in total trips from 1964 to 1977 reflects, of course, the growth in development during this period. The substantial increases in the observed trip lengths, for all three trip purposes, reflect the increased dispersion of activity. Despite these changes, however, application of the 1964-based models to 1977 data closely approximates the observed regionwide trip lengths and interzonal percentages for 1977. It appears that there are no significant biases, except possibly for NHB average trip length, which is underestimated by 10 percent.

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

From the stability tests described in this paper, the following conclusions are drawn:

1. Trip lengths for HNW and NHB trips appear to have more-skewed distributions than those for HBW and TNT trips, which may be a reflection of a short versus long dichotomy within the HNW and NHB categories.
2. Incorporation of a family of travel-function curves, e.g., the Bessel family, selected according to the underlying degree of skewness, appears to be a cost-effective approach to attaining stability across trip type.
3. To attain stability under subarea focusing, refinements to permit explicit treatment of zone-size variation are desirable, if not absolutely necessary.
4. It appears that a key determinant of stability under focusing is the calculation of intrazonal trips, and refinements to allow stratification by zone size appear to be a promising avenue to improved stability.
5. Limited testing indicates that the models described in this paper possess an encouraging degree of stability through time and development patterns.

At least two pressing needs for further research can be identified. The first concerns the fixed penalties, FPENO and FPEND, assessed at the production and attraction end, respectively. If these parameters can be successfully related to real-world trip-end impedances such as parking cost and walking distance, it will be possible to capture the sensitivity of trip distribution to such TSM policies as preferential parking, automobile-free zones, and parking cost strategies. Second, further experimentation with subarea focusing is required to derive generalized rules for handling variation in zone size. The discussion in the section on stability under subarea focusing suggests an approach to improved stability under focusing, but the parameter values derived there may be dependent on the specific problem studied, assumptions concerning approach-link impedances, etc. A generalized strategy for stratifying parameter values by zone size is needed, preferably a method that permits calculation from observable phenomena.

ACKNOWLEDGMENT

The preparation of this report was financed in part through a grant for technical studies from the Federal Highway Administration and the Urban Mass Transportation Administration of the U.S. Department of Transportation and from the U.S. Environmental Protection Agency.

REFERENCES

1. S.M. Howe, T.K. Ryden, and D. Penny. Subarea Diagnostic and Evaluative Procedures for Programming Short-Range Transportation Improvements. TRB, Transportation Research Record 698, 1979, pp. 9-15.
2. N. Nihan and D.G. Miller. The Subarea Focusing Concept for Trip Distribution in the Puget Sound Area. TRB, Transportation Research Record 610, 1976, pp. 37-43.
3. S. Howe and Y. Gur. Trip Distribution in Subregional Analysis. TRB, Transportation Research Record 673, 1978, pp. 165-171.
4. M. Schneider. Access and Land Development. In Urban Development Models, HRB, Special Rept. 97, 1968, pp. 164-177.
5. M. Schneider. Direct Estimation of Traffic Volume at a Point. HRB, Highway Research Record 165, 1967, pp. 108-116.
6. Creighton-Hamburg, Inc. Transportation and Land Development--A Third-Generation Model: Theory and Prototype. Federal Highway Administration, U.S. Department of Transportation, 1969.
7. M.P. Kaplan. Calibration of the Access and Land Development (ALD) Model Travel Function: A Multimodal, Multidimensional Travel Function for Use in Urban Travel Demand Models. Department of Civil Engineering, Northwestern Univ., Evanston, IL, Master's thesis, Aug. 1976.
8. A.G. Wilson. A Statistical Theory of Spatial Distribution Models. Transportation Research, Vol. 1, No. 3, Nov. 1967, pp. 253-269.
9. D.L. Kurth, M. Schneider, and Y. Gur. Small-Area Trip-Distribution Model. TRB, Transportation Research Record 728, 1979, pp. 35-40.
10. B.L. McCarty. Dallas-Fort Worth Urban Area Citizens Survey, 1977: Data Analysis. Transportation and Energy Department, North Central Texas Council of Governments, Arlington, TX, Tech. Rept. 10, Feb. 1978.

New Techniques for Integrating Census Bureau Data and Software with the Urban Transportation Planning System

REBECCA M. SOMERS AND MATTHEW A. JARO

A technology transfer project between the U.S. Bureau of the Census and the Urban Mass Transportation Administration has made Census Bureau software products and simplified Census Bureau data-access procedures available to the users of the Urban Transportation Planning System (UTPS). The Census Bureau software systems for geocoding and computer mapping, as well as utilities for zone definition and area boundary extraction, have been integrated into UTPS

by means of the UTPS Procedure Generation System (UGEN). These software products make census data and software more accessible to the transportation planning process. Geocoding provides linkage to census data areas, user-defined data zones, and geographic coordinates. The mapping software allows quick and easy data display. New software modules for a geographic-data-management system are also being developed. These include the creation of an integrated geographic data base, access methods for the data base, and specific application packages such as address matching.

The Urban Transportation Planning System (UTPS) provides the transportation planner with state-of-the-art methods for the analysis of multimodal transportation systems. UTPS consists of software, documentation, and users' guides and manuals that cover both computerized and manual planning methods. The computer programs are flexible and user oriented; they constitute a major contribution to the current inventory of techniques at the planner's disposal (1). A recent technology transfer project between the Urban Mass Transportation Administration (UMTA) and the U.S. Bureau of the Census has added to this inventory of software by making available software for access to and use of census data.

The Census Bureau provides a wide variety of socioeconomic and demographic data, as well as geographic reference information. In the early 1970s, the Census Bureau developed several computer systems to facilitate the use of its data products. Many of these Census Bureau data and data-application products are useful in the transportation planning process. The systems for address matching and generalized record linkage (UGEO) and line-printer mapping (UCHORO and UGRIDS) have been incorporated into UTPS. Utilities for the definition of planning zones for subsequent address-matching and mapping purposes (UDIME) have also been added to UTPS. The boundary-extraction feature of UDIME is now being added to UTPS. In addition, modules of a geographic data-management system (GDMS) are currently being developed. Facilities for the creation of an on-line geographic data base (GeoModel) and for interactive address matching (ADMATCH'80) have been completed and are ready to be incorporated into UTPS.

Through UTPS, the transportation planner now has convenient access to all of these Census Bureau software products, as well as simplified access to actual census data.

CENSUS DATA IN THE TRANSPORTATION PLANNING PROCESS

Census data are reported in accordance with census geography areas. The boundaries of these areas are generally delineated along the street network and other recognizable boundaries for an area. Thus, the units by which data are reported are areas that can be identified geographically in terms of a linear network. The various census summary tapes contain the statistical data themselves. Specific data files provide data users with geographical reference information concerning the location of these areas.

Census Bureau data are reported according to census geography areas, which range from the block level to the entire United States. These areas are structured in a basic hierarchy of block, block group, tract, county, and state. None of these areas crosses the boundaries of the next-higher area. The areas may be thought of as being "nested." Each area is identified by an identification code for that area type.

There are other area types that are delineated and identified by codes, including standard metropolitan statistical area (SMSA), place, minor civil division and census county division, congressional district, area code, and postal zip code. Some of these represent aggregations of other area types, but they may cross boundaries of the basic area hierarchy.

A geographic base file/dual independent map encoding (GBF/DIME) file is a machine-readable, geographically coded reference file developed by the Census Bureau. It represents a model of the street and block network of an area, usually an SMSA. A GBF/DIME file consists of one record for each street segment. The record contains census geography-area identifiers (i.e., block, tract, county, state, and zip) for the left and right sides of the street segment and the coordinates of the endpoints of the segment. The address range for each segment is also contained on the record (2).

Any of this geographic information may be obtained for any known address within the SMSA through the process of geocoding. For any given address, the street segment on which it is located can be identified. The area code on that specific side of the street segment is then available and may be linked to the original data record that contains the address. In the same way, the coordinates of the intersection at either end of the street segment may be linked to the user's address-related data records.

Although the GBF/DIME files contain geographic coordinates for nodes only, the coordinates of any particular address or of block centroids may be interpolated from these. These points may then be mapped. The GBF/DIME file already contains all the information necessary to map the entire street network of an SMSA or any portion of it. In addition, there are Census Bureau programs that will automatically extract the coordinates of the boundary segments for specified area types. These areas may then be mapped.

DATA INTERFACE WITH UTPS

The transfer of Census Bureau software and data-use technology into UTPS was accomplished by means of the UTPS Procedure Generation System (UGEN). The software provided through this system includes address matching, line-printer mapping, zone definition, and automatic boundary extraction for mapping. Most of this software makes use of GBF/DIME files. The UGEN system handles the execution of the software necessary for any particular process and also handles the necessary manipulation of the GBF/DIME file for that process. However, the existence of the UGEN system itself is practically transparent to the UTPS user. Each application process is operated like other UTPS programs.

UGEN System

The UGEN system was designed to simplify the use of existing Census Bureau software systems that are inherently complex and to automate the specification of these processes for the UTPS user. The usual method of preparing program control cards and the necessary job-control language (JCL) statements for these processes is a labor-intensive and error-prone task. This task may be further complicated for UTPS users when the programs to execute these processes were developed for a context external to UTPS. The user generally prepares a deck of JCL cards and program parameters for an application and submits this deck to the computer-operating system as a job to be executed.

UGEN simplifies this process by acting as an intermediary between the user and the operating system. The user submits the job to UGEN, and UGEN creates a job that is submitted to the operating system. The parameters and specifications that UGEN requires in order to do this are much simpler than those required by the operating system. In the case of the UTPS user, these parameters and control words are in the format required by other UTPS programs. The UTPS user executes a UGEN program (referred to as a macro-procedure) as he or she would any UTPS program. The UGEN macro-procedure then handles the necessary JCL and executes the Census Bureau programs to perform the specific task. Thus, UGEN reduces the drudgery, complexity, and redundancy involved in the specification of sophisticated processes, as well as the learning time required to familiarize oneself with the use of these systems. The UTPS user need not be concerned with the operation of UGEN itself, since UGEN macro-procedures are called on in the same manner as those of other UTPS programs.

The Census Bureau software products that are available through UTPS are address matching, generalized record linkage, line-printer graphics systems, and utilities for the creation of user-defined zones for address matching and mapping (3). These Census Bureau products, which have been integrated into UTPS, facilitate the data input and display functions of the transportation planning process. Geocoding and zone definition facilities make census data and geographic reference information readily available for use in modeling processes. The mapping software provides automatic boundary extraction and graphic display facilities that are easy to operate.

Address Matching

Many transportation and urban planning processes use data files that contain street addresses in the records. For many applications, it is desirable to link certain geographic codes or identifiers to these files. A geographic identifier may be a census tract, a traffic zone, a school district, a Post Office zip code, a policeor fire-reporting district, a block, a congressional district, a pair of geographic coordinates, or any other identifier available. The linking of these geographic codes to data records creates a cross-reference between the data record and other data available according to these geographic areas, such as the census summary tapes or the UTPS zonal data files. Some geographic codes also provide the locational attribute needed to carry out various forms of spatial analysis and mapping.

The process of determining the proper geographic code or identifier for a location defined in terms of address is termed geocoding (or address matching). Geocoding is a special case of the record linkage problem. Record linkage involves two files: a data file that lacks certain information and a reference file that will provide that information. The process involves the comparison of each data file record with records on the reference file in order to find the best match. Once the best reference file record is found, information can be transferred, or linked, to the data file record, thereby providing it with information that it previously lacked. The basis of the linkage process is information that is present in both files, and the result is the transference of information to the data file.

The most rudimentary example of the record linkage process is file updating. In such a case the master file, or data file, must receive information from the transaction file, or reference file. The master file is examined sequentially by using a critical key (such as the record control number) to provide the basis for linkage. When the keys match for a record on each file, the information can be transferred. Another simple example of record linkage is record selection, whereby values on a data file are compared with a list of acceptable values (the reference file). If the data file record contains one of the acceptable values, it will be selected.

Geocoding is a specific application of the record linkage process. The data file, which contains the address information, is matched with a reference file that contains the geographic codes. These codes can then be added to the data file. Address matching is useful for a number of applications. For example, a company may wish to add Post Office zip codes to mailing lists that contain addresses only. The same company may wish to consolidate two mailing lists in order to remove duplicate entries. This entails removing the records that constitute the logical intersection of the two files. In practice, this involves matching, with either file serving as the reference file and with all records that match the data file being discarded. The remaining records (or match rejects) can then be merged with the data file (4).

A transportation planning application may involve geocoding any data file that contains addresses so that traffic zones or census tract codes may be obtained. Once the tract, zone, or other area identifiers have been added to the data file, files that contain data arranged by these areas may be accessed through these codes. For example, origin-destination survey data could be geocoded, thus linking these data to tract data or zone data by means of the geographic code obtained from the matching process. Any other data files containing address information could be geocoded in the same manner. Another important application of geocoding is adding geographic coordinates to a file for the purpose of computer mapping.

In general, geocoding involves linking any type of available geographic code from a reference file onto the appropriate records of a data file. Specifically, the geocoding system available through UTPS links geographic codes found on the Census Bureau's GBF/DIME files with the user's data file. By using the GBF/DIME file as the reference file, the geocoding process involves matching the address (house number and street name) of the records on the data file with the street names and address ranges of the records on the GBF/DIME file in order to obtain the record for the proper street segment. Depending on the side of the segment on which the address falls, the desired type of area code for that side can be linked with the data record. This area code may be for the block, tract, state, county, SMSA, place, minor civil division, zip, or zone identifier. In the same way, the coordinates found on the GBF/DIME file for the particular segment matched may be added to the data record. In addition to the standard census areas and coordinates found on the GBF/DIME files, the user may define planning zones in terms of census areas and may match with these zones.

Geocoding provides address data with the geographic codes needed to access related area data. It also may provide the geographic coordinates for other spatial analyses and mapping. This geographic linking for access purposes makes it easier to incorporate census data into the UTPS modeling systems. Census population, housing, and economic statistics can be accessed at any available level of geography by using the geographic codes obtained through the geocoding process. User-defined planning zones may also be accessed in the same way. Thus, for any particular address, a cross-reference can be established to any data available for the chosen area type into which the address falls--census blocks, tracts, or other planning zones.

The UTPS facility for performing address matching is UGEO, which is a UGEN macro-procedure. UGEO handles the necessary preprocessing of the GBF/DIME file and uses the Census Bureau's ZIPSTAN and UNIMATCH programs to perform address matching. UNIMATCH is a generalized record linkage system that performs all of the various record linkage tasks and the specialized task of address matching (5).

Computer Mapping

It is often desirable to map data that have locational information in order to better understand the spatial characteristics of the data distribution. For example, population or income data could be mapped by census tract to show the distribution across a county. This type of data display by area is termed choroplethic mapping. The variation in data values across areas is represented by different symbolism filling each area. A variation of choroplethic mapping is grid-cell mapping, wherein point data are aggregated to grid cells. For example, average housing values at points representing block centroids may be aggregated to a grid-cell system to show the distribution of such values across a county.

A quick and easy way to produce maps is by means of line-printer mapping packages. Such packages are fairly simple to use and, since most computer installations have line printers, this type of mapping does not require special equipment or personnel.

Two general-purpose line-printer mapping programs have been made available in UTPS through UGEN. These are EASYMAP, a choroplethic mapping program, and GRIDS, a grid-cell mapping program. Both require coordinate and data-value information and a few mapping parameters. The user must decide which of the two mapping techniques is better suited to the presentation of particular data. This choice is based on the number of unique points or areas for which data are available and on their spatial distribution. If areal information is available (such as the boundary coordinates of traffic zones, tracts, counties, or states), then the choroplethic display technique produces the most attractive presentation, provided the areas are not too small or numerous for the scale chosen. If point information is available [such as street intersections (nodes) or centroids of zones], then the grid technique produces an attractive presentation, provided the coverage of the point data is fairly complete with respect to the scale chosen.

The mapping software in UTPS allows quick, easy data display. With very few mapping parameters, choroplethic maps may be produced for area data. The automatic extraction of the boundaries of any specified set of areas frees the user from preparing an input boundary file. For instance, if planning zones had been previously defined, the boundaries could be automatically extracted, and the zone could be mapped by using any zonal data file desired. All census geography areas may also be mapped automatically in this manner. Aggregated grid-cell maps may be produced for point data. The point locations for these data may have been derived previously from a geocoding run. For example, addresses from an origin-destination survey could be geocoded to obtain node coordinates, and the resultant data could be aggregated and mapped at any specified resolution. Street networks themselves may be mapped through a combination of the proper option in either UGRIDS or UCHORO. Again, the coordinates are obtained from the GBF/DIME files.

These mapping systems are used in UTPS through the macro-procedures UGRIDS and UCHORO. Not only is the specification of data format and mapping parameters simplified through the UGEN system, but the formats of these specifications have been made compatible for the two systems. This makes it easier for a UTPS user to learn and to use these mapping packages (6-8).

Zone Definition

Census geography divides space into a system of levels of subareas by which data are collected and reported. These include block, block group, tract, city, county, place, and state. However, these particular areas may not be ideal for every application. In some cases it may be desirable to refer to data arranged by other types of areas.

A utility now available through UTPS allows the user to define zones in terms of census geography. A system of zones may be created by defining each zone as a combination of census geography areas. These zone codes may then be matched in an address-matching pass, just as other geographic areas may. The zone boundaries may also be extracted for mapping, as may other census-area boundaries.

The macro-procedure that provides these utilities is a GBF/DIME file preprocessor called UDIME. According to the options and specifications given, UDIME prepares a GBF/DIME file for geocoding with UGEO, adds zone definitions or redefinitions to this file, and extracts area boundaries for mapping. For example, a particular set of planning zones may be created in which each individual zone is defined as a combination of census geography areas. Addresses could then be geocoded to these zones, rather than to census tracts, blocks, etc. These zone boundaries may subsequently be extracted to form a boundary file for mapping by means of UCHORO.

GEOGRAPHIC-DATA-BASE MANAGEMENT SOFTWARE

Although these software products provide many spatial data-handling and planning-application functions, there is still a need for integrated modules for use in geographic-data-base management. Research and development in this area, whose purpose is to integrate such modules into UTPS, is currently under way, and a geographic-data management system for transportation planning applications is currently being developed. The data base for the system is derived from the GBF/DIME files and is based on the topological relationships inherent in these files. The data-base access methods are based on these topological relationships between map entities and the set relationships between hierarchical systems of nested areas. This data model and set of access methods is referred to as GeoModel.

The GeoModel data base may be accessed directly through any element, whereas a GBF/DIME file must be accessed sequentially. GeoModel represents the creation of a geographic data base from the information contained in the GBF/DIME files and the integration of this information with other user-supplied geographic information, e.g., zones. Such a data base can support low-level data access and manipulation by computer programmers and high-level application software that can be readily used by nonprogrammers. Application software, such as address matching and network editing, are also currently built around this data base. These data-base creation, access, and application functions are available as independent software products themselves. The following sections describe the creation of the data base and the geocoding modules that have now been delivered and will be incorporated into UTPS shortly.

GeoModel

GeoModel is a system of data structures and access methods that provide a geographic data base for a wide variety of planning functions. GeoModel consists of a data storage system; various input, manipulation, and output subsystems managed under a control language; and special-purpose application packages. The purpose of GeoModel is to provide the planner with access to an integrated, flexible geographic data base that does not dictate how the data must be structured or which sort of application must be used. The approach is not to anticipate what the user will want to do, but rather to provide a means to combine primitive access functions in many different ways. The data access facility, implemented as a language that is extensible and recursive, permits the planner to build complex access, edit, and display functio~~ns. The system~~ may then be used in ways that the designer never anticipated. In short, the design philosophy was not to provide highly structured user application packages but to provide basic and powerful data access functions that could be developed and combined to form utilities tailored to the specific needs and applications of various users (9).

The skeleton of GeoModel is created from the Census GBF/DIME file. This file constitutes a model of a map of a street network. It contains the topological relationships between the map elements--street segments, nodes, and blocks--and the census geography-area codes for each block--block number, block group, tract, etc. The GeoModel data base maintains the topological links between map elements, as well as the set relationships between the various levels of geography. An important goal of GeoModel, however, is to insulate the planner, to whatever extent is desired, from the census-defined geography. The planner may define units that are most useful to him or her and may operate exclusively in these units.

Creation of a GeoModel Data Base

The basis for the GeoModel structure is the GBF/DIME file. A program called GMLOAD converts any given GBF/DIME-format file into the GeoModel on-line data base. Several integrated files are created, including a street-segment file, a node file, a block file, and files to support applications such as address matching. These files are interrelated according to the topological relationships on the GBF/DIME file.

Access to the data base is very flexible and, by using the data-access language or any other means developed by the user, complex data-access and manipulation capabilities may be developed. The data base may be accessed directly through any particular element (a street segment, for example) and the related elements may then be found (the bounding nodes and blocks, in this case). Basically, the information available after the initial creation of GeoModel is the same as that on the GBF/DIME file, but the GeoModel data structure makes it possible to access it directly, quickly, and in many different ways.

ADMATCH'80

ADMATCH'80 is an example of how the GeoModel data base may be accessed and used. ADMATCH'80 represents a new concept in geocoding technology. The ZIPSTAN-UNIMATCH operation is based on syntactical and pattern-recognition preprocessing. ADMATCH'80 bypasses all preprocessing and sorting, and it derives its match scores directly from the distribution of address components in a particular city. Thus, ADMATCH'80 automatically adjusts itself to the characteristics of the specific area being processed. Access to the data base is through the phrase files for street names that are to be matched with the incoming address. If the match is successful, the street-segment record for that address is retrieved.

ADMATCH'80 is an interactive address-matching system. It can match a single address or an entire address list interactively or in batch mode. In interactive mode the user will be presented with a list of match candidates for ambiguous addresses. By using ADMATCH'80 and access methods to the data base, related operations (such as intersection matching) may be performed.

SUMMARY

The integration of Census Bureau software products into UTPS has made new procedures and data access available to the transportation planner. Many of these techniques would be complex and time consuming if performed by user-developed methods or even through the use of the available Census Bureau packages. The UGEN system has simplified the use of geocoding and computer-mapping software for the UTPS user. Further developments in the area of geographic-data-management modules and related applications will make even more flexible and diversified techniques available to the transportation planner.

ACKNOWLEDGMENT

The Census-UMTA Technology Transfer Project was funded under contracts with the Office of Planning Methods and Support, UMTA. The software was developed and documented by the Programming Languages Research Branch, Systems Development Division, U.S. Bureau of the Census, under the direction of Matthew A. Jaro. The GeoModel data structure and access methods were designed by Jay R. Baker, under the direction of Matthew A. Jaro. ADMATCH'80 was designed by Jay R. Baker and Matthew A. Jaro. Major theoretical foundations were provided by Marvin S. White, Applications Mathematics Research Staff, Statistical Research Division. Software design, programming, and documentation were performed by Matthew Jaro, Jay Baker, Randall Gacek, Helen Ireland, Charles Murphy, and Rebecca Somers, all of the Programming Languages Research Branch.

REFERENCES

1. Urban Transportation Planning System: Introduction. Urban Mass Transportation Administration and Federal Highway Administration, U.S. Department of Transportation, Aug. 1979.
2. DIME: A Geographic Base File Package. U.S. Bureau of the Census, no date.
3. M. A. Jaro and others. Introduction to Census/UMTA Technology. U.S. Bureau of the Census, Nov. 1978.
4. M. A. Jaro and J. R. Baker. Census Data Interface, Volume I: Geocoding Technology for the Transportation Planner. U.S. Bureau of the Census, Feb. 1979.
5. UNIMATCH: A Record Linkage System. U.S. Bureau of the Census, 1978.
6. EASYMAP Computer Mapping System. U.S. Bureau of the Census, 1978.
7. GRIDS: A Computer Mapping System. U.S. Bureau of the Census, 1972.
8. R. M. Somers. Census Data Interface, Volume II: Computer Mapping for the Transportation Planner.

U.S. Bureau of the Census, May 1979.
9. J. R. Baker. GeoModel: Integrated Data Structures for Representing Geographic Entities. In Proc., Fourth International Symposium on Computer-Assisted Cartography (Nov. 4-8, 1979), American Congress on Surveying and Mapping and American Society of Photogrammetry, Falls Church, VA, 1980, pp. 275-282.

Introduction to Aggregate Data Analysis by Using UTPS: UMATRIX

ROBERT B. DIAL AND LAWRENCE F. QUILLIAN

The effectiveness of any transportation planning operation depends largely on its ability to acquire, manipulate, and present data. The Urban Transportation Planning System program UMATRIX, a powerful data manager, provides the transportation planner with capability in summarization, modification, and display of data. As numerous examples ranging from simple data preparation to application of complex travel-demand-forecasting models demonstrate, the UMATRIX user can access a variety of data forms and can accomplish analytical chores that have been difficult and time consuming. UMATRIX's ease of operation provides an effective and flexible tool for the transportation planner.

The effectiveness of any transportation planning operation depends largely on its data-management ability. In the evaluation of long-range strategies, the analysis of short-range tactics, or merely the observation of the status quo, a planner collects, refines, forecasts, summarizes, and reports data. The Urban Transportation Planning System (UTPS) consists of computerized and manual planning techniques developed, maintained, and distributed by the Urban Mass Transportation Administration (UMTA) and the Federal Highway Administration. New UTPS software that readily creates and manipulates particular types of data considerably reduces the time and cost of transportation planning. This paper describes one of these new programs--UMATRIX.

DATA TYPES

The planner must understand and forecast both the transportation and the demographic characteristics of the region. The data consist initially of network data (which describe the highway and transit infrastructure), survey data (household interviews, on-board surveys, cordon counts, etc.), and data from secondary data sources (Census Bureau information, marketing service summaries, etc). As the planner processes this large data base, he or she creates additional data for purposes of forecasting, plan evaluation, and decision making. Although they are of enormous variety in substance, a large portion of the planner's data falls into two categories: zonal data and interzonal data.

Zonal Data

Zonal data are collected and aggregated into some predetermined geographic "areal unit" configuration. Zonal data (whether forecast or observed) provide such information as population, average income per household, dwelling units, trip attractions or productions stratified by trip purpose, transit route miles, vehicle miles traveled, lane miles, carbon monoxide emissions, energy consumption, and total congestion time for each unit of local geography. Zonal data are read and written by numerous UTPS programs. Among these are UCEN70, MBUILD, AGM, UMODEL, and UMATRIX.

Figure 1 shows a hypothetical five-zone region; population figures are printed along with the zone number. Each zone's specific zonal population value is shown in Figure 2 as an indexed list for the zonal attribute "population". This can be imagined as a single array that has two components: an implicit row index and an explicit cell value. Stored in the array is a list of attribute values (LAV) in which each cell value contains the data for the zone whose number corresponds to the cell's row index (e.g., in cell 5 resides the value of population for zone 5--5900 persons).

Interzonal Data

While zonal data attributes are related to individual entities, interzonal or matrix data are related to a pair of entities. These matrix data relate various characteristics to a given set of origins and destinations. Often these values are summed from the individual link data that constitute the network, and each cell value represents a value for time, trips, fares, modal shares, or distances from one zone to another. Interzonal data constitute the input and output of numerous UTPS programs, notably UROAD, UPATH, AGM, and UMATRIX.

Interzonal data provide a matrix that gives a value for a trip from origin zone I to destination zone J. Figure 3 shows peak-hour home-based-work (HBW) person trip values from zone 5 of a hypothetical region to all other zones. Trips from zone 5 and all other regional trip interchanges are shown in Figure 4. In this case, this matrix of interchange values can be shown as a table of cells that has three components: a row index (I), a column index (J), and a cell value. Thus, for example, for a five-zone region, the matrix contains 25 cell values of trips between each origin and destination zone pair. Row 5 and column 3 refer to the number of peak-hour HBW person trips (230) from origin zone 5 to destination zone 3.

UMATRIX

The UTPS program UMATRIX is a powerful zonal and interzonal data-management system. It provides the planner with complete data-analysis capabilities. It accommodates the large range of data forms required for transportation planning, including matrix data, household surveys, zonal-based demographics, and network characteristics.

The UMATRIX user can access various input data