PROCEEDINGS OF THE

Fourth Annual Conference on Urban Planning Information Systems and Programs

AUGUST 19-21, 1966
BERKELEY, CALIFORNIA

Department of City and Regional Planning and the
Center for Planning and Development Research,
Institute of Urban and Regional Development

UNIVERSITY OF CALIFORNIA, BERKELEY

PREFACE

The Fourth Annual Conference on Urban Planning Information Systems and Programs was held under the joint sponsorship of the Department of City Planning and the Center for Planning and Development Research at the University of California, Berkeley, on August 19-21, 1966.

These Proceedings were produced from master copies which were prepared and edited by the individual participants. Thanks are due to the authors for extending their cooperation toward this effort.

A highlight of the Conference was the establishment of a permanent organization dedicated to the advancement of an interdisciplinary and multiprofessional approach to the planning, designing, and operating of urban information systems. The organization adopted the name: Urban and Regional Information Systems Association, and the following officers were elected to serve as officers of this association:

President	- Edgar M. Horwood
President-Elect	- Barclay G. Jones
Vice-President	- Clark D. Rogers
Vice-President	- Andrei Rogers
Treasurer	- Kenneth J. Dueker
Secretary	- Donald H. Blatt
Board of Directors	- William L. Garrison
	Joel M. Kibbee
	William L. Clark
	George McGimsey
	George S. Duggar

Andrei Rogers
Conference Chairman
University of California
November, 1966

TABLE OF CONTENTS

Page

"A 'Federated' Statewide Information System"
Robert Donati, Lockheed Missiles & Space Company 1

"An Automated Data System: The Los Angeles Approach"
Glenn O. Johnson, Los Angeles City Planning Department 10

"Development of a Management Information System for the New York City Planning Department"
Wilbur A. Steger, CONSAD Research Corporation. 24

"A Systematic Approach to Metropolitan Analysis: Lessons From a Study Employing Census-Tract Data"
Donald L. Foley, University of California, Berkeley 50

"The Economic Component of a Regional Socio - Economic Model"
Robert G. Kokat, IBM Research Team 73

"A Program Approach to Information Systems"
Richard A. Siegel, California State Office of Planning 89

"Keynote Address"
Donald H. Blatt, Director of Management Information, Nassau County, New York 100

"The Role of Models in Setting Values"
Claude Gruen, Urban Economist, Arthur D. Little, Inc., San Francisco, California. 111

"The Present and Futures of Urban Land Use Models"
David R. Seidman, Delaware Valley Regional Planning Commission 119

"Analysis and Design Economics of Time-Sharing Systems"
Robert E. Fagen, Computer Communications, Inc. 129

"The M.E.T.R.O. Urban Game-Simulation, An Experiment in In-Service Training"
Richard D. Duke, Michigan State University 142

"Urban Geocoding Systems and U.S. Census Implications"
William L. Clark, Lecturer in Civil Engineering and Assistant Director, Urban Data Center, University of Washington, Seattle . . .155

"Current Developments in Heuristic Gaming At Cornell University"
Allan G. Feldt, Department of City and Regional Planning, Cornell University 160

"Organization for the Retrieval of Spatially Distributed Data: A Land Use Data System (LUDS)"
Kenneth J. Dueker, Departments of Civil Engineering and Urban and Regional Planning, University of Wisconsin 168

"Spectral Analysis of Spatial Series"
W. R. Tobler, Department of Geography, University of Michigan . . .179

A "FEDERATED" STATEWIDE INFORMATION SYSTEM

Robert Donati*

1. INTRODUCTION AND SUMMARY

Interest is awakening to the potentials of statewide systems for providing and handling governmental information. Throughout the country today, progressive state governments have become well aware of many problems resulting from the proliferation of electronic data processing (EDP) applications and usage on a single-organization basis, and are giving serious attention to more coordinated and intensive development of EDP on a broad, government-wide scale.

Some research projects in recent years have advocated development of fully unified state and local systems with complete centralization of information. This paper explores a different approach – that of centralizing "information about information." The concept is based on creating a federation of autonomous information centers, to be interconnected by a statewide information indexing, switching, and communications network.

Under the federated concept, the statewide information system accepts requests for information from any terminal, determines information sources, obtains and relays the appropriate response to the requestor. The remote centers are individual departments and joint or regional complexes. The network would allow man-machine and machine-machine dialogue throughout the system. The central electronic index would make all data available to network units without physical or organizational centralization of files.

This kind of integrated information system was recommended to the State of California by Lockheed Missiles & Space Company in the study the Company conducted for the State in the first half of 1965. The paper concludes with a description of the development activity for the statewide system which the State has budgeted for the 1966-67 fiscal year; also summarized are two projects in land use and uniform county information systems which are closely related to the evolution of the federated system. All of these three activities should produce tangible products of use to planners within the next year.

---

*Research Specialist, Government Information Systems,
Lockheed Missiles & Space Company, Sunnyvale, Calif.

2. CONCEPT MOTIVATION

Problems facing all levels of government today are transcending the jurisdictions of classically defined departments or specialized organizations. Departments dealing with health and welfare, law enforcement, administration of justice, education, natural resources, transportation, urban renewal, physical development, and the like all need improved, quicker, and routine access to information on a cross-departmental basis. These conditions are prime inducements for development of a statewide information system.

An information clearing-house for data and services rendered to particular persons, places, business firms, governments, and other entities would reduce duplication and make government programs more effective provided, of course, that individual and corporate rights are fully protected. A need also exists for one-time collection of information for any authorized government organization in the format and time required. This requirement is the so-called "single-capture and single-update" of data.

The rudiments of a statewide system are already under development and in operation in many states. Within certain functional chains of state and local jurisdictions, such as law enforcement, administration of justice and education, statewide networks of computers and communications systems are now being established solely for these specific uses.

A large duplication of effort occurs in the proliferation of information systems for single departmental or single jurisdiction use. Nearly all these developments are proceeding with little or no interdepartmental or interjurisdictional coordination.

Some state departments and local governments may never be able to afford computing facilities of their own. Developing an integrated information processing network will give the smaller users access to computing facilities which they could never justify individually.

3. SYSTEM CONCEPT

This paper describes the functions of a statewide information system designed to meet the previously discussed trends and general requirements. The system concept is based on establishing a federation of essentially autonomous government information processing centers located throughout the state and interconnected by a statewide network.

3.1 Elements of the System

The two basic technical elements of the federation are (1) the statewide information indexing, switching, and communication system whose nucleus has been named "Information Central," and (2) the separate EDP facilities to be interconnected by Information Central.

> Information Central will use communications systems to join information processing facilities regardless of their physical location. Essentially, the Central will accept requests for information from any connected terminal, determine where the information is stored, obtain the information from any other centers, and relay the response to the requestor. The Information

Central communication subsystem can also be used for direct exchange of data between two or more organizations.

EDP Centers are computer complexes in certain State and local departments which have heavy and specialized data processing requirements. These facilities could be individual departmental centers, shared facilities, or regional complexes. A State Service Center would be one of these centers and would provide automatic data processing services to organizations lacking these capabilities.

3.2 Questions of Standardization

The federated statewide information system is based on the premise that independent, geographically distinct information processing centers can be effectively integrated into a system without requiring full physical consolidation or loss of local autonomy. However, in order to provide a high degree of automation in the exchange of information, such a system must have an amount of element compatibility, provided through standardization activities of the participants and/or translational capabilities provided by the system. Also required for automatic information retrieval is a master index and terminology thesaurus to all data maintained at the separate centers as well as a common inquiry language.

Information Central has a role somewhat analogous to the control element of the Department of Defense Communication System which allows automatic routing of messages among the three independently developed communications systems of the Army, the Navy, and the Air Force. The federated system is also grossly analogous to the railroad system in the United States today: independently developed, operated, and controlled railroad companies are able to maintain a continuous flow of cars subject only to a few basic constraints, such as track gauge, track curvature, braking and connection systems, and maximum dimensions on cars.

3.3 Advantages of the Federated Approach

This federated system is especially designed to meet the needs of the larger states, such as California, where the government operations and decision-making machinery is large and complex, numerous offices are geographically dispersed, and substantial data processing facilities already exist. The concept also provides voluntary participation of local government without subsuming local EDP as part of a completely rigid state information system.

The development of a federated statewide information system is feasible on the basis of current technology in information processing and data communications. The proposed system can also be evolved from the EDP base presently existing in state organizations.

Although there is an increasing need for improved availability and accessibility to information on an interdepartmental basis, the flow of information is still predominantly in vertical channels among functionally related departments. Functionally structured files are desirable (in place of completely unified files) since most data processing is strongly mission-oriented. Thus, functionally oriented files would need to be created from any large descriptive files to minimize search times and to reduce transmission costs of information used in routine, departmental activities. The federated system recognizes this functional orientation yet provides for efficient interconnection of departmental files.

Geographical centralization of organizational files to construct statewide data banks is unnecessary. The advantages of data banks can be achieved through modern data communication techniques allowing information exchanges without the access and organizational problems associated with complete file consolidations.

The development of a federated information system does not require or hinder reorganization of government. The concept is sufficiently flexible to be responsive to future growth and organizational change.

In sum, a federated system is technically feasible and practical, and can build effectively on the present EDP base. A federated system does not affect government organization and allows individual agencies to retain close control over their operations.

This approach affords the best compromise between centralization and decentralization. There are conditions prevalent within state government today which favor both approaches. This unique feature of the federated approach is that it makes possible the centralization or regionalization of those information activities which are more amenable to this treatment while allowing other information handling to operate in a decentralized fashion when desirable. In this way, a balanced condition can be reached for the entire statewide system.

4. INFORMATION CENTRAL FUNCTIONS

The functions of Information Central provide the missing link in the government information systems picture today – that of physically connecting the individual centers, indexing the available information, and providing an integrated information system based on the greater availability of data.

4.1 Information Indexing. Information Central contains an index of all information to which it has access. Ideally, the total spectrum of state and local government information within the State is classified and indexed, including information to which the Central does not have direct electronic access. The index shows the organizational files in which a particular item of information is stored, the file structures and media, and conditions for accessing and updating the files. For any information item, the index identifies the data which most closely correspond to it in the files of the participating organizations. It is emphasized that Information Central does not normally store data but only the retrieval parameters (such as location, field length, etc.) for data.

4.2 Information Retrieval. Requests for information from both human users and computer systems are received by Information Central in some stylized code or language. The user may request direct retrieval of particular items of information, searches for information, and computations based on retrieval of source data from files.

4.3 Switching and Transmission. Information Central provides the communications network for transmission of data and requests for data among the various computer centers and inquiry terminals connected to the statewide information system. In one mode of operation, assuming equipment compatibility, the Central serves solely as a switching device, directly connecting the requestor with the data processing facilities of the organizations known to possess the desired information.

4.4 File Updating. Updating messages are automatically transmitted to appropriate organizational facilities to indicate changes in information about which they want to be informed. The manner in which a file-updating message is processed locally is at the discretion of the individual organization.

4.5 Information Conversion. Computer programs effect translation between incompatible representations of information when economically feasible. Software also accomplishes conversion of information from one file structure to another. The inquiry language is uniform for all users.

4.6 Transaction Analysis. Message and information flow through the Central is automatically recorded to provide a firm factual and statistical basis on which further system developments can be planned.

5. INFORMATION CENTRAL OPERATION

Information Central operates in the three primary modes.

5.1 Mode I – Response to Queries for Detail Information

In this mode, the total system serves as a fact retrieval system with man-machine dialogue. The requirement here is for the retrieval of specific items of information. The response to the user includes (1) itemized facts (i.e., data) and/or (2) sources of facts, including file location and file search parameters. In the first case the system operates as a pure fact retrieval system, this being the exclusive mode for those terminals with no EDP capability of their own. The second form of response provides information on the sources of data which can then be incorporated into regular computer programs operated at the separate EDP enters of the statewide information system.

5.2 Mode II – Response to Requirements for Input to Operation Programs at Terminals

From those connected departmental files which have random-access storage with priority interrupt in the computer system, Information Central can obtain data for a computer program under operation in another organization. Thus, in this mode the total system operates as an interconnected multiprocessor system performing a variety of data processing functions utilizing information from many files.

The specific parameters required to retrieve particular data can be obtained through a Mode I operation in a form that is directly usable by computer programs in terminal organizations. This feature makes it possible for these organizations to run operational programs and produce reports using up-to-date data from any of the other files in the system, thus reducing the need for human intervention at the point of routine or repeated information retrieval.

5.3 Mode III – File Updating at Primary and Secondary Files

Information Central makes it possible to update the files of one organization using data collected by other organizations. Updating is performed on a strictly voluntary basis on the part of the agency collecting the information and of the agency whose files are to be updated. Each organization chooses the kind of updating messages, if any, that it wants to receive. Each agency also has the discretion of developing its own procedures for processing update messages.

Two primary considerations require the maintenance of certain information in more than one file:

- Some information is used with sufficient frequency that its accommodation in permanent organizational files can be traded off against obtaining it through Information Central each time it is required.
- Certain temporary organizational files may be of relatively small volume compared with the permanent files.

No great need presently exists for information updating on a real-time basis except for some few categories of information as applied by the primary using organization. This is probably true of most data used by an organization other than the original collector. Peak traffic on large data processing systems during the prime shift hours generally dictates that the kinds of applications which are processed be limited to those which require real-time or near real-time reactions. Thus, file-updating is done in most cases on a batched basis during non-peak hours.

6. DEVELOPMENT PLAN

The development of a federated statewide information system for California would require a 10-year period. The development program is initially concerned with specific studies designed to (1) definitize the overall program; (2) establish the broad policies under which the system will be designed, activated, and operated; and (3) establish a pilot program. The early work is strongly addressed to the policy, coordination, and compatibility questions. By the end of the first three years, the Information Central computer facility will have been activated; by the end of five years, the federated system will be operating with limited tie-in to EDP centers.

7. SYSTEM COSTS

The basic philosophy underlying the development of a federated statewide information system is to build upon the existing information systems of a state, adding only those increments needed to convert the aggregation of computer facilities into a statewide system. Consistent with this philosophy, costs were estimated in the case of California for (1) the development of Information Central, (2) those items needed to integrate departmental EDP centers into the network, and (3) the development of certain standard EDP application packages.

The 10-year developmental costs were estimated at $68.4 million and operational equipment rental at $29.8 million, the total developmental cost of the 10-year program thus being $98.2 million. Estimated operating costs after 10 years will be $13.4 million.

The present <u>total</u> information-handling costs for the California government were estimated at $538 million. These figures include hidden costs which are not reflected in the direct costs of electronic data processing, <u>per se</u>. (A review of the departmental functions revealed that the typical state employee devotes nearly half of his time to information-handling activities.) If present trends continue, the total true cost will rise to \$1.8 billion annually by 1980, owing to the increase in state personnel required to provide expanded services to an increasing population.

It was estimated that the development of the Statewide Federated Information System would produce net savings of \$170 million annually by 1980, over and above savings of \$180 million annually anticipated through increased departmental EDP developments.

8. CURRENT ACTION BY CALIFORNIA

The Automatic Data Processing Advisory Committee (ADPAC), a standing committee of top State of California administrators which recommends statewide policy in EDP matters, was charged with primary responsibility for evaluating the recommendations of the Lockheed study. Upon reviewing the responses of all the various state departments and local government, ADPAC recommended to the Director of Finance that initial planning for the Statewide Federated Information System be budgeted for the fiscal year commencing July 1966.

Responsibility for developing the statewide system has been assigned to the General Services Department, which has been given a \$271,000 budget for this work in 1966-67. Four tasks are to be initiated with the following accomplishments expected within this fiscal year:

Task	Products
(1) System design	• State EDP design standards manual • Preliminary design of a prototype operation of a federated information system
(2) File structure study	• Inventory of machine-sensible data • Thesauruses for people, land, and operational data • Semantic usage lists
(3) State-local cooperative studies	• Master index of intergovernmental data exchange • Description of high-volume data exchanges • Inventory of real-time and teleprocessing capabilities in local governments
(4) Planning and control	• Refined implementation plan for the federated system • Policy document for the federated statewide information system

These tasks will provide tangible benefits useful immediately to planners and other government employees while the development of the overall information system is simultaneously advanced.

9. RELATED PROJECTS

The study which proposed development of the federated information system for California also recommended that the State pursue the development of the component departmental or functional information subsystems. It was also suggested that the State agressively promote state-local government cooperation in the development of mutually related data processing applications and that local governments should pursue some joint effort at developing commonly required computer programs. Two major efforts along these lines have been initiated.

9.1 California Regional Land-Use Information System Project

This twelve-month project is the first phase of a longer, three-to-four year effort in developing a rational, integrated land-use information system for governmental and private use in California. The present $330,000 program, for which HUD provided two-thirds of the funds, is under the direction of the State Office of Planning. Other direct government participants are the State Lands Division, the Soil Conservation Division, the Department of General Services, and the County of Santa Clara which is to serve as the demonstration base for the study.

A major purpose will be to establish formal, statewide coordination for the consideration of land-use information issues. Consequently, federal agencies, other local jurisdictions, and national research efforts in this area will be involved in the project.

The project also will identify users and typical inquiries for land-use information, determine data availability, specify the kinds of data to be included in a land-use information subsystem, recommend necessary standards, identify and study relevant issues, and demonstrate rational exchange of information in the field of capital investments.

9.2 California Supervisors Association EDP Project

The County Supervisors Association of California, the official voice of county government in the State, has appointed a full-time EDP committee. The development of specifications for a uniform county information system was inaugurated by this committee in January 1966. A standards subcommittee is making recommendations on programming, analysis, and operating procedures. A subcommittee has been established for each of three major information areas (people, property, accounting-inventory-supply). Membership is composed of representatives of the county departments concerned with the subject area, the delegates having been appointed by the corresponding statewide association of that particular office. County managers and data processing organizations are also represented.

By the end of October 1966, it is anticipated that each subcommittee will have identified every data item required for its subject area, showing present, planned, and future data processing reports and outputs for which the data are required. Further refinement and analysis will occur until the end of the year when a decision will be made on what tasks, such as programming, will be attempted in 1967.

10. CONCLUSION

Planners have been in the forefront in urging the development of government-wide information systems. Now that actual system developments are under way throughout the country, such as the three in California described in this paper, it behooves planners to continue an active role so that their requirements will be met. While it is true that planners use a cross-section of government data there are sufficient differences between information for planning and for government operations to warrant close involvement of planners in any broad information system development.

Much information and facilities for data processing is not suitably oriented for planning and research. Government operations tend to be oriented to recording what has happened on individual entities; planning, however, is more concerned with group and area characteristics and predicting the future. Special requirements for data file structure may be desirable since planners need flexible information retrieval and cross-association capabilities. A good historical base must be retained, and information on a given subject should be comparable between jurisdictions, and, of course, planning information varies between government levels. Such a list of special requirements is only rudimentary, at best, but should serve to indicate the necessity for alertness on the part of the planning community.

A statewide information system has a potential for affording new insight into the emerging problems facing government. Conventional data provided in conventional channels will become less and less adequate as problem boundaries expand beyond the traditional jurisdiction of single functional organizations. Taken as a whole, state government has a substantial proportion of the total data needed by individual agencies. The federated statewide information system unlocks these data and provides an information system that corresponds better to the true nature of problems while retaining control over local operations and decision-making.

AN AUTOMATED DATA SYSTEM: THE LOS ANGELES APPROACH

Glenn O. Johnson*

The City of Los Angeles' experience in the use of Data Processing Equipment dates back over 30 years, since its 1923 application of tabulating equipment in the recording and tabulation of crime statistics. The process of automation of the highly repetitive tasks has been a steady one. Although originally conceived as strictly a tool to handle day-to-day operational tasks, automatic data processing (ADP) is now viewed as a means of achieving a total information system to serve public as well as private operational and future planning needs.

1.0 LOS ANGELES MUNICIPAL INFORMATION SYSTEM

Before discussing the Los Angeles City Planning Department's Information System it is appropriate to discuss the Municipal Information System of Los Angeles, of which the Planning Department's System is a part. Under the policy direction of the Board of Administration of the Data Service Bureau, with technical assistance from the Data Services Working Committee, the City of Los Angeles has embarked on a program for integrating and further automating the City's information handling processes. This program is generally referred to as the Los Angeles Municipal Information System (LAMIS). Practical support to LAMIS was provided following a year of intensive study. In December 1963 the Mayor and City Council approved an ordinance providing for the centralization of data processing systems in the several departments and offices of the City of Los Angeles. The Board of Administration was designated the policy head of a newly created Data Service Bureau. Under the ordinance creating this bureau the City Controller was charged with the interim responsibility of pulling together all of the City's data processing capability into a single and meaningful operating department.

The new Data Service Bureau is serving all operating departments of the City without interfering with their conduct or Charter responsibilities. Presently 45 of the City's departments and offices using over 400 individual computer programs are being serviced by the bureau.

---

*Los Angeles City Planning Department

The Los Angeles Municipal Information System is composed of many related parts or subsystems which are in varying stages of development. Included are: (1) Purchasing and Stores Information System; (2) Budget Formulation and Control System; (3) Fiscal and Accounting System; (4) Personnel Management Information System; (5) Automated Planning and Operations File; and a variety of related information subsystems of the individual operating departments and bureaus of the city government.

A basic objective of LAMIS is to design these subsystems so that they interface with one another to the fullest extent. The subsystems related to City Planning will be commented on in further detail in the succeeding paragraphs.

2.0 LOS ANGELES CITY PLANNING INFORMATION SYSTEM

If the elements of the planning process are considered in relation to automatic data processing applications that are currently in use or being developed it becomes evident that ADP can play an important role in practically every aspect of city and regional planning. In reference to Figure 1, Systematic Planning Process, once the decision to plan has been made and the goals and objectives formulated, the computer has a role to play in:

- Data Inventory
- Data Analysis
- Development and Evaluation of Standards
- Plan Preparation
- Alternative Testing
- Plan Implementation

Figure 1 indicates the areas where ADP has been employed in Los Angeles. These applications are mentioned briefly at this point; they will be classified and discussed in greater detail in the next section of the paper.

A great deal of interest has been generated in the development of data banks to satisfy the inventory requirements for planning. The more sophisticated approaches are designed to provide for the collection, storage, maintenance, and retrieval of pertinent information at a centralized location.

In the areas of data analysis and standards evaluation, computer mapping and graphing techniques, math-models and statistical manipulation programs have demonstrated a great potential in understanding the city in terms of its physical, social and economic character.

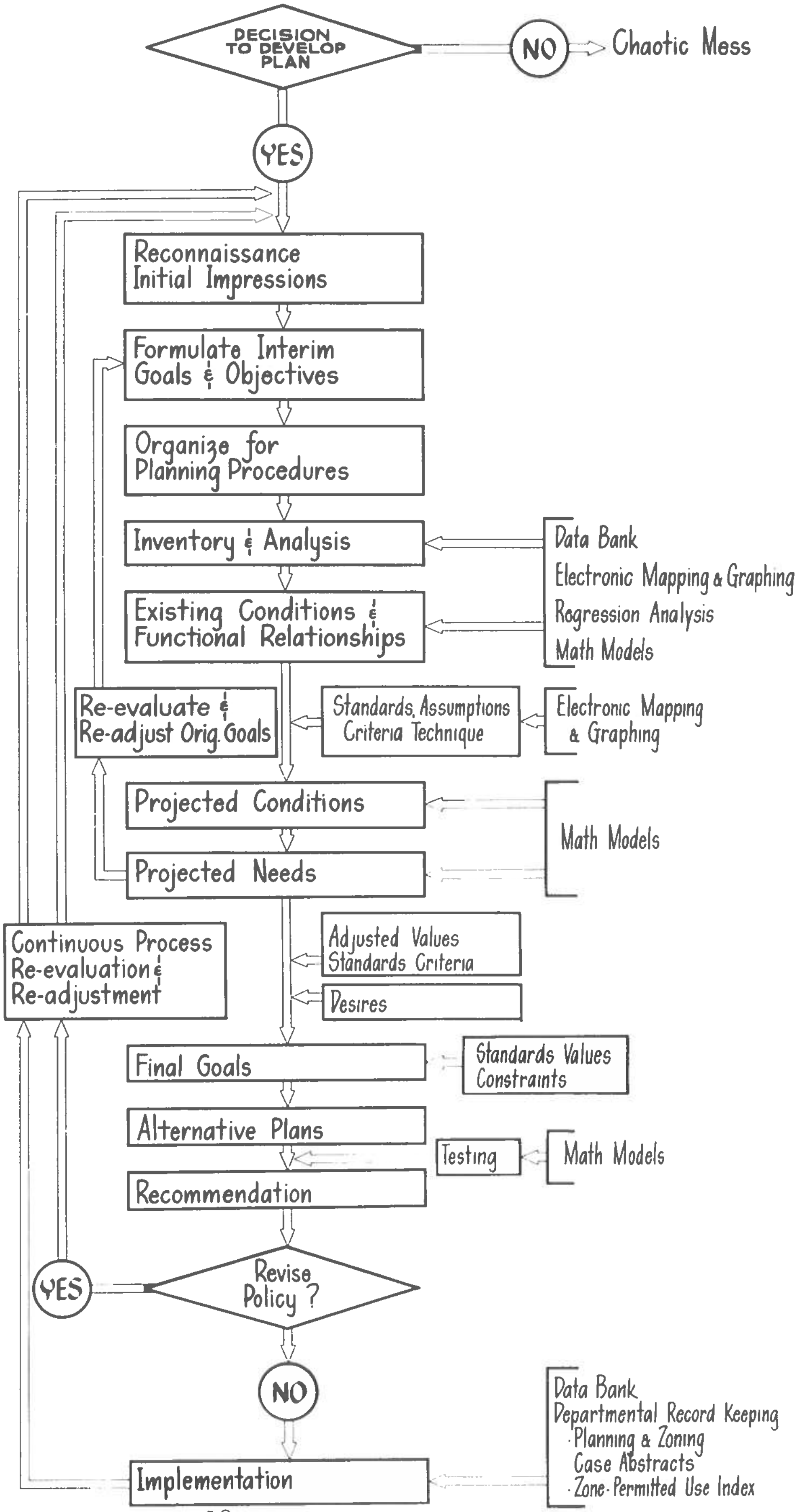
Mathematical models have been employed in the area of projections. One which is currently being used projects population by sex, age and race using birth rates, death rates and migration data. Math-models are also potentially valuable for evaluating alternative comprehensive plan proposals.

Finally, the implementation phase of planning also provides an area in which ADP can be employed. The data bank, for example, can be used to retrieve names and addresses of property owners and to prepare notices concerning public hearings, and also to provide pertinent information during the consideration of zoning matters. Other potential applications are automated abstracts of planning and zoning cases and indexes showing permitted uses in various zones.

FIGURE I

SYSTEMATIC PLANNING PROCESS

SHOWING DATA PROCESSING APPLICATIONS



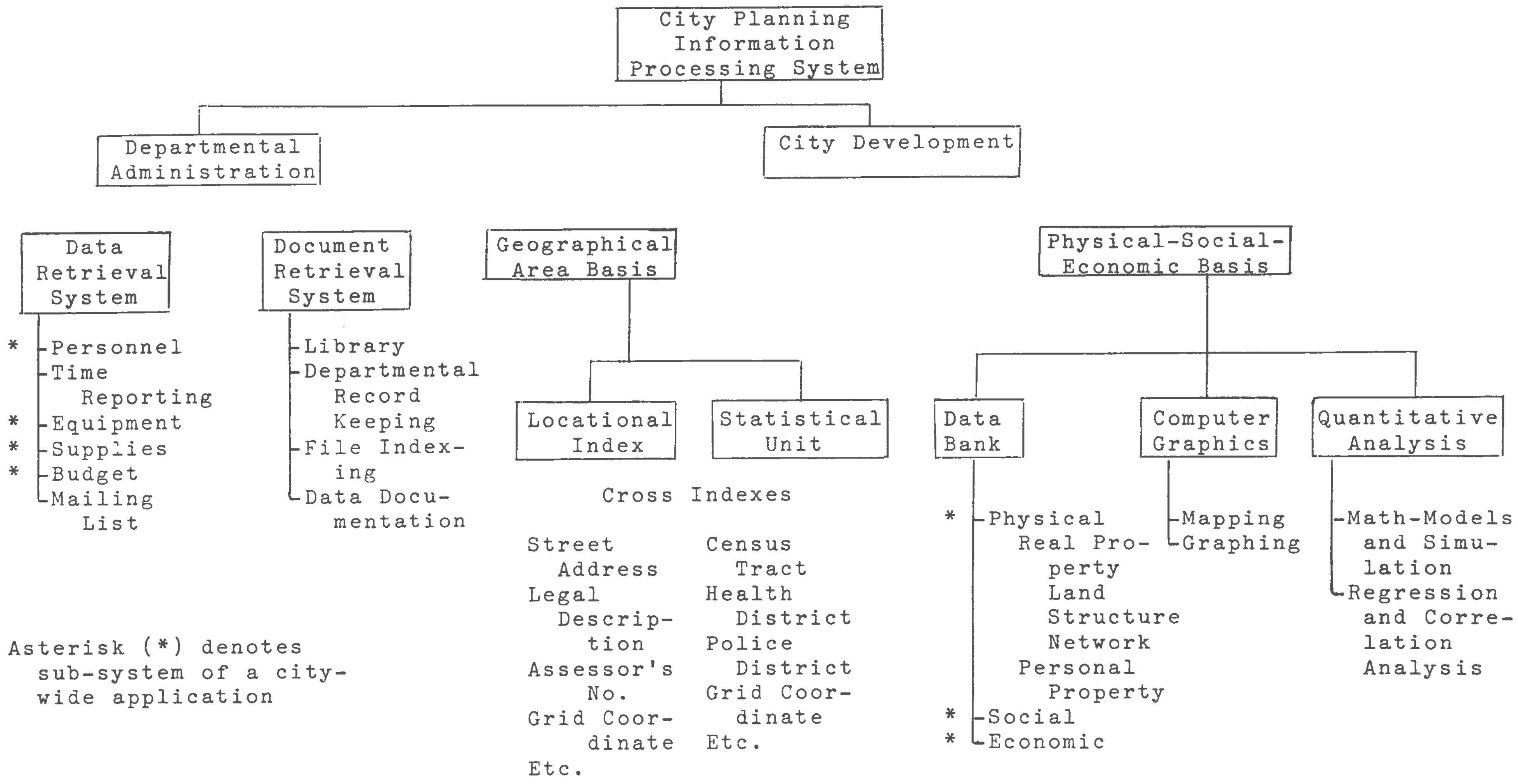
The Los Angeles City Planning Department's Information Processing System has been developed within the framework of LAMIS. Its elements are depicted in Figure 2 on the next page entitled "Los Angeles City Planning Information Processing System - A Classification Of Its Elements." The applications listed include those presently under systems design as well as those that have been fully developed. The elements of an information processing system have been broadly classified into two basic functions--City Development and Departmental Administration. City Development refers to the system elements that describe the processes which in some way relate to the city's growth or decay. Departmental Administration refers to those elements that reflect activities which take place within the department and are administrative in nature. The data processing systems elements of the Departmental Administration Class can be grouped into two general types--a Data Retrieval System and a Document Retrieval System. Such a system consists of selected characteristics that are identified to some statistical or administrative unit of measure. Typical units required for a city planning department are: a person, work assignment, or piece of equipment. A data retrieval system may or may not apply mathematical operations on the values of the characteristics in the system.

2.1 DEPARTMENTAL ADMINISTRATIVE ELEMENTS

Several types of applications have been listed on the diagram under the data retrieval system. Among them are four which have multiple agency implications and consequently might be developed on a city- or county-wide basis as opposed to just a city planning department. Those four applications are: the personnel record keeping system, equipment inventory system, supplies inventory system, and budget control and management system. The Personnel Record Keeping System is currently in a systems design stage of development for the City of Los Angeles. It will consist essentially of an automated record of each governmental employee which includes pertinent characteristics about that employee such as his date of birth and other personal characteristics, work assignment history within the department, history in and outside of government, his interests, educational background and any other pertinent information about the individual employee that would be useful to management. An automated record of this kind will permit the department manager at a moment's notice to retrieve a list of personnel whom he feels would be needed to perform a particular study. For example, he might request a list of all individuals working for a city who have geology majors in college or geology experience. Such persons could then be brought together to resolve a particular problem that is imminent.

Figure 2

LOS ANGELES
CITY PLANNING INFORMATION PROCESSING SYSTEM
A CLASSIFICATION OF ITS ELEMENTS



The Equipment Inventory and Supplies System refers to the common business or accounting applications which are familiar to most people and will, therefore, not require further explanations except to say that the department presently has such systems in tab card form.

Automated methods in budget and management control are relatively new applications particularly in government. Here is an opportunity to automate most of the records or reports that are produced during the budgetary process. Los Angeles has conducted a successful feasibility study of this process and is currently incorporating the system into one bureau of the City, namely, the Bureau of Sanitation. The automation of these records will permit the department manager or team working on the budget to receive quickly pertinent reports and have the capability of analyzing the effect of decisions made during budget preparation and throughout the fiscal year.

A Time Reporting System, currently being handled on a manual basis will be automated soon and will involve the processing of key information with regard to the time spent by employees on particular jobs. By placing this information on cards or tape, periodic reports will be generated easily which will indicate the cost and amount of time spent on a particular job at a particular personnel level such as professional, sub-professional, or clerical.

The Mailing List is another type of application that is simple to develop and very beneficial particularly in the larger agencies. By creating a tape or card record of a name, address, and firm name, along with some kind of code indicating the type of person or firm, it is possible to retrieve a selected list of agencies with whom you may be interested in corresponding. The implementation of such a system can also include the provision for automated addressing of envelopes.

Among the documents that can be handled by a document retrieval system in a planning department are a zoning case file, magazine article, letter, report, photograph, or map. The primary function of this kind of system is to identify a location or pertinent features of the complete text for which a search is being made. Unlike the Data Retrieval System, mathematical operations are normally not applied to the information in this system. One of the applications that falls under the document retrieval system is an automated library. The URBANDOC project is a notable contribution in this area. URBANDOC stands for Urban Documentation and Retrieval System. This system, which is currently a pilot project being conducted in New York, will permit a researcher to obtain a computer print-out of a full bibliographic reference complete with abstract and the designation of what libraries in his region have the available material for reference purposes. A similar system on a smaller scale is currently being considered by the City of Los Angeles.

Another application, the Departmental Record Keeping System, refers to the development of an automated system for recording, maintaining, and retrieving information related to planning and zoning cases. In such a system the essential information for planning and zoning cases with possibly an abstract of the case can be maintained on some kind of disc storage. This permits immediate retrieval of essential information about a prior case and assists an investigator in studying his current case. Further, it allows other departmental personnel to determine the current status of any individual case. For example, using a scope or typewriter device that is remote to the central processor, it is possible to retrieve the current status of a planning case and key information describing that case within seconds. This application is also in the system design stage of the City of Los Angeles.

There are also potential applications in the area of file indexing and mapping. Emphasis has been placed particularly on mapping recently. It's now possible to store maps in core storage and later retrieve them with the aid of ink plotters that are on the market today. Of course this application begins to overlap with the classification on the other side of the diagram of Figure 2 under Computer Graphics. The Datadox System, developed by System Development Corporation, is another useful document retrieval system. Datadox describes the contents of data to any inquirer. It provides the inquirer or potential data user with accurate and timely documentation on the availability of particular data as well as the characteristics and location of these data so that they may be readily accessed. Los Angeles is currently using Datadox as part of its math-model development program.

2.2 CITY DEVELOPMENT ELEMENTS

The elements which relate to the growth and decay of the City identified on the diagram as City Development applications might be considered either on a geographical area basis or a physical, social, and economic basis. On a geographical area basis, the basic problem is concerned with the determination of an appropriate location index or statistical areal unit. By location index, is meant some locational identification of a parcel, such as street address, legal description or map book page and parcel number. Statistical or administrative areal units are the geographical areas used by individual agencies to aggregate data for reporting and analyzing purposes. There are usually several units used in any one jurisdiction. For example, census blocks, census tracts, police reporting districts, sanitation districts, health districts and school districts. A grid coordinate might be used to identify a statistical areal unit, although it can also be used as a location index. Since agencies generally have maintained records on a historical basis by their own particular statistical or administrative unit, a change to a new standardized unit is difficult to effectuate. Consequently, it is necessary to identify individual parcels or other records by all of the statistical or administrative units used in the particular jurisdiction.

Recently, there has been a great deal of interest in the development and use of the coordinate system to identify data. This recording system seems to offer a common denominator between all geographical identification systems. Los Angeles plans to incorporate a grid coordinate system into its data banks which will be discussed in the following paragraphs.

Whether a coordinate system is used or not, correspondence tables are an essential element of any information processing system. These tables show the relationship of the various statistical areal units used within a region.

Applications concerning the physical, social and economic aspects of city development have been grouped under three headings in Figure 2 -- data bank, computer methods of graphic display and quantitative analysis.

The City of Los Angeles' Data Bank endeavor is in the form of the development of an Automated Planning and Operations File, commonly known by the acronym APOF. The history, scope and status of APOF is provided in a report entitled "Final Report of the APOF Conceptualization Study" dated March 31, 1966, prepared by System Development Corporation. The following comments cover many of the significant findings of this report.

"APOF is designed as an automated information system within a municipal information structure that includes other information systems both manual and computer-based. The APOF System collects, stores, retrieves, summarizes, analyzes and provides geographically significant environmental information for city departmental planning and operations."

For the APOF to perform these functions properly certain critical features are required in the system's design. Criteria established by the City for the design and operation of APOF are as follows:

1. "The Automated Planning and Operations File should consist of all pertinent statistical data required to meet the needs of city agencies conducting environmental and operational planning.

2. The file must be responsive to change over time.

3. The city must develop its own capability to operate and maintain the file.

4. System design will be such as to avoid costly duplication or system redesign in later years.

5. The file must store information on each individual parcel of land in the city with complete flexibility in retrieving such information.

6. It must be able to summarize information for any selected area of the city.

7. The file must have the ability to add or delete items of information as required without reprograming."

"To accomplish these objectives the APOF computer processing subsystem must have access to many computer files. Some of these data files will be maintained and updated by APOF itself. Others will be maintained by other automated information systems either managed by the City or by other agencies. Data concerning the physical environment of the City are defined as APOF object data. The APOF System will maintain and update these files. It will have access to but not maintain other files, examples of which are Police, County Health Department, U. S. Bureau of Census, Traffic, and Utilities. Before the APOF System can retrieve information from these files the data will be summarized by appropriate geographic districts so that the protection of private information is assured.

"The information base contained in APOF describes the physical environment of the City. These APOF object data are grouped into land records data and network data. The former are data about land parcels and their attributes. The latter describes physical facilities that are not parcel oriented but occupy space and serve the City. The land records category consists of two data sets. Namely, these are the land parcel data set and the establishment data set. There are seven network data sets -- right-of-way, intersection, water, storm drain, sewer, power, and refuse collection.

"A grid coordinate method is recommended as part of the APOF System for geographic area identification. This method provides a greater degree of processing flexibility than any other geographic indexing scheme.

"In order that the APOF System may properly perform its functions in a user-oriented manner and can be responsive to change over time a system of computer programs called the Generalized Program System is recommended. The nature of the generalized program allows for rapid responses to new requests and is adapted to changes in file content and format. It can also be utilized to process files created by other city information systems.

"The implementation plan for the Total APOF System shows how it evolves over a continuous seven-year period. During this time there are four major milestones at which incremental capabilities become available to city personnel. The first stage of APOF capability is referred as APOF I, which is designed to satisfy primarily the requirements of the City Clerk Land Records Division. The two most significant milestones, APOF I and Total APOF, are summarized below.

"In designing a system to satisfy the Land Records Division requirements, several alternative approaches were considered, some of which provide short-range savings in development at the expense of degraded capability and greater long-range expense. The recommended alternative includes a time shared, remote operation, general programing system and grid coordinates. If the criteria established for APOF System are to be met, this system is essential. Such a system would require two years to develop at an estimated $1.4 million.

"The Total APOF System evolves in three further stages to complete a seven-year development period. The estimated cost for this seven-year development is approximately $3.7 million. Estimated annual APOF operations cost during the seven years raised the total estimate to $5 million. The annual incremental cost for both development and operations of APOF represents less than half of one percent of the current City budget. Once the system is implemented annual operating costs are estimated at $482,000.

"It should be emphasized that the recommended first stage of APOF capability (APOF I) is a complete information system and satisfies applicable system criteria. Since the Generalized Programing System is part of APOF I, the capability of APOF I differs from the later stages only in terms of the information based data content. Its capability is only constrained by the amount and types of information available at a given time. Currently the City of Los Angeles is reviewing the alternatives that have been outlined in this report and in the next few weeks it is expected that a recommendation will be proposed to the City's Board of Administration. Whatever the alternatives will be, it is certain that this system will be an easier adaptive approach, that is, a system of programs that performs file construction, file updating and information retrieval and analysis in a manner specified by the user."

Graphic display is a major concern of any planning agency, so that it is not surprising to find planners interested in the computer methods which have recently been developed. The first experience of the City of Los Angeles in this area was in the use of electronic mapping and graphing programs developed by a staff at the University of Washington under the direction of Professor Edgar Horwood. These programs prepare statistical arrays and frequency distributions, and print data spatially arranged on a page for mapping purposes. A transparent overlay showing geographical area boundaries placed over the page containing spatially arranged data forms a map which can then be duplicated. Using 1960 U. S. Census and other available data, the City Planning Department has developed land use and social-economic profiles of the City.

In connection with the grid coordinate system to be developed as part of the APOF System the City is also exploring the possibility of graphic output. As visualized, the map of the City could be stored in the computer and printed out or displayed on a scope upon request.

There is no doubt that quantitative analysis is playing an increasing role in city planning today. The City of Los Angeles in the 1965-1966 fiscal year embarked on a mathematical development and simulation program. Three basic objectives have been outlined:

1. The first objective is to provide the Comprehensive Planning Section with information that will be helpful in the development of the Comprehensive Sketch Plan for the City of Los Angeles.

2. A second and longer range objective is to begin to build a series of models that will be useful for all phases of city planning to include residential, commercial and industrial land use allocation.

3. The third objective is to build gradually an in-house capability so that the staff will be able to rely on outside assistance to a lesser degree.

It was decided that the best approach was to begin with a model that has already been developed and is in operation. A residential location model was selected. This model was developed by Professor Britton Harris at the University of Pennsylvania. Its purpose is to predict residential location for households that are moving into an urban area. The model requires two inputs -- a projection of households by type according to social-economic characteristics and the number of vacant housing units in each sub-area of the region. A separate population forecast model is being used to determine reasonable exogenous variables. The application of the residential allocation model is providing several by-products which are of major significance to the department. One of these by-products is the development of an automated correspondence table which shows the relationship among statistical areas used by the different governmental agencies in Southern California. Its purpose is to make compatible various types of data gathered on the basis of different areal units in the region. A second by-product is the correlation and regression analysis of data which are necessary inputs for the residential model. Such data include census characteristics, land use, employment and unemployment. The analysis of this data is helpful to the comprehensive planners in their sketch plan process. A third by-product, the development of preference or utility functions for housing, will aid in the department's understanding of housing evaluation. A final by-product is the aforementioned population projection model. A prime virtue of this model is the ability to test different assumptions about migration, birth and survival rates. A series of technical papers will be published over the coming months that will describe the activities of the Mathematical Models and Simulation Development Section in more detail.

3.0 CONCLUSION

In conclusion, the City of Los Angeles has found that data processing applications can be used in a broad range of planning activities. Since many of the applications are relatively simple and can be implemented rather easily while others are complex and require a long period of time and considerable sums of money to develop, there are gaps of time available during the development of the long range applications which permit development of the shorter range programs.

The City of Los Angeles, like most cities, tends to take a rather conservative attitude in implementing many potentially valuable programs. This is understandable from the standpoint that building municipal information systems is a time-consuming and costly effort. Further, to date, no agency can claim to have a fully developed information system. Consequently, a city representative cannot point to any one agency as a prime example of a complete and successfully operating system.

So what can be done to encourage legislators and top city administrators to take better advantage of this new technology? First of all, there needs to be <u>initiative</u> on the part of the technician to continue to develop and promote new techniques. Secondly, <u>departmental support</u> is a necessary prerequisite. This means a director that is receptive to the technician's ideas and willing to support new proposed programs, particularly at budget time. Third, the city officials need to be <u>educated</u> about the new technologies. There are many ways of doing this. Civil service commissions are usually eager to develop and implement educational programs. One that comes to mind, that was developed in Los Angeles, was a course given to the directors of departments and held at a nearby resort area in which the basic fundamentals of data processing were presented. IBM Corporation has exceptionally good training facilities and if your jurisdiction is not near one of their educational centers, there are correspondence courses in which people can take data processing instructions through programmed learning texts. Fourth, is <u>leadership</u> from state, federal and special agencies. Cities tend to look toward government agencies at higher levels. Hence, the successful results from efforts such as those carried on in the transportation studies throughout the country and the Metropolitan Data Center project and the work that is going on in Santa Clara County and the State of California cannot help but encourage city level officials to take advantage of the new technology. Fifth, federal monies to finance these undertakings have been of great help to the counties and smaller cities -- 701 grants, etc. The use of coordinative bodies is a sixth means of promoting the development and use of new skills. Technicians mutually interested in taking advantage of data processing techniques can, working together, provide a significant influence on top administrators while at the same time improving the efficiency and effectiveness of city government. The City of Los Angeles has had such a body for the last several years in the form of the Data Services Working Committee. This committee is the working group for the Policy Committee of the Data Service Bureau. The working group meets periodically when assignments are given to them that involve more than just one agency's data processing program. The purpose of the group is to insure that data processing applications are coordinated. And finally time itself is a factor. Much of the work that has been done to date has been experimental and new. As time goes by and as more and more experiments become tested and proven, successful city officials will recognize the potential and take advantage of this new and rapidly expanding technology.

GOJ
8/66

DEVELOPMENT OF A MANAGEMENT INFORMATION SYSTEM FOR THE NEW YORK CITY PLANNING DEPARTMENT

Wilbur A. Steger*

1. INTRODUCTION

While it would appear from the very existence of this Conference that we "know a lot" about urban information systems, we do not know very much, unfortunately, about the construction of "information systems for urban planners." The two are not the same, by any means. In fact, while urban information systems exist (in their myriad forms) often due to the initial efforts of a planning organization, the use of the systems now in operation (e.g., Alexandria, Pittsburgh, Santa Clara, and Tulsa, to name a few) is just as likely to be by an operating department or line agency, not the planning organization(s).

Not that this is an undesirable situation, since it certainly was at least partially intended that way: but less than delight, I think it is warranted to say, is the reaction of even the most sophisticated planners to the current systems. And for a number of reasons.

I should like to spend my time today discussing some of these reasons, drawing upon the experience of the work program upon which CONSAD Research Corporation has embarked with the New York City Planning Department.

While the CONSAD work program extends into the more operational aspects of the City Planning Department's functions such as zoning and other land use controls which are implemented in its Office of Technical Controls, our discussion here will concentrate on the Office of Comprehensive Planning, its role, and informational needs. In so doing, we shall also be dealing at least partly with two other offices, that of the Executive Director and the Capital Budget Division.

In a statement regarding the objectives of the City Planning program and organization, the Executive Director states:[1]

> Comprehensive planning is the major function of this agency. Comprehensive planning is a sophisticated work process which can produce the information, analytical capability and interrelated planning policies from which an overall strategy of City development can be made to guide specific program proposals and development decisions. Developing such a work process is the primary objective of the planning program ....

---

*CONSAD Research Corporation

By the very nature of the comprehensive planning process, the comprehensive plan would reflect the dynamics of our continually changing urban environment. It cannot be a static, rigid, long-range ideal plan -- although longer range proposals would certainly be included in the plan.

2. SOME OF THE FUNCTIONS OF COMPREHENSIVE PLANNING AS SEEN FROM THE VIEWPOINT OF INFORMATION REQUIREMENTS

The Department summarized its own status with respect to information capabilities, as of September 1965, in this way:[2]

> The Comprehensive Planning Program of the Department of City Planning will require an increased capability to collect, analyze and retrieve information important for planning and decision making. At the present time, the Department's ability to handle data needed for all sorts of planning purposes is limited. We are one of the last of the larger planning agencies in the country not to have its basic information files in machine processing form or to have the capability to carry on planning analyses of various kinds using modern data processing techniques.
>
> Because of this, an important part of the Department's work program will involve building up its data and information handling capability over the next two years. The kinds of investments in machinery and personnel required in this new and constantly advancing field of information handling make it imperative for the department to have a well thought out and carefully conceived plan for designing an information system.
>
> During the next twelve months there are four important goals to be achieved:
>
> 1) To determine a definite direction that efforts to improve our information base and analytical methods should take. Preliminary organizational thinking calls for setting up a new Data or Information Service Division as a major unit of the Department. During the course of the next six months, we want to examine this concept in detail, determine its operating, functional, and staff requirements; how it should relate to other parts of the organization, and what parts of our present staff should be a part of such a unit, if and when set up.

2) To improve our information handling capability by developing data files and computer programs for current planning work and by the specification of requirements for handling the records and information of a limited number of divisions in the department.

3) To make operational an analytical method for estimating population and households in small areas and to begin to lay out other analytical methods to perform other essential planning tasks.

4) To institute a program for training staff in data handling and analytical methods needed for planning.

In this paper, we cannot hope here to describe the numerous functions of the Office of Comprehensive Planning. Suffice to say here that while it covers the gamut of policies which affect the physical and social conditions of the City's peoples -- be they investments in community facilities, housing, transportation, recreational, health, or other facilities and uses of land, the Office of Comprehensive Planning serves: (a) the Office of the Mayor, primarily in an advisory capacity; and (b) the remainder of the Department and Commission as a formulator of short, intermediate, and long-range policy within which more "technical" decisions, such as zoning and mapping can be made.

We would do best, perhaps, if we illustrate some of the various functions of the Office of Comprehensive Planning[3] and thus provide insight into the (present and potential) role of "information." The illustrations include: 1) project evaluation considerations for alternative renewal sites to be selected by New York City; 2) program budgeting; and 3) economic development programs for "urban poverty areas." While these are not "typical" examples in the sense of being "routine," they are representative in terms of their complexity and problem-solving process.

2.1 Example 1. Evaluating Projects for Alternative Renewal Sites: Intelligence Aspects

Evaluating projects for alternative renewal sites to be selected by New York City provides an excellent example of the combination of (a) <u>intelligence</u> and (b) <u>research</u> aspects of the role of comprehensive planning in the context of overall planning for the City.

The "intelligence" aspects are represented primarily by certain symptoms; for example, the short "time-fuse" available to develop the needed information concerning each of the renewal projects to be evaluated. This "needed" information, together with other information, allows the relevant public officials to make "quick" decisions regarding the prospects for each of the projects. The more fundamental indication that this is an intelligence operation lies in the fact that, as a process, New York City Planning is to know -- have at its beck and call -- the most current "information" to make project evaluation, but not know precisely in advance when the occasion to do so will be coming up! Furthermore, the typical difficulties of the intelligence situation are also present; not only is it uncertain which geographical areas and which projects will actually be suggested for evaluation, but it is uncertain (until the time of the "decision meetings") as to: (a) which criteria will be applied in a city-wide context, (b) which information will be acceptable to bring to bear upon these criteria, and/or even (c) which information will be brought into the decision-making process by other public agency representatives.

The "research" aspects of this decision-making process about renewal sites can be easily magnified beyond that which is present at the current time (or perhaps even contemplated for the future). However, it is intended that the work program of the agency will "explore and experiment with new concepts, new techniques, and hitherto unexplored substantive areas for the purposes of improving the long-range capability of the agency to carry on comprehensive planning for New York City."

The intelligence system for performing project evaluation of alternative renewal areas requires certain information which has been produced by previous research, development, and other activities of the Department of City Planning as well as other city agencies. The information for this purpose can be classified into the following categories:

1) Current conditions within the project area (emphasis on areas and clients): this includes information concerning the site of residence, the housing stock of that area, household/housing relationships within that area and employment.

2) Estimates of post-renewal conditions and costs: housing (clearance and rehabilitation), employment, institutions, and costs.

3) The consequences of withholding action: nature of the opportunity, alternatives to Title One renewal, neighborhood expectations, and nature of consequences and investments to date.

4) Relation to a newly emerging renewal framework: the demonstration cities program, and other programs and processes.

5) Relation to overall program balance: benefits to households, benefits to areas, both of a short and long term nature.

One attempting to conceptualize information systems design for this and other comprehensive planning purposes finds various substantive categories around which an information system design can be constructed.[4]

The information for operating this aspect of the overall program could be assigned the following attributes:

1) Population types and groups.

2) Population "needs" (e.g., employment, education, housing, recreation, etc. etc.)

3) Programs (Federal, state and local; employment, recreation, education, housing, etc.)

4) Sources of resources to accomplish programs (public sector; private sector; other).

5) Population activities.

6) Areal characteristics.

7) Time periods.

Taking information as it is produced, or raw data, as a source input to the system, and assigning the above attributes, is one example of a type of an n-dimensional categorization system design. Given the nature of this, or any other categorization scheme, there are experiences (discussed below) of information system designers which offer pertinent advice to the system under ultimate consideration here.

2.2 Example 2. Program Budgeting in New York City.[5] Evaluation Aspects

The top administrator of New York City Planning Department, together with the City Administrator's Office and the City Budget Bureau, have been requested by Mayor Lindsay to serve on a new Policy Planning Council which is to be responsible for developing integrated planning and budgeting procedures now scattered among several city agencies. One of its primary missions will be to implement the Planning, Programming and Budgeting concept recently given such emphasis at the Federal level.

"Program budgeting" is designed to provide an approach to the need for more sophisticated handling of the resource allocation process. It focuses on the decision-making process, particularly on the problems of data and analysis. Its first effort is simply the rational ordering of inputs and outputs, in which the initial emphasis is on the identifiable outputs -- major objectives of government processes. It then attempts to order the inputs -- government activities produced by manpower, material, real estate -- so that comparisons among wide ranges of alternatives are feasible and meaningful.

Program budgeting starts with the structuring of the problem and ends with analysis of the data. Among the analytical tools, cost-benefit or cost-utility analysis that compares benefit or utility (outputs) with resources or costs (inputs) is a most prominent one. Since the objective is to improve the decision-making that occurs in real life, and not in the philosopher's fancy, program budgeting pays special attention to questions of organization and administration, and the politics and pressures of the workaday world.

Implicit in the suggestion that program budgeting be employed in New York City is that New York City government can, in fact, determine its policies most effectively if it chooses rationally among alternative courses of action, with as full knowledge as possible of the implications of those alternatives. The requirement of choice is imposed on it by the fact that the city government is limited by the scarcity of resources. The City must choose not only among various courses of government action, but also between its total program and programs provided by other units and levels of government. The primary purpose of program budgeting is to facilitate the making of these difficult choices.

Planning, programming, and budgeting is the focus of the process of comparison and coordination. In line with the above it involves:

— Appraisals and comparisons of various government activities in terms of their contributions to city objectives.

- Determination of how given objectives can be attained with minimum expenditure of resources

- Projection of government activities over an adequate time horizon.

- Comparison of the relative contribution of private, civic and public activities to the city's objectives.

- Revisions of objectives, programs, and budgets in the light of experience and changing circumstances.

These operations are inherent in any planning, programming, and budgetary process. Program budgeting involves more explicit recognition of the need to perform them than has been traditional. It also involves the application of new analytical techniques as an aid to the exercise of the human judgment on which choices must ultimately rest.

To date, the information requirements for the application of program budgeting are met on an ad hoc basis; no routine set of benefit and cost series or "needs" data is developed within any on-going New York City department. The challenge here is to reduce the degree to which ad hoc practices hold sway, which in turn will ensure the perpetuation of sound program budgeting practices.

2.3 Example 3. Physical Planning for the Development of Urban Poverty Areas: Economic Development for Areas

Plans have been recently offered to help rebuild the physical situation of urban ghettos and regenerate the lives of their inhabitants. Privately financed development corporations have been advanced, for example, to build "adequate housing, parks, and recreation areas, hospitals, demonstration discount stores..."[6] Federal agencies are very involved, also, with more than $1.0 billion devoted in the fiscal year 1967 to the government trying to find more jobs for minorities; Manpower Development and Training, Labor Department, providing on-the-job training for the unemployed and underemployed; the Job Corps and Neighborhood Youth Corps, Office of Economic Opportunity, providing education and work for youth in Corps Centers, or part-time jobs and literacy training; and Work Experience and Training, and Community

Work and Training, Health, Education and Welfare Department, "to provide skills to family heads through vocational training combined with public welfare assistance."[7]

Another ambitious but potentially far-reaching effort is that to be undertaken under the Public Works and Economic Development Act of 1965, with amendments under consideration to assist in making the Act applicable to "urban poverty areas."[8]

The areal boundaries have to, in some way, relate to the "project areas" as defined under the various relevant legislation. The area boundaries as defined in that legislation; the social economic characteristics of the areas under consideration; and the relationship of the proposed project and project area to larger projects and larger project areas: these all have to be considered, in designing economic development projects for urban poverty areas.

It is most important that the public works components of these programs correct and/or address impediments to economic development as these are identified for the urban poverty areas in question. "Low income and high unemployment" (which do not necessarily go together) are concepts too abstract. For immediate implementation, the various types of urban support services, whether these be for health, transportation, education, communications, etc., all have to be "tailor made" to overcome the "impediments to growth" in the overall city and/or metropolitan region -- but be suited to each such area differently.

The development impact of each of the "urban support service" types of public works, i.e., impact upon the urban poverty area and impact upon the surrounding areas, should be estimated with respect to: increasing the permanent full time employment; increasing income level of the residents; improving economic development planning; and, finally, having a demonstration value to other areas.

Analyses of various types have indicated that one concept well-suited for the urban poverty area may be that of a "staging area." People who immigrate from other than urban areas, such as agricultural areas, must have facilities in which to develop and from which opportunities can best be experienced. The present "ghettos" can conceivably, with substantial modification, become successful staging areas for the experiences necessary to "urbanize" non-urban families and individuals. There are significant facilities' implications for converting urban poverty areas to successful staging areas. These will take many planning and engineering resources to implement. It is extremely important that these urban poverty areas properly be seen as systems through which various flows take place. It is also most important to understand these urban poverty areas as subsystems, because, for many purposes, they can be separated from the rest; however, unless the concepts of flows of population and activities through these areas are understood, only disappointment will come from improperly developed measures of changes in the poverty areas themselves.

3. ENVIRONMENT FOR THE USE OF ANALYTIC AIDS: PROPER INFORMATION BASES FOR COMPREHENSIVE PLANNING[9]

The experienced managers of New York City recognize that it is highly unlikely that some unique course of action will "solve" the transportation or economic development or housing problems. New and existing technology must be adopted to transportation, economic development and housing purposes (to mention a few), with greater or lesser effectiveness, depending upon the capability for understanding the changing interaction of the transportation, land use and housing systems with the economy in a society. It is recognized that improvement in urban transportation and housing systems will depend upon an understanding of the consequences for the whole system of changes in parts of the system; upon an understanding of administrative and political processes involved in implementing research funding; and upon an ability to act on the basis of these understandings.

It is also increasingly recognized that it is possible to determine the probabilities of change through large-scale comprehensive analyses which treat the aggregative transportation and housing land use systems as a single system and which consider the system's relationships to its environment. Techniques of this complexity have been possible only in recent years by the introduction of high speed electronic computers and corresponding methods of analysis.

But there are also large deficiencies, both on the part of the analytic methods and in the decision environment which makes for difficulties in application of proper informational and analytic aids to transportation, housing, land-use and economic development problems in the "real world" of New York City.

Some of the environmental problems are more or less obvious:

1) Needed Levels of Resources and Understanding -- With some exceptions, local and state governments throughout the country are deficient in the capability to understand the interaction processes involved. New York City is only recently becoming an exception, and it is taking rapid strides toward meeting the organizational and personnel requirements needed. It is vitally important that executive leadership in local and state government recognize the growing urgency of the need to develop more explicit, rigorous methodology and innovate a spirit for analyzing problems and identifying analytic opportunities.

2) Question-asking Needs -- "Systems analysis" is a method of answering questions posed by comprehensive planning. The answers it gives are only as good as the questions it poses. The wisdom and understanding of those in decision-making roles is therefore the critical element. In the context again of the problems of New York City, the ability to "formulate the questions" is a function of the degree of success in reorganizing the almost one-hundred separate organizations into a more functional and systematic whole.

3) Exogenous Forces: The City and its Surroundings -- Public policy is one of the most significant factors in the development of an efficient operation of New York City's transportation, housing and land use system. These policies determine the role which the units of the various levels of government play in (a) investment and facilities, (b) operation of facilities, (c) levying of charges on users, (d) stimulation of innovation and efficiency through incentives, and (e) establishment of a coordinated and focused program in research. Nevertheless, the formulation of even a comprehensive machinery to assist in urban decision-making would not affect more than a limited percent of the total dollars being invested or expense in the community in any one year. For example, in New York City in 1963-1964, the expense budget was $3 billion and the capital budget was $1 billion. These figures may be compared to 1963 figures of $700, 000, 000 for new privately owned residence buildings, and $100, 000, 000 for commercial structures completed and $100, 000, 000 for public buildings completed. As for wealth figures, it has been estimated that the City of New York's physical assets are valued at between $75 to $100 billion, while the total public capital investment since World War II has been estimated to be slightly less than $10 billion (in 1963 prices).

Another "exogenous force" is the region within which the City is such an important subsystem. Providing an adequate research base for rational municipal policy making will not in itself insure rational municipal policies. For example, probably the most significant portions of transportation policies affecting New York City are made outside of the city administration. Various authorities of at least three state governments and a variety of Federal agencies all play critical roles.

4) The Scope and Detail of Public Policies -- The sheer number and diversity of Federal, state and local programs in any one functional area, such as housing policies and programs, almost defy numerical description. There are more than 100 truly definable "housing policies" which exist as alternatives to the New York City planner. Obviously,

the combinations and permutations of playing these 107 program alternatives against the thousands of potential areas, the hundreds of definable client groups, the numerous time period alternatives, etc., are enormous in number. As a result, the criteria currently being used to determine the amount and mix of expenditures in each of the major housing programs are multiple in value and in turn this makes difficult the job of deciding how much to allocate between the various types of housing, or between various types of urban renewal.

5) The Diversity Factor -- The standards or planning factors which could be used to determine the "needs of the communities," for example, "the cost per dislocatee" or "the number of acres of parks," or "the size of community facilities," etc., are extremely diverse in a city the size of New York. The variance within any one identifiable functional area (e.g., educational facilities) may be as great as the variance between the different types of community facilties.

6) Multiple Objectives and Consequences -- It is not a simple matter to define, let alone measure, the transportation functions between inputs for housing, land use and transportation programs and resulting impacts/output measures. These outputs, effects, or impacts are likely to be of a broadly environmental health nature such as air purification, water purification, city beautification, and the like.

7) Private-Public Relationships -- The private sector-public sector tradeoffs are most unique and telling at the decision level of a city. Alternative incentives to induce private industry to produce low income housing, for example, may be translated into urban policies involving zoning, mapping, industrial clusters, tax policy, etc. The choices here are also enormous, but the close relationships between what happens at the public level and the consequent effects for the private, or vice-versa, are poignantly felt on a moment-to-moment basis at the urban level. The private-public relationships involving social costs, such as the cost of dislocating a business or a family, are also extremely poignant at the urban level. The closeness between the private and public sectors at the urban level causes severe problems for measuring social cost, as is required by sophisticated PPBS methods.

The above are only some of the environmental factors which make detailed and sophisticated application of analytically based information systems difficult at the urban level. While it is difficult to rank these in importance, different urban areas will find these problems more or less severe depending upon circumstances; they also clearly do not make application of analytic methodology and information systems design at urban levels impossible.

4. RELEVANT EXPERIENCE OF OTHER URBAN SYSTEMS DESIGN AND ANALYTIC EFFORTS

Several analytic and information systems design approaches have been attempted, in different ways in different urban and regional areas, throughout the United States; in particular, efforts in New York, Philadelphia, Pittsburgh, and San Francisco, are to be highlighted from an analytic and/or model building standpoint; and Alexandria, Los Angeles, New York, Pittsburgh, Santa Clara, Tulsa, and several others from the "data bank" standpoint.

The New York City Planning System's design effort has benefitted from the experience of these different and widely varying information systems-based applications in urban areas:

4.1 The "Data Bank" Experience --

The "data bank" efforts have had widely differing degrees of success, but all appear to have required a good deal more time for implementation for decision-making than had been initially anticipated. Some of the systems, several years beyond the date originally anticipated, have begun to experience moderate use. One of the fruits of this lesson for New York City is that while there is a need for integrating operating data and planning department data into a fully integrated system eventually, in a large metropolitan area (and in particular in the City of New York), the size of each of the departments is such that each should not and cannot be handled as strictly "black boxes" in some ultimate futuristic design. The innards of the "black boxes" (e.g., the City Planning Department, the Police Department, the Welfare Department, etc.) must be fundamentally understood as subsystems that they are, but at a significant level of detail. In the City of New York, for example, considerable study is taking place within a number of different city departments to better understand the functioning of these, while at the same time some overall system design is being contemplated. Serious errors can be made if the internal workings of important urban planning functions (e.g., planning, capital budget, etc.) are not clearly understood while the overall system design is also being undertaken.

4.2 The City vs. The Region --

While dealing with and understanding metropolitan area problems and processes and structure is very important, modelling and/or analytic efforts devoted to better understanding the CITY is equally important and has not been paid equal attention. Furthermore, analytic models for cities can be quite different from those constructed for metropolitan areas: in particular, the relationships between space, housing, amenities, and transportation require different structural concepts induced, as it were, by the vast differences in densities between the city and the outlying areas.

4.3 Measuring Social Cost --

The construction of measures of social cost is very difficult at the urban level and has been accomplished, at best, very imperfectly. In various analytic efforts, for example, the measurement of "dislocation," as an index, is an attempt to be analytic; it does introduce needed complexity into the measures of relocating people and/or businesses. These indices of dislocation costs also introduce some new qualitative dimensions into the measurements of relocation by raising the notion of different cost weights to be attributed to various client attributes. In a sense, therefore, these index numbers attempt to measure dislocation social costs and attempt to introduce more "realism." Nevertheless, it is most important to realize, in constructing such an informational index, that there are many problems involved in combining the different attributes into an index, or weighting of the various costs and benefits to each client group. At the urban level, there are a large number of reasons for difficulties in constructing social cost indices, in particular where the weight assigned to one characteristic or group might represent an individual person or entity who can have a loud and clear political voice in the urban area.

4.4 "Alternatives" Finding --

A way has to be found better than now exists, to define, parsimoniously, the mixes of alternatives which face urban managers. The method should exhaust the wide range of considerations required. There is obviously no shortage of possible programs to correct or ameliorate the problems of urban areas. What is needed is a proper management over which alternatives are examined. This speaks to the need for the social strategy for experimental design, including the urban managers (including planners) as well as the analysts.

4.5 Classification of Criteria --

Desperately needed are classifications of criteria which can begin to allow us to sort program budgets into dollars for various end state requirements. Possible classes of criteria include:

a) functional budget allocation criteria, (i.e., within or between transportation, housing, public safety, etc.) An example of such a criteria might be "a balanced transportation system" or "integrated housing";

b) urban/regional budget allocation (e.g., the proportion of dollars devoted to basic maintenance of city facilities versus dollars devoted to maintenance of "the impoverished," versus dollars to "develop all the peoples," etc.);

c) subject area classification (e.g., social, physical, fiscal, aesthetic, economic, etc.);

d) absolute vs. relative criteria (e.g., "nobody in essence has less than $2,000 to spend per family spending unit");

e) effects classification (e.g., users versus non-users and within the latter, completely vs. partially collective, etc.).

4.6 Experience in Use of Models --

With regard to the analytic models and in particular the simulation exercises for city planning over the past 5 years, many opposing points of view as between model builders vs. planners vs. urban managers have evolved.[10] These differences range from views about people to views about the theory of urban places to views of the objectives for which the city exists. One important lesson that can be learned from a review of the modelling efforts to date is that the "objective functions" developed do not appear to be close to the heuristic decision process of urban managers/planners, though they capture some of the stated objectives (perhaps the most important ones) of knowledgeable individuals in the land use/housing/transportation fields. Furthermore, the method of maximizing, per the objective function approach, does not appear to satisfy the subtleties of the planner's heuristic decision process. Nevertheless, an explicit allocation process of resources is required at some point in this process. The models have pointed to methods for (potentially) improving the system for allocating resources. Finally, urban models have been more successful in assisting the decision-making processes of the long-range managers of the urban environment (the city planners for the most part), as opposed to short-run urban management of operating department functions.

5. WORK PROGRAM FOR INFORMATION SYSTEMS DESIGN AND ANALYTIC SERVICES

The following is a brief description of the work assignments required to implement the design effort. The assignments have been organized under four general headings - Information Systems Design, Special Analytic and ADP Services, Model Building and Programming, and Staff Seminar Programs.

5.1 Information System Design

This part of the study provides an "impartial evaluation" of the more prominent information needs of the Department, together with an analysis of the way in which data are acquired, processed, communicated, and stored with various units in the Department. The focus of the study has been on those records and documents that are relevant to the major planning and decision-making functions of the Department and the Commission.

The study has been conducted in several overlapping phases:

a. A reconnaissance of records maintenance activities within the Department.

b. A survey of frequently used research materials from all levels of government and including census, buildings, housing, employment, land use, assessment and other data, to determine best methods for their organizations.

c. A survey of information on the status and availability of machine readable data in a selected number of other City departments, concentrating on those departments within the City administration which have records relevant to the Department's work program and which are in machine-readable form (presumably punch cards or paper tape, magnetic tape, or microfilm).

d. Assisting the Department in becoming more acquainted with the ongoing data collection and processing activities of selected outside agencies whose work is closely allied to the Department's comprehensive planning program.

The intention here is to make possible the coordination of work on the design of the City Planning Department's own information system with relevant work being carried out by other agencies and organizations in the City.

Based on the findings of the foregoing studies, recommendations are being made for the ways that the Department should organize the data handling activities, including proposals for an efficient, economical system capable of integrating both manual and ADP procedures, recommended organizational structure, equipment and manpower requirements, and costs.

Included in the products required to implement these information system design recommendations are:

a. Specifications for an initial records management control program. Included in this task has been the determination of such pertinent factors as location, timing, total cost, required staff, equipment and techniques, and operating standards.

b. A System Description Document which describes an overall department of city planning information system, the objective of the document being to specify a system which, when implemented, will provide the most effective information support to the Department's information requirements as is possible within the state of the art in information handling technology under practical budgetary restrictions. The Document also provides a primary input to the Systems Development Approach Plan and is also useful in space, personnel, and budgetary planning. It also facilitates the coordination of data interchange with outside agencies.

c. A System Development Approach Plan which specifies a number of progressive, mutually supporting information system implemenation steps which can be implemented to improve current Department information handling practices. Each step is chosen in a manner which most practically allows the Department to progress towards more effective information handling. Considerations here include budgetary limitations, personnel availability, training requirements, and improvement priorities. The approach plan is also developed so as to minimize disruption of current operations.

d. Operational programs which constitute a portion of the first step towards more effective information system design. The Target I Model (see below) and library programs are examples of these operational programs.

5.2 Special Analytic and ADP Services

The Department required special analytic and automatic data processing services to assist in the completion of projects underway in the work program of the Department. These services have included the preparation of computer programs for analyses of existing data, and assisting the Department in developing the types of analytic tools required for various planning operations. To the extent feasible, programs already developed have been adapted to the Department's planning needs.

The work has consisted of the following elements:

a. Acquisitions - of unit records (punch cards) and tape files; conversion to required formats as may be necessary; provisions of such computer routines and programs as are deemed necessary; card-to-tape conversions of Department material, as has been specified.

b. Library Programs - to be used for statistical analyses and related operations as feasible.

5.3 Model Design and Programming

Demographic models for population and housing analysis have been the primary analytic computer programs provided to the Department by the consultant. These models (labeled "Target I models") are for the purpose of providing a 1965-1966 inter-censal basis for current analysis; it is intended that greater refinement will eventually mean that longer time horizons can be forecast.

a. A validated, operating computer model is being developed for producing information, by small area (e.g., New York City health area) for the year 1965 and for subsequent years for New York City. Such information has included population by five year age cohorts by sex and race; household and family characteristics by age of head, size, race, and income; and specified housing characteristics.

b. A computer model is also being developed for the purpose of measuring the performance on a city-wide and small area basis of various housing programs, governed by alternative operating standards and program budgets. To the extent possible, this model is being developed such that it is sensitive to renewal programs affecting rent structure, housing value, housing condition, new or rehabilitated housing for households of various incomes, housing occupancy, housing demolition, and household displacement.

5.4 Staff Seminar Program

A series of seminars has been developed and conducted for various levels of professional personnel in the Department to provide an introduction to new concepts and technologies important to planning. Topics covered include the following:

Introduction to Concepts of Systems
Management Systems and the Organization
Prospects for Problem Solving
Introduction to Computer Technology
Systems Analysis vs. Operations Research
Information Systems for Management and Research
An Introduction to Gaming and Analytical Modeling
Management Control Tools for Use in Planning Agencies

6. MEASURES OF EFFECTIVENESS AND SYSTEM APPRAISAL

6.1 Possibilities of Developing Measures

Among the real problems facing information systems designers today is the lack of well defined measures to judge how well the ongoing information system -- whatever that may be -- is performing: let alone what any "improved" system design might produce in the way of effectiveness. There are many difficulties in evaluating information systems, including: the lack of well defined objectives; the lack of meaningful models; the lack of meaningful measures; the difficulties in determining the scope of the system; the difficulties in determining test problems; lack of motivation because poor systems design is not "punished"; and lack of quantitative evaluation criteria. In reality, there is still no basis for using the word "science" in "information systems science," since there are not yet any clear ideas about how to characterize information systems, or how to measure their effectiveness in any but the restricted sense of mechanical computation.[11]

Others have been more sanguine about the possibilities for measuring effectiveness. Those who have examined the prospects of measurement have agreed that any discussion of the role of information in a purposeful activity should be based upon some explicit conception of the character of that activity. Therefore, it is extremely important that the utility of having some commonly agreed to framework be a part of the overall strategy for evaluating an information system design.[12]

6.2 System Appraisal

Designers of information systems have in the past been called upon to define the functions, inadequacies, and requirements of the organizations for which the system was to be developed.[13] Experience has proved that external criticism and analysis of these functions, inadequacies, and requirements is unlikely to be well-directed, and, in fact, measures of performance of the information system, it appears, must be based upon intensive internal criticism performed almost exclusively by in-house personnel and only sparingly assisted by outsiders. It is imperative that measures of performance be intimately designed through internal criticism and the mechanisms for achieving this criticism.

7. CONCLUSIONS

The implications of the above is that there is a pertinent body of experience which can be applied to the information systems design problem, such as that of the Department of City Planning of New York City. The experience is perhaps more at the qualitative level than at the quantitative, but also indicated is the great importance of the latter. Experience has also indicated that information systems design for intelligence, research and development functions in organizations is very difficult but that an evolutionary qualitative-quantitative approach can produce significant advances in information system design. Also indicated is the need for agreement between the users' own models of their organization and functions, and the models used by the system designers in establishing the overall approach to systems design. Furthermore, it is very useful to attempt to establish measures of the effectiveness of information system usage in the current system, so as to help diagnose the ills of the system and point the way for certain short-run improvements as well as begin to establish measures which can be used in an ultimate system.

FOOTNOTES

1. New York City - Department of City Planning, "Planning Program and Organization," June 16, 1965.

2. New York City - Department of City Planning and CONSAD Research Corporation, "Work Program for Information Systems Design and Analytical Services," September 7, 1965.

3. More specifically, within this Office, the Division of Renewal and Development.

4. These ideas are adapted from the memorandum to the New York Community Renewal Program staff from N. Douglas, "Background -- 6 Months Work Program," August 26, 1965.

5. For a fuller elaboration of this topic, see W. A. Steger, "Planning, Programming and Budgeting Systems as Seen from the Local Level," talk presented at the Washington Operations Research Council Meetings, Session #2, June 1966.

6. "Kennedy Offers Plan on Ghettos," _The New York Times,_ May 30, 1966.

7. "New Drive to 'Fulfill These Rights,'" _Business Week,_ May 28, 1966, pp. 38-39.

8. "Aid for the Cities," _The New York Times_, May 25, 1966, p. 30.

9. This Section and Section 4 are substantial modifications of W. A. Steger, "Planning, Programming and Budgeting Systems as Seen from the Local Level," _op. cit._

10. W. A. Steger, "The Realities of Simulation Exercises for City Planning," paper presented before the American Statistical Association, March 3, 1966 (N. Y. Chapter).

11. Walter M. Kraise, "The Test of Theory in the Real World," _Proceedings of the Second Congress, Information Systems Sciences,_ Joseph Spiegel and Donald Walker (ed.), Spartan Books, Washington, D. C., 1965, p. 307.

12. Martin L. Ernst, "Evaluation of Performance of Large Information Retrieval Systems," Information Systems Sciences, ibid, pp. 239-240.

13. R. N. Davis, "Classification and Evaluation of Information System Design Techniques," Information Systems Sciences, ibid, pp. 81-82.

REFERENCES

Davis, R. N., "Classification and Evaluation of Information System Design Techniques," Proceedings of the Second Congress, Information Systems Sciences, Joseph Spiegel and Donald Walker (ed.), Spartan Books, Washington, D. C., 1965, pp. 81-82.

Douglas, N., "Background -- 6 Months Work Program," Memorandum to New York Community Renewal Program staff, August 26, 1965.

Ernst, Martin L., "Evaluation of Performance of Large Information Retrieval Systems," Information Systems Sciences, ibid, pp. 239-240.

Kraise, Walter M., "The Test of Theory in the Real World," Information Systems Sciences, ibid, p. 307.

New York City - Department of City Planning, New York's Renewal Strategy/1965, December 1965.

New York City - Department of City Planning, "Planning Program and Organization," June 16, 1965.

New York City - Department of City Planning and CONSAD Research Corporation, "Work Program for Information Systems Design and Analytical Services," September 7, 1965

Steger, W. A., "Planning, Programming and Budgeting Systems as Seen from the Local Level," talk presented at the Washington Operations Research Council Meetings, Session #2, June 1966.

Steger, W. A., "The Realities of Simulation Exercises for City Planning," paper presented before the American Statistical Association, March 3, 1966 (New York Chapter).

A SYSTEMATIC APPROACH TO METROPOLITAN ANALYSIS: LESSONS FROM A STUDY EMPLOYING CENSUS-TRACT DATA

Donald L. Foley*

California has become overwhelmingly metropolitan; nearly 9 out of every 10 residents live in one of the state's dozen-odd metropolitan areas. An understanding of these areas' development patterns is crucial, and, effectively approached, should contribute simultaneously to (a) the descriptive knowledge about each metropolitan area as a distinctive urban unit, (b) the comparative analysis on an intermetropolitan basis, and (c) an overall or conglomerate view of metropolitan California.

In a recent study for the California State Office of Planning, resulting in a two-volume publication, Characteristics of Metropolitan Growth in California [2, esp. chap. 4] , we sought to combine a substantive study of the demographic and housing characteristics of new residential development in several of the larger urbanized areas in California and a methodological exploration of the potentials and problems of fashioning a supporting analytically oriented data system. In this research we deliberately started with, and sought to take optimal advantage of, census-assembled census-tract data available in computer-tape form from the 1960 census.

In this paper we selectively report on this experience, organizing our presentation under three headings: Design of the study, examples of substantive output, and implications for data-system development.

1. DESIGN OF THE STUDY

The 1960 Census marked a very significant expansion in the availability of small-area data at the census tract level. All of California's then 10 Standard Metropolitan Statistical Areas (SMSAs) were tracted in their entirety, most for the first time. In addition, the entire remaining nonmetropolitan portions of California were treated as "pseudo tracts," with data fully comparable to that for census tracts compiled for minor census divisions and most municipalities[6]. The data released on magnetic tape exceeded even that available in published form [4,5].

Since the census-tract data pertained exclusively to population and housing it could best be used in studying residential areas and since for the first time it provided detailed breakdowns for suburban and outlying districts where new growth had been particularly concentrated, we elected to focus on the population and housing characteristics of this recent development. With the full range of data also

*University of California, Berkeley

available for older residential districts, we could, if we chose, feature comparisions between newer and older residential districts. Since the data were largely aggregated at the census-tract level, we were forced to accept this and thus to accept the fundamental idea that we were getting profile views of entire census tracts. Some disaggregation was provided, however, and we sought to take the fullest possible advantage; we shall note below our use of separate statistics for nonwhite and for Spanish-surname white population.

Our analytic strategy led us (1) to maintain the maximum disaggregation possible where it kept open the possibilities of significant cross tabulations in depth; (2) to narrow the number of substantive variables by intuitive and pragmatic judgements as to the relevance of these variables to our proposed analysis. Overall, we pursued an approach that can perhaps best be termed successive subclassification, leading to the deliberate creation of submatrixes. In turn, this both opened up relatively powerful analytic capabilities and also posed difficult computer programming problems as to storage and retrieval.

Let us summarize the data system we created. (We shall restrict our presentation here to strategic considerations and will avoid the temptation to report practical, operating problems which affected so much of our actual work.) We adapted the census-tract data already in its own format, but in so doing we facilitated the use of the data for comparative analytic purposes. Here are the main dimensions of our system:

A. <u>Census tract.</u>

Basic unit, by definition. We treated the separate parts of split tracts (i.e., split to identify parts in different local political jurisdication) as the equivalent of separate census tracts. We eliminated census tracts containing only data for population on vessels at sea and eliminated census tracts with less than 40 occupied housing units or with 10% or more of population in group quarters. When working with minority-group populations, we admitted census tracts only if they had at least 20 housing units occupied by that group.

B. <u>Urbanized area</u>.

We excluded census tracts outside the urbanized area in order to eliminate the sections of the larger SMSA mainly agricultural and non-suburban. We decided to work with the following six urbanized areas, with these populations and numbers of census tracts included:

	1960 Population	Census Tracts
Los Angeles-Long Beach	6,488,800	1,377
San Francisco-Oakland	2,430,700	570
San Diego	836,200	211
San Jose	602,800	157
Sacramento	451,900	77
Fresno	213,400	70

These were the state's six largest urbanized areas (except that San Bernardino-Riverside urbanized area, with population of 377,500, was not included).

C. Index of Recency.

The mean of two percentages, % housing units built 1950-60 and % housing units built 1940-60. It ranges from 0 (if all housing units built before 1940) to 100 (if all housing units built 1950-60), and has a value of approximately 50 for the state as a whole. The use of this index gets around the non-availability of 1950 data for most of the census tracts employed by introducing a measure of recent growth available from the 1960 data alone. We mainly depended upon five Recency categories corresponding to the Index values 0-19.9, 20.0-39.9, 40.0-59.9, 60.0-79.9, and 80.0-100.0, as providing a working range from Earliest Development to Most Recent Development.

D. Composite Index of Economic Level.

The weighted mean of the percentages of owner-occupied and renter-occupied housing units which are above their respective medians (for California as a whole). The weighting takes into account the proportion of housing units that are owner or renter-occupied. The index ranges from 0 (if all housing units are below the medians for the state) to 100 (if all housing units are above the medians); the index has a value of about 50 for the state. As with the Index of Recency, we relied upon five equal-interval categories. We particularly made use of 0-19.9 as low level of home value-rental, 40.0-59.9 as medium level of home value-rental, and 80.0-100.0 as high level of home value-rental.

E. Air distance from center.

The straight-line distance from the centroid of the census tract in question to the centroid of the census tract best representing the heart of the main central business district in the urbanized area. (An exception was made in the Bay Area where distances were to the San Francisco central business district for Marin, San Francisco and San Mateo counties and to the Oakland central business district for Alameda, Contra Costa and Solano counties.) These were calculated by computer from latitude-longitude centroids, census tract by census tract, obtained on punch cards from the Bureau of the Census.

F. Central city vs. urban fringe.

Each census tract or split tract was identifiable as in the central city of the urbanized area (as designated by the name of the urbanized area) or in the remaining outer area.

G. Additional geographic identification descriptors.

Codes representing minor civil division, place size, Congressional district, latitude and longitude of the tract centroid, etc.

H. Population component.

As provided by the census: total population, nonwhite population (derived by subtracting the sum of non-white and Spanish-surname populations from total population) and white population (subtracting nonwhite population from total population). For most, but unfortunately not all, of the variables under I below, there is complete disaggregation by these components. On the magnetic tapes this is available for all census tracts; in published census tract reports this is available only for tracts with relatively large minority-group populations.

I. About 50 additional population and housing variables.

These have been selected from the even larger reservoir of data, tract by tract, on the basis of judged relevance to this particular study. For over 40 of these variables, we had access to or were able to derive the full disaggregation for the five components listed in H. In all cases except population or housing unit count per se, we calculated an appropriate ratio or percentage while also maintaining the original numerator and the denominator in case needed for aggregating statistics for combinations of census tracts.

In all cases we put variables and identifiers into the format required for acceptance by the analytic system, STATPAK, developed by the Computer Center on the Berkeley campus of the University of California and upon which we intended to rely. STATPAK provides packaged program components with capacities for carrying out statistical and graphic-display operations. [1]

The basic data matrix comprised a set of 967 words for each census tract. This, then, varied from well over a billion cells for the Los Angeles urbanized area to nearly 70,000 cells for the Fresno urbanized area. In terms of the dimensions we have just listed and described, this basic matrix was constructed as follows:

A (the census tracts in each urbanized area) by C, D, E, F, G, H, and I (the other variables or identifiers in a single listing, not themselves cross classified).

From this basic matrix we produced sequentially, as required, various submatrixes. The following are the most frequently used submatrixes:

1. A (the census tracts in each urbanized area) by C and I (variable values for the two Indexes plus about 50 population and housing variables) by D (three categories only--high, medium, and low--of the Composite Index of Economic Level).

2. A (the census tracts in each urbanized area) by C and I (variable values for the two Indexes plus about 40 population and housing variables disaggregated by minority-group status of population) by H (each of three minority-group population components).

3. A (the census tracts in each urbanized area) by C, D, and I (as in 1 above) by F (central city vs. urban fringe).

The STATPAK package programs employed had the capacity to carry cross classification further. By retrieving from the main matrix or selectively from designated submatrixes for any given urbanized area, we were able to introduce a further simultaneous cross classification of census tracts by each of the Index values and by air distance from metropolitan center. Thus, drawing upon submatrix 1, comprising a three-dimensional cross tabulation, we proceeded with STATPAK programs to cross tabulate selected I variables in sequence with C (Index of Recency categories), providing a fourth dimension to the cross classification. By introducing comparisons among urbanized areas, we moved to a fifth dimension of cross-classification. The power of such an analytic approach lies in the selectivity it provides us in holding constant various factors while examining the distribution or relationship of other variables on which we particularly wish to focus.

Since we were not only working with a rather large number of variables and with several urbanized areas each including census tracts numbering into the hundreds (and in the case of the Los Angeles urbanized area amounting to some 1,300 tracts), but wanting to produce an output by computer that would be interpretable without further graphic conversion, we particularly developed and came to rely upon a graphic "plot" subroutine in the STATPAK system. We shall show examples of this in the next section reporting substantive findings.

2. EXAMPLES OF SUBSTANTIVE OUTPUT

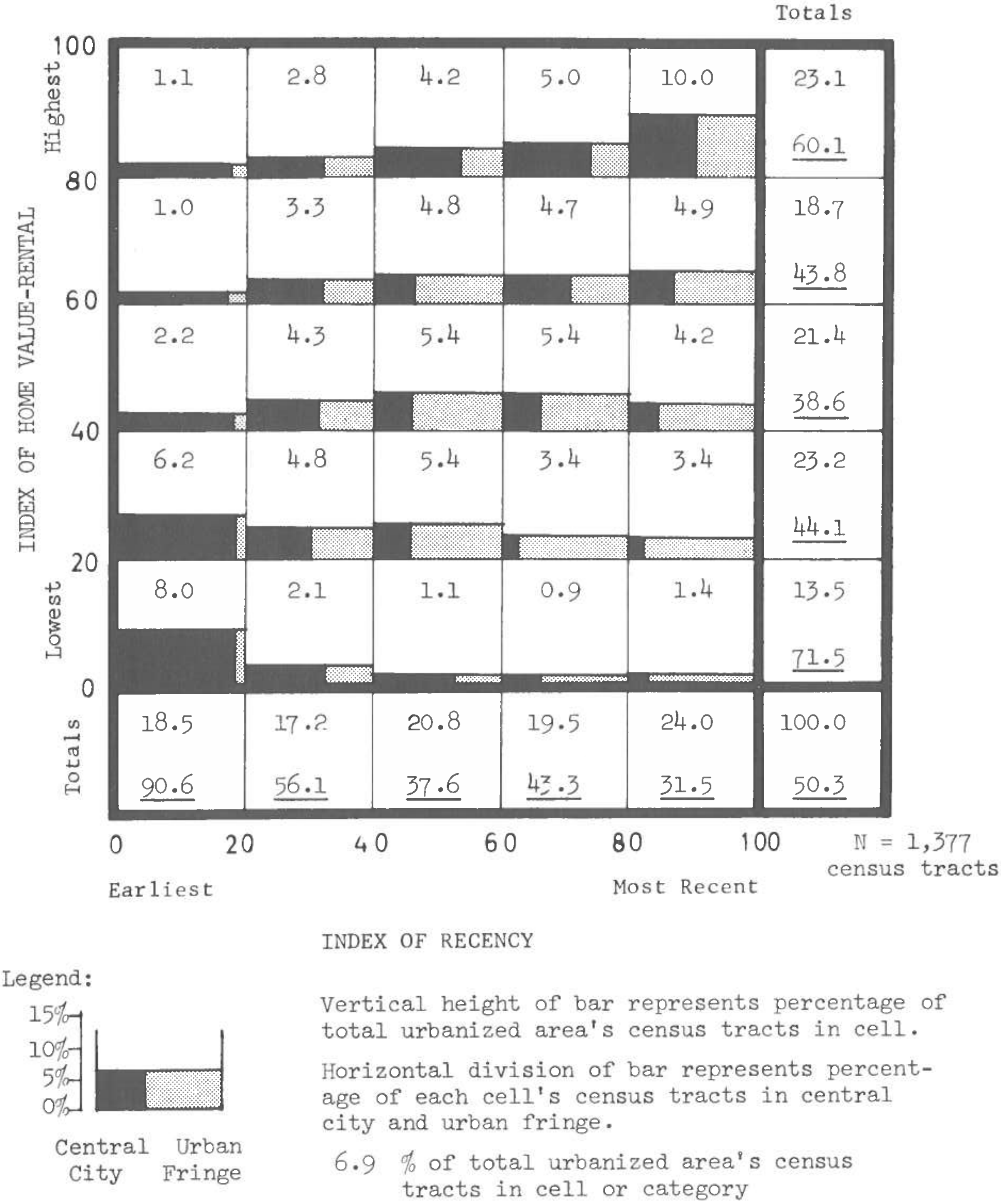
We had two general objectives in our substantive research probing: first, to provide a broad comparative look at the characteristics of recent residential development and, second, to examine the hypothesis that residential areas of recent development are particularly marked by a relative homogeneity of population and housing.

Reflecting the very nature of the census-tract statistics available we were essentially restricted to treating the census tracts as observational units; except for the disaggregation for nonwhite and Spanish-surname White population, we were not able to pull out more than average characteristics for each tract as a whole. To simplify our analysis we mainly resorted to counting the number of tracts with particular cross configurations of characteristics, thus treating the tracts in unweighted fashion. Our main device was to structure a two-dimensional typology and then plot or record the distribution of census tracts within the cells of this typology.

Figures 1 through 4 provide examples of such a two-dimensional

FIGURE 1

DISTRIBUTION OF CENSUS TRACTS BY RECENCY AND BY HOME VALUE-RENTAL
LOS ANGELES-LONG BEACH URBANIZED AREA, 1960



Vertical height of bar represents percentage of total urbanized area's census tracts in cell.

Horizontal division of bar represents percentage of each cell's census tracts in central city and urban fringe.

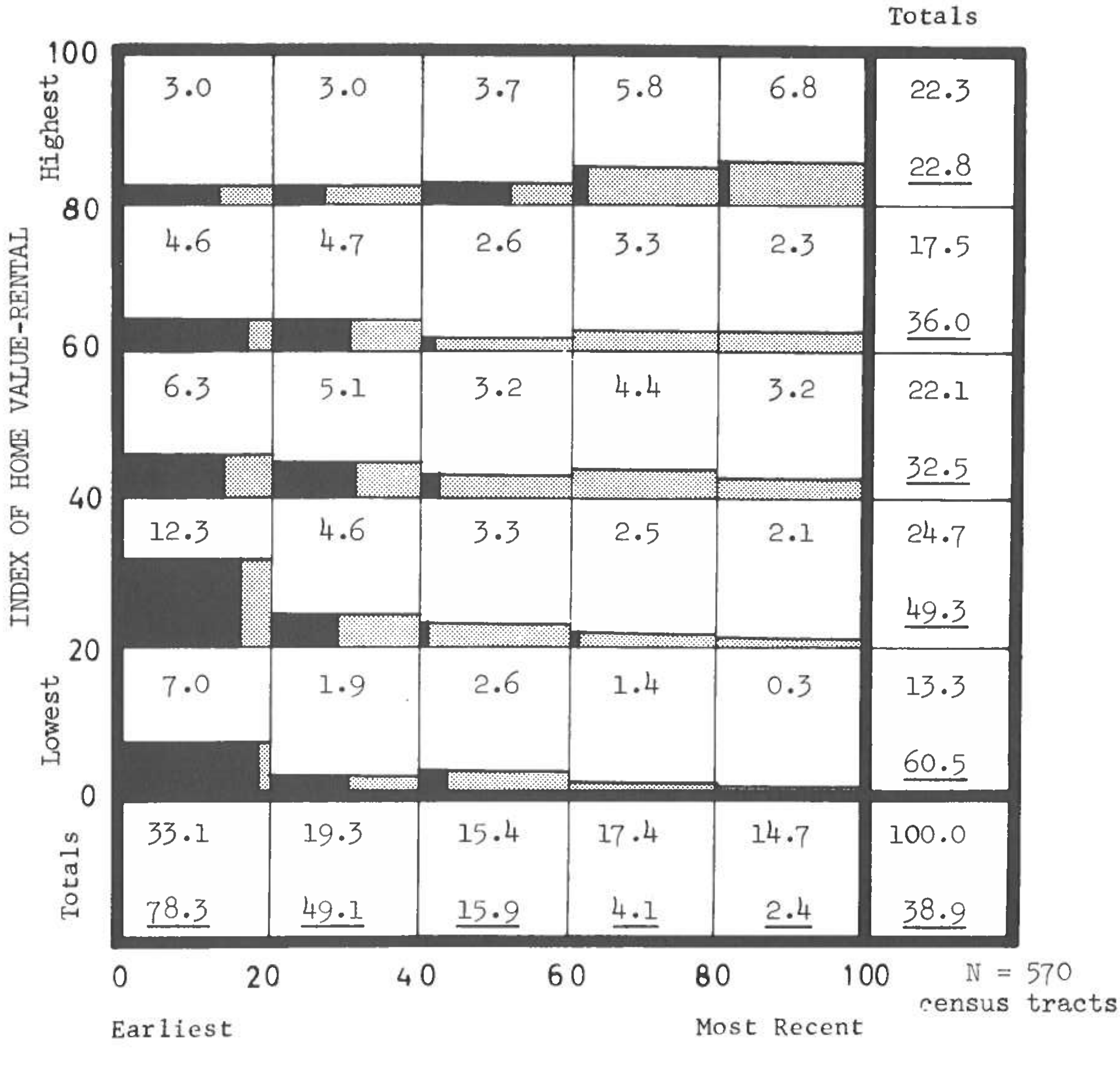
6.9 % of total urbanized area's census tracts in cell or category

42.0 % of each category's census tracts in central city

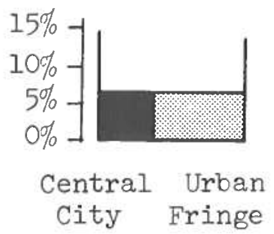
Source: Derived from U.S. Census of 1960 data tapes.

FIGURE 2

DISTRIBUTION OF CENSUS TRACTS BY RECENCY AND BY HOME VALUE-RENTAL, SAN FRANCISCO-OAKLAND URBANIZED AREA, 1960



Legend:



Vertical height of bar represents percentage of total urbanized area's census tracts in cell.

Horizontal division of bar represents percentage of each cell's census tracts in central city and in urban fringe.

6.9 % of total urbanized area's census tracts in cell or category

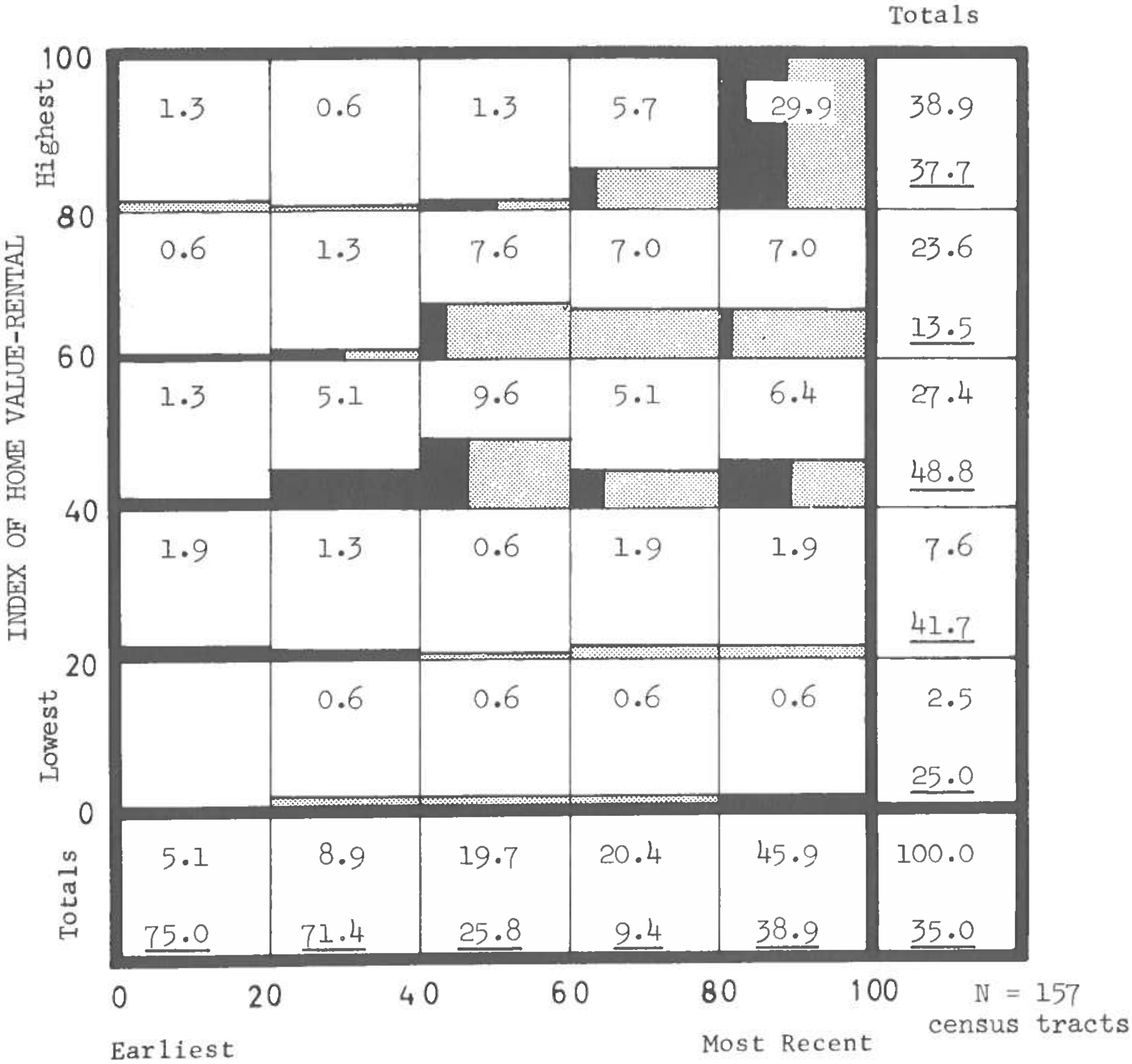
42.0 % of each category's census tracts in central city

Source: Derived from U.S. Census of 1960 data tapes.

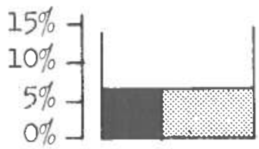
Note: Remainder of Contra Costa County added to Census' definition of San Francisco-Oakland urbanized area.

FIGURE 3

DISTRIBUTION OF CENSUS TRACTS BY RECENCY AND BY HOME VALUE-RENTAL, SAN JOSE URBANIZED AREA, 1960



Legend:



Vertical height of bar represents percentage of total urbanized area's census tracts in cell.

Horizontal division of bar represents percent- of each cell's census tracts in central city and in urban fringe.

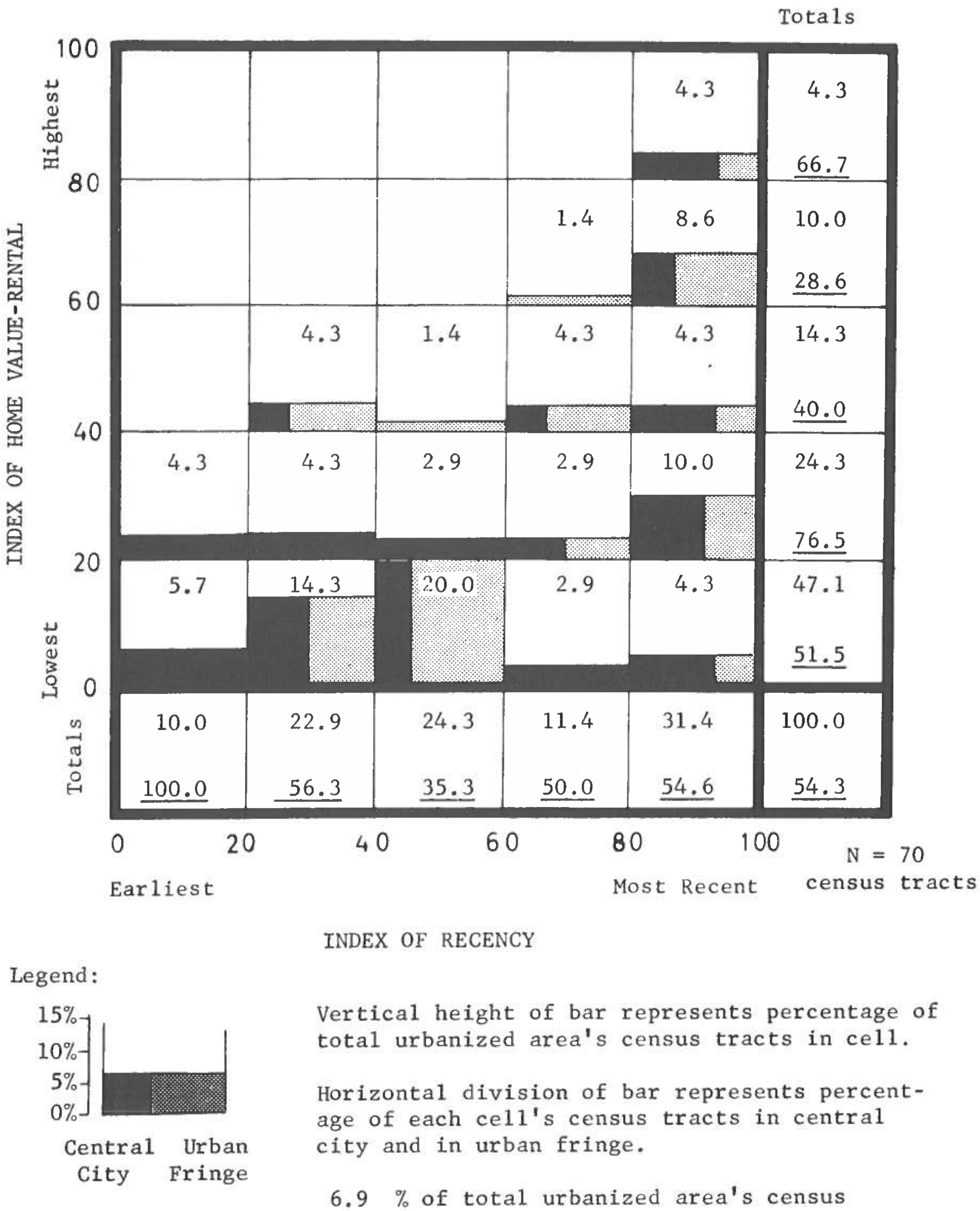
6.9 % of total urbanized area's census tracts in cell or category

<u>42.0</u> % of each category's census tracts in central city

Source: Derived from U.S. Census of 1960 data tapes.

FIGURE 4

DISTRIBUTION OF CENSUS TRACTS BY RECENCY AND BY HOME VALUE-RENTAL, FRESNO URBANIZED AREA, 1960



Vertical height of bar represents percentage of total urbanized area's census tracts in cell.

Horizontal division of bar represents percentage of each cell's census tracts in central city and in urban fringe.

6.9 % of total urbanized area's census tracts in cell or category

42.0 % of each category's census tracts in central city

Source: Derived from U.S. Census of 1960 data tapes.

typological approach in which the two Indexes already described comprise the dimensions; with a 5-by-5 matrix, 25 typological cells are thus created. The height of the bar in each cell is proportional to the number of census tracts for the urbanized area falling within that cell. A further division into two parts--the solid black representing the central city and the grey tone representing the urban fringe--shows their proportionate distribution within each cell of the typology. Fig. 2, for the San Francisco-Oakland urbanized area, demonstrates a metropolitan structure more nearly resembling that of an older Eastern urbanized area than is the case for other California urbanized areas. Note the relatively heavy distribution of census tracts of earliest development in the left tier and the sharp correspondence between central-city tracts and the left two tiers. There are scarcely any central-city tracts of preponderantly recent development in the two right-hand tiers. The Los Angeles-Long Beach urbanized area, as shown in Figure 1, is far more similar to the San Francisco-Oakland urbanized area than one might suspect from stereotyped comments featuring their differences. The Los Angeles-Long Beach urbanized area structure shows relatively less concentration of tracts of early development (the left tiers) and correspondingly greater concentration in the newer development categories (at the right). And, particularly significant, Los Angeles-Long Beach as the designated central cities have far greater representation in areas of recent development than is the case with the San Francisco Bay Area.

The San Jose urbanized area, shown in Fig. 3, provides a contrasting structure. It has relatively few tracts at lower economic levels (in the bottom two tiers) and of earliest development (the left-hand tier). Instead it has its tracts overwhelmingly concentrated in the recent development and relatively high economic status categories. With 30% of its tracts in the upper-right hand cell, this proved too great a concentration even to show to full proportionate height. Note the relatively widespread distribution of central-city census tracts throughout the framework. The Fresno urbanized area, in Fig. 4, by way of still further sharp contrast, has its census tracts heavily bunched at lower economic levels (the bottom tiers) and only in the right-hand tier of cells, representing very recent development, has any strong representation of high economic level development come into evidence.

It will be readily recognized that in this approach we have created social-area categories in the spirit of the Shevky-Bell approach [3], but that we have introduced somewhat different major dimensions. Like Shevky-Bell we have created a typology such that each cell of the typology represents a particular configuration with respect to two main dimensions, recency of residential development and economic level of the housing in the area. This permits us to concentrate on these social spaces, so to speak, and not be unduly concerned with the precise geographic pattern of settlement.

A data system such as that we have employed should also provide a ready base for computer mapping. Indeed, generated typological categories should themselves be readily mappable. We did not carry this phase of work through. Our report did, however, include a number of census-tract maps of the Fresno area so that one could view the geographic pattern as well as the more abstract typological one. Fig. 5 shows the percentage of each census tract's population that is majority-white in character (i.e., population other than Negro, other nonwhite or Mexican American). The large black area north of central business district (actually comprising a number of census tracts although these do not show through the solid black) represents the main area of new development and is overwhelmingly majority-white. In contrast, quite literally across the railroad tracks that parallel the main road through California's central valley (along the diagonal line through the map's center) is an area of seven census tracts shown in white in each of which less than a quarter of the population is majority white. Here are concentrated particularly Negroes, Oriental Americans and some Mexican Americans. Many Mexican Americans also live around the periphery of the urbanized area (that area represents only the middle quarter of the larger district shown on this map) in agricultural employment. Series of maps at an identical scale quite naturally encourage visual comparisons of corresponding patterns.

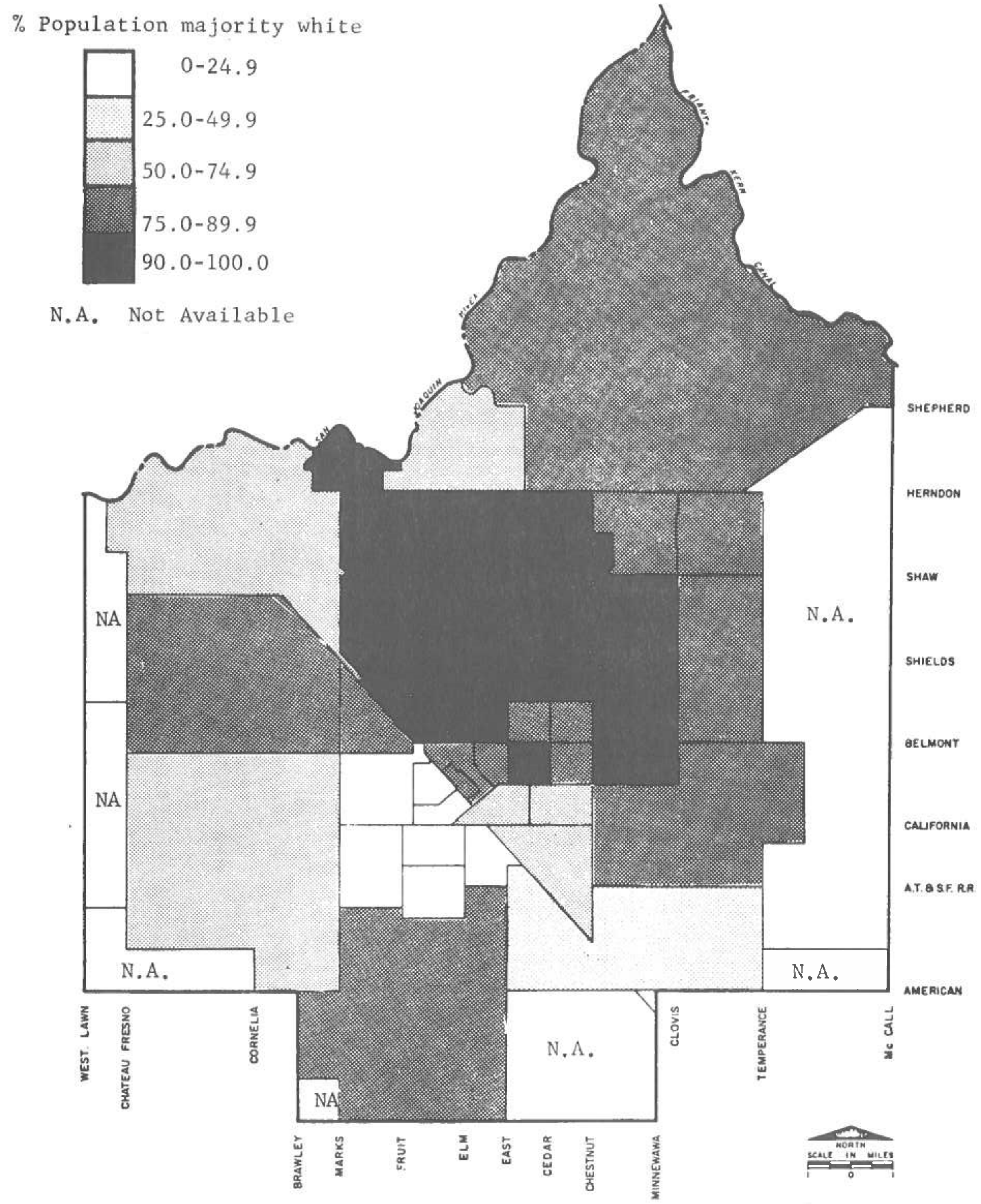
Examples of generalized comparisons of the characteristics of recently developed areas with those of earlier areas--for the Los Angeles-Long Beach, the San Francisco-Oakland, and the Sacramento urbanized areas--are provided as Figs. 6 and 7. Fig. 6 shows that the age composition of recently developed areas is noticeably different from that in areas of early development. Note that this general trend is nearly identical for the three urbanized areas. In each urbanized area the older tracts accommodate some three or four times as great a proportion of older persons and a lower proportion of children (under 20) than do the newer tracts. Fig. 7 shows differentials in automobile ownership. In the tracts of most recent development the proportion of households with access to two or more automobiles very nearly equals the proportion with access to only one, and nonownership of automobiles dwindles to insignificance. The differentials shown on these two graphs are by no means unrelated. Older persons are less likely to drive; the difficulties of getting around in the suburbs without automobiles must in turn discourage older persons from residing there. We developed a large number of these broad comparisons, dealing with some 30 variables.

Graphs such as those just discussed fail of course to get at variations among the newly developed census tracts. To take full advantage of the small-area data at our disposal, we developed a fuller mode of reporting, examples shown as Figs. 8 through 10.

Each of these figures represents a cross tabulation showing the distribution of census tracts for a given urbanized area according to a designated variable on the vertical (Y) axis and according to recency of development, always employed on the horizontal (X) axis. We deliberately chose a format that retained considerable detail but also summarized the overall picture in so simple a manner as to provide virtually a graphic character. The output very nearly reads as a scatter diagram; one is invited to read the numbers as indicating the intensity of the scatter within each cell.

FIGURE 5

MAJORITY-WHITE POPULATION BY CENSUS TRACTS, FRESNO, 1960

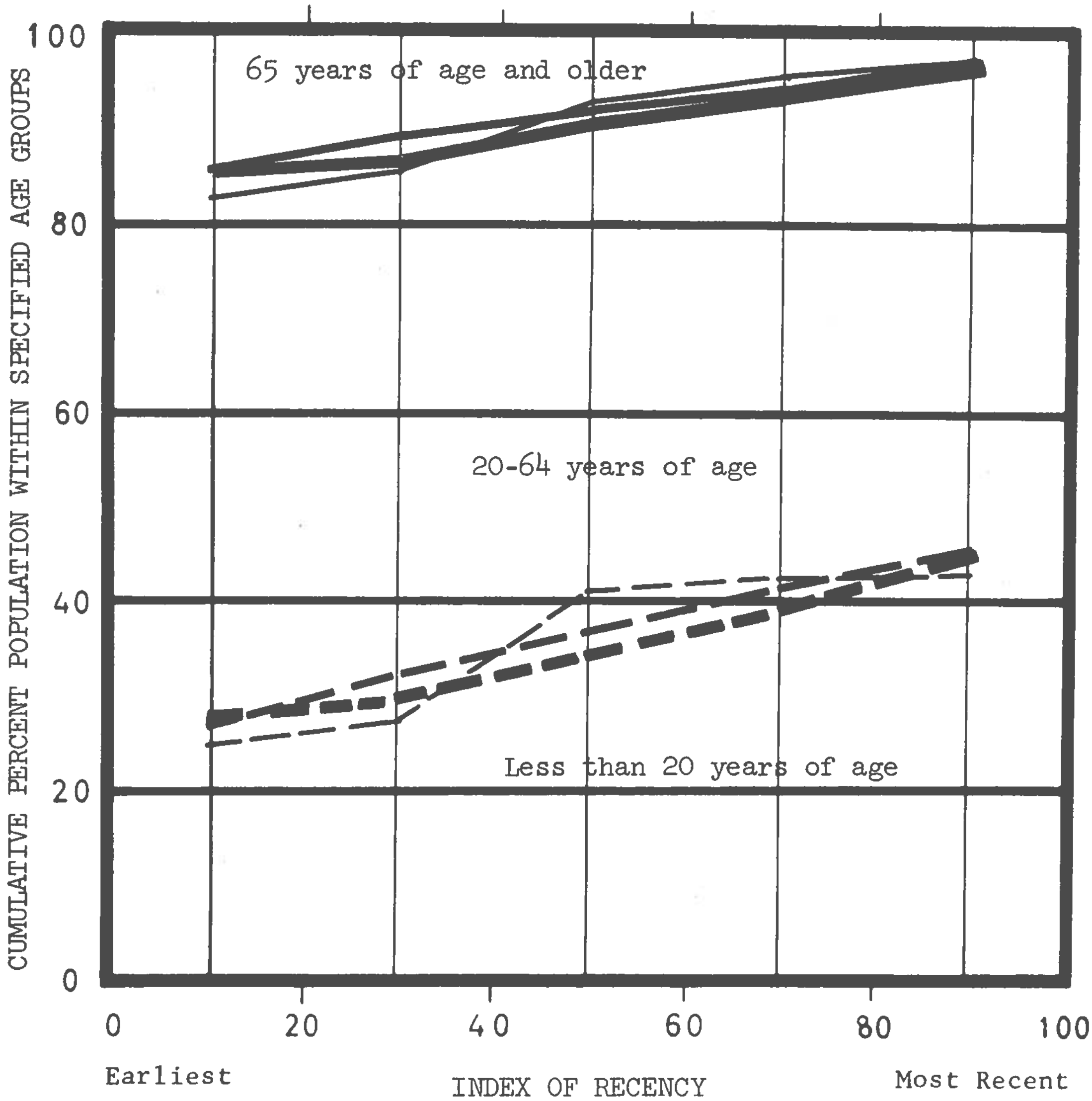


Source: Our calculation from maps produced by Fresno-Clovis Metropolitan Area Project and reproduced in this report above as Figures 4-11A, 4-11B and 4-11C.

Note: "Majority-white population" refers to white population other than that designated as Spanish-surname white in the 1960 Census. Thus it equals, in this case, 100.0% less Spanish-surname white percentage and nonwhite percentage within each census tract.

FIGURE 6

AGE COMPOSITION BY RECENCY OF DEVELOPMENT, LOS ANGELES-LONG BEACH, SAN FRANCISCO-OAKLAND, AND SACRAMENTO URBANIZED AREAS, 1960



Legend: Mean %s

Los Angeles-Long Beach urbanized area

% population less than 65 years of age 90.9%

% population less than 20 years of age 35.7%

San Francisco-Oakland urbanized area

% population less than 65 years of age 90.7%

% population less than 20 years of age 34.8%

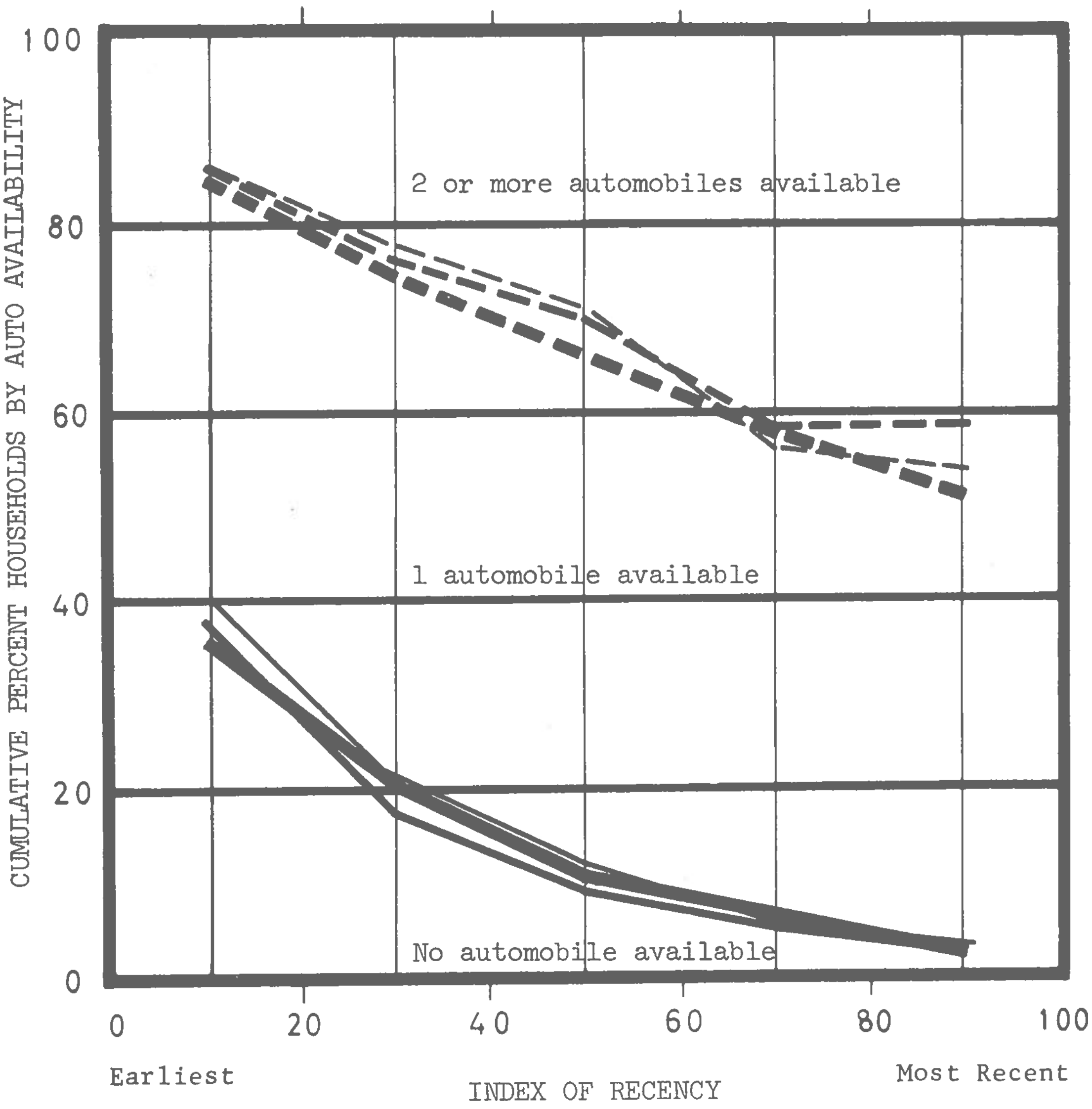
Sacramento urbanized area

% population less than 65 years of age 91.2%

% population less than 20 years of age 36.9%

FIGURE 7

AUTO AVAILABILITY BY RECENCY OF DEVELOPMENT, LOS ANGELES - LONG BEACH, SAN FRANCISCO-OAKLAND, AND SACRAMENTO URBANIZED AREAS, 1960



Legend:

	Mean %s
Los Angeles-Long Beach	
% with less than 2 automobiles available	66.0%
% with no automobile available	14.5%
San Francisco-Oakland	
% with less than 2 automobiles available	72.5%
% with no automobile available	18.8%
Sacramento	
% with less than 2 automobiles available	73.2%
% with no automobile available	21.0%

Let us examine Fig. 8, for the San Francisco-Oakland urbanized area, in which the vertical (Y) axis represents the percentage of the population in each tract that is 65 years of age or older and in which the horizontal (X) axis represents recency of residential development of the tract. Note that on the horizontal axis the first five categories represent 0-19.9, 20.0-39.9, etc. with the fifth 80.0-99.9. To preserve equal class intervals, as expected in the STATPAK programming system, we were forced to place the index value of 100 separately as the sixth category at the extreme right.

The plot at the upper-left hand corner of the page includes all the census tracts. The other three plots represent separate submatrixes for each of three economic levels of home value-rental, as labelled. We deliberately sliced out, so to speak, three of the five economic levels--low, middle and high--and have omitted two further in-between levels. This is because the plot subroutine of the STATPAK system, to avoid the contingency of a figure in any cell exceeding two digits, has the capacity to vary its procedure by letting each figure in the plot represent more than one observation. In the upper left-hand plot of Fig. 8, each plot entry stands for 2 observations or, as need be, an excess single observation; correspondingly in the upper left-hand plot of Fig. 9 (for the Los Angeles-Long Beach urbanized area), each figure represents 5 observations or excess remainders thereof.

These four plot outputs of Fig. 8 show, both in plot form and in the correlation coefficients (r) reported below the plots proper high negative correlations; the more recent the development, the less representation of older population. The plots also show a bunching or homogeneity as among census tracts of recent development with virtually all tracts of very recent development having less than 5% representation of population aged 65 and over; in contrast, there is considerable heterogeneity among tracts of early development, with at least one tract containing in excess of 35% population aged 65 and over. The pattern and the negative correlation persists for each of the three economic levels as well as for the total. Indeed, the highest negative correlation coefficient, -.86, is for the high level of home value-rental where presumably one might expect a somewhat greater freedom of choice by the older persons involved. We conclude that the tracts of older development are carrying more than their share of responsibility for housing older persons and that the residential districts of very recent development, for whatever combination of reasons, are by no means providing their share of housing (and community environmental) accommodations.

Fig. 9 portrays percentage of population in each tract that is Negro on the vertical axis by recency of development on the horizontal axis (this latter the same as in Fig. 8), for the Los Angeles-Long Beach urbanized area. Here the overall negative correlation is generally apparent but less pronounced than in the previous figure. In this case it turns out that the apparent overall correlation masks a more complex relationship with economic level of home value-rental. When we show the plot separately for the 186 census tracts in the low economic level, both the visual plot and the correlation coefficient of +.01 show that the association between

FIGURE 8

POPULATION 65 YEARS OF AGE AND OVER BY RECENCY OF DEVELOPMENT (AND BY HOME VALUE-RENTAL), SAN FRANCISCO-OAKLAND URBANIZED AREA, 1960

Vertical (Y): % Population 65 Years of Age and Over

Horizontal (X): Index of Recency

```
            .
INTERVAL    .  0055   0044   0038
ENDPOINTS   .     0055   0050   0005
            .  +   +   +   +   +   +   +
 105.0  50   +++++++++++++++++++++
               +                          + 0000
  95.0  50   ++                           ++
               +                          + 0000
  85.0  50   ++                           ++
               +                          + 0000
  75.0  50   ++                           ++
               +                          + 0000
  65.0  50   ++                           ++
               +                          + 0000
  55.0  50   ++                           ++
               +                          + 0000
  45.0  50   ++                           ++
               + 1                        + 0001
  35.0  50   ++                           ++
               + 3   1                    + 0003
  25.0  49   ++                           ++
               +32   7   2                + 0041
  15.0  42   ++                           ++
               +55  46  33  25    3      1+ 0165
   5.0  13   ++                           ++
               + 1   2  10  25  35       4+ 0076
  -5.0   0   +++++++++++++++++++++
  PERCEN-      +   +   +   +   +   +   +
   TILES       0  16  26  34  42  49  50
               +       +       +       +
               0.             80.0
                      40.0           120.0
```

TOTAL URBANIZED AREA

N = 570 census tracts

Mean (Y) = 9.34%

S.D. (Y) = 5.79%

Each plot entry = 2 observations

r = -.74

Eta =.73

```
            .
INTERVAL    .  0040   0015   0001
ENDPOINTS   .     0011   0008   0001
            .  +   +   +   +   +   +   +
 105.0 100   +++++++++++++++++++++
               +                          + 0000
  95.0 100   ++                           ++
               +                          + 0000
  85.0 100   ++                           ++
               +                          + 0000
  75.0 100   ++                           ++
               +                          + 0000
  65.0 100   ++                           ++
               +                          + 0000
  55.0 100   ++                           ++
               +                          + 0000
  45.0 100   ++                           ++
               + 1                        + 0001
  35.0  98   ++                           ++
               + 3                        + 0003
  25.0  94   ++                           ++
               +10         1              + 0011
  15.0  80   ++                           ++
               +26   9   4               1+ 0040
   5.0  27   ++                           ++
               +     2  10   8   1        + 0021
  -5.0   0   +++++++++++++++++++++
  PERCEN-      +   +   +   +   +   +   +
   TILES       0  52  67  86  97  98  00
               +       +       +       +
               0.             80.0
                      40.0           120.0
```

LOW LEVEL OF HOME VALUE-RENTAL (0-20)

N = 76 census tracts

Mean (Y) = 9.95%

S.D. (Y) = 7.58%

r = -.60

Eta =.63

```
            .
INTERVAL    .  0036   0018   0018
ENDPOINTS   .     0029   0025   0000
            .  +   +   +   +   +   +   +
 105.0 100   +++++++++++++++++++++
               +                          + 0000
  95.0 100   ++                           ++
               +                          + 0000
  85.0 100   ++                           ++
               +                          + 0000
  75.0 100   ++                           ++
               +                          + 0000
  65.0 100   ++                           ++
               +                          + 0000
  55.0 100   ++                           ++
               +                          + 0000
  45.0 100   ++                           ++
               +                          + 0000
  35.0 100   ++                           ++
               + 1                        + 0001
  25.0  99   ++                           ++
               + 7   4                    + 0011
  15.0  90   ++                           ++
               +28  25  18  12            + 0083
   5.0  24   ++                           ++
               +                13  18    + 0031
  -5.0   0   +++++++++++++++++++++
  PERCEN-      +   +   +   +   +   +   +
   TILES       0  28  51  65  85  00  00
               +       +       +       +
               0.             80.0
                      40.0           120.0
```

MEDIUM LEVEL OF HOME VALUE-RENTAL (40-60)

N = 126 census tracts

Mean (Y) = 9.13%

S.D. (Y) = 5.31%

r = -.82

Eta =.82

```
            .
INTERVAL    .  0017   0021   0032
ENDPOINTS   .     0017   0033   0007
            .  +   +   +   +   +   +   +
 105.0 100   +++++++++++++++++++++
               +                          + 0000
  95.0 100   ++                           ++
               +                          + 0000
  85.0 100   ++                           ++
               +                          + 0000
  75.0 100   ++                           ++
               +                          + 0000
  65.0 100   ++                           ++
               +                          + 0000
  55.0 100   ++                           ++
               +                          + 0000
  45.0 100   ++                           ++
               +                          + 0000
  35.0 100   ++                           ++
               +                          + 0000
  25.0 100   ++                           ++
               +11   2   2                + 0015
  15.0  88   ++                           ++
               + 6  15  18  23   4        + 0066
   5.0  36   ++                           ++
               +         1  10  28       7+ 0046
  -5.0   0   +++++++++++++++++++++
  PERCEN-      +   +   +   +   +   +   +
   TILES       0  13  26  43  69  94  00
               +       +       +       +
               0.             80.0
                      40.0           120.0
```

HIGH LEVEL OF HOME VALUE-RENTAL (80-100)

N = 127 census tracts

Mean (Y) = 7.91%

S.D. (Y) = 5.25%

r = -.86

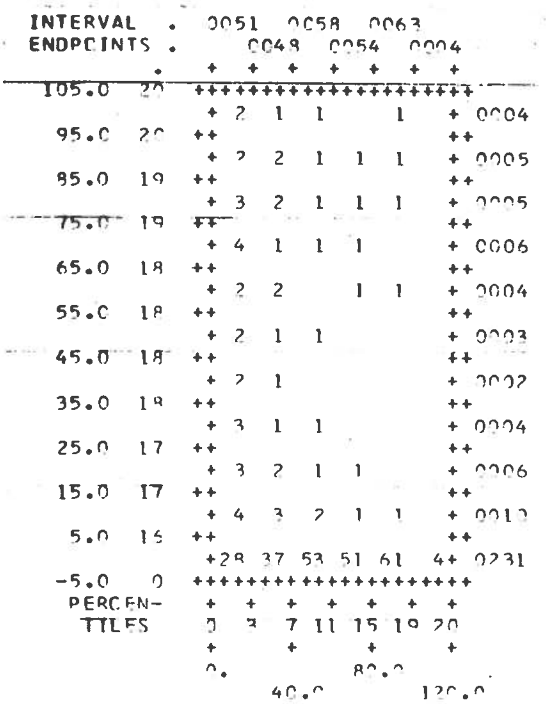
Eta =.85

FIGURE 9

NEGRO POPULATION BY RECENCY OF DEVELOPMENT (AND BY HOME VALUE-RENTAL). LOS ANGELES-LONG BEACH URBANIZED AREA, 1960

Vertical (Y): % Population Negro

Horizontal (X): Index of Recency



TOTAL URBANIZED AREA

N = 1377 census tracts

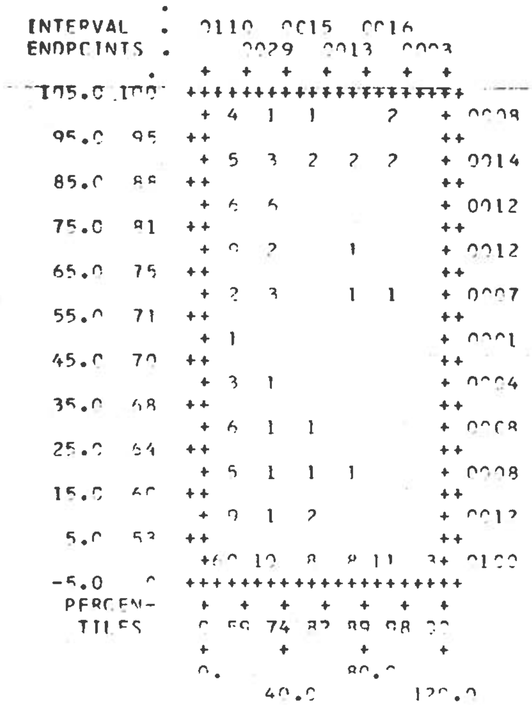
Mean (Y) = 8.20%

S.D. (Y) = 21.99%

Each plot entry = 5 observations

r = -.33

Eta =.37



LOW LEVEL OF HOME VALUE-RENTAL (0-20)

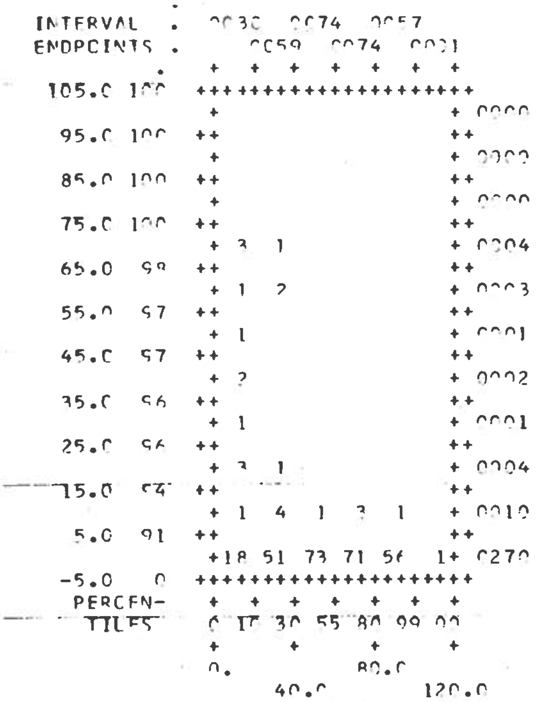
N = 186 census tracts

Mean (Y) = 27.13%

S.D. (Y) = 35.52%

r = +.01

Eta =.22



MEDIUM LEVEL OF HOME VALUE-RENTAL (40-60)

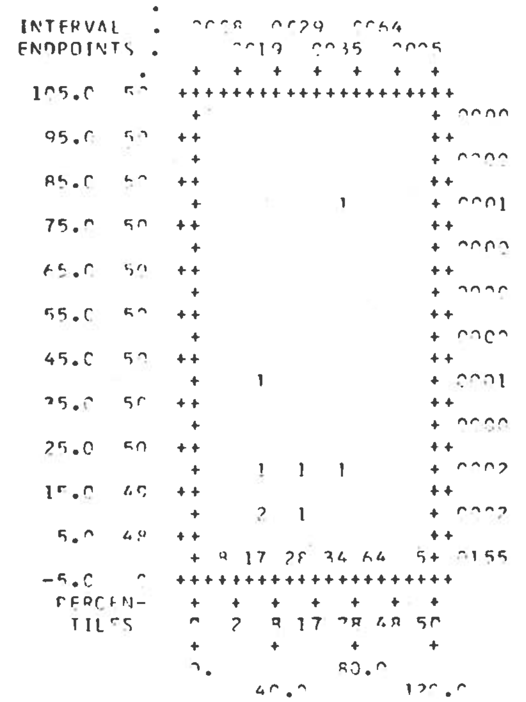
N = 295 census tracts

Mean (Y) = 2.85%

S.D. (Y) = 11.15%

r = -.35

Eta =.45



HIGH LEVEL OF HOME VALUE-RENTAL (80-100)

N = 318 census tracts

Mean (Y) = .94%

S.D. (Y) = 5.35%

Each plot entry = 2 observations

r = -.11

Eta =.16

recency and % Negro drops off almost completely. At the middle economic level, in contrast, there is considerably greater correlation, though by no means significant correlation. The upper left-hand plot, for all economic levels, therefore includes the effect of the relation between economic level and recency. A look at the mean % Negro for each of the three economic levels helps to explain:

186 low economic level tracts	27.1% Negro
295 medium economic level tracts	2.9% Negro
318 high economic level tracts	0.9% Negro
1,377 total tracts of all levels	8.2% Negro

Thus, since new development is generally at medium to high economic levels which have extremely low representation of Negroes, it also follows that new-development tracts will have predominantly low percentages of Negroes. It appears that economic discrimination against Negroes is playing a very significant role as well as, or perhaps coupled with, residential-area discrimination in the newly developed districts.

Fig. 10, for the San Francisco-Oakland urbanized area, shows the percent of adults with a low level of education (eight years schooling or less) plotted against recency of development. But this time we have drawn upon a different set of submatrixes and show separate plots for nonwhite population and for Spanish-surname white population as well as for majority-white population. Here we find that within each of the minority-group components shown there is some tendency for adults with less education to be more concentrated in tracts of earlier development. And this is particularly the case for the majority white population.

Unfortunately, in this brief presentation it is not possible to provide the full feel of intermetropolitan comparisons possible by examining plots for different urbanized areas. In a second-volume appendix to our report we present plot diagrams showing the relationship between each of 30 variables and recency of development, separately for each of the 6 urbanized areas studied. For about half of the variables we show separate plot tabulations for the three population components (as per the example in Figure 10) and for all the variables show separate plot tabulations for the three economic levels (as per examples in Figures 8 and 9). Thus the reader is permitted to make various comparisons of his choice: as between urbanized areas, as between economic level of tract or population component, and as between variables. This encourages persons interested in a particular urbanized area to view that area in perspective.

3. IMPLICATIONS FOR DATA-SYSTEM DEVELOPMENT

In our work we sought to capitalize on the potential power of comparative analysis. We made use of two basic patterns of comparison implicit in the census-tract statistics as assembled and reported by the Bureau of Census: namely, inter-census tract comparability and inter-metropolitan (in our case, inter-urbanized area) comparability. We further introduced other bases for comparison, particularly our two indexes--Recency of Development and Economic Level of the Housing--and systematically classified or crossclassified census tracts according to

FIGURE 10

LOW LEVEL OF EDUCATION BY RECENCY OF DEVELOPMENT (AND BY ETHNIC AND RACIAL STATUS), SAN FRANCISCO-OAKLAND URBANIZED AREA, 1960

Vertical (Y): % Persons Aged 25 Years or Over with Eight Years Education or Less
Horizontal (X): Index of Recency

```
               .
INTERVAL       .  0187  0088  0075
ENDPOINTS      .     0108  0099  0009
               .   +   +   +   +   +   +   +
 105.0 100      +++++++++++++++++++++
                 +                    + 0000
  95.0 100      ++                    ++
                 +                    + 0000
  85.0 100      ++                    ++
                 +                    + 0000
  75.0 100      ++                    ++
                 +                    + 0000
  65.0 100      ++                    ++
                 +                    + 0000
  55.0 100      ++                    ++
                 + 2  1               + 0003
  45.0  99      ++                    ++
                 + 7          1       + 0008
  35.0  98      ++                    ++
                 +22 14   3   1       + 0040
  25.0  90      ++                    ++
                 +60 30 25 12  1    1+ 0129
  15.0  68      ++                    ++
                 +80 49 47 53 36      + 0265
   5.0  21      ++                    ++
                 +16 14 13 32 38    8+ 0121
  -5.0   0      +++++++++++++++++++++
  PERCEN-        +  +  +  +  +  +  +
   TILES         0 33 52 67 85 98 00
                 +     +     +     +
                 0.         80.0
                      40.0        120.0
```

MAJORITY WHITE POPULATION

N = 566 census tracts

Mean (Y) = 12.47%   r = -.46

S.D. (Y) = 8.71%   Eta =.44

```
               .
INTERVAL       .  0154  0039  0013
ENDPOINTS      .     0065  0027  0001
               .   +   +   +   +   +   +   +
 105.0 100      +++++++++++++++++++++
                 +                    + 0000
  95.0 100      ++                    ++
                 +                    + 0000
  85.0 100      ++                    ++
                 + 1                  + 0001
  75.0  99      ++                    ++
                 + 2                  + 0002
  65.0  98      ++                    ++
                 + 5  3  2            + 0010
  55.0  95      ++                    ++
                 +10           2      + 0012
  45.0  91      ++                    ++
                 +27  7  6  2  3      + 0045
  35.0  76      ++                    ++
                 +40 14  9  5  5      + 0073
  25.0  52      ++                    ++
                 +45 23 10  8  2      + 0088
  15.0  22      ++                    ++
                 +13 11  9  6  3    1+ 0043
   5.0   8      ++                    ++
                 +11  7  3  4         + 0025
  -5.0   0      +++++++++++++++++++++
  PERCEN-        +  +  +  +  +  +  +
   TILES         0 51 73 86 95 99 00
                 +     +     +     +
                 0.         80.0
                      40.0        120.0
```

NONWHITE POPULATION

N = 299 census tracts

Mean (Y) = 25.37%   r = -.18

S.D. (Y) = 14.42%   Eta =.21

```
               .
INTERVAL       .  0169  0068  0057
ENDPOINTS      .     0095  0076  0007
               .   +   +   +   +   +   +   +
 105.0 100      +++++++++++++++++++++
                 +                    + 0000
  95.0 100      ++                    ++
                 +           1        + 0001
  85.0  99      ++                    ++
                 +                    + 0000
  75.0  99      ++                    ++
                 + 4  3               + 0007
  65.0  98      ++                    ++
                 + 6  1  4            + 0011
  55.0  95      ++                    ++
                 +17 10  5  5         + 0037
  45.0  88      ++                    ++
                 +25 26 13  9  1      + 0074
  35.0  72      ++                    ++
                 +35 18 15 14 10    1+ 0093
  25.0  52      ++                    ++
                 +34 16 13 15 10    1+ 0089
  15.0  33      ++                    ++
                 +31 15 11 19 27    1+ 0104
   5.0  11      ++                    ++
                 +17  6  7 13  9    4+ 0056
  -5.0   0      +++++++++++++++++++++
  PERCEN-        +  +  +  +  +  +  +
   TILES         0 35 55 70 86 98 00
                 +     +     +     +
                 0.         80.0
                      40.0        120.0
```

SPANISH-SURNAME WHITE POPULATION

N = 472 census tracts

Mean (Y) = 24.63%   r = -.25

S.D. (Y) = 16.58%   Eta =.30

the values of these indexes. We took advantage of the valuable breakdown by nonwhite and by Spanish-surname white population (fully available in magnetic tape format, although in published reports reported only for tracts having large concentrations of these populations); we added a third category, Majority White Population, the residual.

We did not have the possibility of making intercensal year comparisons, mainly because all or main portions of most metropolitan areas in California had not been tracted for the 1950 census. Our capacity to study changes in residential area was thus seriously circumscribed.

What lessons did we learn? What would we recommend as promising future lines of effort? Let us discuss briefly.

1. We are favorably impressed with the power and the relative efficiency of having a standardized approach to the classification and assembly of data on a small-area basis for many metropolitan areas, precisely as the Bureau of the Census is in such an excellent position to do. This provides a set of accounts, so to speak, that are comparable and are thus aggregatable and, within reason, dissaggregatable. When the comparability of data is also extended to comparability of magnetic tape format, then of course common programming can be applied to many or all metropolitan areas, with resulting economies.

2. We are particularly convinced that a large volume of data systematically organized, as is the case with the census-tract data, is never used to full advantage until it is fully and effectively linked with an excellent analytic programming system. We were genuinely pleased with the utility of the STATPAK system of the University's Computer Center; it brought us a sense of push-botton ease in carrying through what otherwise would have been extremely difficult programming. We even raise the point as to whether it may not be more important to focus on the availability of such analytic systems that can be widely shared than necessarily concentrate too exclusively on data-bank systems per se.

3. Clearly, behind both data-assembly and analytic systems lies a tremendously important sphere involving the design of the analysis. We cannot avoid grappling with the fundamental question of what it is that we want to find out. Further, we must be able to decide what analytic strategies are most applicable and to make a rational choice of the strategy to employ. This calls for considerably greater methodological sophistication than is often available on governmental or other staffs.

4. It should be clear that in urban planning we are particularly eager to enhance our capacity to forecast and to simulate. We did undertake a regression-analysis approach to forecasting how typical tracts of recent development are likely to change in their characteristics. This did not prove very satisfactory, mainly because we lacked the full comparative small-area data from any census earlier than the 1960 one. We attempted to simulate change by relying heavily on our Index of Recency

values as applied to 1960 data, but this did not provide a very convincing tool. And at best it approached only the task of forecasting typical change and failed entirely to get at the more realistic question of which census tracts are likely to change in certain ways and which to change in other ways.

5. Although we had hoped to splice additional data to the already available census data, we became so discouragingly bogged down in merely editing and putting into fully usable form the data from the Census Bureau that we were not able to proceed further within our time schedule and financial budget. We obtained, for example, punched cards with estimates for 1965 population and housing counts for the census tracts in the city of Los Angeles, thanks to the cooperation of the Planning Department there. We had intended to run multivariate regression analysis, using population or housing change from 1960 to 1965 as dependent variables to determine which combinations of 1960 indicator variables provided the best preditors of this subsequent change. But we finally stayed with our main intermetropolitan analysis, since we could not obtain fully comparable 1965 estimates for other metropolitan areas.

6. The Census Bureau should be encouraged to provide much more effectively edited and systematized tapes. We spent untold energy merging 100%-data tapes, 25%-data tapes, and latitude-longitude centroid cards. The Census Bureau should also be most pointedly encouraged to develope generally applicable compilor and other analytic programs that can be used with census-supplied magnetic tapes.

7. It seems likely that by the 1970 census smaller geographic units, presumably down to block face, will be used. It is not clear, however, how much data will be supplied and whether this data will be of "marginal" character (i.e., providing the margins only of potential cross tabulations) or whether it will be sufficiently disaggregative to permit cross tabulations within the geographic unit. Presumably the data for block-face areas will permit internal cross tabulations, but we understand that, to avoid disclosures, the Census will release data only for certain combinations of block faces. The questions then arise whether the census tract is also to remain a main areal building block and which internal disaggregations permitting cross tabulation will be provided. For example, we made effective use of the disaggregation by nonwhite and Spanish-surname population. Should we, in the same vein, be encouraging other disaggregation within each census tract?

8. As indicated earlier, we did not get into computer mapping. This is a potentially valuable adjunct of the system. It must be understood, however, that computer mapping involves a very considerable advance commitment to tool up. And for intermetropolitan analysis, this means separate preparations, metropolitan area by metropolitan area. Since the entire state of California is covered by pseudo tracts, it might prove helpful to be able to map nonmetropolitan portions of California. The potential benefits clearly need to be examined as against the costs.

9. Finally, and perhaps merely belatedly catching up with what many others already know, it became clear to us that the costs in time, money, and patience of developing a full system of data and analytic capacity are very great. They would seem to pay off for California Metropolitan Areas only if a strong lead can be maintained by a planning and intelligence unit of the state. It may never prove economical for local agencies or separate firms to attempt to build and maintain full separate systems, although selected larger units (e.g., Santa Clara County and the Bay Area Transportation Study) are making notable progress.

REFERENCES

1. Computer Center, University of California. STATPAK Rulebook. Berkeley: Revised edition, March 1965.

2. Foley, Donald L., Drake, Ronald L., Lyon, David W., and Ynzenga, Bernardo A. Characteristics of Metropolitan Growth in California. Volume I: Report. Volume II: Appendixes. Berkeley: Center for Planning and Development Research, University of California, December 1965.

3. Shevky, Eshref, and Bell, Wendell. Social Area Analysis. Stanford: Stanford University Press, 1955. Also Shevky, Eshref, and Williams, Marilyn. Social Areas of Los Angeles. Los Angeles: Haynes Foundation, 1949.

4. U. S. Bureau of the Census. Contents and Record Layouts of the PB2 Population and Housing Output - 3 Block Matrix. Technical Memorandum No. 29B. Washington, D. C. : March 17, 1961. (Processed.)

5. ____________________. Contents and Record Layout of the Weighted Tallies for Tracts and Pseudo Tracts - 8 Block Matrix. Technical Memorandum No. 291. Washington, D. C. : October 9, 1961. (Processed.)

6. ____________________. U. S. Censuses of Population and Housing: 1960. Geographic Identification Code Scheme. California. Report PHC (2)-6. Washington, D. C. : 1961

THE ECONOMIC COMPONENT OF A REGIONAL
SOCIO - ECONOMIC MODEL

Robert G. Kokat *

INTRODUCTION

A basic area of research in city and regional planning is the development of operational models to evaluate specific programs in advance of their implementation. The importance of specifying operational models is based on the fact that once a program is initiated it may be irrevocable in the short run and produce intermediate impacts on the region which were unexpected and undesired at the outset. For example, Professor Petersen cites the Venezuelan government's housing project in Caracas designed to accomodate 180,000 people. While the physical plant was completed satisfactorily, the social problems that emerged led to a recommendation to suspend all construction of superblocks until a defined housing policy related to the economic and social development of the country was established.[1]

The traditional emphasis on physical plant by city planners and broad economic policy by national planners is seen as outmoded by Petersen. He emphasizes the need for "systems analysis with the computer as the fundamental tool and model building as a routine step, and also that branch of sociology that concentrates on the social setting of ethnic relations, poverty, education, and labor-force recruitment."[2]

Following Petersen's article in the May issue of the AIP Journal, Lowdon Wingo suggests that "simulations of the real world which seek to define causal relationships and predict the consequences of alternative policies" produce outputs "with a high degree of relevance for decision-makers involved in the urban renewal decision process."[3] Although Wingo's article focuses on the development of simulation models for use in urban renewal, his argument and conclusions may easily be extended to include the use of simulations for a variety of policy decisions at the regional and local levels.

---

* The model described in this paper was developed by an operations research team from IBM for the Information Center of the Office of Economic Opportunity, Washington, D.C. I am indebted to the other members of the team, Jose Glasserman, Stan Bender, Charles Dym, Herb Lieberman, Ronald Nuttall, and Marshall Childs, for their thought and suggestions on this section of the model. I am especially indebted to Tim Ling and John Coatti for programming the model.

If Petersen and Wingo are representative articulators of the current needs for new tools of analysis in the planning profession, then the models being developed for the Office of Economic Opportunity are illustrative of recent efforts in this direction. For example, a simulation model just completed for OEO by the Philco Corporation under the name "project scancap," was developed to analyze alternative community action programs to help in selecting that mix of programs that gives the most return per dollar of expenditure, and to analyze the optimal levels of expenditures for given urban communities. This morning's speaker, Wilbur Steger, was the major architect in designing the model for project scancap.

Under the direction of the Information Center at OEO, a regional simulation model for the East North Central states of Indiana, Illinois, Michigan, Ohio, and Wisconsin has been constructed. The model is a research model intended to simulate economic, demographic, and social variables to define causal relationships and attempt to predict the consequences of alternative OEO policies within the spirit of Wingo's suggestions and the pioneering framework of systems analysis implemented by Steger. My discussion this afternoon will be specifically addressed to the economic components of the model.

OBJECTIVES OF THE MODEL

The objective of the economic component of the regional model is to examine the levels of new employment by industry and occupation which are generated as a result of alternative poverty program strategies. Specifically, this aspect of the model is designed to evaluate the impact of different levels and compositions of regional final demands consistent with the expenditures for functional poverty programs on the region's output and employment requirements.

A principal consideration in the design of the model is that poverty may be viewed as both an economic and social problem. The programs initiated by OEO affect not only the economic status of individuals within the region but also their attitudes, desires, education, competence, employability, migration, physical environment, and rate of discrimination. To simulate the possible reduction in the incidence of poverty, the model should specify relationships that at least implicitly attempt to measure effects on these social variables. If we are to provide insights on social consequences, then the economic component must provide outputs which may be translated as inputs to the demographic and sociological components of the model as well as recording the economic impacts of the programs. In addition, the proposed changes in social conditions such as education and discrimination will alter the economic relationships described for the region. In short, the economic component is an element of a larger system, and its design is governed by the operations of that larger system.

Model design is also conditioned by the type of output or report that is desired by the user. Our interest is to describe major segments of a region's economic and social activities over a period of time. In the first phase of the model's construction, we have agreed to limit the time span to three years. Given the initial description of the region, we simulate alternative descriptions that could occur as a result of the war on poverty.

A series of experiments has been designed to permit a systematic analysis of the range of programs under consideration by the Office of Economic Opportunity. The programs to be evaluated may be grouped into three major classes: (1) direct expenditures, (2) public transfers, (3) private investment.

A reference run will simulate the level of economic activity that takes place in the region assuming that all expenditures made by federal, state, and local governments and private firms were and will be made for reasons other than the elimination of poverty. For example, we assume that governments make expenditures for missiles and schools for the purpose of security and education, but not primarily for the purpose of creating jobs for the poor. By contrast, a subsequent experiment will assume increased defense expenditures are made in the region by shifting the national composition of contract awards for the express purpose of creating additional employment opportunities in the region. The intent of the reference run is to establish levels of activity that could reasonably be expected to occur in the absence of specific, identifiable stimuli aimed at generating economic activity for purposes of job creation. The output generated by the reference run will be used to examine the region's sensitivity to marginal gains (or losses) associated with any of the poverty oriented experiments.

The first set of experiments will assume that the external stimulus is provided by direct federal expenditures in the region for the express purpose of aiding those in poverty. Within this set, a few experiments may be directly associated with existing OEO programs while other experiments will be based on programs associated with other federal agencies. For example, we can evaluate the direct and indirect impact on the region of the consolidated efforts of the job corps, neighborhood youth corps, headstart, and community action programs on the one hand, or the impact of an identical expenditure for public housing on the other. While the former programs originate with OEO and the latter with HUD, they are all programs which require expenditures by the government on behalf of those designated in the poverty class. The output from these experiments may be compared with the reference run to establish the incremental gains in employment, income, education, discrimination, and the like.

Another set of government programs which may be tested falls into the category of expenditures not directly affecting the poverty program.

Whereas the former set of programs are either initiating by OEO or by other government agencies on behalf of the poverty program and are intended to directly affect the economic status of the group in poverty, this second set of programs is concerned with unrelated government expenditures which may generate indirect benefits to the group and improve their economic position. For example, an increase in defense expenditures in the region may produce no direct benefit to the poor, but may increase the demand for workers in several unrelated sectors of the region's economy as well as the additional employment in the defense and allied industries. Programs other than defense such as school and highway construction will also be tested within this set of experiments. Again, the outputs generated will be evaluated for marginal gains with respect to the reference run.

Transfer programs represent a third set of experiments to be tested with the model. Whereas government expenditure programs represent demands placed upon the productive capacity of the economy for goods and services, transfer payments made to individuals require no <u>direct</u> exchange of goods and services. The individual receiving the payment produces nothing in exchange but is provided with the ability to purchase products which in turn may increase aggregate demand in the region (depending in part on the manner in which the transfer is funded).

Although the multiplier effect associated with a transfer payment is dampened somewhat over that of a direct government expenditure, it does provide the policy-maker control over the direction and magnitude of his outlay within the poverty group with less variance than an expenditure program. For example, an equal outlay for public employment may generate a higher level of aggregate activity than an increase in social security benefits, but the social security payments may be directed at groups who could not be reached or whose economic position could not be improved under the public employment option. Hence, the transfer payment may be considered a better alternative in some instances because of the distribution effects realized.

Even though the economic framework of the model is limited with respect to identifying the effects on income distribution the sociological and demographic components of the model have been designed to evaluate the short term and longer run implications of the changes in income distribution brought about by modifications in transfer payments programs.

Our last group of experiments will simulate the impact of industrial development in the region. The impacts will be derived by increasing the capacity of selected sectors of the economy through private investment outlays of an equal order of magnitude as those tested for the public programs. Within this set of experiments we shall focus upon investment alternatives which are unrelated to the demands of the previous programs. For example, several of the public policy options outlined above may require or provide the necessary demand incentives for firms to increase their productive capacity within the region. Indeed some of the alternatives may not be operational in the short run due to capacity and/or labor restraints calling for additional investment expenditures. This form of induced investment is handled explicitly in the endogenous component of the model.

COMPONENTS OF THE MODEL

The basic foundation of the economic component of the overall system is a variant of the traditional static input-output model. The input-output approach was selected because it is one of the few macro economic techniques that approximates a complete system and at the same time provides disaggregate detail necessary for the demographic and sociological analysis required by our experiments.

The major purpose of an input-output model is to identify and measure the flow of goods and services between various sectors of the economy for a given period of time. The flows differentiate between intermediate purchasers and final purchasers - an intermediate purchase being defined as one made by steel mills, automobile manufacturers, chemical manufacturers, and the like, and a final purchase as one made by households or governments. This classification scheme enables us to trace the pattern of demand through various stages of processing. In input-output terminology, the intermediate purchases are considered indirect requirements or requirements which result from the demands of final buyers. Hence changes in the demands made by final buyers will alter the requirements of intermediate producers.

A realistic appraisal of the employment, income, and output generated by additional public and private final demands should attempt to measure variations in total demand requirements -- direct as well as indirect requirements brought about by increased intermediate production and employment needs. A measure of total demand requirements is important in evaluating regional impacts since the "local" economy is usually more open to import-export activity than the relatively closed system of the national economy.

Given the degree of geographic concentration of highly specialized industries, a large proportion of the final demand requirements made by residents of the region may be met by firms in production centers located outside of the region. Hence the net effects of an attempt to stimulate demand and increase specific occupational opportunities in the region may be obscured by a failure to understand the inter-regional supply relationships between the region and the rest of the nation.

To derive a systematic method of tracing transactions between the region and the rest of the nation, an inter-regional flow table was constructed. Since the inter-regional flow table is the initial component of our time phased model, we shall devote a few paragraphs to a simplified discussion of the assumptions and methods used to derive the table.

The flow was derived by adapting the national input-output table constructed by the Office of Business Economics of the U.S. Department of Commerce. The national table and articles discussing its development have been published in various issues of the Survey of Current Business.[4] The Commerce model classifies all industrial and business activity into eighty processing sectors on the basis that firms may be grouped into similar sectors with respect to products or services produced on the output side and the similarity of materials purchased for production on the input side. The table lists the eighty industries by rows and columns which display the transactions that take place between the industries during the year 1958. The flow table may be converted to a direct requirements table by a simple process of dividing each entry in a column by its column sum. The direct requirements table indicates the proportion of purchases one industry must make from every other industry in order to produce one million dollars' worth of output. In a sense, each column in the direct requirements table is analogous to a recipe listing relative proportions of the ingredients necessary to produce the final product.

The national direct requirements table is the key component for deriving our inter-regional flow. The column entries in this table are explicit statements of the industries' production functions for the base year's product mix. By applying the national production functions to the region's output levels on an industry by industry basis, a first approximation of the shipments and purchases between the region's industries is established. It is only an approximation since it assumes that all the industries in the region use production techniques that are identical to those used by their national counterparts. This assumption is limiting but not restrictive for regions at the state level. Furthermore, adjustments can be made for industries in the region where variance between regional and national production functions is suspected, by using additional Census data on the size distribution of industry.

A more serious limitation of this method is the assumption of invariance of product mix between regions. This restriction may be reduced by disaggregating the eighty industries or by defining larger regions.

The derived regional flow may be adjusted for imports and exports on an industry basis by summing the flows across the rows and comparing each sum with the region's actual industry output levels. This operation gives us an indication of which industries in the region are net exporters or net importers. In deriving the region's flow, we multiplied each industry's actual output level by the direct requirements proportions in the corresponding column of the national direct requirements table. Since the columns are independent, the row flows derived by this procedure will only sum to the identical values of the actual industry output levels if the region is completely self-contained. In such a case, the relative output levels of the region's eighty industries are identical to the relative output levels of the nation's eighty industries.

If the derived value of the row flow for any industry is greater than the actual output level, then the industry is a net importer from the rest of the nation. The excess of the derived sum indicates that the industry could not meet all of the demands made by the other industries in the region, for if it could meet their demands, then the actual output level would have been greater. Hence the difference recorded must be imported or the other industries could not have operated at the level implied by our derived regional flow. Alternatively, if the derived value for any industry is less than the actual output level, then the industry can meet all of its regional demands and will export the residual to the rest of the nation. For example, the actual total output level of the region's automobile industry greatly exceeds the derived level since the demands by the region's residents represent approximately twenty percent of the nation's demand, whereas the region's automobile industry output represents over eighty percent of the nation's output.

The derivation of the level of imports and exports for each industry in the region is a necessary but not sufficient condition for specifying the inter-regional flow table. If an industry in the region is a net exporter (automobiles), we must identify the consuming industries in the rest of the nation that import the region's surplus products. Similarly, if the region's industry cannot supply the demands of other industries in the region (basic textiles), we must also identify the industries in the region that do the actual importing. After the identifications are made, assignment of the relative amounts of imports and exports must be specified on an industry by industry basis. The identification and assignment process is based on two relatively naive assumptions: (1) that only industries import, not final consumers; and (2) that the imports are assigned to the industries on a proportional basis. The first assumption implies that residents of the region purchase final products from the legitimate institutions of the region. For example, the purchase of a new Ford manufactured in Detroit by a resident of California is usually made from a dealer in California not from a dealer in Michigan. The second assumption requires that a deficit of regional output will be imported by all the consuming industries in proportion to their use. For example, if ten industries purchase steel and the region's steel industry cannot meet their requirements, then all ten industries must import steel from the rest of the nation in magnitudes relative to their initial demands. Assignment on any other basis would require a survey of the actual location of the suppliers of each industry. The limitation of these assumptions is a function of the region that is specified.

For example, if we developed this model for the state of Maryland and the rest of the nation, we would have to include the District of Columbia and two counties in Northern Virginia in the Maryland region since the volume of direct importing by individuals in this area violates the first assumption.

Based on the preceding assumptions, the assignment process may be accomplished with a minimum of effort. The result of the process is the derivation of two additional eighty by eighty tables and a modification of the first approximation of the regional flow table. One of the derived tables records the net imports from the rest of the nation and the other records the net exports. If a supplying industry cannot meet its regional commitments, the original row flows are reduced such that the sum of the row is equal to the actual regional total output level, and the deficits by demanding industries are transferred to the corresponding row in the import table. If the supplying industry has a surplus, the excess is transferred to the corresponding row in the export table and allocated across the row on the basis of the national demand pattern. No adjustment to the original regional flow table is necessary for supplying industries that are net exporters.

The final step in constructing the inter-regional flow table requires specifying the flow between the industries in the rest of the nation. This flow is derived as a residual by subtracting the original regional flow table from the actual national flow table.

The segments of the inter-regional flow table may be summarized as follows: (1) The first eighty rows and columns describe the transactions that take place between the industries in the East North Central region. (2) The last eighty rows and the first eighty columns show the imports from the rest of the nation to the region. (3) The first eighty rows and the last eighty columns identify the exports from the region to the rest of the nation. And (4) the last eighty rows and columns describe the transactions between industries in the rest of the nation.[5] The derived flow table must be consistent with our second major component -- final demand.

In social accounting aggregate final demand is divided into four major components -- personal consumption, private investment, government purchases, and net exports to foreigners. Of these major components, approximately seventy percent of the total is accounted for by personal consumption. Even though aggregate consumption within the region has maintained a fairly stable and steady growth over the past decade, of greater interest is the fluctuation in the relative shares of purchases by income groups. Indeed if we are to examine the impact of income maintenance, housing, education, and training programs on the incidence of poverty, a finer breakdown in personal consumption demand must be specified.

To establish a meaningful disaggregate of the region's consumption function, we examined the variations in relative expenditures by family size and income class using the cross-section sample survey taken by the Bureau of Labor Statistics.[6] The Bureau of Labor Statistics reported expenditure patterns for income groups in thousand dollar increments up to six thousand dollars and in larger increments up to fifteen thousand dollars. The results of this analysis showed that expenditure patterns in the region were clustered into three distinct groups: 0 to $4,000; $4,000 to $8,000; and $8,000 and over. Using these results as guidelines, we established seven consumption functions for the region. The first five are in thousand dollar intervals to $5,000. Even though there was minimum variance in consumption behavior up to $4,000, the added detail was considered necessary for providing supplementary data to the sociological and demographic components of the model. The last two consumption functions ($5,000 - $8,000 and over $8,000), were specified on the basis of a distinct variance in expenditure behavior.

The remaining final demand categories were specified along the traditional lines of accepted social accounting with certain modification to facilitate the inter-industry framework of the flow table. Since there is no single industry in the model that produces investment or government products in the form in which they are consumed, the demands enter the model as first order inputs. A first order input may more easily be illustrated than defined. For example, to specify housing demand -- one sub-category of investment -- we derive the proportions of bricks, lumber, cement, etc. necessary to build the type of housing demanded. The disaggregate products are summed up and stipulated as final demand requirements on the appropriate producing industries. This accounting procedure is also used for weapon systems, sewage systems, schools, dormitories, transportation systems, and the like demanded by the government.

Changes in inventory levels consistent with the overall level of final demand are placed on the appropriate producing industries. Changes in inventory levels were derived endogenously in the model for succeeding time periods, and the values for the initial time period were obtained from Census information. Residential construction, plant construction, producers' durable equipment, and all other construction are specified in one category in the first version of the model. This category will be disaggregated when conversion factors for the individual components of investment are developed.

Government purchases have been differentiated on the basis of federal versus state and local, and defense versus non-defense. For the first version all defense expenditures appear as one category and all non-defense expenditures in another. The last category of final demand, net foreign exports, was obtained by proportionately scaling the national values published by the Office of Business Economics.

OPERATING SEQUENCE OF THE MODEL

In a static model the flows recorded for the intermediate sales and purchases are segregated from the final demands and are used to derive a structural table linking the processing sectors of the economy together. The structural table enables the analyst to derive a consistent set of intermediate demand requirements for any reasonable set of alternative final demands. Under this conceptualization the processing sectors represented by the structural table are the endogenous component of the model and the alternative sets of final demands are placed on the system as exogenous or autonomous inputs. The dichotomy in this formulation is based on the assumption that the intermediate demands generated by the processing sectors of the economy are a function of technology, product mix, and input proportionality while the final demands are based on behavioral relationships that cannot be as rigorously defined. When the stipulated final demands are multiplied by the structural table, the level and composition of purchases which must be made by the endogenous processing sectors are generated. Static analysis is usually performed by comparing the output levels obtained for different sets of final demands where the final demands represent formal statements of alternative policy actions.

A dynamic formulation of an input-output model requires the specification of relationships for the autonomous sectors such that they could be included in the endogenous component of the model permitting the model to produce its own feedbacks on final demand for succeeding time periods. Although many of the conceptual problems of dynamic models have been solved, empiric implementation is not workable at the present time. A compromise between static and dynamic models is represented by the technique of time phasing. In a time phased model, the analyst updates the final demand requirements for succeeding time periods based on the derived outputs of the static model. For example, the derived output levels imply a certain level of employment and income which in turn may be used to specify consumption demand for the next time period. Further examination of the output will indicate necessary investment alternatives to be implemented. A systematic review of one year's output plus some side calculation enables the analyst to derive next year's final demand and determine whether any structural changes should be made. In essence, time phasing is analogous to simulation in that the circuit is closed and the direct and indirect feedbacks are systematically evaluated permitting the model to trace the course of behavior of the variables through time.

In the OEO model, we have experimented with time phasing by writing algorithms or sets of decision rules to systematically review the output of each year's run and update final demands for following time periods. The decision rules incorporated in the final phase of the model are naive statements of behavior and were developed primarily to test the consistency of the system. We shall describe our naive decision rules for updating demand by outlining the operation of the system.

The initialized final demand for the region and the rest of the nation which is represented by a matrix of 180 rows and 12 columns is multiplied by the structural matrix represented by the Leontief Inverse of the inter-regional flow table. The result of this computation is a set of total output levels for the industries in the region and the rest of the nation which is consistent with the initial final demand requirement. These results are printed out for later analysis. The total output levels for the region's industries are converted to labor requirements which become inputs to the sociological component of the system. The region's supply of labor by occupation is matched against the demand requirements and jobs are assigned on the basis of predetermined decision rules. The rules established by our sociologists take account of characteristics of the population and employer preferences with respect to sex, age, education, race, and family composition. After the assignment is made, rates of employment and unemployment by industry and occupation are calculated. From these computations, wages, transfers, taxes, and finally family disposable income are derived. We then translate the income distribution into classes for specifying consumption demands for the next year.

Changes in regional population based on births, deaths, and net migration are taken from the demographic component to determine the second year's level of state and local governments' non-defense expenditures in the region. Using the average per capita expenditure in the region for net gains in population, additional expenditures are computed and added to the original level specified. Net losses in population are used to derive the cut-back in government requirements at a rate of thirty percent of the average per capita expenditure. Thirty percent was selected arbitrarily to reflect the fact that government expenditures will not decline as rapidly as population due to the indivisability of the services and the problem of allocating purchases to current or capital accounts. These decision rules will be revised after we derive a set of functional expenditure categories and compute marginal per capita outlays. The first major category to be disaggregated will be education, and the demands for current and capital purchases will be based on the status of the region's population by age cohorts over time.

Updated investment demand is a lagged variable in the model and is calculated on an industry basis by comparing the net gains or losses in final demand in the previous two time periods. The second year's investment requirements are determined by multiplying the appropriate capital-output ratio for each industry times the observed change in final demand for each industry. The marginal investment or disinvestment requirements are summed and added to the initial level of investment specified, and used to update the original demands. The two major problems with this procedure are that no explicit account is made for unused capacity, and that the magnitude of new investment is based primarily on maintaining the past growth rate. These decision rules will be changed in the next version after we disaggregate investment into residential construction, plant construction, producers' durable equipment, and all other private investment. Changes in residential construction will be a function of population and family formation as well as changes in income. Capital capacity estimates for each industry will be made to determine more realistic demands for new investment, and updated capital-output ratios will be used to reflect the trend to more (or less) capital intensive methods of production. A routine for labor capacity has been written for use in conjunction with the employment supply generated by the sociological component of the model.

The last demand component that is updated endogenously by the model's system is the change in inventory demand. We have estimated the average level of inventory that each industry maintains during the year. Since it is difficult to determine whether the decision to hold inventory is a desired one or not, the estimated levels are established by observing beginning and year end stocks for a five year period. A desired ratio for the stock of inventory to final demand for each industry was calculated from the derived estimates. To compute the change in inventory demand (flow) for the next period, we apply the ratios to the estimated final demand levels which gives us an estimate of the desired ending inventory. The change in inventories for the year is established by subtracting the ending inventory levels from the beginning levels. The computed changes for each industry are entered in the final demand table for the next year.

Federal defense procurement, regional exports to foreigners, and the final demands of the rest of the nation are stipulated exogenously. Defense expenditures in the region are increased at the modest rate of one percent per year; total foreign exports are raised by six percent per year without regard to changes in product mix; and the demands by the rest of the nation on the rest of the nation's industries are increased at three percent per year. The rates of change specified for these three components were estimated by calculating average changes for the period 1955-1965 from published data in issues of the Survey of Current Business.

Total output levels for each industry for the second year are derived by multiplying the updated demand requirements by the structural table. After printing out the second year's output, the entire process is repeated. While the computer program for the model operates as a closed system and year by year output can be obtained indefinitely, we should point out a few major characteristics of the operating logic. First, the input-output component of the model is still static since the final demand requirements are not derived endogenously along with the intermediate demand requirements. Furthermore, net capital formation does not effect either changes in the structural characteristics of the processing sectors nor changes in the domestic import-export pattern. Second, all the relationships specified are linear approximations, and relative changes in the magnitudes derived should be interpreted with care. Finally, no account is taken of relative price changes during the time period simulated.

Although the qualifications to the model may appear restrictive, the system as presently formulated will produce meaningful output for the assessment of policy options. The model is an abstraction of the real world which attempts to express relationships in a simplistic manner, and permit us to focus upon sectors of the region which may experience adverse reactions to any planned policy. Given the magnitudes of the programs to be tested, it is highly improbable that within the space of a three year time period the structure of existing industry or the import-export patterns will be substantially altered. If output from the model indicates massive bottlenecks, then the analyst will have to discount the model's results on the basis that increases in net importing will take place in the short run if excess supply is available in the rest of the nation. At a minimum, the model will identify these bottlenecks. Similarly, the formulation of linear production functions should pose no operational problem if the marginal requirements entered on the system by the experimenter do not call for doubling the region's labor requirements. Again this is a matter of sound economic judgement on the part of the model user. In short, the model is a tool which vastly increases the analyst's ability to trace the interaction of many variables, but it is no substitute for actual evaluation.

The decision rules incorporated in the first version of the system will be revised for internal consistency after test runs are made with actual data. At present all the components of the economic section of the model are running, but must be linked together with the components of the sociological section. Of the three major components in the sociological section, one has been programmed and the other two are flow charted. Consequently, we have decided to test the economic system independently by specifying the linkages with the sociological components with a reduced set of decision rules. We are presently at this phase of our study and plan to have a running version of the entire model by early October.

NOTES

1. William Petersen, "On Some Meanings of 'Planning'," Journal of the American Institute of Planners, Vol. XXXII, No. 3, May 1966, p. 139.

2. Ibid., p. 136.

3. Lowdon Wingo, Jr., "Urban Renewal: A Strategy for Information and Analysis," Journal of the American Institute of Planners, Vol. XXXII, No. 3, May 1966, p. 149.

4. The tables plus a discussion of the national model first appeared in Morris B. Goldman, Martin L. Marimot, and Beatrice N. Vaccara, "The Inter-Industry Structure of the United States: A Report on the 1958 Input-Output Study," Survey of Current Business, November 1964, pp. 10-29.

5. An algebraic formulation of the derivation of the inter-regional flow matrix is shown in the appendix to this paper. The process has been programmed and debugged in FORTRAN IV for the IBM 7094, but as of this date has not been documented.

6. Bureau of Labor Statistics, Consumer Expenditures and Income, Supplement 2 to BLS Report 237-90 (Washington D.C.: Government Printing Office, June 1966), pp. 3-9.

APPENDIX

1. Let A = The national intermediate flow matrix showing the shipments and purchases of the eighty industries

   B = The national final demand matrix where the rows are direct demands on industries and the columns are categories of final demand

   C = The regional final demand matrix

   D = The national direct requirements matrix

   E = The first approximation of the regional intermediate flow matrix

   F = The regional intermediate flow matrix

   G = The regional import matrix

   H = The regional export matrix

   J = The rest of nation intermediate flow matrix

   K = The rest of nation final demand matrix

   L = The inter-regional flow matrix

   V = A vector of total industry output levels for the nation

   W = A vector of total industry output levels for the region.

2. Given: A,B,C,D,V, and W

3. Compute E by scalar-vector multiplication:

$$e_{i1} = w_1\, d_{i1}\ ;\quad e_{i2} = w_2\, d_{i2}\ ;\quad e_{i3} = w_3\, d_{i3}\ ;\ \ldots\ldots\ e_{in} = w_n\, d_{in}$$

$$i = 1,2,3,\ldots\ldots 80 \quad \text{(rows)}$$
$$j = 1,2,3,\ldots\ldots 80 \quad \text{(columns)}$$

4. Compute row summations of E and C:

$$Y_i = \operatorname{Sum}_{j=1}^{n} E_{ij} + \operatorname{Sum}_{j=1}^{m} C_{ij} \qquad n=1,2,3,\ldots.80;\ m=1,2,3,\ldots.20$$

Y = A vector of row summations which equal total regional output levels assuming no imports or exports to the rest of the nation.

5. Compute amount of imports and exports by industry based on the difference between the actual and derived total output levels:

$$R_i = Y_i - W_i \qquad i=1,2,3,\ldots\ldots 80$$

if R_i greater than 0, then industry is a net importer

if R_i less than 0, then industry is a net exporter

6. Compute import matrix G for all R_i greater than 0:

$$g_{ij} = \frac{e_{ij}}{\sum_{j=1}^{n} e_{ij}} \text{ times } R_i \qquad n=1,2,3,\ldots\ldots 80$$

7. Compute regional flow matrix F:

$$F = E - G$$

8. Compute export matrix H for all R_i less than 0:

$$h_{ij} = \frac{a_{ij}}{\sum_{j=1}^{n} a_{ij}} \text{ times absolute value } R_i \qquad n=1,2,3,\ldots 80$$

9. Compute rest of nation flow matrix J:

$$J = A - E$$

10. Compute rest of nation final demand K:

$$K = B - C$$

11. Define L as follows:

$$L = \begin{matrix} F & H & C \\ G & J & K \end{matrix}$$

A PROGRAM APPROACH TO INFORMATION SYSTEMS

Richard A. Siegel*

1. INFORMATION SYSTEMS AND MANAGEMENT CONTROL SYSTEMS

While information systems of the modern variety were not developed without a predetermined use, considerable thought is given to how they can be made to do new things and how they can be applied to new purposes. Invention, in a sense, has become the mother of necessity. There is absolutely nothing wrong with this. It would appear to be natural that men who work with these systems should seek new applications and that men who work with other types of systems should inquire what an information system might do for them.

One area of new application that is especially important and fertile is the one concerned with management of public organizations. Information system specialists would like to develop a system of public management that would utilize the potential of their equipment and expertise. They feel that present public management systems, which were designed long before electronic data processing, are not as efficient and effective as they could be if this potential were utilized. The problem, however, appears to lie in finding such a system.

The need for a framework that mates the capability of information systems with the decision making process that is the essence of a public management system was made explicit by Nathan Grundstein at the last annual meeting of this conference. A quote from his paper will make this apparent.

> "It is the triad of urban management and urban planning and information technology that is my concern. A linkage between them is necessary in order to move towards a kind of urban information system, one for which the designation Urban Management Information System is suitably descriptive. I take it that the design of an information system for urban management is the major technical task presently confronting information technologists and urban planners." (3)

Grundstein goes on to explain the kind of problem faced by public managers.

> "The overriding interest of the urban management decision structure is in the fiscal and operating loads and changes in the operating decision rules of the city agencies. The problems of land use management have become critical in terms of their cash flow implications and their revenue-expenditure implications. Urban management is struggling against the diminishing fiscal flexibility of the city. There is a search for replacement equivalents for lost revenue sources and for incremental revenue sources. There is a concern for realizing the promised fiscal benefits from any investment in physical renewal and land use changes. The concern of urban management with agency

*California State Office of Planning

> operating loads and decision rules reflects a concern with the manageability of urban services activity levels. Urban management is expected to alter agency levels of service activity where these function as undesirable constraints on planning-renewal decisions. Demands for any significant lifting of service level activity, if recognized, must lead to the introduction of a form of management planning, which, in turn, will compel changes in the structural bases of management control of service activities. Radical improvement in the effectiveness with which these services are managed cannot otherwise be obtained."(3)

Now, substitute the word "public" for the word "urban" and abstract from the city and urban renewal context and you have an excellent description of the problem faced by officials at all levels of government. There is a general need to stretch scarce fiscal resources across many requirements and demands. This pressure for increased provision of services is also, as Grundstein states, leading to an increasing recognition that the management process itself must be further rationalized. A widespread search is on for management systems that will lead to a better match between objectives and resources devoted to obtaining them and that promote greater efficiency of agency operation.

It is important for the information system specialist to be aware that this indeed is the case. Theorists and practitioners in many fields are doing work that has strong implications for information systems. The information systems discipline cannot alone solve the problem of creating an effective public management system. The information system specialist must become a team member. He must become allied with others who are working on this problem and make the invaluable contributions that he is uniquely capable of offering.

The lack of such realization has led, I believe, to some false starts and wasted effort. Without the prior guidance for search and collection of information offered by a system of management control, efforts by the information specialist are likely to be channeled into such things as comprehensive data banks. The fantastic technical capabilities of data storage and retrieval systems may lead to attempts to collect and store everything unless there is guidance to the few things that ought to be collected. But the world is so varied and there is so much information to collect that this route must be seriously questioned. The information system specialist should probably, therefore, take his cue from developments in other disciplines. The effective public management information system is one that makes a broad management control system work.

The information system component of a public management control system is enormously important. Without good data, properly classified, and available in a reasonable period of time, no management control system can be effective. A classic example that illustrates this point is the difficulty of implementation of the economist's normative theory of the firm. That theory, simply stated, tells the business firm that its maximum profit level of output is one where the extra costs of producing an additional unit is just matched by the extra revenue that will be obtained by selling it. There is no doubt that the theory offers the best possible decision rule for a management that is interested in maximizing profits. The difficulty is now, and always has been, that information or data of the proper kind is not available. Extra, or marginal, cost and revenue are two types of information that are extremely difficult to obtain. The lack of a good information system is thus the key obstacle to implementation of the theory of the firm.

Having suggested that information system designers should work to operationalize a broader public management system, I would like now to offer the Planning, Programming, and Budgeting system as an example of one that offers excellent guidance for future work. Where public management is already implementing the Planning, Programming and Budgeting system, as is the case with the Federal government and the State of California, the management information system designer has a natural framework to channel his efforts. The conceptual structure of the system, however, may offer good guidance for this purpose even if the particular public agency for which the information system specialist is working is not adopting this approach to management.

The Planning, Program and Budgeting system has several elements where information system requirements are particularly acute. In fact, the information demand of Planning, Programming and Budgeting is enormous. Prior public management systems did not require work objectives of an organization to be stated in measurable form. Provision of that information alone is a challenging task for an information system. The structuring of programs is another Planning, Programming and Budgeting system requirement. But this requires information about activities being performed by diverse organizations that might fall under a common program classification. The information system is thus placed under the requirement of assembling knowledge about what is happening throughout the structure of government that shares some common program attribute. Since no government jurisdiction works in a vacuum, the information system also ought to provide knowledge of activities of other governments that impinge upon the program system of the government in question. (6)

But this does not exhaust the information system requirements of Planning, Programming and Budgeting. In fact, it only begins to specify these. The information system will be required to produce data to make programs operative. We will see examples of this requirement below. The information system will also be required to furnish data on human, financial, and physical resources devoted to programs and the association between the quantities of such resources and various levels of program attainment. In addition, it will have to be ready and able to provide specialized data for the numerous investigations that are carried on under the Cost Effectiveness effort associated with a fully operating Planning, Programming and Budgeting system.

A few quotations from Management Memoranda issued by the California Department of Finance may serve to emphasize the very real nature of these needs. Management Memo No. 66-17 to All State Agencies is appropriately entitled, Programming and Budgeting System--Development of Departmental Workload Information Statements. In that memo can be found the following statements, among others.

> "In the preparation of supporting data for a program statement, all pertinent workload information must be stated. The programming and budgeting system must structure the information to permit subsequent application of analytical techniques."
>
> "Workload information will be discussed in three aspects--program output data, workload plans, and resources input data."
>
> "It is anticipated that in many departments, programs will cut across internal organizational lines and that in the production of workload information statements, a new pattern of information flow will arise. It is suggested that this information system be placed on a formal basis to provide reports to management upon the progress of program managers toward the achievement of program objectives." (Underlining by the author.) (1)

From Management Memo 66-14, we have:

> "The advantage of the programming and budgeting system is that it will enable each agency to attain the following objectives more effectively and systematically:
>
> a. Provide top management with more concrete and specific data on which to base decisions.
>
> b. Define more specifically the objectives of government programs.
>
> c. Provide for a more systematic analysis of alternative courses of action for a comparison of the benefits and costs of alternate choices.
>
> d. Produce a full, continuing cost rather than a partial cost estimate of a program or a change in program.

e. Provide for the review of objectives and the conduct of program analyses and accomplishments on a continuing year-round basis rather than on a crowded schedule as under former budget procedures."(2)

Further quotes from California Department of Finance memoranda could be offered to illustrate the important role of the information subsystem of the Planning, Programming and Budgeting system. But that should not be necessary. Instead, one small quote from a Bulletin issued by the Bureau of the Budget shall be offered to illustrate the point that the same recognition exists at the Federal level.

> "...the planning and budget system in each agency should be made to provide more effective information and analysis to assist line managers, the agency head, and the President in judging needs and in deciding on the use of resources and their allocation among competing claims."(8)

Up to this place we have said that there is a need for an information system that will allow public management to make better decisions. Further, it was suggested that information system specialists might profit from working with management control systems devised by other disciplines. Information system specialists could make a great contribution by making these other systems operational. They are all but slight improvements over present practice without the key contribution that must be made by the information system. The Planning, Programming and Budgeting system was then offered as one which gives strong instructions or clear guides to information system requirements.

2. DEFINITION OF AN INFORMATION SYSTEM

It is apparent from what has just been said that the concept of an information system being used here is not a narrow or simple one. On the contrary, a system is called for that can take concepts of a good public management control framework and translate them into data equivalents. This system must, of necessity, be a large and complex one. Something more is envisaged here then the usual organization consisting basically of systems analysts, programmers, machines and machine operators.

Following Hearle and Mason, we may classify an information system into five components, elements, or stages.(4) These are concerned with data input, storage, processing, output, and communication.

The input element of a public management information system is far more than punch card or tape apparatus to read data bits into other elements but is, instead, the entire effort needed to collect or create data and get it into the system. The machines just mentioned are the end point of the input stage. Prior points include search for data and establishment of either routines or one time efforts to collect data.

The input component could easily be the largest, most demanding of resources, and intellectually challenging of the five. The Office of Business Economics of the United States Department of Commerce could easily be classified as mostly an input activity under the conception used here.

The storage, processing, and output elements are the more mechanical part of the information system that retain data for use, manipulate data, and place the data in a form so that it can be apprehended. Without minimizing the complexities involved at each stage I will say no more about them because little beyond their traditional definitions are required for the purposes discussed in this paper.

The communication elements of the information system is the one which transmits data to the user. This element, again, is more than a mechanical device. Printout mechanisms, cathode ray tubes, and other such devices I would classify in the output stage of the information system. The communication stage is the one which places data in the form most meaningful to the user. Work on the communication function, of course, feeds back to other stages. The way in which the output stage presents data, for example, is controlled by prior instruction from the communications stage. Even more basic than that, the communications elements set the task for the input element by specifying the kind of information needed by the user. This element, then, is really responsible for the design of the entire information system. It translates user requirements into staff, equipment and activity in order to meet those requirements.

3. INFORMATION SYSTEMS AND PUBLIC PROGRAMS

Having defined what I mean by an information system, I would next like to discuss in a definite way how the information system takes its cue from the particular public programs that it serves. The point to be made is that the program, by specifying the information requirements, pretty much specifies the information system. To the extent that a program can operate with presently available data, the input stage of the information system may be relatively simple. Complex programs that attempt new things, on the other hand, may emphasize the input stage to the point where it dwarfs the other stages in size and resource utilization. The age of a program may also be important in this context. New programs require new kinds of data. After they have been operating for a while routines will be established to make this data available on a regularly reported basis.

The proposition that the information system is specified by the program does not work backwards. It is hard to go from an information system to a program. Although this point may appear obscure, I believe that it is responsible for the notable lack of success some economists have had in selling state and local governments on regional income and product accounting information systems. These systems, which are national income and product accounting systems writ small, would be indispensable if they were an essential part of an economic stabilization program. But state and local government do not have much to do in the way of economic stabilization and therefore have little need for the information furnished by a regional income and product accounts information system. Sensibly, therefore, they have not underwritten them. The same may be true of comprehensive data banks. An offer to collect, store, and make quickly available to anyone the information now collected by anyone does not really make anyone excited enough to put his funds to work for the purpose. A focus is needed for data collection. A system of public programs provides this focus. Our data or information systems must serve public management by providing it with the information it needs to answer its great questions.

Two examples of the relation between information requirements and public programs may help to clarify the point being made here. Let us first look at the relation between them in the context of the nation's economic stabilization program. We will only attempt to illustrate and will not therefore spell out the relationship comprehensively.

The Employment Act of 1946 specifies that the Federal government will use all practicable means at its disposal to maintain a high level of employment in the country. (5) This does not appear to be unreasonable to us today because we now have a well articulated theory of how the level of output of an economic system is determined. That theory says, briefly, that the level of output is determined by the magnitude and relation of the spending flows into and leakages of income out of the system. If the system is in equilibrium at a point under full employment the theory says to increase a spending flow or decrease a leakage until the new equilibrium is more to the nation's satisfaction; i.e., closer to full employment. This theory offers the first prerequisite to a public program. It suggests what activity government must follow to cure a felt need.

In equation form the theory is expressed as follows:

$$Y = \frac{a + I + X + G}{1 - c+ct - m}$$

Where: Y = An indicator of aggregate output such as Gross National Product

a = A consumption expenditure constant
I = Private investment spending
X = Exports to foreign countries
G = Government spending

c,t,m = Coefficients of an additional income variable representing proportions consumed, taxed away, and spent for imports.

The above equation expresses the spending flows and leakages in a highly aggregated form. In actuality many different types of consumption, investment spending, government spending, taxes and other variables must be delineated for an effective conduct of stabilization activities.

It should be apparent, even in this brief portrayal of the theory, that an effective economic stabilization program demands an enormous amount of information. One can imagine that many of the people working on the early programs required to furnish this information must have despaired at the possibility of ever providing enough. And yet, today, although the economic stabilization program does not have perfect data, its information system provides enough so that data availability is not a blockage to the effective exercise of public stabilization policy.

The economic stabilization program is now fairly well established. For the purpose at hand, however, it is not a really fair example because it is not a program of a subnational U. S. government, it is not one program in a system of programs comprising the Planning, Programming, and Budgeting system, and it is not typical of such programs because it is not one that proves a public good. It was useful for illustrative purposes, however.

Exhibit 1 illustrates structural detail for a program which does meet most of the requirements just mentioned. Air pollution control is undertaken by state and local governments, clean air is a public good, and air pollution control is but one of many programs mounted by these governments. The exhibit was prepared as part of a study by William Spangle and Associates jointly with Harold F. Wise for the State Development Plan program. (7) The original intent of the exhibit was to illustrate how a broad purpose or general objective of government is refined to ever lower levels of abstraction until the activities and tasks involved become operational. We may use it for a different purpose - as a framework that leads to the specification of information requirements.

Let us take an example from the exhibit. Air pollution control can be accomplished partly by preventing pollutants from entering the air. This requires the establishment of controls of and regulations about pollution emissions and it requires the development of a monitoring system. For fixed source control (control of noxious devices fixed in place), this is accomplished through establishment of discharge requirements; development, installation, and maintenance of emission control systems; and the establishment of an enforcement system. Now, imagine the information required to put this phase of an air pollution control system into operation. Aside from information produced only by basic research, the government would have to know the location and complete description of all devices such as incinerators, furnaces, chemical processing operations, and others that produce air pollutants. A complete description would include type of pollutant and quantity of emission per unit of time, degree of necessity for sustained operation of the device, and other such information as name of the owner, expense of control mechanisms, date and findings of inspection of control mechanism devices, and more.

We need not develop a comprehensive listing. The point is, if the information system technologist wishes to be of real use to public management he must gear his system into important ongoing programs. He must do for all programs what the exhibit does, incompletely to be sure, for air pollution. That is, he must find out just what it is programs accomplish and exactly how they operate. Knowing this he is in a position to develop the kinds of basic data and reports required by public managers so that they may operate the program.

4. THE WAY TO COMPREHENSIVE PUBLIC MANAGEMENT INFORMATION SYSTEMS

Provision of information for specific programs may seem like a different thing than a public management information system that enables scarce resources to be allocated optimally between programs. That was the task of an information system outlined at the beginning of this paper. However, that is not really the case because a fairly good knowledge of the relation between benefit and cost for individual programs is required before we are in a position to provide information that will enable management to make tradeoffs between programs. Program tradeoff, that is, adjusting levels of program activity and support by shaving a bit here and adding a bit there, must be based on good estimates of incremental costs and benefits for units of program activity for all programs in the budgetary system. This means that we take the first steps toward the broad, between program adjustment process by first developing information in detail about single programs. Additionally, the trade-off process only makes sense if each individual program is operating in a very efficient manner. It is the present task of public management information systems to help accomplish this.

In closing, I would like to emphasize the usefulness of the Planning, Programming, and Budgeting system to information system specialists. It offers a framework to guide the collection and classification of data. This is really an invaluable service because without such a framework one really does not know what to select that would be most useful to management. Let the program structure provide such guidance and then devote your energies to developing an information system that will provide high quality, relevant, and timely data. This is an enormous challenge.

References

1. California Department of Finance Management Memo No. 66-17, Programming and Budgeting System -- Development of Departmental Workload Information Statements, June 20, 1966.

2. California Department of Finance Management Memo 66-14, Conversion To The Programming and Budgeting System Directed By Governor Brown, May 17, 1966.

3. Grundstein, Nathan D., "Urban Information Systems and Urban Management Decisions and Control", Proceedings of the Third Annual Conference on Urban Planning Information Systems and Programs, Sept., 1965.

4. Hearle, E.F.R. and Mason, R.J., A Data Processing System for State and Local Government (Prentice Hall, Englewood Cliffs; 1963).

5. I am indebted to Marvin Hoffenberg of the Institute for Government and Public Affairs, University of California, at Los Angeles, for some of the ideas expressed in this section.

6. Rowen, Henry S., Statement Before the Special Subcommittee on Scientific Manpower Utilization of the Senate Labor and Public Welfare Committee. Released May 17, 1966.

7. Spangle, W. and Wise, H. State Development Programming, Working Paper IV, Work Item 202.13. Prepared as a component of the California State Development Plan Program.

8. U.S. Bureau of the Budget Bulletin No. 66-3, Planning-Programming-Budgeting, October 12, 1965.

AIR POLLUTION CONTROL PROGRAMS

EXHIBIT I

GENERAL OBJECTIVES

MAINTENANCE AND IMPROVEMENT OF AIR QUALITY
To ensure levels of air quality compatible with the preservation of public health, including prevention of irritation to the senses, interference with visibility and damage to vegetation.

IMPLEMENTING OBJECTIVES

1st LEVEL / 2nd LEVEL / 3rd LEVEL / 4th LEVEL

- **REMOVAL** (1st LEVEL)
  To remove deleterious substances from ambient air.
  - (NATURAL CLEANSING) (2nd LEVEL)
  - ARTIFICIAL CLEANSING (2nd LEVEL)
    (Inoperative)
- **PREVENTION** (1st LEVEL)
  To prevent pollutants from entering the air supply.
  - CONTROLS AND REGULATIONS (2nd LEVEL)
    To control pollutant generation at its source.

    MONITORING
    To provide the basis for an emergency "alert" before levels of hazard or extreme discomfort are reached, to provide a measure of the effectiveness of other programs, and to record data necessary for the development of standards or the design of control programs.
    Cooperative program:
    State/local/federal
    - ESTABLISHMENT OF STANDARDS FOR AMBIENT AIR QUALITY (3rd LEVEL)
      Bureau of Air Sanitation (DPH)
    - FIXED SOURCE CONTROL (3rd LEVEL)
      - ESTABLISHMENT OF DISCHARGE REQUIREMENTS (4th LEVEL)
        Local Air Pollution Control Districts
      - DEVELOPMENT, INSTALLATION AND MAINTENANCE OF CONTROL SYSTEMS (4th LEVEL)
        Private, if applicable
      - ENFORCEMENT (4th LEVEL)
        Local Air Pollution Control Districts
    - MOBILE SOURCE CONTROL (3rd LEVEL)
      - ESTABLISHMENT OF STANDARDS FOR MOTOR VEHICLE EMISSIONS (4th LEVEL)
        State Motor Vehicle Pollution Control Board
        Bureau of Air Sanitation (DPH)
      - DEVELOPMENT, INSTALLATION AND MAINTENANCE OF CONTROL DEVICES (4th LEVEL)
        - a) DEVELOPMENT OF PROTOTYPES
          Private
        - b) APPLIED RESEARCH AND TESTING
          Motor Vehicle Emission Facility (DPH)
        - c) EVALUATION AND APPROVAL
          State Motor Vehicle Pollution Control Board
        - d) INSTALLATION AND MAINTENANCE
          Private, by licensed stations
      - ENFORCEMENT (4th LEVEL)
        California Highway Patrol
        Department of Motor Vehicles
  - RESEARCH AND DEVELOPMENT (2nd LEVEL)
    To develop methods of controlling pollutant discharge from present sources or to develop economically feasible alternative energy sources.
    - RESEARCH AND DEMONSTRATION GRANT PROGRAM (3rd LEVEL)
      U.S. Public Health Service to public and private contractors in California
    - PRIMARY RESEARCH ACTIVITIES (3rd LEVEL)
      - FEDERAL AIR POLLUTION CONTROL LABORATORY (4th LEVEL)
        U.S. Public Health Service
      - UNIVERSITYWIDE AIR POLLUTION RESEARCH CENTER (4th LEVEL)
        University of California at Riverside
      - PRIVATE INDUSTRY (4th LEVEL)
- **DISPERSAL** (1st LEVEL)
  To reduce pollution levels in areas of high concentration such as inversion basins.
  - (NATURAL DISPERSAL) (2nd LEVEL)

    MICROCLIMATE CONTROL
    To increase the rate of movement of existing pollutant loadings to areas where dispersal of pollutants may more readily take place. (through an increase in upward or outward air currents)
    (Inoperative)
  - GUIDING LOCATION (2nd LEVEL)
    To guide location of new pollution sources within a region to areas where dispersal is most effective or where damaging effects are minimal.
    - LAND USE AND ACTIVITY CONTROLS (3rd LEVEL)
      - LOCAL ZONING ORDINANCES (4th LEVEL)

KEYNOTE ADDRESS

Donald H. Blatt *

The mission of a Keynoter is to strike the right note, be it optimistic or pessimistic, comic tragic, fanciful or realistic. He must combine the note, the melody, with the lyrics; in this case, our key is the lyrics--the messages to be presented by our speakers. Audiences leaving the best musical comedies hum the tunes as they stroll up the aisles if the lyricist is a Noel Coward or a Oscar Hammerstein, the audience may sing the words as well.

I have two songs to sing--one is a melody of hope and realism, the other a marching song. The first is an appeal to us all to organize ourselves as a professional group. The second may become the theme of a symphony, the marching song of the management information revolution.

The first song I sing is that of Professionalism:

I. PROFESSIONALISM

Consider, if you will, the proposition that there is need now for a professional organization devoted solely to the proper development of automated systems technology in government.

There is emerging a breadth of knowledge and experience in automated systems technology as applicable to state and local governments. The sources of this new discipline, which we might call Governmental Information Systems Management, include areas such as basic and applied research and the academic approaches, the practitioners, the "systems approach" as evolved in national defense and missile development, and the role of vendors.

Research and the academic approaches have fostered such developments as the Metropolitan Data Center, the several urban data bank projects, such as those related to urban and regional transportation studies, and the library-sponsored information retrieval projects, (such as Urbandoc); plus the creation of curricula including data processing techniques as tools in solutions of problems of involved calculations and the handling of mass data, both economic and demographic.

The increasing numbers of governmental data processing departments and centers (with 315 computers installed) has contributed to available knowledge and experience, while at the same time has created a shortage of qualified personnel with the necessary broad scope of technical skill, managerial competance and governmental orientation. As more and more governmental jurisdictions place computers on order, whether or not in accordance with a Blueprint-

---

* Director of Management Information, Nassau County, New York

Plan for effective implementation, knowledge and experience are gained--in many cases, the hard way. In too many cases, trial and error is proving to be a costly method of developing a data processing capability. Without a profit motivation to insure proper feasibility analysis, governments are yet able to move on, somewhat blinded by the glamour of "our computer."

Many of the large corporations which have been involved in space age, defense agency supported, missile, rocket, weaponry, control systems, tracking, telemetry, and hard and soft base development, have generated tremendous capabilities in what has come to be called "The Systems Approach." The utilization of techniques of reverse-engineering, PERT, CPM, PERT COST, operations research, linear programming and other similar exotic-sounding, but experience proven methods has been generically typified as the systems approach. Federal agencies, starting with the Department of Defense,[1] have proven the usefulness and the possibilities for management improvement in the use of these techniques.[2] Several state and local agencies are requiring PERT networks and computer programs to be filed with the bid proposals on major projects. On their own computers they can then check the reliability of the schedules and cost figures and establish the base for tracking progress. The wealth of experience and knowledge, developed largely at taxpayer expense by these defense-oriented corporations, is available today to state and local governments alert enough to avail themselves of the opportunity.

The vendors have invested untold sums of money, countless man-hours, -days, and -years; and facilities beyond reckoning in the development of hardware and software packages, many of which have been conceived purely for the governmental market, others of which are easily convertible to governmental applications. (As we know, government covers many areas of function and activity which have direct counterparts in business and industry.) One of the major computer companies has had at least 1500 knowledgable people working for years on the development of software for a new family of computers, and this level of effort continues with this company and will be matched by competitive companies as new product announcements are made. Software, as defined in the Automatic Data Processing Glossary of the U. S. Bureau of the Budget, is "the totality of programs and routines used to extend the capabilities of computers, such as compilers, assemblers, narrators,

---

1. Hitch, Economics of Defense in the Nuclear Age.

2. Allan L. Otten, "New-Look Budget"?, The Wall Street Journal, (April 28, 1966), 1.

routines, and subroutines. (Contrasted with hardware)." It is not adequate, however, only to include as vendors the producers and marketers of computer equipment. There are many companies who sell only software--special languages, information retrieval file organization packages[3], systems support and other services classed as software. Among these services we can include the customer education provided free by manufacturers and at a moderate fee by consultant companies.

In order to capitalize the tremendous investment represented by the experience and knowledge in the preceding four categories, an interdisciplinary organization must be created. This organization should have the following objectives:

1. To establish a common base and a communication medium for the exchange of technical knowledge and experience, including the academic, the practitioner, the technical specialist and the vendor.

2. To establish standards of professional competency, organization, systems development, codes and classifications, glossary and practice.

3. To evaluate the need for and encourage a manpower development program to meet present and future needs.

4. To provide an educational organism for curriculum development, for assisting the practitioners and the academic profession, and for the development of public knowledge of the place of Data Processing in government.

5. To provide the means for continuing evaluation of current and emerging concepts for systems, thus establishing a recognized channel for exposure, innovation, and experimentation, insuring a dynamic, viable organization.

---

3. John K. Parker, "Operating a City Databank," Output (the monthly feature of Public Automation), June 1965.

II. MANAGEMENT INFORMATION SYSTEMS

The second of the major notes I wish to strike is that of the Management Information System. As I indicated previously, this will be the marching song of the current generation of practictioners in the information handling field.

Management Information Systems are based on certain assumptions and requires certain characteristics to be practical tools in the operation of governmental units.

1. ASSUMPTIONS

Among the assumptions we can readily make, based on experience, is that governments generally are relatively large and complex operations. They are relatively large in that the local government unit is, in many cases, the largest single employer within the jurisdictions which it serves. Their budgets may be as large as any single business organization within that jurisdiction. They are complex in that they have a multiplicity of production, service, and job oriented operations. They are not a simple manufacturing organization. They tend to combine many of the facets of complex and integrated business organizations.

Another assumption is that essential information for skillful management is lacking. It is unfortunate that in more cases than not in government there has been no planning or anticipation of the information needs. There has been no definition of the requirements for information. Too many times we see government executives "flying by the seat of their pants," a situation where because of a lack of information a series of crises situations occur. With normal planning and anticipation of the need for information for decision making, crisis after crisis can be avoided. In addition there may be no organization in governments to create information by collating or even initially capturing data so that there can be no byproduct intillegence generated by an information system. We find also that there is a lack of experience, both at the management and the supporting level. At the management level the chief executive is usually too busy with day-to-day operations (and possibly his political aspirations) to take the time to define his information requirements. At the support level we find too often a lack of qualified personnel working within an organizational structure that will provide the expertise for proper design of an information system and the collection of the necessary data.

2. THE INFORMATION PROBLEMS

Information handling problems can be classified in many ways. Among these are 1) a lack of problem definition, 2) the lack of system, 3) the lack of desire to utilize information, 4) the lack of ability to utilize information, and 5) lack of effective communication.

In the first of these problem areas, the lack of problem definition, we are aware that unless there is identification of the information needs there can be no definition of the information handling problem and no postulation of a potential solution. The second of these is a corallary to the first--without problem definition there can be no system; in too many cases in local government there is a complete lack of system, much less a desire to create a complex system which will provide for all of the information needs of the government.

The third, a lack of desire to utilize information, is apparent in many cases where a tight rein on the part of the executive means that he is trying to control, individually, all of the activity. I have seen City Managers who fail to give their department heads sufficient information in terms of all of the directives, administrative statements and similar information statements, with a deliberate intent of making the department heads conform directly to the day-to-day desires of the executive. More recently, government executives have been delegating authority and responsibility, but have not provided the information necessary to implement policies thus established.

The fourth of the problems is a lack of ability to utilize information. This lack is apparent at the top level as well as at the middle management levels, and it is a result of a lack of a professional attitude in the government field as well as a lack of education and qualification for the positions.

The fifth, a lack of effective communication, applies both vertically and horizontally. In many governmental organizations, the only line of communication is down from the top. There is not permitted to be an input from the lower levels and the vertical communication is one directional.

This is also true in many government areas horizontally, in that one department head has no knowledge of the activities of the other department head and is not encouraged by any means to acquaint himself with the problems of his peers and associates.

There are undoubtedly other information handling problems which could be discussed, some of which are of a technical nature. However, it is sufficient for the purpose of this discussion to indicate these five as the major problems in the information handling efforts in government. And this is true to some extent in all levels of government including the Federal.

3. BASIC REQUIREMENTS FOR SOLUTION

There are certain basic requirements for solution to the information handling problems as cited above. Among these are a favorable management climate, resources and systems.

In order for there to be a favorable management climate for the solution of information problems, top management must be aware of the need for information. Top management needs information for the purpose of determining status, such as project status, time considerations, number of personnel, and budget status. It needs information for planning purposes, control purposes and most of all for evaluation purposes, to determine whether or not there is efficient operation and whether the service requirements of the community are being met.

Top management must identify types of information required and establish priorities and schedules. It is not sufficient to say "We need information." Care must be used to define types of information by usage, such as for the purpose of status reporting, planning, control or evaluation. Priorities are necessary because it is not possible under most circumstances to burst forth full bloom with a complete information system. Schedules are required to insure the proper implementation based on priorities.

In order to complete the requirements for establishing the favorable management climate, top management must recognize its responsibility for providing resources in terms of people and money. Not only must there be people at the lowest level collecting information, but there must be people available to design and implement and operate the information systems.

And again a corallary--money must be provided to support not only the people but the equipment necessary for this purpose. Money for the rental of computers, key punches, and other devices that provide elements of the system must be made available.

This climate must be expressed not just in a general sense of saying, "Well, we support the budget," but in terms of policy statements which set forth goals and in general the means of achieving these goals. In addition to making the initial policy statement, however, top management must continue to keep an eye on the development of systems and must periodically review progress so as to determine whether or not the goals are being achieved.

4. RESOURCES AVAILABLE

There are resources available in government for proper solution to information handling problems. Among these resources is the existing organization for production and service. We have a built-in operative network with a mission of providing service to the public on a generally specified level. There is also available the "Systems Approach" to information processing which uses people such as analysts and programmers and machinery such as computers and other equipment to provide the operating system which can interpret data in terms of information and intelligence.

Given the proper organization and the systems approach, it is possible to use basic operational data to create an effective management information system. In fact the management information system can be largely a by-product of the basic operational data which is created during and from normal operations.

There is, however, an additional data generation capability available on the same base of organization and systems approach. It is possible to provide for the creation and collection of additional data, which in itself is not expensive, in order to give a much greater amplitude to the intelligence. It is the intelligence which is the frosting on the cake, making possible interpretation and evaluation as an effective element of the information system.

5. ELEMENTS OF SYSTEM

There are certain requirements for any management information system. Among these requirements is the question of responsiveness to service demands and goals at various levels. It is vital that obligations be met by any system and if the service is based on a legislative requirement, this is readily apparent. However, in many cases, legal obligations for service are not that easily recognizable and must be determined so that the level of responsiveness should and must be part of a good system.

The system should provide communications both vertically and horizontally. The vertical orientation should be such so that work done at the lowest level can be reported at the highest and that policy statements and instructions given at higher levels can be translated in actual performance. The system of work orders, for example, should provide for this vertical communication. In addition, there is horizontal communication, using perhaps a data bank concept, where information collected in one activity in one functional agency, can be utilized by another activity in the same or a different functional agency. For example, information collected for police purposes may be used for probation purposes as well. A good system therefore would provide retrieval of common data for mutual use and accomplish this requirement of horizontal communication.

A good management information system must be flexible to the point of being adaptable to internal restructuring. In fact, the development of a management information system may well result in internal restructuring as a result of better identification of activities and programs. It should be adaptable to external restructuring when new programs require greater depth of information, or when summarization for reporting, such as FBI reports on the part of the Police Department, make it necessary to enlarge the scope of information gathering. In addition to being flexible in the sense above mentioned, the system should be heuristic in nature so that it is possible for the system to be improved through experience and learning. As additional information is created, additional information capability may be generated. This is true, of course, whether the operation of the system be centralized or decentralized and whether the control of the system is centralized or decentralized.

The management information system, although flexible, must be disciplined. In other words there must be a format by which actions are reported so as to guarantee conformity to functional organization as related to service. This requires planning, forecasting, and reporting against the forecast--all of which are useful disciplines in improving management and performance; hence, service to the public.

It should provide analytical capability, such as performance evaluation in both quantitative and qualititative form. For example, it is not sufficient to say that we have swept a hundred thousand miles of streets in a given budget year unless there is known that there are only one hundred thousand miles of streets to be swept. It is much more significant to say by analysis of work reports versus planned program that all streets that needed sweeping were swept when they needed it. This is quantitative and qualitative analysis and evaluation.

The system, of course, must be stable and justifiable. If, for example, a cost accounting system must be created for a basis for the management information system, then the cost of developing the cost accounting system cannot necessarily be charged to the management information system development or the cost might be indefensible. It is probably sufficient to identify the need for a cost accounting system as a good general practice rather than charging it to the management information system directly. However, it is possible to determine the direct cost of information handling and in every case the cost justification should be attempted.

The management information system should be control oriented and therefore be self-policing. In terms of performance and production there are quantitative factors which can establish controls. In terms of inspection there can be qualitative and quantitative factors. In terms of finance, they may be only quantitative. And in terms of planning, there are qualitative factors which can make it possible for the system to be self-evaluating from the control standpoint.

The management information should be financially oriented and budget based; and especially, it must be capable of budget formulation in succeeding budget periods. Because of the financial orientation and because of the nature of government, any system worth its salt must provide for development of working standards and units of work which are applicable to the succeeding budget periods. The properly designed system will automatically create this information and forecast it for succeeding budget periods.

The successful management information system must be production oriented based on standards and on quality control analysis. It is definitely not sufficient to turn out large quantities if the quality of service is the essential. Criteria for standards can be determined. If grass is to be cut, then it is not good practice to cut the grass every week, but to cut the grass only when the grass needs cutting. Standards and quality control can tell the successful manager through the management information system what his performance should be and must be.

The system should be planning oriented so as to provide forecasting of work programs and allow for the assignment of personnel and other resources. It must forecast in terms of long range budgets as well as short-range operational activity for direct production.

The system should provide the data base for management science (operations research) approaches, such as simulations, gaming, linear programming, allocation and others. This becomes a by-product which is sometimes difficult to cost-justify. It is essential for truly scientific management in government.

Above all the management information system must provide a decision base. It should not be a decision maker in itself, but should provide the information and intelligence which the executive working in the proper environment of management climate with the proper resources and the proper system can utilize to improve the entire quality and quantity of service rendered to his public.

III. EXAMPLE--NASSAU COUNTY

As an example of a Planning, Programming, Budgeting System on the local government level, I submit the Nassau County Management Information System, now successfully through the phase of functional organization, system design and system pilot testing. This system is programmed in Autocoder for IBM 1401 tape and is operative. Succeeding phases, over the next one and one-half years will carry the system into the realm of sophistication, represented by the management science techniques suggested above.

The Nassau County Management Information System is essentially a work oriented budget performance program which has been designed specifically as a general purpose system; that is the data fields collected are common to any function or agency of government. The information assembled will accurately report on the work performance and expenditures in various agencies, since specific identification is accomplished by way of a classification matrix (an organization-functional coding scheme). The type and measurement of work is a matter of defining the units of measure, the activity, the program, etc. This approach promotes maximum flexibility for system expansion.

Publication of the statistics of the Nassau County Management Information System as of June 1966 is available on request from J. Ward Wright, Deputy County Executive for the County of Nassau, Executive Building, 1 West Street, Mineola, New York. This is a bound volume incorporating approximately sixty pages of text and exhibits which illustrate the methodology employed in developing the system used as an example in this presentation.

THE ROLE OF MODELS IN SETTING VALUES

by Claude Gruen*

A recent commentary on contemporary systems designs concludes:

> "Its greatest vigor stems from a dissatisfaction with the limitations of man's existing control over his physical environment. Its greatest threat consists precisely in its potential as a means for extending the control of man over man." [1]

Dr. Robert Boguslaw reaches this conclusion after an analysis which argues that the computer-oriented techniques of the systems designer inevitably result in the usurpation of the democratic process which obtains a representation of the values held by the individual members of our society. Systems analysts who model portions of the urban environment can dissociate themselves from his criticism by pointing out that models merely replicate urban systems and do not prescribe alterations in these systems. We can agree with the statement of Dr. Charles J. Hitch and cling to our position as non-normative scientists.

> "The job of the systems analyst is to free the decision-maker from questions which can best be resolved through analysis, leaving to him those more difficult questions which can only be resolved on the basis of judgement." [2]

Nevertheless, if urban models become meaningful planning tools, they will affect the setting of the values which determine the decision-makers' judgments. If models are good enough to be used in public decision making they will make value alternatives explicit, and widen the range of choice open to the decision-makers, but the models themselves will serve any master. They can be used to give the electorate more explicit choice. They can also be used to justify a decision-making monopoly by urban rulers who will be able to claim a new version of the "divine right of kings", a right based on their exclusive ability to communicate with the models.

If the models are to strengthen the democratic process, while concomitantly improving the rationality of public decisions, model builders and users must be aware of the role these naescent tools will play in setting values. Model builders should not take a scientific fifth amendment on the subject of values. Instead, the design of models, and the training of model users must begin to establish a tradition of use that will improve the functioning of our democratic political system. The major criterion of such a system is that it permits the individual members of a society to influence those who make decisions to implement public courses of action whose results will conform to the values held by the individuals.

---

* Urban Economist, Arthur D. Little, Inc. San Francisco, California

Professor Alfred Kuhn defines values as, "... the relative position of a good in a preference ordering, and the higher its position the greater its value.." [3] The meaning he places on the word good encompasses any external which provides utility. "Goods include material things from lollypops to the St. Lawrence Seaway, and non-material things from a word of praise or the scratching of an itch, to the operation of The United Nations or a national security system." [4] Kuhn's definition is an operational version of the definition recommended by a group concerned with defining values that relate to governmental decisions, ".. the normative standards by which human beings are influenced in their choice among alternative courses of action which they perceive.." [5]

The urban political scene has decision-making politicians and parties who seek the endorsement of individual citizens. We do have a democratic system in that it permits the members of society to influence those who make decisions. However, this system is inefficient in the manner it provides information concerning the potentially available public actions and policies for altering the urban environment, and frequently inaccurate in the manner it provides the information needed to evaluate the effects of these alternatives. These information deficiencies inhibit the formulation of action preceding consensus values that reflect the individual values of the urban citizenry.

Let me add, that the problem of discerning and disseminating the needed information has been with man since he began living in groups. Computer models are not likely to put an end to this problem, but they do offer the promise of being able to improve the quality and quantity of the information needed to assign priorities to those actions with a high value-to-cost ratio. At a minimum, we should be able to provide the information required to follow the general rule of decision or rationality stated by Professor Kuhn. "The rule is, starting from any given point, total satisfaction will be increased if a decision is made in favor of any alternative whose value exceeds its costs, when both value and costs are measured forward from the moment the decision is effective." [6] We can stipulate two informational requirements of a democratic political system capable of following this rule:

1. The body politic must be provided with information concerning a wide range of alternate actions or policies. To a great extent the degree of freedom enjoyed by a society is a function of the number of choices open to the citizens of that society.

2. The system must be provided with accurate information relating the costs and benefits of each alternative to the value-sensitive conditions of the urban environment they affect. Meaningful values cannot be asserted in an information vacuum. Individuals cannot impose their values on their elected official unless they have intelligence concerning the impact that alternatives will have on the things they value.

The realities of decision making in our large urban communities do not include unidirectional flows of intelligence concerning values from the voter to the elected official to the planner. Elected officials tend very strongly to shy away from the delineation of an ordinally arrayed and comprehensive set of values based on their perceptions of voter preferences. City planners are not handed a list of values which they can easily convert to goals which can then serve as a basis for the formulation of objectives, with appropriately arranged priorities.

Instead, the urban planner has the choice of presenting plans which he thinks will be accepted and implementable, or stewing in his own frustrations while he futilely waits for a guiding set of previously determined values. I do not suggest that we attempt to change this situation. I do suggest that models can be used to permit the planner and others to do more planning; that is, to present the body politic with more alternatives from which to select.

Only a large number of available alternatives can effectively permit the selection of programs that will respond to the values held by the individuals within the community. If the planner submits only a single plan he has effectively precluded the selection of all preference orderings not included within that plan. The more alternatives he offers, the more likely he is to include those that correspond to the values of the community.

However, the conception of urban plans and policies designed to improve the conditions of life in urban areas is difficult because the factors that produce and maintain these conditions are complex. More plans and policies can be conceived if planners utilize models to augment their own insights and experiences.

If models are to be utilized in this fashion they must do considerably more than forecast the experiences of the past into the future. They must describe the components and relationships of the various subsystems that constitute the urban environment in a manner that will facilitate the conceptualization of public programs. Most importantly, we must begin to model the difficult to replicate, but crucial, social subsystems of the urban whole.

Dr. Ira M. Robinson presented a paper at this conference in 1964 that discussed the development of a simulation model for renewal programing.[7] This model deals with the housing subsystem contained within the urban environment. It replicates those factors that determine housing unit quality, quantity, location, price and type within a geographical unit.[8] It is currently being used by the San Francisco Department of City Planning to test the relative impact of annually suggested changes to the city's development program.

This current use is continuing to exert the same kind of discipline on the model-using team that was present in the model-building effort. The model builders were forced to identify and describe the significant components and relationships of the San Francisco housing market and assign them operational characteristics. These operating characteristics were not necessarily linear cause and effect relationships; instead, many of the most significant were feedback relationships indicating mutual causation. Thus, for example, the relative demand for housing in a neighborhood is influenced by the particular socio-economic groups who are attracted to the neighborhood, while the groups are attracted to neighborhoods partially on the basis of the socio-economic groups currently inhabiting that neighborhood.

The model-using team is constantly retracing the system described by the model in order to interpret the results of model runs. At the same time, they are maintaining the data required to check on the accuracy of the model's predictions, and refine its data inputs. The first of these efforts focuses attention on the key "valves" or linkages that influence changes in urban housing conditions, while the second economically screens out data with a high information-to-cost ratio from the total universe of potentially available statistics concerning the state of the urban environment. Thereby, they facilitate the formulation of a large number of alternative plans and programs for altering housing conditions.

However, the mere proliferation of plans and programs will not improve the workings of a democratic political system unless, as indicated previously, each alternative can be evaluated in terms of the urban conditions it will affect. This is the greatest imperfection of our current method of urban decision making. We frequently fail to evaluate the impact of proposed plans or programs with a sufficient degree of accuracy or comprehensiveness. Perhaps the greatest potential of urban models lies in their ability to pretest the impact of alternatives on a wide range of urban conditions; that is, to explicitly state the effects of each alternative on all the aspects of the urban environment that will be altered in any meaningful manner.

Such evaluations are needed if limited public resources are to be expended on the ameliorization of urban conditions that rank high on the value scale of urban citizens before resources are diverted to less urgent conditions. Equally important, although slightly less obvious, is the need to evaluate each plan or program by considering its effect on all of the conditions within the urban environment that will be altered if the alternative is selected by the decision makers. The complex ecology of the urban environment (or system) predestines the results of each alternative to entail the worsening of some urban conditions and the strengthening of others in varying degrees.

The elected official seeking popular support will be tempted to obscure such complexities by identifying the plan or policy he supports with the improvement of broadly stated sets of urban conditions. Similarly, the planner who wants his plan implemented will tend to ignore all but the positive effects of a proposed action. Together, they can assign goals that are so totally inclusive as to be meaningless. Thus, for example, the original advocates of our slum clearance programs singled out for attention its immediate impact on housing conditions, or said merely that it will be good for the city and reduce unemployment. Such evaluations cannot provide a basis for decisions that will be responsive to the ordinally arrayed preferences of the individual citizens.

Instead, needed evaluations should have included the long-run increase in housing deterioration and rent that will be forthcoming in neighborhoods not directly affected by the clearance programs. The ADL simulation model of the San Francisco housing market describes this indirect effect on housing quality as "accelerated ageing".[9] The abbreviation "acc.ag." is printed out whenever a public action being tested in the model creates the relationships that induce incipient blight.

The model recognizes these relationships because it is programed to go through a series of steps that will identify such a situation. The validity of these steps is dependent upon the model builder's accurate conceptualization of the urban system that determines the quality of the housing stock. The model is programed so that it identifies, in each time period, the location and types of houses that are underutilized by those who can afford them, but are highly desired by households whose rent-paying ability fails to qualify them by a slim margin. It then determines whether the owners of such units would make a greater financial profit if they undermaintained the property so that it would "age" into a lower-condition category. In the event that all the necessary prerequisites are fulfilled, the model types our "acc.ag." next to the print-out tabulation of the appropriate housing type and location. This is what happens when slum clearance programs are simulated without the concomitant provision of new housing. The model also indicates the concurrent increase of certain rents when it is being utilized to pretest programs of this type.

The planner who receives such evaluations for alternative plans faces a difficult task when he is asked to make a recommendation among alternative plans. He can now ordinally array the effects attached to each plan. Some of the urban conditions included within this array will be greatly altered in positive magnitudes, and thus be near the top of his list. Still others will remain virtually untouched, and be in the center of his list, while the array will disclose that certain conditions will be negatively affected if the plan is adopted.

The existence of such an array, even if only loosely quantified, permits a comparison between the values of each alternative and those of the electorate, but it is highly unlikely that each plan can be presented to the voters. This leaves the planner to recommend, and the official to select, from among the alternatives on the basis of the values they feel are appropriate. Their decision explicitly sets the effective values of the community for a particular point in time.

If the values that are established in this manner are not the consensus values of the community, and if the decision makers cannot convince the community that they should be, then the next election will usher in a new group of elected officials. Presumably, the new group of elected officials will once again use the model to evaluate alternatives and be forced to explicitly state the values that form the basis for the judging of alternate plans. Thus, the political process would meet the criterion of a democratic system with increased efficiency because of the information provided by using urban models.

Unfortunately, the model builder cannot merely proceed to build more sophisticated replicas of a wider assortment of urban subsystems while feeling assured that his products will be put to such beneficial use. He must work for such use by beginning to develop a tradition that will encourage public access to model results within an atmosphere that encourages a variety of modeling efforts.

Boguslaw's prediction can come true if we do not work to avoid it. The computer model can serve to give the public more real choice, but it can also provide a sophistic device to a new breed of dictators who will impose their values on the urban community. This can happen if we permit the public to believe that computers can conceptualize or select plans or programs. We must educate the public to suspect those who will seek to emulate the Wizard of Oz and utilize the blinking lights of the computer to shroud their own commands to the public. We can do this by admitting the limitations of the models and clarifying the role of the model builder as a theorist and the model user as an interpreter. The public must be taught to look for the wizard.

However, even if we develop a tradition of model uses based on their ability to assist humans in the formulation and evaluation of plans, there is still another danger. Not only can liars build models, but models do lie. The dependability of every model is a function of the builder's ability to accurately abstract a real work system or set of systems. Such abstractions are particularly difficult when we deal with the social systems that form crucial parts of the urban environment. Therefore, proposed plans should be subject to multiple evaluations and not consigned to a single model, or model user. Furthermore, a variety of models should be utilized to help formulate alternatives.

Perhaps each political party should have a set of urban models, or be allowed to send their own analysts to examine the public models controlled by the incumbants. We must begin to devise institutional constraints designed to encourage the use of models to improve the workings of the democratic process.

Models are just beginning to be used by urban planners and decision makers. They can be used to aid in the conceptualization of innovative plans and programs that will be capable of responding to the values held by urban citizens. They can be used to furnish information about alternative public actions that will permit an explicit setting of judgmental values. These values can be compared with those held by the individual members of the community through the workings of the body politic.

REFERENCES

1. Robert Boguslaw, The New Utopians, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965, pp. 204

2. Charles J. Hitch, Decision-Making for Defense, Berkeley and Los Angeles: University of California Press, 1965, pp.76

3. Alfred Kuhn, The Study of Society - A Unified Approach, Homewood, Illinois: Richard D. Irwin, Inc., and The Dorsey Press, Inc. 1963, pp. 266

4. Ibid pp. 261

5. E. Jacob, James J. Flint, Hedvah L. Shuchman, Values and Their Function In Decision-Making, The American Behavioral Scientist, Vol. V Supplement, May, 1962, No. 9, pp. 10

6. Op cit pp. 274

7. Ira M. Robinson, A Simulation Model of the Residential Space Market in San Francisco, Proceeding of the Second Annual Conference on Urban Planning and Information Systems and Programs, September, 1964, The University of Pittsburgh

8. Cf. Technical Paper No. 8, Model of San Francisco Housing Market, Arthur D. Little, Inc., San Francisco

9. Ibid, pp. 31,32

THE PRESENT AND FUTURES OF URBAN LAND USE MODELS

David R. Seidman*

1. BRIEF DESCRIPTION OF THE ACTIVITIES ALLOCATION MODEL

I would like to begin by describing and discussing the mathematical land use forecasting model we have been working on at the Delaware Valley Regional Planning Commission. We worked for a considerable amount of time on a concept known as the Regional Growth Model, which was intended to simulate the basic underlying economic and social structure relating to the locational or land consumption behavior of households and firms. This effort had to be tabled because of a lack of data of sufficient scope and detail, and lack of sufficient time to develop and test some of the supporting theory.

Beginning around the middle of 1964, we directed our efforts towards what we now call the Activities Allocation Model, a series of seven sub-models, most of which are regression equations. Each of these sub-models determines either the location of a given type of activity--residential, industrial, or commercial--or the amount of land it uses. The model begins with the state of the region in 1960 as defined by a series of variables measured for that year. It proceeds with a series of 5-year recursive steps to 1985; that is, it projects 1965 from 1960 and then 1970 from its 1965 projection, and so forth. Within each 5-year projection it goes through a fixed sequence of the seven sub-models.

For each 5-year recursion, the seven models are run in the sequence below. The models each make predictions by individual districts, of which there are 192 in the nine-county, 4-million persons, Philadelphia metropolitan area being forecast.

1) The residential location model, which determines (for each district) the number of households at the end of each 5-year projection period, by four income classes.
2) The residential space consumption model, which determines the amount of residential land consumed by newly located households.
3) The manufacturing location model, with two employment categories.
4) The non-manufacturing location model, with 13 employment categories.
5) The manufacturing space consumption model.

---

*Delaware Valley Regional Planning Commission

6) The non-manufacturing space consumption model.

and 7) The street area model, which determines the amount of new street area constructed in consequence of other land development.

Since the model was constructed to provide activity distribution and land use forecasts for use in the transportation simulation process, it was considered important to allow the model to be sensitive to the alternative transportation policies being tested by the transportation simulation models.

The method of tying transportation policies to the land use forecasting procedure is via the concept of accessibility. The term "accessibility" has a special and specific meaning in this context. It denotes the ease with which opportunities of a given type throughout the region can be reached from a particular district, and is essentially the average cost-distance travelled (i. e., distance in terms of total time and operating cost) in making a trip to a given type of activity via a given mode of travel. Note that this is a measure of accessibility from one district to the <u>entire region</u>; it is not a measure of the ease of getting from one district to some other district.

Most of the models consist of a non-linear equation in which the parameters are determined by a least-squares technique, using 1950 and 1960 data for this calibration phase. A given model may have several dependent variables, in which case the parameters are independently derived for each of the dependent variables; no simultaneous equation solutions are involved.

2. SUMMARY OF CALIBRATION RESULTS

Initial calibrations were made using quite a large number of independent variables, transformed by a component analysis. (Component analysis is a special case of factor analysis used to obtain a statistically uncorrelated set of independent variables from the original correlated--or collinear--variables.) We found that component analysis did not solve the collinearity problems sufficiently, and we finally wound up with between five and six independent variables in the residential location, residential space consumption, and manufacturing location models; and two independent variables in the non-manufacturing location model. The variables turned out to be many of the traditional ones such as proportion of available land, net residential and manufacturing density, accessibility to households and employment, and proportion of high-income families.

The measure of accuracy used for the calibration runs was R^2, the correlation coefficient squared, also known as the proportion of explained variance. This measure specifies the degree to which the model successfully reproduced the variations between one district and another in the values of the dependent variable in 1960. For the residential and manufacturing location models the measure was based on the amount of growth successfully predicted between 1950 and 1960. For the non-residential location model and the residential space consumption model, the measure was based on the amount of the total pattern in 1960 successfully explained, since neither of these latter models were trend based. The residential location and manufacturing location models achieved R^2's of explained growth of roughly .5 to .7, computed using the 192 districts. The residential space consumption model had an R^2 of .8 . The non-residential location model achieved R^2's, on the total pattern, of between .85 and .95, computed using 69 superdistricts.

3. SUMMARY OF PROJECTION RESULTS

Our projection runs are now virtually complete, and I can report some of the findings of these runs. Our experience has shown that regression models can provide quite rational results at the ring level, and generally at the sector level, of analysis. The results obtained from the model presently are probably about as good as projections obtainable from the best judgments of planners in the region. They have the further advantage of having explicit statements concerning all assumptions of the form and type of interrelationships, all specifications of the relative influence of factors, and all statements of policies inputed, so that the process is reproducible. Furthermore, once the model was programmed and calibrated it was able to provide alternative projections in a very short amount of time. The total amount of time to run one set of projections from 1960 to 1985 was approximately 50 minutes on the IBM 7094 at a cost of roughly $300.

Despite the overall rationality of the outputs, there were some unexpected results which raise or reopen some interesting questions. First, we were surprised to discover that there were no significant differences between the land use patterns forecasted for the different transportation plans and policies inputed. That is, even at the district level most differences were no greater than 5 per cent.

One might conclude from this that the combined relative insensitivity of land use development to the transportation system, and the political constraints narrowing the range of transportation alternatives were such that this small amount of variation in the model represents the true situation. Or one might instead conclude that the model has used accessibility measures which are not adequately sensitive to changes in transportation policy and has selected incorrectly small values for the regression parameters specifying the importance of accessibility.

We did attempt to construct accessibility measures which related to actual trip-making behavior in such a way as to adequately express the sensitivity of trip-makers to changes in the network. Nevertheless the accessibility measures derived were very highly correlated with distance to the Philadelphia central business district, and there were only 10 per cent differences, by and large, between the accessibilities afforded by the extreme alternatives in the 1985 transportation plans. Since we have not yet developed an adequate explanation for the relatively low parameter weights and the relative insensitivity of the accessibility measures, we cannot assert which of these possible conclusions should be drawn. We have assumed as a working hypothesis that the model is telling us something valid and significant.

We were also somewhat surprised to discover that the model wished to project a migration from the central cities of very significant proportions. It was prevented from doing so by various checks built into the projection program which were based on past histories of relative growth and decline. Nevertheless, this again raises two important questions: First, would such declines occur if it were not for the massive intervention of public policy to prevent such projected migrations? Second, what is the best manner of dealing with such changes in public policy when using a model based and calibrated on past trends? Interestingly, at least two other regional agencies have experienced the same difficulty of having their models project outmigrations from the central cities which were not credible or politically acceptable.

Finally, although results at the ring and sector level were generally rational, irrationally large variations were found in the projected locations of residences and employment at the areal level of the 192 districts. Contiguous districts often had unacceptably large variations in their relative growth rates, and it was necessary to constrain projections at the district level by the use of various growth rate feasibility checks based on statistical patterns found in the past.

4. REVIEW OF OTHER WORK DONE ON LAND USE MODELS

Having gone over the results of our efforts at Delaware Valley, I would like to discuss briefly the results of some other researchers in the field. At least four other major efforts have been made to construct regression models of urban growth, of which three were brought to the point where they could provide useful projection output.[1,3,6,7] I would surmise from this past work that using 2000 grids to calibrate on, as was done by Chapin[1], yielded highly significant and rather stable parameters, but an extremely difficult problem in projection because of the enormous variations in the predicted values for various grids. On the other hand, Lakshmanan[7] found that using 20-odd planning areas to calibrate on yielded rational projection output, but the collinearities at this high a level of areal aggregation allowed him to use at the most three independent variables. Certainly Hill[6] and we at DVRPC encountered both these problems using the 150 to 190 districts which we calibrated and projected on. This dilemma is an inherent problem of regression models that has no entirely satisfactory solution.

One might in some cases calibrate at a lower areal level than one projects at, or aggregate the projected output to higher areal levels. I believe it will probably prove most effective when projecting land use patterns by regression models to use these models for projecting at an areal level of, say, 50 subareas in a region like Philadelphia, and then use a much simpler assumption-based, proportioning type of model to prorate these totals to traffic zones or whatever level is necessary for the task at hand.

Perhaps an even larger amount of work has been done on a type of model which I will call, for lack of a better name, behavioral.[8,9,10,11,12] The main distinction I make between regression and behavioral models is that the regression models allow great flexibility in the selection of parameters, and assume simple linear or non-linear relationships between the dependent and independent variables, whereas behavioral models generally involve considerably more complex structuring of the relationships of the variables, and inject considerably greater amounts of the individual analyst's views on the extent and manner in which the independent variables influence the dependents. This distinction is by no means clear-cut, especially when one sees, for example, a model which assumes a quite complicated linear programming structure of decision making, but then uses regression analysis and component analysis on a very large number of independent variables, leaving a great deal of parameter flexibility.[4] Nevertheless, behavioral models have tended more to have parameters assumed or calculated using secondary data sources. The general difficulty encountered by behavioral modelists is that it is in fact difficult to obtain parameters one is confident in, because of the lack of data or satisfactory estimating methods. A primary concern of these model builders has been to find better means of obtaining parameters which represent the region's behavior.

5. POSSIBLE DIRECTIONS OF FUTURE MODELING EFFORTS

Considering the types of difficulties encountered by the regression modelists and the behavior modelists as I have discerned them, I foresee a drawing together of the two types of modeling efforts, in which the regression models are no longer allowed the extreme flexibility of parameter choice that they have had up until now, and will be infused with a greater structural complexity more closely resembling the manner in which urban analysts view the urban system. I see the behavioral modelists using more of the regression types of methods to obtain certain parameters needed. I will give an example of one of the types of synthesis I have in mind:

Say we wish to construct an equilibrium model which predicts the number of clothing stores in a district as a function of the district's accessibility to households and the total number of office employees in the area. We can construct the model with a structure in which we have parameters specifying (1) the rate at which shopping trips starting from households attenuate as a function of distance from home to store; (2) the disproportionate advantage of large centers over small ones; (3) the relative importance to shopping of households vis-a-vis office employees, and so forth. Using the pour-in-kettle-and-stir regression approach, these various parameters can be solved for simultaneously using a least squares criterion which selects that set of parameters which best predicts the actual distribution of clothing stores during the year the calibration data was collected. Doing so would most probably result, however, in a set of relatively unstable parameters, some of whose values could not be considered rational.

Instead of the above traditional approach, we could first use origin and destination data on shopping trips to obtain the attenuation rate parameter. Next we could project the number of trips to various districts using this parameter, and determine, from the deviations between predicted and actual number of trips, what the advantage of large shopping centers over small ones appeared to be, and relate this by a simple regression to the gross density of retail activity in the district. Alternatively we could use an attitude survey of shoppers to determine the factors they consider most attractive in shopping centers, and construct from this a composite measure of attractiveness which is used in lieu of gross density. Finally, having solved for these two parameters, we might use a simple regression on accessibility to households and density of office employees to obtain parameters of their relative weights, or alternatively use data on retail revenue as a function of time of day, or data obtained from questionnaires of customers, to obtain this weight.

The advantage to this sequential method of solving for parameters is that it is likely to yield much smaller amounts of instability in the parameters than that obtained using the first method. Furthermore, to the extent that non-regression methods can be used to obtain parameter values, one has a better indication of the places where the theory is inadequate, since regression solutions for parameters can often conceal incorrect theory by choosing parameter values which compensate to some extent for factors left out or structured wrongly.

6. THE ROLE OF THE SHORT-RANGE MODEL

Let me conclude by discussing one other dichotomy of land use model building, the significance of which I believe has been somewhat overlooked. This is the difference between long-range and short-range forecasting models. (In this context we may consider short-range models to be ones dealing with time periods of one to ten years.) The lack of emphasis on this dichotomy is understandable, since the agencies conducting work on land use models have had as their primary responsibility long-range forecasting, generally forecasts for 20 or 25 years for comprehensive transportation planning. To my knowledge there have not been regionally oriented agencies strongly funded for research, which were responsible for programs requiring short-range forecasting models.

One might expect that the long-range forecasting models could be used as short-range models by using a smaller interval, or just one recursion, since most of them are recursive models projecting in a series of steps. This is true, but conceals the important fact that short-range models can have a much greater flexibility in the choice of structure and especially of variables which they use to make their forecasts. For example, the structure of the short-range forecast model could be based on linear or non-linear trend lines calculated for each small area, which are then modified by changes in various independent variables. Residential growth might, for instance, be projected by a trend line which is increased in slope if accessibility becomes greater, or decreased in slope if available land is significantly lessened.

There are many independent variables which can be used in short-range models and not long-range ones, because long-range models--being recursive--must be able to predict the values of the independent variables in the future. And many variables which strongly influence short-run growth are themselves very difficult to project into the future. Examples of these are the unemployment rate, vacancy rate, quality of school, close proximity of new shopping centers, new transit stops, new elementary schools, etc.

Naturally, in a short-range model special care must be taken to allow the inclusion of new development which is certain to occur, such as those projects which are specified in the capital improvement program. Since most of the larger of these projects are projected forward for between five and ten years, it is likely that the short-range model will be directed much more towards forecasting the occurrences in the private sector. This is in contrast to the long-range model, which must be able to predict events in both the private and public sectors. Thus, there is, in fact, quite a difference in the best approach to constructing short-range and long-range models.

Since there is great potential value in the construction of more sophisticated and accurate short-range forecasting models for use in construction of utilities, retail stores, post offices, elementary schools, and so forth, I hope that foundations and Federal agencies can be found to support research for the application of better techniques to the construction of short-range forecasting models, as they have supported such research in the construction of long-range forecasting models.

7. BIBLIOGRAPHY

1. F. Stuart Chapin, Jr. and Shirley F. Weiss: Factors Influencing Land Development, Institute for Research in Social Science, University of North Carolina, Chapel Hill, August 1962.

2. F. Stuart Chapin, Jr. and Shirley F. Weiss: Some Input Refinements for a Residential Model, Institute for Research in Social Science, July 1965.

3. Charles H. Graves: "Two Multiple Regression Models of Small Area Population Change," in Urban Transportation Planning, Highway Research Record No. 102, 1965.

4. Britton Harris et al: Research on an Equilibrium Model of Metropolitan Housing and Locational Choice--Interim Report, Institute for Environmental Studies, University of Pennsylvania, March 1966.

5. Donald M. Hill: Methodology for Developing Activity Distribution Models by Linear Regression Analysis, Traffic Research Corp., Toronto, Canada, January 1966

6. Donald M. Hill and Daniel Brand: Prototype Development of a Statistical Land Use Prediction Model for the Greater Boston Region, Traffic Research Corp., Toronto, Canada, presented at the Highway Research Board Conference, January 1965.

7. T. R. Lakshmanan: A Model for the Distribution of Urban Activities: Formulation, Evaluation and Re-formulation, Alan M. Voorhees & Associates, Washington, D. C., October 1964.

8. Pierre Larouche: The Simulation of Residential Growth in the Montreal Region, Ministere de la Voirie, Province de Quebec, presented to the Highway Research Board, January 1966.

9. George D. Lathrop, John R. Hamburg et al: "An Opportunity-Accessibility Model for Allocating Regional Growth," in Highway Reasearch Board Record No. 102.

10. A. D. Little, Inc., Model of the San Francisco Housing Market, Technical Paper No. 8, San Francisco Community Renewal Program, January 1966.

11. Ira S. Lowry, A Model of Metropolis, Memorandum RM-4035-RC, Rand Corporation, Santa Monica, California, August 1964.

12. Stephen B. Putman: Industrial Location Model, CRP Technical Bulletin No. 5, Department of City Planning, Pittsburgh, Pennsylvania, December 1963.

13. David R. Seidman: A Decision-Oriented Model of Urban Growth, Delaware Valley Regional Planning Commission, Philadelphia, Pennsylvania, August 1966.

ANALYSIS AND DESIGN ECONOMICS OF TIME-SHARING SYSTEMS

Robert E. Fagen*

1. INTRODUCTION

One of the most important trends in computer utilization today is the time-sharing of a central computer or computer complex by a number of independent users who are on-line and in locations remote to the main computer center.

This trend has opened up many new avenues of computer usage, and is stimulating the invention of a number of new types of computer service that can be supplied to individuals or small firms which would not otherwise have access to a computer; also, within large installations or agencies (government research facilities, industrial concerns, hospitals, universities, and federal, state, and city governmental offices), it is creating new ways of using a large central computer facility which normally would be present anyway. In the latter case, the provision of computation and access to data files for on-line use makes the computer service much more unified and can encourage new organizational and operational techniques such as integrated systems for Management Information and Command and Control functions.

It is obviously impossible to discuss many specific systems in a brief survey; in fact, it would be quite a chore even to list the systems and applications now in use or in advanced development. Also, it is quite safe to say that the surface has barely been scratched. This paper attempts to outline some of the typical features of these systems (from a computer systems design standpoint); this will provide a framework to help visualize the design problems, the economics of use and operation, and finally, future trends in the light of new computer hardware and software techniques.

2. MAJOR ELEMENTS OF A TIME-SHARING SYSTEM

From the point of view of the designer of a remote time-sharing system, there are four major elements to consider: the central computer facility, the communications medium, the remote terminals, and the programming systems. The latter, being the least "tangible," is also the least understood and appreciated, and the most critical element in the design and planning stage of a system as well as in the final checkout prior to operational use. It is also the most susceptible to radical changes upon any changes in operational usage or service provided the network of remote users. It is for this reason that most analyses and designs should start with a specification of the exact service or types of services that are to be supplied the remote users; as in Reference 1, it is quite useful to attempt to categorize these, even though the list of applications services is growing rapidly.

Once a system is installed and operational, the initial design and development problems are over, and the programming or software expense (except for maintenance and additions) no longer enters the picture. The elements that remain - as items of operating cost - are the main computer center, the communications paths, and the terminal hardware. Economic discussions then revolve around these items - although, as will be seen, the allocations of central computer costs depend heavily on how the software is organized. Since this is the

---

*Computer Communications, Inc.

case, a brief summary of design considerations, separated into the four major elements listed above, is necessary before any meaningful discussions of economics can be attempted.

2.1 Central Computer

Although the newer computers are designed specifically with remote time-sharing in mind, almost any of the commercially available solid state general-purpose computers (since 1959 or 1960) are quite suitable - with the possible addition of special multiplexers and/or auxiliary memories - to remote time-sharing. The most significant features that determine how efficiently a central computer system will adapt to such a role are

(1) the input/output processing scheme

(2) the interrupt capability of the central processor

(3) the amount of main memory available, and the bit size of its words

(4) the type and amount of auxiliary memory available, and its mode of access to central memory

(5) the data rates possible through the input/output channels

(6) the instruction repertoire and speed of the central processor.

It is in these areas that the most rapid improvements are being made - and, it is also these characteristics that determine whether a central computer is considered to be "small," "medium," or"large;" in particular, items (3) and (6) are the primary determining factors in both price and scope of the system. Figure 1 indicates the growth history of the "large scale" computers, as measured by these two quantities; Figure 2 is a partial tabulation of some of the more familiar computer systems in the various price and size categories. It is interesting that while the capabilities in all these classes are increasing geometrically with time, the prices tend to remain constant. Thus, the processing capacity of the smallest system today (even in the $50,000 range) approaches the capacity of the largest ($2 to $4 million) system available ten years ago.

Present time-sharing systems utilize main memory - generally fast core with a minimum of 32K words - and some auxiliary memory, such as magnetic drums or disks, with much larger capacity but slower access times. The requirements for such auxiliary memories vary drastically with the type of service being offered which, in turn, determines the extent to which data files and temporary storage for users must be manipulated and the requisite data quantities and rates. Certain applications require only the main core memory, and are the simplest to handle from a software and timing standpoint; however, they are also the rarest, unfortunately - and almost any time-sharing problem carries with it a requirement for data communication between main memory and auxiliary memory. This is one of the trickiest areas in design and software implementation. In future years, it will be greatly affected by mass random access core memories which offer both faster random access and simpler transfer rules to main memory; however, these will not come into operational usage for at least another year or two, and existing systems employ drums or disks where fairly fast auxiliary memory is needed.

Items (1) and (2) are interrelated; both determine the ability to allow central processing to proceed simultaneously with the input/output processing of data from a number of independent sources. In newer systems, the degree of parallelism of input/output processing has progressed to the point where a large volume of data can be moved to and from a diverse set of external sources without any interference with the central computation functions; the central processing itself has also been made parallel in several systems which allow multiple central processors in a single system. Even in systems where a single central processor is employed, the newer computers tend to have sufficiently large main memories (65K or 131K minimum) so that a number of different jobs can reside in memory at once, giving the ability to switch the attention of the central processor from one job to another in microseconds. This ability is essential to time-sharing - and to operating a remote time-sharing network concurrently with a standard workload or "background" jobs.

As can be seen from the above, the central computer facility is actually made up of different pieces of physical equipment; even the "main frame" consists of logically distinct elements such as memory, I/O channels and central processor. Since a time-sharing network "uses up" real time and space on the physical hardware at different rates depending on the mix of problems and services, the determination of system capacity and of cost allocation to users is one of the more challenging technical design problems.

2.2 Remote Terminals

The devices that can be employed by a remote user are growing in number, and can only be discussed in the framework of intended use or service. As in Reference 1, it is useful to lump these uses into broad categories to obtain some understanding of the array of available devices - and foresee the trends in the newer terminals being developed.

A remote terminal can vary all the way from a small general-purpose digital computer with peripheral equipment down to a simple portable "box" with a few keys. At the lower end of the scale, there is a variety of devices, each specialized in nature, and each designed specifically for a single application. Examples of these devices include on-line reservation terminals, stock brokers quotation boxes, data collection devices, etc. Each such device has whatever it needs in the way of keys, controls, and output readout to satisfy the user, and has labels that essentially reflect the language in which the user communicates with the central computer. This "language" is completely determined by the particular application.

At the upper end of the remote terminal spectrum is the "remote computer center" or "remote batch processor." In this case, the remote station is actually a small computer with peripherals that has the capability of submitting batch jobs to a central computer located elsewhere. This usage, is actually an extension of the traditional "closed shop" usage of a computer center; thus, the equipment at the remote sites, and the manner of using it, is typical of that to be found at the central computer site. The remote site loads jobs and receives the output of jobs automatically, the loading and outputting taking place on the remote peripherals, with the job processing occurring at the central computer site. Typical equipment at the remote site includes line printers, card readers, card punches, and a small computer and console typewriter; actually, the "size" of the remote computer is of no significance and can vary from "small" to "large" depending on other uses that may exist at the remote site.

The middle ground between the extremes above is filled by a class of remote terminals serving the application often called "general conversational usage." Here, a user is actually submitting inquiries or jobs of rather short duration, with a requirement for fairly rapid response. Examples of such applications include on-line program debugging, on-line "conversational" execution of FORTRAN (or other problem language) programs, information retrieval and general data inquiry services, etc. Remote terminals that exist now to fill these roles include the teletypewriter, a variety of devices with keyboard and alphanumeric or graphical cathode ray tube (CRT) readout, and modifications of the latter to provide additional input capability such as small card readers or small magnetic tape handlers and added output capability such as small line printers or other hard copy readout. It is fairly clear that this "middle ground" of terminals overlaps the areas mentioned above; at the low end of the scale, a general terminal of this nature can be used for the specialized applications; at the upper end of the scale, a general terminal with small line printer and card reader as attachments begin to have the capability - on a modest scale - to perform remote batch loading.

2.3 Communications

The communications requirements are determined completely by the remote user's input and output data rates. The bandwidth tends to vary in accordance with the spectrum of terminals described above; the specialized "Class A" terminals require as little as 50 to 200 bits/sec since a typical user keys data at a maximum of 10 keys/sec (the equivalent of 60 bits/sec on an alphanumeric keyboard) and output data also tends to be of low rate. The conversational terminals vary from teletype bandwidths (50 bits/sec) if the teletype terminal is used to 2000 or 2400 bits/sec if the CRT keyboards are used. In the latter case, a 2000 bit per second rate will "paint" a new page on the face of the screen in about 3 seconds. Lower rates, such as 200 to 250 bits per second, are also adequate for these terminals provided the user is working with output requirements that add characters to an existing page gradually (which is actually the case in most conversational applications); however, at these relatively low rates, it requires almost a minute to paint a new page of output on a 1000 character screen. Also, if added peripherals such as card readers or small line printers are used, the higher rate of 2000 cps is necessary.

At the upper end of the spectrum, the remote computer centers transmit batch loads in card format, and output large volumes of printer data. Bandwidth requirements are determined by the actual load and the number of peripherals to be driven simultaneously; while the lowest practical bandwidth is 2400 bits/sec, the most practical range is actually from 20,000 to 40,000 bits/sec. Some remote computer centers actually transfer information from auxiliary or central memory to auxiliary or central memory; this usage is not required for the normal batch loading process, but has been used for certain other applications. In these cases, bandwidths from 200,000 cps to several megacycles have been used. These higher bandwidths are also used for transmitting concentrations of lower bandwidth networks from point-to-point.

Figure 3 summarizes some of the ATT Datasets that are available in these bandwidths. The most frequently used are the 103A and 103F sets for 200 to 300 cps, the 201A for 2000 cps, and the 301 (Telpak A) for 40,800 cps. Reference 2 discusses actual field experience with the latter, and the proven capability to drive a number of high-speed card readers and line printers reliably over distances of several hundred miles.

Conservative use of the low bandwidth (103A) datasets allows completely general use of the entire ATT long distance network. For those applications requiring low rates, this makes computer time-sharing as available as any commercial telephones. Several acoustic couplers have been developed that allow the reliable use of any telephone handset, so long as the bit rate is under 250 cps; in effect, this eliminates the need for a dataset installation, and when used with a portable terminal makes every office or home a potential remote site to a central computer located almost anywhere.

2.4 Programming Systems

All time-sharing systems must be serviced by a fairly elaborate programming system that controls all the hardware in the network and provides the required service. As mentioned previously, the programming system is the essential part of any time-sharing system. Typically, the programming system is divided into two functional portions (although sometimes the programming package itself has no identifiable separation into these parts). The first part serves as an executive and monitor that controls communications with the remote network, schedules tasks, stores and retrieves data, and passes assignments to the central processor. The second part is a program that actually performs the desired service or computation; this latter part may include a compiler (if the input language is oriented to some application area, or some general language such as FORTRAN) and other applications programs. The latter functions and programs vary with the application or service to be provided the user; the former are determined by the network "traffic" rules (response requirements, delays, number of users and access rules, etc), and by the mode of operation with other jobs. As this indicates, there are two types of "time-sharing" that may occur - either singly or simultaneously. The first is the sharing of the central computer by the network of users amongst themselves, and the second is the sharing of the central facility between the network and other jobs or users of the central facility. If the network has full use of the facility, the computer is said to be "dedicated;" if it occupies only a portion of time or space with other networks or normal "background" jobs, it is only "partially dedicated." The ability to provide partial dedication in an operational system has not been too prevalent yet, since the older central computers lacked memory capacity and convenient organization and/or software; however, the newer machines and time-sharing systems under development all allow for this capability - this, when fully exploited, drastically affects the economics and cost allocations of the central computer, and also increases the flexibility for any installation.

3. OPERATIONAL ECONOMICS

The cost to each user of a time-sharing system is determined by the same factors that effect the design, namely, the remote terminal costs, the communications tariff, and the cost of the central facility. In this section, these are surveyed briefly, and some forecasts of trends are projected.

3.1 Cost Analysis

The monthly rental that must be charged to an individual user depends on

(1) the cost of the central facility and percentage dedication thereof

(2) the number of users in the network

(3) the cost of the remote terminal hardware

(4) the communications tariff.

(Since most time-sharing services exclude communications costs (the user pays his own line tariffs), these will be discussed separately.)

To start with a simple example, suppose 64 teletype terminals are serviced by a large machine (in the $2 million class) which is fully dedicated to providing the service supplied to the individual remote users. If the service is made available on a full shift basis, the 64 users must contribute the equivalent of full prime shift monthly rental plus the overhead operating expense of the center. Rental plus operating expense for a large installation of this class would run at least $64,000 monthly; thus, the cost per user would be $1000 per month at the minimum. To this must be added the rental of a teletype terminal which averages about $75 per month. Finally, the communications line charges (which depend on distance from the computer center and total amount of usage) must be added. For moderate distances, this could amount to as high as $400 per month on a teletype line for full time usage; for sporadic usage, the charges would run much less, but probably would not amount to less than $50 monthly. Since this total amounts to from $1100 to $1500 monthly, a comparison must be made between the grade and convenience of service so obtained and comparable service (if feasible) that might be obtained by a small scale private computer located at the remote location - since it is now possible to lease rather complete desk size systems in this price range.

This example serves to pinpoint those areas that determine ultimate user cost. Within each area, much variation is possible depending on the exact system configuration and use; therefore, the types of alternatives will be discussed individually.

3.1.1 Central Computer

User costs, for a specific central computer, can be reduced in the following ways

(1) by partial dedication of the computer

(2) by increasing the number of subscribers

(3) by over-subscription, filling time with subscribers who get access to less than full time.

All the above can only be achieved through appropriate time-sharing software, and also might result in degraded service (i.e., slower response, probability of "busy signals," etc) and the cost saving must be traded off against this degradation. The ultimate test, for a particular service, is whether or not the resultant grade of service is tolerable to the particular user.

To pursue the previous example, suppose the same $2 million computer serves 200 on-line users, any subset of 64 being allowed simultaneous access, and the computer itself dedicated only 10% of its time to the on-line network and 90% to a "background" job or jobs. In this case, only $6400 of monthly costs need be allocated to the user network - and since the 200 users get varying amounts of service, each could be charged for time actually used. The average cost could be as low as $32 monthly for an average use of 60 hours out of the prime 176 hours available. It can be seen by this extreme that the communications costs ($50 - $100 monthly) and the terminal hardware costs ($75 monthly) actually outweigh the central computer portion, the composite cost being as low as $150 to $200 monthly.

There are many instances where a choice must be made between a very large central computer - capable of having a sizable network of remote terminals time share with background jobs - and a small computer, completely dedicated and possibly handling fewer remote terminals. In the example above, a small computer would serve as well, provided that (for example)

(1) it could handle 100 users, 32 with simultaneous access

(2) it is fast enough to provide the same grade of service in response

(3) it has a total monthly rental and operating cost less than $3200.

It is clear that with these assumptions, the remote users get the same grade of service at the same cost. Tradeoffs of the type above are fairly subtle, since the ultimate question is the capacity of the central computers - hardware and software - to provide the required service under the requisite time-sharing restraints.

3.1.2 Remote Terminals

User costs in the remote terminal area depend on the type of input data and output data that the user actually requires to do his job. As these requirements are compressed or simplified, terminals can be designed specifically for the requirements - as in the Class A specialized terminals - and can be built to sell for anywhere from $500 to $5000, or to rent for $15 to $150 monthly. The lower limit is reached with devices like a telephone dial or touch tone set - here the input is exceedingly simple, and the output is minimal.

In the field of generalized conversational terminals, the cost depends on the complexity of the input keyboard and output requirements. It is very difficult to find a general device much cheaper than the teletype - which rents for as low as $60 - $75 per month - and provides hard copy output and a fairly general keyboard for input. Requirements for additional keys or for faster and more flexible output tend to increase the cost; the cheapest keyboard/CRT units sell for around $10,000 or rent for about $250 per month. Adding additional units - such as small card readers, tape handlers, or line printers to a conversational terminal tends to increase the cost to the $20,000 and up range - which begins to approach the cost for the small remote computer center or remote batch processor. The latter can be implemented with any of several small computer systems and peripherals which sell for as little as $20,000 to $50,000 and rent for from $500 to $1200 monthly.

For the private user terminals, the trend is to portability, low expense, and convenience. Many terminals now becoming available are combined with telephone couplers that allow the use of any standard telephone and require no installation whatsoever. The lower limit of cost of these simple, portable units is probably around $1000; again, as the output requirements increase, the cost goes up. Units now exist that incorporate portable teletypes and phone couplers that sell for $2800, and eventually portable units with general keyboards, TV output, and phone couplers will be available in the $5000 to $7000 range. All of these will make remote terminal capability available almost anywhere through use of the telephone, with no installation requirements.

3.1.3 Communications

Since computer costs are coming down - to as little as $25 per month for some services - and remote terminal costs are coming down - again to as little as $25 per month for the simplest terminals - the final controlling element of cost will be the communications transmission tariff. As has been mentioned, many portable terminals will allow use of standard telephones - in which case the costs will simply add to the existing phone charges and will depend on usage and distance. In any case, the line charges are always a function of bandwidth, distance, and usage - and the economics of a time-sharing system will often depend on the geographic proximity to the computer center. For example, full time use of a Telpak A line is $15/mile/month; in this case, a remote computer center located 200 miles from the central computer will cost $300/month for communications alone (which, in itself, is more than the remote computer center hardware need cost). On the other hand, for low bandwidth use of a standard phone, a private line - should the center be located in the same city - could be available for as low as $35/month.

3.1.4 Summary

The costs of remote terminals are coming down through reduced central computer costs, cheaper portable terminals, and use of low bandwidth techniques through ordinary telephone lines. This, together with portability and convenience, is rapidly expanding the areas of applications - for total monthly user costs of below $100 - of central computer services which could be made available to remote users. The only factor limiting the rate at which these different services will become almost universally available is the need for systems development - especially of the sophisticated programming systems necessary - to make the hardware installations available and reliably operable.

While these developments are certainly not beyond the state-of-the-art, they do require time and effort. Checkout, in particular, is one of the most time-consuming and, hence, most costly items - since a central facility and remote network must be in place and available all during the checkout period. For a given new system or service, this period can easily last as much as a year, during which no user revenue is available.

LARGE MACHINE GROWTH TREND

YEARS	INSTRUCTIONS PER SECOND	MAIN CORE MEMORY SIZE
1953 - 1958	20,000	1K to 4K
1958 - 1963	100,000	32K
1963 - 1965	500,000	32K to 65K
1965 - 1968	3,000,000	131K to 256K
1968 -	12,000,000	131K + up to 2 million "Mass Core"

FIGURE 1

REPRESENTATIVE SAMPLE OF PRESENT OR RECENTLY ANNOUNCED COMPUTER SYSTEMS SUITABLE FOR ON-LINE TIME-SHARING

SMALL SCALE

MODEL	MANUFACTURER	APPROXIMATE PRICE RANGE*
CDC 1700	Control Data Corp	
IBM 1130	IBM	
IBM 360/20/30	IBM	Systems in this class vary in price from $25,000 to $100,000
DDP 116	Computer Control Corp	
PDP-6	Digital Equipment Corp	
SDS 920	Scientific Data Systems	
SIGMA 2	Scientific Data Systems	

MEDIUM SCALE

MODEL	MANUFACTURER	APPROXIMATE PRICE RANGE*
IBM 360/40/50	IBM	
UNIVAC 494	Sperry Rand	
CDC 3300	Control Data Corp	Systems in this class vary in price from $250,000 to $750,000
Burroughs B5500	Burroughs	
SIGMA 7	Scientific Data Systems	
Honeywell H-400	Honeywell	

LARGE SCALE

MODEL	MANUFACTURER	APPROXIMATE PRICE RANGE*
IBM 360/67/75	IBM	
UNIVAC 1108	Sperry Rand	
GE 635	General Electric	Systems in this class vary in price from $1,000,000 to $5,000,000
CDC 6400/6600	Control Data Corp	
Burroughs 8500	Burroughs	

*These figures are very rough estimates only and do not reflect actual prices from any of the manufacturer's current catalogs, nor the different configurations available.

FIGURE 2

PARTIAL TABULATION OF ATT DATAPHONES

MODEL	LINE AND BANDWIDTH	APPROX DATA SET RENTAL*	APPROX LINE CHARGES*
103A	240 cps - voice grade	$35/mo	$3.00/mile/mo
101C	110 cps - teletype	-	-
201A	2000 cps - voice grade	$75/mo	$3.00/mile/mo
201B	2400 cps - voice grade	$75/mo	$3.00/mile/mo
301B	40,800 cps - Telpak A	$250/mo	$15.00/mile/mo

*These figures are estimates only and do not reflect current official ATT rates.

FIGURE 3

REFERENCES

REFERENCE 1 "Systems Organization," by R E Fagen. Chapter of a book entitled "On-Line Computing," edited by W J Karplus, to be published by McGraw-Hill Co (Fall 1966) as part of UCLA series.

REFERENCE 2 "The Implementation and Operation of Remote Computer Stations," by A A Buchwald, E K Dorff, J D Johnson, T A Johnson, G W Rathjen, and J V Robidoux. Paper presented at the Spring Joint Computer Conference, Boston (April 1966).

THE M.E.T.R.O. URBAN GAME-SIMULATION
AN EXPERIMENT IN IN-SERVICE TRAINING

Richard D. Duke*

+"...man is both a game-playing and a game-creating animal, his capacity to create and play games and take them deadly seriously is of the essence, and it is through games or activities analogous to game-playing that he achieves a satisfactory sense of significance and a meaningful role.

...

"Looked at this way, in the territorial system there is a political game, a banking game, a contracting game, a newspaper game, a civic organization game, an ecclesiastical game, and many others. Within each game there is a well-established set of goals whose achievement indicates success or failure for the participants, a set of socialized roles making participant behaviour highly predictable, a set of strategies and tactics handed down through experience and occasionally subject to improvement and change, an elite public whose approbation is appreciated, and, finally, a general public which has some appreciation for the standing of the players. Within the game the players can be rational in the varying degrees that the structure permits. At the very least, they know how to behave, and they know the score.

...

"...the process of experimentation and adaptation in the social ecology goes on. The piecemeal responses of the players and the games to the challenges presented by crises provide the social counter-part to the process of evolution and natural selection..."

---

* Michigan State University

+ "The Local Community as an Ecology of Games", Norton E. Long, The American Journal of Sociology, vol. 64, Nov., 1958, pp. 251-261.

The various papers presented so far in this conference have exposed you to a wide range of topics relating to urban information systems. If it has been difficult for you, as it has for me, to follow the intricacies of some of the presentations which have been made, consider the plight of the neophyte when first confronted with the conceptualization or application of the notions being elaborated here.

The lay planning commissioner or the elected city politician can be empathized with, if not forgiven, for his sometimes negative reactions to the 1984 world we are so actively trying to create. As we continue our efforts to develop effective urban information systems we must not overlook the dimension of this problem.

We must be able to communicate, reasonably and convincingly, the nature and value of urban models and their required data systems if we expect to obtain financial support for their development and maintenance, and if we hope to see the product of these efforts effectively influence the decision makers on the urban scene.

Toward this end we are developing the M.E.T.R.O. Project, a joint venture supported by the Tri-County Regional Planning Commission, Michigan State University, and the University of Michigan under a grant from the Department of Housing and Urban Development. Using operational gaming techniques to develop an "environment for learning", the instrument is intended to have the capacity of demonstrating to appropriate decision makers, in general terms, the consequences of alternative decision chains on metropolitan growth patterns. This will be accomplished through the use of a simulated, abstracted environment, employing a reduction of time span and dynamic interplay of current decisions with fixed policies. The instrument is intended to simulate growth patterns which would occur naturally, and enable their comparison with planned growth patterns. In addition, M.E.T R.O will be designed to : (1) illustrate the kinds of information which are available to decision makers (e.g., data banks), (2) inform the decision makers about the techniques which are available to evaluate and implement decisions (e.g., mathematical growth models and capital improvement programs), and (3) provide information concerning the implications for urban development which are associated with alternative courses of action.

The term "gaming" is a potentially misleading way of referring to various practical and serious activities. War-gaming, for instance, is an ancient art that is a precursor of our efforts.[1] The parallel is that strategy and tactics are experimented with by top decision makers who are operating in a safe, simulated world, mimicking real world decisions.

It must be clarified, at this point, that M.E.T.R.O. is oriented <u>primarily as a pedagogic device</u>, rather than as a pure simulation. We retain the hope that we will ultimately be able to do valid experiments in simulation. To achieve this, we intend to develop specific hypotheses about the nature of player interaction as we observe the early runs, and then to design and conduct experiments specifically to test these notions. This phase of investigation is not intended under our current financing.[2]

A deliberate progression of experimentation with this technique for training urban planners has been pursued at Michigan State University. The initial effort was to construct a simple hand game, "METROPOLIS I". Subsequently, this was modified, programmed for a computer, and is now operational as "METROPOLIS II". The two instruments have been run repeatedly, under widely varying circumstances. The demand for exhibition runs has now exceeded our ability to comply, if the major effort toward developing a second generation game is not to be impeded.

The M.E.T.R.O. Project, a second generation game, has progressed through the stage of detailed design, and is currently being constructed. When complete, it will be a reasonably sophisticated instrument, flexible enough for operational use. Potential uses are with the formal academic program, with adult education programs, and as an urban plan effectuation tool which enables operating planning agencies to communicate concepts to their lay commissioners and to the local decision makers to whom they are responsible.

Hence, it is our hope that planners trained in applying this technology will be better mediators between the technologies and the laymen. By handling and interpreting the complex problems of urbanism via electronic data processing and simulation, planners will bring into practice problem-solving techniques that have remained at the level of theory. In short, we propose to train practitioners in as thoroughgoing a manner to do in their jobs what systems analysis in aerospace industries has already started: i.e., the application of systems concepts to social problems instead of to missile race problems.

The M.E.T.R.O. Project is designed to quickly and efficiently get across long and short run urban social processes to laymen, politicians, urban planners, and social science students. The M.E.T.R.O. instrument is (1) an interlocking and crosscutting set of games. Each player belongs to two types of teams and is appropriately cross-pressured. There are three prototypical urban area governments (central city, suburb, and urbanizing townships) as one kind of team (each with four functionally differentiated members); and metropolitan-wide professional associations as a second type of team (politicians, planners, land developers and educators, who represent their respective areal teams). The instrument is also (2) an interlocked set of computerized simulation models (voters' pressure groups in the voter response model, exogenous and endogenous firms in a growth model, and a distribution model to tie them together) as well as a data bank and programs for printouts of maps, charts and other display materials to aid the game. Player decisions are basic to the simulation inputs, and simulation results generated each "year" (1½ hours real-time) condition player decisions in turn. M.E.T.R.O. uses from twelve to sixteen players and requires an operating staff of half a dozen. The game and simulations are modelled from Lansing, Michigan, economic, demographic, ecological and political data, slightly abstracted and generalized.

The M.E.T.R.O. gaming-simulation joins the approaches of the Cornell Land Use Game and "METROPOLIS", and then adds refinements of its own to increase the sophistication of the instrument.[3,4] A specific "game" will be played by each team; it is linked into the overall game with realistic interaction between all types of roles. Gamed decisions are injected into the simulation of the metropolitan area, and feedback from the computer relates both to player positions and to the growth pattern that results from the aggregate of gamed decisions. Typical decisions required of each team (different in detail and scope, as appropriate to the particular governmental unit or professional role) will be of three main types: budgets, issues, and policies, both explicit and implicit:

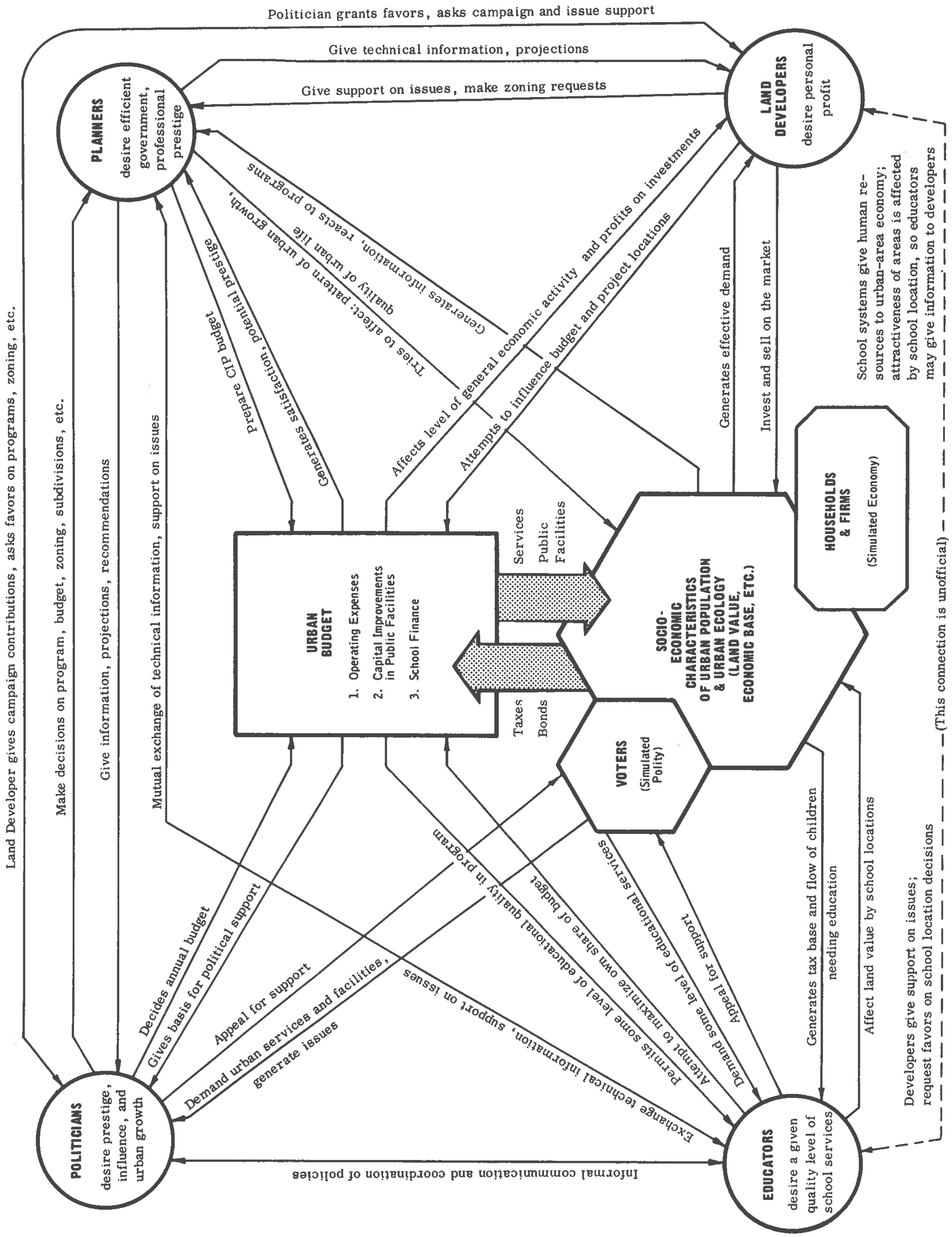
> Budgets -- Each unit of government is accustomed to making an annual commitment of funds. Many policies and relationships are implicit or revealed in the distribution of funds by a unit of government, and these decisions can represent a more meaningful revelation of actual policy than can more explicit statements drawn from these office holders. The budget function will have to recognize the need for the distribution of formal funds - governmental expenditures - as well as that pressure exerted on or by informal funds (influence on or by private investment capital).
>
> Issues -- Each governmental unit will have to be confronted with various issues in the form of a referendum or simply a secret poll on opinions. These will be translated into an index by a conversion program.
>
> Policies -- Each governmental unit has a series of policies under which they operate. Some of these are explicit, many more are implicit, based on tradition or custom for that particular unit. In either case these policies are the standards, or decisions rules, which control the behavior of that unit of government under certain circumstances.
>
>> Explicit Policies -- Are evident in the form of zoning regulations, subdivision controls, utility extension policies, and recorded opinion on issues such as annexation vs. incorporation. There are, of course, many other examples.
>>
>> Implicit Policy -- is evident in more subtle and perhaps more revealing ways. The ratio between actual and assessed value of property types; level, and costs of services provided; extent of professionalization of public employees; and many other factors are indicative of the general state of management which might be expected. The players' reactions to contrived, but typical issues may also be revealing (i.e., the suggestion of some innovative approach for coping with a problem, as the creation of a regional authority).

In each case the results of fresh decisions (annual budget) or of changed policy (e.g., change in ratio of assessed value to real value) must be recorded for each cycle, converted to an appropriate index, and applied, through the use of growth models, to all geographic units having similar characteristics, thus setting the stage for a new machine iteration of the appropriate growth model.

M.E.T.R.O. deals with ideal-types of governments, roles, issues, programs and budgets, all abstracted from Lansing area data. To each modular unit is attributed the gamed results to all like units in interpreting any given run. Similarly, in the simulation, we have abstracted characteristics of organizations and subpopulations in the Lansing area. For example,

FIGURE 1

M.E.T.R.O. Functional Interactions

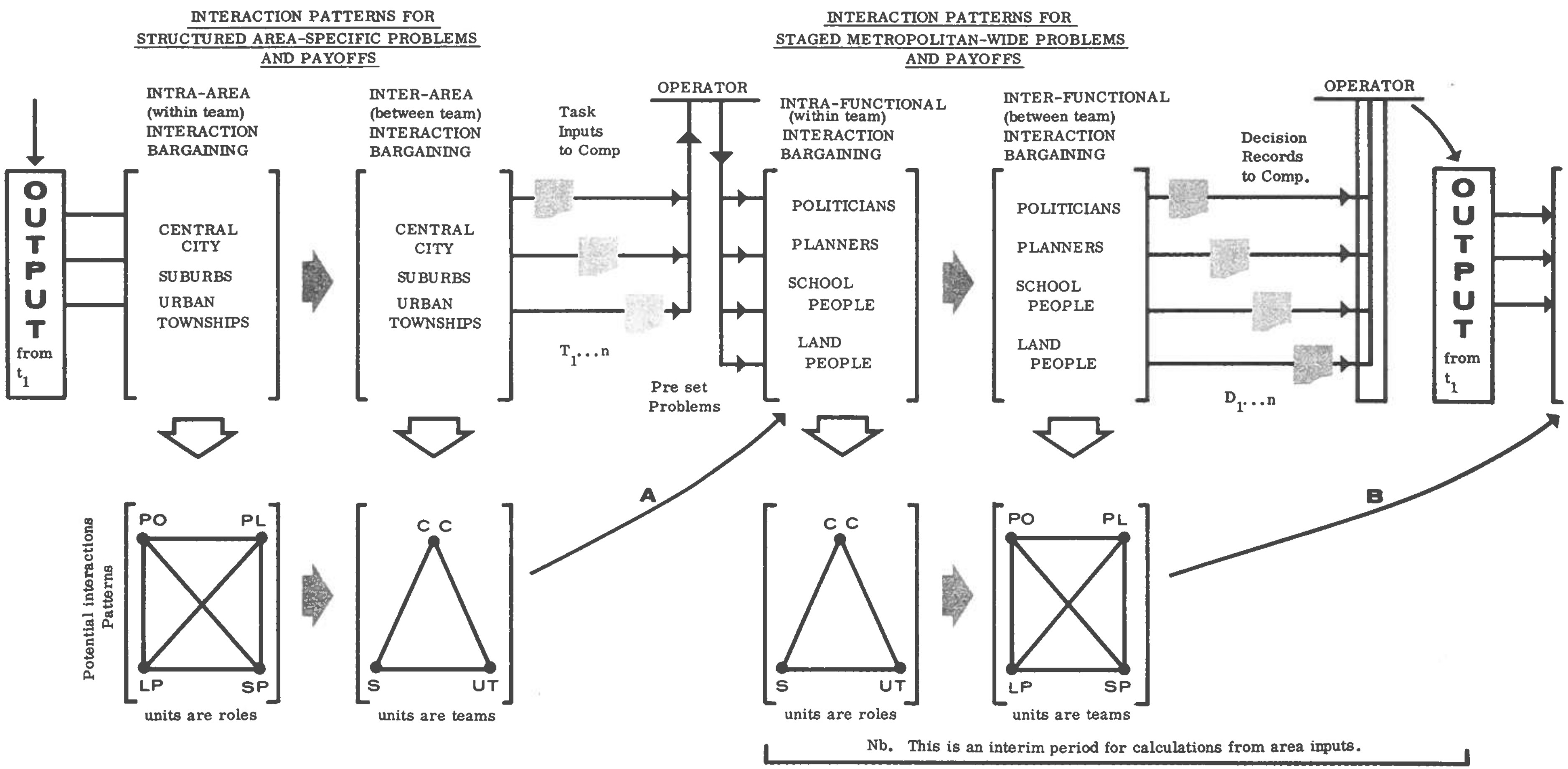


there are three dominant industries: (1) a heavy industrial manufacturing plant that is tied to the fluctuations in the national economy and whose urban area employees are predominantly blue collar workers; (2) a fairly large firm with a slow-growing but highly stable employment outlook (an abstraction of the state government. whose role in the area is like that of a national commercial firm's branch offices); and (3) a high-growth, technologically-oriented "firm" whose major product is concerned with innovation and the knowledge of research and development industries, whose employees are heavily professional, technical and managerial types who are oriented to world and national markets and who are sophisticated in their demands for urban services and culture (this is an abstraction of Michigan State University, comparable in its urban role to aerospace firms and other research and development firms oriented to the new technologies). Our rationale is that basic industry products are unimportant compared to the kinds of employees the firms hire, the inherent stability of their economic activities, their growth rates, and the ancillary firms they support. We give this example in detail to illustrate the processes of ideal-typing that go on. We have similar rationales for selecting five social classes for the region, which are translated into household types for correlation with comsumption patterns, residential mobility, and voter responses on political candidates and issues. For the modelling, regression equations have been used on historical data to establish base-line and response-to-change parameters. The game starts with 1965 and goes on from there, so it is future oriented, and concerned with exploring possibilities of innovations, technological and social, in metropolitan areas.

The major simulation models are: (1) A Monte Carlo type of voter response model which generates turnout and support rates on issues and candidates. Its parameters are largely abstractions of historical tendencies in the Lansing area, and its workings are very sensitive to player decisions in the game; (2) A macro-economic and demographic growth model which generates growth of population, employment and income for the whole metropolitan region. It is a series of lagged finite difference equations relating basic industries, the regional growth pattern, and the pattern of the national economy. It is sensitive to the aggregate decisions of players with respect to investment in firms, housing, capital plant, public facilities, and urban welfare services; and (3) A population and economic firm redistribution model which allocates the above growth to specific areas of the city (consolidated pairs of census tracts) on the basis of transport accessibility to employment and urban services, attractiveness to residential use by social class, density constraints, and the particular locations of facilities, etc., determined by player decisions. It is an elaboration of the T.O.M.M. model developed by J.P. Crecine for a Pittsburgh land-use simulation,[5] and is a complex gravity-type model. In addition to the simulation models, we use computer printout mapping to represent growth of various kinds, or any other variables that are spatially distributed and are held in the data bank. The technique, called "SYMAP", was developed by Howard Fisher of Harvard University. The establishment of a miniature prototype of a computerized data bank in the game means that players can query the system on a variety of issues, population projections, the areas of blight or unemployment, or of recent growth, etc..

FIGURE 2

GENERAL PARADIGM FOR INTERACTION AND
DECISION PATTERNS within a Gaming Cycle (time "t")



A Areal Teams may generate problems for Metropolitan Functions (not pre-programmed by Operator)

B Metropolitan Functional Teams may set up new programs for Areal Teams to decide upon.

Players must buy information, however, so that this becomes a problem of rational adaptation like everything else.

We see the typical M.E.T.R.O. run as involving seven cycles of play in one day, each cycle representing one year. If players are already familiar with the game, all seven cycles may be active play, and sets of seven may be indefinitely strung together over several days to make for many years of hypothetical urban development. But for naive players, the first two cycles will involve learning the nature of the game, its tasks, operations, and strategies. The third cycle will start actual play and this will run five "years" at the end of which the computer will calculate the projected growth of the city on the basis of trends established by the players. Projections will be printed out for the next year, five years, ten years and twenty years hence. These will be part of the critique session that follows the run of the game.

Cycles are carefully elaborated into a tight sequence of activities and interactions among players, and between the player and the computer. The structuring of interaction is facilitated by the layout of the room in which the gaming takes place. There is an inner ring of tables for the first half of the cycle when areal teams meet to make key decisions and an outer ring of tables for the second half of the cycle when players change hats and meet as metropolis-wide teams to discuss area-wide problems. Since the computation time, even given the speeding-up by computerization, cannot be less than half an hour, the meeting of metropolitan functional teams keeps the players on the track during this period, as well as ensuring a useful perspective on the metropolitan area. The intertwining of the operation of the computer (for simulation and game-supporting output) and the gaming allows a maximum of learning activity with a minimum of wasted time.

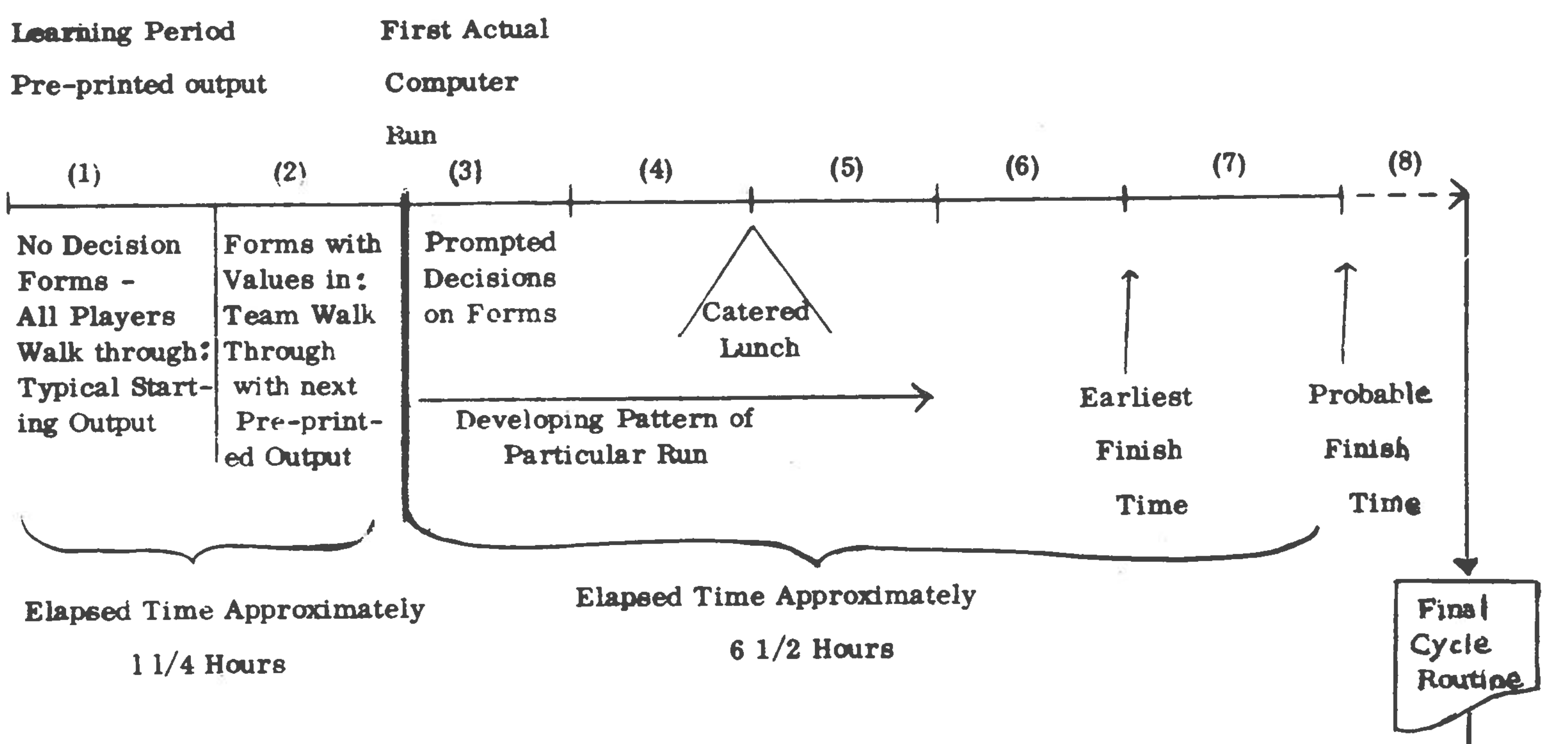
This phasing directly affects player interaction patterns. Their content within areal teams is given in "M.E.T.R.O. Functional Interactions" and the sequences of who interacts with whom is given in "General Paradigm for Interaction and Decision Patterns within a Gaming Cycle". Few unilateral decisions are possible; most player plans and enterprises are highly interdependent, leading to extensive bargaining and coalition formation. Players mutually define their own and others' roles in response to gamed constraints, interaction contingencies, and simulated outcomes. Much of a player's personal standing (liklihood of winning the next election, or personal profit position, etc.) depends on his accurate estimation of how the world will move next year, and his ability to modify, or to take advantage of, the course of events. Though his predictions are aided by data bank outputs and the opinions of his colleagues, his is not a mere optimizing task so much as development of a strategy, bargaining and balancing off power equations in a complex social system. Interaction focuses on the urban budget, upon short-term feedbacks of particular rewards and punishments, and upon the prospects for enhancing most players' chances by growing the area to make a larger base from which to get rewards.

Urban planning of the new type is therefore a twofold process whenever it is oriented to scientific procedures: (1) On the one hand, it

FIGURE 3

M. E. T. R. O. RUN SEQUENCE

(Activity Types for Introduction, Play and Evaluation of Run)



Print-outs:

1. of previous cycle
2. 5 year projection (happy ending)
3. 10 year projection (high, medium, low alternatives for teaching purposes, critique, etc.)
4. 20 year projection
5. Gives what long-term policies are required for each areal and functional team.

INPUTS FOR CRITIQUE

A. If session is not successful, we can have an immediate 1 hour critique (4 P.M. - 5 P.M.)

B. If session generates marked enthusiasm, then more extended critique can follow dinner.

December [illegible] 1965

is concerned to discover rational means of allocating resources to optimize the attainment of diverse public goals. It attempts to evaluate the way in which various public policies, public goods, and gross economic activity patterns will respectively affect (positively or negatively) rational goal attainment. It even compares the short-run policy goals to their long-run effects and the long-run goals of society. (2) On the other hand, planning becomes, in this context, very much of an educational process. Many of the best procedures of national resource allocation are still being discovered and may exist, if at all, only as heuristics in the minds of practioners. Urban policy-making is complex enough so that planners educate themselves merely by the process of exploring the range of decision making tools. They are educators also, not by presenting full-blown plans, but by constructing and explaining a good rationale as a basis for a plan. This introduces concepts of how to think about our urban future and how to think about planning. It sensitizes the public to the alternatives we face in the future.

The M.E.T.R.O. Project is illustrative of one of the most effieicnt techniques for quickly conveying this very complex subject without the problems of a language barrier. Operational gaming involves people in the excitement of a game, but one that is relevant to real problems they face, so that problems of motivating the learner are lessoned. It also gives direct experimental contact with prototypical decision situations, realistic problems and constraints, and prototypical power and conflict relations, all merged into the same framework.

Gaming simulations of the M.E.T.R.O. type will serve educational functions for many populations:

1. Planners -- both professional training for planning students and the refurbishment of skills for professionals in the field, as an introduction to new technologies and the meaning of applied social sciences. For example, training in the use of urban data banks and the array of social statistics they provide.

2. Politicians and Public Administrators -- as a link between planners or technicians and policy makers, it makes professional abstractions and jargon more concrete and pertinent. As an educational tool, it exposes the array of choices actually available and the contributions of some of the new technologies, and it gives the beginnings of rationales for evaluating plan alternatives. Under the heading of elected officials, we must distinguish between the updating of decision-skills of established officials on the one hand, and the familiarization of new officials with the complex problems of urban affairs on the other hand. Relative to this point, M.E.T.R.O. models can provide a fairly good "reality shock" to newly elected officials with strong ideological convictions.

3. Social Scientists -- concerned with urbanism, in political science, economics, sociology, geography, and public administration, can use the M.E.T.R.O. instrument to investigate systematically the effects of hypothetical decision patterns on the urban scene. They can run differect kinds of models in the simulation sector and examine player reactions. They can examine such gaming simulation

instruments for ideas on where data or explanatory concepts are missing; and they can utilize their own M.E.T.R.O. type instruments to explain their disciplines to planners, to students, and to lay persons. When the planning rationale is made more public in this way, better students can be attracted to urban problems and to urban planning.

The recent studies of the potential of a "Statewide Information System for California" by Lockheed Missles and Space Company have focused attention upon the numerous possible benefits, and the immense difficulties, inherent in instigating large scale information systems.[6] One major point which has come to the fore on the basis of this effort is the clear need to develop a new kind of training program. The potential users of such an information system may develop a competence in operating with the system to the point where they even experiment with it to gain confidence in using hardware components that would otherwise be confusing.

A major effort in the gaming simulation research is developing, using, and continually experimenting with a prototype information system. The game's simulation device is based upon a descriptive cross-section of information on the urban-regional environment. Through the operation with and manipulation of this system, players gain an immense amount of experience for transfer to its "real world" counterpart.

In one sense, then, the information system components of the gaming simulation effort provide a unique opportunity for the researchers to experiment with the possible mixes of hardware components to best utilize information in such a system. On the other hand, the gaming instrument offers an excellent vehicle for carrying on the type of personnel training in a variety of information system use which will be required if such systems are to be given a real opportunity for successful implementation.

The M.E.T.R.O. Project has been carried on under the direct sponsorship of the Department of Housing and Urban Development. Several objectives are sought as products of this study effort. First, some improved mechanism is needed for introducing community leaders and businessmen to the local and regional planning programs. Too often, the lack of effective communication of planning program objectives and procedures has stifled the effectuation of plans which evolved from years of work, and at considerable expense to the local community and usually the federal government. Second, there is typically a lack of continuity of planning programs over the long run because of changing political administrations and agency personnel. A technique is needed so that politicians and community leaders can periodically review the format and objectives of the planning program within the context of the continually changing urban situation. Third, new methods for achieving improved understanding of the phenomena of urban growth are much sought after. The device of simulation of urban development through man-machine interaction provides a potentially useful extension of present simulation approaches.

M.E.T.R.O. is devised as an instrument to aid in accommodating the first two of the above concerns. Activities in the gaming-simulation are focused upon the simulated urban region to acquaint participants with current aspects of urban development problems, and to experiment with and react to the simulated environment in order to test ideas and alternative operating procedures. Experience with gaming instruments indicates that the usefulness of bringing groups together into a close knit working environment is a worthwhile product in itself.

1. War gaming is best summarized in: K.J. Cohen and E. Rhenman, "The Role of Management Games in Education and Research", Management Science VIII (January 1961) pp. 131-166; Donald F. Featherstone, War Games (London: Stanley, Paul & Co., Ltd., 1962); and Clark C. Abt, "War Gaming", International Science and Technology (August 1964 p. 29 and ff). On business games see: Cohen and Rhenman, op.cit.; G.R. Andlinger, "Looking Around -- What Can Business Games Do?" Harvard Business Review XXXVI. No. 4 (1958); J.R. Green and R.L. Sisson, Dynamic Management Decision Games (New York: Reinhold Publishing Corp. 1961). Similar to our project in scope is Guetzkow's "Inter Nation Game", cf. Harold F. Guetzkow, et al, Simulation in International Relations, (Englewood Cliffs, N.J.: Prentice Hall, 1963).
For good general review of the considerations involved in gaming-simulations see Richard L. Meier, "Explorations in the Realm of Organizational Theory: IV, The Simulation of Social Organization" Behavioral Science, vol. 6 (July, 1961) pp. 232-247; and for the theoretical basis see R.D. Luce and H. Raiffa, Games and Decisions: Introduction and Critical Survey (New York: John Wiley and Sons, 1957).
The simulation materials in economics are less relevant to teaching, but quite useful are: G.P.E. Clarkson and H.A. Simon, "Simulation of Industrial and Group Behaviour" American Economic Review I, No. 5 (1960) pp. 920-932; M. Shubik, "Simulation of the Industry and the Firm" ibid, pp. 908-919; and G.H. Orcutt, et al, Microanalysis of Socio-economic Systems: A Simulation Study (New York: Harper and Bros., 1961).

2. Fein, Louis, "The Structure and Character of Useful Information-Processing Simulations", undated, 8 pp. mimeo.

3. A political and planning urban game, R.D. Duke, Gaming Simulation in Urban Research, Institute for Community Development, Michigan State University, East Lansing, 1964.

4. Allan G. Feldt, The Cornell Land Use Game, Center for Housing and Environmental Studies, Division of Urban Studies, Misc. Papers #3, Cornell University, Ithaca, New York, 1965.

5. John Patrick Crecine, "T.O.M.M. -- A Time-Oriented Metropolitan Model for Spatial Location" Tech. Bulletin No. 6, Consad Research Corp., January, 1964.

6. ___,"California Statewide Information System Study", Lockheed Missile and Space So., Final Rept. Y8265-5, Sunnyvale, Calif., July, 1965.

URBAN GEOCODING SYSTEMS AND U.S. CENSUS IMPLICATIONS

William L. Clark*

The importance of the technology under development for the 1970 Census to emergent information systems, and the recent advent of successful urban geocoding systems point to some considerations and opportunities which I would like to emphasize to this audience today.

In this paper I will cover four basic points, namely:

1. Definition of a geocoding system;
2. General description of a geocoding system;
3. Brief summary of certain parts of the census system developed for the 1970 Census;
4. The advantage of coordinating a local geocoding system with the U.S. Census system.

DEFINITION OF A GEOCODING SYSTEM

First, let me define an urban geocoding system. Those of you who attended last year's conference may recall a paper devoted to this topic given by Robert B. Dial, and I refer you to that paper for depth reading.[1]

An urban geocoding system is one to retrieve spatially oriented data without preconceived commitment to areal unit definition, such as census tracts, traffic zones, planning analysis areas, etc. The obvious advantages of such a system are:

1. Complete flexibility as to areal definition in the aggregation of data;
2. A release from preconceived notions of a region;
3. The establishment of a directory system substantially independent of the data system;
4. Interrogation and demonstration via map imagery and graphic display.

The apparent disadvantages of such a system are:

1. Relative, rather than absolute location of address within a block face.
2. A requirement for a fair degree of user technical sophistication;
3. Directory system development costs.

GENERAL DESCRIPTION OF A GEOCODING SYSTEM

To clarify the previous points, let me move on to the second topic and discuss the general aspects of the geocoding system. The common denominator for <u>most</u> non-machine-readable data in the street address, but the street itself does not fit very well in automated data handling techniques. There must be a directory of some sort developed which correlates street addresses to some other sort of areal unit such as census tract, police precinct, or any or all such units of interest. This list of areal units could be carried on at great lengths and due to boundary changes would require constant updating.

---

*Lecturer in Civil Engineering and Assistant Director, Urban Data Center, University of Washington, Seattle

The geocoding system, however, links only one areal unit with each address which we call a grid block. A grid block is a very small area covered by one set of grid coordinate intercepts, and is a function of the degree of fineness of the coordinate system being used. In other words, each and every address in the city will have a corresponding unique x, y coordinate which defines one corner of the grid block in which the address occurs. I will describe how we use this coordinate in a moment, but first let me discuss a problem which actually precedes this one.

For many urban areas, certainly not all, there exists some sort of a system in the street addresses. That is, if the address on a given street change uniformly at each intersection, and if even or odd addresses indicate the side of the street, there exists a system which should be taken advantage of.

Given the x, y coordinates at the end points of such a street, and the address delimiters at intersecting streets, for a network of streets, the computer program will form an address directory listing by street, and intersecting streets, the address delimiters, the x, y coordinates of the block centroid and the x, y coordinates of the block face. Incidentally, I will not argue at this point which coordinate system should be used as the geocoding system is independent from any given coordinate system. At any rate, one program of the system is designed to take advantage of any regularity in the street addresses. Ideally, the program would prove to be of tremendous advantage for an urban area whose address system was very regular and the streets laid out in a grid pattern. On the other hand, if an urban area had no system at all to the street addresses, this particular program would tend to be useless, and the entire directory would have to be developed block by block. Realistically an urban area would fall somewhere between these two end points and the program would <u>assist</u> in developing a directory. Also, realistically there is a vast amount of editing to be done, and a second program has been developed to assist in the editing. This is merely a plotting program which plots the street addresses from the directory coordinates and can be compared to the base map for editing purposes.

Now, once such a directory has been developed, and this is by no means a small task, a list of random addresses can be compared with the directory by a third program, and for each address input the program will report back the coordinates of the block centroid and the coordinates of the property itself. Let me say at this point that the coordinates of the property do not exactly describe the location of the property. For example, if a property at 1250 Main Street were entered into the system and in the directory in this particular block face ranged from 1200 to 1299, the coordinate reported for 1250 Main would be half-way between the coordinates of the two intersections. This does not exactly describe the location of the property, but does give each property a unique coordinate.

The fourth and last program in this system is a grid-coordinate-to-polygon program for data retrieval and summarization purposes. Any set of polygons can be described by giving the coordinates of the vertices, be they census tracts, police precincts, trade areas or whatever. A list of addresses which now have coordinates assigned from the previous program can be input to this program which will link addresses and their data to areal units. Hence, once the directory has been developed, data can be retrieved by any areal units desired. Retrieval is completely independent of any preconceived notion of the areal units.

A by-product of this system is a directory compatible with the proposed U.S. Census Street Address Guide for the 1970 Census. This geocoding directory allows you to collect local data independently of the census by street address, for example: building permits, public utility records, health and welfare records, property records, etc., to be used for planning and operational purposes. The system described here will easily allow merging of U.S. Census data with this independent locally collected data.

SUMMARY OF THE 1970 CENSUS SYSTEM

Now, I will leave the subject of geocoding systems for a moment and discuss independently the efforts of the U.S. Census Bureau in tooling up for the 1970 Census, the third point I originally set forth. For detail I refer you to a paper given by William T. Fay at the recent AIP Conference in Portland.[2]

The Census Bureau is in the process of modifying U.S. Geological Survey maps for urban areas of 25,000 or more and intends to finish by 1968. These maps will be sent out to local cooperating agencies for updating, usually by a local planning staff. The Bureau then updates the original tracings and proceeds to develop a computerized address coding guide which will contain, for each block face, the address delimiters, intersecting streets, census tract, zip code, block number and a 5-digit optional field. It will then be necessary to send this information once again to the local agencies for editing. Incidentally, this effort by the local agency has been approved for funding through Section 701 of the U.S Housing Act of 1949, as amended. Copies of the Street Address Coding Guide will be available to the cooperating agencies in machine readable form, once the editing has been recorded by the Census Bureau.

After the 1970 Census it will be possible by communicating to the Census Bureau the correct combination of block faces, to receive back the data by areas of your own choosing, provided the requests do not violate the laws of disclosure. This is certainly a brief description of the U.S. Census efforts, but these are fully documented by William Fay and would take too much time today to discuss in detail.

ADVANTAGES OF COORDINATING GEOCODING WITH STREET ADDRESS GUIDE PREPARATIONS

The main point I would like to emphasize here today is fourth point mentioned initially, namely the advantages of coordinating a local geocoding system with the U.S. Census System. These advantages are:

1. The availability of the maps of the urban areas currently being revised by the Bureau facilitates for the digitizing of street length records in the geocoding system.

2. Local data entered by street address records can be summarized to the block face level automatically via the geocoding system, which will allow a merge with the census data. In other words, it will be possible to take data recorded by address, summarize them, correlate them with the census data and study them in areal units of your own choosing, rather than combinations of census tracts.

3. The operational base for local updating and expansion of both the map file and the street address guide is provided for inter-censal use by local agencies.

4. By local acquisition of U.S. Census data tapes, extremely economical and fast retrieval of data can be provided locally or in selected centers.

5. Both local and census data can be retrieved by polygon file description, or probably in the very near future by CRT - light pen boundary outlines.

6. The marginal cost of geocoding system development is small relative to the vastly expanded data retrieval capabilities and the costs of hand editing the street address guide are absorbed in total systems compatibility expansion.

7. The U.S. Census Bureau will be relieved of some local demands regarding their local block face summary files.

CONCLUSIONS:

1. No doubt information demands of the geocoding system nature discussed here will emerge universally by, or shortly after 1970 with commensurate expenditure demands.

2. Local cooperation with census will be far more significant to census and local agencies if expended to a system of lasting benefit and in keeping with emerging technology. If a local geocoding system is not developed coordinately with 1970 Census effort, a golden opportunity will be missed.

FOOTNOTES

1. Dial, Robert B., Street Address Conversion System, Research Report No. 1, Urban Data Center, University of Washington, Seattle, 1964.

2. Fay, William T. and Robert L. Hagan, "Computer-based Geographic Coding for the 1970 Census", Geography Branch, U.S. Bureau of the Census, Suitland, Maryland 1966.

CURRENT DEVELOPMENTS IN HEURISTIC GAMING AT CORNELL UNIVERSITY

Allan G. Feldt *

Efforts similar to those at Michigan State University have been taking place at Cornell University over the past few years. Our experience supports theirs concerning the utility of gaming devices as teaching instruments and offers complimentary developments in a number of instances. Fairly close cooperation in game developments and uses of the games has characterized the relationship between the two universities over this period.

We have employed three specific games in various parts of the planning curriculum during the past four or five years with considerable success. Our first efforts were with the Planning Operational Game Experiment developed by Francis Hendricks (1) which has been used in several classroom contexts, particularly as a part of the planning field problem. Richard Duke's Metropolis (2) has been run a number of times but generally outside of a normal classroom context with limited student participation. In recent years, however, most of our efforts have been concentrated upon the use and development of still another game which has been developed at Cornell, the Community Land Use Game, more commonly known as CLUG. (3)

Under a recent grant from the Ford Foundation we are now attempting to integrate both the Community Land Use Game and a few other games such as Metropolis more fully into the normal planning curriculum. At the same time we are attempting to develop more advanced versions of CLUG which lend themselves towards the representation of specific real world cities, though on a somewhat more modestscale in the M.E.T.R.O. project developments described by Professor Duke. This game is developing into a device which is highly complimentary to some of the other gaming devices such as Metropolis and is seen by both Professor Duke and me as a supplement rather than as a competitor to his efforts. In several respects the CLUG model is a simple analogue of some of the land use marketing and development subroutines used in METRO and in basic structure is not unlike the Time Oriented Metropolitan Model developed for Pittsburgh by the CONSAD Research Corporation. (4)

I want to concentrate upon our more recent developments in the CLUG-like models but it is first necessary that I provide you with some background information on the operation of the basic model. Although this material has been covered more fully elsewhere, some understanding of the game's operation is necessary in order for you to understand what we are trying to do in the currently developing second generation of these models.

---

* Department of City and Regional Planning, Cornell University

The basic game has now been used for about three years with considerable success. We have noted a high degree of involvement and motivation on the part of most players coupled with highly favorable subjective responses to the learning experience provided by the game from almost everyone who has played for a reasonable length of time. The game has been found to be highly adaptable as a classroom device since we have gone to extraordinary lengths to make it amenable to play without computer assistance and since only eight to ten hours of playing time provide most players with a high degree of acquaintance with the game's operations and potentialities. This leaves adequate additional time for modifying some of the basic parameters and noting the effect of these modifications upon the community developed in subsequent play. The presence of two experienced operators in conducting the game without computer assistance is required and a class period similar to a laboratory period providing for several hours of uninterrupted play is also a necessity.

Upon completion of play, almost all players agree that the experience has been worthwhile and that they have "learned" from the experience. It is difficult, however, to determine exactly what it is they have learned. Most simply they appear to have gained some insight and experience into the actual functioning and growth of a community. Put in other terms, the game appears to provide a grasp of the totality of fairly complex interdependencies and relationships within an artificially created but fairly realistic system model.

Although my information is not at all current, I understand that the Community Land Use Game is now being used in six or seven American universities, in several adult education courses, in one or two professional planning offices, and is under development in planning offices or universities in two or three foreign countries. These adaptations have been made under somewhat trying circumstances since each of the users has had to build his own playing pieces and boards from whatever information he could garner from the printed descriptions. As a part of the work under our current Ford Foundation grant we are attempting to make both programs and kits of the game available at a relatively low cost to potential users. As these kits become available, we hope to have a greater number of adaptations of the game.

As has been pointed out in earlier discussion, the game is fundamentally a communications device, intended for an educational milieu--whether that be in a formal classroom or in actual planning practice. We have not been overly concerned with the model's precision, its detail, or its predictive power, although it does appear to be somewhat stronger in general predictive power than we had originally expected. I want to describe to you briefly what happens in a fairly typical sequence of play of the elementary version of the game in order that you may more fully appreciate what we are attempting to do with the more complex models currently under development. I hope this appreciation will embrace both the fairly strong potentialities of this type of device as well as some of its basic limitations. The simple game I am about to describe is one which would normally take about fifteen to twenty hours of playing time including several hours to teach novices the fundamentals of the game's operation. In the course of this play a community has grown from an empty field to approximately 250,000 population over the course of 25 rounds of play.

At the beginning of the game the players are confronted with a completely open board upon which all land is unowned and is available at some fixed minimum price from the game operator. To keep matters simple, we generally provide a uniform topographic distribution and no important barriers to travel or development such as rivers, lakes, or marshy ground. Again for simplicity's sake we generally provide a relatively simple transportation system consisting of only two modes of transport, that along all grid lines of the board being twice as costly as that along one or two major highways located in a pattern intersecting near the center of the board. One or two transportation terminals are located which provide points of access to the outside world and a utility plant is placed somewhere in close proximity to one of the terminals.

Play begins with players seeking to purchase land by competitive sealed bids on the open market using some part of the limited amount of money provided them at the beginning of play. By majority consent of the players, utility lines must then be provided to those plots of land upon which construction is anticipated in the near future. All utility lines must of course be linked up with the utility plant initially provided. During the construction phase which follows, one or two players will generally construct a small industry while one of the other players will seek to provide residential units to work in the factories constructed. No factory owner can receive income unless he has an appropriate number of employed residential units working for him and no residential owner will receive income unless he has managed to gain employment in some factory or store. At the same time, no player has enough capital to provide both his own factory and his own workers, so some degree of division of labor among the players is required. Players may locate buildings anywhere they please upon the board provided only that the lot in question has utilities provided to it. In order to hold down their transportation costs, however, a prudent player will locate his activity near that other activity or terminal upon the board with which it most frequently interacts. Thus factories locate near the transportation terminal, workers tend to locate near their place of intended employment, and stores tend to locate near the major residential groupings.

During the first few rounds of play, one or two small industries are generally developed providing employment for two to four residential units plus a nonspecialized store to provide goods to the residents and its employees consisting of a single residential unit. Although questions of scale and calibration are not of paramount importance in the basic game being described here, it has generally been found convenient to consider each residential unit as consisting of about 1,000 employed persons and approximately 4,000 total population. Factory owners then receive an income from the bank proportional to their size and number of employees and this money is partially redistributed in turn to their workers. These residential units must then pay some portion of their income to the store owners for goods purchased there and the store owners must in turn meet their own payrolls. Thus a relatively simple input-output matrix operates dictating the ultimate gain for each member of the community derived from a given dollar's input to the industrial sector. Players are then required to pay the transportation costs they have incurred for the types of land they own and their relative locations to each other and then to meet a tax payment to meet the "community's" expenses incurred for providing utility lines and services to the resident population. A relatively straightforward property assessment procedure is followed and applied to a tax rate set by the players themselves. The money collected is used to pay the community's past expenditures and some incurring of community debt is allowed within appropriate constitutional debt limits. Land is then made available for sale on the open market and approximately the same sequence of stages follows in each successive round of play.

Generally, players can reasonably expect to develop a community of approximately 20,000 population within the first few rounds of play. This is usually followed by a few rounds of relative stagnation while additional capital for further development accumulates but by round ten it is not uncommon to have a city of about 50,000 population containing three major industries, several local convenience goods stores, one store offering lower volume specialty items, an office building providing administrative services to other commercial activities, and twelve to sixteen residential units employed in these activities. Successive rounds then begin to produce accelerated development as capital becomes more plentiful but within an additional five or ten rounds usable land becomes relatively scarce. The occurrence of scarcities of capital or land can be manipulated in several directions depending upon the willingness of the operator to employ some form of zoning regulation and/or some form of mortgage financing for the players. Occasional setbacks to the community as a whole and to individual players occur through the use of a pair of dice thrown to determine the usability of each building constructed over the succeeding five rounds. Players roll against the possibility of losing the use of their buildings with odds proportional to the age of their buildings. They are also provided with an opportunity to renovate buildings at substantial cost before rolling, thereby effectively decreasing their age and their probability of loss. Substantial areas of blighted and unusable buildings sometimes occur in communities built and maintained by unlucky or imprudent players.

With reasonable luck and skill, most communities grow to 150,000 to 200,000 population within twenty rounds of play. Continuation very far beyond this point tends to become somewhat awkward in manual operation due to the large amounts of information which must be processed in order to operate the game. Manual games exceeding forty rounds duration have been reported but usually cycling time becomes so slow as to make termination in 25 or 30 rounds desirable. The use of some form of computerized data handling facilities mitigates this problem considerably thereby allowing development of larger communities as well as providing players with some additional insight into the use and potentialities of urban "data bank systems." Programs to handle the information processing for the basic game have been developed and will soon be made generally available although their utility for the game operation without some form of on-line computer access still appears to be low. The recent developments in shared time and remote on-line access in computer hardware seem likely to solve this problem over the next few years.

It has become apparent that the maximum size of community which can be economically developed in an infinite number of rounds regardless of the information processing system available is to a considerable extent determined by the environmental and technological conditions set at the beginning of play. Thus regardless of the number of rounds which the patience or sophistication of a given group of players may allow, the existence of some maximum community size which is attainable under the given environmental and technological conditions begins to become apparent in 25-30 rounds of play. Such a maximum may, of course, be shifted upwards by allowing modifications in the basic parameters to occur during play representing more recent technological innovations and inventions. The fundamentally deterministic nature of these elements does become apparent in prolonged play, however, and perhaps offers an additional important derivative of the use of models of this type.

In some applications of the model, two transportation terminals have been located at opposite sides of the playing board providing for the simultaneous development of two independent communities at the same time. Extended play of these models often produces the growth of a commercial strip type of development along the highway connecting the two communities. In such a situation it is entirely feasible to allow players to experiment with various forms of economic, technological and legal conditions which either promote or inhibit the development of such forms of intercity commercial belts. A number of other possibilities for using models of this type in such an experimental fashion have suggested themselves and a few will be discussed further below.

Over the past year most of our efforts in developing the game have been directed towards applying elaborations of this basic model to specific real world situations. Thus by specifying the particular pattern and range of transportation facilities and the major topographic features of a particular area we are able to play the game under a set of environmental and technological constraints approximating those which actually exist for a particular urban situation. Such a game may be played over any desired historical period by varying the parameters according to the period in question and by beginning play with the existing land use pattern of the city at the beginning of that period.

A model of the Syracuse Metropolitan Area embracing the northern two-thirds of Onondaga County was developed for use in the planning field problem at Cornell during the fall of 1965. This model involved the utilization of the highway, rail, and barge canal transport facilities available to the area in 1960, the major aspects of slope and drainage then existing, and began with an approximation of the 1960 land use pattern for the area in place. The level of complexity of this model was quite high and cycling time required several hours per round on a manual basis. The most obvious conclusion derived from play was that prolonged operation of a game at this level of complexity would be impossible without some form of computer assistance. At the same time, it became apparent that the scale utilized in the elementary version of CLUG was too gross to allow for detailed handling of patterns of intraurban land use development. The model appeared to be most fruitful at this scale for developing regional patterns of growth and expansion.

At the conclusion of five rounds of manual operation, the model employed appeared to calibrate at the scale of approximately two or three years per round based upon normal rates of annual population and economic growth for a city of this size. The pattern of development which was elicited in the game was surprisingly realistic in terms of the actual growth which has occurred in the Syracuse area since 1960 and that which is generally projected over the next ten years or so. This is not to say that the model was able to meaningfully predict exactly where specific types of land use would occur over the period, but rather that the general direction of growth in the area and its relative magnitude regardless of type of land use did appear to be fairly meaningful. Similarly, the demand for additional services and transportation facilities tended to occur in the same general areas where such demands have developed in the real world since 1960. No argument is intended for the utility of such models as actual predictive devices in such a setting but it has become apparent that when handled carefully they behave in a fairly reasonable manner with respect to future growth potentialities.

Some attempts were made to introduce specific planning functions into this model in terms of existing and proposed land use controls, the provision of future services, and so on. The difficulty of operating the model on a manual basis prohibited any detailed exploration of the pedantic value of such innovations in the game. Our limited efforts in this respect, however, made it clear that students forced to interpret their plans directly into an existing urban situation portrayed by the model and then to "live" with these innovations over a brief period of development tend to consider their plans much more carefully both with regard to their feasibility as well as with respect to their possible impact upon seemingly unrelated aspectsof the urban system.

The final observation derived from last year's experience was that the model was surprisingly easy to construct despite the difficulties incurred in its actual operation without computer assistance. It seems clear that the application of a model at this level of abstraction to virtually any city is relatively simple provided that basic data on land use characteristics, topography and transportation facilities are available. A couple of experienced operators could probably adapt the model to any setting desired within a month's time at fairly low cost.

Based upon our experience with the Syracuse model we are now attempting to develop another model for use in this year's planning classes on a more intensive basis than was possible during the previous year. The basic innovation being made is the introduction of full computer accounting and record keeping in order to provide for both more flexible and speedier play. The area being subjected to the model this year is Cayuga County, New York and its immediate environs focusing upon the city of Auburn, New York in its approximate center. Scale has been modified considerably to provide somewhat greater detail in the designation of land use characteristics and a few additional modifications in utility systems, taxation procedures, number of types of land uses, variety of transportation media represented and so on have been introduced. Although representing a smaller aggregate population than in the Syracuse model, the level of complexity entailed in the game's operation will be somewhat higher due to increasing the scale by a factor of about 20, i.e., a basic residential unit in this model represents about 200 persons as opposed to the 4,000 figure utilized in the Syracuse model.

Use of the model will consist of playing one or two rounds each week with appropriate information stored, analyzed and retrieved on computer tape between each round. When on-line computer facilities become available within the next year or so it is intended to increase the frequency and duration of play so as to allow even greater flexibility in replaying the same sequence of development under varying parameters. Considerable use of this model is anticipated under this year's arrangements and some more systematic attempts to test the educational power of the model will be begun. It is expected that as a result of our experience and modifications of the model in a classroom setting over the next two years, models of this type will become a normal part of the educational setting in a number of courses dealing with urban growth and planning at Cornell and several other universities.

With an increasing amount of work being undertaken in the development of heuristic games of this type at several universities and in professional planning offices, some careful consideration as to their possible utility is called for. Based upon the experience at Cornell with the Community Land Use Game and to a lesser extent with Metropolis and the Planning Operational Game Experiment it seems clear that there are at least three general areas in which such games may prove very useful.

Whether in university classrooms, planning offices or in adult education courses, the utility of academic applications of these kinds of models seems to be clear though not yet conclusively proven. Models such as the Community Land Use Game provide students with basic understanding of major types of land use and some of their locational determinants, the meaning and relevance of community interaction data such as those developed in origin and destination studies, the relative effects of various forms of transportation technology on community growth and organization, the operation of simplified forms of input-output economic matrices, the generalized forms of municipal taxation employed in cities, the impact of mortgage finance and variations in the supply of capital on the local economy, the possible uses and impacts of various forms of land use control, and the kinds of problems of congestion and land shortages which often accompany fairly rapid periods of urban growth. When applied to specific cities, the model also tends to provide a convenient framework for the student to familiarize himself with some of the basic attributes of these cities such as the modes of transportation available locally, the general nature of the topography and community layout, the particular patterns of growth which are being experienced locally and the major points of congestion and public demand for services.

Another possible application of models of this type is as a crude form of laboratory experimental device. Thus with a carefully developed model it is possible to experiment with variations in taxation policy, land use control, and other legal and administrative innovations at fairly low cost and without having to worry about the moral, legal and practical problems of conducting such experiments in real world settings. Several attempts to use derivatives of the Community Land Use Game in this fashion are now under way at Cornell and elsewhere, although the practicality of such applications will not become apparent for at least several years.

Finally, some recent thought has been given to the potentiality of developing heuristic gaming models in tandem with more complex mathematical simulation models. If a game model were constructed in a form which provided a simple analogue to an operating simulation model it would appear that it could help to solve one of the major problems with which simulation models have been confronted, i.e., communication. Many simulation models which work quite well and appear to have considerable predictive power do not become adequately used in policy decisions simply because their operation and the meaning of their results is so difficult to understand for everyone except their designers. Although low in predictive power, gaming models are very strong in their ability to communicate operations and outcomes and in such a linkage a game model and a simulation model would both seem very likely to benefit in greater use and understanding of their results.

FOOTNOTES

(1) Hendricks, Francis H. "Planning Operational Gaming Experiment," a paper prepared for the Northern California Chapter, American Institute of Planners' meeting on "New Ideas in Planning," November 19, 1960 (Mimeographed).

(2) Duke, Richard D. Gaming Simulation in Urban Research. Institute for Community Development and Services, Continuing Education Service, Michigan State University, East Lansing, Michigan, 1964.

(3) Feldt, Allan G. The Cornell Land Use Game. Center for Housing and Environmental Studies, Division of Urban Studies, Cornell University, Ithaca, New York, 1965.

(4) Crecine, John P. "A Time-Oriented Metropolitan Model for Spatial Location," Technical Bulletin No. 6, CONSAD Research Corporation, Department of City Planning, Pittsburgh, Pennsylvania, 1964.

ORGANIZATION FOR THE RETRIEVAL
OF SPATIALLY DISTRIBUTED DATA:
A LAND USE DATA SYSTEM (LUDS)

by

Kenneth J. Dueker*

1.0 INTRODUCTION

As Mr. Fagen stated yesterday, in response to a question--the main job is to organize data properly. This I hope to demonstrate.

I view the concern with data on spatially distributed phenomena and data on places, as a problem area separate from the generally recognized areas of scientific computing and business data processing. I maintain that spatial data create special needs for organizing and ordering data.

In my opinion, the spatial data problem area is technically undeveloped. We must improve our technical capabilities.

Data on usage of land are of major concern in the analysis of urban phenomena. Yet, land use data are difficult to organize and maintain. Conventionally, these data are sequentially ordered, by say, location or ownership number. Sequential ordering, as imposed by serial files, such as punch cards or magnetic tape, impose serious limitations upon accessing land use data. For example, if data are ordered by location and data on a particular type of land use are requested, a search of the entire file is necessary. Existing systems for collecting land use data are conceptially weak--constrained by outgrowths of punch card technology.

The objective of this paper is to describe a conceptual approach to facilitate access to land use data for retrieval and updating. To accomplish this objective, list structures are used to organize land use data inorder to facilitate their access.

2.0 LINKED LISTS

A list is defined as a linked sequence of items. Each

The support of the Applied Computer Research Corporation, Seattle, and the National Capital Commission, Ottawa, Canada, is gratefully acknowledged.

* Departments of Civil Engineering and Urban and Regional Planning, University of Wisconsin.

item contains a pointer to the next item on the list. Thus, the items need not be stored in physical sequence. Stored in addressable storage, the next item of an array is implied, whereas the successor item of a list must be specified. Thus, an item on a list contains both a datum and the address of the next item. This address is called the link or pointer. A nonsense list is illustrated in Figure 1.

A powerful feature of list structures is that the insertion and deletion of data on lists is quite easy. Such operations merely involve changing the linkage of the list. To delete the datum in item c of Figure 1 from the list, the link address of cell b is changed from c to d.

Several list processing languages, IPL V[5], SLIP[6], COMIT[7], FLPL[2], LISP[4], have been developed to facilitate manipulation of list structures. In addition, Green[3] and Bobrow and Rapheal[1] have compared various list processing languages. These languages are designed to take care of the routine housekeeping functions of storage allocation, developing lists, and recursive operations on list items. List processing languages are not new. Principally, they were developed to manipulate complex non-numeric data. This symbol manipulation capability is applied to problems of artificial intelligence, theorem proofing, mechanical translation, simulation, and other problems requiring representation of the relational structure of data as well as their symbolic content. List processing languages provide means of forming lists and manipulating lists. However, list structures must be designed with regard to relationships within the data and with regard to desired access to data. The remainder of this discussion is devoted to conceptualizing a list structure for storing land use data.

3.0 SPATIAL CONTEXT

The spatial context of the basic data observations are specified to be uniform regions of contiguous places that are devoted to the same use. In turn, these land use regions are referenced to an arbitrary grid system and like uses are linked or on a list.

Restated, let there be a separate data observation for each homogeneous land use. An entity consists of characteristics for a homogeneous land use. The number of entities in a given area depends on the nature of land use in an area and upon the land use classification system. Also, let the urbanized area be partitioned into large cellular spatial compartments, such as quarter-square miles, square miles, etc. Uniform data regions that are within two spatial compartments are divided into two entities consisting of two pseudo uniform data regions. The organization of these data depends upon the types of queries that the Land Use Data System (LUDS) is capable of handling.

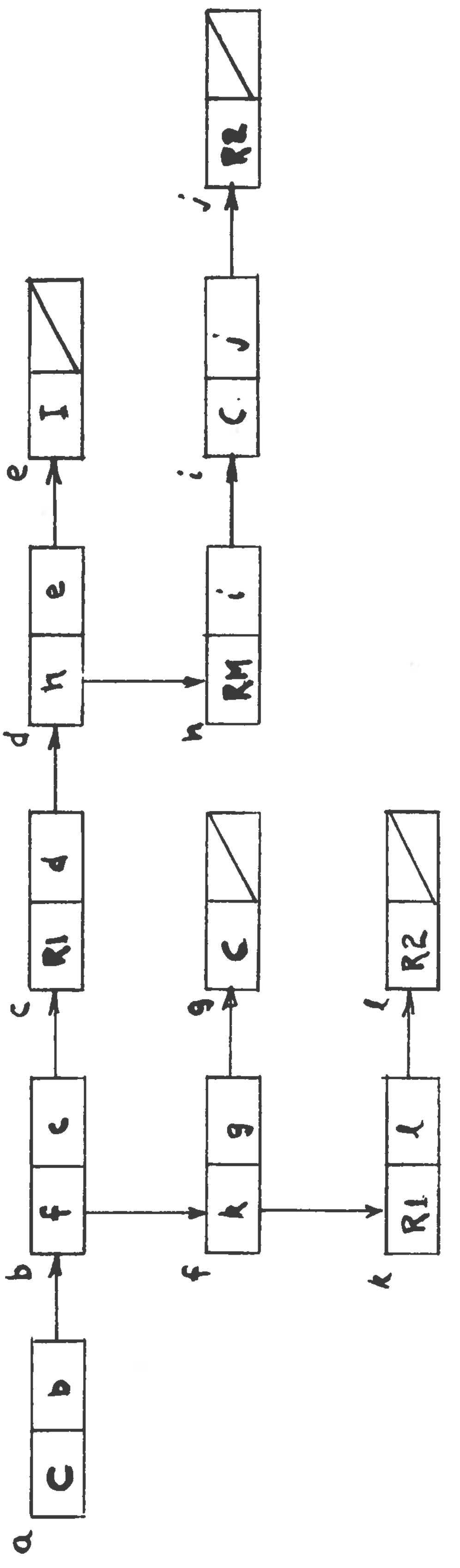


Figure 1. A List

Before further consideration of data organization, the types of queries are presented.

4.0 TYPES OF QUERIES

There are five types of queries to a system such as this. These relate to:

1. Characteristics of land use in spatial compartment x, y.
2. Characteristics of land use of type ℓ.
3. Characteristics of land use ℓ in compartment x, y.
4. Characteristics of query region k (described as a polygon).
5. Characteristics of land use ℓ in query region k.

A design objective is to create a capability of answering queries of this type. What systems requirements do these queries generate? The first type of query requires that the system be capable of accessing the data relating to specific spatial compartments. The second type of query requires a capability to access uses of specified types. The third type of query combines the first two query types for the access and selection of data situated in a specified area of a specific use. The fourth and fifth type of queries relate to needs of accessing land use data within an arbitrarily defined region. The fifth type of query can be considered the general case, with the previous four types being special cases.

Data must be organized so that efficient means for responding to these queries may be designed. Before attempting this, methods of collecting data and updating data must be considered. Consideration of data collection and data updating affects the system design.

5.0 COLLECTING AND UPDATING DATA

The primary concern relating to data collection is that it be systematic, accurate, and efficient. It is proposed that land use data for uniform data regions be collected from maps or controlled mosiacs. These data are organized as entities corresponding to uniform data regions. In addition, each spatial compartment is scanned separately, so that these data are all accessible given the spatial compartment. The order of uses is not important as long as lists are maintained for like uses.

Updating data in LUDS is an important consideration for design of the system. Updating takes several forms. These are generally classified as:

1. Simple Replacement - change characteristics of a land use entity. No change in area or land use classification.

2. Subdivision - replace a land use entity with two or more land use entities for new uniform data regions.

3. Consolidation - replace two or more land use entities with a single land use entity.

A general replacement statement combines these catagories plus more general capabilities.

4. General Replacement - replace m land use entities with n land use entities, where $m \geq n$ or $m < n$.

In general, an updating capability of subdividing areas into different uniform data regions or consolidating areas is desired.

Knowing the types of queries that are asked, the means by which data is collected, and means for updating data, the following data structure is proposed.

6.0 STRUCTURE OF LAND USE DATA

A chained tree is developed to enable access of data in response to the types of questions. At the top of the tree is an A-list. Items on the A-list are accessed given the spatial compartment. There is an A-list item for each spatial compartment and each A-list points to an item on the B-list for the first uniform data region of the spatial compartment. B-list items exist for each uniform data region entity. Items on the B-list point to the next uniform data region within the same spatial compartment, point to the next land use of the same type, and point to the beginning storage location of the actual characteristics of the uniform data region. The actual data characteristics for a uniform data region are not on the B-list, they are on the D-list. An example LUDS tree is presented in Figure 2. In addition, items on the D-list may point to a P-list containing the polygon descriptions for each of the uniform data regions.

A tree, such as the one of Figure 2, is developed by scanning each spatial compartment for contiguous elements of homogeneous land use. An entity for the use is written on the next available position on the D-list. The next available position on the B-list is linked to the D-list record just written. If it is the first use being recorded for a spatial compartment, the B-list address is written on the A-list at an address determined from the mapping function of the spatial compartment value. Also, the B-list address is placed as the

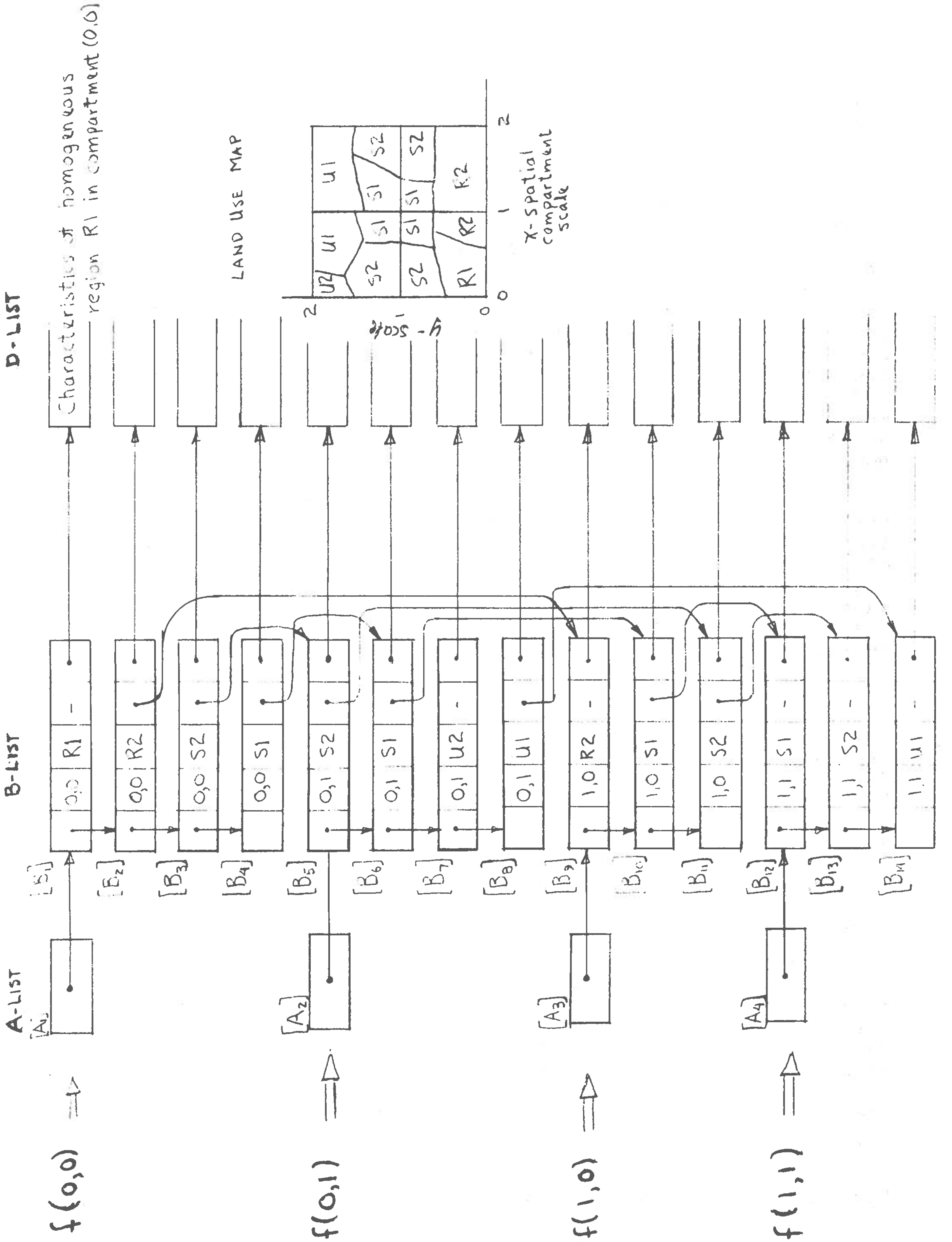


Figure 2 Example of LUDS Tree

fourth element in the immediately prior B-list record having the same land use.

7.0 QUERY RESPONSE

Data structured in this form is easily accessed for response to the types of queries listed above. To determine the characteristics of land use in spatial compartment x, y, an item from the A-list is accessed by a mapping function, $\alpha = f(x,y)$*. The address α contains the location on the B-list containing pointers to characteristics of the first land use type in x,y. The last element of a B-list item points to the location of the actual characteristics of that homogeneous land use region. The first element of a B-list item points to the next B-list item in the same spatial compartment. In this manner, the characteristics of the desired spatial compartment(s) are accessed. Data on land uses in other compartments do not need to be searched.

For the second type of query--to determine the characteristics of land use ℓ--an item of the B-list is accessed. The item to be accessed is determined by retaining a simple array of initial B-list locations for each type of land use. Now, positioned at the first B-list item for land use ℓ the actual data on the D-list can be accessed and the next B-list item for the same land use ℓ is accessible. In this way, just those items relating to land use ℓ need to be accessed.

The third type of query combines the first two. This type of query is concerned with land use ℓ in spatial compartment x,y. This is accomplished by accessing an item on the A-list, as determined by a mapping function that converts x,y to an address. This address points to the first item for compartment x,y on the B-list. Then the B-list items, within compartment x,y, are searched for the first land use ℓ. When found, the fourth element, which points to the next item containing land use ℓ, is used. At this next item, the second element is checked to determine if the spatial compartment is still x,y. If so, transfer is made to the D-list and then to the next item on the land use list.

The fourth type of query is concerned with characteristics of land use within an arbitrarily delineated area, described as a polygon. This type of query is a little more difficult. First the maximum and minimum x-coordinates and y-coordinates must be determined. Then the spatial compartments that these values are within are accessed. Each of the items within these compartments must be tested for inclusion or intersection of the query area polygon with the polygon that define the uniform

---

*This function may be interpreted as a complex number of the form $\alpha = f(z)$ where $z = x + iy$.

data regions.* In this way, land use data that are wholly or partially in a query area are determined.

Finally, the fifth type of query requests data on characteristics of land use ℓ in query area k. Again the query area is represented as a polygon. As in the fourth query type, the appropriate spatial compartments to access are determined from the maximum and minimum x and y coordinate values. Once an item containing land use ℓ is found, the like land use chain can be used to select the remaining items within that spatial compartment to be tested for intersection with the query area polygon. Then the next of the determined spatial compartments is accessed and the process is repeated.

8.0 UPDATING THE LAND USE TREE

The need for updating land use data dictates that these data be stored in list form. In this way, insertion and deletion of items is facilitated. Suppose that the U2 area delineated on the map of Figure 3, by a dashed line represents a new use for that land. Updating to reflect this new use necessitates the modification of existing U1 and S2 regions and the addition of a new U2 uniform data region.

First, the characteristics of the characteristics of the affected U1 and S2 uses must be modified. These are accessed as a type 3 query, i.e. accessing spatial compartment and then land uses U1 and S2.** The appropriate items of the D-list are thus accessed and the new characteristics are written in place of the old.

Then, the characteristics of the new use, U2, are written in the next available D-list location, and a new B-list item is created. References to and from this new B-list item must be made. Figure 3 illustrates the changes to the structure of data. The characteristics of the new use U2 are written in location (D_{15}) and is referenced by the next available location of the B-list, i.e. (B_{15}). A pointer to (B_{15}) is written in the last B-list item (B_8) of spatial compartment 0,1. If there had been a pointer in the like data list, element three, of (B_7) prior to updating, it would be written in element three of (B_{15}). There was none, so it was only necessary to write (B_{15}) in element three of (B_7).

---

*Many of these intersection tests may be avoided if the maximum and minimum x-coordinates and y-coordinates for the uniform data region polygons are among the data characteristics.

**A unique identifier needs to be part of the access key to distinguish between several like uses occurring in the same spatial compartment.

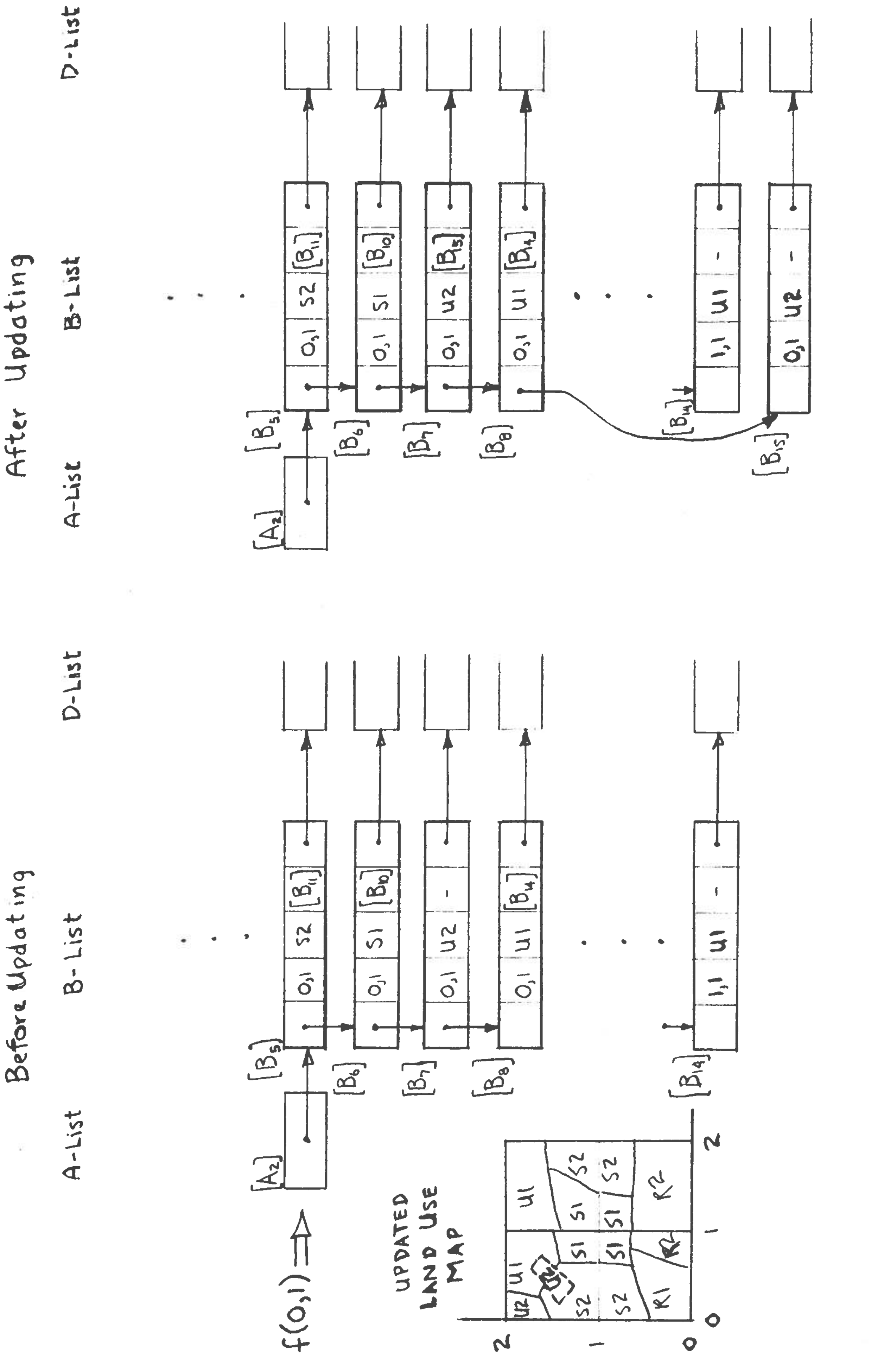


Figure 3 Updating LUDS

9.0 SUMMARY

A system, such as described here, is designed to facilitate access to specified land uses and/or specified locations. It is also designed to facilitate recording of changes in land use. These design objectives dictated the structure of data in terms of lists.

The effect of this is to create the need for a list processor for user access of land use data. The user should not be concerned with transferring from list to list or to different points on lists. This should be automatic given the desired location or land use. A high priority needs to be given to the development of list processing capabilities for manipulating spatial data.

10.0 REFERENCES

1. Bobrow, Daniel G. and Bertram Raphael, "A Comparison of List-Processing Computer Languages", Communications of Association of Computing Machinery, April 1964.

2. Gelernter, H., et al, "A Fortran Compiled List Processing Language", Journal of Association of Computing Machinery, April 1960.

3. Green, B.F., "Computer Languages for Symbol Manipulation", IRE Transactions on Electronics Computers, EC-10, December 1961.

4. McCarthy, J., et al, LISP 1.5 Programmer's Manual, M.I.T. Press, 1962.

5. Newell, A., et al, Information Processing Language-V Manual, Prentice-Hall, (Second edition), 1964.

6. Weizenbaum, J., "Symmetric List Processor", Communications of Computing Machinery, September 1963.

7. Yngve, Victor H., "COMIT", Communications of Association of Computing Machinery, March 1963.

SPECTRAL ANALYSIS OF SPATIAL SERIES+

W. R. Tobler *

Everybody here knows what is meant by a time series. There are many examples: stock market prices, weather records, population of a city over a number of years, and so on. This type of data is conveniently represented in the form of a curve on a graph with time on one axis and the measure of the events on the other axis. By a spatial series, I mean a set of data which uses geographical location for two of the reference axes and has the measure of events on a third axis. An example might be the distribution of population in the United States.

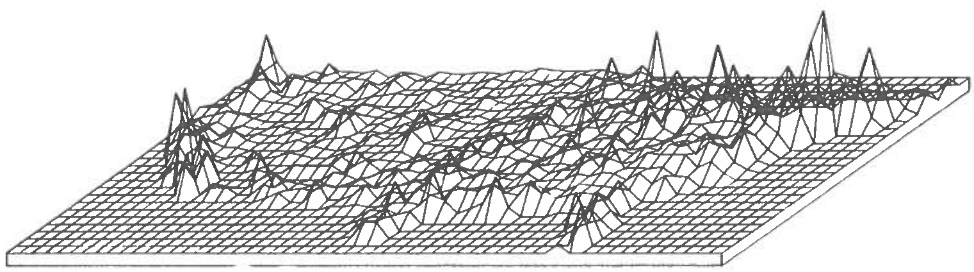


Figure 1: 1960 Population in the United States by one degree squares; square root vertical scale

In recent years a number of individuals have suggested that it would be of value to examine geographically defined data using the

+Supported by National Science Foundation grant GS-1082. The assistance of the University of Michigan computing center in the preparation of the illustrations is appreciated. The data for these examples were assembled by D. Deskins, F. Rens, and R. Yuill.

* Department of Geography, University of Michigan

methods which have been developed for the analysis of time series. What I wish to do today is to explore how this might be done. The subject of my talk, therefore, concerns a statistical analysis procedure. In accord with the expository nature of the discussion it is appropriate to leave out the mathematical details and to concentrate on the conceptual framework.

There seem to be three questions which might be asked: First, just what is time series analysis? Second, is the hypothesis really justified; that is, are there interpretations which make this type of analysis of spatial series worth doing? Finally, if there are such interpretations, why has this method of analysis not already been applied to spatial series?

Taking the first question first, "What is time series analysis?" No simple answer can be given. It is an entire set of methods of analysis developed over about two centuries. I will mention some of these methods in a moment but delay to point out why time series analysis should differ from other types of analysis. Suppose one has a collection of paired observations to be examined. In most situations one can scramble the pairs in any manner and this has no effect on the results. If, however, the observations are from a time series there is an essential order to the data and it is not sensible to scramble the pairs. Somewhat more precisely, the statistician considers that the observations are independent in the one case, but that they are serially correlated in the other case. Exactly the same situation pertains in the analysis of spatial series, where the location of the events is considered an essential part of the observations.

An important strategy employed by statisticians is to decompose data into independent components. For example, in the analysis of variance, one separates the "between cell" variance from the "within cell" variance. In regression analysis the total variability of the data is partitioned into an "explained" component and a "residual". In principal components factor analysis the data is reorganized into "loadings" on orthogonal vectors in a multidimensional factor space, and so on. Mathematically, these methods assume the independence of the individual components and that the total variability of the original data can be expressed as a linear combination of the individual components.

In an exactly comparable manner, time series data can be decomposed into components. These are then usually called trends, for example, secular trends and seasonal trends. A simple way to think of this is to consider the graph of the time series. What is being said is that the original curve can be decomposed into a number of independent simpler curves. Each of these simpler curves explains, in the statistical sense, certain proportion of the variability of the original curve. Adding up all of the simpler curves gets one back to the original curve, but this is not as important as the fact that one may be able to assign a meaningful interpretation to the more "significant" of the simpler curves.

In practice the procedure involves the use of Fourier or trigonometric series of varying amplitude, wave length or frequency, and phase. Unfortunately this is historically also closely connected with the analysis of strictly periodic phenomena, and therefore often regarded as not being appropriate for social data (10,11). The more modern analysis is based on concepts taken from statistical communications theory (3,11,12). Here the concept is not so much periodicity, but rather that any time series, under stationarity assumptions, can be considered to consist of components whose variances can be assigned to frequency bands. This is known as spectral analysis (1).

Computationally, the main difference is that the average dependence of events at one time on events at another time (that is, the autocorrelation of the data), rather than the original data, are submitted to Fourier analysis to obtain a continuous spectral density function.

If one has two sets of time series data, one can compare them by analyzing the relation between the components, rather than comparing the observations directly. This is known as cross-spectral analysis. In terms of the curves, this amounts to calculating the correlation between all of the corresponding pairs of simpler curves.

My second question relates to the possibility of applying these procedures to spatial data. Of particular importance are the interpretations that one might give to such analyses, since mathematically there is no difficulty in substituting a spatial variable for the temporal variable - distance along a highway, for example. The extension to two spatial variables is also direct. Graphically one can imagine that the "surface" is decomposed into simpler independent surfaces.

A very simple example of the application of some of these notions is the moving average. The illustration shows a two dimensional version

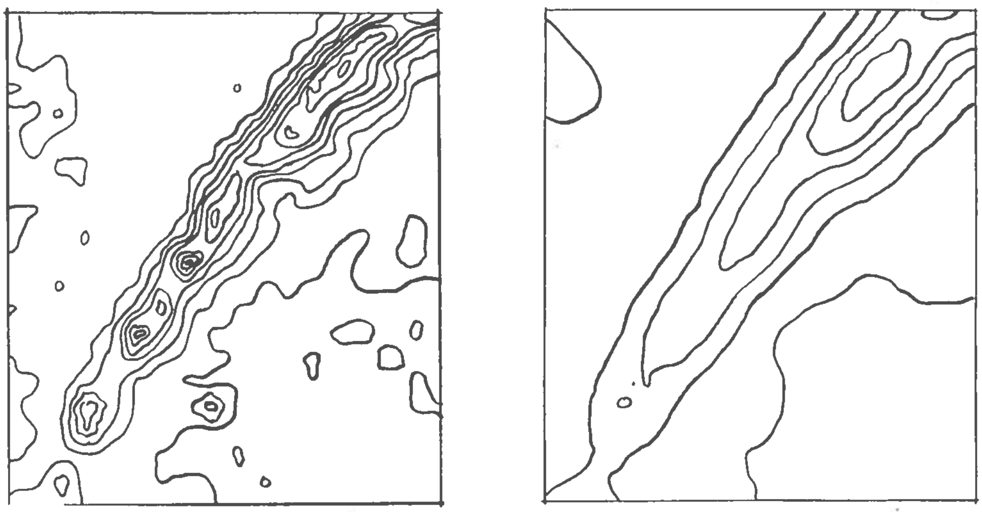


Figure 2: Original and filtered spatial series shown in contour map form

applied to a map. It can be seen that the effect is to simplify the contours, i.e., we have applied a low-pass spatial frequency filter (5). One can argue that the spatial processes which we are attempting to understand contain small scale disturbances. These minor fluctuations distract from our ability to comprehend the general phenomena, and are therefore appropriately eliminated. Alternately, one may be interested in these details; subtracting one map from the other brings these to the fore by acting as a high-pass filter.

A related example is geographical trend analysis, as recently described in the literature (2,4). A practical application of this technique might be the greater acceptance and availability, of geographical data which have been "adjusted" for national, regional, or metropolitan trends, just as economic statistics are now adjusted for seasonal trends. The net effect of this might be a reduction of geographical complexity toward the "homogeneous plain" concept now employed in theoretical work of some elegance. Presumably such adjustments for geographical trends would aid in the evaluation of numerous important practical problems of development and planning.

The notion of persistence in a time series can also be adapted to spatial series. We all recognize that near places are more like each other than are widely separated places. The degree and decay rates of such relatedness can be measured by the two-dimensional autocorrelation function. Using this reasoning a geographically auto-regressive aspect can be built into land use (or other) models. Stated in another way, events at any location might depend on the situation at neighboring locations in a more explicit manner than in the potential model.

Another interpretation is that these procedures, specifically the cross-spectra, allow the comparison of two sets of spatial data; two maps if you will. A further interpretation involves the notion of a transfer function. Consider the geographical distribution of population in say, Detroit at several time periods.

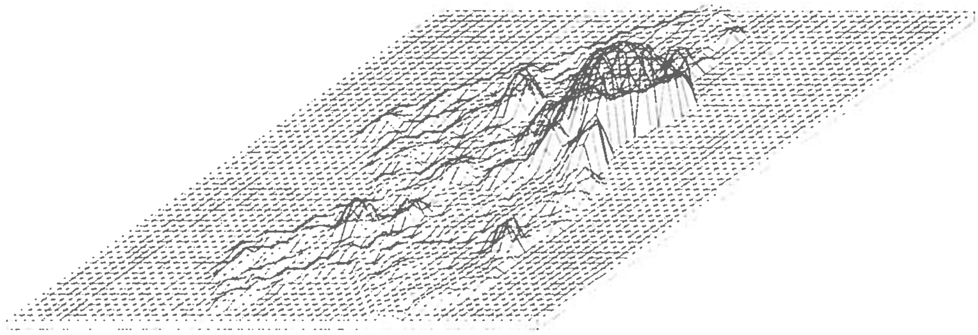


Figure 3: The Input Signal. 1930 Detroit population distribution; 1.5 mile grid; square root vertical scale; "flat" areas: data not available

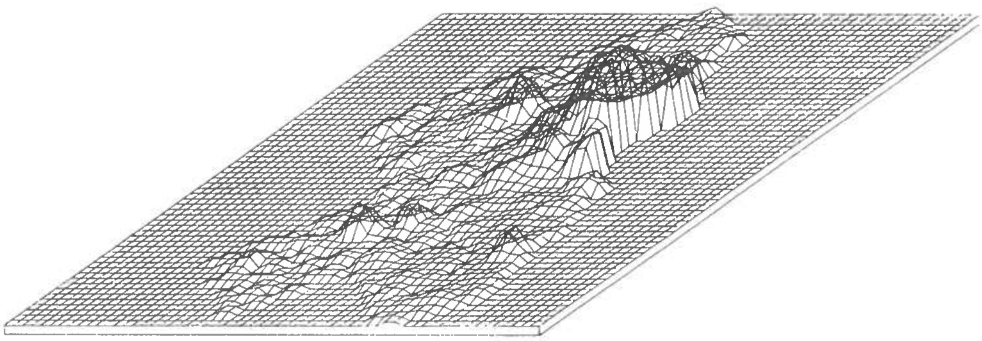


Figure 4: The Output Signal. 1940 Detroit population distribution. Parameters as in Figure 3.

The distribution at one time period can be considered to have been obtained somehow from the earlier period. Stated in another way, the population distribution in 1930 can be considered an input to some process, and the distribution in 1940 can be considered the output from this process:

1930 POPULATION DISTRIBUTION → [?] → 1940 POPULATION DISTRIBUTION

The details of the process are complicated and largely unknown. Both the 1930 and the 1940 distributions can, however, be analyzed in terms of their spectral density functions. Given these two spectra it is possible to calculate the spectral density function of the intervening process. Any realistic urban growth model must necessarily have a comparable transfer function. A crucial question, of course, is whether the process changes between successive time periods:

1930 → [BOX A] → 1940 → [BOX B] → 1950 → [BOX C] → 1960

Oversimplifying greatly, if A operates as an amplifier, does B also act as an amplifier?

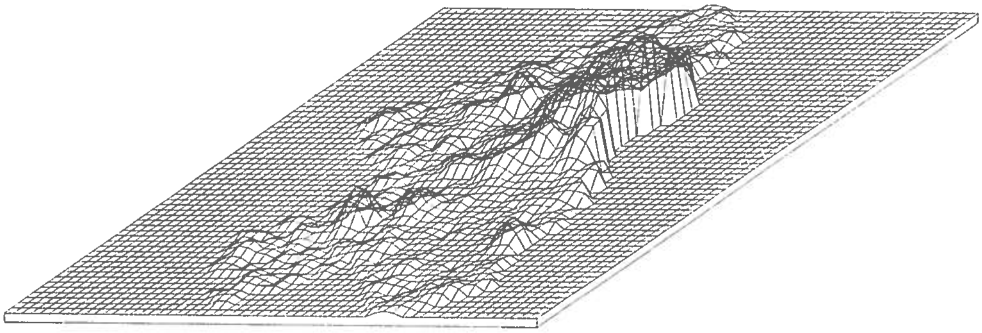


Figure 5: 1950 Detroit Population Distribution

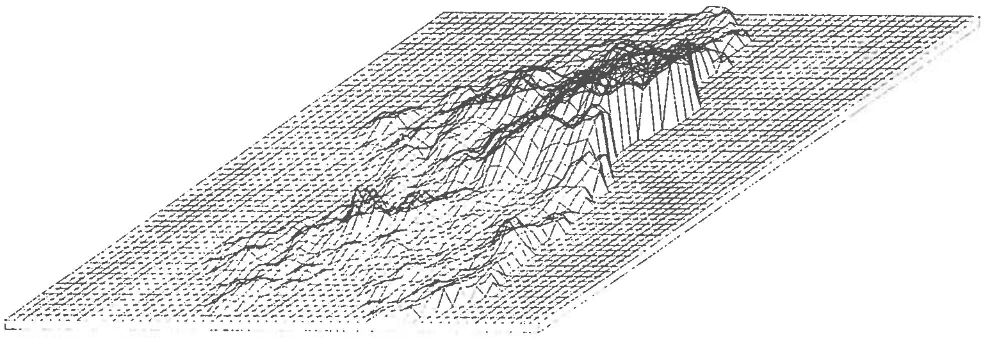


Figure 6: 1960 Detroit Population Distribution

My third question was why these procedures have not been applied in practice, if they are really as valuable as some people would lead one to believe. I leave it to you whether the elementary examples I have suggested would be useful.

Some people might suggest that the lack of application is due to computational difficulties. I do not believe this to be the case, and the methods have been employed in geophysics, particularly oceanography (7,8) and meteorology (6), and geology (9) to a lesser extent. A much more serious difficulty is the lack of adequate data. For the simpler analyses the locational coordinates are required, but this is not sufficient for practical application of the more complicated analyses. In time series analysis a suitable set of data might consist of a minimum of 200 observations taken at equally spaced intervals. In the two-dimensional spatial case a comparable set of data would consist of observations taken at equal intervals in a North-South direction and at equal intervals in an East-West direction. The number of intervals in each direction should be large since the computation of the autocorrelation function has the effect of drastically collapsing the data during the averaging process to about one tenth of the original number. By large I mean between 200 and 1000 cells in each direction.

To the best of my knowledge the first population data collected by square cells was in 1852. Today a number of transportation studies have adopted this practice but it is still not common. It may not be possible to explore the full ramifications of the initial hypothesis until this problem has been overcome.

References:

(1) J.S. Bendat, and A.G. Piersol, Measurement and Analysis of Random Data, New York, J. Wiley, 1966.

(2) R.J. Chorley and P. Haggett, "Trend Surface Mapping in Geographical Research", Transactions, British Institute of Geographers, 37(Dec. 1965), pp. 47-67.

(3) C.W. Granger, Spectral Analysis of Economic Time Series, Princeton University Press, 1964.

(4) P. Haggett, Locational Analysis in Human Geography, St. Martin's Press, 1966.

(5) J.L. Holloway, "Smoothing and Filtering of Time Series and Space Fields", Advances in Geophysics, 4(1958), New York, Academic Press, pp. 351-389.

(6) J.A. Leese, and E.S. Epstein, "Application of Two-Dimensional Spectral Analysis to the Quantification of Satellite Cloud Photographs", Journal of Applied Meteorology, 2(Oct. 1963), pp. 629-644.

(7) M.S. Longuet-Higgins, "Statistical Analysis of a Random, Moving Surface", Philosophical Transactions, Royal Society of London, Series A, No. 966, Vol. 249, pp. 321-387, 1957; also, "Statistical Properties of an Isotropic Random Surface", Ibid., Series A, No. 975, Vol. 250, pp. 157-174, 1957.

(8) W.J. Pierson, ed., "The Directional Spectrum of a Wind Generated Sea as Determined from Data Obtained by the Stereo Wave Observation Project", Meteorological Papers, New York University, Vol. 2, No. 6, 1960.

(9) F.W. Preston, and J.W. Harbaugh, "BALGOL programs and geologic application for single and double Fourier series using IBM 7090/7094 computers", Kansas State Geological Survey, special distribution publication 24, 1965.

(10) Karl Stumpff, Grundlagen und Methoden der Periodenforschung, Springer, 1937.

(11) J. Tukey and R.B. Blackman, The Measurement of Power Spectra, New York, Dover, 1959.

(12) N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, Cambridge, MIT Press, 1949.