised just how well it worked.

We picked a subject for the first study, and I'm going to try to summarize that to illustrate the process, but I don't want to get bogged down too much in all the details because of time constraints. There is a relationship between energy and the economy (Fig. 4). There's a history as to why we chose this topic: it complements other studies that are going on. It's obviously of some importance, and it meets the criteria for EMF issues: it is important; there are many models that address it; it's controversial, so we can get interested people; and everybody has his own opinion about what the answer is.

We obtained half a dozen models (Fig. 5). We started out with a little more than that, but some of the models turned out to be still in the conceptual stage and actually didn't fit, in the sense of being able to produce numbers. Some modelers were busy and couldn't handle our requirements. But we did get these six to participate, survive the process, and produce the numbers.

While you're reading, let me summarize the characteristics of these models pointing out the variance there. There are different modes of aggregation. Hnylicza has two sectors, energy and everything else; PILOT is an optimization model; Hudson-Jorgenson is a general equilibrium system; we have optimal control approaches; Kennedy is a fixed coefficient system, etc. The models are different, but they all have explicit representation of the energy sector in the economy. We're trying to look at that link.

Then we looked at the models closer and found out they were all the same, that Samuelson's text really does influence the way people think about the economy. If you look at the accounting structure in these models, as opposed to how the links are modeled, you see this: (Fig. 6.) producers and consumers

ENERGY AND THE ECONOMY

THE FIRST EMF STUDY

-- IS GROWTH IN ENERGY CONSUMPTION ESSENTIAL FOR GROWTH IN THE ECONOMY, OR IS THERE FLEXIBILITY FOR ADJUSTING ENERGY UTILIZATION WITHOUT IMPEDING ECONOMIC MOMENTUM?

-- WHAT ARE THE LINKS BETWEEN THE ENERGY SECTOR AND THE REMAINDER OF THE ECONOMY?

-- HOW STRONG IS THE POTENTIAL FEEDBACK FROM ENERGY TO THE ECONOMY?

THE FORUM CONDUCTED TESTS OF SIX DIVERSE MODELS TO ANSWER THESE QUESTIONS.

FIGURE 4

PARTICIPATING EMF MODELS

HUDSON - JORGENSON

Developed at Harvard for the Ford Policy Project and reported in A TIME TO CHOOSE. An econometric model with 9 basic sectors.

WHARTON

Developed at Wharton EFA under the direction of Prof. LARRY KLEIN. Extends the 50 sectors of the Wharton annual model to include energy detail.

HNYILICZA

Developed by Dr. E. HNYILICZA of the MIT Energy Lab. A fully general equilibrium system aggregated to 2 sectors, energy and all other inputs

PILOT

Developed by Prof. GEORGE DANTZIG at Stanford University. Determines activity levels of 23 economic sectors to optimize total consumption.

KENNEDY - NIEMEYER

Developed by Drs. M. KENNEDY and V. NEIMEYER at the University of Texa Concentrates on the impacts of capital and energy in a 9-sector aggregation of the economy. Applied in FEA studies of economic impact of nuclear moratorium.

DRI - ILLINOIS - BROOKHAVEN

Developed through a cooperative effort at Data Resources, Inc., the University of Illinois, and Brookhaven Laboratory. Combines the aggregate substitution of the HUDSON-JORGENSON model with 100-sector detail in the economy and energy inputs. Used in preparation of ERDA plans.

Figure 5

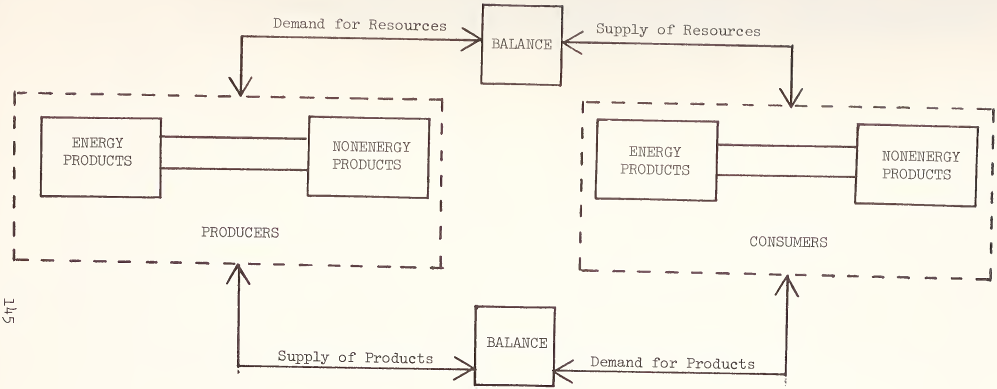


Figure 6 Flow of Products and Resources

labor and goods, energy goods, nonenergy goods, substitution back and forth. You can build a little taxonomy, and then ask, how is it different from the others? And then you can start identifying the key things that are going to drive the models, and where they tend to be the same. That's a nice thing, and we've developed it further. We have a little paper which goes through the process of describing each one of these six models, using this taxonomy.

Then we start testing the models (Fig. 7). This is a little history. It sets up a straw man. We focus on total energy and GNP, and find that they moved together in the past. The question is, are they going to move together in the future? That's the straw man: there's going to be a one-to-one relationship between energy and the economy. If you reduce energy input or dramatically increase its price, does that reduce GNP in the future?

First, we step back a little bit in order to deal with these models in the context of a different question: suppose we change the GNP by some process, e.g., productivity or population, but we don't change the scarcity of energy, do the models show that an increase in economic activity increases the energy demand? We ran the test and we got the affirmative answer (Fig. 8).

But, in practice, we find out something more. All models take the population and the labor force growth as exogenous, all models take technological changes as more or less exogenous, and the sum of these gives the rate of growth in the GNP. With the same assumption for each model, you get the same output. We tested this from model to model, and came out with almost the same numbers.

So we're not using the models to forecast GNP by itself. They're not designed to do that. We're trying to look at the effects of energy scarcity. Just this fact turns out to be a major source of information for the model users.

Looking at the question of energy scarcity, we find that you can develop a simple, intuitively appealing model, which explains what these detailed models are doing under assumptions of reduced energy availability or increased energy prices. The explanation is robust across all six models. I will quickly state what the simple model is; it would take a lot longer to go into the details.

The first point to observe is that the value shares, the expenditures on energy in our economy, are small (Fig. 9). That is important, at least for small changes. The energy values share is only four percent of the economy. It follows that a 10 percent change in energy input produces only a four-fifths of one percent change in the output of the economy, i.e., the value share is the elasticity of output with respect to input. This can be formalized in a simple model, summarized in the Fable of the Elephant and the Rabbit. If you take one elephant, which is the economy, and one rabbit, the energy sector, and put them together to make a stew, you might expect if would come out tasting very much like elephant stew. That analogy is intended to illustrate the importance of the value share (Fig. 10). What we have here is this simple model: (Y) is the aggregate of the economy, (E) the aggregate input of energy

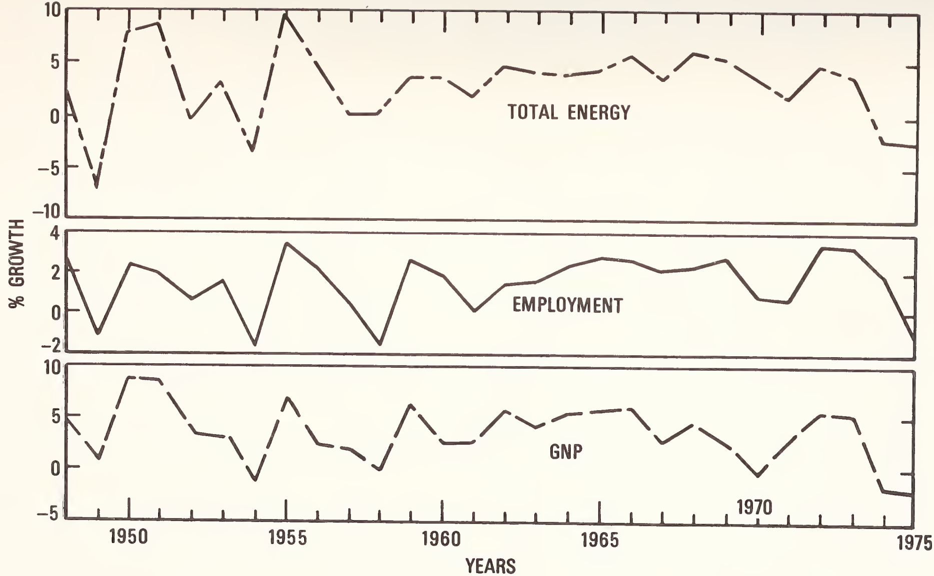
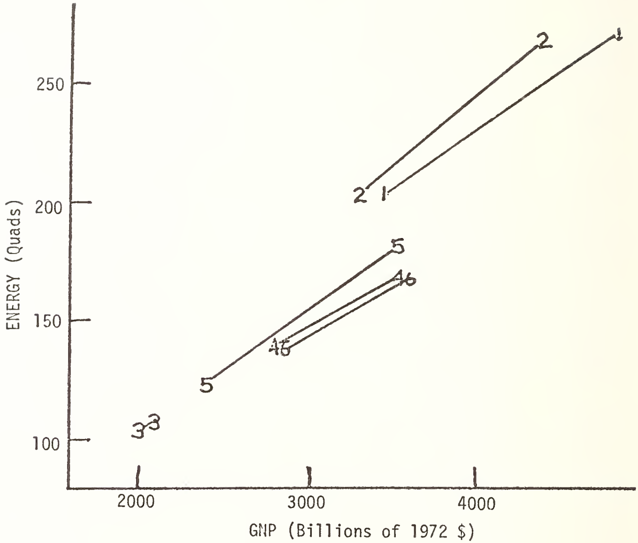


Figure 7 Annual Growth (%): U. S. Energy Consumption, Employment, and GNP

ENERGY RESPONSE TO ECONOMIC ACTIVITY



1: PILOT (2010) 2: KENNEDY-NIEMEYER (2010) 3: WHARTON (1990)
4: HUDSON-JORGENSON (2000) 5: HNYILICZA (2000)
6: DRI-BROOKHAVEN (2000)

Figure 8

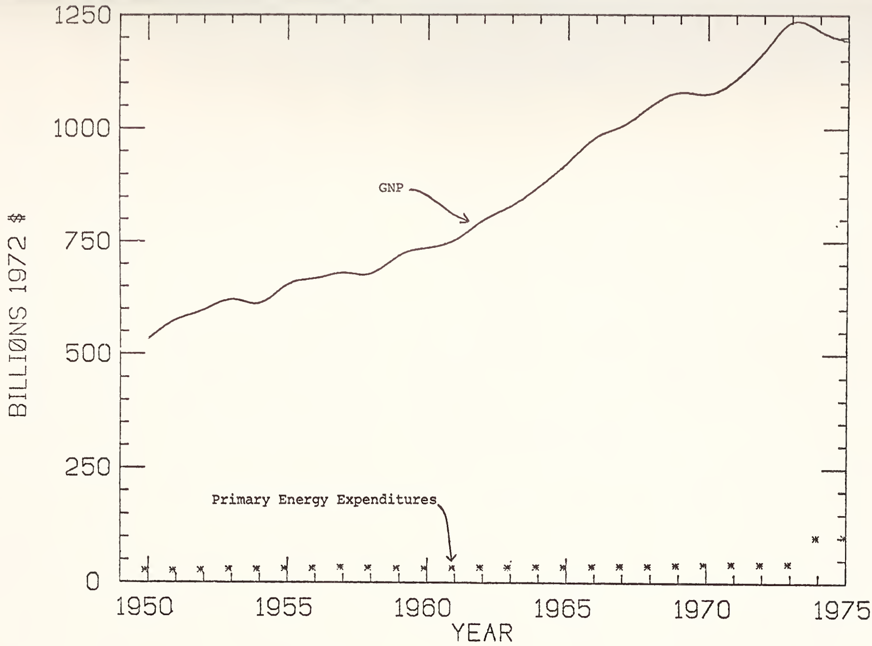

1250
1000
750
500
250
0
BILLIØNS 1972 $
GNP
Primary Energy Expenditures
1950
1955
1960
1965
1970
1975
YEAR

Figure 9

1. $Y = F(E,R)$

2. $Y = P_E\, E + P_R\, R$

3. $\underset{E,R}{\text{Max}}\; F(E,R) - P_E\, E - P_R\, R$

4. $\dfrac{\delta F}{\delta E} = P_E$

5. $\dfrac{\delta F}{\delta E} - \dfrac{E}{Y} = \dfrac{P_E\, E}{Y} = s$

6. $\dfrac{Y}{Y_0} \approx \left(\dfrac{E}{E_0}\right)^s$

Figure 10 Elephant and Rabbit

and (R) the aggregate input of everything else. We're assuming a functional relationship between these inputs and outputs. The value of the inputs equals the value of the outputs. And we assume that the economy optimizes or makes efficient choices, to determine marginal conditions. As a local approximation, the ratio of the value of inputs to the value of output is approximately estimated by the value share. It's a simple analysis and, as a local approximation, it is very good. It is the support for the statement that small changes in energy input produce much less than a proportional change in the output.

This model has been used extensively in other analyses. It is the heart of the economic analysis of the Ford Foundation-Mitre study, which came out recently, NUCLEAR POWER: ISSUES AND CHOICES. This simple model is a good starting point, but it's not the final answer. It doesn't recognize the effects of less than perfect substitution. We might find that, in the use of energy, we couldn't completely substitute capital and labor, and it becomes necessary to look at how much substitution could occur and how it would affect the economy. If you take the simple model and use the production function, we can develop a beginning representation of the flexibility of input use (Fig. 11). There is a measure of index of substitution, and it turns out that the models are very different in the way that they treat this substitution, either by assumption or because of empirical work. This is very important, and the index of substitution becomes the index of that economic impact. For low values of the index, say 0.1, a 50 percent reduction in energy input can yield a 28 percent reduction in GNP, and at the high range, say 0.7, there is only a one percent reduction in GNP. The models are very sensitive to this particular parameter.

This, again, is to prove that we can analyze the American economy in two equations (Fig. 12). Here we have a specific formula for the production function, and this is the elasticity of substitution (σ). You take the marginal conditions, manipulate that equation to get something in terms of energy, Y, σ, and E, and make some plausible assumptions about other inputs; we can change the energy input and solve the system for GNP. The picture looks like this (Fig. 13), energy vs. GNP. There are certain important points here. One is that the relationship is insensitive to small changes in energy input, no matter what you assume. If you take away one Btu, the benefit lost is a little output, but you also don't have to pay for that Btu of energy. At the margin, these are equal, so the derivative of GNP is zero. That argument already is a revelation: many people don't think of the problem this way.

But for the larger energy changes, the impact depends very crucially on this assumption about the elasticity of substitution. At the lower values, the GNP drops off quite fast. In the higher range, it looks like you really can reduce energy input without much impact on the economy.

This overview prepares the way for comparing the models. We now run the models, and they produce data. We take the data, the output of the models, and we can estimate the elasticity of substitution implicit in the models, and use their graph to compare the flexibility in the models. We find that the models fall into two categories, based on their assumption (Fig. 14). There are two models which assume that there is no substitution

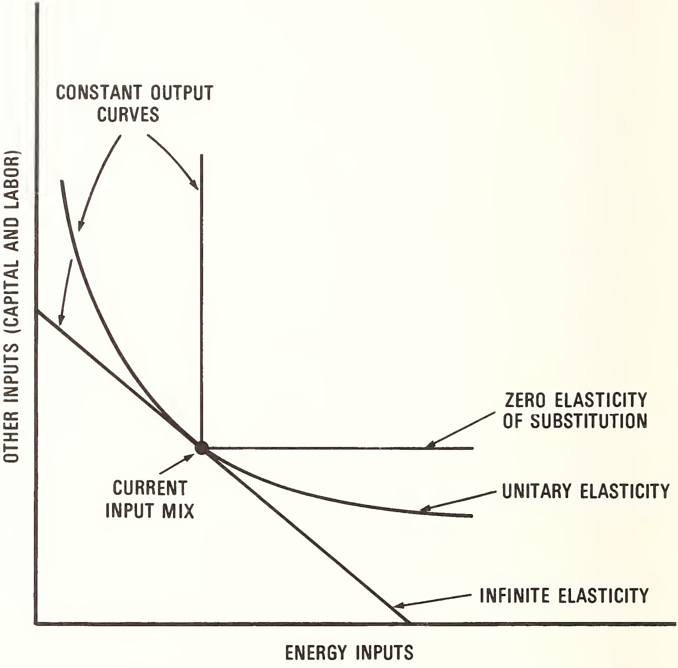


Figure 11 The Elasticity of Substitution Concept

1. $Y^{\frac{\sigma-1}{\sigma}} = a\, E^{\frac{\sigma-1}{\sigma}} + b\, R^{\frac{\sigma-1}{\sigma}}$

2. $\frac{\delta F}{\delta E} = a \left(\frac{Y}{E}\right)^{1/\sigma} = P_E$

3. $E = a^{\sigma}\, Y(P_E)^{-\sigma}$

4. $\sigma \neq 0, 1, \infty$

Similar relations for R fix parameters for a baseline forecast.

Figure 12 Production Function Analysis

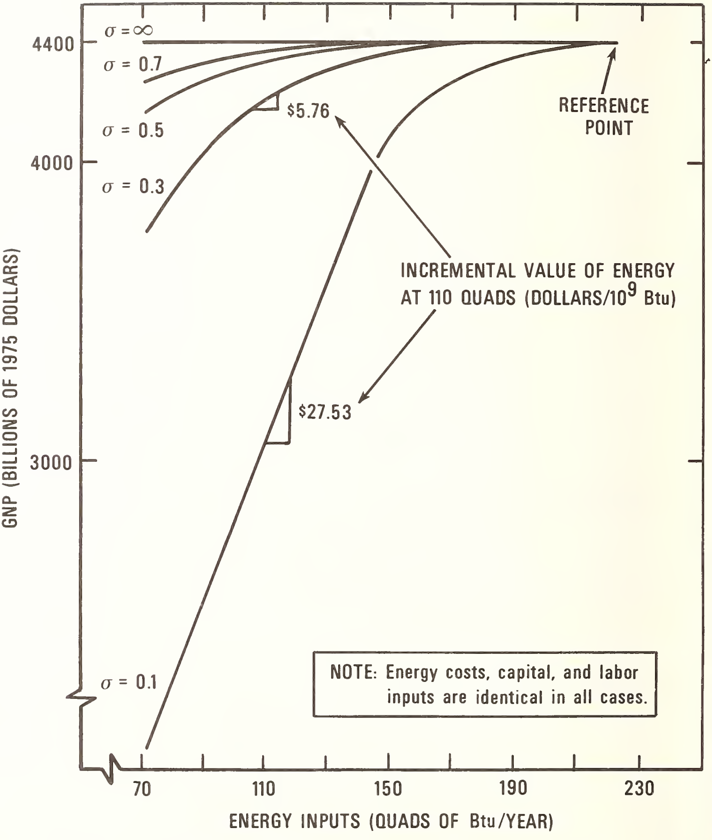


Figure 13 Economic Impacts of Energy Reductions in the Year 2010 for Various Elasticities of Substitution (σ)

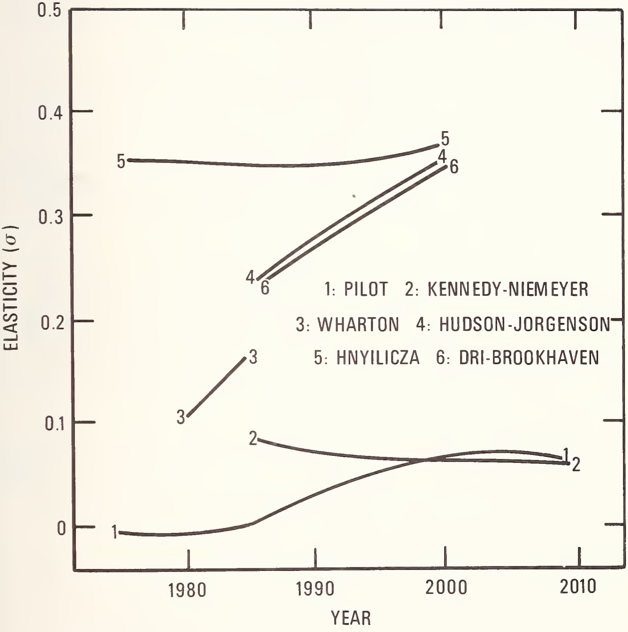


Figure 14 Aggregate Elasticity of Substitution

explicitly, except for a couple of very small sectors. When we estimate their elasticities, we get a very low value. The rest of the models have fairly flexible structures, and they yield higher elasticities. This is the Wharton Model, and these data were available only yesterday. In the long run, these four models all end up in about the same range. If you remember, in the previous figure, that is a fairly high elasticity of substitution. It indicates that there is a great deal of flexibility in the economy, and the reduction in energy produces economic impacts that are small proportionally, albeit large absolutely.

When we tried to validate this simple model as an explanation of the full models, we found it didn't work too well. Initially, this was puzzling, until Dale Jorgenson suggested an interaction, over time, with capital formation. We are changing energy input over time, and that reduces the marginal productivity of capital. Investors want to keep their rate of return about the same, so they lower their investments. Over time, less capital accumulates, and you end up in 2010 with a lower productive capacity. That's why the GNP drops off more than the simple little analysis illustrates, where capital is held constant.

Fortunately, this explanation could be tested in our simple models. Just keep the rate of return on capital constant in the production function. We did this, and it turned out that Jorgenson was right, and it makes a big difference (Fig. 15). There's the worst case. You can see first the case where the capital inputs stay constant. GNP is very sensitive, but the model does not pick up the roughly four percent drop in the GNP that Jorgenson's model predicted. But, when we kept the capital return constant, we got much closer to the result of the full model.

Where this leaves us is that for our original purpose, which is basically pedagogical, we have a usable, simple few equations that summarize acceptably the behavior of the detailed models. We start with a value share, and a measure of substitution. This is the key to the explanation of the aggregate behavior of these models. We can compare those models in a reasonable way; people can understand them and then, hopefully, use them.

The modelers are in agreement with this analysis, but they recognize that it might appear in opposition to the conventional wisdom, that energy is important. We are trying to develop other explanations; one of them is that a small percentage of a big number is still a big number. Even though large reductions in energy may produce only a one, two, or three percent reduction in GNP, which doesn't sound like much, it is a large absolute impact. In present value terms, it is a lot more than ERDA's budget, a lot more than 10 years worth of ERDA's budget. So it is a significant impact, and we should worry about it.

A limitation of this summary is that the aggregate analysis doesn't tell you what is happening in individual sectors. We might be curtailing the aluminum industry, and this could be a serious problem.

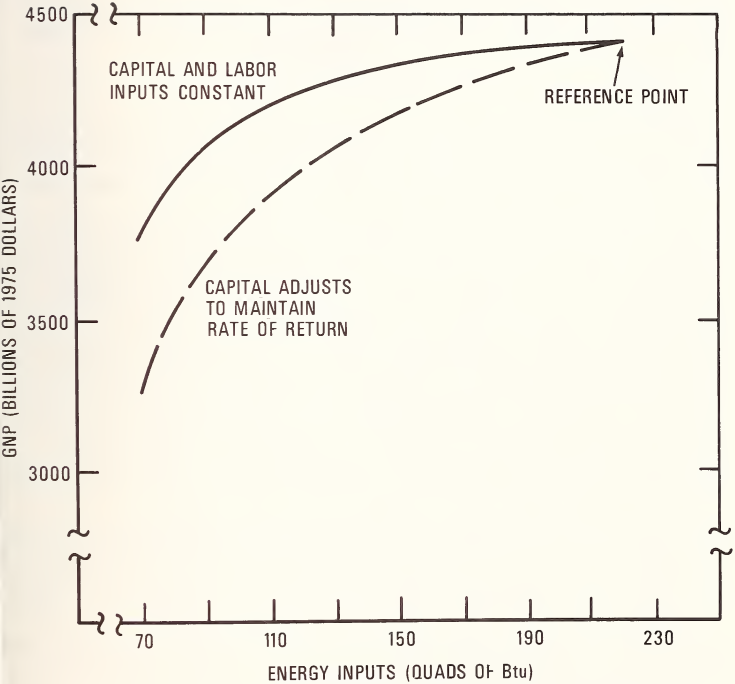


Figure 15 Economic Impact of Energy Scarcity in the Year 2010 for Alternative Capital Assumptions (Elasticity of Substitution σ = 0.3)

This summarizes our main conclusions (Fig. 16). These meet the criticism of Dave Kresge: to say something positive about the models. This is something positive. The models provide meaningful analysis about the links among energy, capital, labor, and any other materials, in order to determine potential GNP. Of course, I've used a special adjustive, "potential" GNP. What I mean by this is that all models assume full employment. They're all long-run models. Many Congressmen aren't much interested in the long-run, full employment, and so every decision they make is dictated by what happens in the short-run employment rates.

QUESTION: What is the meaning of "meaningful?"

HOGAN: Well, meaningful in my terminology means that the resulting relationships are intuitively plausible, given the theory which you adopt, that you can relate the results to the data, and that, across a range of models which have very different aggregation levels and structure, you get the same kind of results being produced.

The simple analysis, the simple model that I talked about, aids in understanding. You can understand, more or less, what is going on in terms of substitution and the capital-energy links. We can explain what is happening in the aggregate sense in these very detailed, complicated models. We also tried to list the things that need to be done. This list actually is quite long. Suppose we have another embargo, what's going to happen to the economy? Well, that's not a long-run question; it's a short-run question and the results are very different, and these models aren't capable of dealing with that. They don't talk about income distribution; they don't talk about the distribution of ownership of different kinds of industries, and how that affects the economy, etc. We could go on; there are many problems that you might think that the models could deal with, given that they concentrate on energy and the economy, which they don't. They simplify the world in order to analyze a very important question. In any model, by definition, making a simplification means leaving out many things.

The decision making group of the advisory panel, when reviewing this study, had a lot of constructive comments. But it was quite clear that these omissions in the models dominated their concern. They are interested in these long-run issues; they don't want to throw away this information, but they are 10 times as interested in the short-run events.

QUESTION: If they lacked those, if the model didn't have any, was it known before the analysis, in general --

HOGAN: Oh, they certainly were known to the modelers, but they were not known to the people in the "group of users." The users were making statements like, "This really is helpful to me to understand the model, to have it expressed in a way that the terminology is usable... It's a tremendous effort to explain things in simple terms." The users don't deal in modeler's jargon. We have taken the view that it's the modelers problem, and that we have to deal with that because you cannot expect the decision makers, or even their senior staff people, to become sophisticated in the mathematics. We've got energy modeling forum

EXECUTIVE SUMMARY

ENERGY AND THE ECONOMY

The Electric Power Research Institute created the Energy Modeling Forum (EMF) to improve the usefulness of energy models.

The major conclusions derived from the models of energy and the economy include:

> "In the presence of constant energy prices, increases in economic activity produce similar increases in energy demands..."
>
> "Higher energy prices or reduced energy utilization need not produce proportional reductions in aggregate economic output..."
>
> "The models do show some reductions in economic output resulting from higher energy prices. The magnitudes of these reductions are very sensitive to the substitution assumptions implicit in the models."
>
> "The benefits of energy substitution may be lost in part if energy scarcity impedes capital formation."

Figure 16

to simplify constantly and reduce the jargon. It's time consuming and very demanding but the users appreciate it. I think it did help in in bridging this kind of a communication gap.

QUESTION: What size effort is involved in this?

HOGAN: There are about 30 people in the working group directly, and I would say that 15 of those people worked on it fairly seriously and 15 came to the meetings to help with the critique, writing, and so forth.

QUESTION: Were any of these full-time?

HOGAN: No, not full-time. There were about three people who worked on it full time, and then the rest of those 15 people who spent maybe as little as two or three days a month and as much as a couple of days a week, over a period of about six months. A couple of the modelers, in particular, who had large systems and were having a hard time implementing the tests were really spending a lot of time working on it.

QUESTION: I thought some modelers and analysts recognize what you said, i.e., that some models don't do well with environmental impact and new technologies, but they say the biggest problem is trying to crank that analysis in with the availability of data.

HOGAN: We don't understand it. Environmental impacts that most people worry about, with few exceptions, are very much local problems. It's not the number of power plants so much as it's the number of power plants and where do you put them. And so, if your model doesn't distinguish between locations, it's not telling you what's going on in the environmental problems.

QUESTION: So what can you do?

HOGAN: I don't have any answers to that right now. All I am doing is summarizing something everybody knows and observes. But we're not saying very much constructive. The best thing you can do is to talk about emissions as opposed to pollution, and emissions are very hard numbers to convey, to assign any meaning to.

QUESTION: The answer that was given back to me was, okay, now it's hard enough to try to predict the number of power plants, the nuclear power plants that are going to be in the country in the year 2000, but to try and predict what part of the country and what their base is going to be, the language, I think, is going to be even more complex.

HOGAN: I don't have any input on that either on the models that you are talking about. I'm not surprised. It's not the kind of thing where I can think of an alternative. Certainly, it's going to be with us, and someone, somehow will make his own trade-offs. It will need more work on model and data development.

So this is very different, you know, from what Dave Kresge was talking about. As a matter of fact, in terms of understanding the models, in terms of complexity, it creates many more complicated questions from very simple questions that we asked, and it is complicated to do it, because we have to have several different models simultaneously and then try to explain them all with some kind of a framework. But the models, themselves, are much more complicated, and we haven't scratched the surface yet, even with the energy-economy models.

QUESTION: This activity sounds like it might be an interesting program. Do you get a sense that there are some special features here, partly because of the needs of special interests?

HOGAN: No, I don't think there's any special characteristic. This is not model assessment or evaluation. What we're really trying to do is explain what's out there. For example, all these runs were produced by the modelers: it's completely a backroom operation. We say, this is the scenario we want you to run, and they did exactly as expected. They call up and say, well, I'm going to be late because I ran it and it didn't work right, and I've got to fix these things, and so forth. What I want them to give me is their best view of the proper use of their model, not run a rigidly controlled, scientific test of the code. The modelers certainly weren't bashful about fixing up places where they had some trouble, but that wasn't what we were trying to look at. We're not trying to communicate what we think the models say. We might take it for granted that everybody knows everything about a model that's been around for years, but it is clear that everybody does not know, even though the model has been used for a long time. The EMF is trying to improve communications and understanding of what exists.

mod
and
of
rem

tha
dep
muc
def
thi
mod
As
mode

You
rec
it
pol
a fu
expa
mode
the
poi
your
as i
over
that
to l
what
you
trib
is b
degr
eith
ofte
betw

put
the
were
that
the

MODELS IN THE POLICY PROCESS: A FRAMEWORK

Brian Crissey

I would like to address my remarks from two different areas: one is the model and policy process work that is reflected in the book of the same name, and the second is the model analysis that took place about two years ago as part of my dissertation at Johns Hopkins. I would like to start off with a few remarks.

A large-scale problem does not need a large-scale model. The reason I say that is that there is a scale of appropriateness for anything that you do, depending on how much you know about what you're doing. If you don't know very much, or if the data is inaccurate or there are problems with the generation or definition of the variables and so on, then obviously you start combining these things and computing, calculating, et cetera. For increasingly more complex models, the validity with which your are treating these answers will decline. As a small example, Figure 1 is a conceptual diagram of error propagation in models.

You start out with an initial value of some answer coming out of some model. You're using the model to decide between two different policies that are to be recommended, Policy A or Policy B. Notice that when you run it with Policy A, it comes out below where Policy B comes out. Therefore, Policy B is the better policy. But if you then apply numerical analysis to obtain confidence levels as a function of propagation of errors, you notice that the confidnce intervals expand almost exponentially over the computed time, especially for iterative models. It's not so bad for models that compute a specific year without using the previous year's input. But you can see there that the outputs are really points from distributions that depend upon how accurately you built the model, your errors, the errors in the data, and so forth, so that the outputs behave as if they were means of distributions that overlap. You can imagine that they overlap so much that the outputs behave as if they were means of distribution that overlap. You can imagine that they overlap so much that you really have to look at the probability that Policy B is better than Policy A, in which case what you really have conceptually is a three-dimensional surface of merit where you have the Policy B distribution projected along the X axis and Policy A distribution projected along the Y axis. To determine the probability that Policy B is better than Policy A would be equivalent to passing a vertical plane at 45 degrees through that surface and computing the volumes under the surface on either side. Since the means are often fairly close and the variances are often quite large, in many cases the models cannot tell you much about deciding between policies.

Another example of this is what we call a Bonini paradox. C. P. Bonini put out a model for a firm back in the '60's that would enable him to reproduce the behavior of that firm, and thereby determine what the problems of the firm were, and why it wasn't making money, et cetera. He built this model so well that he was able to duplicate almost all of the important characteristics of the behavior of the firm. When he finished he couldn't understand his model,

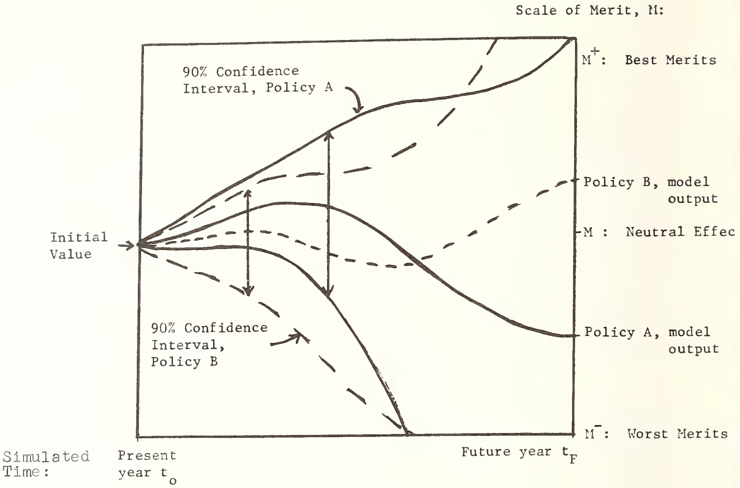


Figure 1

Error Propagation Effects in a Policy Model

A policy model is employed to recommend Policy A or Policy B according to the projected merits of each policy in future year t_F . A superficial look at the model's output for year t_F shows Policy B having positive merits and Policy A negative merits, yet one cannot definitely conclude that B will be better than A because propagated errors have caused the confidence interval around the output to greatly increase over simulated time.

because it was such a complex firm. He could reproduce the firm but he could understand the thing, so it didn't help much. The model needs to be at the level of your understanding. Models are for learning, so learn from them but don't believe them. That's another axiom that I would throw out at you.

As models get larger their maintenance becomes a problem. There is a phenomenon you might call the owner-builder phenomenon. A person builds a model and essentially owns it by putting his name on it. As models age they tend to get larger (I've seen very few get smaller), and the older they are the more complex they are. If the owner-builder then decides to leave and someone else is asked to take over, you've got a very complex model that is very difficult to understand. Usually the owner-builder is about the only person that can understand the model in sufficient depth to answer all the questions that might come up about it. And as a result, the owner-builder then becomes an expert on the model or the field to which the model applies. And from this you could say that although experts create models, models create experts and we have examples of this that are shown in the book, MODELS AND THE POLICY PROCESS. One short example is that of Jan Leendertse, a hydrologist. He built a model of the Jamaica Bay area near New York City; a very fine three-dimensional complex model of the flow of pollutants in a tidal estuary. We went to New York and asked a lot of people in the government various questions about this model and how it has been used to make policy decisions. We found five different areas where policy decisions have been made and for which people asserted that the model had been used to make those decisions. In fact, in none of those cases was it true that the model had been run and then the results of the model had been used to make the decision. What had happened was that in every case the policy situation was moving so fast that the model could not be updated, run and validated fast enough to respond to the ever-changing demands in the policy situations. So in fact what happened was instead of the model being run, the owner-builder, Leendertse, was asked his opinion of what would have happened if he had run the model. So he said, "Well, I think it would have done this...." Chances are he was at least in the right ballpark because, as a result of working very closely with Jamaica Bay and his model, he did gain a pretty good understanding of the physics of the Bay. He knew where the dead spots were and where the flows were and so on. He did become an expert as a result of the model.

There is a phenomenon called the artichoke phenomenon -- which is not our term. That is the name for the process by which models grow in response to criticism. Somebody says, oh, but you didn't include X and so you slap another horny plate on there. Eventually you have the artichoke which is all sharp and pointy on the outside and the really interesting things are down inside where you can't see them. The thing keeps on growing until finally there are so many points on the outside that no further critics wish to handle it, hence the criticism stops and the model is complete. Further, the model builder is in complete control of the use of the model.

In my experiences at looking at various models, trying to analyze them, and in my own model analysis of the MacAvoy-Pindyck model, I found that it does take a considerable length of time to get into these models. The figure of five or six months was put out just yesterday, and I think Dave Kresge was right on that. It took me about six months to totally understand the MacAvoy-Pindyck model which, as you know, is a large model in terms of the number of equations, there are several thousand equations. The question that obviously comes from this is, "If this is the prerequisite for understanding models as they are built today, who can afford this, and who is going to pay for this kind of analysis?"

QUESTION: Could you say a few more words about the term "understanding?"

CRISSEY: O.K. In my case, what I meant by that was knowing precisely what every equation was in the computer code of the model, why it was there and how it behaved independently of the rest of the model. That's all.

QUESTION: So you think it was an understanding of the static structure rather than dynamic?

CRISSEY: Right. So that you could answer questions like whether the model builder assumed variable X is equal to A or B -- where is that in the model? Oh, that's equation No. 4, for example. You can go right there and say, well, here is the equation and obviously they assume it equals A. I was interested in that level of understanding: what assumptions was the modeler making? The behavioral understanding is another step beyond that. So it does take maybe a half man-year to understand most large models.

Another factor in the size of the models is that models that are expert-dependent (owner-builder type models) are vulnerable to the whims of the modeler. As an example of this, I was in the Army a couple of years and I was dealing with computer modelers from the Pentagon. One of them was responsible for a pretty large complex processing program that took input records and output records of people coming into and out of the service, promotions and pay records, and all sorts of things and kept track of them. He built the program in such a complex way that he was able to insert at one place, essentially Statement 100 to back up all the tapes and erase them, and in another place put in a statement that said if a random number equals time-of-day then go to Statement 100. About three years after he left the Pentagon that struck and backed up and erased all the tapes. This, of course, is a rather strange example. I don't think modelers are malicious the way draftees sometimes are in the Pentagon, but the fact remains that any time a particular person has solitary control over the complex instrumentalities of the program or the model, then you have to wonder what is in there. And I wonder also why there has not been more talk about structured programming for models and things like this that are designed to bring the logic of the program out into the open and make it very clear, so that another person can pick that up and go from the top down and understand what the actual program of the model is.

I think Dave Kresge mentioned that there is prestige in model building and no prestige in model analysis or running. That's another thing that I think we have to consider in all this. Until we change that or do something about it we are going to continue having a lot of duplication. We found in the "Models and Policy Process" book that there is much reinvention of the wheel that goes on, and I think this inefficiency is directly related to the fact that models are overly complex, the artichoke phenomenon, lack of documentation, lack of structured programs and things like that.

QUESTION: Is reinvention necessarily bad? The way you described the models, it's less effort to reinvent one.

CRISSEY: That's why they did it, I'm sure. What I would suggest is that it shouldn't be the case and if we had started from the very beginning using standard model building procedures, maybe top-down structured programs and things like that, it would make another person more likely to be able to get into the model in less time. Then there would be less reinvention of the wheel. But at the present time I think there is very little choice. Most people do reinvent the wheel.

QUESTION: I could think nothing less creative than a bureaucratic structure on how to build models.

CRISSEY: Well, all right. We can talk about that later.

Let me go into model systems. There is no such thing as validity for models of real systems. The reason is that validity means passing all possible tests and you can always generate more tests. One test you can generate is the wait-and-see test and that doesn't help. That's usually the one test that ultimately must be passed in order to have valid models, but then it's too late to change the model if it is wrong. So validity is, I think, an inappropriate concept. I think a better concept is confidence in a model and confidence is raised by passing more and more of these tests. If you pass a hundred of them, the chances are likely that you will pass the next one, more so than if you can only pass one test--

QUESTION: Did you say competence or confidence?

CRISSEY: Confidence in the model, its operation, structure, theory, data and all that sort of thing.

Another aspect of this is that I am dealing with policy models, models that are designed to be useful in the establishment of new policies, creation of recommendations to policy makers and so on. The very essence of policy is that there is controversy, that people disagree about a lot of things. As a result, there are really no answers to these controversial areas, there are only opinions. And as a result, a model of a policy area is really just mechanized opinion and we ought not to forget that. You can take any area that is being assessed by a policy model and go back to the policy side and look at the debate on Capitol Hill or wherever and see who is saying what and what they are disagreeing about. You can get a list of these policy areas where there is contention, and so long as there are experts differing on Capitol Hill, it is presumptuous for a modeler to go in and decide for himself that the controversy should be resolved this way or that way. If he does that, then he has mechanized his own opinion. That is what his model will be and that is the way his model will be received when he tries to use it to make policy recommendations. So let's not forget that.

QUESTION: I think we have a basic disagreement. You're, I think, putting forth the hypothesis that the model is to support any one opinion or any one policy or both of them. It is my contention that a model is there to find the implications of various policy proposals and find out the good and the bad of the various alternatives.

CRISSEY: That's right, I'll agree with that. You are, however, thinking about what comes OUT of the model, while I am thinking about what goes INTO the model -- the assumptions that drive the model's results towards some policies and away from others. There is often an implicit assumption among modelers that in the best of all possible worlds, the model speaks truth and then that truth gets put into the policy. But in the real world, models often speak opinion disguised as truth, while policy makers listen only to what they want to hear.

Let me tell you a little bit about a case-in-point that I have a lot of experience with and that's the MacAvoy-Pindyck natural gas model. If you just didn't know anything about models and you just read the policy debate on Capitol Hill regarding natural gas you'd find out that the MacAvoy-Pindyck model is the deregulation model. It comes out and says "deregulate." It doesn't say that there are pros and cons about the thing, but it really takes a stand and it says we should deregulate. That may or may not be the right thing, but the fact is that it does take a stand and there are many other models that are used that way. Once MacAvoy and Pindyck selected a range of alternatives to look at, that range of policies was processed through their mechanized opinion model and it was ranked by a chosen scale of merit. They come out in favor of deregulation, because one of their opinions that was mechanized in the model was that the only important thing about natural gas was the equalization of supply and demand. They sought only to minimize excess demand for natural gas. There are many other values and criteria, like social equity or resource conservation that might be considered on Capitol Hill that were not considered in the model. There are many other examples like that. If market clearing is your only criterion and you're going to evaluate the entire spectrum of policies, then one is going to come out on top. That one turns out to be the same one that the modelers favored before they built the model. The model did not form their opinions. Their opinions formed the model.

QUESTION: Are you sure MacAvoy-Pindyck favored derregulation?

CRISSEY: I'm basing this on what MacAvoy wrote before he built the model.

QUESTION: Earlier you said, Brian, that the more you run the model, run different kinds of problems, the more confidence you get. Why shouldn't some of the other people run the model with different sets of input and thereby get different outputs?

CRISSEY: This is the area of third-party model analysis. Other people ought to get involved because different people have different opinions and if you went into the model you'd do something different than if I went in because you see it differently. Everybody sees it differently. So yes, I would agree that others should run the model.

QUESTION: One other thing about the MacAvoy-Pindyck application that you alluded to is that the caveats get lost along the way. The theory was reasonable and the application had the right expected values but they had low statistical significance in some of the crucial parameters.

CRISSEY: Right. One example is the controversy about whether the natural gas industry is competitive or whether it's monopolistic or oligopolistic or whatever. They did a little bit of analysis on that and decided it was competitive. They then built a model on that line despite the continuing debate on Capitol Hill. If Congress had decided in that debate that the industry was thoroughly competitive, then the MacAvoy-Pindyck model might have been applicable. But without going back to their model and having a little dial on the input that says the industry was monopolistic or oligopolistic, or whatever, the model is not up to the demands of policy.

Three examples of how opinions differ in models are counter-modeling, adversary modeling and multi-modeling. These are terms that we use in the book, and counter-modeling is our term for taking a particular model such as one that is now on the shelf and putting a different opinion into it. "I don't think he does the pricing right, so I will put in some pricing feedback right here." Counter-modeling means taking that model and fixing it a little bit and running a policy on it, getting different answers and then coming back to the original modeler and saying, "Well, look, here is your model, this disproves your theory because I did this." A lot of that is documented in the book and it's directly a question of the difference of opinions that go into the models.

QUESTION: Does this correspond to sensitivity analysis?

CRISSEY: Not in the usual sense of the term because sensitivity analysis usually embodies the structure of the model and the differences of opinion are usually broader than that, like competition versus oligopoly. That is a structural thing. We take the whole structure of the model and follow the same kind of idea where we have the structure of the model for this opinion, the structure of the model for that opinion and so on, and then the sensitivity to opinions is what you're testing. But this is different than the sensitivity to a specific coefficient in a model where the structure is invariant.

QUESTION: I was wondering if it was different in some way -- and it's not clear to me on what that is. Changing the essential structure of the model, how does that differ?

CRISSEY: O.K. In your model there are the standard parts that anybody doing modeling in that area might agree to include. But there are some smaller parts that will reflect the various opinions that come from the policy base. It is these that are changed to assess the sensitivity to opinion.

QUESTION: Then it's important that the model be designed for modularity in structure?

CRISSEY: Exactly. That's right. That's an important conclusion you should draw from this. Any time you are doing a model of a policy area, you need to go back to the policy area itself, examine its base and see what the contention points are, what things are being debated, and make very sure that in your model you can change the assumptions relevant to the policy disagreement. Because if you cannot change those you're fixed in an invariable base. You are not going to be able to reflect the diversity of opinion that you need to.

Adversary modeling is similar to counter-modeling, but you use totally different models. A real quick example of that is a coal power plant that was going to be built south of Baltimore a few years ago called Brandon Shores. It had to fit the new Maryland Power Plant Siting Act, so the Baltimore Gas and Electric Company hired N.U.S., which is right down the highway here, to do a computer model of the air pollution impact of this new plant. They came out with very nice results: the new plant is going to clean up the air. The State of Maryland was still skeptical, so they hired the Applied Physics Lab and Martin-Marietta to do models of that situation and, as long as everyone was doing modeling, the Bureau of Air Quality Control came in with their model. We had four different models in here, each one trying to say what the air pollution impact of this plant would be. There were differences of opinion that were subtle until I went in and actually saw what the models assumed. These differences made the final answers, in some cases, differ significantly between models depending on whether they were sponsored by somebody who wanted the plant built or somebody that wanted air quality pretty high. And some of these assumptions were very gross, like do we assume that this plant is a plant all by itself or do we include in our analysis the dirty coal plant next door, which is really the same piece of ground but it's called a different name? Is it one composite plant that's half dirty, half clean, or is it only one new clean plant? Which sets of data do you use? What sort of meteorological data do you use? In one model there is a certain class of air turbulence that was ignored because it only occurred five or ten percent of the time. The trouble was that that was the class of air turbulence where the smoke plume most often touched the ground. If you ignore that one class even though it's infrequent, the aggregate over time is going to be affected in terms of the air pollution concentration at ground level. In adversary modeling, different models and different assumptions just come at each other with different opinions. From the viewpoint of the policy arena the result seems to be that it is obvious that these models are mechanized opinions because they don't agree; they're way off from each other. So as long as this is still the case, we still have work to do in designing models that are politically defensible.

Multi-modeling I won't say much about. Bill Hogan talked about that a little bit without naming that term yesterday. Multi-modeling is using a lot of models together in consort to try to achieve consistency and agreement, and to see what the differences in the models are.

QUESTION: You said as far as policy modeling is concerned you still think you have to do considerable work in that area. You're always going to have policy modeling --

CRISSEY: What I was talking about is building models in such a way that from the very beginning they can reflect adequately the demands from the policy arena. You can build models so that you wouldn't expect to have the situation where another model could come up and destroy your model by coming out diametrically opposed to you. You ought to be able to say, "Well, your asssumption was this, so we'll twiddle the assumption here." You should get something approximating their answer, because the differences are largely in the opinions and assumptions you work with.

QUESTION: I know that, but I daresay for any time somebody comes up with a model that has a certain opinion, somebody else can make a model that will come up with the opposite.

CRISSEY: That's right. That's why we need top-down structured programming and obvious clarity in structure of models. In that way we can identify the assumptions that explain the differences and direct the attention of people towards making the right assumptions.

QUESTION: I think there is a point there. You're sort of implying that mechanized opinion is what's going on now and that's probably bad and that maybe we ought to try to get away from that, and yet the other view is that you want to just raise people's consciousness and say that that's what it's always going to be, but that's not bad, and therefore you want to try to understand that.

CRISSEY: Yes. Let me take a middle ground. I agree with both of those. I would say that there's nothing bad in mechanized opinion if there is nothing else. And therefore, we ought to just raise people's consciousness, accept that and deal with it. But we can do more than mechanize a single opinion. What is the range of opinions in the problem area, and how will they affect the structure of the model? We can be straightforward on that and I think it may be a big step forward, to try to match the demands of the policy process.

Figure 2 is an example of a diagram much like what Dave had on the board yesterday, that comes from the work I did two years ago. Let met briefly take you through this and start with what we will call the "reference system" (1). This is the thing that you are trying to model. In this case it's the natural gas supply and demand. The modelers (2) have perceptions about the reference system and they create a policy model (3). The policy analysis area (11) will eventually make policy decisions. From this area is derived a set (5) of politically viable policies that can be considered. These are called the "policy options or levers." One principle that needs to be followed in policy modeling is that

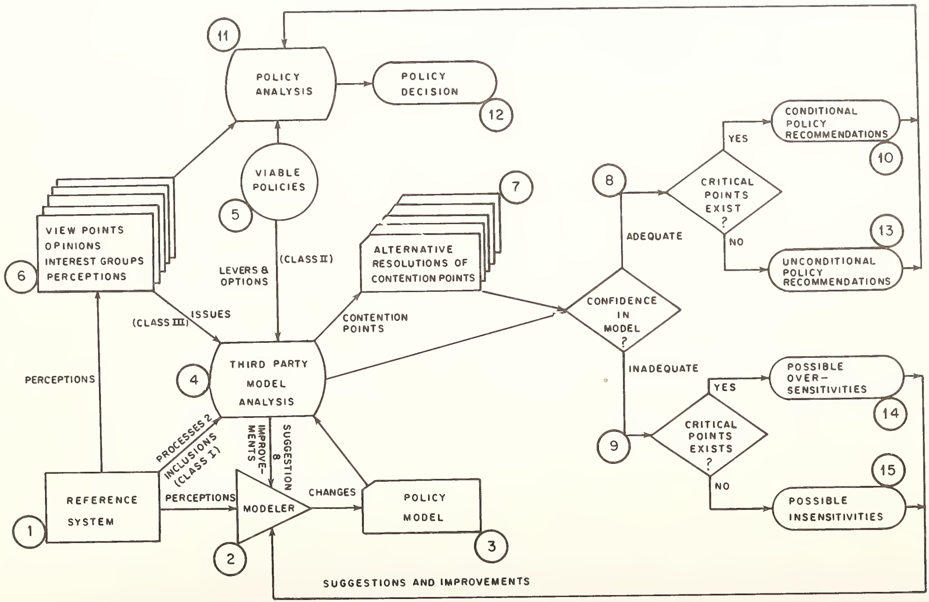


Figure 2 The Rational Framework for the Use of Computer Simulation Models in a Policy Context

when you look at the policy process and you're trying to make a model that's going to be useful to it, you had better make sure that you can push the right buttons in your model. Looking at the reference system is like making sure that a car can move forward and turn. Looking at the viable policies is like making sure that it has a steering wheel, that it has the thing that you need to have in order to be relevant to the policy arena. So there are many examples of models that will tell you a lot about a particular area but they won't tell the policy maker anything about choosing between Policy A and Policy B, because they're invisible to the model.

QUESTION: I guess what you're saying is you should have a listing of what the policy questions are before you start out?

CRISSEY: That's right. That's where you're going so I think that's where you have to start.

QUESTION: Do you think that the government in their RFP's would state what policies they want to attempt to examine with the models?

CRISSEY: No, I'm suggesting something different. In natural gas, if you go to Congress and look at the total set of bills which have ever been produced on natural gas, you will see that they are all variations of several themes, and there are few new themes. Once in a while there is a new theme, but if you can match all the past themes, all new legislation is some kind of a complex combination of the old themes.

There are viewpoints, opinions, interest groups, and perceptions that are affecting the policy arena. These create issues or contention points which can be identified in the model. The various points of view on the issues can be associated with alternative resolutions of the contention points. And each of these various resolutions of contention points can be applied to that model to see what is the effect on the model. By looking at the effect on the model deriving from points of view, third-party analysis ought to be able to raise or lower confidence in the model.

A critical point is a contention point which is such that if you shifted the opinion on that point (you shifted the resolution of that contention point in your model) then the policy conclusion of the model is shifted significantly. "Significantly" is a relative term.

If you have a natural gas model and you find that by shifting say, from the assumption that the industry is monopolistic to the assumption that it's competitive, the choice between deregulation and regulation flips in desirability, then you have a critical point. You ask the same question of the model whether you have confidence in the model or not. If you do have a critical point and you have "adequate" (relative term) confidence in your model, then you can make a "conditional policy recommendation" which says, "If the natural gas industry IS competitive, then we ought to deregulate. If it IS NOT competitive then we ought to regulate." That is something the model can say that is useful in the policy debate. The model cannot really say whether the industry is competitive or monopolistic, because that is what the policy arena is trying to decide. This is an attempt to find out what can models say and be straightforward about. It can't say everything; it can't answer all the questions; but you can see what it can answer. If you have "adequate" confidence in the model and it has no critical points (i.e., the model is such that you can reflect any of the relevant opinions, and the model always comes out in favor of the same policy), then you ought to be able to make an "unconditional policy recommendation" based on the degree to which you have confidence in your model.

If you don't have "adequate" confidence in your model then the answers to these questions about the effect of point of view on a model still tell you something about your model. There is a lot of gas in the ground, but what if a natural gas model indicated that future gas production will continue its past behavior, even if the doomsayers are right and there is ZERO gas in the ground? (This actually is the case with the MacAvoy-Pindyck model.) Certainly one would conclude that the model was insufficiently sensitive to physical limitations of resources, especially in an era of great differences of opinions as to the extent of undiscovered resources.

Some states produce large quantities of natural gas for use within the sta The price of this gas in these states is not regulated, hence it is already as high as the future deregulated price. What if a natural gas model were to assum that the production of gas in these states is a direct function of the REGULATED price of interstate gas? (The MacAvoy-Pindyck model does this.) Whenever a dere lation policy is simulated enormous amounts of gas pour forth from these states, despite any change in the regulatory environment of its producers. Certainly one could conclude that the model was overly sensitive to the price of interstate natural gas. Observations such as these come back to the modeler, who will change the policy model. Then it must be analyzed again, for it's a different model.

QUESTION: You started assuming there that this third party was an objective, if you will, independent modeler that does all the various analyses. I wou say that probably in more frequent terms, each of the opposite points of view would have his own model and the third party would really be an arbiter between the models. Is this true?

CRISSEY: Sure. The Brandon Shores case examined in our book is a good example of that happening. We call it adversary modeling. In that case, those who had to decide whether to grant a license to the plant had to arbitrate between three models that were discrediting one another. Any policy model can be discredited, because they are simplifications and because they are mechanized opinions. There is always some test that you can come up with to embarass it. You can make any model look as bad as you want. Comparative analysis is, I think, the direction we have to go. Which model gets worse faster as it simulat is the question.

QUESTION: The right criteria is the next best alternative?

CRISSEY: Yes.

QUESTION: I think what a model analysis ought to be doing is saying what's driving that model. What is the key variable? Is it competition, is it non-competition or what, when it is all done? People have to think in simplistic terms. Was there a key driving variable in the MacAvoy-Pindyck model? Wasn't that competition versus oligopoly?

CRISSEY: There are several critical points in the model as I have mentioned, and competition, was one I considered, but the determination was only "probably critical" because I wasn't able, in the time that I had allotted, to restructure and reestimate the entire model to see what it would look like if I had assumed that the natural gas industry was oligopolistic. Unless one can represent a point of view in a model, he cannot know its impact. Competition is almost undoubtedly a critical point, but I didn't prove that it was.

The moral is that models should be designed to be able to reflect all major points of view that are relevant to an issue the model is trying to address.

QUESTION: It may seem like a quibble but it seems to me that you preface everthing with the word "policy." But I don't see how the word policy affects anything that you said in any special way, that is, this applies to any kind of model. Every model is a policy model insofar as it studies the effect of some decision variable on a response. What I don't know is if we have a value of the meaning of the word "policy," or is there something you're saying that I don't see?

CRISSEY: The processes I have described are applicable to any model of any reference system about which people have differing points of view.

QUESTION: I think what you've put out is good practice for anybody that's doing modeling.

1.

Asc
cei
In
sta
who
At
eng
be
the
fai
rel
of
fre
or s
pri
mat
or s

lea
to a
mode
in d
wil
EVEN
with
the

clea
Admi
eval
1985
have
et a
for
cost
of h
model

for s

STRATEGIES IN MODEL MANAGMENT

John Mulvey

1. MODEL EVALUATION

In his highly successful book and British Broadcasting series, "The Ascent of Man," J. Bronowski [4] describes the collapse of the 150' vaulted-ceiling cathedral at Beauvais, France in 1284 A.D. shortly after it was built. In contrast, the 125' ceiling at Rheims (less than 100 miles away) has remained standing over 700 years. These structures were built by guilds of freemasons who roamed across Europe, exercising judgments based on previous experiences. At that time, formal mathematical reasoning was generally not used; the engineering discipline was in its infancy and the building stresses could not be calculated. Thereby, a project was labelled a success solely by standing the test of time, for example, Rheims; and conversely, a project became a failure when the implementation failed, for example, Beauvais. There was no reliable way for predicting success or failure beforehand. Today, the builders of mathematical models assume a role similar to that of the Renaissance freemasons. In model building, there are no commonly accepted principles or standards to describe the process of developing a good model. Besides prior experiences, the scientific journals are available as sources of information; however, these journals usually provide only theoretical proposals or short descriptions of successful implementations.

As further evidence, the training of MS/OR specialists is geared to learning a set of non-overlapping skills. How many of us have been exposed to an academic course which considers the process of evaluating competing models? Given a single decision problem, two practitioners who are steeped in diverse techniques such as mathematical programming and simulation will invariably develop models which use their particular expertise - EVEN THOUGH THE REAL PROBLEM IS IDENTICAL. Nothing is inherently wrong with this bias, of course, provided that a methodology exists for evaluating the competing designs. This presentation is a first step in that direction.

The use of mathematical models for decision making in U. S. society is clearly increasing. On the Federal Government level, the Federal Energy Administration (now the Department of Energy) employs a linear program for evaluating the effects of energy policies on the U. S. economy in 1980, 1985 and 1990 (National Energy Outlook [19]). Manpower planning models have been studied by the U. S. Navy for setting promotional policies (Charnes et al. [6]). For many years, corporations have employed simulation models for developing planning strategies (Ackoff [1]). Decisions involving the cost of air pollution (Cohen and Hunter [8]) and for controlling inventories of human blood (Frankfurter et al. [11]) have been based on computerized models. The list is endless.

The models which will be discussed are mathematical programming models for scheduling personnel. I do not consider the ideas offered below to be restricted to these models; however, the discussion has been limited to a single class of models because of the well-defined objectives which mathematical programs display.

2. APPLICATION IN THE CONTEXT OF PERSONNEL SCHEDULING

To illustrate how a comparison effort should take place, I will briefly describe a real-life scheduling problem and then present three potential formulations for this problem. It involves the annual scheduling of faculty and courses with the Graduate School of Management (GSM) at UCLA.

In 1973, the Graduate School of Management revamped their MBA curriculum. This necessitated a centralization of the annual scheduling of faculty to courses and time periods (quarters). Scheduling had previously been conducted by each department in relative isolation since faculty and courses were uniquely assigned to individual departments. The integration of these subschedules was primarily carried out by Ida Fisher (an administrator) in conjunction with the department heads. However, the new MBA program had a considerable number of overlapping courses, and the idea of coordinated scheduling was central to this plan. The large size of the problem (100 faculty/500 courses courses/3 quarters) required that a computerized system be developed. The goal of this system was to assist Ida Fisher in scheduling the faculty.

Ida's decision problem is typical of manpower planning and scheduling -- balancing the needs for personnel with the resources available and the preference of the people assigned. Three related formulations for assisting Ida will now be presented.

A. The Network Formulation

The structure of the network model is shown in Figure 1. Each faculty member is provided with a faculty node and three related faculty/quarter nodes on the left-hand side of Figure 1. Each course is provided with a course node and up to three related course/quarter nodes on the right-hand side of Figure 1. The model determines the optimal matching of the left- and right-hand sides. Variables are defined as flows across the arcs; the flow on the arcs of the network is in course-quarter equivalents. The flow on these arcs is restricted by lower and upper bounds [the values of the numbers in parentheses (1, 2) in Figure 1 indicate a lower bound equal to 1 and an upper bound equal to 2]. Thus, in Figure 1, Buffa is assigned a total of 5 courses for the three quarters, since the arc connecting the source node to Buffa's node has a restriction (5, 5). Courses are similarly constrained.

Information concerning the needs and desires of the students can be used to determine the lower and upper bounds on the number of sections of each course offered per academic year, and by quarter. These restrictions appear as upper and lower bounds on the areas on the right-hand side of Figure 1. A forecasting model in conjunction with student questionnaires generates the menu of courses to be taught.

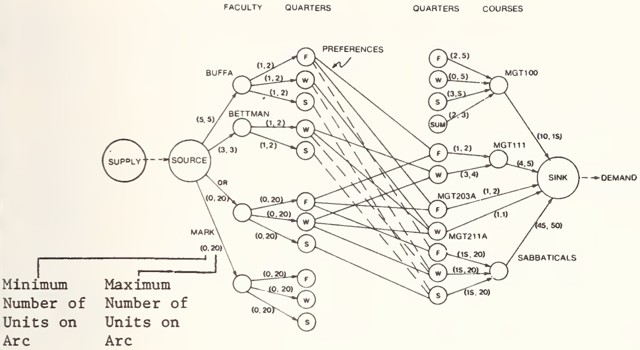
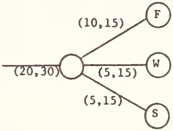
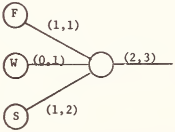


Figure 1. Network Representation



(a) Teaching Assistants



(b) MGT 200A

Figure 2. Individual faculty/course nodes and related arcs

Figure 2 portrays several examples of how lower and upper bounds of the flow on the arcs can be useful in achieving various objectives. As illustrated in Figure 2(a), the total number of course sections to be offered by teaching assistants during the year is restricted to between 20 and 30. However, any one quarter cannot have more than 15 course sections offered by teaching assistants because of the capacity restrictions of the other arcs.

Similar restrictions determine the number of offerings of the courses. For example, MGT 200A will be offered either two or three times during the academic year as shown in Figure 2(b). One section will be offered during the fall as indicated by the corresponding minimum and maximum flow restriction of one. At least one section will be offered during the spring quarter, and a third section may be offered during either the winter or spring. The determination of whether this third section will actually be offered, and during which of the two quarters, will be made by the model, based on the availability of faculty resources. Thus, the user is able to incorporate many options within the context of a simple network model

The objective function for this model is maximizing the preferences of the faculty for teaching certain courses and at the same time satisfying the arc restrictions. Faculty preferences for courses are determined through an annual faculty questionnaire. The preference weights range from minus 2 to plus 2, and are assigned by the faculty members. The administrators review these preferences and occasionally revise the weights to reflect teaching ability and student input. (For further details, see Dyer and Mulvey [10].) It should be noted that the network model is a special case of a linear program and that highly efficient strategies are available for solving this type of problem.

B. An Integer Program

The original model formulation (see Mulvey [18]) took the form of an integer linear program. The network constraints and the objective function just described were an essential part of this formulation. In addition to these network conditions, an expanded set of restrictions was incorporated into this model. Restrictions such as the following were allowed:

1. If course A is taught by Professor X then course B must be taught by Professor X.

2. Professor X could teach two sections of course A in the fall quarter or one section in the winter quarter.

3. Professor X wishes to teach one course from the set (A, B, C, D, E).

4. Professor X will teach one section of course B and only if Professor Y does not teach course B.

Faculty members were asked questions of this sort via a detailed questionnaire and their response formed the basis for modeling the extra constraints.

Although the resulting model was large for general integer programming, I developed a specialized enumeration procedure to capitalize on the structure of the problem.

C. An Auxiliary Model

The third formulation (Figure 3) incorporates considerably less detail than the previous two models. This model is an aggregation of the pure network model A. To derive this model first observe that the faculty members are not uniquely assigned to departments but can be clustered into areas of common interest. Instead of four unique nodes for each faculty member, these individuals are replaced by a faculty group or cluster node. For instance, one group is the "finance" faculty. Faculty members who are considered close to the finance group, i.e., those able to teach finance courses, are assigned to that group node. In an entirely analogous manner, the individual courses are assigned to course-group nodes. All arces in network model A linking faculty/quarters with course/quarters are preserved in the aggregate model. For instance, an arc from Professor Smith (Group A) to course MGT 101 (Group I) in the fall quarter would be assumed by the arc (A, I). An arc in the aggregate network will typically replace many arcs in the original network. The preference weight for the new arc is a simple weighted average of the arcs it replaced.

Following Geoffrion [13] I call this formulation an auxiliary model. It possesses the structural characteristics of the original network model A, but the size is greatly reduced in the number of arcs and nodes.

3. INGREDIENTS FOR COMPARISON

I now take up the issue of how to select one of the three candidate formulations, given our knowledge about the scheduling environment and the computer codes which are available for solving these problems. (As an aside, how would you go about choosing?)

Examine this issue with respect to five critical ingredients or dimensions for comparison as depected in Figure 4. Two of these dimensions (computational burden and user friendliness) deal with the computer software for solving the optimization problem; two dimensions (realism/complexity and information requirements) involve the underlying mathematical models; and one dimension (performance) involves both.

It should be obvious that the objectives implied by these dimensions are often conflicting. To conceive a totally realistic model, i.e., one which duplicates the original system to arbitrary precision, usually conflicts with the objective of building an affordable system; implicit or explicit tradeoffs in these objectives is inevitable. The goal of this section is formalizing this process by analyzing the five dimensions for models A, B and C.

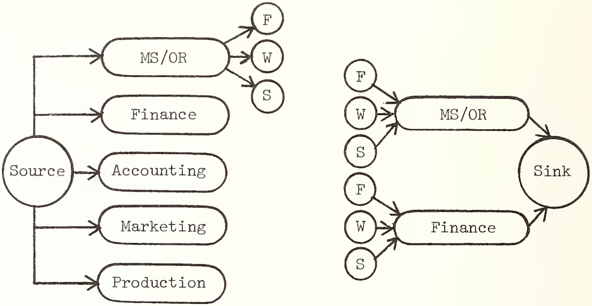


Figure 3 An Auxiliary Model for the Personnel Scheduling Example

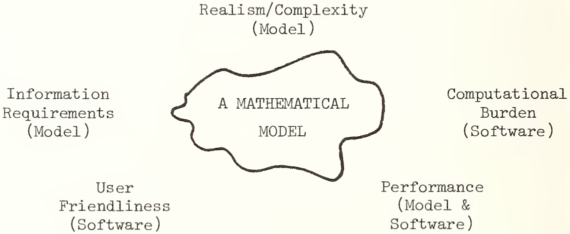


Figure 4 Five Dimensions for Evaluating Mathematical Models

A. Performance

By performance, I refer to the usefulness of the information which Ida receives by the model in her task of scheduling the faculty. Since these models deal with personal preferences, the data are soft and the schedule cannot be determined by a single solution from any model. The purpose of modeling this situation is gaining insights, and the amount of understanding which results from solving each model measures its performance.

For each model, it would appear that the goal is maximizing faculty happiness. However, it was assumed that the objective of maximizing faculty happiness and student satisfaction are complementary. Faculty members generally prefer teaching courses that are consistent with their professional abilities and teaching styles. Likewise, students generally prefer instructors who are enthusiastic about a course and its contents. While there may be some exceptional cases, it was not felt that these occurrences justify the burden of collecting additional information beyond simple expressions of faculty preferences. Also, as previously mentioned, information concerning the needs and desires of the students were used to determine the lower and upper bounds on the number of sections of each course offered per academic year, and by quarter. Implementation later showed that these assumptions were correct.

From Ida's perspective, the generation of a completed schedule and a method for altering the computer-generated results were crucial to its performance. Since she was not mathematically inclined, the equations defining the constraints were not very helpful to her. However, by studying the computer-generated assignments and the list of faculty members who were eligible to teach various courses, she quickly ascertained what had happened within the optimization routine. She quickly learned by trial and error.

The following ordering of the models can be thereby established: Model A (=P) Model B (>P) Model C, where (>P) means "possessing greater performance capabilities." Since Model A and Model B developed complete schedules, whereas Model C provided summary area coverage information, the performance of A and B proved to be superior to C from Ida's perspective. Unfortunately, as we will soon see, the performance ranking does not tell the entire story; the selection problem cannot be based solely on performance.

B. Realism/Complexity

I define the realism of a model to mean the relative closeness of the mathematical form to the situation which is being modeled. How well does the model mirror reality? In general, it has been my experience that the realism and complexity are synonymous -- the more realism, the more complexity is required.

Assuming that the information gathered from the student and faculty questionnaires is basically sound, it is a trivial matter to establish a rank ordering of the candidates. Model B (>R) Model A (>R) Model C, where (>R) means "more realistic than." The integer programming formulation B has the most general structure; it can accommodate any situation which can be handled by either of the other models. The integer program required about 5,000-6,000 variables and 1,500 constraints. Approximately 150 of these constraints were non-network.

Next in realism is the network formulation A; it has more capabilities than the auxiliary model which is an aggregate subset of it, but less realism than the integer program. The network consisted of 5,000-6,000 arcs and approximately 1,200 nodes.

Remarkably, because the auxiliary model is an aggregation of the network model, the amount of detail which is lost by using the auxiliary model instead of the unabridged network can be precisely measured. Hence, the smaller auxiliary model could be employed as a surrogate for the unabridged network, and the loss in accuracy measured. (For further details about the theory of aggregation, see the work of Geoffrion [12] and Zipkin [24].)

Unfortunately, in many instances a simple ranking such as this is not obvious. Elements of one model may be more realistic than elements of a competing model, and vice versa, and a serious complication is added to the decision of selecting alternatives. A mechanism for describing the extent of these differences is sorely needed and should be an important topic for future research.

C. Information Requirements

The amount of information which is collected and processed can impose a considerable burden on the user. In many applications, the sheer weight of this data may lead to the ultimate demise of an implementation. Thus, the information requirements must be considered when performing an evaluation of competing models.

Again, a simple rank ordering can be found for our scheduling problem. Model B (>I) Model A(>I) Model C, where (>I) means "requires more information." Model C requires the least amount of information and is the most desirable with respect to this characteristic; the bulk of the data can be gathered by asking area and curriculum representatives instead of individual faculty members. It should be noted that Models A and B require comparable information regarding faculty preferences. Each faculty member assigns preference weights (betwen minus 2 and plus 2) to their list of eligible courses. In addition, Model B requires data concerning "if-then" and other non-network restrictions.

D. User Friendliness

User friendliness is a term coined by Harlan Crowder to represent the inherent ease (or lack of ease) which is encountered when running a computer system. Many programs are difficult and awkward to use on a regular basis, and the criterion of user friendliness must be taken into account in the selection process. Otherwise, a perfect model may be developed, but the unavailability of correspondingly perfect software may prevent its use.

The introduction of software into the decision of selecting the best model complicated the problem. The development of a model is no longer time invariant; the "best" model today may be considered inferior when additional software becomes available. Also, a systematic way of measuring the freindliness of a system is difficult because of the very subjectivity of this concept. Nonetheless, this criterion is essential and cannot be avoided.

In 1973, the compter systems which were available for preforming the optimization were RIP30C (a general integer programming package (Geoffrion and Marsten [15]) and an advanced out-of-kilter method (superkilter) developed by Barr, Glover and Klingman [3]. Since the network system possessed a data base management facility, we considered it to be better than the integer programming system with regard to user friendliness. Thus, the models were ranked in 1974 with respect to this dimension as: Model A (=F) Model C (>F) Model B, where (>F) means "more friendly than."

E. Computational Costs

Another important consideration relating to the available software is the computational cost of solving the model. The cost of pre-processing data must be included in this criterion, as well as Ida's time spent in running the program.

Using the software mentioned in the previous section, the following relationships were evident: Model C (<C) Model A (<C) Model B, where (<C) means "costs less than." I estimated that the integer programming system would cost at least $250 for each feasible solution, hence it was too costly to locate the optimal solution. The paper by Mulvey [18] indicates an approach for decomposing the problem into a series of smaller problems, but even these subproblems cost almost $50 to solve.

In contrast, the network model could be solved much more cheaply. A typical Model A problem with 1,200 nodes and 6,000 arcs costs approximately $5.00 to solve, including the cost of preprocessing and postprocessing the data. The cost of finding a solution to Model C was even less -- about $.50 per run. For these reasons, Model B was deemed the most expensive, followed by Model A, and that followed by the cheapest Model C.

F. Selecting the Best Alternative

Given the comparative rankings along the five critical dimensions, the model builder must trade off these objectives in order to solve this multi-attribute problem. Figure 5 illustrates the elements of this decision. Each of the dimensions is labeled with an index corresponding to the previously described ordering. An A indicates that the model was the best of the three candidates, a B indicates a second-place score, and a C indicates third place. Depending upon the relative importance of the dimensions, any of the three models could be chosen as the most appropriate. For instance, if realism was crucial and far more important than the computational cost, Model B would be selected. On the other hand, if computational cost were deemed more important than realism, Model B or C would be selected.

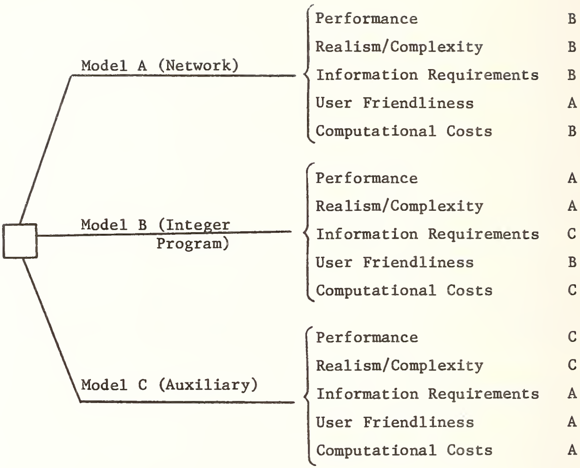


Figure 5. A Decision Tree for Selecting the Best Model

It is interesting to note that the network Model A seemed to be a good compromise between the greater detail and costs of the integer program and the reduced detail and cost of the auxiliary model, and the network Model A was eventually selected for use at UCLA.

This system has been in successful operation at UCLA for five years. The cheapness of the solution program and the flexibility of the support software gave Ida the flexibility of solving many partial scheduling problems. She used the computerized results in conjunction with her extensive understanding of the environment to schedule faculty by an iterative and interactive approach. She was able to accommodate the confounding non-network constraints (Section 2.B) by hand.

An inherent difficulty with including these non-network constraints into a model was the gaming of the system which occurred when these conditions were mathematically "forced" into the constraint set. Take the situation where Professor Jones knows that he is the only person able to teach MGT 200 and this course must be offered next year. Suppose in addition that Professor Jones does not want to teach MGT 401. He can rig the results by including a constraint of the form: if MGT 200 then not MGT 401, and he will be excluded from teaching MGT 401. Many subtle variations on this theme are possible when these types of constraints are allowed. An advantage of network formulation A is the requiring of manual intervention by Ida for each of these non-network constraints.

As a matter of record, the computational costs played a crucial role in the historical decision. Before the super-kilter program was available, a sample network was solved with the SHARE version of the out-of-kilter algorithm. The cost of this single run was $60.00. At that time, the auxiliary model was conceived as a possible alternative to the full-scale network. Fortunately, I was able to receive a version of super-kilter in time. This illustrates the dynamic nature of the selection problem; what is a good model today may become outdated tomorrow.

4. RECOMMENDATIONS

In conclusion, I suggest two different views of models and make recommendations relative to these views. On the right, modeling can be considered as a scientific process in which a set of objective principles guides the evaluation process. On the left, modeling can be considered within the domain of engineering in which heavy doses of judgment are tempered by professional standards.

A. Models as Science

An often-stated advantage of using mathematical models for decision making is the historical information which lingers after the model is used, i.e., its track record. By tallying the correct as well as the incorrect decisions, the models can, in theory, be ranked for accuracy, reliability and consistency. Unfortunately most decision systems are being constantly modified. The sheer ease with which basic assumptions can be altered in most large-scale systems is too tempting for the users. Thus, the evaluation process becomes confounded, and the empirical results may become misleading.

The computer can be used as an ideal experimental laboratory: conditions can be isolated and controlled; replication, the keystone of scientific activity, can be usually guaranteed by careful planning, the experimental design can be detailed, step-by-step; objectives are usually defined to .001 precision or better. Yet the scientific method is rarely linked with model comparisons since the above mentioned standards never seem to be fully accounted for. The decision makers must recognize these limitations and require model developers to justify their models using a sound scientific approach. As a first step, I recommend that whenever computer codes are used for implementing a model, every effort be made to distribute the code to interested parties. It is difficult, if not impossible, to render cross-model comparisons without having possession of all applicable codes. As a second step, the construction of a set of valid benchmark problems would facilitate these comparisons.

B. Models as Engineering

When models are currently evaluated, the usual statements about how they differ revert to an extensive "shopping list" of the assumptions which are required by each model. After reviewing a typical menu of normality, linearity, negative cross and own elasticities, and so on, most decision makers are left with a hollow feeling. A technical reply is, "So what does all of this mean?" As an answer, the typical expert is likewise left with little besides tiredly repeating the list of assumptions.

Instead of beginning an evaluation with such an enumeration, I believe that a mathematical model is a dynamic entity which must be evaluated by seeing its performance. An analogy is made with the American Ballet Theater. Surely, you would not begin a critical analysis of this troupe with the heights and weights of the dancers. Instead, you would watch the dance under a variety of operating conditions: romantic, classical and modern ballet styles, and you would observe other troupes perform identical suites so that reference criteria could be established. A model is not unlike a dance troupe in that it cannot be evaluated in isolation or without seeing it in action. Likewise, it often does not possess a scientifically precise single answer because it is a simplification of reality. A model should be subjectively evaluated and rated by "model critics," and their critiques should be made generally available. Yet the critics can be wrong; what is well-suited for one decision maker may be entirely unsuited for another.

The users of these models must begin to recognize these limitations and require model developers to justify their recommendations on a subjective or qualitative basis and in a manner which can be readily understood. This evaluation should be conducted in addition to the scientific evaluation previously mentioned. To begin the process, I recommend that university courses be developed under the heading "model appreciation," similar to the art appreciation courses taught in schools of fine art. Students should be exposed to a variety of approaches for solving a single problem. In this way, the modelers will develop an understanding of the pros and cons of alternative techniques and thus be in a better position to evaluate the ensuing recommendations.

As a second recommendation, I would like to see auxiliary (or prototype) models built prior to the construction of large-scale models. These simplified versions would keep the structure, but not the size, of the unabridged model. Perhaps a requirement for such a model could be included in the guidelines which NBS is considering.

To underscore the disparity between engineering and science and to return to the analogy of the medieval cathedral builder, I am reminded of my first course in engineering at the University of Illionis in which we were shown numerous films of engineering failures, such as bridges collapsing during violent storms and earthquakes. Even today, engineering design is subject to uncertainties, and judgment still plays an important role. Can mathematical models be considered any better?

QUESTION: There is a related problem and that's institutional expertise by methodology. If I had four different proposals to answer a certain problem, I can tell you ahead of time which model would be used. People almost always use the same model. That's a problem.

MULVEY: At first, operations research employed an interdisciplinary approach in which people of different expertise were brought together.

QUESTION: I think in the old days of defense systems analysis, at least as personified by the Rand Corporation, you did see that. You did see interdisciplinary teams working -- and we seem to have gone away from that for some reason.

QUESTION: But we still have interdisciplinary teams. Nonetheless, anyone interested in an interdisciplinary team will develop one model.

MULVEY: I'm not sure there is an answer to this problem. It has led to a lack of integrated research. Yet some of the efforts which are taking place in energy are very impressive.

QUESTION: When you say validate and verification of a model, what does that mean in addition to seeing whether the equations are correct?

MULVEY: I would execute the model with benchmark test problems which were developed prior to completion of the model. In other words I would conduct computational experiments of various kinds.

QUESTION: But you didn't test whether the model represents the real world. That's what we call parallelization and we do it.

MULVEY: Well, it's very difficult to verify this type of model.

QUESTION: I think that presents a basic problem. For more or less with any optimization model, you have in principle the same sort of empirical validation problem.

MULVEY: How do you judge the success of a model? It has been implemented for several years as a policy vehicle, so it met one criterion of success. The users are happy with the results of the model, that is to say they kept renewing the project. We were also happy since the model was being used.

Something else which often happens is that influential decisions are often made by programmers, either mathematical or computer, without consulting the ultimate decision maker. Hence, a great deal of policy is made by the programmers.

QUESTION: They are never recorded, that's the problem. They're implicit, but nobody tells you about them.

MULVEY: That's right. It's very difficult to properly consider all of the factors.

QUESTION: If they're recorded, it's not bad. But if they don't record it, you don't know about them.

QUESTION: What do you mean by computational experiments of different kinds, were they algorithmic or computational applications of a specific technique?

MULVEY: No, I would take two models and compare their results on a common set of inputs.

QUESTION: We have had early discussions of large models that were difficult to communicate in our discussions of the artichoke. I think one option which was mentioned was to have a small model documented and presented first, and then when people start asking questions, such as you left this, that and the next thing out, pull out your other one. You understand it, you understand what has been happening and its right because we have a larger model to defend it.

MULVEY: I'll end here.

QUESTION: Could you summarize in a sentence about the strategy of model management?

MULVEY: In my presentation, I suggested that five dimensions were important (i.e., performance, realism and complexity information requirements, user friendliness, and computational costs) for cross-model comparison. Over time competing models will alter their relative positions with respect to these dimensions because of new information, improved software capabilities, and changes in the underlying structure of the problem. The users of these models must be aware of the dynamics and develop suitable strategies for managing the available models.

QUESTION: Do you really believe that informational requirements and complexity tradeoff? Isn't it also true in modeling that complexity can be substituted for information? That you can indeed achieve a lot what you normally take as input or at least estimate it, and fuse parts of it within the model structure itself with quite heavy demand of the product for information.

MULVEY: I indicated a commonly observed, but not perfect, correlation between informational requirements and complexity.

QUESTION: You cannot say that large information means you have little complexity?

MULVEY: Sure, otherwise I would replace information requirements and complexity with a single dimension. There are five separate considerations.

REFERENCES

1. ACKOFF, R. L., A CONCEPT OF CORPORATE PLANNING, Wiley-Interscience, New York, 1970.

2. BALINSKY, W. and REISMAN, A., "Some Manpower Planning Models Based on Levels of Education Attainment," MANAGEMENT SCIENCE, 18, 12, August 1972, pp. B691-B705.

3. BARR, R. S., GLOVER, F. and KLINGMAN, D., "An Improved Version of the Out-of-Kilter Method and a Comparative Study of Computer Codes," MATHEMATICAL PROGRAMMING, 7 (1974), 60-86.

4. BRONOWSKI, J., THE ASCENT OF MAN, Little, Brown and Company, 1973.

5. BURACK, E. H., STRATEGIES FOR MANPOWER PLANNING AND PROGRAMMING, General Learning Press, Morristown, N.Y., 1972.

6. CHARNES, A., COOPER, W. W. and NIEHAUS, R. J., "Studies in Manpower Planning," U. S. Navy Office of Civilian Manpower Management, Washington, 1972.

7. CHARNES, A. and COOPER, W. W., "Goal Programming and Multiple Objective Optimization," EUROPEAN JOURNAL OF OPERATIONS RESEARCH, 1, 1977, 39-54.

8. COHEN, A. A., and HUNTER, Jr., A. P., "An Input-Output Analysis of the Costs of Air Pollution Control," MANAGEMENT SCIENCE, 21, 4, December 1974.

9. CROWDER, H. P., "Impact of Future Computer Technology on Mathematicl Programming," IBM Technical Report, Yorktown Heights, N.Y., 1977.

10. DYER, J. S. and MULVEY, J. M., "An Integrated Information/Optimization for Academic Planning," MANAGEMENT SCIENCE, 22, 12, August 1976.

11. FRANKFURTER, G. M., KENDALL, K. E., and PEGALS, C. C., "Management Control of Blood Through a Short-term Supply-Demand Forecast System," MANAGEMENT SCIENCE, 21, 4, December 1974.

12. GEOFFRION, A. M., "Customer Aggregation in Distribution Modeling," Working Paper No. 259, Western Mngt. Science Inst., UCLA (October 1976).

13. GEOFFRION, A. M., "The Purpose of Mathematical Programming Is Insight, Not Numbers," INTERFACES, 7, 1, 1976, 81-92.

14. GEOFFRION, A. M., DYER, J. S., and FEINBERG, A., "An Interactive Approach for Multi-Criterion Optimization, With an Application to the Operation of an Academic Department," MANAGEMENT SCIENCE, 19, 4, (1972), 357-368.

15. GEOFFRION, A. M. and MARSTEN, R. F., "Integer Programming Algorithms: A Framework and State-of-the-Art Survey," MANAGEMENT SCIENCE, 18 (1972), 465-491.

16. KEENEY, R. L. and RAIFFA, H., DECISIONS WITH MULTIPLE OBJECTIVES: PREFERENCE AND VALUE TRADEOFFS, John Wiley and Sons, 1976.

17. LEWIS, C. G., ed., Manpower Planning: A BIBLIOGRAPHY, English University Press, London, 1969.

18. MULVEY, J. M., "Preliminary Application of a Resource Allocation Procedure for the Educational Sector," Discussion Paper No. 21, O.R. Study Center, Graduate School of Management, UCLA (March 1972).

19. 1976 NATIONAL ENERGY OUTLOOK, Federal Energy Administration, U.S. Government Printing Officel, #041-018-00097-6.

20. NEMHAUSER, G. L. and NUTTLE, H. L. W., "A Quantitative Approach to Employment Planning," MANAGEMENT SCIENCE, 11, 8 (June 1965), B155-B165.

21. SCHROEDER, R. G., "Management Systems Design: A Critical Appraisal," in APPLYING ANALYTIC METHODS TO PLANNING AND MANAGEMENT, New Directions for Institutional Research, 13, Spring 1977.

22. Workshop on Utility and Use of Large-Scale Mathematical Models, sponsored by the U. S. Department of Commerce, National Bureau of Standards, Washington, D.C. 20234, April 28-29, 1977.

23. YOUNG, G. K. and PISANO, M. A., "Nonlinear Programming Applied to Regional Water Resource Planning," WATER RESOURCES RESEARCH, 6, 1970, 32-42.

24. ZIPKIN, P. H., "A Priori Bounds for Aggregated Linear Programs with Fixed-Weight Disaggregation," Technical Report No. 86, School of Organization and Management, Yale University (January 1977).

INTI

cons
and
fiel
ever
of s
are
curr
capa
also
not
or e
and
comp

of (
that

or (

I
and w
for m
ing t
impor

1
2
3

Br
like
in the

SOFTWARE REQUIREMENTS FOR AN IMPROVEMENT IN TRANSFER AND ADAPTABILITY OF MODELS

Siegfried Dickhoven

INTRODUCTION

Due to the extreme complexity of large-scale mathematical models, their construction use and implementation must involve computers. A closely linked and parallel development to the field of computerized modeling has been the field of software-instruments for modeling purposes. This development, however, has always time-lagged advances in modeling applications. The production of such modeling-instruments is rather expensive and therefore, these efforts are done generally after a certain period of urgent need. Compared to the current state-of-the-art in software techniques and actual operating system-capabilities, most of this modeling-software is somewhat antiquated. This is also due to the fact that many of these modeling software-instruments have not been developed by computer scientists, but by economists, social scientists or engineers, who are usually funded for their material research capabilities and not for the development of software-instruments. Nevertheless, these non-computer experts have developed high quality software-systems.

Current modeling trends ('2nd-generation modeling') are in the direction of (1) consolidation (making model-building more a science and less an art) that can be characterized by slogans like

- model comparison, review and evaluation
- user (decision maker) - participation (education)
- development of test - (implementation) methodologies

or (2) they lead to more experimental directions described by

- models of enormous dimensions of size
- the use of more sophisticated methods
- the application of optimization techniques
- the combining (linking) of different approaches ('eclectic approach').

To compare these trends with existing modeling software on the one hand, and with new software technologies on the other, leads to a set of requirements for modeling software. In my presentation however, I will focus on the following three areas which, in respect to our present activities, seem to be very important for an improvement of the transfer of modeling 'know-how':

1. modularization
2. software interfaces
3. wide range processors.

But, before dealing with these items and their impacts on modeling, I would like to give some background information on my institution, our specific role in the modeling scene, and our recent activities in this field.

GENERAL OBJECTIVES OF GMD-IPES

The 'Gesellschaft fuer Mathematik und Datenverarbeitung (GMD)' with headquarters at Schloss Birlinghoven near Bonn is a large-scale research institution (subordinated to the Bundesministerium fuer Forschung und Technologie) pursuing application-oriented basic research, applied research and development in the field of data processing. The GMD research and development activities cover the whole range of hardware, software, and applications, and their role in government and society. They also include advisory activities and contract work, in particular for the public sector. The GMD comprises eleven research institutes and five departments, with an overall staff of about 600 people, of whom about 250 are scientists.

The 'Institut fuer Planungs- und Entscheidungssysteme (IPES)' is one of these research institutes. It has the general objectives to improve the methodological and technical instruments for computer-aided political planning, and to analyze the impacts on the politico-administrative system caused by the increased use of data processing. It arose from a GMD working group for planning projects which had designed and implemented an information system for the integrated activity planning of the Federal Government that has been installed and used in the Chancellor's office since 1973. The Institute's research program comprises the following activity areas:

- Computer-based political planning (policy advice)
- Collection and review of socio-economic models and their support by software
- Development of methods for analysis and design of planning organizations in government
- Data processing as a communications medium for political planning
- Political impacts of increasing use of data processing.

OUR ROLE IN THE MODELING SCENE

Regarding this research program and our past and continuing projects in the field of socio-economic modeling, our specific role in this field includes the following.

We are model builders in that as we have developed some simulation models for several ministries in the Federal Government. Among these are a dynamic group-model (Markov-type) to simulate the effects of different personnel-policy scenarios on the mobility behavior of scientific personnel in large-scale research centers like ours: [1] a structure-model of the German labor-market, using the 'System Dynamics' methodology; [2] and a model for the prediction of medium- and long-term budgets of a Federal transfer law (Federal Student Aid Program) within the limits of the present law, as well as of projected amendments; [3] using the microanalytic modeling approach, which seem to us to be the best methodology for such purposes; [4] the origin of this last activity was a review of a similar model developed by a private research institution contracted by the Federal Government. This leads to our next position in the field of socio-economic modeling, the role of a model reviewer.

The just mentioned activity of revising a special model in educational planning included verification, rebuilding and validation of the model, as well as making it work on the ministry's computer facilities and the designing a more comfortable user-interface. The Institute for Planning and Decision Systems (IPES) also undertook a survey on the use of data processing in the political planning of the German Federal Government. This survey reported on the actual state, as well as discussing several future scenarios of DP-applications in political planning [5]. It showed that Federal agencies had many technical difficulties with models developed by others that used various modeling philosophies and different computer- and operating system-environments. According to our objective to improve the software support for modeling (producer of modeling tools), we undertook a second survey on existing modeling-instruments. In this survey, we reviewed the development of software in this field and compared the status of existing instruments with modern software techniques and capabilities used with other DP-applications [6, 7]. We analyzed about a dozen econometric systems, and roughly a half dozen systems or packages in each of the fields of system dynamics, microanalytic (individual) simulation, and method base systems. The study produced the following results:

- The scene in this software sector is very diverse. This is of course, a natural phenomenon that always occurs in any relatively new and rapidly developing area of data processing.

- Most of the software instruments make use of what are now somewhat antiquated techniques, compared to those which should currently be possible using modern operating systems.

- The development of software instruments in this area normally proceeds in two steps. At the beginning, most efforts are spent on increasing methodical support to satisfy the needs of the model builders, while interfaces to facilitate communication and understanding for nontechnical users are often neglected. But, after the first versions are in use, and more users who did not participate initially in the model development become involved, problems of handling it become more and more important for further development. This second step however is often never fully carried out, because the original software concept does not allow it, or the software developers are the only users of their system, or because of inadequate resources.

One of the few exceptions in this field of software systems is the TROLL-System for econometric applications, developed at MIT and which is now held by the National Bureau of Economic Research (NBER), Cambridge, Mass [8]. Besides having rather good methodical facilities in the field of econometrics, TROLL supports its users by a wide range of operating system functions (for example: data management, special edit-functions, macro- and default-facilities, monitoring). It uses rather modern computing techniques, although its line-concept for dialogues and its vast amount of different commands show that the system has reached the limits of its growth.

THE DEVELOPMENT OF A GENERAL 'MODEL BASE SYSTEM (MBS)'

Based on our analysis of the actual and future trends in modeling towards sophistication and consolidation, and facing their needs for software support with the stated capabilities of existing modeling software, we decided to develop a special 'operating system for socioeconomic models'. It is called 'Model Base System (MBS)'. The MBS shall contribute to a consolidation in this field and also provide facilities for experimental modeling [9].

MBS includes well tested and widely used construction-oriented systems (languages) such as DYNAMO, FORTRAN, MEBA (a German econometric system), and interfaces to some data base systems, as well as to data analysis packages. Thus, MBS can support the following groups of modeling activities:

- specification of formal model structure and of structural parameters [using DYNAMO, FORTRAN, etc. for basic (low level) structures and a dynamic linking language of MBS for meta structures]
- generation and loading of initial data
- call and processing of models of different types
- model-linkage
- adaption of external models
- data analysis and report generation
- mangement and documentation of data and models

In developing such an 'operating system' for simulation models, we hope to enable (as far as possible) non-computer experts to work with different models and model types, at least on the meta level, in a rather simple and almost uniform manner.

In regard to this development we think that the following three concepts:

- modularization
- software interfaces
- wide range processors

are rather important for improving the transfer and adaptability of models, because they are suppositions for making models a transferable product.

While modularization is understood here as a concept to characterize the transferable good (the models) and not a concept to develop the modeling tool, the other two concepts refer to output- (input-) characteristics of the modeling tools and to their performance requirements.

These topics will be discussed separately, although there are many tradeoffs between them. They are also discussed in relation to our current activity, the development of the Model Base System that is to be produced for two Federal ministries, the needs of which are to some extent quite different from those of pure scientific environments.

MODULARIZATION

With the increasing scale and number of socio-economic models, the application of modularization techniques becomes more and more apparent. This well approved technique for managing large and complex systems is, to some extent, also used in the field of modeling.

Well known applications like the Mesarovic/Pestel World Model [10], Project LINK [11], the Formula Bank Project [12], the MEBA-Specification-Stock [13] use the modularization concept in a substantial way. This is probably the most important way of controlling model complexity available to both model builders and model users. But, without a formal and software-oriented concept of modularization, all technical transfer problems cannot be resolved, unless the transfer takes place in the same software system, or at least in the same methodology.

With this concept, for example, it will be quite convenient to combine complementary or to compare similar model-parts (modules) that are written in the same language (e.g., DYNAMO) and that belong to the same methodology. But there will arise severe technical problems like respecification, translation and manual data transfer, if the modules to be combined or to be compared are of different types or in different languages (as they will be rather often in 2nd generation modeling).

To avoid these unnecessary technical problems it is often and erroneously assumed that the obvious solution is to create a new language or modeling system that combines the advantages of all (or some) different systems and methodologies and meets all their different needs. Although a lot of the modeling systems that we have analyzed started with this pretension, they have either led to a general purpose language like FORTRAN or PL 1 (and are completely useless) or they fail; and even software giants could not accomplish such a system for the modeling scene.

To overcome these technical difficulties with 2nd generation modeling-activities, we consider a better concept that integrates some already existing modeling languages of different methodologies and retains all capabilities and conventions of the approved modeling systems, inclusive of reporting and running specification, if needed by its users. Only new capabilities like linking and scenario generation, or very homogenous and to be extended facilities like reporting, documenting and analyzing, get a new and largely uniform kind of use. The realization of that concept in our MBS looks roughly like the following [14]:

The basic elements (elementary modules) of socio-economic models are so-called 'partial simulation operators'. They are generally time dependent and are a special form of a general operator that transforms a set of input quantities according to its transformation prescriptions into a set of output quantities for one and only one (time) step. This reduces the building process of models to the construction of the model-'core' or model structure. General tasks like data transfer, run- and time-loop-control are performed by a central simulation-processor.

These elementary modules are specified either in MBS or in modeling systems like DYNAMO III-F [15], the econometric systems IAS [16] (developed by the Institute for Advanced Studies, Vienna, and also used for project LINK at Bonn University) and MEBA [17] (used in Bundesministerium fuer Wirtschaft), a microanalytic system like MASS [18], and the general purpose language FORTRAN IV. FORTRAN-modules, however, have to be prepared before their adaption by the MBS-System according to some formal prescriptions. In this adaption-process there is also generated a user information block, called 'Kommunikationsteil' containing

- name and brief description of the module,
- list of control variables, and
- description of the module's data-interface for the meta-construction (e.g., linking) level.

The "Kommunikationsteil' of each module is produced under control of the MBS-user, who defines, besides other, the list and (new) names of variables for the meta-construction level.

On this meta-construction level, the modules can be linked together, or with special reporting or analyzing modules, or with data elements of a data base. Special model runs including conditions and systematic search may also be specified on this level. (This second level is very appropriate for 2nd generation modeling activities.)

This formal concept of modularization preserves the approved construction environment of experienced model builders' as well as providing capabilities for the computer-aided transfer of models (and models of different approaches). By providing a largely uniform manner of meta-construction and run specification (especially for 'production runs'), we hope that MBS will contribute to a better user-participation and more transfer of know-how between model builder-groups.

SOFTWARE-INTERFACES

To improve the transfer of models, the modularization concept leading to more formal uniformity of model builders' products is the obvious and direct way. Besides, this direct way yields another chance to promote model (-know how-) transfer, i.e., a better transfer of modeling-tools. Looking at the vast variety of different modeling software systems (and we probably know only the peak of an iceberg), the impression arises that there are many efforts wasted in a manifold reinventing of the wheel. This phenomenon is quite common in almost every new and rapidly growing area of data processing (see for example the rather young history of data base systems): but the time is ripening for a consolidation in the development of modeling tools, too.

The software interfaces concept, of course, cannot be the concept for consolidation in this area, but it is a possible first step towards consolidation. The linking of different modeling tools, especially of central tools (construction systems) with peripheral instruments like data base systems, report generators or analysis packages, within one operating system environment does not create severe problems for a computer scientist. For a model builder, however, it is really a great problem and therefore very seldom applied. By developing interface-programs the situation would become much better.

Two kinds of such interfaces are possible in relation to the kind of linking:

(1) Direct interfaces between modeling instruments: This kind of linking is more important for combinations of central (construction) systems with peripheral tools like report generators, because most central modeling systems are rather poor in these peripheral (post-processing) tasks. This direct interfacing has been done, for example, by the Urban Institute, Washington, D.C., who have linked their microanalytic modeling system MASH with the time series package TSP of the Brookings-Institution and with the report generator TPL of the U.S. Department of Labor to produce tables ready for printing [19]. This Table Producing Language again is based on an already existing data base system and linked with the statistical package SOUPAC from the University of Illinois [20].

(2) Interfaces into a general purpose high level (HL) language:

Although direct interfaces are also possible between central modeling tools, they are not recommended for these purposes. While interfaces between central and peripheral systems only have to provide (numerical) data-transfer between different instruments, an interface between central systems generally has to transfer programs. This makes it much more complicated and too expensive for only one connection. The detour into high level languages like FORTRAN has the advantage that the produced outputs can also be used by those people who work with this high level language as their modeling tool. This kind of interface can be devleoped either as an Input-Interface (able to adopt programs of that HL-language) or as an Output-Interface (producing HL-language programs) or as both. Applications of this interface type exist for example for the DYNAMO-F Language (type: Input-Interface) and for the Viennese econometric system IAS (here: Output-Interface) having both FORTRAN-interfaces.

We think that efforts in this direction will also increase the impetus of model builders towards standardizing in this field, because the use of linked modeling tools will lead to many more technial transfer problems, which may be overcome by the setting of interface standards.

WIDE-RANGE PROCESSORS

Wide-range processors are understood here as modeling instruments that support the set-up and processing of models of different methodological approaches, e.g., system dynamics- and econometric- or microanalytic- and econometric-models. The linking of one central with one or several peripheral modeling tools is not such an instrument, though it broadens the spectrum of working with models (but only with models of one approach). This rather new type of modeling software will become rather important for almost all 2nd generation activities.

Except for some microanalytic systems like MASH [19] and MOVE [21], which have been combined with econometric modeling nearly from their begining, there exist only rudimentary systems of this type like SIMA [23] or RSYST [23]. While SIMA is a system that has created an overall concept for econometric- and system dynamics-models, the RSYST-System comprises different, but newly developed subsystems for each approach. The Model Base System is similar to the latter, but it provides existing subsystems for the different approaches. It also has a two level concept that makes clear differences between elementary construction (1st generation activities) and meta-construction (2nd-generation activities). Such multi-level processors for modeling activities, including additional levels for the writing of methods by its users, will become the modeling tool of the future, as indicated by experimental systems like the ACOS SYSTEM [24] or the KARAMBA-Concept [25].

Second generation modeling often deals with models developed by others and wide-range processors are the specific tools for this kind of modeling activity. Such processors enforce the transfer of models, as well as accelerating the process of consolidation and experimentation in modeling. They will also contribute to standardization in modeling and to uniformity of software-development for modeling purposes. While the technical problems will enforce the users' wish for standardization, the high barrier of development costs will automatically lead to a concentration on development of such modeling tools.

CONCLUSION

The three software concepts mentioned here involve other, more special software requirements like information management, design or user interface, or programming and documentation standards. These will also contribute to an improvement of model transfer and adapatability. I will only mention the use of graphics for the recognition of the underlying model structure and for the presentation of results [26], and the application of structured programming that would make, for example, FORTRAN-modules easier to be read by non FORTRAN-programmers.

Since I am attending a workshop of the U. S. NBS, I should close with some remarks about standardization in modeling. Though this idea has more problematic layers, I will only refer to some technical (low level) aspects of modeling standards.

As already pointed out, these aspects i.e., primarily in the definition of standards for the formal interfaces of modules and of different modeling tools. For data interfaces, these standards will include prescriptions for type, size dimension and naming of data, including rules for documenting the data interface and/or the purpose and use of the underlying tool. The data interface of modules will probably be treated quite similarly. Subsequent standards will probably include uniform procedures for use of the different tools and for a more uniform formal structure of the different modules.

But the setting of standards, even at this low level, is a long and necessarily a cooperative process. To support this process, we have initiated a discussion circle by holding a workshop on modeling software [27], in which developers and users of models, as well as developers and users of modeling tools, are involved. This first meeting made the participants aware of the variety and future trends of modeling tools and of the needs for formal standards. Assisted by out pilot users, we are now defining some formal interface and documentation standards for our Model Base System that will be applied and tested for several different models of the Federal Government. The experiences of this application will be discussed at our next meeting and we hope that they will lead to some 'informal technical standards'.

REFERENCES

[1] S. Bruechkel; W. Schwarz: 'Fluktuation in Grossforschung-seinrichtungen -eine personalstatistische Analyse', Bonn 1975.

[2] J. Henize: 'Ein Arbeitsmarkt- und Inflationsmodell' in: J. Baetge: 'Systemtheorie und sozial-oekonomische Anwendungen', Berlin 1976.

[3] GMD-IPES: 'Computer-gestuetztes Planungssystem fuer das BAFoeG - Methodischer Ueberblick und Benutzeranleitung (BAFPLAN-Version 1)'. Internal Report IPES.77.202, Bonn 1977.

[4] G. H. Orcutt; S. Caldwell; R. Wertheimer: 'Policy Exploration through Microanalytic Simulation'. Washington, D.C. 1976. J. A. Pechmann; B. A. Okner: 'Who bears the tax burden?', Washington, D.C. 1974. H. J. Krupp: 'Moeglichkeiten der Verbesserung der Einkommens-und Vermoegenspolitik', Goettingen 1975.

[5] P. Hoschka; U. Kalbhen: 'Datenverarbeitung in der politischen Planung', Frankfurt 1975.

[6] S. Dickhoven; W. Kloesgen; U. Pankoke; W. Schwarz: 'Software fuer sozio-oekonomische Modelle', Report IPES.76.101, Bonn 1976.

[7] W. Kloesgen: 'Leistungsspektrum sozio-oekonomischer Modellsoftware', Report IPES.76.103, Bonn 1976.

[8] NBER: 'TROLL-Reference Manual (Standard System)' Cambridge, Mass. 1974-76.

[9] S. Dickhoven; W. Kloesgen: 'Grundlagen fuer die Entwicklung eines Modellbanksystems', Reprint IPES.75.102, Stuttgart 1976.

[10] M. Mesarovic; E. Pestel: 'Manking at the Turning Point', Readers Digest Press, New York 1974.

[11] R. J. Ball (ed.): 'The International Linkage of National Economic Models', Amsterdam 1973.

[12] Study Group for International Aanlysis Bergasse 16, A-1090 Vienna, Austria (Project started 1976).

[13] H. Enke; M. Kroeber-Weitz; D. Luedecke, W. Natzmer; K. U. Schweizer:

'Materielle Ausgestaltung des Modellrahmens - Teil 7: zusammenfassende Darstellung der brauchbaren Schaetzungen aus den einzelnen Teil-bereichen', MEBA-Report No. 13 Tuebingen 1976.

[14] GDM-IPES: 'Modellbanksystem (MBS) Leistungsbeschreibung', Internal Report IPES.77.203, Bonn 1977.

[15] A. L. Pugh: 'DYNAMO III - users manual', MIT, Cambridge, Mass., 1975.

[16] K. Plasser: 'IAS-Benuetzerhandbuch, IAS-Level 2, 7', Wien 1976.

[17] B. Schips: 'Das Leistungsspektrum der oekonometrischen Methodenbank (MEBA)' in: [27]

[18] A. Glazer; H. Jaramillo; P. Nelson; G. H. Orcutt: 'MASS, a Micro-Analytic Simulation System', working paper, New Haven, 1976.

[19] G. Sadowski: "MASH - A Computer System for Microanalytic Simulation for Policy Exploration', The Urban Institute, Washington, D.C. 1977.

[20] U.S. Dept. of Labor: 'The Development and Uses of Table Producing Language'. Washington, D.C. 1975.

[21] R. Brennecke: 'Das MOVE-System, Ein Prozessor fuer oekonometrische Anwendungen' in: 'Datascope' 5 (1974) 13, Frankfurt 1974.

[22] H. Maier: 'Computer-Simulation mit dem Dialogverfahren SIMA', Bassel 1976.

[23] R. Ruehle: 'RSYST I-III, experience and further development' in: 'Atomkernenergie' 26 (1975) Stuttgart 1976.

[24] NBER: 'ACOS - An Overview', Cambridge, Mass. 1976.

[25] R. Hueber; P. C. Lockemann; H. C. Mayr: 'Architektur von Methodenbanksystemen', Internal Report, Universitaet Karlsruhe 1976.

[26] W. Appelt; P. Krause; P. Wisskirchen: 'Graphische Datenverarbeitung beim Design von Modellsoftware' in: Proceedings des V. internationalen Kongresses: 'Datenverarbeitung im europaeischen Raum', Wien 1977.

[27] S. Dickhoven (ed.): 'Modellierungssoftware', Proceedings der GMD-Tagung: 'Status und Anforderungen auf dem Gebiet der Modellsoftware', Report IPES.76.102, Bonn 1976.

pa
dir
imp
put
top
R &
Edu
mad

1.

cha
val

sta

sui

proc
use
and

dete
best
conf

what

for

FUTURE DIRECTIONS

The final Workshop session drew on the preceding material and on the participants' insights in an attempt to identify the principal research directions and procedural improvements that, if pursued, would aid in improving the utility of large-scale models. Many ideas and concepts were put forth; there was not time enough to discuss each in much detail. The topics can be grouped under five somewhat overlapping areas: Technical R & D, Conceptual Explorations, Guidelines, Systems Management, and Education. We next use these headings to structure a list of the points made during the discussion:

1. TECHNICAL R & D

a. Sensitivity analysis (SA) procedures, including sensitivity to changes in the structure and complexity of models; SA as an aid in model validation; the interpretation of SA; and error estimates using SA.

b. Development of test-case generators to aid in producing model statistics.

c. How to improve the transferability of models.

d. How to improve the modeling process and modeling environment by suitable enhancements of languages and operating systems.

e. Aids for algorithm development and algorithm standardization.

f. Improving the people-model (computer) relationship by better procedures for output interpretation and presentation; development and use of user-computer interactive procedures for variations in data input and selection of solution.

g. Mathematical and computational considerations in modifying deterministic models to reflect the stochastic nature of problems; how best to generate and communicate probabilistic results; error analysis and confidence intervals.

h. How to improve model documentation; the value of documentation; what to document; dynamic (up-to-date) documentation procedures.

i. The development of model evaluation methodology.

j. The development of a taxonomy for model "types," purposes, and for relations between and among models.

2. CONCEPTUAL EXPLORATIONS

a. Criteria for and types and levels of model evaluation; depth and type of "appropriate" evaluation as a function of what elements; how to evaluate a model and the criteria for assessing a model's "credibility."

b. How to verify and validate (V and V) a model; what must modelers do for V and V; how do evaluators determine if V and V have been done; statistical and other tests for V and V; the running of extreme cases and special scenarios; the creation of "adversary" problems to be used in model acceptance-testing and in model evaluation procedures.

c. Improving the capability to derive understanding and insight from models; the behavioral aspects of modeling and modelers; the role of models in public debates.

d. Determining the appropriate scale of a model; complexity versus simplicity depending on model use; the relationship between the purpose and structure of a model and its computational requirements.

e. The differentiation of models "of" (research models) versus models "for" (application models) in terms of their respective requirements for documentation and evaluation.

f. Documentation requirements for a model as a function of model purpose, dissemination and training needs, model complexity, and one-time use versus continuing or diverse application (perhaps by users other than the original sponsor or developer).

g. Need for experiments to measure the effects and effectiveness of guidelines and the value of documentation; how to select or generate a suitable sample of projects for such experiments.

3. GUIDELINES

a. Content of model management guidelines and/or standards for model development. (See GAO guidelines in Appendix.)

b. Criteria for specifying and applying model management guidelines to a particular model based on the model's importance in the decision environment; need for flexibility in applying guidelines to a particular modeling situation; evolutionary nature of guidelines.

c. Guidelines for data source management; procedures for data updating and verification.

d. Use of phases and checkpoints in model-development management in ways that balance the concerns and needs of both sponsor and developer (contractor); ability to measure the accomplishment of a phase or passing a checkpoint; relationship to the RFP statement of work; Contract Officer Technical Representative (COTR) procedures for monitoring phases or checkpoints.

e. What should be the process of a model review; criteria to be used when reviewing RFP, proposals, progress and completion; should reviews be POST HOC or continuing, in-house or external; on what basis is a model reviewed; how to guard against biases; how to develop a review process acceptable to developers (contractors); the feasibility of developing a model contractor performance record; use of such a record, and criteria for contractor evaluation.

f. What is model documentation and what documentation should be a part of a modeling project; need for documentation guidelines that are sensitive to model purpose, use, and training needs; what is the process of documentation; how to determine if given documentation is adequate for a particular model; how to measure the cost of documentation; concept of model life-cycle cost.

g. How to accomplish the training needs for a model implementation; what materials are required.

h. Minimum set of guidelines, standards, and documentation to allow for portability; programming languages and their impact on portability.

i. Behavioral considerations in applying guidelines and in model development; biases of COTR, developer and user; biases of review panels.

4. SYSTEMS MANAGEMENT

a. Library of standardized routines and languages, of data on costs of models to be used for subsequent cost estimation, and of model applications.

b. Library of models for dissemination and maintenance/updating of programs, reports and data.

c. Need for an American Society for Testing Mathematical Models (ASTMM); issues of organization, cost and scope.

d. Development of a "modeling" newsletter describing applications and related material.

5. EDUCATION

a. How to introduce and explain stochastic concepts associated with model results to students, public and users; concepts of confidence levels and risk; removing the mystique of modeling; increasing the understandability (transparency) and understanding of models.

b. Education by instruction in methodology or through learning-by-doing or by case studies; on-the-job programs; university programs for students; review of available literature in this area; development of case studies.

c. Setting up intern programs; pre and post Ph.D. training; MBA program practicum.

- - - - - - - - -

N.B. It should be remembered that (by design) the workshop's attendees consisted mainly of professional DEVELOPERS of models, rather than users or evaluators. It is therefore not surprising that the topics listed above under "Technical R & D" elicited the greatest enthusiasm and consensus. Issues of "Systems Management" and of "Education," though of intellectual interest, were not of primary concern to the bulk of this audience. On the other hand, the matter of "Guidelines" was of substantial practical and professional concern. It was considered very important that Government-supported modeling efforts not find themselves under the "dead hand" of rules which failed to take into account the great diversity of size, purpose and innovativeness among modeling activities; which left model developers at undue risk of mid-term cancellation on the basis of arbitrary or vague criteria; and which might for the most part have their administration entrusted to persons without the professional experience or self-confidence to exercise fully such flexibility as was in principle permitted. It was also considered important that guidelines and their application should be based on solid logic and empirical knowledge rather than on unproven assumptions or "folk wisdom;" many of the items under "Conceptual Explorations" were proposed and supported as steps towards establishing such a rational basis for guidelines.

WORKSHOP IMPRESSIONS

Saul I. Gass

Due to time limitations, the Workshop was unable to address many of the issues related to improving the utility of mathematical models. In the final summary session, possible issue research directions were organized under five headings: Technical R&D, Conceptual Explorations, Guidelines, Systems Management and Education. That session's discussions are described in the preceding section.

Of the many issues that need to be clarified and resolved, two received the most discussion in terms of praise and abuse. Praise for model documentation and related standards (if supported by a proper level of funding and if done gradually), and abuse for model management guidelines (at least in terms of the GAO report; see Appendix).

The need for improved, more detailed model documentation appears to have no opposition, although there is some concern as to whether or not such documentation will increase the utility of the Government's modeling activity. There are no known studies that compare the benefits and costs of model documentation. Documentation standards need to be developed and tested. The evolution of such standards must involve both model users and developers.

Given a promulgated set of model documentation standards, their complete or partial adoption should be based on model size, value, complexity, resources, etc. The adherence to model documentation standards and the delivery of related model documents should be required by the contract RFP or grant statement. Precedent for this type of action exists. Many Government agencies impose computer programming documentation specifications on their contractors, although it is our impression that such specifications have not been standardized. A possible paradigm for cross-agency standardization is the NBS FIPS 38: GUIDELINES FOR DOCUMENTATION OF COMPUTER PROGRAMS AND AUTOMATED DATA SYSTEMS. A basic, initial item of research is the determination of the scope and content of model documentation standards that would then lead to the development of a comparable FIPS 38 guideline for mathematical models.

Workshop participants offered no opposition to the general concept that complex mathematical models should be documented to enable others to understand and use the model. The Workshop did not delve into particulars. However, concerns were expressed as follows: (1) attempts to require "full blown" documentation for all modeling efforts; (2) documentation resources would have to be increased, possibly causing a decrease in funds and personnel available for model developmental and implementation; and (3) for urgent modeling activities carried out in haste to aid in resolving immediate policy or strategic problems, documentation of the model and/or its enhancements will always be relegated to the future -- but the future usually requires the solving of additional immediate problems or no further use of the model, thus documentation that is useful for outsiders will never get completed or possibly never even initiated.

This last item is a major concern. Complex decision-aiding models that mushroom in a "fire-drill" mode, as well as those that evolve in a more leisurely fashion, are used by agencies to justify particular programs or decisions. The "opposition" (OMB, Congress), without proper model documentation, cannot understand the rationale of the decisions and must counter the agencies while lacking full information. An approach must be worked out (even prior to the final development of model documentation guidelines) for these modeling efforts to be supported at a level that allows for documentation to be developed during and beyond the model development phase. At the agency level, the writing of model documentation does not necessarily offer it any immediate benefits. In fact, the availability of documentation might even work against its strategy in developing a decision model, one that might be biased towards a particular resolution. And complete documentation would increase the cost to the agency.

Prior to the Workshop, the GAO report, WAYS TO IMPROVE MANAGEMENT OF FEDERALLY FUNDED COMPUTERIZED MODELS, was distributed to the participants. The report is an attempt by the GAO to describe an approach to computerized model development that would improve the management of modeling projects and make such models more responsive to user needs. The report represents a formalization of good modeling practices, but the specific approach described had not previously been debated by the modeling community (developers and users). There is a serious question as to whether the GAO approach could be implemented in the real world of contracts and grants, as applied to the development of complex models.

As the GAO report has had limited external distribution, and as model management procedures is a key item in improving model utility, one purpose of the Workshop was to discuss such procedures using the GAO approach as a reference point. (We note that the majority of the Workshop speakers were model developers, with the total set of participants split between model developers and Government users.)

Although the GAO report was not reviewed point by point, the developer participants were rather vigorously opposed to the GAO five-phased approach that calls for specific intermediate review steps that could require the cancellation of the project. The basic developer concerns are the ability to sustain a viable project under the threat of cancellation (both in a business and technical sense) and the restraints imposed by managment procedures on technical innovation and advancing the state-of-the-art. It is clear that any Government attempt to formulate model management procedures must take into account the needs, interests and concerns of the model developer community.

The use of complex decision making models by the Government has caused the interest in procedures for model evaluation to increase. There is a need for Congress and others to obtain third-party independent evaluations of the executive branch's model-based programs and decisions. There is no set approach to the model evaluation process. However, the workshop participants did receive material reviewing the need for model evaluation and a suggested methodology for use as a basis for discussion in the Workshop. Time did not allow much discussion of this material, but one session did review the need for and approaches to policy model evaluation.

Based on comments during the Workshop, we have the impression that the need for third-party evaluation by developers is not appreciated. The community of model developers includes members exhibiting the full range of abilities: excellent to poor. The developers who participated in the Workshop have consistently produced superior complex models for their clients. They fail to recognize that most models are developed by those with lesser skills and experience. However, no matter what talent produces a model, the resultant model must be able to undergo a close scrutiny in terms of verification and validation. The Government must be able to evaluate a model so as to make some statement as to whether the model can be used with confidence in a decision environment.

A similar comment applies to the need for a model managment procedure. A small class of superior model developers do not need such procedures and would probably be constrained by their imposition. This class of modelers includes, in general, the innovators and frontier advancers. But again, most models are produced by a less talented class. Hence, the Government needs to establish some mechanism for improving the management of its modeling activities.

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

12.

13.

14.

BIBLIOGRAPHY

MODEL EVALUATION AND ASSESSMENT

1. "Advantages and Limitations of Comptuer Simulation in Decision Making," B-163074, U.S. GAO, Washington, D.C., May 3, 1973.

2. "Computer Simulations, War Gaming, and Contract Studies," B-163074, U.S. GAO, Washington, D.C., February 23, 1971.

3. "Improvement Needed in Documenting Computer Systems," B-115369, U.S. GAO, Washington, D.C., October 8, 1974.

4. "Models, Simulations, and Games - A Survey," M. Shubik and G. D. Brewer, R-1060-ARPA/RC, Rand Corporation, Santa Monica, CA, May 1972.

5. "Federally Supported Mathematical Models: Survey and Analysis," G. Fromm, W. L. Hamilton and D. E. Hamilton, Stock No. 038-000-00221-0, U.S. GPO, Washington, D. C. 20402.

6. "Project Independence Report," Federal Energy Agency, Washington, D.C., November 1974.

7. "Review of Selected Army Models," Department of the Army, May 1971.

8. "Computer Simulation Methods to Aid National Growth Policy," Committee on Merchant Marine and Fisheries, 94th Congress, U. S. GPO, Washington, D.C., 1975.

9. "Criteria for Assessing Model Validity for Managerial Purposes," R. E. Schellenberger, DECISION SCIENCES, Vol. 5, No. 5, 1974.

10. "Improvement Needed in Managing Automated Decision Making by Computers Throughout the Federal Government," FGMSD-76-5, U.S. GAO, Washington, D.C., April 23, 1976.

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

12. "Auditing a Computer Model: A Case Study," Division of Financial and General Management Studies, U.S. GAO, Washington, D.C., May 1973.

13. "Ways to Improve Managment of Federally Funded Computerized Models," LCD-75-111, U.S. GAO, Washington, D.C., August 23, 1976.

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

15. "Guidelines for the Practice of Operations Research," The Operations Research Society of America, OPERATIONS RESEARCH, Vol. 19, No. 5, September 1971.

16. "Evaluation of Complex Models," Saul I. Gass, COMPUTERS AND OPERATIONS RESEARCH, Vol. 4, No. 1, pp. 27-35, March 1977.

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

18. EVALUATION RESEARCH, C. H. Weiss, Prentice-Hall, Inc., Englewood, Cliffs, NJ, 1972.

19. HANDBOOK OF EVALUATION RESEARCH, Vols. I and II, Edited by E. L. Struening and M. Guttentag, Sage Publications, Beverly Hills, CA, 1975.

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

21. "Mathematical Modelling for Management," S. Eilon, INTERFACES, Vol. 4, No. 2, February 1974.

22. DESIGN AND USE OF COMPUTER SIMULATION MODELS, J. R. Emshoff and R. L. Sisson, The Macmillan Co., NY, 1970.

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

24. A GUIDE TO MODELS IN GOVERNMENTAL PLANNING AND OPERATIONS, S. I. Gass and R. L. Sisson, Sauger Books, Potomac, MD, 1975.

25. SYSTEMS SIMULATION, G. Gordon, McGraw-Hill Co., New York, NY, 1969.

26. "Validation Problems in Games and Simulations," C. Hermann, BEHAVIORAL SCIENCE, Vol. 12, pp. 216-230, May 1967.

27. "Diogenes Revisited -- The Search for a Valid Model," P. W. House, SIMULATION, October 1974.

28. "Models and Managers: The Concept of a Decision Calculus," J. D. C. Little, MANAGEMENT SCIENCE, Vol. 16, No. 8, April 1970.

29. "Practical Aspects of Simulation Models," G. A. Mihram, OPERATIONAL RESEARCH QUARTERLY, Vol. 23, No. 1, March 1972.

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

31. "Validation," R. Van Horn, Chapter in THE DECISION OF COMPUTER EXPERIMENTS, T. H. Naylor, Ed., Duke University Press, Durham, NC, 1969.

32. "Validation of Simulation Results," R. L. Van Horn, MANAGEMENT SCIENCE, Vol. 17, No. 5, Janury 1971.

33. COMPUTER SIMULATION TECHNIQUES, T. H. Naylor et al., John Wiley and Sons, New York, NY, 1966.

34. SYSTEMS SIMULATION, R. E. Shannon, Prentice-Hall, Englewood Cliffs, NJ, 1975.

35. "A Policy Model Appraisal Paradigm," O. P. Hall, Jr., POLICY SCIENCE, 6, 1975, pp. 185-195.

36. MODELS IN THE POLICY PROCESS, M. Greenberger, M. A. Crenson and B. L. Crissey, Russel Sage Foundation, NY, 1976.

37. HOMILIES FOR HUMBLE STANDARDS, D. T. Ross, Communications of the ACM, Vol. 19, No. 11, November 1976.

38. "Survey of the State-of-the-Art: Social, Political, and Economic Models and Simulations," C. C. Abt, et al., Abt Associates, Cambridge, MA, 1965.

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

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

41. POLITICIANS, BUREAUCRATS AND THE CONSULTANT -- A CRITIQUE OF URBAN PROBLEM SOLVING, G. Brewer, Basic Books, 1973.

42. SYSTEMS ANALYSIS IN PUBLIC POLICY, I. Hoos, University of California Press, Berkeley, CA, 1972.

43. EVALUATION OF AIR TRAFFIC CONTROL MODELS AND SIMULATIONS, DOT-TSC-FAA 71-7, NTIS No. AD-733 755, DOT, Systems Analysis Division, Transportation Research Center, Cambridge, MA, June 1971.

44. Set of 39 NSF Evaluations of Policy-Related Research. See for example: REVIEWS AND CRITICAL DISCUSSIONS OF POLICY-RELATED RESEARCH IN THE FIELD OF POLICE PROTECTION, S. I. Gass and J. M. Dawson, Mathematica, Inc., October 1974. (This also contains a listing of the 39 studies.) Available through NTIS, Springfield, VA.

45. LARGE-SCALE MODELS FOR POLICY EVALUATION, P. W. House and J. McLeod, John Wiley and Sons, New York, 1977.

ut
Go
di
mod
li
(3
des
and
res

siv
int
of

PROGRAM

WORKSHOP ON THE UTILITY AND USE OF LARGE-SCALE MATHEMATICAL MODELS

This Workshop was organized to examine the problem of improving the utility and use of large-scale mathematical decision models in the Federal Government. Recent Government sponsored surveys and reports have indicated dissatisfaction among model users and developers with many aspects of the modeling process. Principal areas of concern include the lack of: (1) guidelines for model development and management, (2) documentation standards, and (3) model evaluation procedures. The program of the Workshop, Figure 1, was designed with these concerns in mind, as well as the broader issues of use and utility of decision models and the confidence to be placed in their results.

The Workshop's speakers and participants were selected for their extensive experience in the development and use of mathematical models, and their interest in furthering professionalism in analysis and modeling. The names of the attendees are listed in Figure 2.

WORKSHOP ON THE UTILITY AND USE OF LARGE-SCALE
MATHEMATICAL MODELS

Thursday - April 28, 1977

Welcome	Dr. Alan J. Goldman 9:00 - 9:10
The Workshop Issues	Dr. Saul I. Gass 9:10 - 9:30
Review of the DOD Modeling Effort and Modeling as a Profession	Dr. Garry Brewer 9:30 - 10:15
(Coffee)	10:15 - 10:30
Review of the non-DOD Modeling Effort	Dr. Gary Fromm 10:30 - 11:00
Issues Facing Model Developers: Presentation and Panel	Dr. Seth Bonder Dr. Dennis Meadows Dr. Dan Maxim Mr. Alexander Pugh III 11:00 - 12:30
(Lunch)	12:30 - 1:30
Model Implementation	Dr. Richard C. Larson 1:30 - 2:00
Transfer of Models to Agencies of Local Government	Dr. Jan M. Chaiken 2:00 - 2:30
The PTI Experience	Dr. Jack Barrett 2:30 - 3:00
The FEA Project Independence Model Experience	Dr. Harvey Greenberg 3:15 - 3:45
The EPRI/NBER Energy Model Assessment Project	Dr. David T. Kresge 3:45 - 4:15

FIGURE 1

The Energy Modeling Forum ....Dr. William Hogan
4:15 - 4:45

Summary ....Dr. Saul I. Gass
4:45 - 5:00

Friday - April 29, 1977

Models in the Policy Process:
A Framework ....Dr. Brian Crissey
9:00 - 9:30

Strategies in Model Management ....Dr. John Mulvey
9:30 - 10:00

Software Requirements for an
Improvement in Transfer and
Adaptability of Models ....Dr. Siegfried Dickhoven
10:00 - 10:30

(Coffee) ....10:30 - 10:45

Guidelines, Standards, and Management
Improvements for Modeling Activities:
Discussion and Summation ....Dr. Saul I. Gass
10:45 - 1:00

FIGURE 1 (Continued)

WORKSHOP PARTICIPANTS

Dr. Jack Barrett
Public Technology, Inc.
1140 Connecticut Ave., N.W.
Washington, D.C. 20036

Dr. Seth Bonder
Vector Research, Inc.
P. O. Box 1506
Ann Arbor, MI 48106

Dr. Garry Brewer
School of Organization & Mgmt.
Yale University
1891 Yale Station
New Haven, CT 06520

Dr. Jan Chaiken
Systems Science Dept.
Boelter Hall
UCLA
Los Angeles, CA 90024

Mr. Wallace Cohen
Office of Program Analysis
US GAO
441 G. St., N.W.
Washington, D.C. 20548

Dr. Brian Crissey
National Academy of Sciences
Watergate Building
2600 Virginia Ave., N.W.
Washington, D.C. 20037

Dr. Siegfried Dickhoven
IPES, Postfach 1240
Schloss Birlinghoven
D-5205 St. Augustin 1
Bonn Germany

Dr. Gary Fromm
Stanford Research Institute
1611 N. Kent St.
Arlington Va.

Dr. Saul I. Gass
University of Maryland
National Bureau of Standards
Washington, D.C. 20234

Mrs. Judith F. Gilsinn
Applied Mathematics Division
National Bureau of Standards
Washington, D.C. 20234

Dr. Neal Glassman
Operations Research Program
(Code 434)
800 N. Quincy St.
Office of Naval Research
Arlington, VA 22217

Dr. Alan J. Goldman
Applied Mathematics Division
National Bureau of Standards
Washington, D.C. 20234

Dr. Harvey Greenberg
Federal Energy Administration
12th & Penn. Ave., N.W.
Washington, D.C. 20461

Mr. William G. Hall
Applied Mathematics Division
National Bureau of Standards
Washington, D.C. 20234

Mr. Arthur Hawnn
Dept. of the Army
Corps of Engineers
Washington, D.C. 20314

Dr. Karla Hoffman
Applied Mathematics Division
National Bureau of Standards
Washington, D.C. 20234

FIGURE 2

Dr. William Hogan
Dept. of Economic Eng. Systems
Stanford University
Stanford, CA 94305

Dr. John Honig
7701 Glenmore Spring Way
Bethesda, MD 20034

Mr. R. H. F. Jackson
Applied Mathematics Division
National Bureau of Standards
Boulder, CO 80302

Mr. Barry J. Kennedy
Dept. of the Army
Corps of Engineers
Washington, DC 20314

Mr. Philip Kiviat
FEDSIM/CA
Dept. of the Air Force
Washington, DC 20330

Dr. David Kresge
National Bur. of Econ. Res.
575 Technology Square
Cambridge, MA 02138

Dr. Richard Larson
O.R. Center (24-215), MIT
Cambridge, MA 02139

Mr. Robert Lilly
Route 1, Box 344
Frederick, MD 21701

Dr. Dan Maxim
Mathematica, Inc.
P. O. Box 2392
Princeton, NJ 08140

Mr. James McLynn
DTM
4905 Del Ray Ae.
Bethesda, MD 20014

Dr. Dennis Meadows
Systems Dynamics Group
Dartmouth College
Hanover, NH 03755

Dr. John Mulvey
Harvard Business School
Boston, MA 02163

Mr. Alexander Pugh, III
Sloan School of Management
MIT
Cambridge, MA 02139

Mr. Paul Roth
ICST
National Bureau of Standards
Washington, DC 20234

Mr. Roy Saltman
ICST
National Bureau of Standards
Washington, DC 20234

Mrs. Patsy Saunders
Applied Mathematics Division
National Bureau of Standards
Washington, DC 20234

Mr. Bruce Thompson
US GAO
441 G St., N.W.
Washington, DC 20548

Mr. Jim Wiggins
USGAO
441 G St., N.W.
Washington, DC 20548

Dr. Christoph Witzgall
Applied Mathematics Division
National Bureau of Standards
Washington, DC 20234

Mr. Peter Wood
Bureau of Labor Statistics
Washington, DC

FIGURE 2 (Continued)

NBS-

BI

4. TI

7. AU

9. PE

NA
DE
WA

12. SP

15. SU

16. ABS
liter

T
t
w
a
W
p
d
e
m
g
v
a

17. KEY
sepa

la
st

18. AVAIL

Fo

Or
20

Or
VA

NBS-114A (REV. 9-78)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS SP 534	2. Gov't. Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

Utility and Use of Large-Scale Mathematical Models
Proceedings of a Workshop held at the
National Bureau of Standards
Gaithersburg, Maryland April 28-29, 1977

5. Publication Date

May 1979

6. Performing Organization Code

7. AUTHOR(S)

Edited by Saul I. Gass

8. Performing Organ. Report No.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, DC 20234

10. Project/Task/Work Unit No.

11. Contract/Grant No.

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

Operations Research Division
Center for Applied Mathematics
National Engineering Laboratory
National Bureau of Standards, Washington, D.C. 20234

13. Type of Report & Period Covered

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

Library of Congress Catalog Card Number: 79-600045

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

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

The Workshop on the Utility and Use of Large-Scale Mathematical Models, held at the National Bureau of Standards, Gaithersburg, Maryland (April 28-29, 1977), was a "first" in its purpose: to examine the problem of how to improve the use and utility of large-scale mathematical models in the Federal Government. The Workshop speakers addressed specific problems areas of concern, including: the present status of model use in DOD and non-DOD applications; issues facing developers; problems of model implementation; transfer and development in the energy field; model assessment and evaluation; use in analysis; comparison of models; management of the modeling process; model software and documentation; guidelines, standards and management improvement activities. This proceedings volume presents the papers and much of the discussion presented at the Workshop, along with a summary of directions for needed research.

17. KEY WORDS *(six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons)* Documentation; energy; evaluation; guidelines; implemenattion; large-scale; management; mathematical models; policy analysis; software; standards; transfer.

18. AVAILABILITY [XX] Unlimited	19. SECURITY CLASS (THIS REPORT)	21. NO. OF PRINTED PAGES
☐ For Official Distribution. Do Not Release to NTIS	UNCLASSIFIED	217
[X] Order From Sup. of Doc., U.S. Government Printing Office, Washington, DC 20402, SD Stock No. SN003-003-	20. SECURITY CLASS (THIS PAGE)	22. Price
☐ Order From National Technical Information Service (NTIS), Springfield, VA. 22161	UNCLASSIFIED	

USCOMM-DC

Stock Number 003-003-02060-5 – Price $4.25

JOUR
of the
and c
engine
includ
compu
with r
the ba
cluded
related
a spec
citation
NBS r
domest
$3.75 f

Note: T
Section
matical

DIMEN
This m
enginee
consum
with pr
highligh
protectio
abateme
formanc
grams i
of matte
strume

Annu

Monogra
ture on
d techn

Handboo
al prac
n with
d regul

cial P
nsored
blicatio
ket ca

pplied
, and
mists, I
. others

tional S
data
rials,
ated.
nated
dard D

following
phies a
ganic I
survi
$25.0
ded N
l subs

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH—The Journal of Research of the National Bureau of Standards reports NBS research and development in those disciplines of the physical and engineering sciences in which the Bureau is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology, and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent NBS publications in NBS and non-NBS media. Issued six times a year. Annual subscription: domestic $17.00; foreign $21.25. Single copy, $3.00 domestic; $3.75 foreign.

Note: The Journal was formerly published in two sections: Section A "Physics and Chemistry" and Section B "Mathematical Sciences."

DIMENSIONS/NBS

This monthly magazine is published to inform scientists, engineers, businessmen, industry, teachers, students, and consumers of the latest advances in science and technology, with primary emphasis on the work at NBS. The magazine highlights and reviews such issues as energy research, fire protection, building technology, metric conversion, pollution abatement, health and safety, and consumer product performance. In addition, it reports the results of Bureau programs in measurement standards and techniques, properties of matter and materials, engineering standards and services, instrumentation, and automatic data processing.

Annual subscription: Domestic, $11.00; Foreign $13.75

NONPERIODICALS

Monographs—Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NBS, NBS annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a world-wide program coordinated by NBS. Program under authority of National Standard Data Act (Public Law 90-396).

NOTE: At present the principal publication outlet for these data is the Journal of Physical and Chemical Reference Data (JPCRD) published quarterly for NBS by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements available from ACS, 1155 Sixteenth St. N.W., Wash., D.C. 20056.

Building Science Series—Disseminates technical information developed at the Bureau on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NBS under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The purpose of the standards is to establish nationally recognized requirements for products, and to provide all concerned interests with a basis for common understanding of the characteristics of the products. NBS administers this program as a supplement to the activities of the private sector standardizing organizations.

Consumer Information Series—Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order **above** *NBS publications from: Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.*

Order **following** *NBS publications—NBSIR's and FIPS from the National Technical Information Services, Springfield, Va. 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. Register serves as the official source of information in the Federal Government regarding standards issued by NBS pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NBS Interagency Reports (NBSIR)—A special series of interim or final reports on work performed by NBS for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Services (Springfield, Va. 22161) in paper copy or microfiche form.

BIBLIOGRAPHIC SUBSCRIPTION SERVICES

The following current-awareness and literature-survey bibliographies are issued periodically by the Bureau:

Cryogenic Data Center Current Awareness Service. A literature survey issued biweekly. Annual subscription: Domestic, $25.00; Foreign, $30.00.

Liquified Natural Gas. A literature survey issued quarterly. Annual subscription: $20.00.

Superconducting Devices and Materials. A literature survey issued quarterly. Annual subscription: $30.00. Send subscription orders and remittances for the preceding bibliographic services to National Bureau of Standards, Cryogenic Data Center (275.02) Boulder, Colorado 80302.

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

OFFICIAL BUSINESS

Penalty for Private Use, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE
COM-215

U.S.MAIL

SPECIAL FOURTH-CLASS RATE
BOOK