1134

TRANSPORTATION RESEARCH RECORD

Transportation Data

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1987

Transportation Research Record 1134
Price: $10.50
Typesetter: Lucinda Reeder

modes
1 highway transportation
2 public transit

subject area
12 planning

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Transportation data.

p.cm -- (Transportation research record, ISSN 0361–1981 ; 1134)
ISBN 0-309-04518-5
1. Traffic engineering--Congresses. 2. Traffic surveys--Congresses. 3. Roads--Congress. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 1134
[HE332]
380.5 s--dc19
[388] 88-12444
CIP

Transportation Research Record 1134

The **Transportation Research Record** series consists of collections of papers on a given subject. Most of the papers in a Transportation Research Record were originally prepared for presentation at a TRB Annual Meeting. All papers (both Annual Meeting papers and those submitted solely for publication) have been reviewed and accepted for publication by TRB's peer review process according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Research Council, or the sponsors of TRB activities.

Transportation Research Records are issued irregularly; approximately 50 are released each year. Each is classified according to the modes and subject areas dealt with in the individual papers it contains. TRB publications are available on direct order from TRB, or they may be obtained on a regular basis through organizational or individual affiliation with TRB. Affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Contents

Foreword

In their paper entitled The 1984 Home Interview Survey in the Dallas-Forth Worth Area: Changes in Travel Patterns, 1964–1984 Norris and Shunk identify some important changes in trip generation rates. Due to demographic and labor force changes in the Dallas-Forth Worth area during the 1970's and 1980's, the number and share of home-based nonwork person-trips per household declined. The shares and numbers of home-based work and nonhome-based trips increased during this period. The reduction in home-based trip rates has important implications for urban development and transportation planning.

In his paper entitled Priority-Setting Procedures and Scarce Data: The Synthetic Solution, Fricker presents a case study to illustrate the actual implementation of the synthetic method whereby missing data can be synthesized to permit implementation of priority setting or a pavement management system. The author's method is applicable to roads, bridge projects, or any ranking procedure that involves multiple criteria and incomplete data.

In their paper entitled A Database Management System for Traffic Engineering Departments of Small and Medium-Sized Cities, Baass and Allard describe a database management system based on dBASE III PLUS that includes an integrated urban data bank and three types of software for information systems management, information retrieval, and interfaces with applications programs. A sector of Quebec, Canada, was used to test the feasibility of the system.

In his paper entitled Adequacy of the Sample Size and Accuracy of the Highway Performance Monitoring System for Use in Texas, Memmott covers some of the characteristics of the Highway Performance Monitoring System (HPMS) sample in Texas and the validity of the recommended method in selecting the sample size for each state. A recommended method to correct the deficiencies in the sample size technique is also presented along with some comparisons of HPMS output with data from other sources.

In their paper entitled Direct Data Entry Using Microcomputers—A Travel Survey Pilot Project, Steuart et al. discuss a pilot project in the Toronto region origin-destination travel survey using direct data entry and microcomputers. Interviewers using a computer program were compared to interviewers using conventional pencil and paper forms. Comparison of the two methods proved the automated process just as expedient as the manual method during actual telephoning, but far superior during sampling processing.

In his paper entitled Management of a Small Home Interview Travel Survey Using a Microcomputer, Eash discusses how a microcomputer and general purpose business software were used in a small telephone survey of household travel behavior. Principal applications were selection of sample households, mailings to sample households, scheduling interviews, checking the status of each interview, developing the completed interview data sets, and reporting of survey results.

In his paper entitled A Quick Cluster Control Method: Permanent Control Station Cluster Analysis in Average Daily Traffic Calculations, Albright discusses a statistical procedure to improve the accuracy of average daily traffic (ADT) counts. While previous procedures used a nearby permanent control counter to adjust short-term counts to ADT, the new procedure used cluster analysis of permanent control counters. Clustering is based upon functional classification of roadways. Benefit of cluster analysis has typically been a 3 to 4 percent per year improvement in ADT accuracy, resulting in 15 to 20 percent accuracy between traffic count years.

The 1984 Home Interview Survey in the Dallas-Fort Worth Area: Changes in Travel Patterns, 1964–1984

BAHAR B. NORRIS AND GORDON A. SHUNK

A recent travel survey in the Dallas-Fort Worth metropolitan area identified some important changes in trip generation rates. Due to the demographic and labor force changes in the Dallas-Fort Worth area during the 1970s and 1980s, the number and share of home-based nonwork person-trips per household declined. The shares and numbers of home-based work and non-home-based trips have increased during this period. The reduction in home-based trip rates has important implications for urban development and transportation planning. According to the 1984 Dallas-Fort Worth travel survey, the average household-trip rate was 8.68 trips/day, which was fairly stable since 1964. Person-trips per person and vehicle-trips per person, however, increased since 1964, reflecting the smaller household size and automobile occupancy rates of recent periods. The results of the 1984 travel survey also indicated that average trip length in the metropolitan area was about seven mi, average trip duration was 17–19 min, automobile occupancy rate was 1.13 for work trips and 1.5 for nonwork trips, transit mode share was 1.7 percent, and peak-hour travel times were 7–8 a.m. and 5–6 p.m.

Some schools of transportation planning and travel forecasting contend that travel behavior is stable over time and that travel forecasting parameters can be held constant without need for repeated surveys *(1)*. Though this assumption may prove true in stable areas, the assumption must be validated in a region experiencing major change. A new travel survey was conducted in the Dallas-Fort Worth area in 1984 because of anticipated major demographic changes involving household size and work habits. It is the premise of this paper that even though overall person-trip rates may be reasonably stable over time, changes in the composition of person-trips in the last two decades have been significant enough to warrant updating the existing travel forecasting models.

The purposes of this paper are, first, to shed light on the recent changes in trip patterns due to a relative increase in home-based work (HBW) trips and an absolute decline in the average number of home-based nonwork (HNW) trips per household; second, to provide summary information on the findings of the 1984 Dallas/Fort Worth travel survey; third, to discuss some methodologically relevant issues that could be helpful for future surveys; and fourth, to discuss the implication of the changes in household travel patterns for the model refinement efforts of a transportation planning agency.

North Central Texas Council of Governments, Transportation and Energy Department, P.O. Box 5888, Arlington, Tex. 76005–5888.

In the next section an overview of the changes in the frequency and composition of person-trips due to the shift away from HNW trips is offered. The Dallas-Fort Worth area travel survey results are compared with recent survey results from San Francisco, Denver, Houston, and Atlanta.

In the following section, the salient features of the North Central Texas Council of Governments (NCTCOG) survey are reviewed and 20-year longitudinal comparisons for such travel characteristics as peak-hour travel time, travel duration, trip length, automobile occupancy rates, and travel mode are provided. When available, comparable data for other metropolitan areas are also provided.

In the next section, certain methodological issues that have implications for trip frequencies and composition are discussed, for example, the implications of the choice of the survey time period for frequency of school trips and the impact of trip linking on the share of each trip purpose.

In the final section, the survey results are applied to model validation and calibration.

THE 1984 NCTCOG HOME INTERVIEW SURVEY: WHAT TRENDS CAN BE DETECTED ABOUT CHANGING TRAVEL PATTERNS IN THE POST-1980 PERIOD?

A frame of reference for the 1984 travel survey is created by looking at the changes in the Dallas-Fort Worth area travel patterns since 1964 and finding commonalities with other metropolitan areas. Implications of the changes for model refinement efforts are then examined.

The Survey

Between April 23 and July 13, 1984, interviews of 2,471 households were conducted at their places of residence, and demographic and travel information for the household members 5 years old and over were recorded. Some 20,200 person-trip records were created for all the vehicular trips made by 6,403 persons in the sample households. The only exceptions to vehicular trips were walk-to-work trips. Weighted by the total number of households in the NCTCOG transportation study area, the sample households represent over 1 million households and approximately 10 million trips in 1984 (Table 1).

TABLE 1 PERSON-TRIPS BY PURPOSE—1964 AND 1984 TRAVEL SURVEYS *(2, 3)*

WEIGHTED SAMPLES

	1984	1964	
	Person Trips	Person Trips	Percent Change
HBW	2,600,433	1,026,338	153.37
Column %	26.99	20.90	
HNW	4,601,088	2,915,587	57.81
Column %	47.75	59.40	
NHB	2,433,963	964,735	152.29
Column %	25.26	19.70	
Total	9,837,655	4,906,660	100.5

* Person trips include all in-vehicle trips (driver, auto and carpool passenger) and transit trips. They exclude taxi, school bus, walk, bicycle, and "other" modes. This information does not reflect "external" related travel.

Changes Since 1964

The last time a comprehensive travel survey was conducted in the region was 1964. The data in Table 1 show that in the 20-year period the total number of trips doubled from 4.9 to 9.8 million, an annual increase of 5 percent. The greatest change occurred in HBW and NHB trips—each showing an increase of over 150 percent in the 20-year period. Home nonwork trips, which include trips that originated in homes for shopping, school, and personal business purposes, showed a relative decline, as indicated by a reduction in their share of total trips from 59 to 48 percent of all person-trips. Home-based work and non-home-based (NHB) trips increased their combined share from 41 to over 52 percent of all person-trips (Table 1).

Paralleling the shift in the composition of trips is a change in the frequency of person-trips. On the one hand, there has been a slight decline in the average number of person-trips per household from 9.12 in 1964 to 8.68 in 1984 (Table 2), and on the other hand, an increase in the number of person-trips per person from 2.73 in 1964 to 3.40 in 1984 (Table 3). Similar shifts occurred in other metropolitan areas.

The shift in the relative frequency of the three trip purposes in the past two decades is partially explained by the changes in the demographic profile of the region. Increased labor force participation by 1980 of women from about 40 percent to over 58 percent of the population aged 16 years and over is one of the reasons for the increased share of HBW trips. The 1980 census of the Dallas-Fort Worth SMSA shows that an average of 1.44 persons per household were employed in 1980, compared to 0.92 persons in 1960. This increase in the number of working people per household has resulted in more than a twofold increase in the work force in the 20-year period at an average growth rate of 6.1 percent annually, whereas the population aged 16 years and over has less than doubled at 4.45 percent annually.

Another demographic transformation in the region that has contributed to shifts in trip patterns is the reduced household size. In 1964, the average household size in the region was 3.22, whereas in 1984 it declined by 21 percent to an average of 2.55 persons. The combined effects of a higher labor force participation rate and smaller household size produced fewer HNW trips as household members reduced the frequency of departures from home by consolidating their trips. The extent of this shift in the frequency of HNW trips is reflected in the travel surveys of other metropolitan areas.

Commonalities with Other Metropolitan Areas

The trend shown in the 1984 NCTCOG survey towards a reduced share of HNW trips is supported by other travel surveys in the 1980s. The San Francisco Bay Area Travel Survey of 1981, the Denver Regional Council of Governments 1985 Travel Survey, and the Atlanta Area Survey of 1980 all show declining absolute numbers of HNW trips per household (Table 2). In all these surveys, the decline in the frequency of HNW trips has contributed to a slight decline in the overall number of trips per household. Furthermore, the parallel decline in the relative importance of HNW trips in these areas is shown by the declining share of HNW trips as a percentage of all trips, from a high of 59 percent in 1964—in the case of the Dallas-Fort Worth area—to a low of 47 percent in 1984 (Table 4).

TABLE 2 PERSON-TRIPS PER HOUSEHOLD—COMPARISON OF THE 1984 NCTCOG SURVEY WITH OTHER SOURCES *(2–7)*

	HBW	HNW	NHB	ALL TRIPS
Dallas/Fort Worth - NCTCOG Survey,* 1984 (2)	2.29	4.32	2.07	8.68
SDHPT Survey, 1964 (3)	1.91	5.42	1.79	9.12
% Annual Change, 1964-1984	0.99%	-1.01%	0.78%	-0.25%
San Francisco Bay Area Survey,* 1981 (4)	1.89	4.49	2.35	8.71
1965	1.86	5.01	1.91	8.78
% Annual Change, 1965-1981	0.001%	-0.63%	1.41%	-0.001%
Denver Regional COG Survey,* 1985 (5)	1.96	3.40	1.97	7.33
1971	2.03	4.31	1.75	8.09
% Annual Change, 1971-1985	-0.25%	-1.51%	0.90%	-0.67%
Houston-Galveston Area Survey, 1984 (6)	2.11	3.89	2.92	8.92
Atlanta Area Survey, 1980 (7)	1.95	4.45	1.87	8.27
1972	2.03	4.80	1.65	8.48
% Annual Change, 1972-1980	-0.33%	-0.61%	1.11%	-0.21%

* Based on Linked Trips

TABLE 3 DISTRIBUTION OF TRIP PURPOSES—COMPARISON OF THE 1984 NCTCOG SURVEY WITH OTHER SOURCES *(2–7)*

	HBW	HNW	NHB	ALL TRIPS
Dallas/Fort Worth - NCTCOG Survey, 1984 (2)	0.90	1.69	0.81	3.40
- SDHPT Survey, 1964 (3)	0.57	1.62	0.54	2.73
% Annual Change, 1964-1984	2.90%	0.20%	2.50%	1.23%
San Francisco Bay Area, 1981 (4)	0.74	1.75	0.91	3.40
- Survey, 1965	0.60	1.61	0.61	2.82
% Annual Change, 1965-1981	1.46%	0.56%	3.06%	1.31%
Denver Regional COG Survey, 1985 (5)	0.77	1.33	0.77	2.87
- Survey, 1971	0.70	1.60	0.60	2.90
% Annual Change, 1971-1985	0.71%	-1.21%	2.02%	-0.07%

Implications of the Survey Results for Model Refinement

There are modeling implications in the shift in the composition of person-trips away from HNW trips. The pronounced downward shift in the share of HNW trips has been large enough that the compensating gains in HBW and NHB trips have not succeeded in offsetting the HNW loss. The fact that this pattern has been displayed in all four metropolitan areas for which comparable data are available suggests that the trip generation implications of the shift should not be ignored. It is also important to note that an income variable in trip generation cannot totally capture this change in travel over time, due to the counterintuitive decline in HNW travel with increases in real income.

A convincing case, therefore, can be made for the need for new travel data in the 1980s for planning agencies that calibrated their travel models with the 1960s survey results. The consistency of the trend among the metropolitan areas examined suggests that (a) overall trip rates are more stable than the disaggregated rates by purpose, (b) for a given survey year, rates are more consistent across regions than over time, (c) the slight decline in the aggregate trips per household is the result of the shrinking household size rather than reduced travel, and (d) due to declining household size and automobile occupancy rates, person-trips per person and vehicle-trips per person have increased considerably since the 1960s.

TABLE 4 PERSON-TRIPS PER PERSON—COMPARISON OF THE 1984 NCTCOG SURVEY WITH OTHER SOURCES *(2–5)*

	HBW	HNW	NHB
Dallas/Fort Worth - NCTCOG Survey, 1984 (2)	26.99	47.75	25.26
- SDHPT Survey, 1964 (3)	20.90	59.40	19.70
% Annual Change, 1964-1984	1.46%	-0.98%	1.41%
San Francisco Bay Area Survey, 1981 (4)	23.64	49.68	26.67
1965	24.25	54.00	21.75
% Annual Change, 1965-1981	-0.16%	-0.50%	1.41%
Denver Regional COG Survey, 1985 (5)	26.0	47.0	27.0
1971	23.0	57.0	20.0
% Annual Change, 1971-1985	0.93%	-1.25%	2.50%
Atlanta Area Survey, 1980 (7)	23.58	53.84	22.58
1972	23.99	56.56	19.45
% Annual Change, 1972-1980	-0.21%	-0.60%	2.01%

TABLE 5 AVERAGE AUTOMOBILE OCCUPANCY RATES BY PURPOSE—WEIGHTED 1984 SAMPLE *(2, 3)*

Purpose**	Average Auto Occupany		Percent
	1984	1964	Average 1964-84
HBW1	1.07	1.36	-21.32
HBW2	1.15	1.24	- 7.26
HBW3	1.14	1.16	- 1.72
HBW4	1.13	1.12	+ 0.01
HNW	1.53	1.70	-10.00
NHB	1.36	1.46	- 6.85
TOTAL	1.36	1.52	-10.53

* Weighted average of vehicle occupancies for self-driven auto trips. For occupancy rates of 9 plus, the average is assumed to be 10.

** HBW1, HBW2, HBW3, and HBW4 refer to HBW trips made by individuals with annual household incomes falling in the first quartile (less than $15,000), second quartile (between $15,000 and $29,999), third quartil (between $30,000 and $39,999), and fourth quartile (over $40,000), respectively.

CHANGES IN THE TRAVEL PATTERNS IN THE DALLAS-FORT WORTH AREA: 1964–1984

In this section a profile of travel behavior in the NCTCOG region is created that shows how frequently, at what time of the day and how far people travel, what the automobile occupancy rates and mode choices are, and how all these characteristics have changed since 1974.

Who is Most Likely to Travel: The Influence of Household Size, Automobile Availability, and Income on Trip Rates

Larger households and those with a higher income or more cars are more likely to make trips. Households of six or more members make an average of 15 trips/day, whereas single-member households make an average of 3.25 trips/day. Trips per person, however, decline as household size increases, reflecting the economies of scale in trip making that accrue to larger households. This trend is illustrated by the decline from 3.25 trips/person for single-member households to less than 2.4 trips/person for households of six or more members. The decline in the average household size since 1964—from 3.22 to 2.55 persons in 1984—has consequently led to an increase in person-trips per person. Table 4 shows the changes in person-trips per person in the Dallas-Fort Worth area, San Francisco, and Denver. Person-trips per person increased for all trip purposes in the Dallas-Fort Worth region and the San Francisco area, but decreased for HNW trips in Denver.

The increased availability of automobiles to households has also led to the increased number of person trips. The average

TABLE 6 VEHICLE-TRIPS PER HOUSEHOLD AND PER PERSON—1964 AND 1984 NCTCOG HOME INTERVIEW SURVEYS *(2, 3)*

	HBW	HNW	NHB	ALL TRIPS
Vehicle Trips Per Household, 1984	2.03	2.82	1.52	6.37
1964	1.62	3.19	1.23	6.04
% Change, 1964-1984	25.30	-11.60	23.58	5.46
Vehicle Trips Per Person, 1984	0.80	1.10	0.60	2.50
1964	0.48	0.95	0.37	1.80
% Change, 1964-1984	66.67	15.80	62.16	38.89

TABLE 7 PERCENTAGE DISTRIBUTION OF TRIP START TIMES—1984 HOME INTERVIEW SURVEY *(2)*

A.M. TRIPS												
TIME	12-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
HBW	0.58	0.22	0.21	0.28	0.59	2.40	10.66	19.27	9.64	2.70	1.49	1.67
HNW	0.46	0.21	0.09	0.00	0.03	0.28	0.93	5.70	7.29	3.71	4.99	4.94
NHB	0.15	0.15	0.10	0.02	0.02	0.15	0.42	1.48	3.04	5.16	7.31	10.50
TOTAL	0.41	0.20	0.13	0.08	0.17	0.81	3.38	8.22	6.83	3.81	4.66	5.49
P.M. TRIPS												
TIME	12-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12
HBW	2.45	2.04	2.80	6.08	11.94	12.94	4.94	2.00	1.33	1.69	1.24	0.82
HNW	4.06	4.36	5.30	8.10	7.59	8.36	10.17	8.62	5.77	5.82	2.28	0.97
NHB	14.68	9.88	9.10	8.98	7.43	7.56	3.88	4.47	2.55	1.45	1.29	0.20
TOTAL	6.35	5.15	5.61	7.79	8.70	9.37	7.18	5.80	3.77	3.61	1.75	0.73

number of automobiles per household in 1984 was 1.84 compared to 1.33 in 1964, an increase of 3.8 percent in the 20-year period. The number of automobiles per person had a steeper increase of 76 percent from 0.41 automobile/person in 1964 to 0.72 automobile/person in 1984. The greater availability of automobiles has boosted trip rates, both directly by reducing the number of zero-car households, and indirectly by reducing automobile occupancy rates. Table 5 shows that since 1964 the average automobile occupancy rate has declined by 10.53 percent from 1.52 to 1.36 persons/automobile trip. The growth in vehicle-trip rates per household and per person shown in Table 6 reflects the combined effects of lower household size and automobile occupancy rates.

Higher-income people are also more likely to make trips. Households in income Quartiles 3 and 4 are at least twice as likely to travel as those in Quartile 1, a relationship that is true regardless of the fact that households in higher-income quartiles are larger on the average.

How Far, How Long, and at What Time Are People Most Likely to Travel?

An important product of the survey was identifying the time and duration of traffic peaks. The morning peak hour was 7–8 a.m. and the afternoon peak hour was 5–6 p.m. (Table 7). The peak hour in the region varied widely by purpose. The peak hour for HBW coincided with the peaks for total trips (7–8 a.m. and 5–6 p.m.) and the peaks for HNW trips occurred later (8–9 a.m. and 6–7 p.m.). NHB trips, on the other hand, peaked between 12 noon and 1 p.m.

In 1984, households in the Dallas-Fort Worth area had an average trip length of 6.9 mi. From Table 8, the average length of HBW trips was greater for higher-income quartiles, and HBW trips tended to be longer than HNW and NHB trips. Since 1964, the average distance traveled increased by 21.05 percent from 5.7 to 6.9 mi.

The average trip duration, as perceived and reported by the respondents in the 1984 travel survey, was 17.4 min for all

TABLE 8 AVERAGE TRIP LENGTH (MI)—1984 HOME INTERVIEW SURVEY *(2, 3)*

WEIGHTED SAMPLE

Purpose***	1984* Travel Survey (miles)	1964 Travel Survey (miles)	Percent Change 1964-84
HBW1	8.26	8.50**	---
HBW2	10.09	8.50**	---
HBW3	10.64	8.50**	---
HBW4	11.43	8.50**	---
HNW	5.30	4.80	10.42
NHB	6.50	5.70	14.04
Weighted Average	6.90	5.70	21.05

* 1984 trip lengths are based on origin/destination tables obtained from the Work Place Survey.

** 1964 HBW average.

*** HBW1, HBW2, HBW3, and HBW4 refer to HBW trips made by individuals with annual household incomes falling in the first quartile (less than $15,000), second quartile (between $15,000 and $29,999), third quarter (between $30,000 and $39,999), and fourth quartile (over $40,000), respectively.

purposes combined (Table 9). The average duration was longest for work trips, particularly work trips made by higher-income households. There are problems of validity associated with perceived travel time as an index of actual travel duration. The NCTCOG regional travel model simulation of the 1984 travel duration, which is based on origin-destination trip tables, is methodologically more accurate and is slightly longer (by an average of 19 min), but has an intrinsic flaw in that it does not distinguish between peak and off-peak trip duration. Because most HNW and NHB trips are likely to be taken during off-peak hours, the implications of an upward bias in the model's estimate of travel duration for HNW and NHB trips should be taken into account.

How Are Trips Most Likely to be Made: Changes in Travel Mode, 1964–1984

The dominant travel mode in the NCTCOG region was self-driven automobile trips. In 1984, 78.2 percent of all person-trips were made by automobile drivers, 20.1 percent by automobile passengers, and 1.7 percent by transit riders. In 1964, these proportions were 63.8 percent for automobile drivers, 33.0 percent for automobile passengers, and 3.2 percent for transit passengers (Table 10). The increased proportion of automobile driver trips and the declines in automobile passenger and transit trips were consistent for all trip purposes.

IMPACT OF SURVEY METHODOLOGIES ON TRIP FREQUENCIES AND COMPOSITION

In this section, a number of methodological issues that have a potential impact on the frequency and composition of trips in a survey are discussed; specifically, the impact of the choice of the survey time period on the frequency of school trips and the impact of trip linking on the composition of trips by purpose are examined.

The Choice of the Survey Time Period

The 1984 NCTCOG survey was conducted over a 4-month period extending from April 23 to July 13. The survey lasted

TABLE 9 AVERAGE TRIP DURATION AS PERCEIVED BY THE SURVEY RESPONDENTS—1984 HOME INTERVIEW SURVEY *(2)*

WEIGHTED SAMPLE*

Purpose**	Average Duration (Minutes)
HBW1	19.71
HBW2	22.66
HBW3	22.69
HBW4	24.30
HNW	15.23
NHB	12.50
All Trips (Weighted Average)	17.44

* Includes self-driven and passenger auto trips, carpool and motorcycle trips only. Excludes bus, school bus, taxi and other (truck, air and external trips)

** HBW1, HBW2, HBW3, and HBW4 refer to HBW trips made by individuals with annual household incomes falling in the first quartile (less than $15,000), second quartile (between $15,000 and $29,999), third quartile (between $30,000 and $39,999), and fourth quartile (over $40,000), respectively.

TABLE 10 PERSON-TRIPS BY MODE AND PURPOSE (as percentages of all trips)—1964 AND 1984 HOME INTERVIEW SURVEYS *(2, 3)*

WEIGHTED SAMPLE

Mode / Purpose	Auto (Self-Driven)	Auto** Passenger	Bus	Total Person Trips
HBW (1984)	88.62	7.94	3.44	100.00
(1964)	78.80	14.40	6.80	100.00
HNW (1984)	68.64	30.07	1.29	100.00
(1964)	57.30	40.10	2.60	100.00
NHB (1984)	85.22	14.02	0.76	100.00
(1964)	67.90	31.10	1.00	100.00
TOTAL (1984)	78.22	20.04	1.74	100.00
(1964)	63.90	33.00	3.20	100.00

* Person trips include all in-vehicle trips (by the driver or passenger, auto and carpool) and transit trips. They exclude taxi, school bus, walk, bicycle, and "other" modes.

** Includes carpool/vanpool passengers.

longer than was originally intended due to delays in the completion of some interviews. The closing dates for the public school districts in the region began around May 25, 1984, and ended by the first week of June. Prolonging the survey time over a period when the nature of trips and their purposes changed created problems of estimation bias. These problems had to be addressed when the survey results were analyzed and the trip generation model was calibrated.

The average number of trips per household in the period before schools closed (8.67 trips) was 28 percent higher compared to those made after schools closed (6.77 trips). The difference between the two periods was most pronounced in the HNW trips, which included school trips. By the t-test, trip means of the two periods showed significant differences, particularly for HNW trips. Future survey design should avoid a time period that would overlap with school closing dates and should estimate travel demand based on a period that would uniformly fall within school periods. This practice is particularly justified when, as in the case of the Dallas-Fort Worth area survey, the previous surveys have been conducted during the school year. The Dallas-Fort Worth survey was not intended to last beyond the end of May; future surveys should be planned with up to a 50 percent time buffer to assure no overlap.

Trip Linking: Impact on the Composition of Trips by Purpose

The NCTCOG trip rates were based on linked trips, a mechanism that was designed to generate a more accurate distribution of trips by purpose. The impact of linking procedures on the final distribution of trips was that more trips were allocated to HBW and HNW purposes with the loss of some NHB trips. Therefore, survey results based on linked trips need to be carefully examined when compared with those based on unlinked trips.

Trip linking impacts the distribution of trips by purpose by linking out a small portion of the intermediate trips made on the way to the final destination. Trips that are linked out are those made for the purposes of serving passenger, changing mode, or riding, before the principal trip purpose was realized. Of the original 20,200 trips recorded in the survey, 4.3 percent were trips for the purpose of taking a passenger to a destination on the way to the tripmaker's principal destination. Another 2.5 percent of the trips had mode-change and ride purposes associated with them before the final purpose was realized. On the whole, 6.8 percent of the trips were linked out and as a result the distribution of the final trip purposes changed. For instance, if a person left for work but on the way dropped a child at school (a serve-passenger linkage), the linking procedure would link out the latter trip so that only a home-to-work trip would remain. Similarly, a passenger with a driver dropping off another passenger (a ride linkage) or a person who is dropped off at a bus stop for a ride to the final destination (a mode-change linkage) all made trips that were linked out, and only the originally intended purpose of the trip (e.g., work) was recorded.

The impact of trip linking on the composition of the Home Interview Survey trip purposes was to increase the share of the HBW trips from 23.5 to 27.5 percent of all the trips. For HNW trips, the impact was less pronounced because the linking increased the share of these trips from 47.5 to 48.6 percent of all trips. For NHB trips, the impact was a sharp decline from 29 to 24 percent of the trips after the linking was completed.

CONCLUSIONS AND IMPLICATIONS FOR MODEL REFINEMENT

An evaluation of the NCTCOG travel survey results and comparisons with the results from other metropolitan areas indicate that though overall trip rates per household tend to remain constant over time, the changing demographics of the recent decades have altered the composition of trips. Changes in labor force participation rates and in household size in the past two decades have spearheaded a major transformation in the composition of trips, and HNW trips have declined both in absolute terms, from over five trips per household to about four, and in relative terms, from nearly 60 percent of all trips to about 47 percent. Reduced household size and automobile availability have also led to pronounced increases in person-trips per person and vehicle-trips per person.

Cross-sectional comparisons led to the important conclusion that trip rates are more likely to be stable between regions than over time. The practical implication of this consistency is that smaller planning agencies with inadequate survey funds can borrow rates from regions of similar size and transportation characteristics.

The survey results also have implications for model refinement and recalibration. For instance, new trip production rates for the NCTCOG trip generation model will be calculated, and whether the increase in NHB trips warrants disaggregating them into work-based and non-work-based trips will be investigated. Furthermore, changes in average trip length and automobile occupancy rates have implications for the vehicle-miles of travel generated by a given model. These and many more modeling issues will be examined in detail in light of the findings of the travel survey.

REFERENCES

1. K. Boyle. *Stability of Trip Generation Rates Over Time and Over Space: A Literature Review.* New York State Department of Transportation, Albany, N.Y., 1984.
2. *1984 Home Interview Survey.* North Central Texas Council of Governments, Arlington, 1986.
3. *1964 Home Interview Survey.* State Department of Highway and Public Transportation, Arlington, Tex., 1983.
4. H. P. Kollo and C. L. Purvis. Changes in Regional Travel Characteristics in the San Francisco Bay Area: 1960–1981. Washington, D.C., Aug. 1983.
5. Survey Reveals Metro Area Travel Patterns. Denver Regional Council of Governments, Denver, Colo., Dec. 1985.
6. *1984 Regional Travel Survey Executive Summary.* Houston-Galveston Area Council, Houston, Tex., May 1986.
7. Development of Trip Generation Models—Atlanta Metropolitan Area. Barton-Aschman Associates, Inc., Atlanta, Ga., Sept. 1985.

Priority-Setting Procedures and Scarce Data: The Synthetic Solution

JON D. FRICKER

Many public agencies would like to implement a systematic project priority procedure or pavement management system (PMS), but lack the data to do it. This paper presents a way to synthesize the missing data to permit implementation of priority setting or a PMS, in turn providing valuable guidance to the data collection effort. The magnitude of this effort can be minimized if the agency knows which data are of immediate importance. The use of synthetic data makes that knowledge available. A case study is presented to illustrate actual implementation of the synthetic method and to analyze the results. This method is applicable to road and bridge projects, or any ranking procedure that involves multiple criteria and incomplete data.

There is the presumption on the part of most advocates of highway maintenance priority-setting procedures and pavement management systems (PMS) that the prospective user—be it a city, county, or state—has a comprehensive, up-to-date road network database. But our experience in Indiana indicates that most counties have not made developing and maintaining such a database a high priority. In fact, a recent survey *(1)* determined that only about 15 of Indiana's 92 county road departments had access to computers of the type normally used to store such data. Recently, however, a growing number of counties have expressed an interest in systematizing the selection of road projects for major maintenance and repair. Do these counties have to wait until a complete database is assembled? For some counties, the magnitude of the data collection effort for traffic volumes alone would create interminable delays in implementing a priority-setting system or PMS. As a technology transfer (T^2) center project, the author helped develop a set of three simplified county-level, priority-setting techniques having the following principal attributes:

1. They are easy to understand and easy to use.
2. They not only can perform acceptably when a county's database is far from complete, they can be used to direct the county's data collection efforts to minimize use of resources and maximize the quality of the resulting project priorities.

A description of the three simplified methods has appeared in the literature *(2, 3)*. As a sequel to the T^2 project, a new data collection effort under way in LaPorte County, Indiana, was monitored and guided in a way that would demonstrate the second attribute listed. In this paper, methods for realizing this attribute and results to date are described.

Highway Extension and Research Project for Indiana Counties and Cities, School of Civil Engineering, Purdue University, West Lafayette, Ind. 47907.

SYNTHETIC DATA

Many factors can be used to characterize the level of need for maintenance or repair activities of a road segment. The most common factors are pavement condition (*PCR*), traffic volume (*ADT*), safety (*HAZ*), and project cost *(2)*. Among these, traffic volume is perhaps the most difficult and time-consuming factor for which to collect data that are accurate enough to use in priority-setting methods.

In LaPorte County, only 90 of 668 road segments had up-to-date traffic volume counts at the start of this study. In the next section, the steps in a procedure to generate synthetic traffic volumes that not only allow the priority-setting methods to be implemented, but also can focus a traffic volume count program on the segments most critical to the priority-setting process are listed. This method can be adapted to the collection of other factor values as well.

Generating Synthetic Volumes

The steps needed for generating synthetic traffic volumes are as follows:

1. Establish a list of homogeneous road segments (i.e., segments thought to have similar factor values and conditions along their lengths). Subsequent data acquisition may indicate that segments should be subdivided or joined.
2. Identify those segments in the road list that lack up-to-date volume counts. Because these segments must be assigned synthetic volumes, they will be referred to as "synthetic segments" until their actual traffic volumes are determined.
3. Ask one or more knowledgeable persons such as the county road supervisor and his foremen to place each synthetic road into one of three strata—low, medium, or high—based on their perception of the average daily traffic (*ADT*) of each segment. The following *ADT* values have been found useful in defining these volume strata:

$0 < \text{low } ADT < 200$; $200 \leq \text{medium } ADT < 1{,}000$; $\text{high } ADT \geq 1{,}000$

but any number of strata and their boundaries can be adopted. If the knowledgeable persons are considered to have good sense of *ADT* values on synthetic roads, more than three strata can be defined. If subsequent actual counts frequently fail to confirm their judgment, adjacent strata can be combined to provide a wider target. In this paper, a three-stratum case will be assumed.

4. Place each of the actual segments—segments for which actual up-to-data *ADT* values exist—into the appropriate stratum, as defined in Step 3. Once all actual segments are placed in their respective strata, calculate the mean *ADT* value of the actual segments in each stratum: $\bar{v}_{lo}$, $\bar{v}_{med}$, and $\bar{v}_{hi}$.

5. Move those segments not in need of major maintenance or repair to a separate routine maintenance list.

6. Assign all synthetic segments in the low-volume stratum the temporary synthetic *ADT* value of $\bar{v}_{lo}$; assign medium-volume synthetic segments an *ADT* value of $\bar{v}_{med}$ and high-volume synthetic segments an *ADT* value of $\bar{v}_{hi}$.

7. Using these and any other needed synthetic factor values, activate the priority-setting process. Even if a majority of road segments are using borrowed synthetic factor values, there will be sufficient diversity in the combination of factor values (assuming at least three factors) to avoid a large number of ties in the project ranking that results. This will be demonstrated later in the case study example.

8. For those synthetic segments that rank high in the priority list, determine the actual values of their factors as soon as practical. A high rank means the segment would be close enough to the top to be included in the county's work plan, given its budget. This condition can be determined by estimating the project cost for each segment in the priority list, beginning at the top and accumulating project costs until the budget is exceeded by, say, 10 percent.

9. If, for some reason, it is not practical to replace a high-ranking synthetic segment's synthetic values with actual values, the segment's synthetic values will nevertheless be updated as more actual counts are taken and the stratum averages are revised. Before calculating the revised stratum averages, check the new counts to determine if any synthetic segment was placed in the wrong stratum. If so, move the misplaced segment to the proper stratum, based on the boundary values established in Step 3.

10. Repeat Steps 6 to 9 until no synthetic segments appear high in the priority list, or until it is not practical to replace the synthetic factor values of high-ranking synthetic segments with actual data.

This series of steps will probably involve several iterations, so it is especially important to use priority-setting methods that are fast, economical, and flexible. The methods described in the literature *(2, 3)*, one of which is used in the case study example to follow, have these properties.

A CASE STUDY

The author had the opportunity to test his scheme in a realistic setting. LaPorte County, Indiana, which had just been the subject of a project to develop a simplified road project priority programming procedure *(3)*, was about to embark upon a large-scale traffic volume counting program. The county road supervisor agreed to periodically supply the author with the newly acquired counts and, whenever possible, acquire counts on roads suggested by the author. The rest of this section is a summary of how the synthetic volume count idea was implemented in LaPorte County, structured on the 10 steps listed in the previous section.

Step 1: Establish Homogeneous Segments

A total of 668 such segments were identified, based on a listing supplied by the Indiana Department of Highways, and entered into a datafile.

Step 2: Identify the Synthetic Segments

Of the 668 segments, only 90 had reliable, up-to-date volume counts. This meant there were 578 synthetic segments at the beginning of the study.

Step 3: Place Each Synthetic Segment Into a Stratum

Using the stratum boundaries

$0 <$ low *ADT* < 200; $200 \leq$ medium *ADT* $< 1{,}000$;
high *ADT* $\geq 1{,}000$

the county road supervisor and his staff placed 36 of the original 668 segments into the high-volume stratum, 242 into the medium-volume stratum, and 390 into the low-volume stratum.

Step 4: Calculate the Synthetic Volume for Each Stratum

Once the stratum assignments were made, the synthetic volumes

$\bar{v}_{lo} = 116.26$; $\bar{v}_{med} = 387.69$; $\bar{v}_{hi} = 4{,}027.33$

were calculated from the mean volume of the actual segments in the corresponding strata.

Step 5: Remove Segments in Good Condition From Datafile

A total of 220 of the 668 segments had *PCR* values of 4 or better and no apparent hazardous conditions, so these segments were moved to a routine maintenance datafile. Of the 448 remaining segments, 374 did not have current volume counts. Some 16 of the 90 segments with current counts were placed on the routine maintenance list, but their actual *ADT* values were still used to calculate synthetic volumes.

Step 6: Assign Synthetic Volumes to Synthetic Segments

One of the $\bar{v}$ values found in Step 4 was assigned to each of the 374 synthetic segments remaining after Step 5, based on the synthetic segment's stratum membership established in Step 3. Until this time, these 374 segments had no *ADT* values associated with them.

Step 7: Determine Project Priorities

Although all three simplified methods developed in the literature *(2, 3)* were used successfully in this step, only the percentile priority-setting technique will be shown in this paper. The top of the output file, containing the top 20 segments in the

TABLE 1 STAGE 0 PROJECT RANKS BY PERCENTILE METHOD

Rank	Segment No.	*PCR*	*ADT*	*HAZ*	Average Percentage
1	347.	1.00	672.00	1.00	97.22
2	283.	2.00	387.69	1.00	84.38
3	178.	2.00	387.69	1.00	84.38
4	209.	1.00	4,027.33	0.	77.37
5	276.	1.00	4,027.33	0.	77.37
6	385.	1.00	4,027.33	0.	77.37
7	612.	1.00	4,027.33	0.	77.37
8	155.	1.00	647.00	0.	74.80
9	154.	1.00	387.69	0.	72.75
10	20.	1.00	387.69	0.	72.75
11	34.	1.00	387.69	0.	72.75
12	55.	1.00	387.69	0.	72.75
13	213.	1.00	387.69	0.	72.75
14	232.	1.00	387.69	0.	72.75
15	85.	1.00	387.69	0.	72.75
16	120.	1.00	387.69	0.	72.75
17	331.	1.00	387.69	0.	72.75
18	340.	1.00	387.69	0.	72.75
19	341.	1.00	387.69	0.	72.75
20	345.	1.00	387.69	0.	72.75

NOTE: The next 428 segments are not shown in this table.

Segments with *HAZ* = 0 and *PCR* ≥ 4.0 have been moved to routine maintenance list.

Factor	Input Weight	Norm Weight
PCR	1.5	33.3
ADT	2.0	44.4
HAZ	1.0	22.2
COST	0.	0.

Stage 0 ranking, is shown in Table 1. At this point, there are 24 segments with *PCR* = 1.00 *ADT* (synthetic) = 387.69, and *HAZ* = 0., tying them for ninth position—at least until their synthetic volumes can be replaced with actual counts. There are similar ties throughout this first priority list, but already the relative needs of the county segments have begun to emerge.

Step 8: Acquire Actual Counts to Replace Synthetic Volumes on High-Ranking Synthetic Segments

In a county with over 1,000 mi of roads (over 85 percent of which did not have up-to-date counts), it is useful to know which segment counts will have immediate importance. The road supervisor can weigh this guidance against the desire to avoid placing counters at widely scattered locations throughout the county. The LaPorte County's first volume count update contained 168 new actual segment volumes that affected 8 of the 18 synthetic segments in the top 20 of the first priority list (Table 1).

Step 9: Check for Misplaced Segments, Then Update Synthetic Volumes

The stratum assignments in Step 3 were based on the best available judgment, but 5 of the 168 new actual volumes indicated a need to reassign the segment involved. After the reassignments, the new (Stage 1) synthetic volumes became

$$\bar{v}_{lo} = 115.77;\ \bar{v}_{med} = 447.46;\ \bar{v}_{hi} = 2{,}936.58$$

Thus, even those 10 synthetic segments in the current top 20 that did not get actual volumes for use in the next stage of priority setting will be using updated synthetic volumes.

Step 10: Repeat Steps 6–9

During the course of this study, the county provided 12 sets of volume updates. The most important updates, in terms of size and significance, are presented in Table 2. After four updates, 322 of the county's 668 road segments had received actual volume counts and only one synthetic segment (No. 208) remained in the top 20. If project priority decisions were made at this point, the data in Table 3 or its full 448-segment version would be used as the basis. A total of 7 of the top 20 and 12 of the top 33 segments in Table 1 or its longer version remain that highly ranked in Table 3. This means that some original synthetic volumes were fairly accurate but, more important, even the relatively inaccurate ones provided a focus for the county's volume-counting program. In fact, with at most 50 counts, the county could provide the percentile ranking method with sufficient *ADT* data to produce a list that would not have any synthetic segments.

AN APPRAISAL OF THE SYNTHETIC SOLUTION

Use of synthetic volumes appears to be an effective way to (a) minimize delays in implementing priority-setting procedures

TABLE 2 TRENDS IN SYNTHETIC VOLUMES

Stratum:	Low			Medium			High		
	n_k	n_s	New Avg.	n_k	n_s	New Avg.	n_k	n_s	New Avg.
Stage 0	39	-	116.26	39	—	387.69	12	-	4027.33
Stage 1	66	0	115.77	76	14	447.46	26	4	2936.58
Stage 2	77	0	109.96	87	10	455.76	29	3	2931.10
Stage 3	136	0	94.83	138	2	427.74	31	0	2458.32
Stage 4	140	0	94.17	150	0	428.56	32	1	2515.56
Stage 12	267	0	89..08	164	0	422.45	31	1	2347.32

n_k = number of actual segment volumes used to calculate stage k stratum average, which will be used for priority-setting in stage k+1

n_s = number of segments placed in top 20 by percentile method using synthetic volumes ("New Avg.") calculated after the previous stage

Stage 0: Stratum averages based on original 90 "actual segments"

Stages 1-12: Priority-setting done using stratum averages from the results of the previous stage, including counts made since the previous stage

TABLE 3 STAGE 4 PROJECT RANKS BY PERCENTILE METHOD

Stage 0 Rank	Segment No.	*PCR*	*ADT*	*HAZ*	Average Percentage
1	347.	1.00	811.00	1.00	96.03
114	169.	2.00	606.00	1.00	84.23
7	612.	1.00	4,249.00	0.	77.49
32	348.	2.00	381.00	1.00	75.69
10	20.	1.00	927.00	0.	74.40
96	163.	1.00	788.00	0.	73.61
8	155.	1.00	782.00	0.	73.51
9	648.	1.00	637.00	0.	72.71
3	178.	2.00	286.00	1.00	72.64
103	93.	1.00	524.00	0.	72.22
9	154.	1.00	521.00	0.	72.12
9	527.	1.00	444.00	0.	70.73
2	283.	2.00	187.00	1.00	68.04
33	565.	2.00	5,704.00	0.	66.99
34	42.	2.00	3,312.00	0.	66.69
34	208.	2.00	2,515.25	0.	66.39
45	58.	2.00	2,117.00	0.	65.80
256	148.	2.00	1,501.00	0.	65.30
9	656.	1.00	416.00	0.	65.06
256	620.	2.00	1,404.00	0.	65.00

NOTE: The next 428 segments are not shown in this table.

Segments with *HAZ* = 0 and *PCR* ≥ 4.0 have been moved to routine maintenance list.

Factor	Input Weight	Norm Weight
PCR	1.5	33.3
ADT	2.0	44.4
HAZ	1.0	22.2
COST	0.	0.

and (b) maximize the effectiveness of a volume-counting program. To learn more from the LaPorte County case, some of the elements of the process should be examined.

The volume trends contained in Table 2 are shown in Figure 1. The expectation was that the synthetic volumes would (a) approach the actual stratum averages asymptotically from above or below, or that (b) there might be some straddling of the actual average. Asymptotically from above appears to apply to the high- and low-volume trends, whereas the medium-volume plot may be straddling the eventual actual medium-stratum average. However, not each stage has average-volume calculations based on the same increase in the number of actual segments in a stratum. For example, the update for Stage 3 in Table 2 contained many more segments than for Stage 4, so a bigger change coming into Stage 3 in Figure 1 was expected. However, the overall trends from Stage 0 to Stage 12 were instructive:

- The high-volume stratum average started out 1,680.1 vehicles per day (vpd) higher than its current Stage 12 value, an error of +71.6 percent. By Stage 4, the overestimate was reduced to 168.24 vpd, an error of +7 percent.
- The original medium-volume estimate was 34.76 vpd below its latest value, an error of −8.2 percent. By Stage 4, the estimate was 6.11 vpd (+1.5 percent).
- The first low-volume average was a 27.18-vpd (+30.5 percent) overestimate, when compared to the Stage 12 figures. By Stage 4, the difference was reduced to 5.09 vpd (+5.7 percent).
- By Stage 4, all but one of the high-ranking synthetic segments were converted to segments with actual volume counts.

It is also interesting to note that the original high-volume overestimate with respect to the Stage 12 $\bar{v}_{hi}$, which would have a significant effect on the priority rankings, was reduced from 71.6 to 25 percent after only 14 synthetic high-volume segments received actual counts at Stage 1 (see Table 2). Again, the synthetic solution makes possible efficient use of resources and information of immediate value.

The effectiveness of the synthetic volume method depends to a large extent on the proper assignment of segments to strata in Step 3. Table 4 presents the accuracy with which the synthetic segments were assigned to their correct strata. After Stage 1, 11

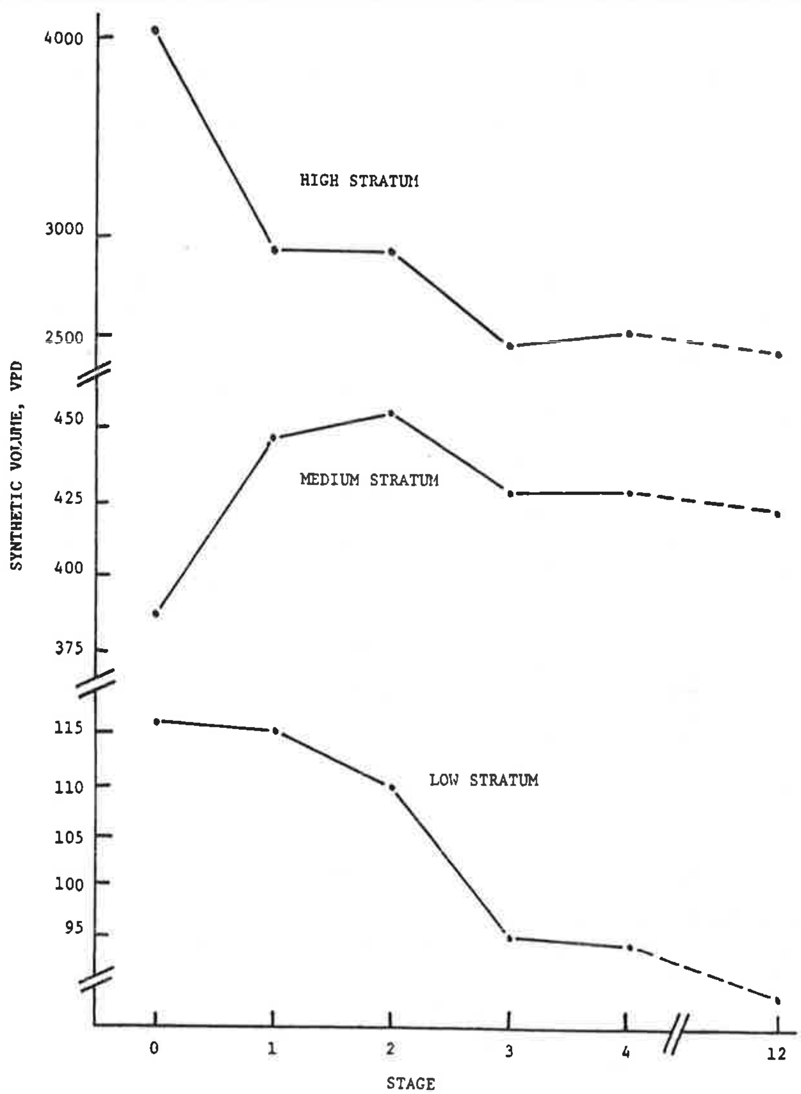


FIGURE 1 Plot of synthetic volume trends.

segments thought to belong to the medium stratum had actual counts that caused them to be reassigned to the low stratum. Another 7 medium segments turned out to have high *ADT* values. Out of 78 counts made at Stage 1 in preparation for Stage 2, 29 segments had to be moved. In most cases, these were segments with actual volumes near a stratum boundary. Only rarely to were drastic reassignments (low to high or high to low) necessary, and these were often due to data entry errors. Experience to date indicates no reason to use more than three strata or to revise the stratum boundaries in the case study.

TABLE 4 DEGREE OF ACCURACY IN ASSIGNING SEGMENTS TO STRATA

	Assigned Stratum	Revised Stratum		
		High	Medium	Low
Stage 0	High	36	0	0
	Medium	4	237	1
	Low	0	0	390
Stage 1	High	39	0	1
	Medium	7	219	11
	Low	2	8	381
Stage 2	High	43	2	3
	Medium	1	218	8
	Low	0	1	392
Stage 3	High	34	6	4
	Medium	5	203	13
	Low	2	32	369
Stage 4	High	39	2	0
	Medium	2	239	0
	Low	1	7	378
Stage 12	High	38	0	1
	Medium	0	231	5
	Low	0	3	390

One more way to examine the evolution of the *ADT* datafile from largely synthetic to primarily actual data is through some measure of error that quantifies the relationship between each newly counted segment's actual volume and its most recent synthetic volume. The equation used for each stratum at each stage was

$$D = \sum_{j \in S} (syn_j - act_j)/n \qquad (1)$$

where j indicates a segment that has just received an actual volume act_j at Stage k; n is the number of such segments; and syn_j is Segment j's (and Stratum S's) most recent synthetic volume. Because $syn_j = syn_S$, Equation 1 was simplified to

$$D = syn_S - (1/n) \sum_{j \in S} act_j \qquad (2)$$

and used to produce the entries in Table 5. Because this measure is based only on the segments receiving actual counts at the current stage, and is not cumulative, some volatility might be expected. The low (0–200 vpd) and medium (200–1,000 vpd) strata seem to have well-behaved D values, despite an occasional synthetic segment assigned to the wrong stratum. The high-volume (≥1,000-vpd) stratum has the most room for variation, as is evident in Table 5. At each stage, the synthetic counts are generally much higher than the actual. At Stage 3, there were an unexpectedly high number of segments that were either misassigned (Table 4) or had actual *ADT* values just over 1,000 vpd, which caused the high D value. The D value emphasizes the need to get actual volume counts for any synthetic segments with high preliminary priorities, especially if they are listed as high-volume segments. They are likely to be carrying an overestimated *ADT* and, therefore, too high a ranking. Thus, low D values indicate good judgment in choosing the number of strata and in assigning segments to them. Consistently high D values may indicate the need for more strata with smaller ranges, but as long as high-ranking synthetic segments get actual counts before decisions are made, the impacts of these inaccuracies are negated.

TABLE 5 AVERAGE DIFFERENCE BETWEEN NEW ACTUAL AND PREVIOUS SYNTHETIC VOLUMES

Stratum	High	Medium	Low
Stage 0	---	---	---
Stage 1	2065.97	-108.08	-2.09
Stage 2	52.91	-65.63	40.68
Stage 3	1591.43	68.95	30.15
Stage 4	651.32	66.41	-13.84
Stage 12	0	174.43	37.37

NOTE: Negative entry in table means average synthetic volume was lower than average actual volume by that amount. Large entries are usually due to one or more segments having been placed in the wrong stratum.

The most important analysis of the synthetic method takes place using the project ranking lists themselves for successive stages. The synthetic method exploits the facts that most high-ranking synthetic segments have overestimated *ADT* values, and after the first one or two stages it is rare for a holdover synthetic segment to move into the top 20 or 30. Thus, it takes only a few stages (i.e., iterations through Steps 6–9) to develop a reliable database for ranking the most deserving projects, even if data were scarce at the start. In the case study, only three or four stages were necessary.

THE FINISHING TOUCHES

Throughout the discussion of the case study, the only factor given synthetic values was *ADT*, but in Tables 1 and 3 *COST* is given a weight of zero and is not included in the ranking process. This exclusion is because to expect county officials to maintain ongoing cost estimates for all road segments is unreasonable. Instead, it is wise to wait until the candidates for major roadwork have been identified, such as in Table 3. Then the highest-ranking segments can receive as detailed a cost estimation effort as desired. The decision makers can go down the latest ranked list, approving and skipping projects until the accumulated costs of approved projects are about to exceed the budget. Another approach is to enter these cost values into the priority-setting procedure, assign *COST* an appropriate factor weight, and examine the results. Using 1.0 as the *COST* weight, the percentile method produces the data of Table 6. The priorities are quite similar to those in Table 3, although some high-cost projects including synthetic segment No. 208 slip down a few positions.

This last-minute approach to *COST* values contrasts with the synthetic approach to *ADT* values, but the philosophy is the same: obtain the most useful information for the least effort. It would be possible to develop a synthetic cost function based, for example, on a segment's length, width, *PCR*, *HAZ*, and *ADT* values. If *COST* is believed to be an important factor in sorting out project ranks during the early iterations or stages, the function could provide approximate (synthetic) *COST* values that would be replaced by actual estimates for the top 20 or so segments. The synthetic *COST* value, however, seems much more postponable than does the *ADT* value.

COMMENTS

This paper has mentioned only road project priorities at the county level, but there is nothing about the synthetic method to prevent its use for bridge project prioritization or by cities. What is important to realize is that the synthetic method does require multiple applications of whatever priority-setting procedure is adopted. Therefore, that procedure had better not be too expensive or inconvenient to run several times over a short period of time. The percentile ranking method used in this paper's case study is one of several that is quick and economical. In the statistical analyses summarized in Tables 2, 4, and 5, the synthetic method is an approximation procedure. On the other hand, the actual case study rankings (Tables 1, 3, and 6) are encouraging in their consistency and logic.

TABLE 6 STAGE 4 PROJECT RANKS WITH COST DATA INCLUDED

Segment No.	*PCR*	*ADT*	*HAZ*	*COST*	Average Percentage
347.	1.00	811.00	1.00	7.80	96.90
169.	2.00	606.00	1.00	17.20	86.99
612.	1.00	4,249.00	0.	33.80	81.25
348.	2.00	381.00	1.00	86.00	79.58
20.	1.00	927.00	0.	83.60	78.56
648.	1.00	637.00	0.	8.40	77.78
155.	1.00	782.00	0.	146.30	77.42
163.	1.00	788.00	0.	761.40	77.21
154.	1.00	521.00	0.	15.20	77.17
93.	1.00	524.00	0.	134.70	76.44
527.	1.00	444.00	0.	93.50	75.38
283.	2.00	187.00	1.00	62.70	73.47
565.	2.00	5,704.00	0.	690.10	71.75
42.	2.00	3,312.00	0.	154.80	71.71
58.	2.00	2,117.00	0.	92.00	71.39
656.	1.00	416.00	0.	21.00	71.34
148.	2.00	1,501.00	0.	61.90	71.22
603.	1.00	408.00	0.	15.60	71.10
303.	2.00	1,087.00	0.	54.70	70.69
528.	3.00	1,379.00	1.00	108.70	70.27

NOTE: The next 428 segments are not shown in this table.

Segments with *HAZ* = 0 and *PCR* ≥ 4.0 have been moved to routine maintenance list.

Factor	Input Weight	Norm Weight
PCR	1.5	27.3
ADT	2.0	36.4
HAZ	1.0	18.2
COST	1.0	18.2

ACKNOWLEDGMENTS

This study was supported by funds provided through the Highway Extension and Research Project for Indiana Counties and Cities, School of Civil Engineering, Purdue University. The author's deepest gratitude goes to Gene Shurte, Road Supervisor, LaPorte County, Indiana. Thanks to his extraordinary cooperation, the ideas in this paper have had an excellent opportunity to be demonstrated and evaluated. The computer programs needed to repeatedly update the stratum averages were developed and implemented by Huey-Kuo Chen, whose diligence and precision is gratefully acknowledged.

REFERENCES

1. Computers in Highway Organizations. *Pothole Gazette*, Vol. 4, No. 4, Highway Extension and Research Project for Indiana Counties and Cities, School of Civil Engineering, Purdue University, West Lafayette, Ind., Dec. 1985.
2. J. L. Shaffer. *A Methodology for Determining and Prioritizing County Highway Network Needs.* MSCE thesis, Purdue University, West Lafayette, Ind., May 1986.
3. J. L. Shaffer and J. D. Fricker. Simplified Procedures for Determining County Road Project Priorities. In *Transportation Research Record 1124.* TRB, National Research Council, Washington, D.C., 1987, pp. 8–16.

A Database Management System for Traffic Engineering Departments of Small and Medium-Sized Cities

KARSTEN G. BAASS AND BRUNO ALLARD

Most traffic information systems now in use are inadequate. The data they contain are often inaccessible, incompatible with other databases, and rarely up to date. Furthermore, the data are rarely fully exploited because they are usually in manual form. The proposed system attempts to solve these problems. It is a database management system based on dBASE III PLUS that includes an integrated urban data bank and three types of software for information systems management, information retrieval, and interfaces with application programs. A sector of the city of Quebec was used to test the feasibility of the system. Although limited, the use of the data shows that the proposed system has distinct advantages over existing approaches. The system will become even more attractive as cheaper, more powerful, and more versatile microcomputers enter the market.

In the transportation planning and traffic engineering field, information needs are of a great variety and consequently data at different levels of detail and precision are needed. One can distinguish between four principal groups of information users, which correspond to the levels of hierarchy of the transportation administration. These are the executing personnel at the technical level, the planners at the tactical level, the executives having administrative or managerial authority in the organization at the strategic level, and finally the users from outside the organization who represent an important and noticeable group due to the great public interest in the transportation domain. Figure 1 shows the differences in the information needs of these groups.

The technical personnel, responsible for the daily operation of the system, require detailed information that should be precise and up to date. This basic information is generated and collected mostly by the department itself. The need for data at this level is constantly growing due to the introduction of computerized procedures and expert systems for design and operation. On the other hand, planners need data coming from various sources (e.g., from other city departments) that may be less detailed and less recent. The administrative or managerial authority of the department requires synthesized and more global information, which should be available rapidly in order to allow optimal decision making based on facts. Finally, external users such as citizen groups, consulting engineers, and more frequently lawyers need particular and precise information. These users are often poorly served by the existing information systems and it is here where the public realizes its inadequacy. A good information system should satisfy all these different needs.

Ecole Polytechnique de Montreal, Department of Civil Engineering, P.O. Box 6079, Station A, Montreal H3C 3A7, Canada.

Computerized information systems have evolved rapidly; from application program oriented databases to complex database management systems (DBMS). Figure 2 shows the basic principle of such a system; the literature *(1–3)* can be consulted for further information.

Today many preprogrammed DBMS packages are widely available on large-scale computers. These have many interesting features, supply information in a user-friendly way, and supply information that one could otherwise obtain only by time-consuming and expensive programming. These programs can be used efficiently for the management of urban transportation data.

Some of the more important objectives (which can be attained by certain commercial DBMS) to pursue in the design of an urban information system are the following:

- Independence of data. (Two data are independent if modifications can be carried out on one, without regard to the other.)
- Nonredundancy of data. (This is an important requirement with respect to storage and to the updating process.)
- Interrelation between different urban data by common references.
- Integrity of data.
- Modularity of the system.

From the user's standpoint a good DBMS should provide

- Multiuser capability.
- Simplicity of use and comprehensive documentation (full-screen editing, on-screen display format generation, report generator, easy query facilities, etc.).
- Restructure ability. (Because the user's need may change in time, the system should allow addition to or deletion of data records.)
- Security of data by a password or other control procedure.
- Easy updating of data.

Due to the commercially available DBMS, the complex duties of data management become easy. These programs offer enormous advantages in the field of traffic engineering where great quantities of data have to be stored, treated, and analyzed. Because the traffic engineering department is part of the urban

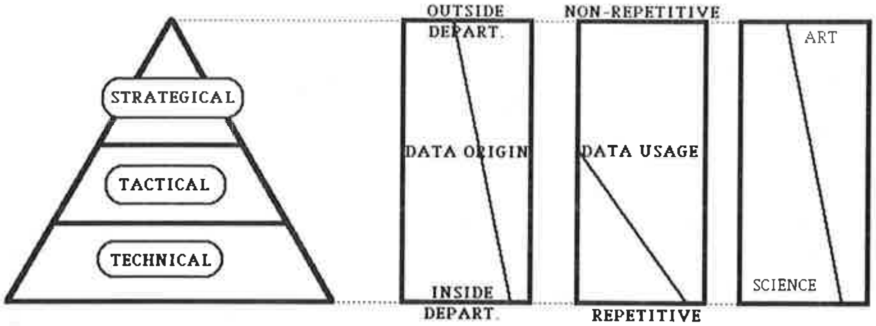


FIGURE 1 The different information needs.

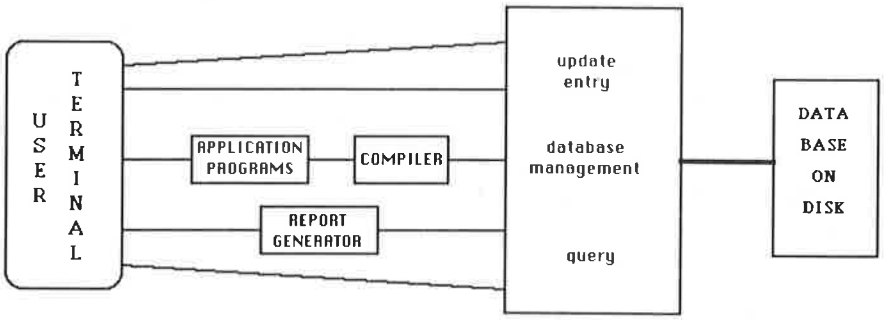


FIGURE 2 A database management system.

services and one wishes to solve urban problems in a more global way, a traffic information system should be integrated into the data of the other urban services to avoid the problems of redundancy and difficulties in data transfer between services. Figure 3 shows how one could design this urban information system (UIS). Common references, such as the links and geocoded nodes of the street network, would permit exchange and integration of data between city departments. All departments are using the same programs for interactive interrogation and exploitation of the database. Control, security, updating, and data entry depend on one authority. In theory, this UIS is ideal. It is fast and efficient, centralized, integrated, and modular and makes maximal use of the data that are so expensive to collect and to keep updated. But there are only a few municipalities undertaking the task of establishing such a system due to the large expenses. The administration of a city touches so many domains that an adequate design for all users seems to be very complex. So complex, indeed, that the creation of two systems,

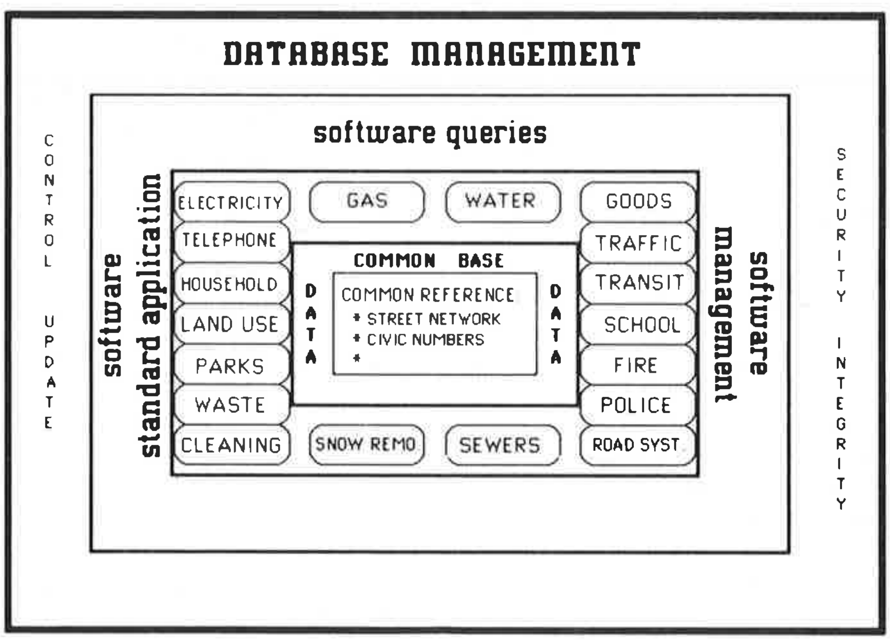


FIGURE 3 Diagram of an integrated urban information system (UIS).

one for planning and one for urban operation purposes, is often proposed.

From a practical standpoint, the centralized UIS is often questioned because the user is dependent on a central system over which the user has no control, especially if the user's queries are not considered a priority by the UIS administrator.

Finally, the mere size of such a centralized data management system seems to contribute to its failure. Recent developments in the field of networking of microcomputers indicate that the centralized approach could soon be replaced by a decentralized one.

For these reasons, the development of a global UIS does not seem to be the way to follow for a municipality with little or no experience in the domain of database management. To avoid the problems related to the implementation of a global information system, a municipality could start temporarily with a DBMS in each and every one of its departments.

A central authority should assure that these data systems remain compatible with one another (e.g., by providing a common geocoded network as a reference). Communication between databases of the different departments would then easily be possible. This procedure would, however, not eliminate redundancy of certain data and would cause some problems in updating elements common to all databases.

But because there are relatively few common data (at least at the operational level where the availability of recent data is a must), this approach would represent important advantages compared to the situation presently encountered in many municipal departments. This partial database would allow the start of computerization (once data are stored on disk they can always be reused or restructured), and would be a transitional stage necessary to fill the gap between the manual information system and a fully integrated UIS.

THE INFORMATION SYSTEM FOR A TRAFFIC ENGINEERING DEPARTMENT

The decentralized approach in the design of a DBMS in traffic engineering was adopted. In an early version, started at the end of the 1970s, the problem was attacked by programming on a centralized mainframe computer. Part of the interactive query and management programs were written. But soon it appeared that an information system would necessitate a significant investment in programming effort or would require the acquisition of expensive DBMS programs. Management often depends on the expertise of programmers and many wish to preserve that dependent relationship. Software houses offering programs for mainframe computers have traditionally not offered packages that permitted computer end users to bypass the programmers. The first database systems were thus complex and difficult to use. Only recently, systems on microcomputers as sophisticated or more so than those on large computers became available.

It was decided that an information system based on a commercial DBMS on a microcomputer should be designed, supposing that microcomputers would develop rapidly with respect to their storage capacities and rapidity of execution. It was also expected that the available DBMS would also become more and more sophisticated and would grow with the needs of the user community.

It was then decided to use dBASE II *(4)*, which was widely available, and even considering the limitations of this early version, it appeared after having conducted a pilot study that the commercial packages were in fact satisfactory for the purpose.

The dBASE system is a relational database management system in which data are stored and related to other data by relations. Under certain conditions, the files can be considered mathematical relations, and the mathematical theory of relational algebra can be applied to problems dealing with the data in these files. One of the basic advantages of a relational database is that one does not have to anticipate all the needs when the files are set up. The dBASE system is now in its third version (dBASE III PLUS), and nearly all of the objectives cited earlier have been attained by the program. A continuity between different versions is guaranteed, so that data can easily be accepted from the older versions.

The management of the data bank (including security, updating, data entry, modification of the data structure, adding new data items, etc.), the query language, and the report preparation are all well developed.

There is a query facility for nonprogrammers to view, enter, and update the data, and a report generator orienting the information to paper instead of to the screen. A special programming language as part of the DBMS can be used to program any report based on the data. The DBMS allows the data to be read and data files to be produced for application programs in BASIC, FORTRAN, Lotus, or other languages. One basic requirement in a practical environment is to allow access to the database by several users at the same time. The network environment is presently in a rapid state of evolution and the newest version of dBASE allows multiuser operation.

There are in fact a number of access and update security codes that not only protect the data values but also relationships between data. Files are protected by data encryption. They cannot be read until they are deciphered or decrypted. A user's profile is created with an encryption code and a password. No user can gain access to the DBMS unless he provides a log-in that can be validated. There are eight different privilege levels for reading, updating, extending, and deleting data where Level 1 is the fullest privilege. These privileges are established by the database administrator and fixed in the user's profile.

The built-in local area networking (LAN) (Figure 4), together with the password protect system, allows some network users only to enter data, whereas other users create and maintain the databases, command files, or applications. Figure 1 shows the scheme whereby different codes are attributed to different levels of users.

There is also a run time package that allows running a compiled version of the dBASE application program without buying the DBMS program. This package makes the application more efficient and faster.

THE DATABASE

Data Integration

Because the goals of data sharing between different city departments and future integration of the data should be pursued, the most basic question to answer is how to establish the basic relation by appropriate linkages.

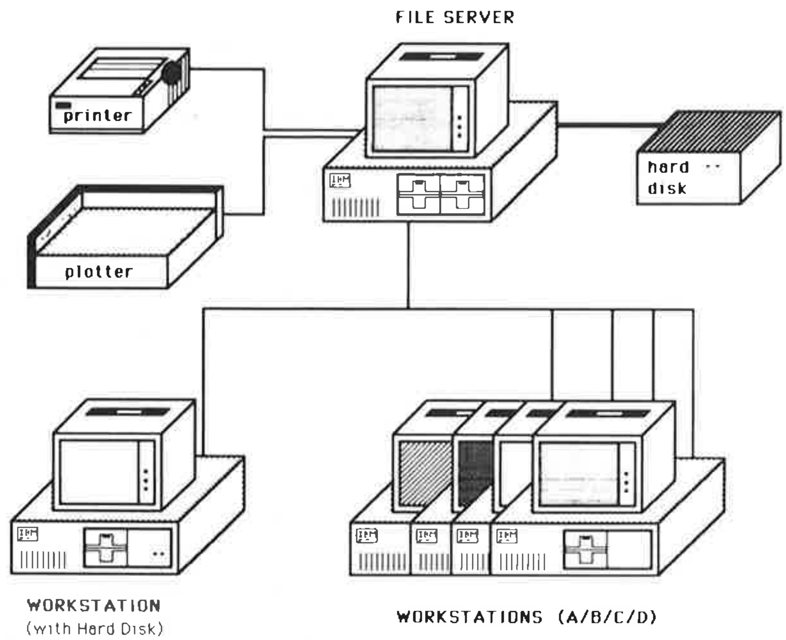


FIGURE 4 The LAN environment.

Urban information pertains to different entities of which Nora *(5)* has identified several:

- Individuals,
- Corporations,
- Public equipment,
- Buildings,
- Building lots, and
- Street segments.

It appears that geographical coordinates are the best, although not exclusive, means of urban data integration because most of the entities have in fact a spatial dimension. Traffic and transportation data are primarily associated with road segments, which can be geocoded by x–y coordinates and by map digitizing after having represented the street network as a network composed of links and nodes. A road segment is thereby represented by two nodes at the street intersections with intermediate nodes, if necessary, for a better graphical representation.

The Dual Independent Map Encoding (DIME) developed by the U.S. Bureau of the Census in 1970 is one such system *(6)*. The proposed reference system is similar to the DIME system (see Figure 5) and uses the Urban Transverse Mercator (UTM) system coordinates, allowing an integration of Canadian census data into the database.

The Contents of the Traffic Engineering Database

There are two kinds of data related to traffic engineering, data related to an intersection (e.g., traffic light parameters) and data related to street segments. Some rules for data collection and data entry are necessary, but most departments have adopted a way similar to the one shown by Figure 5 to enter segment-related data. There is a special application program that calculates for each item entered in the database its x–y coordinates

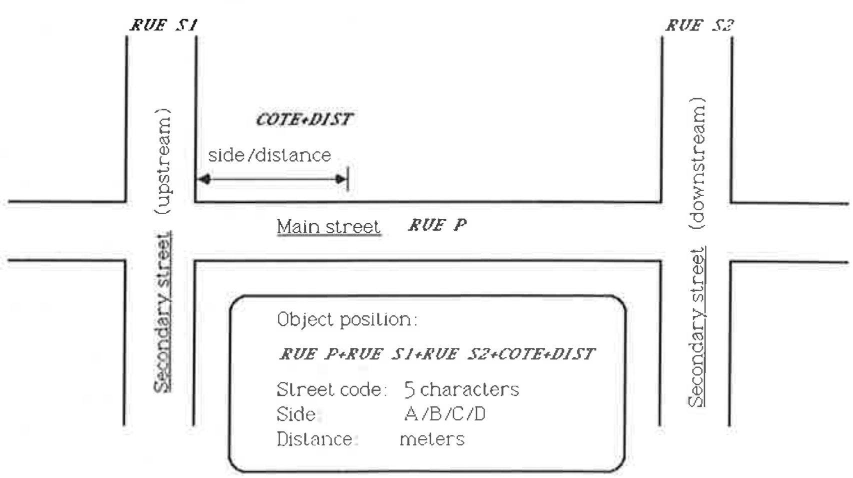


FIGURE 5 Description of the reference system.

based on the coordinates of the nodes describing the segment. This allows an easy and fast graphical scaled representation of the data.

The following data are considered.

Segment-related data

- General data

  Civic numbers
  Private entrances
  Street lamps
  Bus shelters

- Geometry

  Width
  Nodes
  Links
  Islands
  Turning lanes

- Traffic control devices

  Location
  Traffic sign type
  Conditions day and night
  Maintenance dates
  Pavement markings
  Sign posts

- Traffic information

  Volume
  Traffic composition
  Spot speeds
  Overall travel speeds
  Delays
  Capacity
  Class of the street
  Parking

- Accident information (information corresponding to the standard accident report used in Quebec)
- Pavement nature and road surface condition

  Roughness
  Structure
  Bearing capacity
  Cracking
  Maintenance

Intersection-related data

- Geometry

  Turning radius
  Islands
  Sidewalks
  Nodes and links
  Lanes and turning lanes

- Traffic signs

  Location
  Type with code
  Condition of sign
  Maintenance
  Pavement markings
  Sign posts

- Traffic lights

  Cycle
  Splits
  Phases
  Controller
  Traffic light type and condition

- Traffic information

  Volumes
  Traffic composition
  Capacity
  Saturation flow
  Buses
  Parking

This information, when gathered and stored in the DBMS, will contribute to better management, will permit identification of deficient system elements, and will allow prioritizing and scheduling maintenance work. This information will be most helpful for answering queries from the public and will also be useful in legal problems when it has to be shown that the traffic inventory is up to date and stored and treated on a modern and efficient state-of-the-art system.

GENERAL DESCRIPTION OF THE SYSTEM

The initial problem of design was to define an efficient structure of data storage because the way data were stored influenced the efficiency of the DBMS. Yet the structure had to be flexible enough to adapt to the changing demands of the user. This requirement was met because dBASE allowed easy modification of the data structure. Certain weaknesses of the information system appearing during its use could thus be corrected in a convenient way.

After the starting screen giving the logo of the system designed by Allard *(7)* appears, the user is asked to fill out the log-in menu with a code and a password. This menu enables the user to change certain features as, for example, the colors of the menus and the databases to be used (e.g., if there are different city departments).

There are two modes of operation of the system, by menus and by macrocommands. Menus facilitate greatly the use of the program, but the user is limited to the options offered on the menus, underlining the necessity of providing an adequate selection for the user. In fact, the options on the menus were chosen by the potential system users. Standardization and simplicity of the menus were achieved by designing all menus corresponding to the standard layout shown on Figure 6. The menus were designed with the help of an on-screen cursor design program [e.g., QUICKCODE *(8)*].

There are several levels of menus distinguished by different color arrangements defined in the user's profile. The hierarchy of the first two levels is shown in Figure 7.

The main menu in Figure 8 shows the basic options available at the moment that correspond to the basic tasks, data management, traffic studies, and analysis.

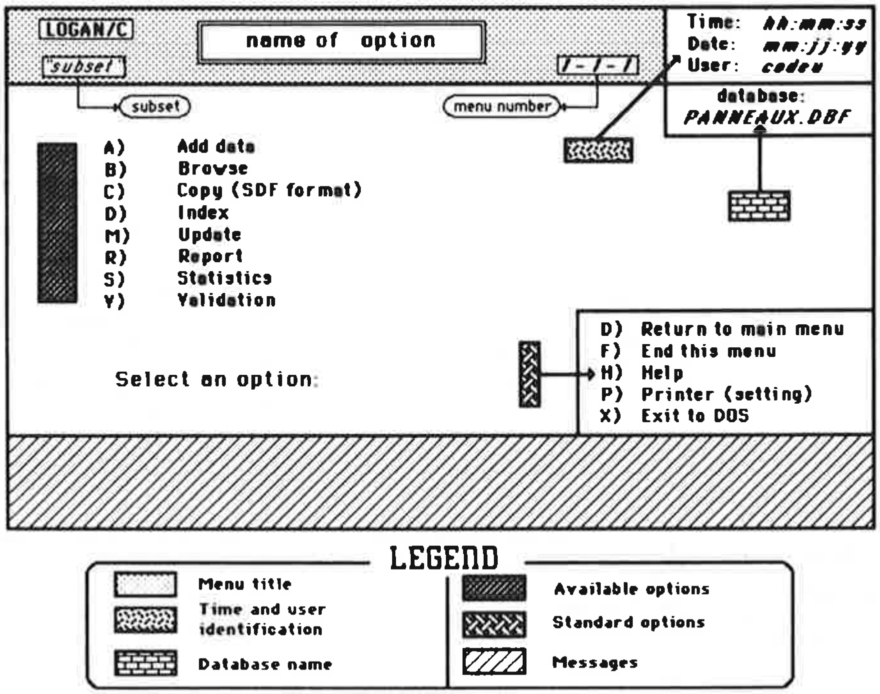


FIGURE 6 The standard for menu design.

Options 1, 5, and 6 allow data entry, validation, and updating. Option 2 can be chosen to establish a new database as a subset of the original one. Option 3 represents the query part, for which direct commands available by dBASE permit the display, on screen or in print, of the contents of the database or part of it. On an intermediate level, between these in-house commands and the application programs, are the preprogrammed repetitive reports generated by the report generator of dBASE. The design of these reports requires some knowledge of the dBASE programming language. The elements described are necessary for the inventory and operation of the traffic service and as such are useful on the technical level, but graphical output is also useful.

These applications cannot be done directly by existing commands of the database management system. These applications are accessible by choosing Option 4 of the main menu. The programs are written in FORTRAN and data are extracted interactively from the database.

The remaining options facilitate work with the system, for example, by giving explanations about the functioning of the system (i.e., the HELP option). LOGAN/C was designed to be used in the French language but the menus in the following application examples were translated for easier understanding.

Examples of Work To Be Done by the Information System

Updating the Information in the Database

This task is done by choosing Option 1 on the main menu. Menu 1–1 (Figure 9) appears on the screen, from which Number 1 is chosen for updating civic numbers.

This choice produces the Menu 1–1–1 shown in Figure 10 and displays nine different options, from adding of data to updating, validating, browsing through the database, and so on.

Option A on this menu produces a lower level of menu, which permits the addition of new information (see Figure 11).

Predefined and Ad Hoc Analysis

Option 3 from the main menu gives access to analysis on the database. Certain analyses are required periodically. A program extracting the information and displaying it in a certain manner on screen and by printer is preprogrammed to facilitate the production of these repetitive reports. Four typical examples are given on the menu in Figure 12.

One of the advantages of the DBMS is the possibility of querying the database and extracting the desired information directly. Thus, a subset of data is obtained without complex programming. Boolean operators of relational algebra can be used to create new associations or to obtain a subset of data. Two examples of such ad hoc analyses may illustrate its potential.

In the first example, a list is needed of all traffic signs on Côte d'Abraham, whose five-letter code is CABRA. The database PANNEAU (or traffic sign) is queried.

The commands are as follows:

```
USE PANNEAU
LIST FOR RUE1='CABRA'
```

Figure 13 shows the variable names and database names used in the examples.

In the second example, traffic signs are obtained that have a day visibility code of 3 or more on Côte d'Abraham and are upstream from the intersection with the street with code name SVALL between 20 and 1000 m from that intersection. Figure 14 shows the menu that achieves this and that also displays the

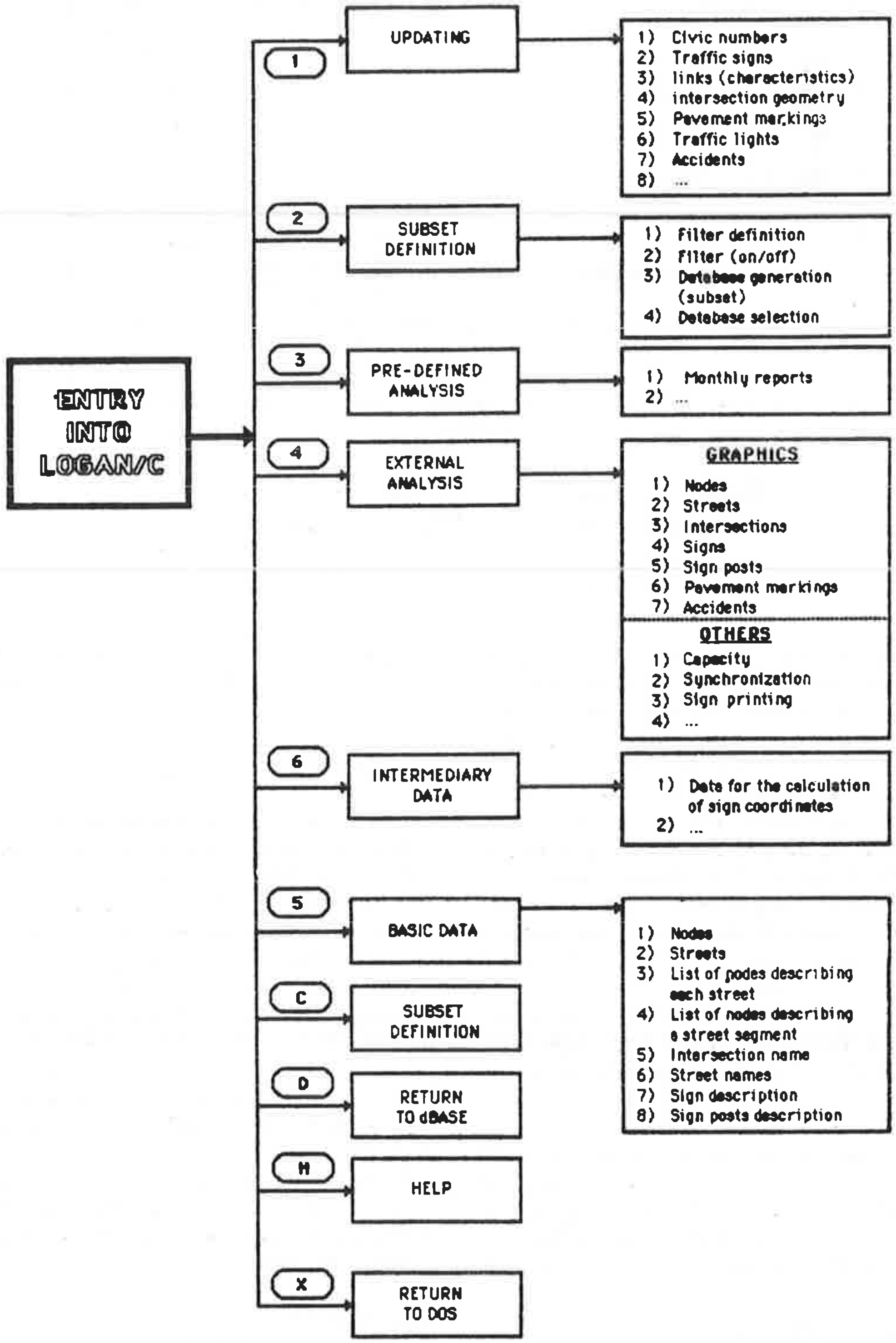


FIGURE 7 The first two menu levels of the information system: (a) first menu level, (b) second menu level.

```
.........................................................................
|                     ....................            | Time:    14:50:08 |
|                     |  L O G A N / C   |            | Date:    06/20/86 |
|                     ....................            | User:    BRUNO    |
.........................................................................
|
|        1)      Updating data
|        2)      Subset definition
|        3)      Pre-defined analysis
|        4)      External application
|        5)      Entrance of basic data
|        6)      Intermediairy data
|
.........................................................................
|        C)      Configuration/user profile
|        D)      End of dBASE
|        H)      Help
|        X)      Return to DOS
.........................................................................
```

FIGURE 8 The main menu of the information system.

```
.........................................................................
|  logan / C     ..........................            | Time:    13:59:05 |
|                |    DATABASE UPDATE     |            | Date:    06/20/86 |
|                ..........................    1-1     | User:    BRUNO    |
.........................................................................
|
|      1)        Civic numbers
|      2)        Traffic Signs and sign posts
|      3)        Links (characteristics)
|      4)        Intersection geometry
|      5)        Pavement marking
|      6)        Traffic lights
|      7)        Accidents on links
|      8)        Traffic counts
|                                               ..........................
|                                               | D) Return to main menu
|                                               | F) Exit this menu
|                                               | H) Help
|                                               | X) Return to DOS
.........................................................................
```

FIGURE 9 Menu giving access to the different databases.

```
.........................................................................
|  LOGAN / C        ....................            | Time:    13:55:12 |
|                   |  CIVIC  NUMBERS  |            | Date:    06/20/86 |
|                   ------ update -------   1-1-1   | User:    BRUNO    |
.........................................................................
|                                                   |  database:
|     A)      Add data                              |  CIVIQUE.DBF
|     B)      Browse                                ..................
|     C)      Copy (SDF format)
|     I)      Index
|     M)      Update
|     R)      Report
|     S)      Statistics
|     V)      Validation
|     W)      Subset                        .........................
|                                           | D) Return to main menu
|                                           | F) Exit this menu
|          Select an option:                | H) Help
|                                           | P) Printer (setting)
|                                           | X) Retunr to DOS
.........................................................................
```

FIGURE 10 Menu for work on the civic numbers.

```
..................................... :  ADD  :......................................
| LOGAN / C            ....................                 | Time:      14:59:05 |
|                      | CIVIC   NUMBERS  |                 | Date:      06/20/86 |
| :_____:              ----- update -------        1-1-1.A| User:      BRUNO    |
....................................................................................
|                              Census track ID                                   |
|         Codes                                        Full names (output)        |
| main street:                  :_____:     |   :                            :   |
| second. street(upstream)      :_____:     |   :                            :   |
| second. street(downstr.)      :_____:     |   :                            :   |
| street side:                  :_:         |                                    |
| distance from intersect.      :______:    |                                    |
| (upstream)                                |                                    |
| . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .|
|                               Address number                                   |
|                                                                                |
|     civic number:             :____:                                           |
|     fraction code:            :_:                                              |
|                                                                                |
....................................................................................
```

FIGURE 11 **Entrance form for data on civic numbers.**

```
....................................................................................
| LOGAN / C       ...........................                 | Time:     15:34:05 |
|                 |  PRE-DEFINED  ANALYSIS  |                 | Date:     06/20/86 |
|                 ...........................      1-3        | User:     BRUNO    |
....................................................................................
|                                                                                  |
|      1)   List of signs requiring maintenance                                   |
|      2)   Number of signs per type                                              |
|      3)   Number of fatal accident per street                                   |
|      4)   Number of accident per type                                           |
|                                                                                  |
|                                                 ................................ |
|                                                 | D) Return to main menu        |
|                                                 | F) Exit this menu             |
|                                                 | H) Help                       |
|                                                 | P) Printer (setting)          |
|                                                 | X) Retunr to DOS              |
....................................................................................
```

FIGURE 12 **Predefined analysis.**

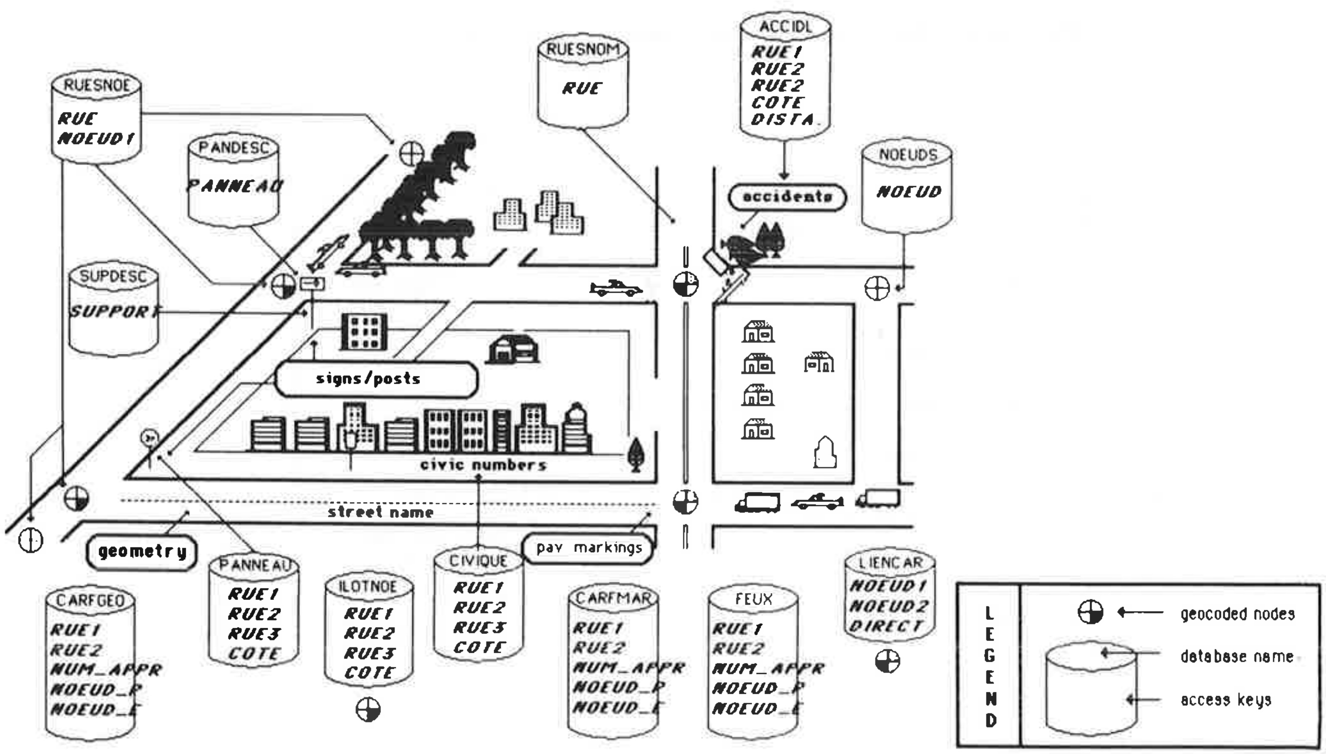


FIGURE 13 **Names of variables and databases.**

```
LOGAN / C    ------------------------------         | Time:    11:24:35
             |  DEFINITION / UTILISATION   |         | Date:    06/21/86
             ------    OF  SUBSET     ------    1-2  | User:    BRUNO
------------------------------------------------------------------------

     1)        Definition of filters on database
     2)        Filters (on/off)
     3)        Generation of new databases   [*]
     4)        Choice of database subset

                                          -----------------------------
                                          | D) Return to main menu
                                          | F) Exit this menu
               * Use option 1 to specify  | H) Help
                 the conditions           | P) Printer (setting)
                                          | X) Retunr to DOS
```

FIGURE 14 Menu used for the definition of a condition for extraction of data from the database.

```
FIELD NAME    /  DESCRIPTION               | FIELD NAME  /  DESCRIPTION
        type                               |         type
RUE1:    (C5) Main street name             | D_FABRICA(D) Fabrication date
RUE2:    (C5) Second. street name (up)     | D_INSTALL(D) Installation date
RUE3:    (C5) Second. ... (downstream)     | D_ENTRETI(D) Maintenance date  .
COTE:    (C1) Street side (A/B/C/D)        | HAUTEUR (N2) Hight  (cm)
DISTANCE(N6) Dist. from intersect.         | DIM_LARG(N2) Width
PANNEAU (N4) Sign number                   | DIM_LONG(N2) Length
CO_FLECHE(C1) Arrow code (from sign)       | C_REMPLAC(N1)Reason for maintenance
SUPPORT (N4) Sign post number              | FABRICANT(N1)Supplier
SURFACE (N1) Sign post surface             |
ETAT_JOUR(N1) Sign visibility: day         |
ETAT_NUIT(N1) Sign visibility: night       |
VISIBILIT(N1) Sign reflectorization        |
-------------------------------------------------------------------------
 Example: RUE1='CABRA' .OR. RUE1='SREAL' .AND. RUE2='SVALL' .AND. DISTANCE > 20
.AND. DISTANCE < 1000 .AND. D_ENTRETI < CTOD("10/20/85").AND. ETAT_JOUR > 3
            ENTER YOUR CHOICE
```

FIGURE 15 Menu used to define a subset on the database.

```
---------------------------------- : ADD :----------------------------------
 LOGAN / C      ----------------------              | Time:     14:59:05
                | D E F I N I T I O N |              | Date:     06/20/86
                ------  subset  -------      1-2-1.  | User:     BRUNO
--------------------------------------------------------------------------
                                                     | database name
 Subset title:                :_________:            |    FILTRE.DBF
 Subset name                  :________:             ---------------------
 Description of subset        :_:
 (optional)
--------------------------------------------------------------------------
     Database name            :ACCIDENT:

     Conditions
 :_______________________________________________________________:
 :_______________________________________________________________:
```

FIGURE 16 Menu for the definition of a subset of the accident database.

```
LOGAN / C                                              | Time:    10:21:15
            |  EXTERNAL  ANALYSIS   |                  | Date:    06/20/86
                                              1-4      | User:    BRUNO

       G)   GRAPHICS                     A)   Others
   1)    Nodes                       1)  Capacity analysis
   2)    Streets                     2)  Traffic light coordination
   3)    Intersections               3)  Printed list of traffic signs
   4)    Traffic signs               4)  Printed list of sign posts
   5)    Sign posts
   5)    Pavement markings
   6)    Accidents
                                         | D) Return to main menu
                                         | F) Exit this menu
                                         | H) Help
                                         | P) Printer (setting)
                                         | X) Retunr to DOS
```

FIGURE 17 Menu to give access to the application programs.

names of the variables and their description, thus rendering research in the dictionary unnecessary.

Another interesting feature is presented by Option 2 of the main menu, which permits generation of subsets of data by defining filters on the databases. Figure 15 is the corresponding menu, and Figure 16 is displayed after having chosen Option 1 in Menu 1–2.

In many cases in traffic engineering, it becomes necessary to use the data as inputs to application programs. Option 4 of the main menu gives access to a number of application programs that will be increased as the system grows. Figure 17 illustrates this menu.

There are several graphical programs to display the information contained in the database. The first program gives a map of the geocoded network (Figure 18).

The second program draws the intersections in two ways, one with traffic markings and dimensions and the other with the existing traffic control signs. One of these options is shown in Figure 19.

Option A3 permits use of a program that demonstrates the coded traffic sign information in a stylized way on a line printer. Signs are not drawn but are represented by their code; their positions, nevertheless, are to scale. Figure 20 shows an example.

The explanations of the sign codes are printed beneath the drawing. Details of the program, which works in an interactive mode and can easily and rapidly access the data, are given by Gourvil *(9)*.

In traffic engineering, many programs exist for different tasks. For example, for traffic light coordination there are TRANSYT and PASSER and for volume data treatment there is COUNTS. These programs are generally prepared by companies outside the traffic engineering department; data entry for them is extremely rigid and time-consuming, diminishing their usefulness. However, the integrated database contains all the needed information for these programs except certain parameters that have to be entered interactively. There is a special interface program that automatically extracts the necessary data and asks the user to add other needed parameters. This process makes the data independent of the special application programs. The methodology of the data extraction is similar for different programs. Work is underway to design an efficient program that prepares data for application programs. Presently there are two interfaces, one for coordination of traffic lights on an artery (algorithm of Little, *10*) and one for the treatment of traffic volume data. The principle of use of a program interface is briefly described by an example of the traffic light coordination. Necessary data are parameters like progression speeds, cycle lengths, volumes, distances between intersections, and time splits at traffic lights. All data except the speed and cycle length, if the existing values are not used, are contained in the database. The user has to specify the name of the artery to be synchronized in a five-letter code and the beginning and ending intersection of the coordination. The interface program asks for cycle lengths and for speeds. Outputs can be directed to screen or to the printer.

These menus and underlying programs contribute to an easy and pleasant use of the information system, even if the present system is in the developmental phase. The following objectives are attained:

- Better use (analysis and extraction) of the collected data.
- Reduction of the traffic data collection work.
- Standardization of present working procedures.
- More frequent use of the application programs and easier program development.
- In the future, use of the information system for planning purposes and for decision making.

The successful implementation of such a computer-based information system depends on the participation and continuous consultation of the potential users during its development. The system has to adapt to the demands of the user (and not vice versa) and has to grow with the needs. One should not believe that a complete system can be developed in a single step. Given the modularity of the chosen DBMS that allows easy adding and changing of the data and data storage structure, the information system can, in a first step, be conceived globally, limiting data introduction to the presently available data and allowing a progressive enlargement of the database with future integration into other urban databases.

Stepwise introduction of such a system also alleviates certain psychological difficulties to the user created by the introduction of a global system that needs a large amount of data without producing an immediate beneficial and practical result.

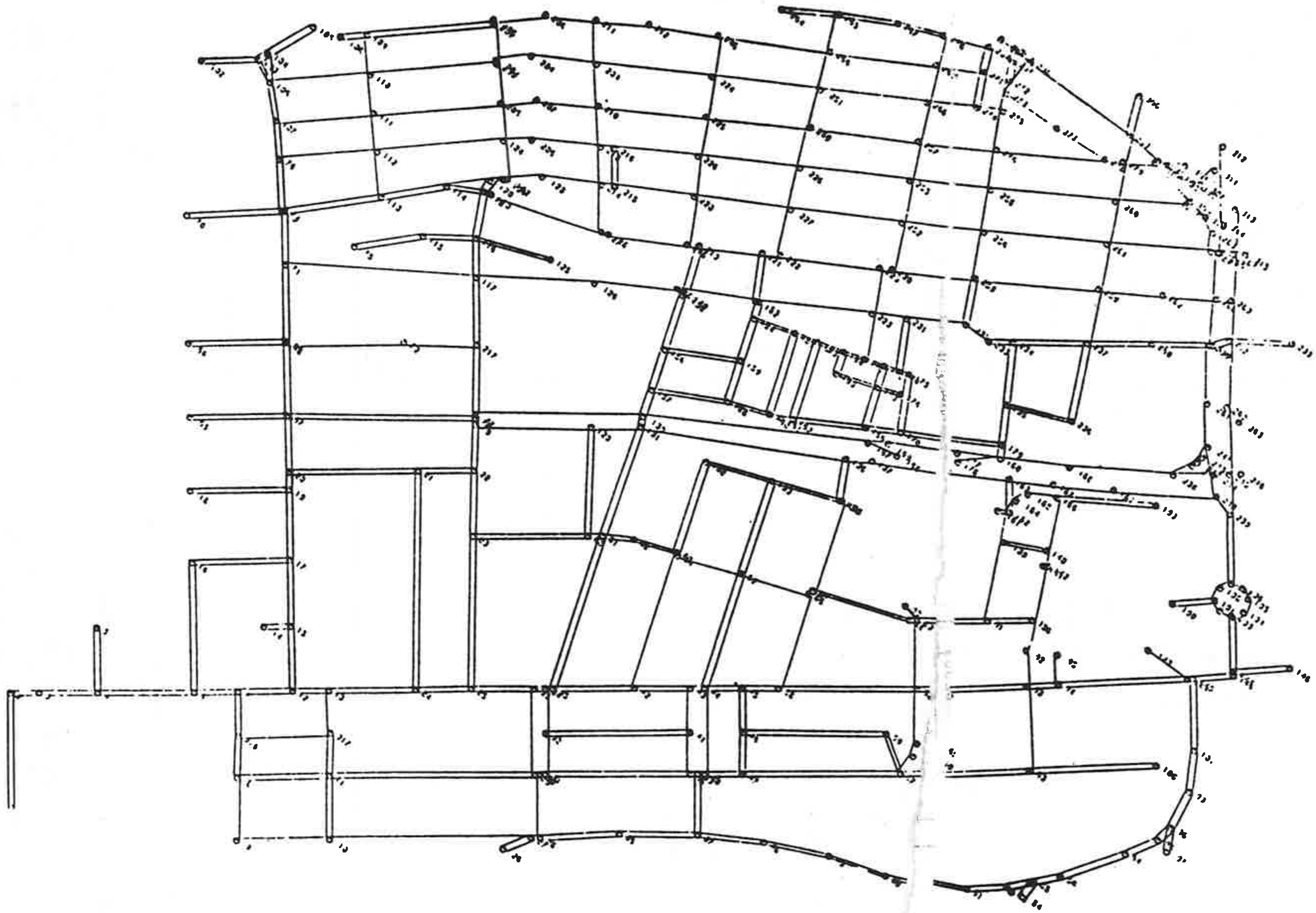


FIGURE 18 Drawing of the network used in the pilot study.

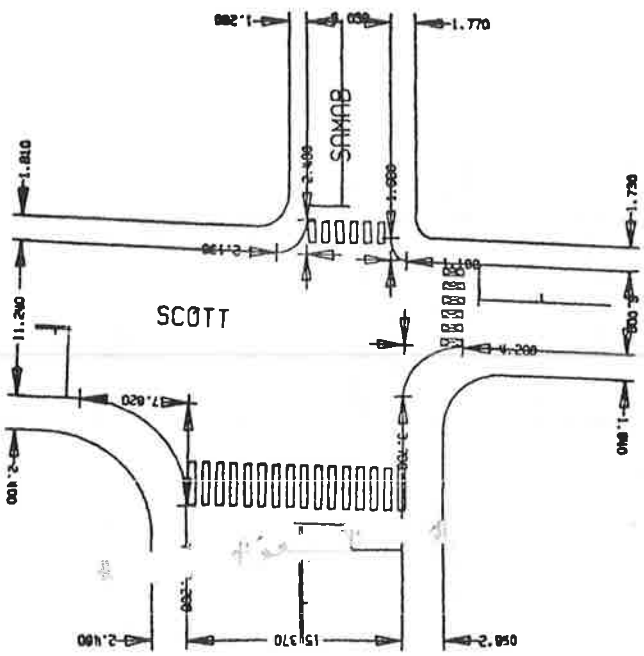


FIGURE 19 Intersection with pavement markings.

THE APPLICATION

The system was initially designed for small or medium-sized cites. It was applied to a sector of the city of Quebec with a network of 322 nodes and 290 geocoded links. The area of the sector was approximately 1.2 km^2; there were 2,040 traffic signs.

Data were collected from a traffic sector in Montreal for another application intended to prove the feasibility of the system in a larger city.

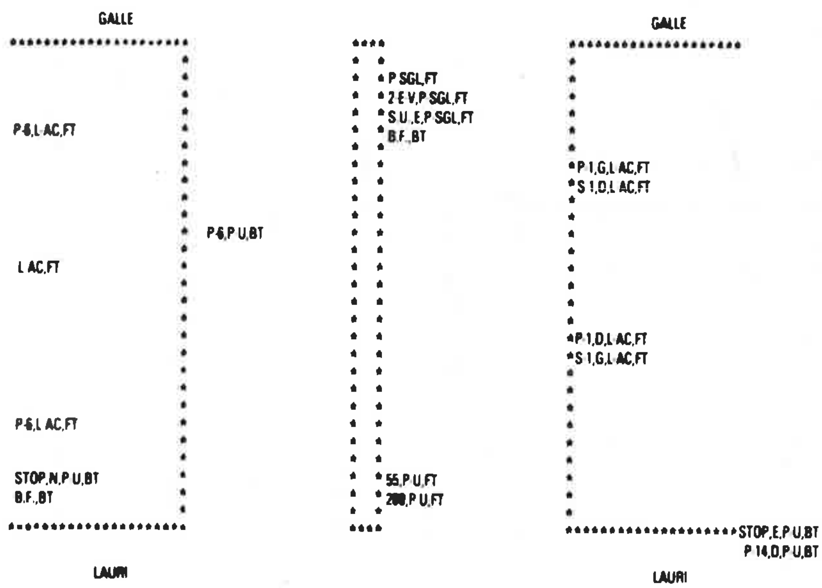
The information system resides on an IBM AT computer with color graphics and a 20-MB hard disk. Response time, even for the longest search through the file of traffic signs, is good.

CONCLUSION

The DBMS programs are in rapid and constant evolution and provide more and more useful commands and user-friendly information treatment. As microcomputers become more powerful and storage capacities increase constantly, the proposed approach becomes feasible also for medium- and larger-sized municipalities. The advantages of the proposed information system based on the well-known and widely distributed DBMS are numerous.

Data are not redundant and are easily kept up to date, and the access is controlled by password security. Data or combinations of data can be treated directly by means of available commands or can be used for application programs. Due to a common link, the information is accessible to all city departments, improving planning, operations, and decision making.

A global, centralized urban information system would be



PANNEAUX

ABBREV.	CODE	DESCRIPTION
STOP	1	ARRÊT/STOP 600X600
S.U.	50	SENS UNIQUE 900X300
	55	TERRE-PLEIN; CONTOURNER PAR LA DROITE 600X750
	290	OBSTACLE; ÉVITER PAR LA DROITE 300X900
P-1	901	STATIONNEMENT INTERDIT EN TOUT TEMPS
P-6	906	STATIONNEMENT INTERDIT DANS CETTE RUE
P-14	914	POSTE EN COMMUN - TAXIS
S-1	941	STATIONNEMENT PERMIS 10 MINUTES EN TOUT TEMPS

SUPPORTS

ABBREV.	CODE	DESCRIPTION
P-U,BT	1	POTEAU EN «U», BORD DU TROTTOIR
P-U,FT	11	POTEAU EN «U», FOND DU TROTTOIR
L-AC,FT	34	LAMPADAIRE D'ACIER, FOND DU TROTTOIR
P-SGL,FT	41	POTEAU DE SIGNAL LUMINEUX, FOND DU TROTTOIR
B.F.,BT	46	POTEAU D'INCENDIE, BORD DU TROTTOIR

FIGURE 20 Representation of the position of traffic signs on a line printer.

theoretically perfect. However, in practice it is difficult, time-consuming, and costly to develop from scratch. With the objective of integration in mind, stepwise development is preferable.

It is recommended that each municipal department develop a database at its rhythm, corresponding to its needs, and following a development plan that ensures a perfect compatibility of data coming from different departments. This recommendation would allow common use of the data, possible because of recent development of the local-area networks.

The pilot project has demonstrated the feasibility of the approach. Supplementary development in the field of interfaces between application programs and the information system will be carried out in order to improve the usefulness of the system and make it an efficient tool for economic decision making and providing aid for planning, development, and day-to-day operation of the traffic engineering system.

REFERENCES

1. C. J. Date. *An Introduction to Database Systems*. 3rd ed. Addison-Wesley, Reading, Mass., 1982, Vol. 1.
2. D. Kroenke. *Database Processing: Fundamentals, Design, Implementation,* 2nd ed. Science Research Associates Inc., Chicago, 1983.
3. D. Kruglinski. *Data Base Management Systems. A Guide to Microcomputer Software*. Osborne/McGraw-Hill, Berkeley, Calif., 1983.
4. *dBASE III PLUS The Data Management Standard.* Ashton-Tate, Torrance, Calif., 1985.
5. H. Nora. Méthodologie pour la planification des systèmes intègres d'information municipale; application au cas des villes nouvelles. *Proc., 4th European Symposium on Urban Data Management.* Planning and Transport Research Computation, Inc., Madrid, 1974.
6. J. Silver. *Special Report 187: Geographic Data Resources for Small and Medium-Sized Urban Transportation Planning Programs.* TRB, National Research Council, Washington, D.C., 1980, pp. 65–69.
7. B. Allard. *Système d'information municipal en circulation sur microordinateur.* M.A.Sc. thesis. Ecole Polytechnique de Montreal, Canada, 1985.
8. *QUICKCODE III. The dBASE III Program Generator.* Fox & Geller, Elmwood Park, N.J., 1984.
9. L. Gourvil. *Conception d'une banque de données de circulation comme sous-ensemble d'un système d'information urbain.* M.A.Sc. thesis. Ecole Polytechnique de Montreal, Canada, 1982.
10. J. D. C. Little et al. Synchronization of Traffic Signals for Maximal Bandwidth. In *Highway Research Record 118.* HRB, National Research Council, Washington, D.C., 1966, pp. 21–47.

Adequacy of the Sample Size and Accuracy of the Highway Performance Monitoring System for Use in Texas

JEFFERY L. MEMMOTT

In this paper, some of the characteristics of the Highway Performance Monitoring System (HPMS) sample in Texas, the validity of the recommended method to select the sample size for each state, a recommended method to correct the deficiencies in the sample size technique, and some comparisons of HPMS output with data from other sources are discussed. The current HPMS sample in Texas includes a rural area, small urban area, and 30 urbanized areas. Each area is sampled by functional class and average annual daily traffic (AADT) volume group within each functional class. Calculating the sample size within each volume group is not correct and generally results in samples being too large in the smaller AADT volume groups and too small in the larger groups. In the larger groups it is possible to have a calculated sample size of 1 or 2 no matter how many sections are in the group. A method is recommended to overcome the deficiencies of the current sample size procedure. The recommended technique involves sampling at the functional class level and distributing the sample to the AADT volume groups. The output data of the HPMS analysis programs are also compared to data from other sources. The HPMS estimate of Texas 20-year needs is compared to the 20-year project list for Texas, which is calculated independently of HPMS by district personnel in Texas. Even though the comparison is not complete, HPMS tends to estimate larger rural rehabilitation needs but smaller rural added-capacity needs than are contained in the 20-year list. In comparing HPMS output to performance measures such as fuel consumption and accidents, the estimates are reasonably close in most cases, indicating some level of confidence in the program assumptions and calculations.

The Highway Performance Monitoring System (HPMS) *(1)* was developed by FHWA to provide Congress and others timely and accurate information on the public highway system. This information covers the condition of the existing system as well as future anticipated needs and the effects if future funding does not cover those needs.

The HPMS program covers two major areas. The first is data on the highway system. A sample of highway sections is used to represent the entire system. Detailed data are collected by the states on these sample sections. This information and a small amount of data on all highway sections are sent to FHWA each year. Statistical methods are used to select the sample size, based upon functional category and average annual daily traffic (AADT) volume groups. A random selection process is used to select the samples, which are then maintained over time. A new sample is not taken each year, but the same sample is maintained with only additions or deletions made that are necessary to conform to the statistical procedure.

A package of computer programs to analyze the sample data has also been developed. The programs provide an analysis of the current or existing condition of the highway system and a number of options to look at future needs as well as impacts of different funding limitations. The basic procedure the computer packages use is first to estimate the current condition of the sample highway sections. Those conditions are then compared to minimum tolerable conditions tables. For those sections that have values less than those minimum values, an improvement is simulated. Both the type of improvement needed and the construction cost are estimated internally within the program. If a funding limitation is imposed, the program selects the highest-ranked needed improvements until the funds for that period are exhausted. The other improvements are deferred until the next funding period.

Some of the characteristics of the sample sections selected in Texas, a deficiency in the sample size procedure, and a recommended method to correct that deficiency are discussed. Some comparisons of the HPMS output are made in an attempt to determine how well the sample represents the entire Texas public highway system.

HPMS SAMPLE

Current Sampling Procedure

The sampling procedure FHWA recommends for the HPMS sample is relatively simple. All highway sections are first categorized into rural, small urban, or urbanized areas. The urbanized areas are handled either collectively or as individual areas. Currently, there are 30 designated urbanized areas in Texas that are sampled separately. Each area is then broken down into functional classes and then into AADT volume groups within each functional class. The objective is to get as close as possible to homogeneous groups of highway sections because the sample will represent all highway sections in a group. The 1983 sample size and mileage for Texas are presented in Table 1. The local functional class is not sampled in HPMS.

Each one of the volume groups within each functional class is sampled separately, with a minimum of three sample sections in each volume group. If the total number of sections is less

Texas Transportation Institute, The Texas A&M University System, College Station, Tex. 77843.

than three, then all sections should be sampled or combined with the next adjoining group. The recommended sample size for each AADT group is determined by the following formula taken from Appendix G of the HPMS Field Manual *(2)*:

$$n = F/\{1 + 1/[N(F - 1)]\}, \quad n \geq 3, \tag{1}$$

where

n = required sample size,
Z_α = value of the standard normal statistic for confidence level α (two-sided),
C = AADT coefficient of variation,
F = $(Z_\alpha C/d)^2$,
d = desired precision rate, and
N = universe or population stratum size.

FHWA has recommended values for both Z_α and d based on functional class with generally greater desired precision and confidence levels for higher functional classes. The critical parameter in the equation is the coefficient of variation. FHWA recommends that a coefficient of variation be calculated for each sampled group of sections and that the calculated value be used in Equation 1. FHWA also provides a table of coefficients of variation for states that do not or cannot calculate their own.

There is a fundamental problem with using the AADT coefficient of variation to estimate the sample size within each volume group. The coefficient of variation, calculated by dividing the standard deviation by the mean, is a measure of dispersion, in this case the dispersion of AADT within each volume group. The problem is that the range of AADT values is restricted. Therefore, the mean AADT within each group will be confined to that range. As sample sizes are calculated from larger-volume groups, the mean increases with correspondingly smaller coefficients of variation. A smaller coefficient of variation in Equation 1 results in a smaller required sample size.

An example of the impact of restricted volume groups on sample size is presented in Table 2. This example shows a hypothetical situation for rural Interstate highways but the conclusions apply to any functional class. It is assumed there are 100 and 1,000 sections in each volume group, and for simplicity sections are distributed uniformly within each group. (All assumed distributions produce similar results.) The impact of volume group is dramatic, with a required sample size of 78.5 or 265.3 in Volume Group 1 and only 1.2 in Volume Group 9. Everything else is the same among Volume Groups 1 to 9 except the AADT range of each group—Volume Group 1 ranges from 0 to 10,000, Volume Group 9 from 80,000 to 90,000. The sample should be about the same size or even larger in the larger-volume groups because there are more congestion and pavement deterioration.

The original purposes of dividing functional classes into volume groups for sampling purposes was to ensure some samples were being taken from the relatively small number of highway sections at the larger volumes. Although that is a worthwhile goal, forcing Equation 1 to do more than it was designed to do is not the answer. There is another problem at the larger-volume groups. With small coefficients of variation at larger volumes, the universe number of sections becomes insignificant in determining the sample size. In Table 2, the sample size converges for both the 100-section column and the 1,000-section column. For Volume Groups 7–9, the sample size is the same for both universe sizes.

Another way of looking at the same problem is the size of the confidence interval as compared to the volume group range. The confidence interval is defined as a percent of the universe mean in that volume group. For example, if a 90–5 precision is

TABLE 1 HPMS SAMPLE IN TEXAS (1983)

Functional Class	Rural				Small Urban				Urbanized			
	Sections		Mileage		Sections		Mileage		Sections		Mileage	
	Total	Sample	Total	Sample	Total	Sample	Total	Sample	Total	Sample	Total	Sample
Interstate	1,342	142	2,267	1,137	141	34	142	98	527	110	654	453
Other Freeway					44	24	53	48	581	140	590	406
Principal Arterial	7,929	370	8,069	2,843	3,335	231	981	300	3,852	531	2,479	995
Minor Arterial	7,557	132	6,994	1,153	4,071	93	1,227	70	5,245	501	3,675	639
Urban Collector Rural Major Collector	36,381	128	34,953	705	3,229	160	1,142	92	6,785	670	4,032	525
Rural Minor Collector	16,053	169	18,467	684								
Total	69,262	941	70,751	6,522	10,820	542	3,545	610	16,990	1,952	11,429	3,019

*Includes All Public Roads Except Local Functional Class.

specified the sample mean will be within ±5 percent of the universe mean 90 percent of the time. If a sample is drawn 100 times, the sample mean is expected to be within 5 percent of the universe mean 90 times. Confidence intervals are generally defined around the sample mean because the universe mean is not known. But, in this case, the universe mean AADT is generally known, so it is valid to define the confidence interval around the universe mean.

The problem is that the confidence interval tends to cover a larger portion of the volume group range at higher-volume groups; in some groups the confidence interval is actually wider than the volume range. Even though volume group ranges tend to increase somewhat as AADT increases, the mean AADT increases much faster, resulting in wider confidence intervals. One example can give an indication of the problem. The precision level for minor arterials in individually sampled urbanized areas is 70–15. The Houston urbanized area has 154 sections in Volume Group 5 of the minor arterial functional class, with a mean AADT of 17,003. The confidence interval is then 17,003(1 – 0.15) to 17,003(1 + 0.15), or 14,453 to 19,553. However, the AADT range of Volume Group 5 is 15,000 to 19,999. Only a small fraction of the volume group, from AADT 19,553 to 19,999, is not covered by the confidence interval, but none of the 154 sections are in this part of the range.

Because all sections are within the confidence interval, it would be impossible to select a sample with a sample mean outside the confidence interval, even if the sample size were 1. Because the precision criterion only requires that 70 percent of the sample means fall within the confidence interval, the required sample size using Equation 1 is less than 1, in this case $n = 0.2936$.

Just the opposite occurs at the smallest volume groups, with narrow confidence intervals in relation to the volume group range, and a corresponding increase in the required sample size. For example, in Volume Group 1 of Houston urbanized minor arterials there are 40 sections with a mean AADT of 1,426. The same 70–15 precision level applies, so the confidence interval is 1,212 to 1,640. The volume group range is 0 to 2,499, with the confidence interval covering only 15 percent of the range. As a result, the required sample size is relatively large ($n = 11.34$), even though only 70 percent of the sample means are required on average to fall within the confidence interval.

Proposed Sampling Procedure

It is clear from the analysis of sample size that it is not appropriate to use Equation 1 to determine the sample size within each volume group. However, the formula is valid for the entire range of AADT within each functional class. Equation 1 could be used to determine the sample for each functional class, for example, rural Interstate, rather than within each volume group. Calculating the sample size at the functional class level rather than at the volume group level tends to increase the sample size. The precision level could be adjusted to keep the overall sample size for each functional class approximately the same size as the current sample, but would probably not be advisable in the case of Texas because of the low sampling rates in some rural functional classes.

It should be noted that the current sample covers a higher percentage of the highway mileage than the highway sections, but the percentages are still low in both rural major collector and rural minor collector. In addition, the mileage percentage is higher because samples were extended to include parts of adjacent sections that exhibited similar characteristics. Covering a higher percentage of the highway mileage in this fashion does not necessarily improve the sample because the sample is chosen randomly. If enough samples are taken, the sample

TABLE 2 RANGE OF SAMPLE SIZE USING FHWA FORMULA

Volume Group	Range of ADT (Thous.)	Coefficient of Variation (Uniform Dist.)	Hypothetical Rural Interstate Sample Size: 100 Sections in Each Group	Hypothetical Rural Interstate Sample Size: 1,000 Sections in Each Group
1	0 - 10	.577	78.5	265.3
2	10 - 20	.192	28.8	38.6
3	20 - 30	.115	12.7	14.2
4	30 - 40	.082	6.9	7.3
5	40 - 50	.064	4.3	4.4
6	50 - 60	.052	2.9*	3.0
7	60 - 70	.044	2.1*	2.1*
8	70 - 80	.038	1.6*	1.6*
9	80 - 90	.034	1.2*	1.2*

*Since the formula would give a sample size of less than three in these cases, the minimum sample size of three would be used in actual application.

tends to represent the sections that are not in the sample. Extending some samples may bias the sample if some group of highways tends to be extended more than others. For example, if sample sections in West Texas tend to be extended more than samples in East Texas, the sample would be biased towards conditions and needs in West Texas.

If Equation 1 is used to calculate the sample size at the functional class level and the current precision rates are used, the required sample size for use in Texas increases. Table 3 presents the increased sample size needed to use HPMS on the state highway system at the state and district levels. It is not necessary to increase the samples on public highways off the state system and these are not included in Table 3.

There is another advantage in calculating the sample size at the functional class level rather than the volume group level. In Appendix I of the HPMS Field Manual *(2)*, sample size requirements for estimating proportions are discussed. The sample size must be large enough to detect a 10 percent change in proportions with 80 percent confidence. A formula and a graph are provided for estimating the required sample size. The problem is that the sample size for proportions is to be estimated at the functional class level, which makes it incompatible with Equation 1, which is currently used at the volume group level.

Under current procedures, Equation 1 is used to calculate a sample size for each volume group. The sample sizes are then summed to a total for each functional class. The sample size is then checked for the accuracy requirements for proportions. If additional samples are required, then the additional samples are distributed proportionately among the volume groups within the functional class.

If both procedures used the functional class level as the basis for estimating sample size, there would be no need to go through a multistep process; the one that gave the greater required sample size would be used. Equation 1 would be used if

$$F \geq 83.2(N - 1)/N \quad (2)$$

Otherwise the proportions formula, which assumes the required precision of ±10 percent with 80 percent confidence, would be used:

$$n_p = 83.2/(1 + 83.2/N) \quad (3)$$

where n_p equals the required sample size for estimating proportions.

After the sample size is determined for each functional class, the sample must be allocated to each volume group. There are several ways this allocation could be done. One simple way is to allocate the sample proportionately based upon the total sections in each volume group. Another technique is called the optimal allocation because it minimizes the variance for a given sample size. The weights for each volume group are the number of sections times the standard deviation. The problem with both of these techniques is that in some functional classes the number of sections in each volume group varies dramatically. For example, in rural major collectors there are 30,021 sections in Volume Group 1; 6,331 sections in Volume Groups 2, 3, and 4; and only 29 sections in Volume Groups 5, 7, and 8. This difference results in small sample sizes for larger-volume groups, even though requiring a minimum sample size of three does reduce the problem somewhat. For these situations of unequal numbers in different volume groups, the allocation could be structured so that the larger-volume groups receive a larger representation in the sample, which can be accomplished by distributing the sample over the volume groups weighted by vehicle-miles.

$$n_{ij} = n_j(DVM_{ij}/DVM_j), \quad n_{ij} \geq 3, \quad (4)$$

where

n_{ij} = required sample size for Volume Group i in Functional Class j;
n_j = required sample size for Functional Class j, calculated from Equation 1;
DVM_{ij} = daily volume in vehicle-miles for Volume Group i in Functional Class j; and
DVM_j = total daily volume in vehicle-miles for Functional Class j.

TABLE 3 RECOMMENDED CHANGE IN SAMPLE SIZE FOR USE IN TEXAS

	Rural	Small Urban	Urbanized	Total
Total Sections State System	64,852	6,373	5,323	76,548
Current Sample (1983)	913	328	715	1,956
Recommended Sample for State-wide Use	1,840	643	1,182	3,665
Recommended Sample for District Use	5,652	1,271	1,358	8,281

Each of the procedures for allocating the sample to volume groups has certain advantages and has to be studied further. Some combination of the techniques may be most useful.

Limitations of Sample

Whereas the recommended changes to sample size calculations will reduce or eliminate some problems with the current procedures, they will not eliminate all problems with samples in general or with this particular sample. A sample selected to represent the entire universe of highway sections cannot represent it perfectly. There is some error any time a sample is used. In addition, a sample cannot be used to describe individual sections outside the sample. For example, the sample cannot be used to pinpoint all sections that need improvements. A 100 percent sample or an inventory of the highway section would be required for that purpose.

Another aspect of this particular sampling procedure is that it is only sampling AADT. For example, a desired precision of ±5 percent with 90 percent confidence as used in Equation 1 applies only to how good the sample AADT mean is as a measure of the population AADT mean. In this case, the sample AADT mean would be within ±5 percent of the population AADT mean with 90 percent confidence. That question by itself is usually trivial because the population AADT mean is usually already known. The population AADT mean is used to calculate the coefficient of variation, which in turn is used to calculate the sample size. The assumption is that AADT is a good predictor of items that are not known in the population, such as pavement condition and future anticipated congestion. In the next section, some comparisons of the HPMS output that have some implications for the accuracy of the Texas sample are examined.

HPMS OUTPUT

Comparison to 20-Year Plan

As mentioned in the introduction, HPMS consists of a package of computer programs to analyze the sample data collected in each state. Table 4 presents a summary of the output for Texas using the default assumptions and parameters. In order to adjust a number of these parameters for Texas conditions, data are being collected.

Table 4 also uses a type of analysis that allows for analysis over four 5-year periods. Improvements are simulated for each period on each sample section with a deficiency during that period. That also allows for more than one improvement on a particular section. It would be possible, though unlikely, for an improvement to be simulated on a particular section in each 5-year period. Table 4 also assumes no funding or right-of-way restrictions, so it represents the total 20-year needs of Texas as represented by the sample and the assumptions of the model. The calculated needs are very large over the 20-year period,

TABLE 4 HPMS OUTPUT WITH FOUR FUNDING CATEGORIES ON TEXAS STATE HIGHWAY SYSTEM

Rural

Highway Type	1985–1989		1990–1994		1995–1999		2000–2004		Total	
	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost
Added Capacity	659	1,353.2	427	833.1	931	1,730.6	733	1,345.0	2,750	5,261.9
Upgrade to Standard	15,809	3,282.6	9,030	1,224.6	3,001	472.8	2,213	329.6	30,053	5,309.6
Rehabilitation	5,879	4,701.6	5,727	3,925.3	4,687	3,166.9	2,987	1,794.8	19,280	13,588.6
Resurfacing and Traffic Engineering	15,140	1,400.6	11,531	820.2	18,051	1,521.7	21,893	1,467.2	66,615	5,209.7
Total	37,487	10,738.2	26,716	6,803.2	26,672	6,892.0	27,825	4,936.5	118,700	29,369.9

Urban

Highway Type	1985–1989		1990–1994		1995–1999		2000–2004		Total	
	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost
Added Capacity	1,077	17,501.8	244	3,222.5	490	11,474.7	399	9,897.1	2,210	42,096.1
Upgrade to Standard	424	277.0	204	94.7	55	32.7	8	5.3	691	409.7
Rehabilitation	222	351.2	212	302.8	222	232.3	5	8.3	661	885.6
Resurfacing and Traffic Engineering	1,865	1,023.1	1,509	1,086.7	1,975	1,376.4	2,019	1,554.9	7,368	5,041.1
Total	3,588	19,153.1	2,170	4,706.8	2,742	13,107.3	2,432	11,465.5	10,932	48,432.7

about $30 billion in rural areas and $48 billion in urban areas. It is also interesting to note how the amounts change over time. The first 5 years are the largest, indicating a backlog of current needs of $11 billion in rural areas and $19 billion in urban areas.

Table 5 presents a comparison of the HPMS output to the 20-year project list for Texas. This 20-year list does not include any funding restrictions. HPMS does not estimate new location construction needs, and the 20-year plan does not include maintenance activities such as resurfacing, so they are not directly comparable, but for some categories of projects some comparisons can be made. In the 20-year HPMS output comparison, HPMS is predicting much larger rehabilitation needs (specially in rural areas), larger added-capacity costs over less mileage, and somewhat smaller upgrade-to-standards costs over more mileage.

The biggest discrepancy in construction costs comes in the urban added-capacity category. The cost per project is more than three times greater in HPMS than in the 20-year plan. Even though different design standards and traffic growth could be responsible for some of the differences, they would probably not be sufficient to explain such a large difference. This would indicate a need to examine the assumed construction costs in HPMS and to revise them to more closely reflect Texas costs. The biggest difference in mileage comes in rural estimates, with HPMS predicting larger mileage needs for rehabilitation and upgrade to standards, and less for added capacity. One of the reasons for larger mileage needs in HPMS is because up to four improvements can be simulated on each section over the 20-year period, whereas the 20-year plan includes relatively little staging and is restricted to added-capacity stages of construction. Another reason may be the way projects are

TABLE 5 COMPARISON OF HPMS 20-YEAR IMPROVEMENT ESTIMATES WITH 20-YEAR PROJECT LIST FOR TEXAS

	Rural				Urban				Total			
	HPMS		20 Year Plan		HPMS		20 Year Plan		HPMS		20 Year Plan	
	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost
Added Capacity	2,950	5,261.9	5,805	8,737.2	2,210	42,096.1	3,224	18,184.3	4,960	47,358.0	9,028	26,921.5
New Location			4,984	3,937.9			738	3,944.8			5,611	7,882.7
Upgrade to Standard	30,053	5,309.6	16,021	5,331.4	691	409.7	1,652	1,484.6	30,744	5,719.3	17,672	6,816.1
Rehabilitation	19,280	13,588.6	4,337	1,337.1	661	885.6	808	373.6	19,941	14,474.2	5,144	1,710.7
Resurfacing and Traffic Engineering	66,615	5,209.7			7,368	5,041.1			73,983	10,250.8		
Total	118,700	29,369.9	31,036	19,343.6	10,932	48,437.7	6,421	23,987.3	129,632	77,802.6	37,457	43,330.9

TABLE 6 COMPARISON OF HPMS 5-YEAR IMPROVEMENT ESTIMATES WITH 20-YEAR PROJECT LIST FOR TEXAS

	Rural				Urban				Total			
	HPMS		20 Year Plan		HPMS		20 Year Plan		HPMS		20 Year Plan	
	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost	Miles	Cost
Added Capacity	659	1,353.2	5,805	8,737.2	1,077	17,501.8	3,224	18,184.3	1,736	18,855.0	9,028	26,921.5
New Location			4,984	3,937.9			738	3,944.8			5,611	7,882.7
Upgrade to Standard	15,809	3,282.6	16,021	5,331.4	424	277.0	1,652	1,484.6	16,233	3,559.6	17,672	6,816.1
Rehabilitation	5,879	4,701.6	4,337	1,337.1	222	351.2	808	373.6	6,101	5,052.8	5,144	1,710.7
Resurfacing and Traffic Engineering	15,140	1,400.6			1,865	1,023.1			17,005	2,423.7		
Total	37,487	10,738.2	31,036	19,343.6	3,588	19,153.1	6,421	23,987.3	41,075	29,891.3	37,457	43,330.9

developed. For example, if a highway needs upgrade to standards or rehabilitation, and if there is a chance added capacity may be required in the future, it may be included in the proposed project even if the added capacity by itself is only marginal.

A better comparison may be possible from the data presented in the lower portion of Table 5. This compares the first 5 years of HPMS output to the 20-year plan. The reason this may be a more valid comparison is because the 20-year plan tends to concentrate on current needs or anticipated needs in the near future. In rural areas, HPMS is predicting much smaller added-capacity needs and higher rehabilitation needs, with upgrade to standards almost the same. The urban comparisons are all similar, with added capacity and upgrade to standards showing the largest difference.

With adjustments to the assumptions and parameters, HPMS appears to have potential for being used to estimate current and future highway needs in Texas. Eventual discrepancies have to be evaluated critically to determine if some correctable systematic error is being introduced.

Output Evaluations

Some of the output from the HPMS analysis can be checked against other sources to determine how well those values are being calculated within the program and how well the sample sections represent the universe of highway sections in Texas. These comparisons are presented in Table 6.

As can be seen in Table 7, vehicle-miles and injury accidents are being predicted almost exactly, but larger errors occur in fuel consumption, property damage accidents, and fatal accidents. The output appears to be doing a reasonably good job of calculating these values, and even though HPMS would probably not be used to estimate these numbers, it indicates that overall the sample appears to be representative of the entire highway network.

CONCLUSION

The HPMS sample data and analysis package was designed to provide pertinent information on the current status of the highway system and estimates of future needs. It represents a

TABLE 7 COMPARISON OF HPMS OUTPUT IN TEXAS

	1983 Value[1]	1983 HPMS Estimate[2]
Fuel Consumption (Gallons of Fuel per 1000 vehicle-miles)	73.1	99.6
Total Highway Fuel Consumption (Millions of Gallons)	7,953.3	10,537.7
Vehicle-Miles (Billions)	108.8	105.8
Accident Rate (per 100 million vehicle-miles)		
Property Damage	234.2	326.7
Injury Accidents	106.5	105.5
Fatal Accidents	2.6	3.6
Total	343.3	435.8

[1] Fuel consumption figures and vehicle miles taken from Highway Statistics, 1983 (3), and adjusted to exclude local functional class. Accident rates taken from Motor Vehicle Traffic Accidents, 1983 (4), and includes statewide accidents divided by total vehicle miles including local functional class.

[2] HPMS samples cover all functional classes except local functional class and are expanded to represent all state highway sections excluding local functional class.

significant improvement over previous attempts to measure highway needs and should be a valuable tool in future years for highway-related policy planning.

There is, however, an error in the use of the sample size formula at the volume group level within each functional class. The problem is masked because there tend to be far fewer highway sections in higher-volume groups, so the bias of the formula in calculating higher sampling rates for lower-volume groups is not readily apparent.

At the national level, the bias in sampling rates is probably small compared to the size of the sample and may not be causing distortions in the estimates, but at the state or substate level it could have some impact. In addition, the problem will be increasing in the future as traffic volumes increase and more highway sections move into larger-volume groups. The error should be corrected before it becomes a significant problem.

The comparison of HPMS output with data from other sources in Texas gives some degree of confidence in the output and a good base to build upon. It is evident, however, that a larger sample is needed for use in Texas, specially in the rural areas. Adjustments are also needed in the design standards, minimum tolerable conditions, and construction costs assumed in the model.

There are several changes currently being made to the HPMS analytical package by FHWA that should improve the accuracy of the estimates and the usefulness of the data. These changes include improved pavement deterioration curves, more flexibility in the summary output, and the possibility of using benefit-cost criteria in selecting improvements with a budget constraint. It is hoped that efforts to improve the model will continue in the future.

ACKNOWLEDGMENT

This paper was prepared in cooperation with the Texas Department of Highways and Public Transportation and the FHWA. The contents of this paper reflect the views of the author, who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views and policies of the Texas Department of Highways and Public Transportation or the FHWA.

REFERENCES

1. *Highway Performance Monitoring System Analytical Process. Volume 2: Technical Manual.* Office of Highway Planning, FHWA, U.S. Department of Transportation, March 1983.
2. *Highway Performance Monitoring System Field Manual.* Office of Highway Planning, FHWA, U.S. Department of Transportation, Jan. 1984.
3. *Highway Statistics, 1983.* FHWA, U.S. Department of Transportation, 1983.
4. *Motor Vehicle Traffic Accidents, 1983.* Texas Department of Public Safety, Austin, 1983.

Direct Data Entry Using Microcomputers—A Travel Survey Pilot Project

GERALD N. STEUART, BRUCE MORI, AND PETER J. NOEHAMMER

One aspect tested in the pilot project for a Toronto region origin-destination travel survey was direct data entry using microcomputers. Four interviewers utilized a computer program that performed (a) file management of the sample, and (b) data prompting for the interviewer. Twelve interviewers used conventional pencil and paper forms. A comparison of the two methods proved the automated process just as expedient as the manual method during actual telephone interviewing, but far superior during sample processing. Large potential benefits from direct data entry are possible after interviewing is complete as paper forms must further be entered onto a computer by keypunch. The costs associated with providing each interviewer with a microcomputer and program balance the costs of keypunching paper forms. No extra cost or risk over reliable manual methods is evident for direct data entry to automate the survey process. Furthermore, productivity gains during interviewing provide the potential for direct data entry to surpass conventional manual methods in terms of cost-effectiveness.

Transportation planning in virtually every community in North America has been formulated on the basis of one, and usually only one, large-scale and comprehensive inventory of urban travel. Most of these studies were carried out in the 1960s using home interviews as the data collection technique. Information regarding all trips made by each member of a household during a specified 24-hr period was recorded. The resulting travel data have become the foundation on which much of transportation planning is based.

Toronto is typical of this historical development. The first and only comprehensive inventory of urban travel in the Toronto region was the Metropolitan Toronto and Region Transportation Study (MTARTS) Home Interview Survey, which was carried out in 1964. This information has served as the basis for most of the large-scale transportation planning efforts since that time. After approximately 20 years, the information is of questionable value as a basis for estimating current urban travel. The data will always be useful for looking at time trends; however, to serve this purpose new origin-destination data must be collected.

All transportation planning agencies in the Toronto region have realized the inadequacies of 20-year-old information and the limitations of incomplete information. They have formed a committee to promote a coordinated effort to improve the situation. The Toronto Area Transportation Data Collection Steering Committee has the objective of determining the best collective course of action to improve the quality of travel data in the region. The committee has recommended a comprehensive origin-destination (O-D) travel survey for a universe of approximately 4 million people encompassing six regional municipalities in the Toronto area.

The Ontario Ministry of Transportation and Communications financed a pilot project *(1)* in March 1986 aimed at testing several proposed features of a full-scale study. Primary features tested in the pilot project were a proposed set of questions to obtain pertinent household, person, and travel information; a procedure for informing the sample of an impending interview; a test of the sampling procedure; and most important from the standpoint of this paper a test was carried out in which some of the interviewers used microcomputers to enter travel data directly to computer files. Interviews were conducted March 4–9 and 18–20, 1986.

DIRECT DATA ENTRY

The objective of the microcomputer experiment was to verify that direct data entry could be used in a travel survey. In this paper, the findings of an experiment in which microcomputers were used by an interviewer to enter travel information directly to a computer file while a home interview was taking place over the phone, that is, direct data entry, are reported. Observations and conclusions are aimed specifically at the feasibility of using direct data entry for a full-scale O-D study.

The data collection process in a large-scale O-D survey, whether performed by direct data entry or by more conventional manual methods using paper forms, contains three distinct phases:

1. Sample processing. Assemble a sample of the universe of households and continuously update the status of each record in the sample;
2. Interviewing. Select a record from the sample, prompt an interviewer with questions to be asked, and record the data;
3. Data processing. Check the answers received for validity, assign a numerical code for spatial locations, and record all information in computer files.

These phases are closely related, as the output of one serves as input to the next. O-D surveys in the past have done all record keeping and data recording on paper forms. This method requires an extensive filing system for managing the paper forms, both completed and to be completed.

Conceptually, the entire process could be automated on a computer system with a network of terminals on either a

Department of Civil Engineering, University of Toronto, Toronto, Ontario, M5S 1A4, Canada.

mainframe or minicomputer. However, the process must operate without interruption. Unfortunately, the development of a large-scale computer system with multiple on-line users, including networking of terminals, is usually subject to a breaking-in period during which the system is unreliable. It is required that the system development costs be rationalized for one short application. This process normally implies the use of existing software packages.

A better strategy is to use the computing device to aid or improve already proven techniques without taking the risks associated with developing a complex computer network in a short period of time. The most common and most widely used software packages for data management are available for microcomputers. An added incentive for developing microcomputer procedures is the potential for use in other cities. Other O-D studies have a better potential for access to compatible microcomputers than identical minicomputers or mainframe computers with compatible operating systems.

The rate at which interviews are conducted can be adversely affected by computer system failures and operator errors. Thus, contingency plans and error recovery strategies are important components of this project. In addition, the computer hardware or software should never dictate the pace of the interview. The computer program should not jeopardize or hinder the rate at which an interviewer can process the sample. The computer screen should merely provide the electronic counterpart of a paper form, and the program should handle sample updating and database management tasks normally done manually by filing forms.

During the pilot project, all aspects of the actual interviewing were the responsibility of a market research firm. This included the hiring, training, and supervising of all interviewers. Approximately 20 interviewers were trained for the project and 5 were selected for direct data entry. Four of these were regular interviewers, each one using the same computer and sample every session, while the fifth became the supervisor. Instruction on the use of the computer was carried out by the authors of this paper. In addition, project members were available in person early in the project to detect problems. On the basis of these observations and input from the interviewers, several minor modifications were made to the computer program during the course of the pilot project.

The computer program was written in a popular database language available for microcomputers operating under MS–DOS. The program was then compiled into an executable program for distribution. This step allowed the necessary features of the database language to be used without having to purchase a program license for each computer, and did not infringe on copyright regulations. The program was configured for an IBM-compatible personal computer with at least 256 kilobytes of random access memory, a monochrome monitor, and two 360-megabyte floppy disk drives. The interviewer was given two disks every evening. One disk contained the program and the interviewer's sample file, with a record of all attempts to contact each household. A record was ignored after five unsuccessful contact attempts. The second disk recorded all the data collected during the course of interviewing and was replaced every evening. An update of the global database could be made every day.

SAMPLE PROCESSING

The first phase in an O-D survey is to assemble a sample of the universe of households in the study area, assuming the study area has been defined and a sample size selected. The sample should be obtained in the form of a data tape and loaded onto the appropriate file management system in random order with sequential sample numbers. The first task of file management is to provide input to the printing of preinterview letters. It is possible to automatically print for mailing out an appropriate number of preinterview letters containing a current date once the sample is set up in a computer file.

The pilot project was conducted on a sample of households from eight postal districts consisting of approximately 5,000 records. This sample was selected sequentially from the current telephone directory by Tele-Direct, a subsidiary of Bell Telephone, and prepared on magnetic tape. Every fourth record was copied to a separate computer file, which became the sample for the direct data entry experiment. The remaining records were to be processed using conventional paper forms.

The sample was divided into two sets. Preprinted letters from one set were sent out on Friday and Saturday, March 1 and 2, 1986. These letters notified households of interviews to be conducted March 4–9, inclusive, for travel on March 3–7, inclusive. Interviews were not conducted during the following week, which coincided with the public school's March break. The second set of letters was to have been mailed out on March 11; however, a sufficient sample was available from the first mailing. As a result, the second mailing for the remaining 3 days of interviewing was never carried out. Some adjustments to the sample allocation were necessary because direct data entry processed records faster, thereby leaving insufficient sample for three remaining evenings of interviewing. One in every four records remaining in the paper form sample was transferred to the direct data entry sample. The expanded sample proved advantageous because the additional 220 transferred records could be distributed according to the needs of each interviewer. Although the interviewers were initially assigned the same number of records, each processed at a different rate.

There is no reason to believe that direct data entry will decrease the length of time an interviewer spends actually conducting interviews on the telephone. Productivity gains, if there are to be any in direct data entry, are likely to come from efficiently handling the records in the sample and automatically sequencing the calls for the interviewer. For this reason, considerable effort was spent on file management by the program developers.

Two aspects of automatic file management of the sample records improved interviewer productivity and improved the likelihood of making contact with the respondent. Strategies for attempting repeated contacts were automated for "no answer" and "busy" situations. "Call backs" were used to accommodate respondents at a more convenient time, or to obtain complete information from a member of the household who was not at home. For a "call back" situation, the record and an optional message were automatically redisplayed for the interviewer at a prespecified combination of time and day when interviewing was taking place. "No answers" would automatically reappear on the screen, either the same evening or the next evening, after a specified period of elapsed time. Noninterview time consumed 66 percent of the interviewers' time for the conventional paper entry mode of Operation 1. File management by computer has the potential to reduce this time, and reduce the average time per completed interview by as much as 2 min. Furthermore, the interviewer would spend more time interviewing and less time handling forms.

Management of the sample files began with the preparation of the database file containing the sample that included a series of fields from the information provided by Tele-Direct:

- Respondent's name
- Street address (and apartment number, if applicable)
- Municipality
- Postal code
- Telephone number

Each record was then completed by adding the following fields:

- Sample number
- Five repetitions of three fields (initially blank or zero)
  - Data of response
  - Time of response
  - Response code
- Interview completion time

An evening's interviewing begins with the program's searching the sample file for the first record that has not been completed. It subsequently displays the information on the screen in the form shown by Figure 1. The left-hand side of the form contains a display of the information for all the fields of that record. As such, the form displays a complete history of any previous action. The upper right-hand side of the screen displays the current date and time taken from the computer's internal clock. The box on the lower right contains a reminder of all the valid actions the interviewer can take on the sample by depressing the appropriate key. These actions, or responses, were taken only after the interviewer had dialed the telephone number displayed.

One additional precautionary step was taken in the pilot project. The sample number, respondent's name and telephone number, and the response were written down on paper for every telephone call made by the interviewer. This procedure was done primarily for supervisory staff records, but also served as a precautionary measure in the event of a computer malfunction.

Response 1 was assumed to be the most common response and given the first position. This response was designated if the respondent confirmed that the name and address were correct and that they were willing to be interviewed. The record was then internally marked in a manner to prevent it from reappearing on the screen.

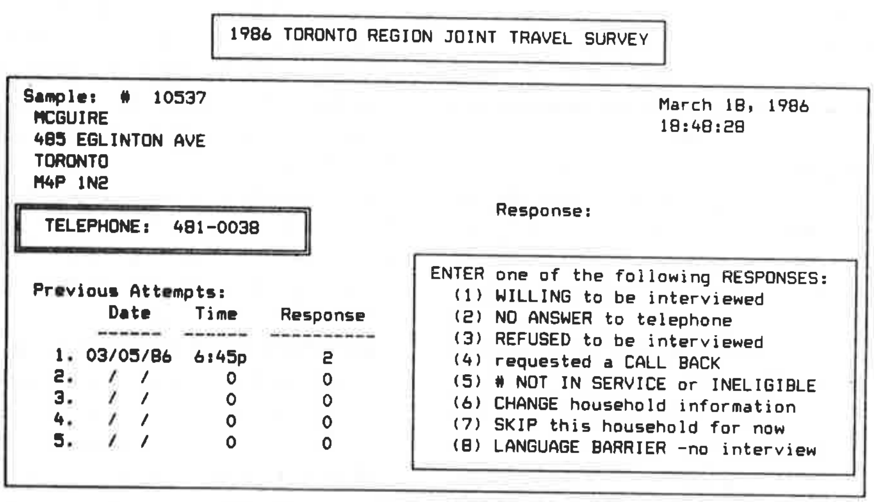
Response 2 was designated if the telephone was not answered after an appropriate number of rings. The final version of the program, used in the last few days, attempted to patch a strategy onto the existing program structure that would redisplay "no answer" records 2 hr later.

Response 3 was designated if the respondent, for whatever reason, refused to be interviewed. The record was then internally marked in a manner to prevent it from appearing on the screen again.

Response 4 turned out to be a useful device enabling the sample record to reappear on the screen at a designated time. A comment field, which recorded information to be presented on the screen to remind the interviewer of details regarding the call back, was also incorporated at the request of the interviewers. The procedure operated as follows when Response 4 was selected. A new screen appeared asking the interviewer for a date and time of the call back. A default value of the current date was displayed and could be overwritten. The interviewer was then given the option of entering three short lines of comments. The screen then displayed the next sample record to be interviewed. The record registering a call back came back on the screen at the appropriate time, as the program searched the call back file for dates and times.

Response 5 was designated if the operator came on the line to advise that the number was not in service, the number represented a business rather than residential phone, or the respondent had moved out of the study area but kept the same telephone number. The record was then internally marked in a manner to prevent it from reappearing on the screen.

Response 6 allowed the interviewer to change the informa-


```
1986 TORONTO REGION JOINT TRAVEL SURVEY

Sample:  #  10537                                   March 18, 1986
 MCGUIRE                                            18:48:28
 485 EGLINTON AVE
 TORONTO
 M4P 1N2
                                               Response:
 TELEPHONE:  481-0038

                                      ENTER one of the following RESPONSES:
 Previous Attempts:                     (1) WILLING to be interviewed
          Date      Time    Response    (2) NO ANSWER to telephone
        -------   ------   ----------   (3) REFUSED to be interviewed
  1. 03/05/86    6:45p        2         (4) requested a CALL BACK
  2.   /  /        0          0         (5) # NOT IN SERVICE or INELIGIBLE
  3.   /  /        0          0         (6) CHANGE household information
  4.   /  /        0          0         (7) SKIP this household for now
  5.   /  /        0          0         (8) LANGUAGE BARRIER -no interview
```

FIGURE 1 Direct data entry—Screen 1 format.

tion in the fields describing the name and address of the respondent. This action was important because the home address was stored in memory and became part of the trip record. When the editing was complete, the original screen returned and the interviewer could select the appropriate response to proceed.

Response 7 allowed the interviewer to skip the current sample record. The record would not appear until the computer was turned off and the start-up procedure carried out.

Response 8 was designated if the interviewer encountered a language that was indeterminable, or if the pool of interviewers did not include anyone with sufficient knowledge of the language to conduct an interview. The record was then internally marked in a manner to prevent it from appearing on the screen again.

Every action an interviewer took on a record in the sample was timed to the nearest minute by the computer's internal clock and recorded in the sample file. The recorded data showed file management to be important. Of the 142 hr that interviewers spent processing sample records, just less than 67 hr were spent entering trip records on the basis of information gathered over the telephone. The remaining 75 hr were spent attempting calls, encouraging interviews, and doing other file management tasks.

INTERVIEWING

Computer specialists immediately visualize a serial process whereby the interviewer, when telephone contact is established, would be presented on a visual display terminal with the name of the potential respondent, then with a series of precisely worded questions each followed by a prompt for data entry. Each entry would be checked for validity, then either rejected and reprocessed or stored in a database file. Unfortunately, travel survey interviews seldom progress sequentially as answers are provided in a conversational form as opposed to a rigidly structured form. The interviewer is used to, and benefits from, seeing the complete household's travel behavior rather than simply a portion of a single trip. Travel survey interviewers have always been accustomed to seeing the evolution of a household's complete trip-making behavior on a single form. This procedure provides information about duplicate trips by different members of the household and allows the interviewer to recognize previous errors and to move directly to correct them. Interviewers are also better able to encourage information and help the respondent along because they are fully aware of which trips have been made. Should a series of conventional, sequential forms be used, many errors would unfortunately go unnoticed, and much would be left to the memory of the interviewer to go back to the intended form.

This approach, however, requires that interviewers be conversant with the fundamentals of microcomputer operation and be able to type about as quickly as they can write. Experienced interviewers in market research have traditionally used a pencil and paper. Fortunately, there are two compensating forces at work. First, interviews are conducted in the evening, allowing for a larger selection of experienced interviewers, some of whom will be able to type. Second, there is a growing interest in the field of market research for direct entry of information at a computer terminal. This fact implies that more experienced interviewers will be able to use direct entry and that many others are interested in developing the skill.

Another implication is the need for an interviewer to be able to move freely and quickly from field to field on the interview form. The interview form will be large and complex by database management standards and is not of fixed length. The length of the form will depend on the number of persons and the number of trips made by each person. Most database management software allows the operator to move sequentially from field to field; however, moving on the screen to another section of a large and complex form of unspecified length may be time-consuming. Training an interviewer to move the cursor about on the screen with the use of key strokes and to bring up a new screen when required must be accomplished in training sessions.

The screens designed for this project proceed in the following manner. When an interviewer selects Response 1, indicating a respondent is willing to be interviewed, a new screen appears on the monitor as shown in Figure 2. The screen represents two database files. The box on the left contains all information to be collected about the household (the number of persons and number of automobiles). The screen prompts for this information and when entered stores the information in the household database with the sample number as an identifier.

The screen then displays, one line at a time in the right-hand box, a request for information about each person in the household. The computer automatically numbers every person and requests a name or other identifier for the interviewer to refer to that person (free format, eight characters maximum), approximate age (two-character numeric with 99 indicating refusal), sex (M or F), employment status (P part time; F full time; S student; H home; U unemployed), and driver's license (Y or N). A 9 entered in any field denotes a nonresponse or unknown information. The interviewer is able to edit any of the information about any person while in this box; however, the interviewer cannot change any information in the previous box. Each field is checked for valid information according to the indicated allowable codes. When the interviewer enters information for the last person correctly, all information remains on the screen and a request for trip information is presented as shown in Figure 3.

The first line of trip entry begins with a numeric identifier for the trip, which is automatically assigned by the computer, followed by a numeric field for person number. The origin of the first trip taken by the person in question is recorded in a

```
Sample No. 10541    Tel.:474-0156                          Surname: TURNER
              #  Name     Age S Em DL  #  Name  Age S Em DL  #  Name  Age S Em DL
#Peopl  2  | 1 FRANK     35 M  F  Y |                     |
#Autos  2  | 2 MARY      32 F  H  N |                     |
```

FIGURE 2 Direct data entry—Screen 2 format.

```
Sample No. 10541     Tel.:474-0156                              Surname: TURNER
              #--Name--Age-S-Em-DL--#--Name--Age-S-Em-DL--#--Name--Age-S-Em-DL--
#Peopl  2  |1 FRANK    35 M  F  Y |
#Autos  2  |2 MARY     32 F  H  N |

--#--------ORIGIN---------------DESTINATION----------P-TIME--MDES-TRANSFER-RTIME-
  1 123 ROEHAMPTON CT AGIN  EATON CENTRE TORONTO    1  920a D
  1 EATON CENTRE TORONTO    CITY HALL TORONTO       1  130p D
  1 CITY HALL TORONTO       123 ROEHAMPTON CT AGIN  3  500p D
 MD1[D][            ] MD2[ ][          ] MD3[ ][          ] MD4[ ][          ]
```

FIGURE 3 Direct data entry—Screen 2 format, including trip records.

free format field of 22 characters maximum. The address information of the home location was previously stored in computer memory and can be recalled anytime by the interviewer by simply selecting Function Key F9 on a standard keyboard. The screen then prompts for the destination of the trip, which is also entered into a free format field of 22 characters maximum. The program automatically stores this field in computer memory and defines Function Key F10. This action facilitates the use of F10 on the next trip to recall the previous destination, as it was often the origin of the next trip. The screen then prompts for trip purpose, which is a numeric code selected from a table presented on a card and usually placed on top of the monitor. The next field to be entered is the starting time of the trip using a 12-hr clock with a.m. and p.m. designations (the field is four numeric characters followed by one alphabetic character, limited to a or p). A 12-hr clock was chosen over the conventional 24-hr clock used in most travel surveys because it was more convenient for the interviewers and would also eliminate any errors made while converting from a 12- to a 24-hr clock. The request for entry then moves to the second line to record modal information on every segment of the person's trip. Trips are allowed a maximum of four segments made up of any combination of the available modes. Information is entered beginning with a single character field designating the mode

B Bus
S Subway
G GO train (regional commuter system)
V Via rail
D Automobile driver
P Automobile passenger
T Taxi
C Bicycle
O Other
9 Unknown or not available

followed by a free-format field of 10 characters maximum for a description of a public transit mode, such as a bus name or number. The transfer field is an eight-character free-format field used to record the location of transfers between public and private modes of transit. This field was included at the request of the local transit authority. The last field is used to record the return time, when appropriate, in the same format as the previous time field. A return time is entered if the following trip is identical to the current trip but in the reverse direction. The fields can be selected out of order with the use of the keyboard's arrow keys. The last field can always be reached in one step by depressing the END key on a standard keyboard.

The efficient use of screen space is demonstrated by the manner in which trip information is first recorded and then displayed as subsequent trips are recorded. When the interviewer moves on to recording the next trip, which is automatically done by depressing the RETURN key from the last field on a line, the next trip entry form overwrites the second line of the previous trip entry. However, at the same time, the four single-character designators of mode for each segment of the previous trip are displayed in a space that was left vacant beside the departure time field. In this manner the interviewer has all information about every trip displayed on one line, with the exception of the descriptions of the modes.

Trip information can be edited at any point in the process by selecting Function Key F1. A prompt asks, "Which trip?" The operator then selects the trip number to be edited and enters the new information. Function Key F1 is also used to move to the next household sample record when required and to exit the program when an evening's interviewing is complete.

TABLE 1 DIRECT DATA ENTRY—SUMMARY OF INTERVIEW TIMES

AUTOMATIC DATA ENTRY
Number of Interviews and Average Interview Times

Interview Day	Station 1 No.	Station 1 Avg.	Station 2 No.	Station 2 Avg.	Station 3 No.	Station 3 Avg.	Station 4 No.	Station 4 Avg.	All Stations No.	All Stations Avg.
04	10	8.50	9	10.33	6	18.00	16	10.00	41	10.88
05	16	7.94	9	13.11	8	12.75	8	9.88	41	10.39
06	18	5.22	13	7.54	9	14.11	17	5.76	57	7.32
07	6	9.67	6	9.67	4	16.00	6	4.67	22	9.46
08	19	7.74	16	7.75	11	11.55	20	7.60	66	8.34
09	19	6.89	13	7.23	8	9.25	12	9.17	52	7.86
18	25	4.88	15	7.67	7	18.00	16	6.88	63	7.51
19	23	5.39	18	8.61	12	12.08	18	7.44	71	7.86
20	18	4.89	15	6.00	11	11.64	12	10.29	56	7.67
All days	154	6.49	114	8.42	76	13.04	125	8.20	469	8.48

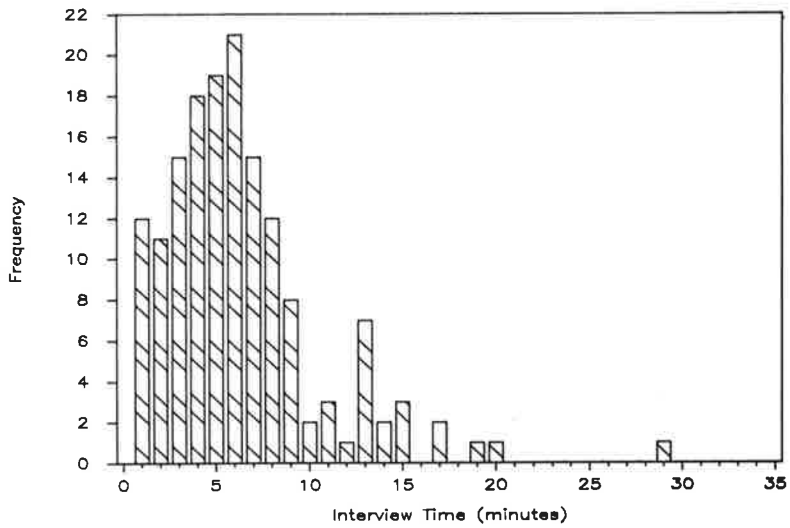


FIGURE 4 Direct data entry—Histogram of interview times, Station 1.

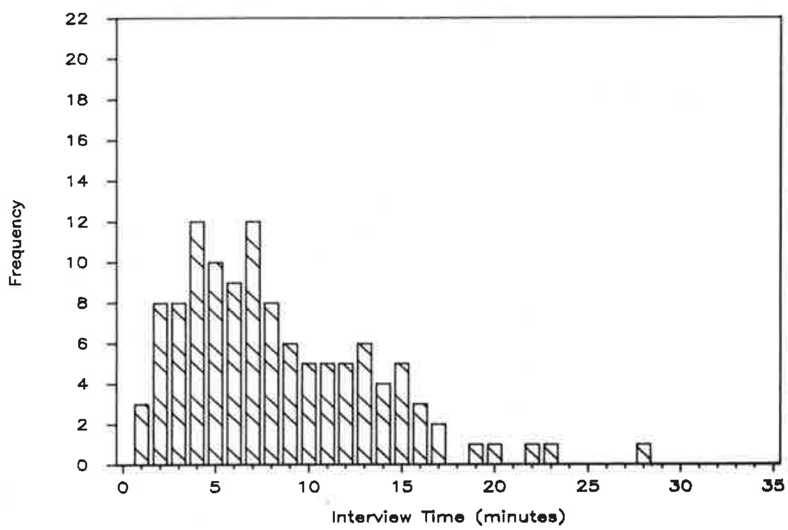


FIGURE 5 Direct data entry—Histogram of interview times, Station 2.

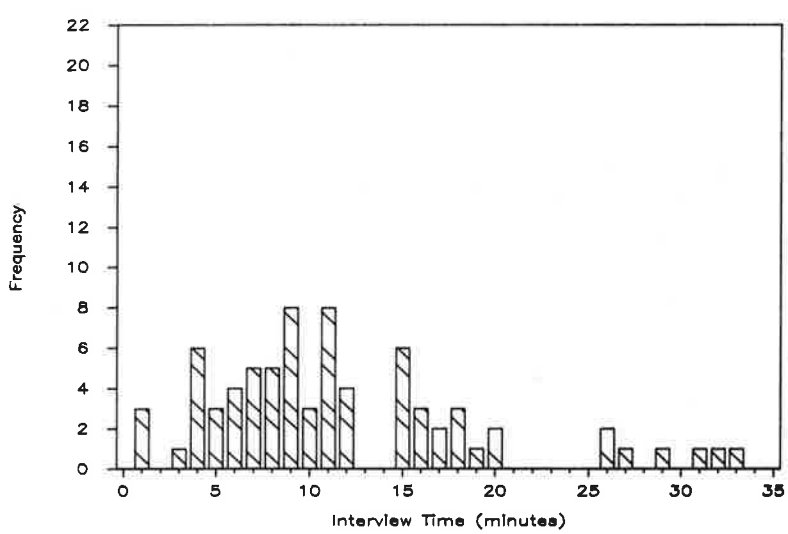


FIGURE 6 Direct data entry—Histogram of interview times, Station 3.

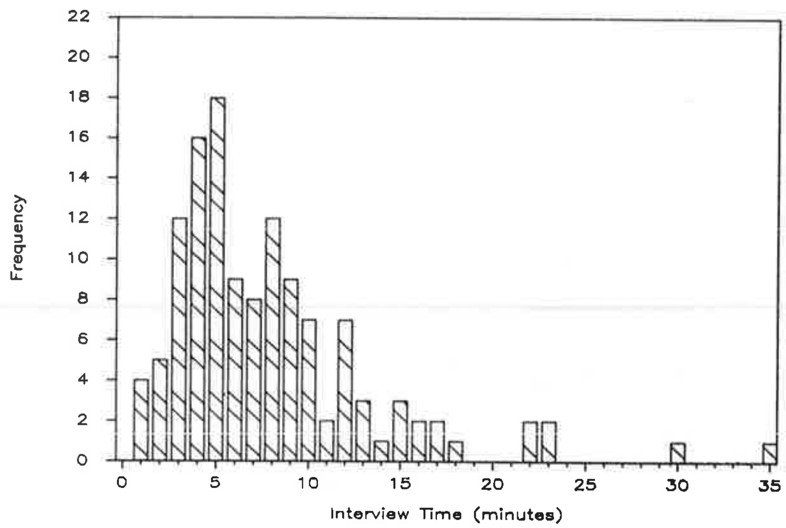


FIGURE 7 Direct data entry—Histogram of interview times, Station 4.

TABLE 2 DIRECT DATA ENTRY—SUMMARY OF ACTIVITIES FOR ALL STATIONS

TORONTO TRAVEL SURVEY
PILOT PROJECT
Summary of Automatic Data Entry Activities
'Day' designates the interview day
ALL STATIONS

Response Code	Day 03/04	Day 03/05	Day 03/06	Day 03/07	Day 03/08	Day 03/09	Day 03/18	Day 03/19	Day 03/20	Total
1	41	41	57	22	66	52	63	71	56	469
2	52	29	33	25	46	42	44	63	67	401
3	16	5	10	4	9	10	9	4	8	75
4	32	35	49	38	67	44	46	53	36	401
5	4	3	4	4	3	4	8	7	10	47
8	0	1	0	0	1	3	1	3	4	13
Total	145	114	153	93	192	155	171	201	182	1406
[1]Total eligible contacts	57	47	67	26	76	65	73	78	68	557
[2]Refusal rate	0.28	0.11	0.15	0.15	0.12	0.15	0.12	0.05	0.12	0.13
[3]Response rate	0.72	0.87	0.85	0.85	0.87	0.80	0.86	0.91	0.82	0.84
Completions per interviewer hour	2.31	3.50	3.06	2.08	3.50	3.29	4.21	4.24	3.45	3.32
Minutes per completed interview	26.0	17.1	19.6	28.9	17.1	18.2	14.3	14.2	17.4	18.1

Total number of records processed in 9 days of interviewing: 801

[1]Completed + refusals + language barriers
[2]Refusals/total eligible contacts
[3]Completed/total eligible contacts

Interviewers worked for several evenings before they were comfortable with the compact design of the screens. The initial reaction was for additional prompts and descriptions of each field or step in the process. By the end of the first week, however, they were able to read the complete history of a household's travel information at a glance. When asked in the second week, all interviewers agreed that the screen and computer entry was easier to use than paper forms for this unstructured type of interview. Elaborate screen prompts are not desirable for travel surveys for which the same interviewers are using the same program for up to 10 weeks (i.e., the expected duration of the Toronto O–D Survey).

The rate at which interviewers can process records is made up of two components, interview times and record-processing times. The automatic recording of begin and end times for every interview allows a detailed study of interview times. On the other hand, record-processing times can only be examined in combination with interview times in the form of total records processed in the total number of hours that interviewers spent working. Average interview times for every station and every period of interviewing are presented in Table 1. Improvement is apparent as the interviewers became more experienced. The average interview time settled down to a value of approximately 8 min per interview.

A more complete presentation of interview times is presented in Figures 4–7. The area under the histograms represents the total number of interviews complete during the pilot project. Stations 1, 2, and 4 appeared to interview at similar rates, much more efficiently than Station 3. The interviewer at Station 3 never became completely comfortable with direct data entry using a microcomputer; this lack emphasized the need for interviewers to have at least some typing ability.

Table 2 presents data on the number of completed interviews processed per hour. The best test of the process occurred on the evenings of March 18 and 19 because the interviewers had become more experienced with direct data entry and software problems were solved by this time. The results of March 20 were not as favorable because they represented the final night of interviewing. Consequently, most of this particular evening was spent trying to contact households that had not been reached on previous attempts. The average processing time settled down to a value close to the average of 14.1 min per completed interview experienced for 25,000 interviews in Vancouver in 1985. However, considering the problems Station 3 was having with the process and the precautions taken to keep a record of each sample on paper, an average processing time of 12 min should be achievable using direct data entry when interviewers are screened and become familiar with the process.

DATA PROCESSING

Even though significant productivity gains are possible with direct data entry due to efficient file management, major cost savings for the automated process are realized during the data processing phase. These savings are possible because the data are handled only once in the automatic process. The same data are handled by at least two other persons in the manual process in which the data are entered from paper form at a keyboard using double entry for verification. Apart from the potential manpower savings, the automatic process should also produce higher quality data by eliminating errors associated with double or triple handling.

Based on an O-D travel survey conducted in Vancouver in 1985, estimated costs for data entry were compared to

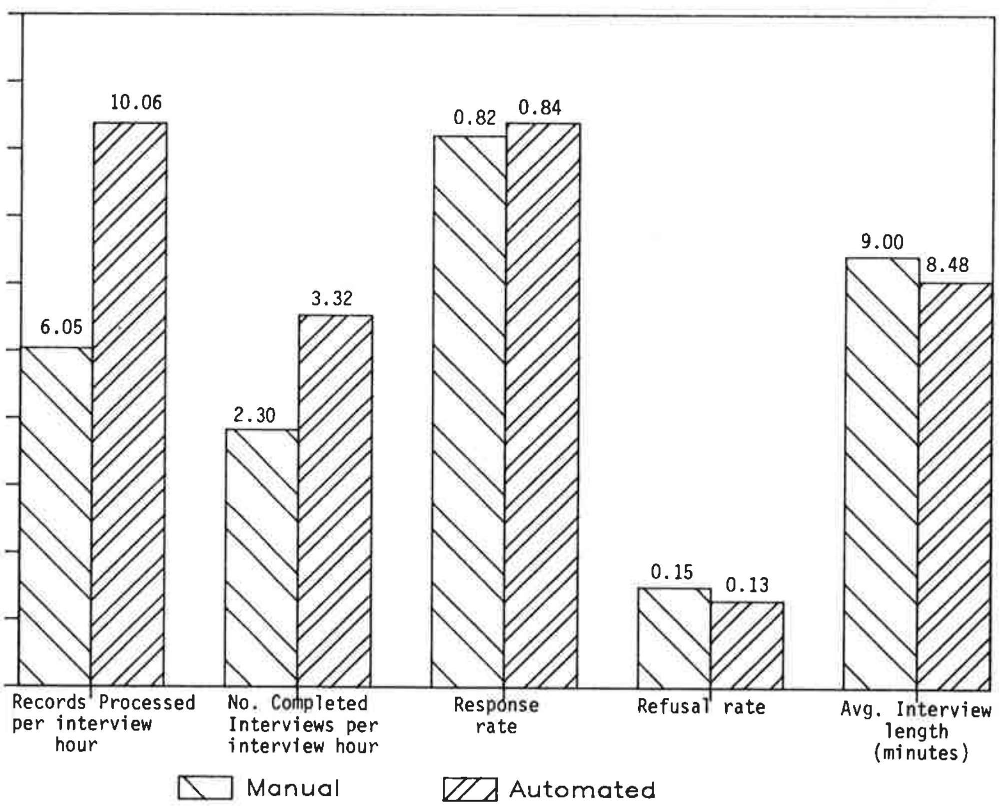


FIGURE 8 Comparison of survey methods—Manual versus automatic.

estimated microcomputer rental and program development costs associated with a full-scale Toronto region O-D travel survey *(2)*. It was calculated that the cost of direct data entry is virtually identical to the expense involved in keypunching paper forms. These cost estimates are based on the same number of interviewers working on both methods. One should be careful to observe, however, that improved productivity of interviewers using direct data entry truly makes this method more cost-effective (i.e., direct data entry produces more completed interviews for the same cost as manual methods).

RESULTS

Figure 8 shows the benefits involved with direct data entry and automatic file management on microcomputers. The comparison is drawn with the paper form method in order to show marked differences in processing times, but similar response, refusal, and interview rates. During the experiment, there were 67 percent more sample records processed per interviewer-hour and approximately 50 percent more completed interviews per interviewer-hour using the direct data entry method.

CONCLUSIONS

The direct data entry experiment, as part of the pilot project for a Toronto region O-D travel survey, provides new optimism for improving survey efficiency. The experiment proved ideal for comparing reliable manual methods (i.e., paper interview forms) to the automatic mode of data collection and entry. The two groups of interviewers allowed a fair comparison to be drawn between conventional and new methods. Productivity gains using direct data entry amounted to about 50 percent more interviews per hour. Sample processing is greatly improved by the use of an automated method (67 percent more records processed), whereas actual interviewing time, response rates, and refusal rates are comparable for both methods.

The largest benefit from direct data entry surfaces after interviewing is complete. Information collected in the automated process needs only to be coded to a zone system or geocoded; however, the manual method requires paper forms to be entered by keypunch onto a computer storage device. This step not only introduces potential error but also tremendously increases cost. This additional cost is virtually identical to the cost of providing each interviewer with a microcomputer and program if the same number of interviewers are used for both methods. Direct data entry using microcomputers can be incorporated in O-D surveys at no extra cost or risk over reliable manual methods. Furthermore, productivity gains using an automated method provide the potential to greatly surpass conventional methods in terms of cost-effectiveness.

ACKNOWLEDGMENT

The research described in this paper was funded by a research contract with the Ontario Ministry of Transportation and Communications. The views expressed in this paper are those of the authors and not necessarily those of the Ministry.

REFERENCES

1. *MTC Origin Destination Survey Pilot Study.* Environics Research Group Ltd., Toronto, 1986.
2. G. N. Steuart. *Survey Coding and Processing Techniques: Applications to a Travel Survey for the Regions of Durham, Halton, Hamilton, Peel, Toronto and York.* Ontario Ministry of Transportation and Communications, Toronto, 1986.

Management of a Small Home Interview Travel Survey Using a Microcomputer

RONALD EASH

In this paper, how a microcomputer and general-purpose business software were used in a small telephone survey of household travel behavior is discussed. Principal applications covered are (a) selecting sample households; (b) mailing to sample households; (c) scheduling interviews; (d) checking the status of each interview; (e) developing the completed interview data sets; and (f) reporting survey results. Staffing requirements and expenditures to complete the survey are also presented. This form of computing is well suited to data collection projects of this size.

The Chicago Area Transportation Study participated in a project to evaluate a commuter rail service extension in northeastern Illinois. The agency's responsibilities were to develop and calibrate a mode choice model to estimate ridership for the proposed commuter rail service. The resulting model includes choice of suburban station, choice of access mode to the suburban station, and choice of central business district (CBD) station to final destination egress mode—the sequence of decisions typically faced by commuter rail riders.

No data set with the detail needed to calibrate this type of model was available, so some data collection had to be included in the project. The use of a current database for the study area also strengthened the study's recommendations.

In this paper, a telephone survey of approximately 200 households to collect data on daily household travel behavior is discussed. This survey was managed with the aid of a microcomputer and commercially available microcomputer software. A small personal computer (PC) was used for (a) selecting sample households; (b) preparing personal form letters and mailing labels; (c) scheduling interviews; (d) tracking the status of each household's interview; (e) compiling data from the completed interviews; and (f) reporting survey results.

THE SOO LINE COMMUTER RAIL FEASIBILITY STUDY

Figure 1 shows the study area for the project, which covers 12 townships, each approximately 36 mi^2 in area. The track for the proposed service extension is owned by the Soo Line Railroad, which discontinued passenger service on this line in the 1950s.

The lead agency for the Soo Line commuter rail feasibility study was the Northeast Illinois Rail Corporation, now known as METRA. The consulting team assembled for the study was headed by R.L. Banks and Associates. JHK and Associates and CATS were subcontractors to R.L. Banks and responsible for the ridership estimates and model development.

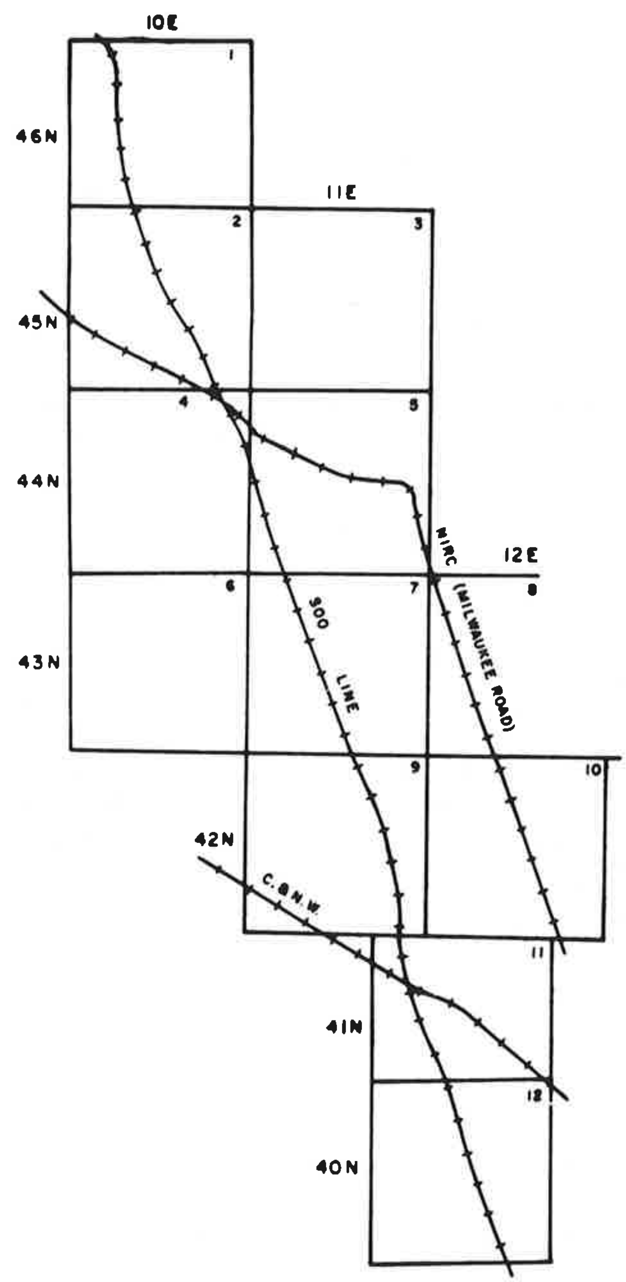


FIGURE 1 The Soo Line Corridor study area.

Chicago Area Transportation Study, 300 W. Adams Street, Chicago, Ill. 60606.

OVERVIEW OF THE SURVEY

For model calibration, the trip data set obtained from the surveyed households needed to have coverage in two dimensions. Trips had to be distributed across the study area townships and all major travel modes. Minimum numbers of responding households were required for each township and work trip travel mode, so selection of households to contact depended on the interviews completed while the survey was under way.

After a household was selected, a contact letter was sent to the household slightly in advance of the designated survey day. Accompanying this letter was a form that household members could use to keep track of their survey day travel. The evening following the survey day, the household was contacted by telephone. The person who answered was asked to report their household's travel during the previous day using the completed trip reporting form to recall trips. The interviewer would record the household's trips in a standard format, resolve inconsistencies in the reporting of trips, and try to aid in the recall of trips.

SELECTION OF HOUSEHOLDS TO BE CONTACTED

Households were selected for sampling from three sources depending on the probable mode chosen for household work trips. Sample households were obtained by (a) random sampling of households within townships; (b) random sampling of households on monthly ticket lists for existing commuter rail services in the study area; and (c) on-board solicitation of volunteers on several feeder and express bus lines in the corridor. The approximate numbers of these three types of households in study area townships were determined from the 1980 census Urban Transportation Planning Package *(1)*.

The variance in the estimate of a population characteristic obtained from a stratified sample is minimized by proportionally sampling each stratum according to the product of the fraction of population in the stratum times the standard deviation of the characteristic within the stratum *(2)*. For households selected randomly inside townships, the required number of sample households in a township was computed from the number of households in the township and the variance in township journey-to-work travel times reported in the census. Township household income as reported in the census was the population characteristic used to determine the sample size in the two categories of households with transit work trips.

In the case of the randomly selected households, it seemed desirable to have large samples in townships with varied travel behavior, which is partly measured by the variance in work travel times. Transit-oriented households were sampled to increase the sample sizes in townships with a variety of households. For this purpose, income seemed as good a measure of household variety as any other characteristic available from the census. Table 1 presents the desired household sample sizes assuming 210 (rounding produces a total of 212) completed interviews.

APPLICATION OF THE MICROCOMPUTER TO SAMPLE SELECTION

Randomly sampled households within townships were selected in the following way. Random *X–Y* coordinates were generated to identify points within air photographs. The nearest intersection to the randomly generated map coordinates was then located. One of the approaches into the intersection was chosen at random, and the range of addresses on this intersection approach determined between the selected intersection and the adjacent intersection. An address in this range was randomly generated and the nearest address that belonged to a private household selected from a reverse (address to telephone number) telephone directory *(3)*. The directory name and address were used for correspondence with the household.

Several short BASIC programs were written to generate map coordinates, select the intersection approach leg, and determine an address within an address range. In situations where there were multiple telephone listings for the same address, indicating more than one household at that address, selection of the household to contact was again at random.

Households from the monthly commuter rail ticket lists were chosen using randomly generated page numbers and page positions in the monthly ticket holder printouts. The printout page numbers and page positions were also selected with a short BASIC program.

TABLE 1 DESIRED HOUSEHOLD SAMPLE BY LOCATION AND WORK MODE

	Township												
Selection of Households	1	2	3	4	5	6	7	8	9	10	11	12	Total
1. Randomly by Township	2	5	3	2	6	3	6	12	22	15	26	28	130
2. Rail Monthly Tickets	0	2	0	1	2	1	3	12	13	10	12	5	61
3. Bus Riders	2	1	1	0	1	1	0	5	4	6	0	0	21
Total	4	8	4	3	9	5	9	29	39	31	38	33	212

MANAGEMENT OF THE INTERVIEWS

Figure 2 is the format of the data file for mailings of survey materials and the scheduling of interviews. This file was prepared with the microcomputer program PFS:FILE *(4)*, an inexpensive (the full retail price was $125) database management program with limited features. This is one of a number of such programs that can be purchased in this price range with comparable features.

The file format shown in Figure 2 is designed the same way one might prepare a set of index cards, whereby each card keeps track of one contacted household. The format in Figure 2 is for households randomly selected by township. The format for households from the monthly ticket lists and bus riders is identical except for a field that identifies the source of the household.

The name, address, and telephone number of the household in this file are copied from the reverse telephone directory listing. The state and last name of the household, which would appear to be redundant, are included so that the file can be used for mailing labels and personalized form letters.

Fields for scheduling the telephone interview dates are listed in the format underneath the address information. Interviewers are identified in the file along with whether the interviewer could contact the household and, if contacted, whether the household cooperated in the survey. This information allowed for some limited evaluation of interviewer performance. At the end of the format is the final disposition of the household. Is the household included in the data set? unreachable after three call-backs? not a household? or not cooperative?

Household travel was surveyed only on midweek days, Tuesday through Thursday. Phone interviews were usually scheduled for Monday, Wednesday, and Thursday evenings after the dinner hour. In order to complete the mailing of survey forms, households had to be assigned a survey day about 1 week in advance of the interview.

The scheduling of households depended on (a) the availability of telephone interviewers to work the evening following the survey day; (b) the running tabulation of successfully completed interviews by township and mode; (c) the number of households to be recalled (households that had received the mailing but could not be contacted on their assigned survey night); and (d) previous experience with refusal rates. After an evening of interviewing, the file of contacted households was updated the following morning.

MAILINGS TO SAMPLE HOUSEHOLDS

A personalized form letter and a form for tabulation of household trips were mailed to each contacted household. The mailing was scheduled so that the form and introductory letter arrived no more than a day or so in advance of the survey date. Mailing labels were generated from the survey household file discussed, again using the PFS:FILE program. The survey date field was used to pull the addresses of scheduled households for the mailings from all households in the file.

Form letters were printed for each household using the word processing program PFS:WRITE *(5)*, which was compatible with the contact household file. The inside address and salutation were copied from the contact household file and filled out as the letters were printed. The selection of which form letters to print was controlled through the survey date field in the household data file. Labels for the interview forms used by the phone interviewers were also created from the contacted household file.

RESOURCES REQUIRED FOR THE INTERVIEWING

Due to the small size of the survey, only two or three phone interviewers were needed each evening. The interviews were carried out over a 3-week period in the early fall of 1985. During the 3-hour interview period (6:30 to 9:30 p.m.), an

```
                          RANDOM SAMPLE HOUSEHOLDS

Township Number:           Household Number:
==============================================================================
Directory Name:
Address Name (If Different):
Address1:
Address2:
City:                                           State:            Zip:
Telephone Number:                           Last Name:
==============================================================================
Scheduled Survey Date:                      Scheduled Survey Call Date:
1st Call Back:             2nd Call Back:             3rd Call Back:
==============================================================================
Scheduled Interviewer:
Completed Interview:      Unable to Reach:      No Longer a Residence:
==============================================================================
Call Back Interviewer:
Completed Interview:      Unable to Reach:      No Longer a Residence:
==============================================================================
Interview in Sample:
```

FIGURE 2 Data file format for contacted households.

experienced interviewer could complete about 10 interviews. Allowing for refused interviews and available call-back interviews, approximately 40 to 50 households received mailings for each survey day.

With the exception of the interview staff and a small amount of staff time to choose the contact households, only one staff member was required to run the survey. A single staff person familiar with database management and word processing software can easily handle the survey mailings and scheduling of interviews for this size of survey. Coordination of the mailings, interview scheduling, and completed interviews—critical for this type of survey—are difficult when the mailings and data processing require several agency staff.

COMPILATION OF SURVEY DATA SETS

Three data sets for model calibration were prepared from the surveyed households. Data were tabulated by household, by person for each adult member of a household, and by trip for all household trips that could reasonably be completed by commuter rail. All surveyed households are combined in these three data sets.

```
                         HOUSEHOLD DATA SET:
=========================================================================:
Random Sample Households Township Number:
Sample Transit Households Township Number:       Transit Source:
=========================================================================:
Household Number:      Address1:
City:                                            Zip:
Home Q-Sec:
-------------------------------------------------------------------------:
Composition:

Adults:         Children:           Enfants:         Female Head:
Elderly:        Handicapped:        Drivers:
-------------------------------------------------------------------------:
Employment:

Emp. Full-Time:          Emp. Part-Time:          Retired:
Student Full-Time:       Student Part-Time:
-------------------------------------------------------------------------:
Vehicles:
Automobiles:            Trucks/Vans:
```

FIGURE 3 File format for the household data set.

```
                     PERSONAL CHARACTERISTICS FILE:
=========================================================================:
Random Sample Households Township Number:
Sample Transit Households Township Number:          Transit Source:
Household Number:
=========================================================================:
Adults:                     Person Number:
Sex1:    Relationship1:     Employment:    Drive:    Elder:    Hand:
-------------------------------------------------------------------------:
Children:                   Child Number:
Sex2:        Age:            Relationship2:
-------------------------------------------------------------------------:
Work Trip:                                        Residence:
Travel Time:                                      City:
Principal Mode:                                   Zip:
Work City:                                        Home Q-Sec:
Work Intersection:
Work Q-Sec:
-------------------------------------------------------------------------:
Commuter Rail:   Line:                           Station:
Access Mode:                Parking Cost:              Transit Fare:

                             Soo Line Trips

Work Trips/Month:           Nonwork Trips/Month:
```

FIGURE 4 File format for the personal data set.

```
                              DRIVE ACCESS FILE:
==========================================================================;
Random Sample Households Township Number:
Sample Transit Households Township Number:          Transit Source:
Household Number:                    Person Number:
==========================================================================;
Closest Station:            RR Line1:           RR Station1:
Distance1:                  Parking Fee1:          Train Freq1:
--------------------------------------------------------------------------;
Next Closest Station:       RR Line2:           RR Station2:
Distance2:                  Parking Fee2:          Train Freq2:
Time Savings2:                       Cost Savings2:
--------------------------------------------------------------------------;
Lower Cost Station:         RR Line3:           RR Station3:
Distance3:                  Parking Fee3:          Train Freq3:
Time Savings3:                       Cost Savings3:
--------------------------------------------------------------------------;
Bus Access Station:         RR Line4:           RR Station4:
Distance4:                  Bus Fare:              Bus Freq:
Stop Distance:
--------------------------------------------------------------------------;

Walk Access Station:           RR Line5:           RR Station5:
Distance5:                                              Train Freq5:
-------------------------------------------------------------------------;
Kiss and Ride Station:         RR Line6:           RR Station6:
-------------------------------------------------------------------------;
Auto Passenger Station:        RR Line7:           RR Station7:
-------------------------------------------------------------------------;
Home Q-Sec:
Work Q-Sec:
=========================================================================;


                                 LINE-HAUL FILE:
==========================================================================;
Auto:
Line-Haul Time1:
Destination Parking Fee:
--------------------------------------------------------------------------;
Other Transit:                                Stop Distance2:
Line-Haul Time2:                              Other Freq:
Line-Haul Fare2:
Walk Distance to Destination2:
--------------------------------------------------------------------------;
Commuter Rail:
Line-Haul Time3:
Line-Haul Fare3:
Bus Distance to Destination:                  Added Transit Fare:
Bus Stop to Destination:
Walk Distance to Destination:
--------------------------------------------------------------------------;

                          GENERAL TRIP CHARACTERISTICS:
==========================================================================;
Work Trip:
Trip to CBD:
==========================================================================;
```

FIGURE 5 File format for the trip data set.

These files were again assembled using the program PFS:FILE. No attempt was made to encode the data directly into the microcomputer files during an interview. Interviewers recorded the results of the interview on an interview report form and the data files were then transcribed from these forms.

Figure 3 shows the format for the household data set. This data set describes the household, its composition, employment characteristics, and vehicle ownership. There are a total of 198 records (a segment of the file containing the items in the figure), one for each completed interview.

The format for the personal data set is shown in Figure 4. There is one record for each adult household member and the file describes the individuals and their work trips. It contains 421 records from completed interviews.

The last data set format in Figure 5 is the trip data set. This file includes all reported trips that conceivably could ride existing commuter rail lines in the study area. Generally, an acceptable trip for this file is long enough and directionally oriented so that a significant part of the trip can use available commuter rail services. There are 152 trips in this file, somewhat less than one per household.

The trip data set has three sections, for commuters who board at suburban stations, line-haul travel, and general trip characteristics. The first section of the file summarizes how the individual reaches a suburban commuter rail station when the line-haul mode choice is commuter rail. The line-haul portion of the file describes the line-haul trip segment and notes how travelers reach their final destination from the point where they depart a line-haul travel mode. This point is usually a parking garage or lot, a rail transit station, bus stop, or downtown commuter rail station. Two other trip characteristics, work or nonwork and CBD destination or non-CBD destination, are noted in the last general trip characteristics section.

The model calibration requires that additional hypothetical trips be entered into the trip file. For example, the complete trip file for model calibration has to include the trip characteristics of current commuter rail trips hypothetically diverted to the automobile driver mode. The cost estimates that follow do not include the costs of estimating the characteristics of these alternative mode trips.

STAFFING AND COSTS OF THE SURVEY

Three hundred seventy-eight households were contacted, and 198 interviews were successfully completed. There were also a few partially completed interviews that produced useful data for one or more of the files.

The survey covered four bimonthly agency accounting periods in September and October 1985. The staff-person-hours presented in Table 2 are tabulated from these four accounting

TABLE 2 RESOURCES EXPENDED ON THE HOME INTERVIEW SURVEY

1. Staff Person-Hours

Work Item	Per Contacted Household	Per Completed Interview
1. Selection of Contact Households	0.24	0.45
2. Mailing of Survey Forms	0.15	0.29
3. Phone Interviews	0.28	0.53
4. Compilation of Survey Data Sets	0.16	0.31
Total	0.83	1.58

2. Approximate Manpower Cost (Average Wage Rate Plus Overhead at 120 percent)

Work Item	Per Contacted Household	Per Completed Interview
1. Selection of Contact Households	$7.00	$13.00
2. Mailing of Survey Forms	$4.40	$8.50
3. Phone Interviews	$8.20	$15.60
4. Compilation of Survey Data Sets	$4.70	$9.10
Total	$24.30	$46.20

periods. Staff-person-hours are further subdivided into four survey work elements identified by sections in this paper. Slightly more than 1.5 staff-person-hours were required to complete an interview. This amount covers selection of the household through the entry of the household, person, and trip data into the microcomputer files, but does not cover any analysis of the data or model calibration activities.

These staffing requirements were converted to dollar estimates using $26,000, the average CATS annual salary at the time of the survey, and an agency overhead rate of 120 percent. These figures are presented in the bottom half of Table 2. For this survey, the cost of data collection is just under $50 for each successfully completed household interview.

The bar chart in Figure 6 shows the staff-person-hours broken down by work element and accounting period. This chart shows how expenditures for these four work elements vary over the time required to complete the survey. Staff-person-hours for sample selection and mailing of survey forms largely precede the actual interviews, whereas the bulk of the expenditures for compiling the data into the microcomputer data sets follows the interviews.

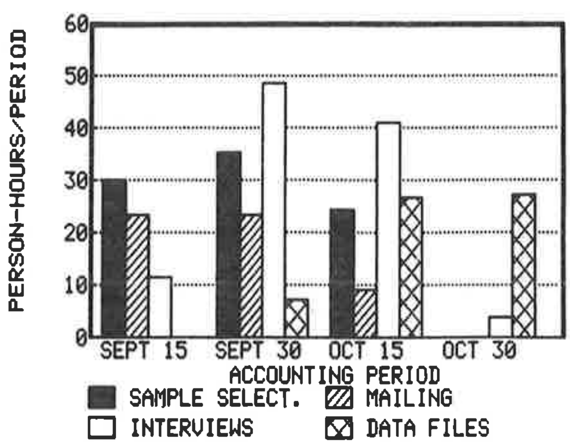


FIGURE 6 Survey staff-person-hours by accounting period.

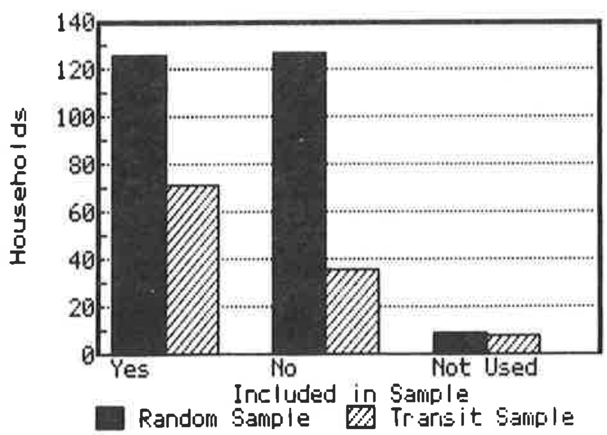


FIGURE 7 Business graphics display of survey data—survey households.

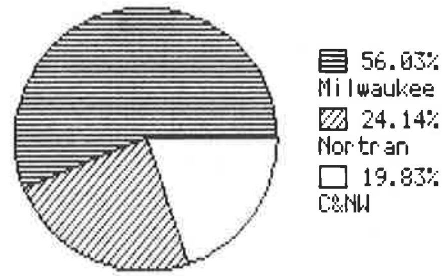


FIGURE 8 Business graphics display of survey data—survey transit households sample source.

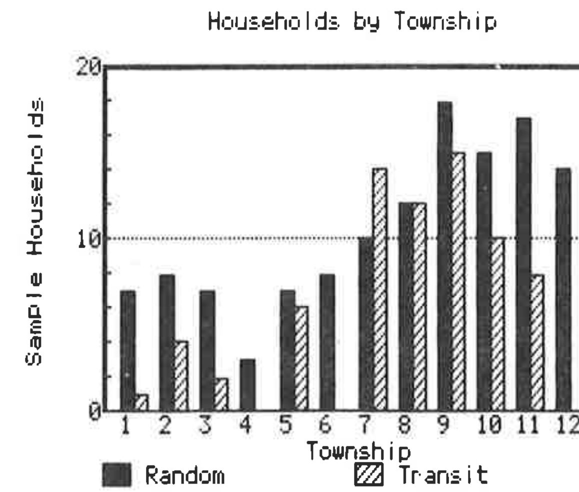


FIGURE 9 Business graphics display of survey data—households by township.

DISPLAY OF SURVEY DATA

Figures 7–9 provides several examples of the use of a business graphics microcomputer program, in this case PFS:GRAPH *(6)*, to display the survey data. The graphs in Figures 7 and 9 summarize the sampled households in the survey. The bar graph in Figure 7 shows the number of responding and refusing households according to how the household was selected to be contacted. The pie chart in Figure 8 is the number of sample households obtained from the monthly tickets and bus ridership (the NORTRAN wedge in the pie chart). Successful interviews by townships are shown in the bar chart of Figure 9.

SUMMARY AND CONCLUSIONS

This paper has demonstrated how the use of a microcomputer and general-purpose business microcomputer software can assist typical data collection in a metropolitan planning organization. The survey involved a small number of professional staff. The manager of the project directly managed the survey and performed the microcomputer applications presented in the paper. He always had access to the microcomputer because it was located in his office. The project manager had several years' experience in mainframe computing, a year or so of experience with a microcomputer, and some familiarity with the business software and its capabilities before undertaking the project. The priority was to complete the survey, not to learn about personal computing.

The following conclusions can be drawn from this experience:

1. There is no need for specialized software or computing skills to support reasonably sized data collection efforts for many transportation planning applications, given the capabilities of general-purpose business microcomputer programs and the ease with which this software can be used.

2. The microcomputer greatly enhanced the project manager's ability to respond to situations that arose during the survey, particularly in the scheduling of households to contact as interviews were completed.

3. The microcomputer permitted a great deal of control over the survey because it reduced the number of administrative and data collection staff needed for the survey. The survey was not subject to delays due to other agency data collection and data processing activities.

4. The completion of the survey within a tight time limit was to a large extent made possible by the availability of the microcomputer. Access to the machine, data files, and software were unrestricted and work could continue outside normal business hours when it was necessary to catch up to schedule.

REFERENCES

1. *Urban Transportation Planning Package, 1980 Census: Technical Documentation for Summary Tape.* Journey-to-Work and Migration Branch, Population Division, Bureau of the Census, U.S Department of Commerce, Washington, D.C., 1983.
2. M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand.* M.I.T. Press, Cambridge, Mass., 1985.
3. *Chicago North Suburban Street Address Directory.* Reuben H. Donnelley Corp., Chicago, Ill., 1985.
4. *PFS: File: User's Manual.* Software Publishing Corp., Mountain View, Calif., 1984.
5. *PFS: Write: User's Manual.* Software Publishing Corp., Mountain View, Calif., 1984.
6. *PFS: Graph: User's Manual.* Software Publishing Corp., Mountain View, Calif. 1983.

A Quick Cluster Control Method: Permanent Control Station Cluster Analysis in Average Daily Traffic Calculations

DAVID ALBRIGHT

A foundational piece of information in highway planning is average daily traffic (ADT). The way in which permanent control counters are used to adjust short-term traffic counts to ADT impacts current and forecast traffic analysis. ADT values affect geometric design of roadways. Too-low adjustment and forecast of ADT values can result in inadequate lanes or thickness of lanes, causing low levels of service and surface failure. Too-high ADT values can result in overbuilding a facility. The New Mexico State Highway Department has implemented a statistical procedure to improve the accuracy of ADT values. Whereas the previous procedure was to use a nearby permanent control counter to adjust short-term counts to ADT, the new procedure uses cluster analysis of permanent control counters. The clustering is based on functional classification of roadways. The benefit of cluster analysis has typically been a 3 to 4 percent per year improvement in ADT accuracy, resulting in 15 5o 20 percent accuracy improvement between traffic count years. When forecast to design year for a roadway, the method can improve volume accuracy by more than a lane of traffic. The quick cluster control method provides significant improvement in traffic accuracy without additional staff requirements.

Highway planning agencies attempt to use the best possible methodology at the lowest cost. This policy is particularly true in a period of budgetary constraint. To be acceptable, new methodologies must be demonstrated as providing significant improvement in accuracy, and must be implemented with current or reduced personnel. In this paper, a procedure to facilitate decision making in one area of improved highway planning methodology is outlined. The procedure, referred to as quick cluster control, demonstrates the accuracy and facilitates implementation of cluster analysis in traffic volume estimation.

One of the foundational pieces of information in highway planning is average daily traffic (ADT). ADT is an estimation of the daily traffic volume on a given roadway. Where there is a permanent counter, an hourly record of traffic is recorded for each day of the year. The daily totals may be summed, then divided by 365, to obtain annual average daily traffic (AADT). Where there is no permanent counter, short-term coverage counts for an area are conducted. Generally, there is a cycle of months or years between counts. The short-term count is taken over a period of days. A permanent counter is used to adjust the short-term count to an estimated ADT. In part, the accuracy of an ADT is affected both by the accuracy of the short-term count and by the factor used to adjust the count to ADT. A historical record of ADTs is maintained and updated annually.

Why is the accuracy of ADT important? The ADT values for a roadway affect current and planned traffic control devices. The historical trend in ADT values is used to forecast traffic volume to roadway design year. Error in current ADT expands significantly as the data are forecast to the design year. Forecast volumes affect decisions on the usefulness of alternative relief routes and on the geometric design of relief routes. Forecast ADT based on annual record entries is also an essential portion of average daily load (ADL) values, which affect structural characteristics. Inaccuracy of ADT, current and forecast, can result in inadequate lanes or thickness of the lanes, causing low levels of service and surface failure. Too-high estimates of ADT can result in overbuilding a facility. In either circumstance, ADT error is costly. For this reason, ADT values must be prepared as accurately as possible.

The current ADT factor method used by the New Mexico State Highway Department is calculated as shown in Equation 1.

$$\text{ADT} = \frac{\text{(Usable hours from coverage count)}}{\text{(Total from same hours from control station)}} \times \text{(ADT at control station for the previous year)} \quad (1)$$

When a short-term coverage count is taken, the total count is divided by the same period traffic counted at an individual permanent counter. A counter is selected as controlling that area, and is called the control station. A proportion is established of ADT at the permanent counter to the ADT at the coverage site. This is then multiplied by the ADT at the permanent counter site during the previous year. The result is a factored ADT used to estimate traffic volume at the coverage site.

Individual permanent counters are used to factor coverage counts throughout the state, and to make annual ADT adjustments when counts are not taken. Individual counters must be used to control extensive areas of roadway.

There is a 5- to 6-year cycle between coverage counts. Between the short-term coverage counts, how are ADT values determined for entry into the record? Each year, the permanent counter controlling the roadways in a given area is reviewed for percent change. This change ratio is then applied throughout the area controlled by that individual permanent counter.

How is a permanent counter station assigned to control a roadway? Usually the physically closest permanent counter on

New Mexico State Highway Department, 1120 Cerrillos Road, Santa Fe, N. Mex. 87504.

the same or nearby roadway is used as the control. There is sometimes a review of similar characteristics between the site of the permanent counter and where the short-term count is taking place. An Interstate location would not, for example, be used to control a nearby access road.

How are new permanent counter locations determined? The 5-year plan for highway construction and maintenance is reviewed for additional permanent control sites. If a roadway is being reconstructed, it is a relatively inexpensive addition to install a permanent counter. Before development of the present procedure, there was no cluster analysis of permanent counter locations. For example, two additional urban sites may be needed to establish a statistically valid group for that type of roadway. There could be a phase-out of current counter locations where there are unnecessary counters. Need for permanent counters on urban roads and overcounting of rural primary arterials were discovered while conducting the cluster analysis.

Over what period are the coverage counts conducted? The current method can incorporate traffic volumes 7 days of the week. Short-period 48-hr counts are the rule.

A major difficulty with the current New Mexico State Highway Department method is that the nearest permanent counter, even if it appears similar, may not be characterized by similar growth as the area it is used to control. Currently, permanent counter locations are used in two ways that affect the accuracy of ADT values. First, when a count is taken, the count is modified by the permanent counter that controls that area. Second, in the absence of a blanket or coverage count, the change in the permanent counter is used to calculate record ADT values. Although there undoubtedly is error in the adjustment during the count year, the primary concern of the quick cluster control method is error in noncount years as individual permanent counter change ratios are applied.

Cluster analysis of permanent counters is an alternative to the current individual counter method. Cluster analysis is placement of objects into groups (or clusters) suggested by information about the objects. In clustering, groups are defined by the data. Similarities and differences between objects may not be immediately apparent. Cluster analysis involves finding the greatest similarity among objects within a group, and the greatest difference between groups. All clustering methods are based on the same agglomerative procedure. Each observation, or permanent counter, begins in a unique cluster. The two closest clusters are merged to form a new cluster. Merging of closest clusters is repeated until only one cluster remains. Clustering methods differ in how closest clusters are defined. There are then statistical and application techniques employed to help determine the optimum number of groups.

Clustering was applied to the permanent counter locations in New Mexico, excluding those counters on the Interstate. The statistically based clustering results were then compared with results of clustering by functional classification of roadways on which the permanent counters are located.

The decision to explore clustering and to develop a quick cluster control method was based in part on the apparent deficiencies of the current method. The decision was also based on the cluster analysis recommended by the FHWA and used by other states. An outline of federal and state effort is helpful before detailing cluster analysis using New Mexico data.

What has the federal government encouraged and states other than New Mexico accomplished in use of cluster analysis? Much of the impetus for state-level research and development of cluster analysis has come from the federal government. A standard reference text is the *Guide for Traffic Volume Counting Manual (1)*, published in 1965. The technique identified for clustering was labor-intensive, unlike the rapid computer routines available today. The method involved color-coding a specific monthly group map for stations and groups of stations. Despite the lengthy procedure, open to a variety of manual errors, the document was helpful.

The Guide *(1)* identified use of a monthly traffic ratio (MTR), the ratio of monthly average weekday traffic (MAWDT) to AADT. It then documented two important characteristics that led the way to clustering of traffic volume data:

- The pattern of monthly variation of traffic volume persists over long stretches of highway, and
- The pattern of monthly variation of traffic volumes persists over long periods of time.

The group mean factor from cluster analysis may be applied to short-term counts located on road sections that fit within the group.

The Guide *(1)* recommended a minimum number of permanent counter locations within each group. It suggested four was the minimum for valid results. More permanent counter locations were recommended.

An updated version of the *Guide for Traffic Volume Counting Manual (2)* was published in 1970. It was clear the MTR values should be applied to short-term counts as an adjustment factor, and daily adjustment factors were ruled out. The Guide *(2)* averaged urban and rural MTR values to approximate suburban area ratios. There was a significant amount of refinement remaining.

The FHWA published the *Guide to Urban Traffic Volume Counting (3)* in 1981. This document notes that from three to five groups will be determined when 4-year averages of monthly data are used. The resulting MTRs are applied to weekday volume measurement as an adjustment factor. The Guide *(3)* encouraged the short-term weekday volume count as being no less than 48 hr in duration. It was noted that 48-hr counts pick up an additional 2 percent accuracy over 24-hr counts.

In 1985, the FHWA published the *Traffic Monitoring Guide (4)*. Appendix A of this Guide *(4)* outlines a clustering procedure for permanent counter locations. Using Statistical Analysis System (SAS) procedures, two criteria were used to select the optimum number of groups. A statistical indicator called the "cubic clustering criterion" was used. When this statistic reaches a minimum value, the corresponding number of groups is considered optimum. Additionally, when the R^2 statistic is significant, but the gains become modest, the optimum number of groups is indicated. When the results of these two rule-of-thumb tests vary, judgment is used to select the optimum number of groups. This judgment is based upon familiarity with the data and permanent counter locations.

What has been accomplished on the state level? In West Virginia during 1969, Pant and Wegmen *(5)* in *A Multiple Linear Discriminant Function Analysis* used classification count data for grouping truck weight stations. In 1971, Lieder *(6)* examined and amplified this work in *A Grouping Procedure*.

Pant established five groups initially, then reduced this number to four. When he tried various combinations of variables for defining groups, ADT and percent trucks gave inconsistent groupings.

This grouping procedure was successfully used by Pant and Wegman *(5)*. Lieder *(6)* modified and applied it to Texas truck weight and Maryland traffic data. Lieder found that the sum of least square differences provides good groups "if good judgment is exercised as to the cutoff," in optimum number of groups. Lieder also recommended that a minimum of 3 years of data should be used in the analysis. In other words, groups can be established using this procedure, although it requires judgment. The specification of multiple-year average data proved valuable in Lieder's work.

Sharma and Werner *(7)*, in Alberta, Canada, used cluster analysis in *Improvements in Travel Monitoring and Data Aspects of the Energy Problem*, published in 1981.

Sharma and Werner reviewed the existing literature, and focused on the *Guide for Traffic Volume Counting Manual (2)* and its use of MTR values. Sharma and Werner grouped 45 permanent counter locations. Of these, 44 were provincial primary highways. The other was secondary. Sharma and Werner used hierarchical grouping and wrote, "It has to be emphasized here that this method . . . does not indicate specifically what the optimum number of groups is for the study objectives."

Sharma and Werner *(7)* noted that the errors associated with each step of the grouping process identified a range of groupings. Establishing an effective range reduced the impact of subjective consideration in selecting an optimum number of groups. Their conclusion was that the optimum number of groups using this data was between 6 and 10. Still, in application, some specific grouping must be used. They selected the middle of the range, 8 groups.

This optimum number of groups was based on comparisons of mean monthly traffic factors. However, within the eight, Sharma and Werner *(7)* designated four primary groups, which were used for further analysis.

Sharma and Werner *(7)* analyzed the characteristics of the primary groups, which they termed "rationalization of resulting groups." Sharma and Werner were interested in different combinations of trip characteristics, such as purpose and length, among the groups. They used origin and destination studies as the basis of the analysis. From this work, the ADT values were improved, but they did not find a clearly distinguished load classification breakdown among the groups.

In Louisiana, Shah and Hirschmann *(8)* applied cluster analysis to permanent counter locations in *Analysis of Routine Traffic Count Stations to Optimize Locations and Frequency*.

Shah and Hirschmann *(8)* used 3-year traffic counts as the basis for the cluster analysis, which used functional classification. They wrote, "A clustering system crossing functional classes and/or parish lines was ruled out. A review of the results indicated that this method of analysis could not lend any valid interpretations in the determination."

Shah and Hirschmann *(8)* also noted that this was a limited study, not intended to be a complete statistical analysis. They used 3-year average statistics in accordance with observations from Dreusch's *(9)* analysis in Missouri in 1966 and Lieder's *(6)* 1971 research.

Hartgen and Lemmerman *(10)* wrote *Streamlining Collection and Processing of Traffic Count Statistics*, published in 1982. The work identifies application of ADT clustering techniques in New York State.

The motivation for their work was reduction in seasonal control stations. Through grouping assignment and other measures, they reduced counts by 35 percent with little or no reduction in information accuracy.

New York began with 12 factor groups when first applying cluster analysis. In 1975, this number was reduced to 8; in 1982, to 4. The judgment of Sharma and Werner *(7)* in moving from 8 groups in 1975 to 4 groups in 1982 paralleled the experience in New York.

After the clusters were compiled, it was determined that there was an unnecessarily large number of stations in some groups. Permanent counter stations were eliminated without negatively affecting the sample.

The four groups Hartgen and Lemmerman *(10)* determined by cluster analysis were

1. Urban or commuter,
2. Low-flow nonrecreational (less than 1,000 veh/day),
3. Rural long-distance, and
4. Recreational.

The variability of the estimates for the grouping ranged from 2 percent in Group 1 to 20 percent in Group 4.

After the groups were formed, they were analyzed by location, design, and use characteristics. After a fairly extensive analysis of variables, the results were interesting. "No combinations of these variables would perfectly distinguish the groups, but the urban/rural nature and access control of each highway appeared to have the greatest potential for making an initial distribution."

The New York application of cluster analysis used MTR values that could be characterized by urban or rural nature, and access control. This distribution is essentially by functional classification of roadways, with some reduction of the access and mobility distribution used in functional classification *(11)*. Other variable analysis was instructive, but did not enhance the groups.

The New Jersey Department of Transportation uses cluster analysis. In correspondence from Louis Whitely, Project Engineer, it was noted that clustering is done each year. Some shifting of counters among groups occurs. The grouping analysis is maintained through the rest of the year.

The New Jersey Department of Transportation is moving to a functional classification pattern for the groups that will appear as

1. Rural Interstate,
2. Urban Interstate,
3. Other rural,
4. Other urban, and
5. Highly recreational.

One of the implications of the cluster analysis has been the phasing-out of unnecessary permanent counter locations. The number of new locations required for some groups was far outweighed by the number of locations that could be removed from other groups. The result has been increased accuracy with smaller amounts of data that must be manipulated.

By the early 1980s, the federal guidelines and state applications of those guidelines had become refined. The technique of clustering permanent counters was demonstrated in a variety of settings. The use of computers made cluster analysis a relatively easy procedure, requiring small personnel time commitment after the initial analysis was complete. What has the use of clustering shown? A minimum of 3 years' data, with a minimum of four to six counters in each cluster, provides helpful MTR values. Finally, the use of clusters based on functional classification should be carefully examined from work in Louisiana and New York, as well as developing procedures in New Jersey.

If cluster analysis can be successfully conducted, will this methodology significantly improve the accuracy of ADT calculation? Based on the answer to this question, a quick cluster control method may be developed.

Cluster analysis was used to define groups of permanent control stations in New Mexico. All permanent counter locations not on the Interstate were clustered. Not all New Mexico permanent counter stations have full-year data available. Some 28 stations had sufficient data for analysis.

The SAS PROC CLUSTER was used. This program performs hierarchical clustering of observations using 1 of 11 agglomerative methods. The clustering method selected was Ward's *(12)* minimum variance method. Ward's, along with squared Euclidean distances, generally produces the best clustering results. Ward's method is based on minimizing the within-cluster sum of squares.

The initial cluster was for 1-year data. The most recent year for which there were complete data was used. A total of 12 monthly traffic factors were developed for each station. Initially no multiple-year averaging was done.

What did the analysis show? The clustering produced eight groups of permanent control stations. What were the indicators used to identify the optimum number of groups? The R^2 statistic was confident at the 95 percent level at eight groups,

TABLE 1 CLUSTER ANALYSIS GROUPS, BASED ON 4-YEAR AVERAGE DATA

Group	Functional Classification	General Location
Group 1		
C7S	Rural primary arterial	North of Santa Fe–NU Limit
A10S	Rural minor arterial	South of Taos
A40	Rural primary arterial	NM 68 north of Taos
A41	Rural primary arterial	NM 44 at Cuba
A87	Rural primary arterial	North of Silver City
A107E	Rural primary arterial	North of Alamogordo to Cloudcroft
A113	Rural primary arterial	West of Ft. Sumner
A105	Rural primary arterial	East of Ruidoso
Group 2		
A115	Rural minor arterial and major collector	North of Clovis
A34N	Rural primary arterial	North of Cedar Crest
A107S	Rural primary arterial	North of Alamogordo to Tularosa
SFC	Urban primary arterial	Cerrillos Road in Santa Fe
SFD	Urban primary arterial	St. Francis Drive in Santa Fe
B119	Rural primary arterial	South of Roswell
A32S	Rural minor arterial and major collector	South of Armijo
A119	Rural minor arterial and major collector	South of Roswell
CBD–B	Urban minor arterial	Canal St. near Plum in Carlsbad
Group 3		
B59	Rural primary arterial	West of Aztec
CLO–A	Urban primary arterial	West of Grand Avenue in Clovis
B124	Rural minor arterial and major collector	South of Lovington
A59	Rural primary arterial	South of Farmington
A–74	Urban minor arterial	Southwest of Las Cruces
ALB–AA	Urban minor arterial	US 85 in northwest Albuquerque
A–124	Rural minor arterial and major collector	South of Hobbs
ALB–SM	Urban primary arterial	San Mateo in Albuquerque
Group 4		
A25W	Rural primary arterial	US 56 at Abbott
B40	Rural primary arterial	US 84 south of Jct. NM 96
A28	Rural primary arterial	East of Raton

but dropped below this level when at seven groups. There was a significant change in the semipartial R^2 from the eighth to seventh group.

An important observation in selecting optimum group size is the cubic clustering criterion (CCC). This statistical indicator was referred to previously in discussion of the 1985 edition of the *Traffic Monitoring Guide (4)*. CCC is based on the assumption that a uniform distribution on a hyper-rectangle will be divided into same-sized hypercube-shaped clusters. Performance of the criterion, evaluated by Monte Carlo methods of

TABLE 2 MONTHLY TRAFFIC RATIOS FOR CLUSTER ANALYSIS GROUPS, BASED ON 4-YEAR AVERAGE DATA

	Group 1	Group 2	Group 3	Group 4
January	.8196	.9045	.9127	.905
February	.8284	.9292	.9301	.9195
March	.9163	.9847	.9673	.9613
April	.9078	.9937	.9959	.9911
May	1.0112	1.0284	1.0248	1.0495
June	1.1063	1.0603	1.0502	1.0622
July	1.2029	1.0669	1.0535	1.0637
August	1.1897	1.059	1.0517	1.0617
September	1.043	1.0097	1.03	1.0094
October	.9994	.9944	1.0135	1.0017
November	.9218	.9713	.9749	.9563
December	.9156	.9573	.9710	.9825

TABLE 3 MONTHLY TRAFFIC RATIOS FOR FUNCTIONAL CLASSIFICATION GROUPS, BASED ON 4-YEAR AVERAGE DATA

	Group 1	Group 2	Group 3	Group 4
January	.82036	.90406	.72369	.68703
February	.83461	.92645	.71501	.62619
March	.9136	.97577	.87733	.77827
April	.92669	.99466	.79259	.73566
May	1.03769	1.03345	.92189	.92400
June	1.11942	1.05956	1.12623	1.13344
July	1.20107	1.05946	1.43332	1.63359
August	1.18041	1.05633	1.43569	1.57033
September	1.03756	1.01803	1.04732	1.16169
October	1.01296	1.00308	.93376	1.00964
November	.9553	.96968	.78611	.79651
December	.92296	.96541	.82738	.90812

expected R^2 values, may tend to be conservative given the number of permanent counter locations being clustered *(13)*. PROC TREE was also used. This procedure creates dendograms using output from PROC CLUS.

The SAS statistics manual and CCC technical manual identify selection of the optional number of groups from a plot of the CCC. There should be one number that increases in a slight peak. In use of this statistical indicator, a plot of the CCC has proven more useful than the CCC value alone. From a plot, multiple peaking characteristics may be quickly identified that may raise questions concerning the optimum number of groups. In most instances, there is no difference between the lowest CCC value, as suggested in the *Traffic Monitoring Guide (4)*, and clusters indicated by a plot of the CCC and cluster group dendogram. When a difference exists, it is by one or two groups. The optimum number is resolved by reference to the R^2 value. In the initial cluster analysis, the CCC plot progressed consistently only through the eighth group.

Graphically, and from key statistics in the computer output using 1-year data, the identified optimum number of groups was eight. The same process of selecting optimum group size was used in all subsequent cluster analyses.

The next step in the cluster procedure was to run multiple-year data. SAS PROC CLUS and PROC TREE were used with 4-year average ADT data. Table 1 is a list of the four groups established by cluster analysis.

The MTR by group, based on 4-year average data, is shown in Table 2. There is limited application of these MTR values. An area where an MTR value is applied must have an established history to group it using the CCC.

Of the four clusters identified, two clusters had adequate or more than adequate stations for the sample. Groups 1 and 2 had an adequate sample of permanent counter locations, whereas Groups 3 and 4 did not have an adequate number of samples. Groups with an adequate sample of counters had a small standard deviation among individual counter MTR values. Where there was a small sample as in Group 3, the standard deviations increased. This fact indicated the importance of an adequate number of observations within groups formed by cluster analysis.

Group 2 is on the borderline for an adequate sample. The absence of one of these counters in 1981 caused the Group 2 ADT adjustment factor to be invalid for that year.

Next analyzed were clusters based on functional classification. This analysis established four functional classification groups, not including rural and urban Interstate. Table 3 presents MTR values for each group. The groups are shown in the following:

Group No.	*Functional Classification*
1	Rural primary arterial (Functional System Code 02)
2	Rural minor arterial and rural major collector (Functional System Codes 06 and 07)
3	Urban primary arterial (Functional System Codes 12 and 13)
4	Urban minor arterial (Functional System Code 16)

All current permanent counter stations not on the Interstate are on roadways that fit one of these categories. Stations by funcitional classification group are as follows:

Group No.	*Permanent Counter Locations*
1	C7S, A40, A41, A87, A107E, A34N, A107S, B119, B59, A59, A25W, A28, B40, A105, A113
2	A115, A32S, A119, B124, A124
3	SFC, SFD, CLOA, ALBSM
4	A10S, CBDB, A74, ALBAA

Identified in the analysis were two rural and two urban groups: rural primary arterials, rural minor arterials with major and minor collectors, urban primary arterials, and urban minor arterials with major and minor collectors. The rural road MTR values for primary arterials are distinct from MTR values for minor arterials and collectors. Other rural roads, a category used in previous studies, was not indicated. Because of the limited urban sample, use of two urban groups rather than a single group of other urban roads cannot be established with the same certainty. On the basis of the rural data, however, two urban groups are retained until further permanent counter data are available.

The MTR by functional classification corresponds closely to the ratio by cluster analysis for Groups 1 and 2. Where there is an inadequate sample (e.g., Groups 3 and 4), the MTR values are not similar. This fact reinforces the conclusion that a representative sample of no less than five, and preferably more than five, permanent counter sites is needed in each group.

There is a higher within-group standard deviation for groups based on functional classification than those based on PROC CLUS and CCC. The standard deviation is considerably higher for Group 1, only slightly higher for Group 2.

Given similarities in groups and MTR values, a major advantage of clustering by functional classification is application. Any count site can be immediately identified by functional classification, and the corresponding group MTR applied in ADT calculation. When short-term counts are taken in areas with uncertain permanent control stations, the group mean for counters with the same functional classification may be used to factor the count to ADT value and to adjust record ADT data between count years. This approach is also attractive because it is simple to apply and there is an improvement in methodology and results without adding staff time.

Clustering by functional classification provides usable results, with immediate site application for ADT calculation. A quick cluster control method can be identified to implement clustering.

The quick cluster control method is based on comparing cluster with individual permanent counter ADT adjustments. Record data are used in the comparison. In selection of comparison sites, the period between counts at a given location should be 4 to 5 years, and a new count should have been recently taken.

As has been discussed, there must be at least 3 years' data for a minimum of five permanent counts in each functional classification cluster. The usable clusters must have the same functional classification as the sites selected.

From an initial count at the comparison sites, each year the record ADT data are adjusted or factored by an individual permanent counter. After a new blanket count was conducted at the comparison points, the count was adjusted and a new ADT value established. The difference between the new count-based ADT and the ADT based on individual permanent counter factoring is the factoring error. In some instances, the difference between the factored ADT and count-based ADT is slight. In other instances, the difference is significant.

Cluster by functional classification provides a basis for factoring the ADT between blanket counts, as well as adjusting the count to ADT. At whatever year there is an available cluster factor, it is used to adjust the record ADT to the following year. The cluster factor adjustment is made from year to year, until the same comparison with recent count-based ADT can be conducted as was done with the individual permanent counter. The difference between the cluster-factored ADT and recent count-based ADT establishes the cluster factoring error for that comparison point.

The individual permanent counter and the cluster analysis factored ADT values can then be compared. The factoring errors can also be compared.

Where the difference between the cluster ADT and permanent counter ADT is less than 100 vehicles, the conclusion is that the roadway is equally controlled by both. The traffic recording devices are accurate roughly within a range of 50 veh/day. Because the error of the two counts being compared may vary in the opposite direction, improvement of less than 100 veh/day is discounted. In applying New Mexico cluster factors, there was a limited number of years' adjustment, particularly for Group 2, rural minor arterial and major collector. In the future when more years' adjustment is made, the difference between cluster and individual permanent counters will be more significant in some areas currently equally controlled. In some areas the result of more years' data will be controlled by an individual permanent counter. In other areas there will be more accurate control by cluster analysis.

When the cluster analysis ADT was closer to the recent count-based ADT and the difference was greater than 100 vehicles, the area was considered controlled more adequately by cluster analysis. When the permanent counter ADT was closer to the recent count-based ADT and the difference was greater than 100 vehicles, the area was considered controlled more adequately by individual permanent counter. Where two adjacent points of comparison on the same roadway were controlled by different procedures, a major generator or crossroad was selected between the points of comparison. This procedure was used to separate the control areas.

Comparable benefits of cluster analysis or individual permanent counter analysis can be shown in several ways. The comparisons are based on the last-factored ADT. The numeric difference between the individual counter-factored ADT and the cluster-factored ADT is determined. If the number is less than 100, the procedure stops, and the site is considered equally controlled. If the number is larger than 100, it is divided by the recent count-based ADT. The resulting percent closer to the count-based ADT that one factor method is than the other is an overall measure of accuracy.

This overall measure of increased accuracy for the factoring method represents between 2 and 3 years of ADT adjustment. The number of years of ADT adjustment is divided into the overall measure of accuracy. The result is the annual improvement in ADT accuracy one factoring method provides over the other. This annual improvement may then be multiplied by the number of years an ADT must be factored before another blanket count is taken. This procedure gives a total between-count improvement of accuracy of one method over another at a given point of comparison.

Another way in which the benefit of clustering is determined is in forecasting. The log data for a given point from year to year are used in multiple regression analysis to forecast ADT to design year. Best-fit regression analysis can be used with no R^2 value of less than 0.90. A forecast for each point of comparison is conducted, once with the permanent-counter-factored ADT values and once with the cluster-analysis-factored ADTs. The difference in results will be substantial at many points of comparison.

The cluster control method used in New Mexico is quick because it accepts initial and subsequent count factoring using the individual counter method. It also accepts whatever between-count years' data have been factored by individual counters when cluster data are unavailable. As much of the current methodology is accepted as possible, while demonstrating the benefits of clustering. This approach tends to favor the existing method. The distinction in design year projections for ADT would be more significant if the initial and all subsequent ADT values were adjusted by cluster analysis. The tendency to favor the existing methodology is acceptable because it facilitates the quick calculation of control areas, and communicates to decision makers that clustering improves accuracy.

Site-specific examples of the quick cluster control method are helpful. Two areas were selected, the city of Taos, and Chaves County, New Mexico. The sites were selected because there was a lengthy period between counts, adjusted recently on the basis of a new count.

Where clustering improved ADT factoring, typically 3 to 4 percent per year improvement in accuracy was provided. This resulted in an average of 15 to 20 percent improvement in accuracy between blanket counts. At some locations sampled, the increase in accuracy through clustering reached 8 to 10 percent per year, or 40 to 50 percent improved accuracy between counts. Where permanent counter controls had large error in adjusting ADT, the resulting forecast using inaccurate record data overestimated future demand by more than a lane of traffic. The use of record ADT data for design year forecasts makes critical the use of cluster adjustment at appropriate locations.

Use of clustering by functional classification as a standard procedure in ADT calculation and factoring can improve ADT accuracy. Use of the quick cluster control method can encourage and facilitate use of clustering in traffic volume estimation.

REFERENCES

1. *Guide for Traffic Volume Counting Manual.* U.S. Department of Commerce, Bureau of Public Roads, 1965.
2. *Guide for Traffic Volume Counting Manual.* Vol. 20, Appendix 26, U.S. Department of Transportation, March 1970.
3. *Guide to Urban Traffic Volume Counting.* U.S. Department of Transportation, Sept. 1981.
4. *Traffic Monitoring Guide.* U.S. Department of Transportation, June 1985.

5. S. K. Pant and F. J. Wegman. *Multivariate Statistical Techniques for Grouping Loadometer Stations and Development of a Computer Algorithm for Routing Trucks in West Virginia.* West Virginia University, Morgantown, 1969.
6. N. Lieder. *A Grouping Procedure.* Highway Planning Technical Report, No. 22. FHWA, U.S. Department of Transportation, March 1971.
7. S. Sharma and a Werner. Improved Method of Grouping Provincewide Permanent Traffic Counters. In *Transportation Research Record 815*, TRB, National Research Council, Washington, D.C., 1981, pp. 12–18.
8. S. C. Shah and J. J. Hirschmann. *Analysis of Traffic Count Stations to Optimize Locations and Frequency.* Louisiana Department of Transportation and Development, Baton Rouge, 1981.
9. R. L. Dreusch. Estimating Annual Average Daily Traffic From Short-Term Traffic Counts. In *Highway Research Record 118*, HRB, National Research Council, Washington, D.C., 1966, pp. 85–95.
10. D. Hartgen and J. Lemmerman. *Streamlining the Collections and Processing of Traffic Count Statistics.* Transportation Analysis Report 19. New York Department of Transportation, Albany, Aug. 1982.
11. *A Policy on Geometric Design of Highways and Streets.* AASHTO, Washington, D.C., 1984.
12. *SAS User's Guide: Statistics.* Statistical Analysis System Institute, Cary, N.C., 1982.
13. W. Sarle. *Cubic Clustering Criterion.* SAS Technical Report A108. Statistical Analysis System Institute, Cary, N.C., 1983.