Computerized Method for Updating Planning Data Bases Used in Travel Demand Forecasting

LARRY W. McPHERSON, CLINTON L. HEIMBACH, AND LARRY R. GOODE

Data bases that have been created for use in urban transportation studies describe the urban environment for only one point in time. With the exception of simple manual techniques, no provision has been made for updating these planning data bases on a continuing basis. The objective of this study was the design and implementation of a computerized information system capable of supporting the continuing socioeconomic data requirements for the travel demand forecasting phase of the urban transportation planning process. Agency operating and administrative records from state and local governments serve as system input. The system performs a geographical analysis that determines the home base locations for these records and, hence, the data that they contain. Subsequent to this geographical analysis, the system aggregates the data on a small area basis required by the travel demand models in the planning process. Requirements for data confidentiality established by law for certain socioeconomic data led to the development of a mathematical model that predicts the income variable. The model is based on harmonic analysis and is specified as a Fourier series. The transportation planning information system developed during the course of this research is capable of synthesizing, on a small area basis, the demographic and employment data used in the transportation planning process for urban study areas that have populations less than 500,000. This conclusion is supported by the findings from system implementation and testing in Greensboro, North Carolina.

Data bases created for urban transportation studies describe the urban environment for only one point in time. With the exception of simple manual techniques, no provision has been made for updating these planning data bases on a continuing basis. The difficult, and sometimes impossible, task of continually updating the transportation planning data base is reflected in both direct and indirect costs. Direct costs are straightforward expenditures of time or money. An example of direct cost is the 8 person-months required to manually update the employment and dwelling unit inventory for an urban area that has a population of approximately 200,000 persons. Indirect costs accrue whenever the transportation planner cannot respond to the transportation needs of the urban area because the planning data base is not kept in an accurate, complete, and current condition. Indirect costs tend to be a function of how well the initial planning data inventories are updated.

Currently, no means exist, except on an ad hoc basis of staff estimates, for keeping the planning data inventories of household size, household income, and car ownership updated. Even though quantification of these indirect costs is difficult, over time the indirect costs would exceed the direct costs involved in updating the planning data base. Therefore, this study focused on the development of a computerized system capable of monitoring and recording those changes in the urban environment that can then be used to update the transportation planning data base periodically. The updated transportation planning data base can then be used to produce estimates of travel demand derived at the household level, or any other higher level of aggregation desired.

LITERATURE REVIEW

Writers in the field of transportation engineering generally agree that the ultimate success of the urban transportation planning process (UTPP) depends in large part on an information system that can provide a source of current, accurate, and timely planning data. These same writers are in less agreement, however, about the design specifications for such an information sytem. Regardless of the various features that could be manifested in such a system design, the consensus was that any information system designed to support UTPP must be capable of monitoring and reflecting socioeconomic change that may occur in single-family households throughout the planning area under consideration. The single-family household is the basic unit of analysis used in the travel demand phase of UTPP and is the reason that any well-designed data monitoring and retrieval system must be capable of quantifying change in any of several household characteristics, including (a) family size, (b) family income, and (c) family car ownership. Street addresses were the suggested means for linking or referencing these data to specific household units in the planning study area under consideration.

The literature review disclosed no general theory of information system design that could serve to establish an operational framework for a particular data system capable of supporting the data requirements for transportation planning. Recent work, however, in the disciplines of information and computer science allows a logical and common-sense design approach to the urban transportation planning information system (UTPIS). One such conceptual design by Peat, Marwick, Mitchell and Company (PMMC) demonstrates such an a posteriori approach to UTPIS development (1).

The major findings of the PMMC study that are germane to the research reported herein are as follows:

1. The potential for using secondary data sources (i.e., government records) to update transportation planning data bases and
2. The use of computerized geocoding techniques for linking these data sources to geographical areas of interest.

METHODOLOGY

The design and development of the North Carolina Urban Transportation Planning Information System (NCUTPIS) required two analytical techniques. They can be listed in order of importance to this research effort as: (a) computerized geocoding and (b) harmonic analysis.

Computerized Geocoding

The single most important element in the design of NCUTPIS was the geocoding subsystem that can assign locational (areal) codes to government source-data records automatically. A geocoding system provides the mechanism for linking together a variety of diverse data sources that may have only one common link (e.g., street addresses). After the data have been linked to specific areal units (e.g., traffic zones) they can be aggregated to any other higher level of spatial detail. The geocoding process provides a convenient framework for continuous monitoring and updating of transportation planning data. The address matching and geocoding process is illustrated in Figure 1 (2).

Figure 1. Address matching and geocoding process.

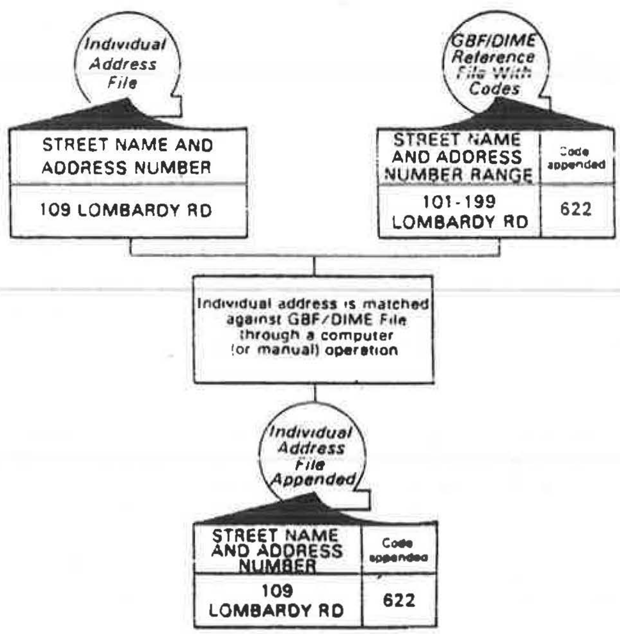


Figure 1 shows that address matching is simply the process of matching data records in two files (address and reference) on the basis of street name and number. In the example 109 Lombardy Road (address file) is first matched to a range of addresses 101-199 along Lombardy Road (reference file). After the match has been made the geographic identifier (geocode) 622 is appended to the matched record. This process of geographic identification is referred to as geocoding.

Once a file of individual addresses has been geocoded, any data related to those addresses can be summarized according to any geographical or areal units of interest. For example, the total number of vehicles can be tabulated according to the traffic zones in which they are registered.

A computerized geocoding system is inherently complex but potentially powerful. An example of such a system is the UNIMATCH geocoding procedures developed by the U.S. Census Bureau. UNIMATCH is complex but has a powerful potential for geocoding. The U.S. Department of Transportation (USDOT) sponsored a project in 1978 aimed at facilitating use of that system. The UNIMATCH system includes procedures for address matching, printer graphics, generalized record linkage, and programmable text generation (3). The technology transfer project sponsored by the USDOT consisted of installing these procedures in the Urban Transportation Planning System (UTPS) environment for convenient use by transportation systems analysts, engineers, and planners. This interagency project was successful and has resulted in the most sophisticated, generalized record linkage system currently in existence in the field of transportation (4). These address matching-geocoding procedures were first field tested successfully at the North Carolina Department of Transportation's (NCDOT's) Computing Center in June 1979.

Harmonic Analysis

Harmonic analysis was the second most important analytical technique used in developing NCUTPIS. Harmonic analysis was used to express the income-estimating model as a Fourier series. Functional expansions in Fourier series and their uses are part of a branch of mathematics known as orthogonal functions. These functions have many important engineering applications because of their curve-fitting characteristics. Often a Fourier series is written for some function whose values are given numerically. The process of finding the Fourier coefficients for such a function is called harmonic analysis (5).

Many applied mathematical texts treat the situation where a given function [f(X)] can be expanded over the interval (-L, +L) and into a series of the type:

$$f(X) = (a_o/2) + \sum_{n=1}^{\infty} a_n \text{Cos}\,[(n\pi X)/L] + \sum_{n=1}^{\infty} b_n \text{Sin}\,[(n\pi X)/L] \quad (1)$$

and a_n and b_n are given approximately by

$$a_n = 2/K \sum_{P=o}^{K-1} f(X_p) \text{Cos}\,[(n\pi X_p)/L] \quad (2)$$

and

$$b_n = 2/K \sum_{P=o}^{K-1} f(X_p) \text{Sin}\,[(n\pi X_p)/L] \quad (3)$$

when the expansion interval (-L, +L) is divided into K equal parts of length $\Delta X = 2L/K$ and X_p (P = 0, 1, 2, . . . K-1) are the finite division points.

Income Model Construction by Harmonic Analysis

Manual construction of income models can be a tedious and error-prone task. The computer program HARANAL was written to help with this developmental task. The program is designed to read tabular percentages, fit the data with a harmonic series, and prepare a printer plot of the results (6).

Data for the calibration of income models are usually available from comprehensive origin and destination surveys. Other sources of data are the 1970-1980 Census Urban Transportation Planning Packages. Car ownership-income models have been developed from census data for a number of urban centers across the United States, including the North Carolina cities of Greensboro, Winston-Salem, High Point, and Raleigh (7). None of these models, however, was based on mathematical analysis.

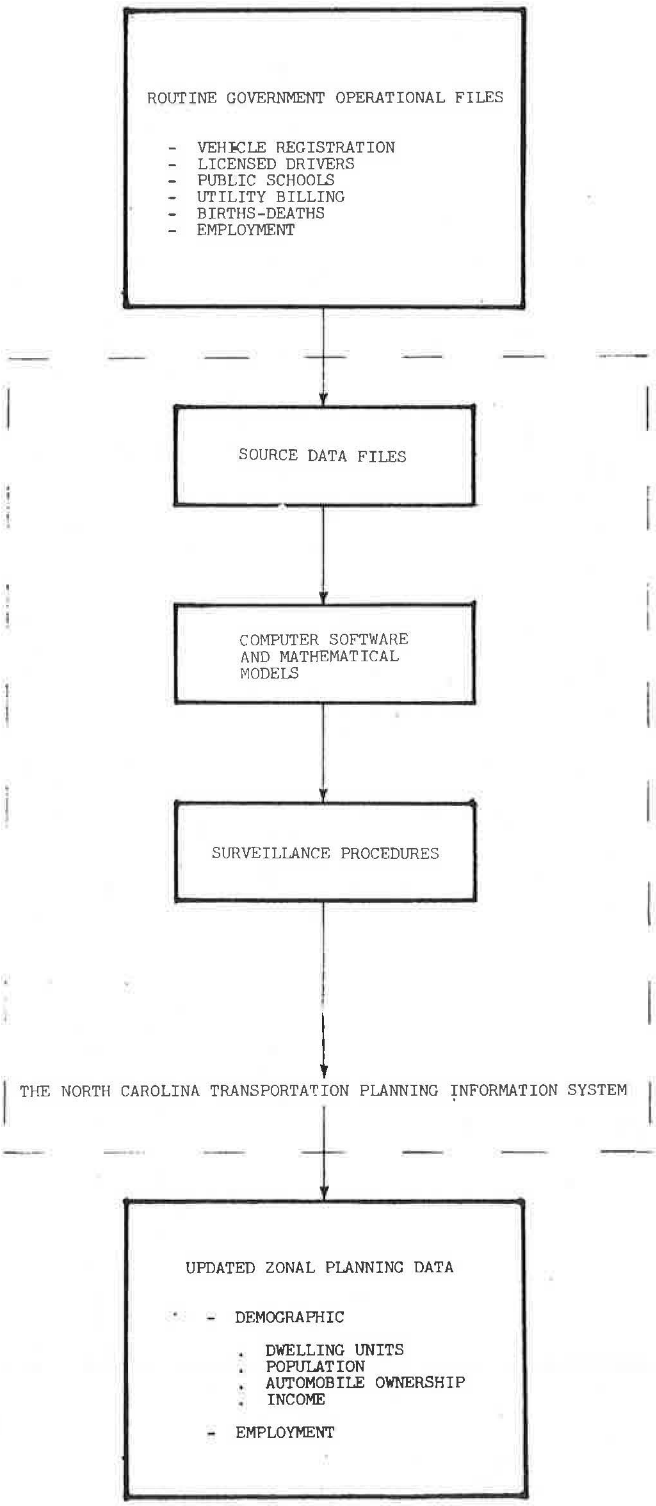
CONCEPTUAL STRUCTURE OF NCUTPIS

Although the conceptual structure of the transportation planning data system developed during this research was based on the PMMC study, it is distinguished from that work in three important aspects.

1. In the design of NCUTPIS major emphasis was placed on the travel demand data requirements of North Carolina's current urban transportation planning process. No attempt was made to develop a system that would meet all of the urban center's potential land use and transport planning data requirements. Design emphasis was placed on meeting the immediate needs of the continuing UTPP and also on providing the flexibility for future expansion of the system to include the collection of other high-priority data items.
2. An attempt was made to make NCUTPIS more user-oriented by use of the geocoding procedures found in the familiar UTPS environment.
3. NCUTPIS includes computerized techniques for development of an income model.

The conceptual structure of NCUTPIS that was designed to produce current planning data used for travel demand is shown in Figure 2. Figure 2 illustrates four major components. The first component consists of six source data files. All of these source files can be obtained from the routine operating or administrative records of state or local government agencies. These six files are used to feed original source data to NCUTPIS . Table 1 summarizes the data contents of each file and also indicates ownership and geographic coverage.

Figure 2. Conceptual structure of NCUTPIS.



The second major component of NCUTPIS consists of all the computer programs needed to extract, format, merge, sort, and summarize the planning data. This software is contained in the UTPS Macro Procedure Generator System (UGEN) or maintained in the utility procedure library of NCDOT's Computing Center.

Mathematical models make up the third major system component. As expected, limitations exist on the use of original source data. In many situations legislation imposes confidentiality requirements on the use of agency data that relate to employment, income, revenue, or sales. Mathematical models can be constructed and used in a manner that complies with established confidentiality requirements. The income-estimating model used in NCUTPIS is a good example of the synthesis of sensitive data.

The fourth system component consists of two surveillance subsystems designed to produce current demographic and employment data on a small zonal basis. The demographic subsystem is designed to produce the following data for each traffic analysis zone in the study area:

1. Total number of single family housing units,
2. Total population,
3. Population less than 6 years old,
4. Population 6 to 17 years old,
5. Population 18 to 64 years old,
6. Population more than 64 years old,
7. Total number of vehicles,
8. Number of vehicles by age group,
9. Number of vehicles by weight classification,
10. Number of trucks,
11. Number of pickup trucks,
12. Number of motorcycles, and
13. Average income.

The employment subsystem is designed to produce estimates of employment by place of work for each traffic analysis zone broken down by the standard industrial classification (SIC) categories.

Source Data Files

Six magnetic tape, source data files are used as input into NCUTPIS and are listed in Figure 2. Table 1 summarizes the data content of these source files.

Motor vehicle registration was obtained directly from NCDOT's Division of Motor Vehicles vehicle master file. Data contained in this file can be geocoded and summarized by traffic analysis zone. These summaries can serve as direct input for travel demand models used in the continuous planning process. They also serve as the primary input to the income estimating model.

Data on dwelling units were obtained directly from public utility water billing records. Other public utility billing records, including telephone and electricity, can provide additional housing data. These billing records are geocoded and summarized by traffic analysis zone. The housing counts derived from these billing records serve as direct input into travel demand models.

Data on the population age group 16 years and older were gleaned from the NCDOT's Division of Motor Vehicles' licensed driver master file. This file required significant modification before it was used in NCUTPIS. The driver file normally resides in disk storage. The driver records had to be extracted and output as a tape file before they could be geocoded by NCUTPIS. Population data on the age group 5 to 15 were obtained from public school registration records. Usually two separate school files must be geocoded, one from the county and one from the city. Population data on the age group birth through 5 were obtained from the Department of Human Resources.

Table 1. Data file source, ownership, geographic coverage, and use restrictions.

Source of Data	Ownership	Type of Data	Geographic Coverage	Restrictions
North Carolina DOT's Division of Motor Vehicles vehicle master file	NCDOT	Vehicle registration by type and owners' addresses	State and county	Confidential data—release controlled by Division of Motor Vehicles
Local utility billing master file	City	Dwelling unit addresses	City	None
North Carolina DOT's Division of Motor Vehicles licensed drivers master file	NCDOT	Licensed drivers' addresses	State	Confidential data—release controlled by Division of Motor Vehicles
North Carolina Public Schools	County and city	Student addresses	County and city	Confidential data—release controlled by county or city public schools
North Carolina Department of Human Resources birth/death file	Department of Human Resources	Addresses of births and deaths	State	Confidential data—release controlled by Department of Human Resources
North Carolina State Employment Security Commission summary file	N.C. State Employment Security Commission	Address of employment by SIC code	State and county	Confidential data—release controlled by state law through Employment Security Commission

Employment data are the single most difficult data that the transportation planner has to assemble on a zonal basis. The most promising sources of employment data are the files maintained by state departments responsible for administering unemployment insurance programs (i.e., the State Departments of Labor and Employment Security). NCUTPIS extracts employment data from the Employment Security Commission's master file.

Collectively, the six source data files shown in Figure 2 provide the current, raw data input to NCUTPIS. This system extracts and then, on a zonal basis, summarizes the socioeconomic data found in the source files. These data can then be used directly or indirectly as input to travel demand models.

Computer Software

The set of computer programs used to support the operations of NCUTPIS includes two major components:

1. The address matching and geocoding procedures and
2. The generalized utility procedures used for the sorting, merging, summarizing, manipulating, and analyzing tasks.

These procedures are currently maintained at the NCDOT Computing Center. This center will supply program documentation on user request.

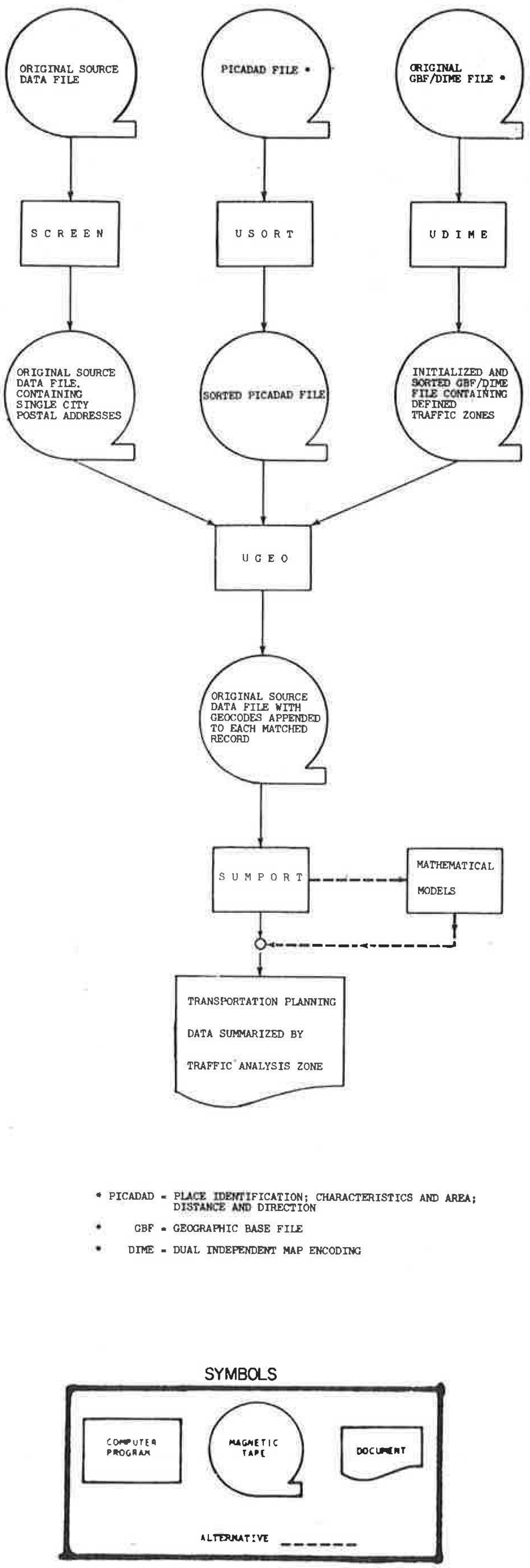
Address Matching and Geocoding Procedures

The process of record linkage is sometimes referred to as file matching. File matching involves the transfer of data from one file to another when certain matchkeys and predefined criteria are met. Address matching is an important special case of record linkage. Four UTPS computer programs accomplish the address matching and geocoding tasks (3):

1. UGEN--The UTPS procedure generator (UGEN) was written so that some severe limitations of IBM procedures (PROCS) could be overcome. These limitations include (a) no conditional generation of data description (DD) cards, (b) PROCS could not be used to generate or modify program data cards, and (c) PROCS cannot invoke other PROCS. UGEN overcame these limitations by providing a method for the UTPS user to enter familar UTPS Job Control Language (JCL) and parameters and have the required JCL and parameters for the record linkage programs generated. UGEN input may consist of several jobs where each job may contain several steps. The EXEC cards may contain references to programs, catalogued procedures, or MACRO PROCS (any procedure recognized by UGEN) that can be interpreted by UGEN. UGEN outputs a file that is read by the operating system and executed as a sequence of one or more batch jobs.
2. USORT--USORT is a MACRO PROC designed to prepare a single sort JOB step for execution. USORT calculates required work space automatically. This PROC requires no SYSIN file for the sort parameters because they can be given as EXEC keywords. The USORT PROC is used to sort the PICADAD file in preparation for input into the UGEO PROC.
3. UDIME--The UDIME MACRO PROC prepares GBF/DIME files for input to the UGEO MACRO PROC. The original GBF/DIME files are created by the CUE (correction, updating, and extension) program from the census. These files must be initialized by the ZIPSTAN and GBFSPLIT programs and then sorted. UDIME accomplishes these job steps with a minimum of user specifications. UDIME outputs a permanent file that becomes the GBF input file for the UGEO program. UDIME also provides the single most important design feature for NCUTPIS. It provides the user with the capability of defining traffic analysis zones in terms of the geography contained in the GBF/DIME file. As with the other MACRO PROCS, UDIME must be invoked with UGEN.
4. UGEO--The UGEO MACRO PROC, like other UTPS PROCS, has a compilation and an execution phase. The execution of UGEN constitutes the compilation phase. The execution phase consists of a series of nine complex steps of computer processing available through a single EXEC card and several keywords. UGEO requires two reference files and one data file as input. The reference files are PICADAD and GBF/DIME for the particular urban center for which data are being geocoded. PICADAD is used to obtain place codes for each city name. UGEO outputs a file that contains the original data records, each of which are appended with a matchkey composed of standardized address fields and any geocodes that the analyst has requested.

In summary, UGEN was developed to simplify the use of computerized address-matching systems. The computer processes in the system are inherently complex. System users required automation of these processes in order to prevent a labor-intensive and error-prone operational task. Hence, UGEN was designed as the interface between the user and the computer operating system (OS). UGEN prevents much of the drudgery, complexity, and redundancy involved in user specification of sophisticated computer pro-

Figure 3. Flowchart for address matching and geocoding process within UTPS environment.



cesses. The UTPS user prepares a card deck with the step that invokes UGEN via a catalogued JCL procedure. The data for this step include the user's whole job. UGEN reads the user's JCL and UTPS parameters, generates a JCL job stream, and later submits this job to the computer OS if no errors are detected (3).

Generalized Utility Procedures

Several utility programs were developed to support the UTPS computer programs discussed. The general purpose of these utility programs is to prepare the source data for input into NCUTPIS and to summarize and report on the geocoded output data. One special purpose program provides the user with a generalized curve-fitting procedure and is capable of preparing a computer plot of a user-specified income model.

Currently, four utility programs support the UTPS geocoding programs:

1. SCREEN--SCREEN reduces a source data file that contains addresses for several cities to a data file that contains addresses for a single city. This file can be geocoded much faster than can the unreduced file.
2. REFORMAT--REFORMAT reformats records from the data file that may not be suitable for input to the UTPS geocoding programs. For example, the street name field may precede the street number field. The UTPS geocoding programs require the switching of these fields by the REFORMAT program.
3. SUMPORT--SUMPORT provides for the summary, reporting, manipulation, and analysis of the data that are located in the geocoded records output from the geocoding programs.
4. HARANAL--HARANAL provides a generalized curve-fitting capability that is based on Fourier series and harmonic analysis. The output includes a harmonic equation of the fit and a printer plot of the calculated and observed tabular values. This procedure can prepare a printer plot of a user-specified income model.

The sequencing of both the utility and UTPS programs is shown in Figure 3. This sequencing is intended to manifest the essential aspects of the data acquisition tool (NCUTPIS) that was designed during this study to support the travel demand models used in UTPP.

Income Models

Development of a mathematical model to estimate income became necessary because of confidentiality requirements established by law. Federal and state revenue agencies are prohibited by law from releasing their income tax files to other agencies, even for planning purposes. If not for this constraint, an income file would have existed from which zonal incomes could have been extracted directly. Fortunately, the confidentiality constraint can be overcome by use of the income-estimating model.

Surveillance Subsystem

The fourth system component was designed to collect zonal summaries of urban transportation planning data on the two functional categories of demography and employment. The structure of each subsystem is shown in Figures 4 and 5.

Demographic Subsystem

The demograhic summary procedures will produce 13 items of demographic data annually for each traffic analysis zone in the study area.

The demographic subsystem structure requires five of the source data files (shown in Figure 4) as subsystem input.

Employment Subsystem

The employment subsystem is designed to produce zonal summaries of employment broken down by SIC codes. The input to this subsystem was the North Carolina State Employment Security Commission master file, which includes the number of employees for all establishments except small businesses and government agencies. The employment subsystem structure is shown in Figure 5.

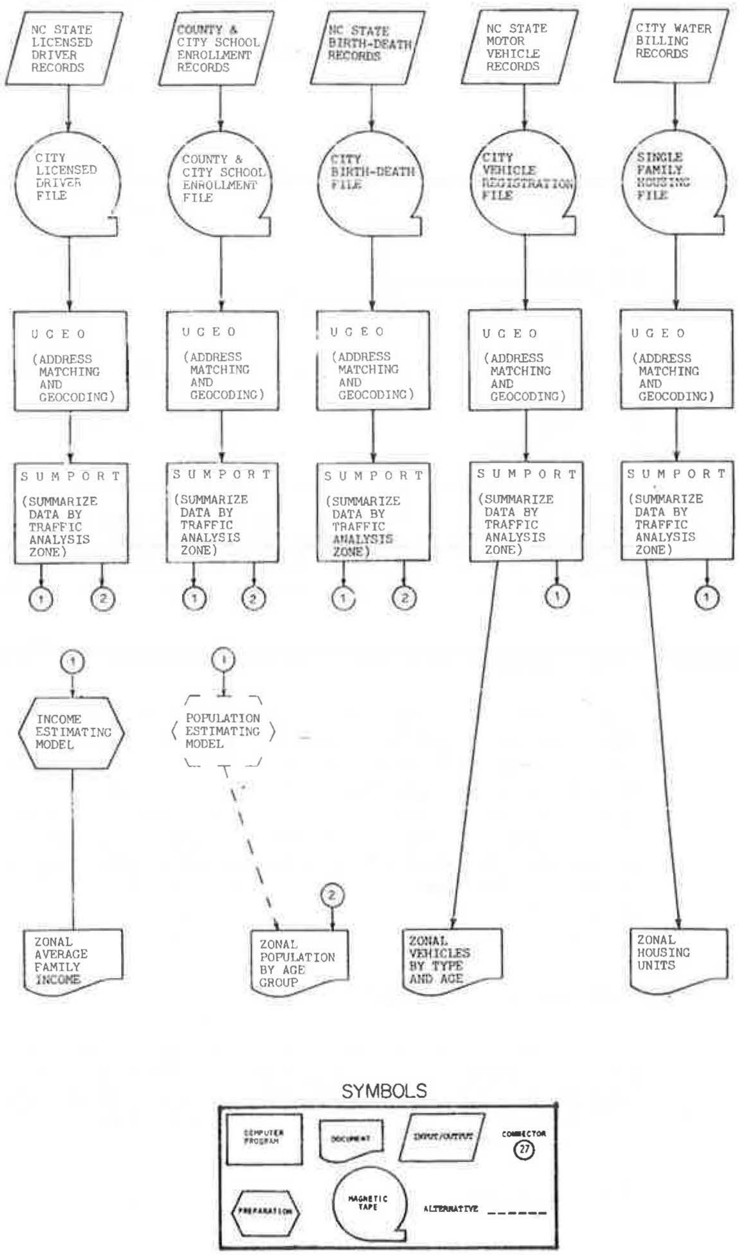
IMPLEMENTING AND TESTING NCUTPIS

The North Carolina urban center chosen for initial implementation of NCUTPIS was Greensboro. For the purpose of analysis the Greensboro transportation planning study area was divided into 302 traffic zones. Testing of NCUTPIS involved nothing more than determining if the system could provide summaries of population, single-family housing units, vehicle ownership, average family income, and employment for each of the 302 traffic zones in the Greensboro transportation planning area.

Demographic Subsystem

The demographic subsystem was initially tested by using the vehicle registration file for Guilford County as source data. The total number of vehicle records in that file was more than 200,000. The program SCREEN was used to reduce this to 122,727 records that contain only Greensboro addresses. This source data file along with the PICADAD and geographic base (GBF) files for Greensboro were

Figure 4. Demographic surveillance files.



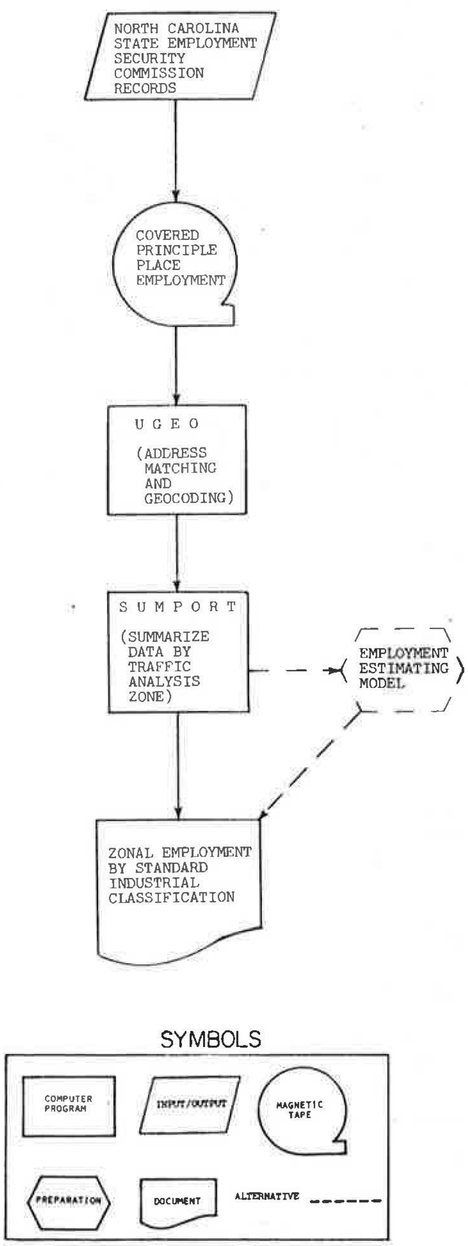
input to the UGEO procedure. The output results from this procedure are shown in Figures 6 and 7.

The entire demographic subsystem of NCUTPIS was tested in a similar manner. Each new source data file was input to the UGEO procedure. The output from this procedure was then summarized over each traffic zone. The first 25 zonal summaries for the demographic subsystem are given in Table 2. Average family income for these traffic zones (as estimated from the Greensboro income model) are also included in this table.

The following findings are suggested relative to implementation of the computerized data retrieval system (NCUTPIS) and are based on the data sets and analytical procedures used during this investigation.

1. The match rate for the vehicle registration file was 80.5 percent. The matched records total included 89,844 matched with certainty and 8,973 matched with uncertainty.

Figure 5. Employment surveillance files.



2. The match rate for Greensboro's water billing file for dwelling units was 93.2 percent. The matched records total included 51,421 matched with certainty and 1,821 matched with uncertainty.

3. The match rate for licensed drivers in the Greensboro area was 77.2 percent. The matched records total included 96,916 matched with certainty and 15,950 matched with uncertainty.

4. The match rate for county students living in Greensboro could not be determined because of the many non-Greensboro addresses contained in the file. The file contained a total of 28,031 student records of which 5,090 were matched to Greensboro addresses with certainty and 4,570 were matched with uncertainty.

5. The match rate for Greensboro city students was 93.8 percent. The file contained a total of 22,835 records of which 20,241 were matched with certainty and 1,170 were matched with uncertainty.

6. The match rate for births in the Greensboro area was 65.9 percent. This birth file contained 10,325 records from 1974 through 1979 of which 6,233 were matched with certainty.

7. Harmonic analysis provided a convenient mathematical framework for the construction of the income model.

8. Estimates of zonal income averages are easily obtained from the income model.

9. NCUTPIS is completely user-oriented relative to the UTPS environment.

Employment Surveillance Subsystem

The employment subsystem was tested in the same way as the demographic subsystem. Greensboro employer records as maintained by the North Carolina State Employment Security Commission served as the input data file. A match rate of only 49.2 percent was

Figure 6. Vehicle registration address matching run statistics matched with certainty.

```
                    RUN STATISTICS
122,727 INPUT RECORDS READ--(DATAIN)
      0 INPUT REJECT RECORDS READ--(REJTIN)
122,727 INPUT RECORDS SELECTED FOR PROCESSING
 23,689 REFERENCE RECORDS READ  (REFERIN)
  5,559 UNIQUE REFERENCE FILE KEYS
      0 TEMPORARY OUTPUT BLOCKS WRITTEN-(WKFILE)
      0 TEMPORARY INPUT BLOCKS READ--(WKFILE)
     20 REFERENCE CANDIDATES IN MAXIMUM DOMAIN
      3 REFERENCE CANDIDATES IN AVERAGE DOMAIN
      2 REFERENCE CANDIDATES FOR AVG DATA RECORD
 89,844 OUTPUT RECORDS WRITTEN--(MTCHCUT)
 32,883 OUTPUT RECORDS WRITTEN--(REJTCUT)
  3,897 CRITICAL FIELD REJECTS  (CR)
 28,014 SEARCH FIELD REJECTS    (SR)
    972 GROUP REJECTS           (GR)
      0 SELECT REJECTS          (SE)
      0 DATA FILE SEQUENCE CHECKS
 89,844 RECORDS MATCHED         (MT)
      0 EXACT MATCHES
                    MATCH RATE =   73.20 PER CENT
```

obtained during the geocoding process. The primary reason for this low match rate was the large number of postal addresses found in the record address field. Postal addresses cannot be processed by the present version of NCUTPIS's matching procedure. Until this address-matching problem is solved, NCUTPIS cannot be used in extracting zonal employment data. The only other method that can be used to obtain employment data is the manual inventory or survey.

Figure 7. Vehicle registration address matching run statistics matched with uncertainty.

```
                  RUN STATISTICS
32,883 INPUT RECORDS READ--(DATAIN)
     0 INPUT REJECT RECORDS READ--(REJTIN)
-----------------------------------------------------
32,883 INPUT RECORDS SELECTED FOR PROCESSING
23,689 REFERENCE RECORDS READ--(REFERIN)
 2,301 UNIQUE REFERENCE FILE KEYS
-----------------------------------------------------
     0 TEMPORARY OUTPUT BLOCKS WRITTEN-(WKFILE)
     0 TEMPORARY INPUT BLOCKS READ--(WKFILE)
    24 REFERENCE CANDIDATES IN MAXIMUM DOMAIN
-----------------------------------------------------
     5 REFERENCE CANDIDATES IN AVERAGE DOMAIN
     1 REFERENCE CANDIDATES FOR AVG-DATA RECORD
32,883 OUTPUT RECORDS WRITTEN--(MTCHOUT)
-----------------------------------------------------
     0 OUTPUT RECORDS WRITTEN--(REJTOUT)
 3,897 CRITICAL FIELD REJECTS    (CR)
18,356 SEARCH FIELD REJECTS      (SR)
-----------------------------------------------------
 1,657 GROUP REJECTS             (GR)
     0 SELECT REJECTS            (SE)
     0 DATA FILE SEQUENCE CHECKS
-----------------------------------------------------
 8,973 RECORDS MATCHED           (MT)
     0 EXACT MATCHES
                       MATCH RATE =   27.28 PER CENT
```

SUMMARY

The main function of transportation planning is to predict future traffic flows and patterns so that a system plan can be developed that will provide an acceptable level of service relative to the predictions. Tremendous advances have been made in the art of travel forecasting during the last 30 years. The cost of performing travel forecasting has also been tremendous. A major part of this cost involves the collection of socioeconomic data that are the determinates of existing traffic patterns in the urban area.

The objective of this study was the development of a computerized data retrieval system that would obviate the need for costly and time-consuming surveys. The development of such a data system required the construction and calibration of an income-estimating model.

The analysis of system implementation suggests the following conclusions:

1. The implementation of NCUTPIS shows it to be capable of collecting the demographic and some of the employment data used in the continuing UTPP. NCUTPIS, therefore, accomplishes the objectives that were set forth for this research. No attempt was made to provide a system design that could serve all the data requirements for all types of urban planning. The data system, however, can serve as a basis for any future expansion that may be desirable.
2. Conclusions as to how well NCUTPIS performed its function of data acquisition tool could not be drawn because no equivalent data base exists to serve as a bench mark for measuring its performance. What can be concluded is that NCUTPIS provides an operationally feasible method for maintaining, on a small area basis, an urban transportation planning data base by using data extracted from the routine operating and administrative records of state, county, and local government agencies.

Table 2. Demographic subsystem output.

Greensboro Traffic Zone No.	Total Vehicle Registration	Total Single Family Dwelling Units	Licensed Drivers	County Students	City Students	Births from 1974 to 1979	Avg Family Income ($)
1	340	421	130	63	1	10	4,800
2	201	80	17	4	0	0	22,500
3	35	60	20	1	0	0	3,600
4	80	150	101	83	9	6	3,500
5	188	104	85	39	8	10	16,000
6	173	154	68	24	2	2	6,400
7	107	148	68	5	7	2	4,500
8	204	155	268	2	169	22	7,500
9	502	352	706	62	29	32	8,100
10	62	15	73	4	3	6	22,500
11	661	447	793	171	94	50	8,500
12	231	81	105	9	11	9	22,500
13	494	366	521	2	54	24	7,500
14	502	332	595	11	40	27	8,900
15	898	723	1,050	42	142	103	6,800
16	681	416	482	13	209	32	10,500
17	60	109	94	0	24	10	3,550
18	42	94	31	0	0	0	3,500
19	14	78	16	0	8	0	3,500
20	1,377	998	1,227	63	411	73	7,800
21	277	311	471	4	484	35	6,100
22	330	241	443	13	275	16	8,900
23	1,122	867	1,273	16	317	128	7,400
24	112	77	66	0	18	12	8,300
25	168	227	222	37	14	4	4,600

3. Implementation of NCUTPIS demonstrates the importance of intergovernmental and interagency working agreements. The overall effectiveness of this data system depends to a great extent on these agreements.

This study should provide useful guidelines for implementation of similar information systems in other urban centers. Note that the necessity for developing an income model was revealed in this study. Income data were not previously available on a continuing basis from any other source. If income is to continue to be the primary predictive variable used in the simulation of urban travel demand, then additional research must be performed to substantiate the findings of this investigation. If subsequent substantiation cannot be furnished by research, then a suitable substitute for the income variable should be found. This substitute must be easily monitored and collectible on a small area basis. One possible substitute that has these characteristics is family life-cycle.

REFERENCES

1. Peat, Marwick, Mitchell and Co. Conceptual Design for an Urban Transportation Planning Information System. Office of Highways Planning Division, FHWA, Jan. 1973.
2. GBF/DIME System Description and Uses. Geography Division, Bureau of the Census, U.S. Department of Commerce, Feb. 1978.
3. M.A. Jaro, J.R. Baker, and R.M. Somers. Introduction to Census/UMTA Technology. Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Nov. 1978.
4. R.B. Dial. The Urban Transportation Planning System (UPTS): Philosophy and Function. Office of Transit Planning, UMTA, June 1975.
5. R.E. Gaskell. Engineering Mathematics. Dryden Press, New York, 1958.
6. L.W. McPherson. Toward the Development of an Urban Transportation Planning Information System. North Carolina State Univ., Raleigh, PhD Dissertation, 1981.
7. Urban Planning Division, Federal Highway Administration. Trip Generation Analysis. U.S. Government Printing Office, Aug. 1975.

Publication of this paper sponsored by Committee on Transportation Information Systems and Data Requirements.