TRANSPORTATION RESEARCH RECORD 1090

Statewide Data Collection and Management Systems

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1986

Transportation Research Record 1090
Price $11.20
Editor: Elizabeth W. Kaplan
Compositor: Marion L. Ross
Layout: Theresa L. Johnson

mode
1 highway transportation

subject area
11 administration

Transportation Research Board publications are available by order 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

NOTICE: The Transportation Research Board does not endorse products or manufacturers. Trade and manufacturers' names appear in this Record because they are considered essential to its object.

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

Statewide data collection and management systems.

(Transportation research record, ISSN 0361-1981 ; 1090)
1. Traffic surveys. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 1090 380.5 s 87-1862
[HE369] [388.3'14]
ISBN 0-309-04110-4

Transportation Research Record 1090

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

A Computerized Highway Link Classification System for Traffic Volume Counts

NICHOLAS J. GARBER AND FARAMARZ BAYAT-MOKHTARI

Most state transportation agencies are making a major effort to reduce their annual expenditures on traffic counts yet maintain the desired level of accuracy. These state agencies are therefore developing traffic count programs based on collecting data on a statistically selected sample of highway sections. The assumption is that if road sections can be put into groups such that each group contains sections of highways with similar characteristics, then data collected on a statistically selected sample of sections in any one group will provide traffic data representative of all sections within that group. The main variables used for this grouping are now the FHWA classification system and the average annual daily traffic (AADT) of the road. Unfortunately, the FHWA system can be subjective in cases, and in most cases the AADT of the highway section is unknown or wrong. Estimates of the coefficients of variation of the AADT of groups formed by the standard procedure have therefore tended to be high, which means that large sample sizes are needed to obtain the required accuracy. Consequently, the cost of collecting annual traffic data has not been reduced significantly. In this paper is presented a clustering technique that does not require a knowledge of the link AADTs but does require the use of certain characteristics, such as terrain, land use, and vehicle mix, which are shown to be surrogates of the AADT and can be easily obtained. The technique was used in grouping highway links in the Richmond area of Virginia, and it was found that estimates of the coefficients of variation from sample data were much lower than those recommended by the FHWA. It is concluded that the required sample sizes for annual traffic data collection are lower and this is reflected in lower costs.

Estimates of annual average daily traffic (AADT) volumes are important to the planning and operation of state highway departments. These estimates are used in planning new construction; in improving existing facilities; and, in some cases, in allocating maintenance funds. It is therefore important that any method used to obtain the estimates provide data of sufficient accuracy for the intended use. The importance of having reliable and current data on traffic volumes at hand is generally recognized, and over the years data collection programs have tended to expand. This expansion has led to large amounts of money being spent annually for the collection and analysis of traffic data. Renewed efforts are, however, now being made to reduce the annual expenditure on traffic counts yet maintain the desired level of accuracy. Most of this effort has been focused on developing statewide traffic count programs that collect traffic data on groups of statistically selected highway sections and assume that the average of each group is representative of the volume of sections that have similar traffic characteristics *(1–3)*. The first step in developing such a program requires the classification of highway sections or links into clusters such that all links within each cluster have similar traffic volume characteristics.

The primary factors commonly used for grouping highway links are the functional class of the link as given in the Highway Performance Monitoring System (HPMS) *(4)* and its AADT. The use of these factors, however, presents some problems. First, in several cases the assignment of a particular functional class may be subjective because it is sometimes difficult to differentiate between functional classes such as minor arterials and major collectors and major and minor collectors. In addition, even highways of the same HPMS functional class may not have similar traffic characteristics (e.g., seasonal variation) if they are located in different parts of a state. Second, the AADT at each link is required for that link to be properly assigned to a particular cluster. Unfortunately, in most cases the AADT is not known, and engineers then have to make assumptions based on experience. A highway link classification sytem based solely on these two factors may, therefore, give clusters or groups that include links that have different traffic characteristics. Because the accuracy of any counting program is highly dependent on developing clusters that contain only highway links with similar traffic characteristics, a suitable classification system to achieve this is essential. Such a classification system has been developed for the state of Virginia's rural highways as part of a major study to develop a statewide traffic count program.

Although the procedure was developed primarily for rural roads, it can also be used for urban roads if a set of appropriate guidelines for link identification is developed for urban roads.

DEVELOPMENT OF CLASSIFICATION SYSTEM

The classification system was developed primarily to eliminate the necessity of knowing the AADT of a link before it could be assigned to a group and to avoid sole reliance on the FHWA functional class stratification. The following steps were taken in the development of the system:

1. Link definition and identification,
2. Identification of significant variables that influence AADT, and
3. Link clustering.

Link Definition and Identification

The first step is to break down each highway in the rural area of the state into short, homogeneous sections known as highway links.

N. J. Garber, Virginia Highway and Transportation Research Council and Department of Civil Engineering, University of Virginia, Charlottesville, Va. 22901. F. Bayat-Mokhtari, Virginia Highway and Transportation Research Council, Box 3817, University Station, Charlottesville, Va. 22903.

The main requirement for a link is that each point on it have the same traffic characteristics, such as AADT and daily, weekly, and seasonal variations in traffic volume. The following three basic guidelines were used to identify a link:

1. Freeways and Interstates—Any section of the highway between any two consecutive interchanges is considered a link. This satisfies the main requirement because traffic volume changes cannot occur between consecutive interchanges on these highways.
2. Arterials—Any section of road between any two consecutive major intersections or between any two consecutive intersections if the length is 2 mi or greater is considered a link. This is based on a survey that indicated that the minimum travel distance on such facilities is usually greater than 2 mi.
3. Collectors—Any section of road between any two consecutive major intersections is taken as a link, but with the condition that a new link should start whenever there is a change in the physical appearance of the highway. For example, a new link will begin at a section where the highway changes from a two-lane undivided highway to a three-lane highway with the middle lane used for turning movements.

By using these guidelines and the road inventory mileage record file for each highway district, each rural highway was divided into a number of links.

Each link was identified by its maintenance jurisdiction (i.e., state, county, or incorporated area), route number, county in which the link is located, and a sequence number identifying all of the links belonging to the same highway and located within the same county. In addition, the length of each link is given together with its starting point and end point.

Identification of Significant Variables

Because the main objective was to develop a clustering system that does not initially require the AADT of each link, but at the same time will produce groups that consist of highway links with similar AADTs, it was necessary to identify those variables that have a significant effect on AADT so that they could be used as surrogates of AADT in the clustering system.

A detailed search of the literature indicated that the following candidate variables have some impact on AADT and other traffic characteristics:

1. Locational characteristics
   - Urban versus rural
   - Terrain
   - Area land use
2. Design characteristics
   - Number of lanes
   - Access control
   - Lane and roadway width
3. FHWA functional classification
4. Traffic composition
   - Percentage of passenger cars
   - Percentage of out-of-state passenger cars
   - Percentage of trucks with three or more axles
5. Posted speed limit

To identify which of these variables have a significant effect on AADT, statistical tests were carried out using AADT data for 1977 through 1980 obtained at 112 permanent count stations in Kansas, Maryland, and Virginia. Before any statistical test was carried out, however, it was decided to combine all of the variables under Item 4 into a single variable, defined as "functional use," by considering the individual variables in the following manner:

PC_{ij} = percentage of passenger vehicles in total traffic on link i in state j;
POC_{ij} = percentage of out-of-state passenger vehicles in passenger vehicle traffic on link i in state j;
PHT_{ij} = trucks with three or more axles as a percentage of total traffic on link i in state j;
$\overline{PC}_j, \overline{POC}_j, \overline{PHT}_j$ = average of PC_{ij}, POC_{ij}, and PHT_{ij}, respectively (i.e., state average for PC, POC, and PHT); and
$SPC_j, SPOC_j, SPHT_j$ = standard deviations of PC_{ij}, POC_{ij}, and PHT_{ij}, respectively.

Limits of the mean for a state plus or minus one standard deviation were used to determine whether a particular variable for a given link was high, average, or low with respect to that of the state in which the link is located. For example,

if $PC_{ij} > \overline{PC}_j + SPC_j$, the link PC is high;
if $PC_{ij} < \overline{PC}_j - SPC_j$, the link PC is low; and
if $\overline{PC}_j - SPC_j < PC_{ij} < \overline{PC}_j + SPC_j$, the link PC is average.

Similar limits were defined for POC_{ij} and PHT_{ij}. Table 1 is a matrix of the predominant combinations into which a given link fell and the five types of functional uses that were obtained. These are referred to in this study as recreational, local service, long-distance service, industrial, and local commercial links.

TABLE 1 FUNCTIONAL USE CLASSIFICATION

PC_{ij} (%)	POC_{ij} (%)	PHT_{ij} (%)	Functional Use
High	High	Low	Recreational
High	Low	Low	Local service
High	Average	Low	Long-distance service
Average or low	Low	High	Industrial
Average or low	Low	Average or low	Commercial

- Recreational links—Links in this category have a relatively high volume of out-of-state passenger cars, which may easily be affected by seasonal factors. The exception to this is links located in the vicinity of state boundaries. In general, seasonal characteristics have a significant impact on traffic volume on these links.
- Local service links—These links are used mainly by residents of the area for commuter trips and exhibit relatively little variation in traffic volume throughout the year.
- Long-distance service links—The traffic characteristics of these links are similar to those of the local service links, but they contain a larger portion of long-distance commuter trips.
- Industrial links—These have a relatively high percentage of

heavy trucks and are typically on highways connecting major industrial cities.

- Local commercial links—These links have an average percentage of trucks with three or more axles but a relatively high percentage of pickups. Business trips are likely to be predominant on these links.

The set of candidate variables was further reduced by discarding the locational variable (urban versus rural), because all links considered were located in rural areas, and by discarding access control, because this is somewhat related to the FHWA functional classification system.

The remaining candidate variables were then tested for significant effect on AADT using an analysis of variance (ANOVA). The significant variables identified were

- FHWA functional class,
- Functional use as described in this paper,
- Land use of the county in which the link is located,
- Population of the county in which the link is located, and
- Type of terrain.

These variables were therefore used in the clustering technique described in the following subsection.

Link Clustering

McQueen's K-means method was used as the clustering technique *(5)*. This technique provides for the assignment of a set of m data units to a number of clusters such that data elements within any given cluster are "similar to" or "near" each other. The first K data units are initially selected as K clusters of one member each. The distances between all paired combinations of the K clusters are then computed. If the smallest distance is less than a predetermined minimum (c), the two associated clusters are merged. The centroid of the new cluster is determined and the process is repeated until the distance between the centroids of any two clusters is greater than c. The distances between each of the remaining (m–K) data units and the centroids of each of the clusters already formed are computed and each data unit is assigned to the cluster with the nearest centroid (i.e., minimum d) if this distance is less than c. After each assignment, the centroid of the gaining cluster is computed. If the distance to the nearest centroid of any data unit is greater than a refining parameter (R) where $R \geq c$, then that data unit is taken as a separate cluster.

In this study the squared Euclidean distance in n dimensional space, defined by Equation 1, was used as the basis for representing "similarity" or "nearness" among the data units.

$$d2_{ij} = \sum_{h=1}^{n} (x_{ih} - s_{jh})2 \qquad (1)$$

where

$d2_{ij}$ = squared Euclidean distance,
x_{ih} = value of variable h for case i,
x_{jh} = value of variable h for case j,
$d2_{ij}$ = squared Euclidean distance between case i and j, and
n = number of variables

The number of clusters obtained is dependent on the values of c and R. In this study R was taken as equal to c. The links were initially clustered for a minimum value of 0.01 for c, which gave the largest number of clusters. The number of clusters was then gradually decreased by gradually increasing the value of c.

It can be seen that the variables identified as significant are mainly nominal variables (e.g., terrain and land use) and cannot, for this reason, be used directly in Equation 1. It was therefore necessary to convert these nominal variables to interval variables. The procedure adopted to achieve this was to represent each category of the nominal variable by 1 percent of the average AADT of the continuous count stations in that category. This provides for a common measure of all variables, and at the same time employs the relative impact of each category of each of those variables on the AADT. As an example the computation carried out for terrain is shown in Figure 1. Based on these computations,

	Type of Terrain		
	Flat	Rolling	Mountainous
Sample Size	37	51	24
$\bar{x}$ AADT (1979)	3035	4831	5071
$\bar{x}$ AADT (1980)	2976	4674	5179
$\bar{\bar{x}}$	3006	4753	5095
Code	30	48	51

FIGURE 1 Sample computation for converting nominal variables to interval variables (from continuous count stations in Maryland, Kansas, and Virginia).

the following codes were used to represent the x_{ih}'s of the significant variables.

- Terrain
  - Flat, 30
  - Rolling, 51
  - Mountainous, 24
- Population of county in which link is located
  - <10,000, 20
  - 10,001 to 20,000, 42
  - 20,001 to 30,000, 38
  - 30,001 to 40,000, 46
  - 40,001 to 50,000, 93
  - 50,001 to 100,000, 68
  - >100,000, 74
- Land use
  - Agricultural, 23
  - Industrial, 63
  - Service, 30
  - Mining, 40
- Functional use
  - Recreational, 31
  - Local service, 59
  - Long-distance service, 37
  - Local commercial, 25
  - Industrial, 10
- FHWA classification
  - Interstates, 93
  - Principal arterials, 42
  - Minor arterials, 31
  - Collectors, 19

TABLE 2 WEIGHTING FACTORS FOR SIGNIFICANT VARIABLES

Variable Type	F to Remove Based on AADT Data for 1977	1978	1979	1980	Average F	Weighting Factor
Functional use	11.00	11.86	10.08	11.9	11.38	12
FHWA functional classification	13.03	11.86	12.24	9.89	11.01	11
Population of county	4.4	5.09	5.2	4.68	4.84	5
Terrain	2.97	2.98	2.38	2.69	2.76	3
Land use	0.33	0.70	0.74	0.68	0.71	1

These values may be considered representative and can be used by any state because they were computed from data obtained at continuous count stations located in different parts of the country. On the other hand, specific values for a particular state may be computed if there are an adequate number of continuous count stations located in the state.

Another problem that had to be overcome before the technique could be used is related to the differences in the order of magnitude and dispersion among the values of the n variables. It was noted that, if the order of magnitude of the range of values that a particular variable takes is much larger than the range of values for other variables, the value of the "Euclidean distance" between data units will be based on that variable. To overcome this problem, the values of the variables were standardized using Equation 2.

$$Z_{ih} = (x_{ih} - \bar{x}_h)/SD(x_h) \tag{2}$$

where

Z_{ih} = standard value for variable h for case i,
x_{ih} = mean value of variable h for case i,
$\bar{x}_h$ = mean value of variable h, and
$SD(x_h)$ = standard deviation of variable h.

It was also noted that the variables used do not all have the same degree of impact on the AADT of a given link. Therefore a weighting factor had to be included for each variable. A stepwise regression analysis was used to determine the relative influence of each variable on the AADT by assigning the average value of "F to remove" of each variable as the weighting factor for that variable (Table 2). These factors may be used, or may be determined for a given state using the same procedure if data are available at an adequate number of permanent count stations.

The squared Euclidean distance used in the McQueen's K-means technique is therefore

$$d^2_{ij} = \sum_{h=1}^{m} W_h(Z_{ih} - Z_{jh})^2 \tag{3}$$

where W_h is the weighting factor for variable h and m is the number of variables.

A FORTRAN computer program was written for executing the whole procedure based on the flowchart shown in Figure 2.

RESULTS

The methodology was tested using the Interstate, arterial, and collector roads in the Richmond district. At total of 363 highway links were obtained using the guidelines presented earlier. Figure 3 is a plot of the number of clusters versus c and indicates that the rate of increase of the number of clusters is relatively high for values of c less than 2.5 and low for values of c greater than 6.5. These values correspond to nine and four clusters, respectively, and suggest that a reasonable number of clusters is between four and nine.

Because the coefficient of variation of the AADTs within any given cluster is an indication of how successful the clustering procedure is, data on average daily traffic were collected on a sample of links in each cluster for a system of eight clusters, and the coefficients of variation were estimated for each cluster. The results obtained, given in Table 3, indicate that all of the coefficients of variation were lower than the recommended FHWA values.

CONCLUSION

It has been demonstrated that the clustering procedure presented in this paper can be used to form groups of highway links with similar traffic characteristics, without the necessity of first assigning a value for the AADT of each link. The coefficients of variation of the average daily traffic obtained for a test run in the Richmond district show that coefficients well below those recom-

TABLE 3 ADT ESTIMATED COEFFICIENTS OF VARIATION FOR EACH CLUSTER OF A CLUSTERING SYSTEM OF EIGHT CLUSTERS IN THE RICHMOND DISTRICT

Cluster No.	Predominant Type of Highway	No. of Links Sampled	ADT	COV
1	Interstate	5	8,601	0.13
2	Principal and minor arterials	8	6,500	0.16
3	Interstate	9	13,910	0.20
4	Major collectors	6	6,630	0.14
5	Principal and minor arterials	7	12,737	0.16
6	Major collectors	10	3,584	0.18
7	Interstate	8	33,799	0.38
8	Minor arterials	8	1,708	0.18

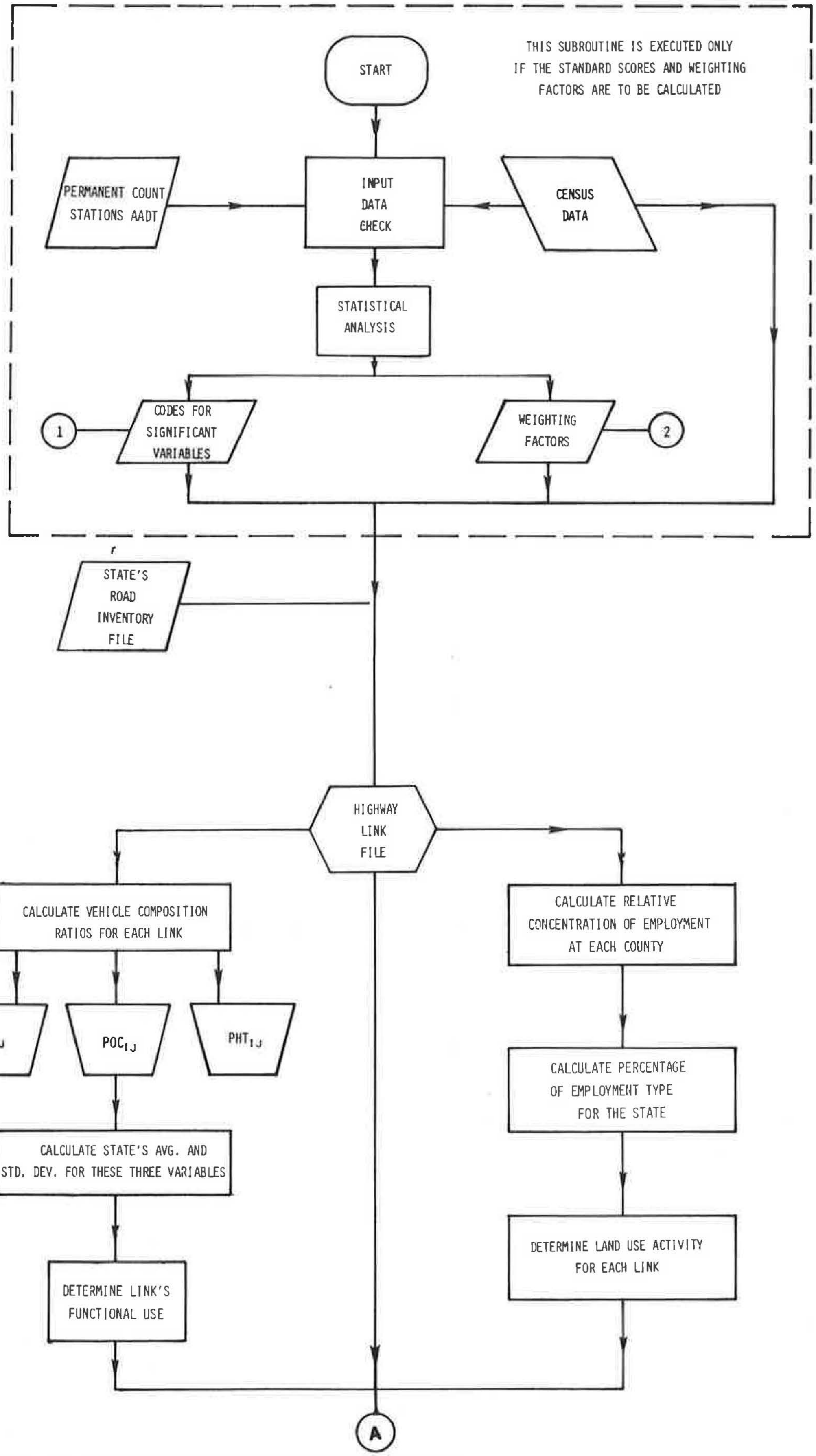


FIGURE 2 Schematic representation of proposed highway classification algorithm.

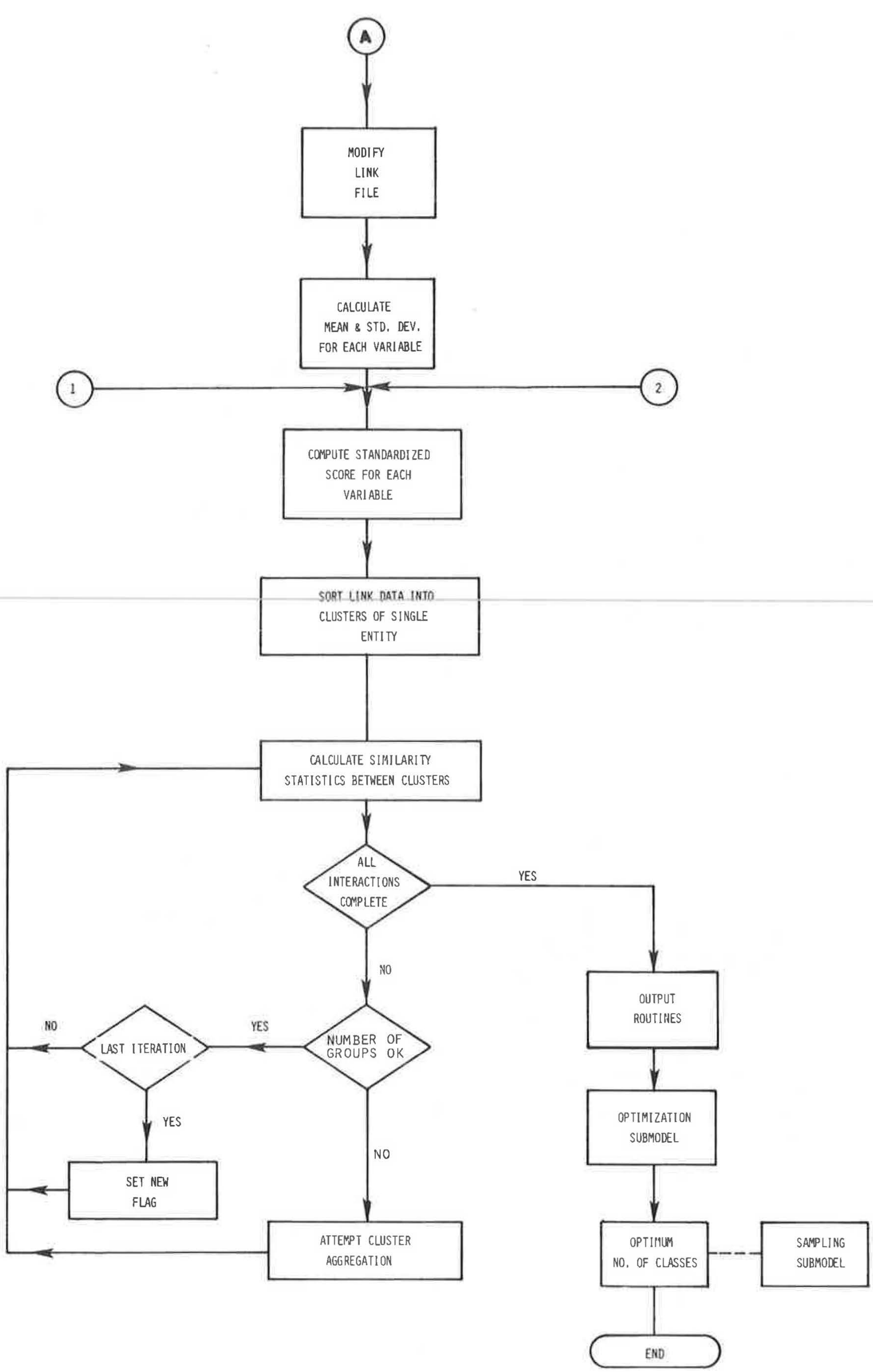


FIGURE 2 *continued*

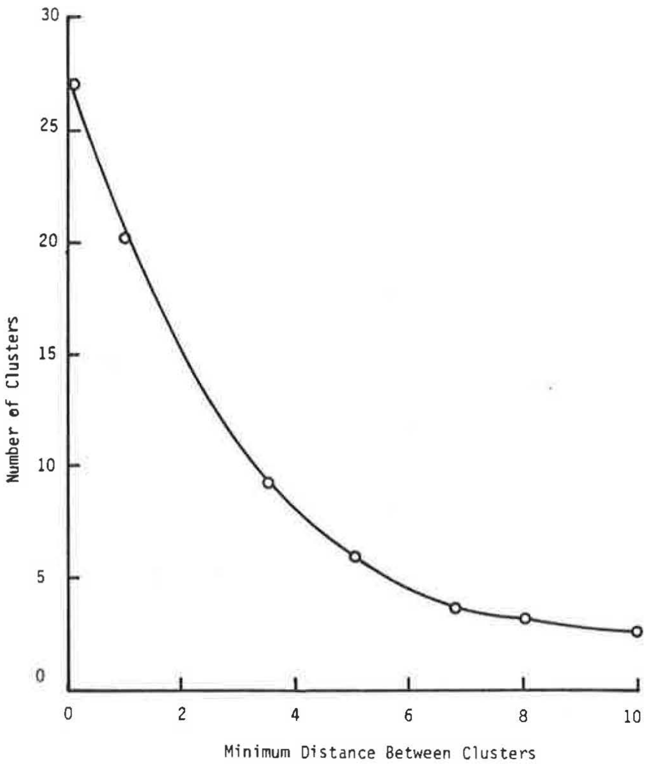


FIGURE 3 Number of groups versus *c* for Richmond district.

mended by the FHWA can be obtained. The procedure requires that all highways being considered for grouping be divided into homogeneous links such that the traffic volume characteristics on any one link do not vary along that link. Input data for each link include the population of the county in which the link is located, the terrain of the area, the land use, and vehicle type. This clustering procedure will develop groups of highway links such that volume data collected at a statistically selected sample of links within a given group will give a good indication of the average AADT of the links in that group. Because the coefficients of variation of the AADTs in each group will be low, when obtained by this procedure, the number of links required to be sampled for a given level of accuracy will be relatively small and will therefore result in cost savings.

REFERENCES

1. *Guide to Urban Traffic Volume Counting.* FHWA, U.S. Department of Transportation, Sept. 1981.
2. *Development of a Statewide Traffic Counting Program Based on Highway Performance Monitoring System.* FHWA, U.S. Department of Transportation, March 1984.
3. P. Garwyn. *Accuracy of Annual Traffic Flow Estimates from Automatic Counts.* TRRL Supplementary Report 515. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
4. *Highway Performance Monitoring System.* FHWA, U.S. Department of Transportation, Sept. 1980.
5. W. J. Dixon et al. *BMDP Biomedical Computer Program P.* University of California Press, Los Angeles, 1979.

Trip Generation for Special-Use Truck Traffic

DAN R. MIDDLETON, JOHN M. MASON, JR., AND T. CHIRA-CHAVALA

Special-use truck traffic is the traffic associated with the processing and transporting of timber, grain, beef cattle, cotton, produce, sand and gravel, and limestone. Industry and vehicle characteristics for each of these six commodities were determined. The impact of each special-use activity center was assessed in terms of trip generation. Specific activity centers were selected for each industry. Number of trips generated, radius of influence, loads, vehicle configuration, and seasonal variations were determined for each selected activity center through agency and industry contacts and field studies.

The term "special-use" has been coined to designate truck traffic that has atypical travel patterns, trip lengths, truck configurations, and axle loads. The travel patterns of these vehicles tend to be cyclic in nature; in some cases the trip is made several times in a typical day. Trip lengths are relatively short, usually less than 100 mi. The origin and destination may remain the same month after month, but eventually either the origin or the destination will change. Axle loads, although generally not well documented, are in many cases greater than normally expected. Trips generated by these special-use activity centers pose problems for the planning, design, and maintenance of the highways that serve their needs.

To determine the definitive elements of these isolated traffic demands, the Texas State Department of Highways and Public Transportation (SDHPT) initiated a 4-year study to evaluate the impact of special-use truck traffic. The predecessor to this study was a comprehensive evaluation of the effects of oil field development on roadways in the state of Texas *(1)*.

The special users identified in this study fall into the two broad categories of agriculture and surface mining. A list of specific commodities was refined as industry characteristics were determined; the selected commodities are

Agriculture	*Surface Mining*
Timber	Sand and gravel
Grain	Crushed stone
Beef cattle	
Cotton	
Produce	

METHODOLOGY

Four basic steps were followed to accomplish the objectives of the study:

1. Select special-use industries,
2. Determine industry characteristics,
3. Determine vehicle characteristics of selected industries, and
4. Determine trip-making characteristics.

D. R. Middleton and T. Chira-Chavala, Texas Transportation Institute, Texas A&M University System, College Station, Tex. 77843-3135. J. M. Mason, Jr., Post, Buckley, Schuh & Jernigan, Inc., 5405 Cypress Center Drive, Suite 200, Tampa, Fla. 33609-1066.

Select Special-Use Industries

Selection of special-use industries began with the identification of industries whose specific commodities and activity centers uniquely affected the highway system in Texas. The activities surrounding oil field development and production were studied initially; agricultural product movement and quarrying and mining remain as unique special-use generators.

The list of specific commodities selected for study was refined as industry characteristics were determined. Some commodities were found to be more significant than others. Evaluation of the impacts of activity centers for special-use commodities such as uranium ore and poultry showed a relatively small number of trips generated in comparison with other commodities. In addition, poultry was not a weight-intensive (high-density) commodity.

Determine Industry Characteristics

Several public agencies were contacted to acquire available information about the selected commodities. For the timber industry, for example, the Texas Forest Service and the Forest Service provided printed information, maps, and names of private firms.

Site visits to various activity centers, which included interviews with key industry personnel, were conducted. Activity centers are defined as points where commodities are processed or handled. These centers often served as focal points for mode transfer.

The processing-activity phase of special-use commodities usually had more than one activity center that could be chosen for evaluation. Therefore a selection process involving the following criteria was established to identify the appropriate activity center.

1. The site must be a "primary" operation in total processing of the commodity,
2. A "significant" number of trips must be generated by the commodity, and
3. The commodity must represent a fairly widespread problem in the state.

Although these criteria were not easy to quantify, they were suitable for establishing the primary processing point of the identified commodity. The selected activity centers for the chosen commodities are as follows:

Commodity	*Activity Center*
Timber	Mills
Grain	Elevator
Beef cattle	Feedlot
Produce	Distributor
Cotton	Gin
Sand and gravel	Pit
Crushed stone	Quarry

Telephone interviews were also used to supplement on-site interviews. These interviews often yielded a range of answers to a standard set of questions depending on the specific industry's size, differing climates, differing harvesting or mining techniques, effects of rail, and economic conditions during the life of each company.

The disparity of information received verbally indicated that field surveys were necessary to supplement the on-site office and telephone interviews. A comprehensive data collection plan was developed for this purpose. State maps similar to Figure 1 were used to depict the location of activity centers and the intensity of activity for a particular commodity. These exhibits were supplemented by site-specific maps from the Texas Forest Service *(2)* or lists of activity centers *(3)*, which provide the following information for each county: name of firm, mailing address, telephone number, and in many cases an indication of size.

Determine Vehicle Characteristics Associated with Selected Industries

Methods used to acquire information about vehicles used in the special-use industries were telephone requests for literature from Texas truck and trailer dealerships, office and field interviews of industry personnel, vehicle classification counts at activity centers, information from other ongoing truck-related research, and information from state departments such as the Texas Department of Public Safety (DPS), License and Weight Division. The vehicular information gathered included AASHTO classification, vehicle dimensions, engine and drive train characteristics, load-carrying capacities, typical axle loads, and percentage vehicle distribution. Table 1 gives vehicle dimensions, vehicle descriptions, and carrying capacities for selected commodities.

A sample of axle weights was collected as part of another

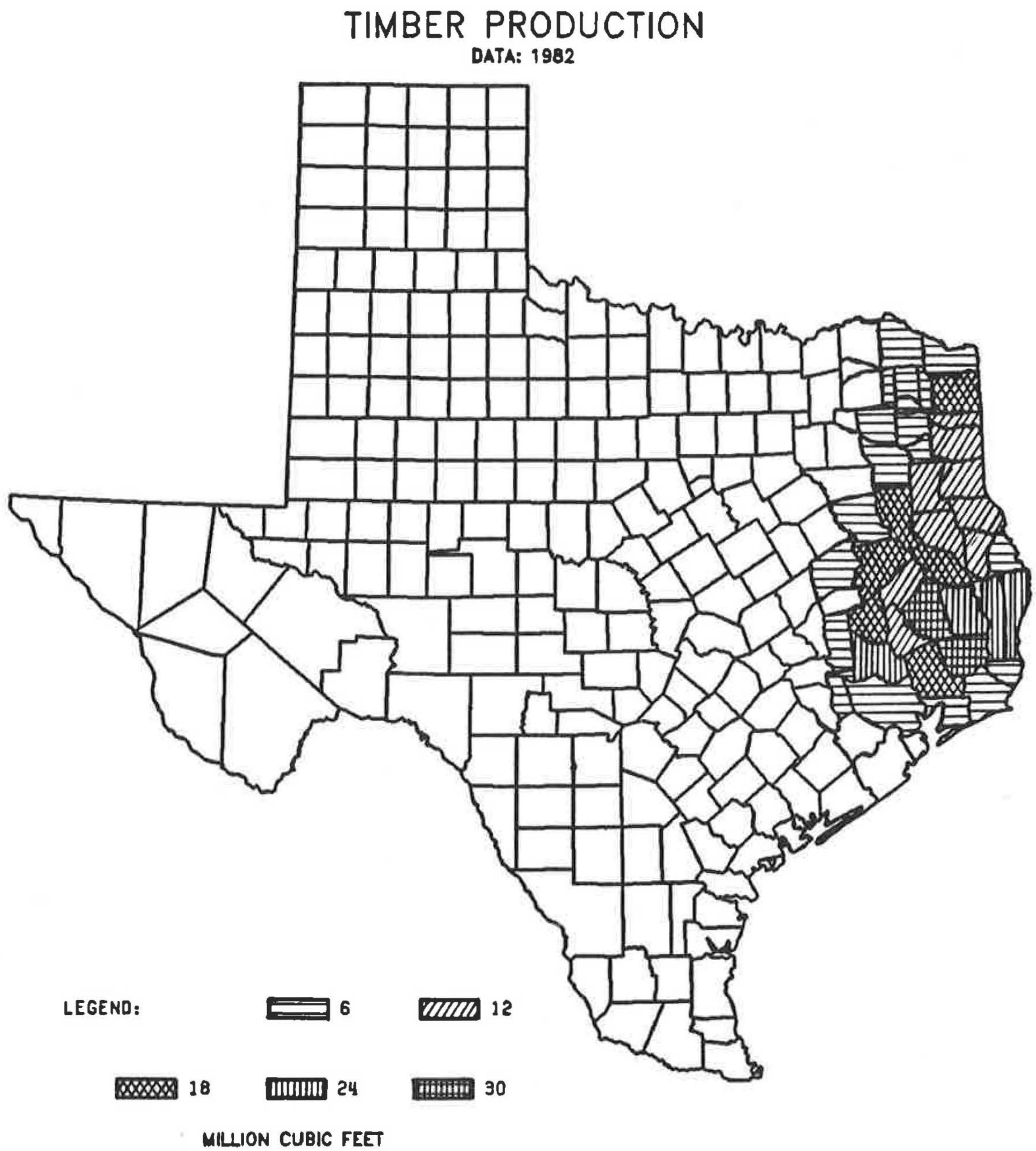


FIGURE 1 Location and intensity of timber operations.

TABLE 1 TYPICAL TRAILER AND TRUCK BED DIMENSIONS

Commodity	Vehicle Description	Width (ft)	Length (ft)	Overall Height (ft)
Timber	3-S2 fold-up	8.0	35–45	—[a]
	3-S2 pulpwood	8.0	35–40	—
	3-S2 (chip) van	8.0	38–40	12.8
	3-S2 flatbed	8.0	32–45	4.7[b]
	SU-1	8.0	8–10	9.6
	SU-2	8.0	10–14	9.6
Grain	SU-1	8.0	8–10	9.6
	SU-2	8.0	10-14	9.6
	3-S2 grain	8.0	39–42	9.6
Beef cattle	3-S2 possum belly	8.0	44–50	13.5
	3-S2 grain	8.0	39–42	9.6
	SU-2 grain	8.0	10–14	9.6
Produce	SU-2	8.0	10–14	9.6
	Tractor/field trailer	7.0	10–12	8.0
	3-S2 reefer	8.0	30–50	12.5–13.5
Cotton	SU-2 module	8.0	37.5	13
	Field trailer	7.0	—	—
	3-S2 flatbed	8.0	32–45	4.7[b]
	3-S2 van	8.0	38–50	12.8
Sand and gravel	SU-2 with pup	8.0	24–28	8
	3-S2 dump	8.0	24.3–35	8.9–9.7
Crushed stone	SU-2 with pup	8.0	24–28	8
	3-S2 dump	8.0	24.3–35	8.9–9.7

[a]Data unavailable.
[b]For flatbeds, the height is the floor level.

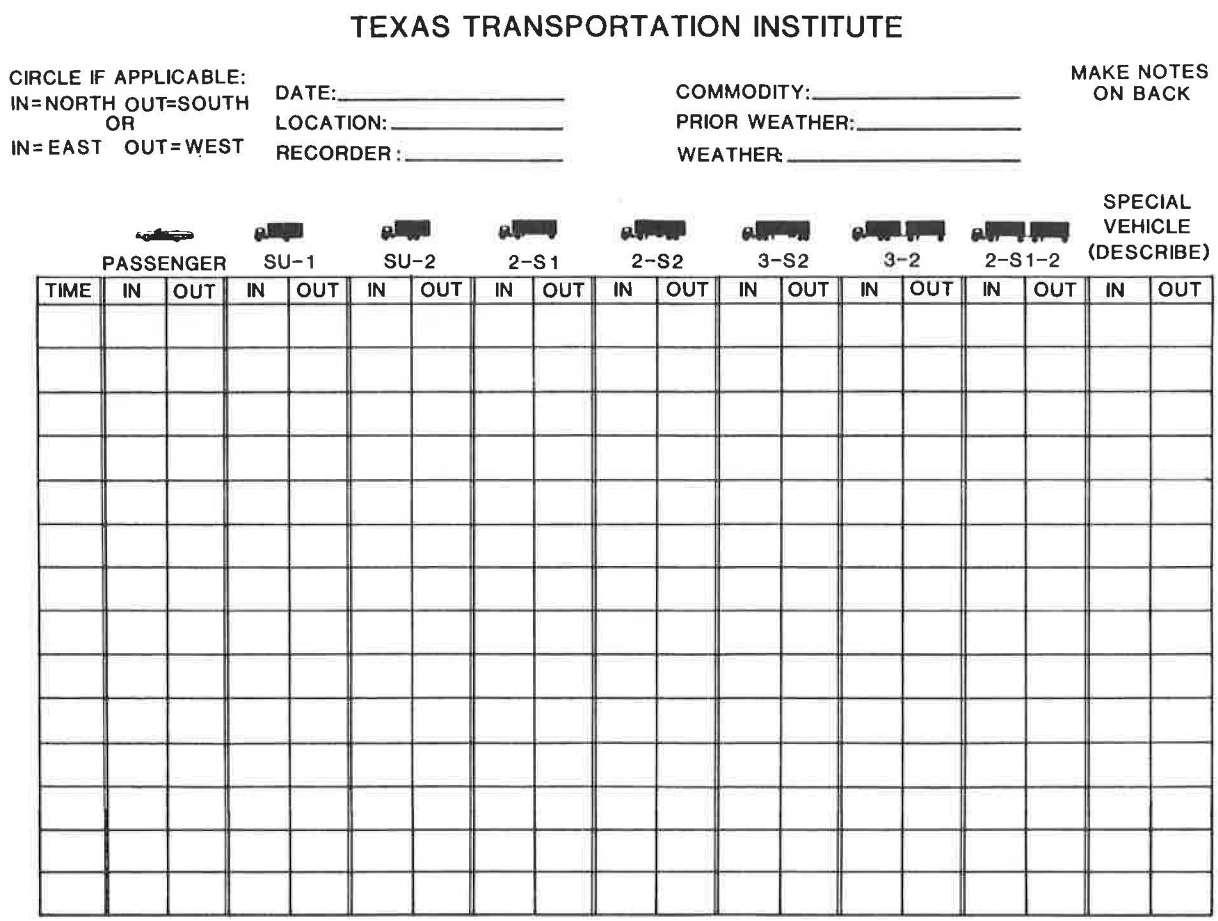
research project that evaluated truck tire pressures on Texas highways *(4)*. The procedure involved project staff working with DPS License and Weight personnel at weigh strips or other acceptable locations and using semiportable scales in an ongoing enforcement effort. Typical weights of special-use vehicles determined during a 2-year period in areas near the special-use activity centers are given in Table 2.

Determine Trip-Making Characteristics

Specific information sought was radius of influence and trip generation rates. Radius of influence represented the maximum distance from an activity center at which vehicular traffic is generated. For trucks, it is usually thought of as the haul distance from the loading site (timber cutting site) to the load destination or unloading site (timber mill).

A range of values for both radius of influence and trip generation rates was gathered from several on-site office interviews. These trip generation figures were supplemented by manual and machine traffic classification counts at selected activity centers. Trip-making characteristics as well as typical vehicle weights determined by office interviews are given in Table 2. A summary of vehicle classification information acquired through office and field interviews is given in Table 3.

To determine which activity centers to study throughout the state, a random selection procedure was used. This involved



TEXAS TRANSPORTATION INSTITUTE

CIRCLE IF APPLICABLE:
IN=NORTH OUT=SOUTH
OR
IN=EAST OUT=WEST

DATE:
LOCATION:
RECORDER:

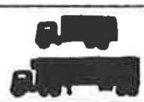
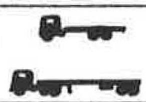
COMMODITY:
PRIOR WEATHER:
WEATHER:

MAKE NOTES ON BACK

	PASSENGER		SU-1		SU-2		2-S1		2-S2		3-S2		3-2		2-S1-2		SPECIAL VEHICLE (DESCRIBE)	
TIME	IN	OUT	IN	OUT	IN	OUT	IN	OUT	IN	OUT	IN	OUT	IN	OUT	IN	OUT	IN	OUT

FIGURE 2 Standard manual count form.

TABLE 2 SPECIAL-USE COMMODITY SUMMARY FROM OFFICE INTERVIEWS

COMMODITY (Activity Center)	COMMON TRUCK SILHOUETTES	AVG. DAILY TRUCK TRIPS(a)		TYPICAL RADIUS OF INFLUENCE (MILES)	TYPICAL WEIGHT RANGE (1000 lb) (b)		PEAK SEASON
		AVG	LG		TANDEM AXLE	GROSS VEHICLE	
TIMBER MILLS - Paper		350	550(c)	50	32-38	74-86	Mar.-Nov.
- Plywood		150	350	50	32-38	74-86	Mar.-Nov.
- Particle Board		150	300	50	32-38	74-86	Mar.-Nov.
- Sawmill		150	250	50	32-38	74-86	Mar.-Nov.
GRAIN Elevator		200	400(c)	20	32-40	76-90	May-July
BEEF CATTLE - Feedlot		60	90	600	32-35	76-82	Yr. Round
PRODUCE - Distributor		200	500	20	32-36	72-80	Mar.-Apr.
COTTON - Gin		40	110	20	30-36	72-82	Sep.-Dec.
SAND/GRAVEL - Pit		600	1,100(c)	60	32-36	70-80	Mar.-Nov.
LIMESTONE - Quarry		1,400	2,800(c)	120	32-36	70-80	Mar.-Nov.

(a) One-way trips, i.e. one origin, one destination
(b) Based on experience, conversation, limited weight information, other research projects.
(c) Truck trip generation depends on percent rail. Rail is assumed to have negligible influence at these activity centers.

developing sample plans for the commodities of interest in such a way that the truck trip generation factors obtained would represent the entire state and activity centers of varying sizes. In most cases, a two-stage process was used to select activity centers for manual and machine counts. The first step involved random selection of counties in which the commodity was produced. The second stage involved a random selection of activity centers from the selected counties.

The number of activity centers selected throughout the state was

Commodity	*No. of Sites*
Timber	13 (mills)
Grain	12 (elevators)
Beef cattle	10 (feedlots)
Cotton	12 (gins)
Produce	6 (distributors)
Sand and gravel	15 (pits)
Limestone	15 (quarries)

Of a total of several hundred possible activity centers statewide, 83 were selected for observation. A manual count procedure was used almost exclusively because of the difficulty of finding automated count stations that would clearly represent only traffic generated by the activity center.

The standard manual count form is shown in Figure 2. Information recorded included date, time, and number of vehicles entering and exiting the site by AASHTO classification. Other information recorded during the count was location, name of recorder, weather on count day and 2 days before, and additional information gathered through interviews of industry personnel.

For each site selected, a vehicle classification count was made using 15-min intervals for all traffic entering and leaving the facility during a total time period of 1 day. This meant on-site observation at any given site for from 8 to 18 hr. Typical AASHTO vehicle classifications used were PC (passenger car); SU-1 (single-unit truck with two axles); SU-2 (single-unit truck with three axles); SU-2 with pup (SU-2 pulling two-axle trailer, surface mining applications); 2-S1 (two-axle tractor, one-axle semitrailer); 2-S2 (two-axle tractor, two-axle semitrailer); 3-S2 (three-axle tractor, two-axle semitrailer); 3-2 (three-axle truck, two-axle trailer); and 2-S1-2 (two-axle tractor, one-axle semitrailer, two-axle trailer).

TABLE 3 SPECIAL-USE TRIP GENERATION FROM INTERVIEWS (peak season)

Location (activity center)	No. of Passenger Cars per Day	Percentage of One-Way Truck Trips per Day: Single Unit	3-S2	Other Trucks	Maximum Daily Truck Trips[a]
Timber mills					
Pulpwood	1,800	21	75	4	550
Plywood mill	650	5	90	5	325
Particle board	650	12	84	4	300
Large sawmill	—[b]	4	80	16	250
Average sawmill	240	7	80	13	150
Grain elevator	8–10	88	12	0	400
Beef cattle					
Large feedlot	110–140	11	89	—	90
Average feedlot	50–60	16	84	—	50
Produce					
Large distributor	400–500	40	40	20	500
Average distributor	120	23	59	18	220
Cotton gin	—	65	13	22	110
Sand and gravel					
Large pit	140	5	95	—	1,000
Average pit	—	5	95	—	550
Limestone					
Large quarry	800	5	95	—	2,000
Average quarry	300	5	95	—	1,000

[a]One-way trips (i.e., one origin and one destination).
[b]Data unavailable.

TABLE 4 MANUAL CLASSIFICATION TRAFFIC COUNTS

Activity Center	Size	Average Percentage of Combination Trucks	Average Percentage of Single-Unit Trucks	Total Truck Trips[a]
Timber mills				
Pulpwood mill	Large	83	17	291–435
Plywood mill	Average	80	20	64
	Large	92	8	196–281
Particle board mill	Large	83	17	305–362
Sawmill	Small	54	46	65
	Average	77	23	82
	Large	79	21	161–264
Grain elevator	Average	24	76	133–313
	Large	58	42	349–570
Produce distributor	Small	24	76	23–34
	Average	69	31	125
	Large	44	56	340–379
Sand and gravel pit	Small	25	75	58–128
	Average	92	8	97–137
	Large	85	15	240–775
Limestone quarry	Small	64	36	42–63
	Average	12	88	122–194
	Large	60	40	147–474

Note: Based on preliminary survey data, subject to change.
[a]One-way trips—one origin and one destination (entering plus exiting).

Results of these classification counts are given in Table 4. The reported values are initial counts for 1 day at fewer than the total number of selected sites. Differences between the values quoted in interviews and the actual field counts were expected. Additional site-specific classification counts will be conducted in future years. However, several factors must be recognized in dealing with special-use commodities. Inclement weather such as heavy rain often slows processing of such commodities. Fluctuations in the demand for a commodity such as crushed stone in a particular geographic area also affect production rates. Another noteworthy point is that interview information was not meant to be precise; an approximate range of values was sought for comparison.

SUMMARY AND CONCLUSIONS

Special-use truck traffic as considered herein involves the traffic associated with the transporting of timber, grain, beef cattle, cotton, produce, sand and gravel, and limestone. This traffic is likely to be unique in vehicle distribution, axle configuration, axle loads, and seasonal fluctuations. Trips generated by special-use activity centers pose problems in the planning, design, and maintenance of the highways that serve their needs.

The impact of the various special-use activity centers must be evaluated in terms of automobile and truck trips generated per unit time, radius of influence, and seasonal fluctuation. Trip generation rates in the range of from 100 to 400 trips per day were found at many activity centers. The radius of influence or the haul distance of these trucks is usually in the range of from 20 to 100 mi. The peak period of haul in the state of Texas for most of these commodities is March through November.

Vehicle classifications by interview and field counts indicated that the predominant AASHTO classification was 3-S2. 3-S2s were usually more than 80 percent of the total truck traffic generated by the activity centers surveyed. Single-unit trucks were also found in all commodity movements evaluated, and in larger numbers at grain elevators, produce distributors, and cotton gins.

Trip generation rates are currently lacking at industrial sites. Site-specific, special-use truck traffic information is so scarce as to be practically nonexistent. The vehicle classification, traffic count, and commodity movement information provided in this paper begins to fill the void in current trip generation data. At least 2 more years of field counts are planned at the selected activity centers. Annual and seasonal variations are anticipated; economic shifts may also alter the initial findings substantially. The results nonetheless provide guidance for estimating the magnitude of the impact of the identified special-use traffic generators.

ACKNOWLEDGMENT

Information presented in this paper represents the initial efforts of an ongoing study. The methodology and findings reflect the characteristics of the particular commodities found in Texas.

This phase of the project was sponsored by the Texas State Department of Highways and Public Transportation and the Federal Highway Administration. The views, interpretations, analyses, and conclusions expressed or implied in this paper are those of the authors. They are not necessarily those of the sponsors.

REFERENCES

1. J. M. Mason, B. E. Stampley, and H. C. Petersen. *The Effects of Oil Field Development on Flexible Pavement Networks.* Research Report 299–6. Texas Transportation Institute, Texas A&M University, College Station, Aug. 1984.
2. D. J. Skove. *Harvest Trends 1983.* Publication 134. Texas Forest Service, Texas A&M University, College Station, 1982.
3. *Texas Forest Industries Directory 1982.* Texas Forest Service, Texas A&M University, College Station, 1982.
4. F. L. Roberts, J. T. Tielking, D. Middleton, R. L. Lytton, and K. Tseng. *Effect of Tire Pressures on Flexible Pavements.* Research Report 372–1F. Texas Transportation Institute, Texas A&M University, College Station, Dec. 1985.

A Statistical Approach to Statewide Traffic Counting

STEPHEN G. RITCHIE

A statistical framework that can be used for analysis of statewide traffic count data is described. A basis for designing a streamlined and cost-effective statewide traffic data collection program is also provided. The procedures described were developed as part of an in-depth evaluation study for the Washington State Department of Transportation. They were used to develop recommendations for an improved, statistically based, statewide highway data collection program. The program is intended to be implemented readily and is consistent with the FHWA Highway Performance Monitoring System and the recent FHWA draft Traffic Monitoring Guide. Several modifications (improvements) to the statistical framework of the latter for volume counting and vehicle classification were investigated, particularly methods of deriving estimates of annual average daily traffic (AADT) from short-duration axle counts at any location on the state highway system. AADT estimates can be derived for each vehicle type, if desired. The estimation of associated seasonal, axle correction, and growth factors is also described. The methodology enables the statistical precision of all estimates to be determined. The results obtained from applying these procedures to Washington State traffic data are presented.

For many years state departments of transportation (DOTs) have had responsibility for collecting a large amount of highway data. This has been undertaken to assist planning, design, and operations functions, as well as to comply with requirements and needs of other agencies including those at the federal level. However, collection of large amounts of data is costly. In a climate of increasing fiscal austerity at all levels of government and in all program areas, it is important not only that the right type of data is collected but that data are collected efficiently. Moreover, the data should meet the needs of the users with respect to type, amount, form, accuracy, and availability. A statewide highway data collection program should satisfy these criteria in an up-to-date and cost-effective manner.

In this paper a statistical framework that can be used for analysis of statewide traffic count data is described, and a basis for designing a streamlined and cost-effective statewide traffic data collection program is provided. The procedures described were developed as part of an in-depth evaluation study for the Washington State Department of Transportation (WSDOT) and were used to develop recommendations for an improved, statistically based, statewide highway data collection program (see paper by Ritchie and Hallenbeck in this Record).

Several studies have been reported in recent years that relate to general efforts to develop more cost-effective approaches to statewide highway data collection. These include the work of Hallenbeck and Bowman *(1)*, who proposed a general statewide traffic-counting program based on the Highway Performance Monitoring System (HPMS) *(2)*; the study by Wright Forssen Associates *(3)*, which evaluated, and developed improvement recommendations for, the highway data program of the Alaska Department of Transportation and Public Facilities; and work by the New York State Department of Transportation to streamline and reduce the cost of its traffic-counting program *(4)*. Although each of these studies provides useful background and guidance, the conceptual basis of Hallenbeck and Bowman *(1)*—utilizing the HPMS framework for purposes of statewide highway data collection—was explored in this study. There are a number of other relevant and useful works in the general area *(5–13)*. A comprehensive account of sampling theory as it has been developed for use in sample surveys is given by Cochran *(14)*.

In this paper, a statistical framework is presented for volume counting and vehicle classification, particularly for deriving estimates of annual average daily traffic (AADT) from short-duration axle counts at any location on a state highway system, using Washington State and WSDOT as a case study.

Department of Civil Engineering and Institute of Transportation Studies, University of California, Irvine, Calif. 92717.

ANNUAL AVERAGE DAILY TRAFFIC

Basic Model

A basic model for estimating AADT for a particular highway segment based on a single, short-duration count is

$$AADT = VOL(F_S)(F_A)(F_G) \tag{1}$$

where

- VOL = average 24-hr volume from a standard WSDOT 72-hr Tuesday-Thursday short count;
- F_S = seasonal factor for the count month;
- F_A = weekday axle correction factor if VOL is in axles; equal to 1 if VOL is in vehicles; and
- F_G = growth factor if VOL is not a current year count; equal to 1 otherwise.

To determine the relative precision of an estimated AADT from Equation 1, the coefficient of variation (ratio of standard deviation to mean) must be found. This can be obtained from the following approximate expression:

$$cv^2(AADT) = cv^2(F_S) + cv^2(F_A) + cv^2(F_G) \tag{2}$$

where each cv^2 is the squared coefficient of variation of each variable. Thus the coefficient of variation of the AADT estimate is

$$cv(AADT) = [cv^2(F_S) + cv^2(F_A) + cv^2(F_G)]^{0.5} \tag{3}$$

The relative precision (percentage) at a 100 (1 − α) percent confidence level is then given approximately by

$$\text{Precision } (AADT) = \pm\ 100 Z_{\alpha/2} cv(AADT)\ \% \tag{4}$$

where $Z_{\alpha/2}$ is a standard normal statistic corresponding to the 100 $(1 - \alpha)$ percent confidence level (found in tables of any statistics book).

Also, a 100 $(1 - \alpha)$ percent confidence interval is defined approximately as

$$AADT \pm Z_{\alpha/2} AADT\ cv(AADT) \quad (5)$$

The Z-statistics corresponding to 95, 90, and 80 percent confidence levels are 1.96, 1.645, and 1.282, respectively.

Seasonal Factor Analysis

Factor Grouping

The data for analyzing seasonal factors were basically obtained from WSDOT *Annual Traffic Reports (15)*, which list the monthly permanent traffic recorder (PTR) traffic volumes throughout each year.

Several alternative methods for performing seasonal factoring were evaluated. The primary ones considered were

- Continued use of existing WSDOT Data Office procedures (see paper by Ritchie and Hallenbeck in this Record),
- Cluster analysis of PTRs,
- Procedures suggested in the FHWA draft counting guide *(13)*, and
- A revised FHWA procedure using linear regression.

The chosen strategy was the fourth of these options. The approach uses the basic method recommended by FHWA. The state highway system is stratified by geographic region and functional classification. The strata are then examined to determine which have similar seasonal patterns and might therefore be combined. PTR data from 1980 through 1984 were used to calculate the appropriate factor groups. The chosen groups were

- Rural Interstates,
- Urban roads,
- Other rural roads in the northeastern part of the state,
- Other rural roads in the southeastern part of the state,
- Other rural roads in the northwestern part of the state,
- Other rural roads in the southwestern part of the state, and
- Central mountain passes.

With the exception of the central mountain group, each factor group is defined by functional class of road and county boundaries. (Note that the urban group contains all state highways classified as urban regardless of county location.)

The advantages of the adopted approach are that

- The seasonal factors are statistically valid, meaning that the precision associated with any AADT estimate based on these factors can be calculated;
- The overall errors associated with this approach are equal to or smaller than the errors associated with any other seasonal factoring approach considered; and
- The factoring procedure is transparent to any user of volume information and thus allows the recalculation of the raw traffic count at some later time if desired.

Each of the other seasonal factor procedures had drawbacks that were judged unacceptable. For example, in the case of cluster analysis,

- The clusters computed were not consistent across years (i.e., PTRs changed groups from year to year), which means that roads should change groups as well, but no method was available to make that adjustment each year (see paper by Ritchie and Hallenbeck in this Record);
- Individual road sections are not easily or accurately assigned to cluster groups, irrespective of the difficulties mentioned previously; and
- The total error in the AADT estimate (including seasonal variation, daily variation, and variation in the axle correction factor) was only marginally better than that obtained by the recommended approach before inclusion of the indeterminate error that is present as a result of the first two points.

Regression Models

Seasonal factors for each month of the year were therefore derived for each of the seven factor groups described earlier. The modified FHWA approach adopted basically involved a regression analysis for each factor group for each month of AADT versus the average 24-hr short-count volumes that could be formed for each PTR from 72-hr Tuesday-Thursday counts in that month. The resulting regression coefficient of the short-count volume is then the derived seasonal factor for that factor group and month. This approach corresponds to the manner in which short counts are actually taken and converted to AADT estimates by WSDOT.

The first seasonal factor regression model estimated was as follows (note that the constant term is suppressed):

$$AADT = \beta\ VOL + u \quad (6)$$

where *AADT* and *VOL* are as defined previously, β is the regression coefficient (seasonal factor) to be estimated, and u is the error term. Such an equation would typically be estimated by ordinary least squares *(16)*. However, one of the required assumptions of that method is homoscedasticity, which means that the variance of the error term (u) is constant regardless of the magnitude of *VOL*. It often happens that this assumption is not valid (the case of heteroscedasticity) and the model must be reduced (by a transformation) to a form in which the error term does have a constant variance.

Estimation of Equation 6 revealed the presence of heteroscedasticity for some factor group and monthly traffic count data sets. Further, a consequence of this problem was that estimated variances would be biased and would underestimate the true variance. To address this issue, a commonly used transformation was employed to reduce Equation 6 to a homoscedastic form. It was assumed that the variance of the error term was known up to a multiplicative constant:

$$var\ (u) = \sigma^2\ VOL^2 \quad (7)$$

Dividing through Equation 6 by *VOL* yields

$$AADT/VOL = \beta + (u/VOL) \quad (8)$$

Substituting $e = u/VOL$ gives

$$AADT/VOL = \beta + e \quad (9)$$

where

$$\begin{aligned} var(e) &= (1/VOL^2)\, var(u) \\ &= (1/VOL^2)\, \sigma^2\, VOL^2 \\ &= \sigma^2 \end{aligned}$$

Thus, the variance of the error term (e) in Equation 9 is constant (σ^2) and ordinary least squares estimation methods can be applied. The form of Equation 9 is now so simple that computerized regression packages are not really required. The estimation results can be obtained as follows:

$$\hat{\beta} = \sum_{i=1}^{n} (AADT/VOL_i)/n \quad (10)$$

$$\hat{\sigma}^2 = \left\{ \sum_{i=1}^{n} [(AADT_i/VOL_i) - \hat{\beta}]^2 \right\} /(n-1) \quad (11)$$

$$var(\hat{\beta}) = \hat{\sigma}^2/n \quad (12)$$

and the t-statistic on $\hat{\beta}$ is

$$t_{\hat{\beta}} = \hat{\beta}(n^{1/2})/\hat{\sigma} \quad (13)$$

In Equations 10 and 11 the subscript i refers to each short count in the month for the factor group, and n represents the number of counts.

Finally, the relative precision of the AADT estimates must be derived. When the seasonal factors from Equation 9 are applied to counts in the following year, the value of the ratio $AADT/VOL$ in the equation is forecast. Therefore the appropriate variance measure is the variance of the prediction error for the forecast ratio of $AADT$ to VOL. It can be shown that this variance is given by

$$\sigma^2 (1 + 1/n) \quad (14)$$

for each factor group and month. The required coefficient of variation for Equation 3 is then

$$cv(F_S) = \hat{\sigma}\, (1 + 1/n)^{0.5}/\hat{\beta} \quad (15)$$

It is interesting to note that this theoretically derived result is equivalent to that obtained by more qualitative reasoning *(1, 13)*.

Results

The seasonal factors for 1984, derived using the procedures described, are given in Table 1 for April through September (the period when WSDOT performs the vast majority of its traffic counting) and in Table 2 for October through March. Because of the high variability of factors for the central mountain group, this group was treated separately.

The coefficients of variation, based on Equation 15, are given in Table 3. These have been used to calculate relative precision levels of April through September AADT estimates, as given in Table 4, without incorporating axle correction or growth factors.

It is also interesting to note how the AADT precision levels vary as a function of the number of PTRs in each factor group. Little improvement in relative precision was obtained beyond about six to eight PTRs per group. Thus, in terms of statistical precision of AADT estimates only, little is gained by having additional PTRs. However, as discussed by Ritchie and Hallenbeck in this Record, there may be other reasons for maintaining large numbers of PTRs in any group, such as the automatic collection of vehicle classification data.

TABLE 1 1984 SEASONAL FACTORS FOR APRIL THROUGH SEPTEMBER

	Month					
Group	April	May	June	July	August	September
Rural Interstate	1.132	1.126	0.960	0.907	0.849	0.990
Urban	0.966	0.952	0.903	0.894	0.878	0.907
Northwestern	1.023	0.995	0.921	0.848	0.812	0.957
Southwestern	1.087	1.055	0.935	0.823	0.769	0.925
Southeastern	1.137	1.077	0.956	0.896	0.855	0.979
Northeastern	1.025	0.927	0.895	0.754	0.779	0.882

Axle Correction Factor Analysis

Axle correction factors are required to convert short-count volumes to AADT estimates when those short counts are obtained using equipment that records axles rather than vehicles. Calculation of the factors requires vehicle classification information (percentage of vehicles in each class) as well as knowledge of the number of axles per vehicle in each vehicle class.

The average number of axles per vehicle (A_V) in a given factor group (typically highway functional class) is given by

$$A_V + \sum_{C} (Axles_C)\, (P_C) \quad (16)$$

where $Axles_C$ is the number of axles per vehicle in Class C and P_C is the proportion of vehicles in Class C (system-level estimate). The variance of A_V is then given by

$$var(A_V) = \sum_{C} (Axles_C)^2\, var(P_C) \quad (17)$$

where $var(P_C)$ is the variance of Vehicle Class C proportion, from a vehicle classification study.

Thus the coefficient of variation of A_V is

$$cv(A_V) = \left[\sum_{C} (Axles_C)^2\ var(P_C)\right]^{0.5} / \left[\sum_{C} (Axles_C)(P_C)\right] \quad (18)$$

However, the desired axle correction factor (F_A) is actually the inverse of A_V:

$$F_A = A_V^{-1} \quad (19)$$

It can be shown by a first-order Taylor series approximation that

$$cv(F_A) = cv(A_V) \quad (20)$$

TABLE 2 1984 SEASONAL FACTORS FOR OCTOBER THROUGH MARCH

Group	Month October	November	December	January	February	March
Rural Interstate	1.274	1.220	1.116	1.554	1.425	1.238
Urban	1.045	1.006	0.935	1.088	1.033	0.988
Northwestern	1.236	1.124	1.067	1.296	1.558	1.075
Southwestern	1.467	1.283	1.067	1.408	1.259	1.145
Southeastern	1.500	1.318	1.043	1.595	1.472	1.259
Northeastern	1.339	1.176	0.981	1.200	1.184	1.163

TABLE 3 COEFFICIENTS OF VARIATION OF 1984 SEASONAL FACTORS, $cv(F_S)$

Month	Factor Group Rural Interstate	Urban	Northwestern	Southwestern	Southeastern	Northeastern
January	0.172	0.090	0.149	0.216	0.196	0.074
February	0.150	0.073	0.105	0.154	0.190	0.100
March	0.113	0.057	0.102	0.147	0.180	0.146
April	0.109	0.062	0.095	0.132	0.144	0.123
May	0.089	0.070	0.078	0.108	0.138	0.080
June	0.064	0.057	0.095	0.082	0.118	0.077
July	0.057	0.063	0.092	0.077	0.115	0.104
August	0.064	0.042	0.090	0.143	0.090	0.097
September	0.090	0.059	0.069	0.129	0.112	0.086
October	0.167	0.112	0.150	0.217	0.239	0.176
November	0.255	0.090	0.130	0.186	0.250	0.115
December	0.078	0.073	0.084	0.114	0.088	0.083

This result permits the coefficient of variation of the axle correction factor to be derived readily from Equation 18 for insertion into Equation 3.

Table 5 gives the estimated axle correction factors for eight functional classes of highway, together with relative precisions and coefficients of variation.

Growth Factors

Growth factors often represent a relatively minor part of the factoring process to obtain AADT estimates from short counts. However, at times an old count must be converted to a more recent AADT by means of a growth factor. Several methods exist for estimating growth factors. In general, the approaches are fairly crude ways of attempting to account for traffic growth or decline over time. The analysis discussed in this section was exploratory only, although the results appear reasonable.

Simple growth factors were estimated for each of the previously identified seasonal factor groups for 1982–1983 and 1983–1984. The factors were obtained by forming the ratio of AADT in the more recent year to that in the earlier year for each PTR in a group and applying the regression analysis procedure discussed previously. In one group there was one PTR, and in a second group no PTR, for both years, so that coefficients of variation of the factors (F_G) could not be formed. Table 6 gives the estimated growth factors for each period together with their coefficients of variation.

TABLE 4 RELATIVE PRECISION (%) OF SEASONALLY ADJUSTED AADT ESTIMATES FROM SHORT COUNTS IN EACH MONTH (without incorporating axle correction or growth factors)

Month	Factor Group Rural Interstate	Urban	Northwestern	Southwestern	Southeastern	Northeastern
April	18	10	16	22	24	20
May	15	12	13	18	23	13
June	11	9	16	13	19	13
July	9	10	15	13	19	17
August	11	7	15	24	15	16
September	15	10	11	21	18	14

Note: 90 percent confidence level.

TABLE 5 AXLE CORRECTION FACTORS

Functional Class	F_A[a]	Percentage Precision[b]	$cv(F_A)$
Rural Interstate	0.423	10.2	0.062
Rural principal arterial	0.461	8.8	0.053
Rural minor arterial	0.471	4.8	0.029
Rural collector	0.459	10.7	0.066
Urban Interstate	0.454	3.9	0.023
Urban principal arterial	0.463	6.8	0.041
Urban minor arterial	0.482	2.1	0.013
Urban collector	0.495	1.6	0.010

[a]Weekday factors.
[b]90 percent confidence level.

TABLE 6 GROWTH FACTORS

Group	1982–1983		1983–1984	
	F_G	$cv(F_G)$	F_G	$cv(F_g)$
Rural Interstate	1.065	0.020	1.024	0.037
Urban	1.175	0.306	1.046	0.066
Northwestern	1.052	0.110	1.016	0.055
Southwestern	1.059	—	1.094	—
Southeastern	1.041	0.060	1.041	0.042
Northeastern	—	—	—	—

VEHICLE CLASSIFICATION

Data Analysis

Because of the limited nature of vehicle classification counts taken by WSDOT in recent years, the best available data set for statistical analysis was from a 1980–1981 study that was done for FHWA. Unlike volume counts, which utilize a system of PTR stations for continuous monitoring, it is not presently possible to derive vehicle classification seasonal factors for conversion of a single (say 24-hr) classification count to an annual average estimate for a given highway segment. Rather, the data available permit only an approximate systemwide plan to be developed for an annual counting program on different functional classes, in order to derive annual average vehicle classification results. Improvements to the department's current vehicle classification activities are discussed further by Ritchie and Hallenbeck in this Record.

The 1980–1981 data consist of 248 manual 24-hr vehicle classification counts. The data were collected at 31 locations across the state with 4 weekday counts (one per season) and 4 weekend counts (one per season) at each location. For purposes of analysis, the data were reduced to six vehicle types:

1. Cars,
2. Two-axle trucks,
3. Three-axle trucks,
4. Four-axle trucks,
5. Five-axle trucks, and
6. Trucks with six or more axles.

In addition, a slightly more detailed set of functional classifications than was used in the seasonal factor development was retained for initial analysis. These functional classes consisted of eight groups: Interstates, principal arterials, minor arterials, and collectors for both rural and urban locations.

The principal analysis method used was a two-stage cluster sampling approach with multiple strata. The first set of strata corresponded to functional classes. Within strata, the primary sampling units or clusters were possible count locations, and the secondary or elementary sampling units were days at each location (required to be the same at each location in a stratum). The second stratification was introduced with respect to weekdays and weekend days because vehicle classifications were noticeably different across these strata; truck percentages were often considerably lower on weekend count days. The population sizes for each stage were taken to be the number of HPMS population sections in each functional class in the case of locations and, at the second stage, simply the number of weekdays or weekend days, or both, in a year. Allowance was also made in the analysis for the unequal size of the second-stage units (as is often assumed in cluster analysis) due to the daily variations in traffic volume throughout the year.

Within each functional class, and for each Vehicle Class C, the average (weighted) vehicle proportion (P_C) was estimated as

$$P_C = \left(\sum_{i=1}^{n} p_i\right)/n \tag{21}$$

where

$p_i = w_1 p_{i1} + w_2 p_{i2}$;
p_i = proportion at location i;
p_{i1} = weekend proportion at location i

$$= \left(\sum_{k=1}^{m_1} C_{ik1}\right) \Big/ \left(\sum_{k=1}^{m_1} X_{ik1}\right);$$

p_{i2} = weekday proportion at location i

$$= \left(\sum_{j=1}^{m_2} C_{ij2}\right) \Big/ \left(\sum_{j=1}^{m_2} X_{ij2}\right);$$

C_{ik1} = total number of vehicles of type C at station i on weekend day k;
C_{ij2} = total number of vehicles of type C at station i on weekday j;
X_{ik1} = total number of vehicles at station i on weekend day k;
X_{ij2} = total number of vehicles at station i on weekday j;
p_{i1k} = proportion observed on weekend day k;
p_{i2j} = proportion observed on weekday j;
m_1 = number of weekend days at each location;
m_2 = number of weekdays at each location;
w_1 = 2/7;
w_2 = 5/7; and
n = number of count locations.

The variance was obtained from

$$var\,(P_C) = (1 - f_1)(s_1^2/n) + [w_1^2(1 - f_{21})s_{21}^2/(nm_1) + w_2^2(1 - f_{22})s_{22}^2/(nm_2)] \tag{22}$$

where

$f_1 = n/N,$
N = population size of HPMS segments for functional class,
$f_{21} = m_1/104,$
$f_{22} = m_2/261,$

$$s_{21}^2 = \left(\sum_{i=1}^{n} s_{21i}^2\right)/n,$$

$$s_{21i}^2 = \sum_{k=1}^{m_1} (p_{i1k} - p_{i1})^2/(m_1 - 1),$$

$$s_{22}^2 = \left(\sum_{i=1}^{n} s_{22i}^2\right)/n,$$

$$s_{22i}^2 = \sum_{j=1}^{m_2} (p_{i2j} - p_{i2})^2/(m_2 - 1), \text{ and}$$

$$s_1^2 = \sum_{i=1}^{n} (p_i - p_c)^2/(n - 1).$$

Thus the coefficient of variation of the estimate is

$$cv(P_C) = [var(P_C)]^{0.5}/P_C \tag{23}$$

The relative precision (percentage) at a 100 (1 − α) percent confidence level is then given approximately by

$$\text{Precision } (P_C) = \pm\ 100\ Z_{\alpha/2}cv(P_C) \tag{24}$$

In addition to this analysis approach, which distinguishes between counts on weekdays and weekends by introducing sample stratification, estimates for P_C were also calculated without this stratification by pooling weekday and weekend counts at each location. For this simpler formulation, P_C is calculated from

$$P_C = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} C_{ij}\right)\bigg/\left(\sum_{i=1}^{n}\sum_{j=1}^{m} X_{ij}\right) \tag{25}$$

where

C_{ij} = total number of vehicles of type C at station i on day j,
X_{ij} = total number of vehicles at station i on day j,
$f_1 = n/N,$
$f_2 = m/365,$
m = number of days sampled at each station, and
n = number of count locations.

The variance of P_C is then calculated from

$$var(P_C) = (1 - f_1)(S_1^2/n) + f_1(1 - f_2)(s_2^2/mn) \tag{26}$$

where s_1^2 is as previously defined, and

$$s_2^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} (p_{ij} - \bar{p}_i)^2/[n(m - 1)],$$

$$\bar{p}_i = \sum_{j=1}^{m} C_{ij}/\sum_{i=1}^{m} X_{ij}, \text{ and}$$

$$p_{ij} = C_{ij}/X_{ij}.$$

The coefficient of variation and precision of P_C are then calculated as before by Equations 23 and 24, respectively.

Results

Table 7 gives the classification count results for each functional class. These averages are based on the weighted weekday and weekend counts. Table 8 gives the relative precision of these results at a 90 percent confidence level. Clearly, the precision of the estimates for large trucks (five or more axles) is relatively poor, although this was not unexpected given the limited nature of the counts and the inherent variability of truck travel as a percentage of total daily volume. Table 9 gives the coefficients of variation for each vehicle class proportion.

TABLE 7 PERCENTAGE OF VEHICLES BY TYPE IN EACH FUNCTIONAL CLASS

	Vehicle Class					
Functional Class	1	2	3	4	5	6
Rural Interstate	87.0	3.1	0.6	0.3	8.3	0.8
Rural principal arterial	90.3	3.2	1.0	0.1	5.0	0.3
Rural minor arterial	92.2	2.9	0.9	0.1	3.5	0.5
Rural collector	89.3	3.5	3.0	0.3	3.6	0.3
Urban Interstate	91.1	2.8	0.7	0.4	4.5	0.4
Urban principal arterial	90.8	3.1	0.6	0.2	4.9	0.4
Urban minor arterial	94.4	2.8	0.8	0.2	1.7	0.2
Urban collector	95.1	3.4	0.4	0.1	0.9	0.1

The estimation of annual average daily truck traffic (AADTT) volume can be accomplished readily by applying the analysis results and extending the AADT estimation equations:

$$AADTT = VOL\ (F_S)(F_A)(F_G)(P_C) \tag{27}$$

where P_C is the appropriate vehicle proportion estimate from Table 7 and all other notations are as defined previously. It must be remembered that this AADTT estimate is based on system-level vehicle classification data not a specific truck count for the section where the volume count *(VOL)* was taken.

The coefficient of variation can be obtained from

$$cv(AADTT) = [cv^2(F_S) + cv^2(F_A) + cv^2(F_G) + cv^2(P_C)]^{0.5} \tag{28}$$

where $cv(P_C)$ is as given in Table 9. The relative precision at a 100 (1 − α) percent confidence level is then given approximately by

$$\text{Precision } (AADTT) = \pm\ 100\ Z_{\alpha/2}\ cv(AADTT)\ \% \tag{29}$$

TABLE 8 RELATIVE PRECISION (%) OF VEHICLE CLASSIFICATION RESULTS

Functional Class	Vehicle Class 1	2	3	4	5	6
Rural Interstate	4	11	13	35	35	33
Rural principal arterial	3	7	50	43	43	48
Rural minor arterial	2	9	22	45	33	68
Rural collector	7	29	82	62	91	69
Urban Interstate	1	8	13	22	20	14
Urban principal arterial	3	17	22	39	41	40
Urban minor arterial	1	26	31	67	19	44
Urban collector	1	25	35	43	34	86

Note: 90 percent confidence level.

TABLE 9 COEFFICIENTS OF VARIATION FOR VEHICLE PROPORTIONS FROM TABLE 7

Functional Class	Vehicle Class 1	2	3	4	5	6
Rural Interstate	0.024	0.068	0.079	0.213	0.215	0.201
Rural principal arterial	0.018	0.044	0.303	0.263	0.259	0.294
Rural minor arterial	0.010	0.057	0.134	0.271	0.201	0.416
Urban Interstate	0.007	0.050	0.077	0.131	0.119	0.088
Urban principal arterial	0.018	0.103	0.134	0.237	0.247	0.241
Urban minor arterial	0.008	0.157	0.187	0.405	0.114	0.266
Urban collector	0.007	0.150	0.216	0.260	0.207	0.522

As an example, consider the calculation of an annual average daily five-axle truck volume on a rural Interstate segment, based on a short duration axle count in June:

Average 24-hr volume (VOL) = 50,000 axles,

F_S = 0.960 (Table 1),
F_A = 0.423 (Table 5),
F_G = 1.0 (because this is a current-year count),
P_C = 0.083 (Table 7),
$cv(F_S)$ = 0.064 (Table 3),
$cv(F_A)$ = 0.062 (Table 5),
$cv(F_G)$ = 0.0 (because an estimated factor is not used), and
$cv(P_C)$ = 0.215 (Table 9).

Thus, from Equation 21, the estimate of daily five-axle trucks is

$$AADTT = 50{,}000\,(0.960)(0.423)(1.0)(0.083)$$
$$= 1{,}685 \text{ five-axle trucks.}$$

From Equation 22, the coefficient of variation of this estimate is

$$CV(AADTT) = [(0.064)^2 + (0.062)^2 + (0.0)^2 + (0.215)^2]^{0.5}$$
$$= 0.233.$$

Finally, from Equation 23, the relative precision of this estimate at a 90 percent confidence level is

$$\text{Precision } (AADTT) = \pm 100\,(1.645)(0.233)\ \%$$
$$= \pm 38.3\ \%,$$

which means there is 90 percent confidence that the true value of AADTT is within about 40 percent of the estimate of 1,685 five-axle trucks per day.

Sample Design

The results obtained from these analyses of vehicle classification data provided some basis for developing the study recommendations for this data item (see paper by Ritchie and Hallenbeck in this Record). Some of the findings related to design of a sample for collecting vehicle classification data are presented in this subsection.

Of interest is how the statistical precision of classification estimates is affected by sample size and choice of confidence level. To gain further insight into these relationships, a number of tabular and graphic reports were generated.

For example, Table 10 gives the variation in precision achieved with a number of different sample designs in the case of rural Interstates. These results are based on a cluster analysis, as before, but with pooled weekend and weekday counts without stratification. It can be seen that the precision levels are more sensitive to the number of locations chosen than the number of days surveyed per location. For a given number of classification counts, the results indicate that it is better to take all of those counts at different locations, with only one count per location, on randomly chosen days during the year.

TABLE 10 RELATIVE PRECISION (%) OF RURAL INTERSTATE VEHICLE CLASSIFICATIONS FOR DIFFERENT SAMPLE DESIGNS

No. of Locations	No. of Days	No. of Counts	Vehicle Class 1	2	3	4	5	6
2	1	2	9	37	81	95	105	105
2	5	10	6	23	39	68	71	69
4	1	4	7	26	57	67	74	74
4	5	20	4	16	27	48	50	49
8	1	8	5	18	40	47	52	52
8	5	40	3	11	19	34	35	34
20	1	20	3	12	25	29	33	33
20	5	100	2	7	12	20	21	21
40	1	40	2	9	18	20	23	23
40	5	200	1	5	8	13	14	14

Note: 90 percent confidence level.

To avoid the added complexity and cost of having to take at least two counts per location (one weekday, one weekend) at every sampled location, as required by the stratified cluster analysis procedure, it was decided that, for purposes of sample design and implementation, a pooled cluster analysis approach should be used without stratification by day of week. All that this would mean in practice is that the count day or days at a location would be chosen randomly from all days in the year. Given the nature of the data on which the analyses were based and the interim nature of any recommended manual count program [due to introduction of automatic vehicle classifiers by the department (see paper by Ritchie and Hallenbeck in this Record)], this approach was judged appropriate.

Also investigated was the effect of both confidence level and number of counts (or locations counted) on the precision of vehicle proportions. Achieving both smaller precision levels and higher confidence levels requires that more counts be taken. In the case of five-axle trucks on rural Interstates it was noted for example that the major improvement in precision came from taking approximately 20 counts and that the improvement in precision for successive counts was relatively small. However, the magnitude of the precision was still undesirably high. The implication is that, to achieve precise results, a much larger number of vehicle classification counts than the department currently collects are required. The detailed recommendations that were developed on the basis of these results are reported by Ritchie and Hallenbeck in this Record.

CONCLUSIONS

A rigorous statistical approach to statewide data collection and program design permits the estimation of data precision and can provide a rational basis to assist in allocating limited resources among the various possible data collection activities. A statistical approach is also important because the desired precision and confidence level have a major impact on sample design and cost. There is little point in collecting more precise sample data at a higher level of confidence than is required by the data users, particularly when considerable cost savings can be realized by using smaller sample sizes. Conversely, when resources are limited and insufficient for the desired sample size, trade-offs between precision and level of confidence can be made explicit. Further discussion of this issue is presented in a companion paper by Ritchie and Hallenbeck in this Record.

A statistical framework for volume counting and vehicle classification, and particularly for deriving estimates of AADT from short-duration axle counts at any location on a state highway system, has been presented. AADT estimates can be derived for each vehicle type, if desired. The estimation of associated seasonal, axle correction, and growth factors was also described. The methodology enables the statistical precision of all of these estimates to be determined.

ACKNOWLEDGMENT

Many individuals from WSDOT, the Washington State Transportation Center (TRAC), and FHWA contributed significantly to this study, and their input is gratefully acknowledged. This paper is based on a research project conducted by the University of Washington and TRAC for the Washington State Transportation Commission and Department of Transportation, in cooperation with the FHWA, U.S. Department of Transportation.

REFERENCES

1. M. E. Hallenbeck and L. A. Bowman. *Development of a Statewide Traffic Counting Program Based on the Highway Performance Monitoring System.* FHWA, U.S. Department of Transportation, 1984.
2. *Highway Performance Monitoring System Field Manual for the Continuing Analytical and Statistical Data Base.* Office of Highway Planning, FHWA, U.S. Department of Transportation, 1984.
3. Wright Forssen Associates, Comsis Corporation, and TRANSPO Group. *Highway Traffic Data Study,* Vol. I: *Final Report.* Alaska Department of Transportation and Public Facilities, Planning and Programming Division, Southeast Region, 1983.
4. D. T. Hartgen and J. H. Lemmerman. "Streamlining Collection and Processing of Traffic Count Statistics." In *Transportation Research Record 928,* TRB, National Research Council, Washington, D.C., 1983, pp. 11–20.
5. Peat, Marwick and Mitchell. *Guide to Urban Traffic Volume Counting.* Final Report. FHWA, U.S. Department of Transportation, 1980.
6. J. DiRenzo, L. Bowman, and M. Hallenbeck. *Review of the Department's Traffic Count Program.* Pennsylvania Department of Transportation, Harrisburg, 1983.
7. John Hamburg & Associates. *Improved Methods for Vehicle Counting and Determining Vehicle-Miles of Travel—Final Report on State of the Art.* NCHRP, TRB, National Research Council, Washington, D.C., 1982.
8. John Hamburg & Associates. *Improved Methods for Vehicle Counting and Determining Vehicle-Miles of Travel—Traffic Counting Program Design.* NCHRP, TRB, National Research Council, Washington, D.C., 1982.
9. L. T. Hoang and V. P. Poteat. "Estimating Vehicle Miles of Travel by Using Random Sampling Techniques." In *Transportation Research Record 779,* TRB, National Research Council, Washington, D.C., 1980, pp. 6–10.
10. L. M. Rudman. "Vehicle Kilometers Traveled: Evaluation of Existing Data Sources." In *Transportation Research Record 726,* TRB, National Research Council, Washington, D.C., 1979, pp. 19–28.
11. D. L. Green and A. S. Loebl. *Vehicle-Miles of Travel Statistics, Lifetime Vehicle-Miles of Travel and Current Methods of Estimating Vehicle-Miles of Travel.* Oak Ridge National Laboratory, Oak Ridge, Tenn., 1979.
12. J. P. Mahoney. *Traffic Data Needs for Pavement Rehabilitation.* Washington State Transportation Commission, 1983.
13. *Draft Traffic Monitoring Guide.* Office of Highway Planning, FHWA, U.S. Department of Transportation, July 1984.
14. W. G. Cochran. *Sampling Techniques.* John Wiley & Sons, Inc., New York, 1977.
15. *Annual Traffic Reports 1980–1984.* Planning, Research and Public Transportation Division, Washington State Department of Transportation, Olympia, 1983, 1983, 1983, 1984, and 1985.
16. G. S. Maddala. *Econometrics.* McGraw-Hill Book Company, New York, 1977.

Classifying a Rural Road Network for Traffic Counting

CHRISTO J. BESTER AND J. D. DE B. JOUBERT

The evolution of a classification system for the South African rural road network is described. An initial attempt to classify a provincial network on the basis of the trip length and trip purpose of traffic, as judged subjectively by local road officials, was tested in a pilot study on traffic counting. The results of the study show that there is a good correlation between average trip length, as calculated from an origin-destination matrix, and various traffic patterns. A final classification system, based on these trip lengths, is recommended.

As a result of the variation in traffic volumes, the annual average daily traffic (AADT) at a specific location can only be determined accurately by means of permanent traffic counters (PTCs). These, however, can only be afforded at a limited number of locations and therefore the traffic on the remainder of the links in a network is counted for short periods only. The AADT on these links is estimated by means of expansion factors derived from the data obtained from the PTCs.

The purpose of the classification of the road network is to group the links according to the uniformity of their traffic patterns to ensure that the expansion factors are calculated from the appropriate group of PTCs and applied to the correct group of short-term counts. Traffic patterns can be described by the regular variation in daily, weekly, and seasonal traffic volumes.

There are various methods of grouping PTCs on the basis of their monthly variation in traffic flows *(1-3)*. However, for the vast majority of road sections, the traffic patterns are unknown, and assigning them to a specific group is difficult. Moreover, when no PTCs have been in operation in a certain area the problem of classifying the network becomes more complicated.

The evolution of a process for classifying the South African rural road network for traffic counts is described in this paper. The process started off by using a framework in which subjective judgment and an intimate local knowledge of the nature of traffic were the main components. This framework was then tested in a pilot study, the results of which were used to develop a final procedure for classification based on trip length.

In South Africa rural traffic counting is the responsibility of the provincial road authorities. Over the years different systems have evolved. In one province 40 links are counted in a revolving system; each link is counted for 1 year every 5 years. In another the 10 PTCs are manually operated for 18 hr a day, and in a third about 90 percent of the PTCs are located on roads with essentially similar traffic patterns. Some systems give the results in terms of equivalent vehicle units (with a heavy vehicle representing three cars) and others give the total number of vehicles and a percentage of heavy vehicles.

Because of the discrepancies in the presentation and accuracy of traffic counts, all attempts at countrywide road planning *(4, 5)* have been severely hampered. The Committee of State Road Authorities (CSRA) therefore decided to form a subcommittee to investigate and report on a uniform traffic-counting system for all South African rural roads. This has also been the main objective of a research project of the Rural Transport Group at the National Institute for Transport and Road Research (NITRR).

Procedures for rural traffic counting in the Northern Hemisphere are well established *(6)*. With a few exceptions these could be adopted for use in South Africa. The main problem, however, was to decide on a uniform classification system for the rural network, which again was a prerequisite for determining the locations of the PTCs.

The main difference between the proposed procedure for South Africa and the traditional procedures used elsewhere is that in South Africa factors are calculated for seasons (four or five per year) instead of months. The seasons are determined as follows: First, the *E*-days are identified; these are exceptional days, usually public holidays, the days following or preceding them, and school holidays, when traffic deviates significantly from the normal patterns. The remaining days with normal traffic, or *N*-days, are then divided into counting seasons with uniform weekly traffic patterns. The short-term counts take place on *N*-days only. In Figure 1 the *E*-days and *N*-days are shown for a specific road.

SUBJECTIVE CLASSIFICATION

Classification of a road network for traffic counting is based on the traffic patterns, which in turn are affected by trip purpose and trip length *(3)* of the traffic on a specific link. With this in mind, the following framework was drawn up for the classification of the South African rural road network:

- Class A: Roads on which commuter traffic over distances less than 100 km is mainly concentrated.
- Class B1: Roads carrying long-distance intermetropolitan traffic closer than 20 km to the major cities.
- Class B2: Roads carrying long-distance intermetropolitan traffic farther than 20 km from the major cities.
- Class C1: Roads carrying interregional, medium-distance traffic.
- Class C2: Roads carrying traffic mainly between neighboring towns.
- Class D1: Collector roads in an intensive agricultural area (crops).
- Class D2: Collector roads in an extensive agricultural area (livestock).
- Class E: Roads with exceptional traffic patterns, such as extremely high volumes of intermittent recreational traffic.

No PTCs would be used on Class-E roads unless it was important to justify a continuous count for its own sake.

To test this framework it was decided to do a pilot study in the Orange Free State.

C. J. Bester, National Institute for Transport and Road Research, CSIR, P.O. Box 395, Pretoria 0001, Republic of South Africa. J. D. de B. Joubert, University of Pretoria, Pretoria 0001, Republic of South Africa.

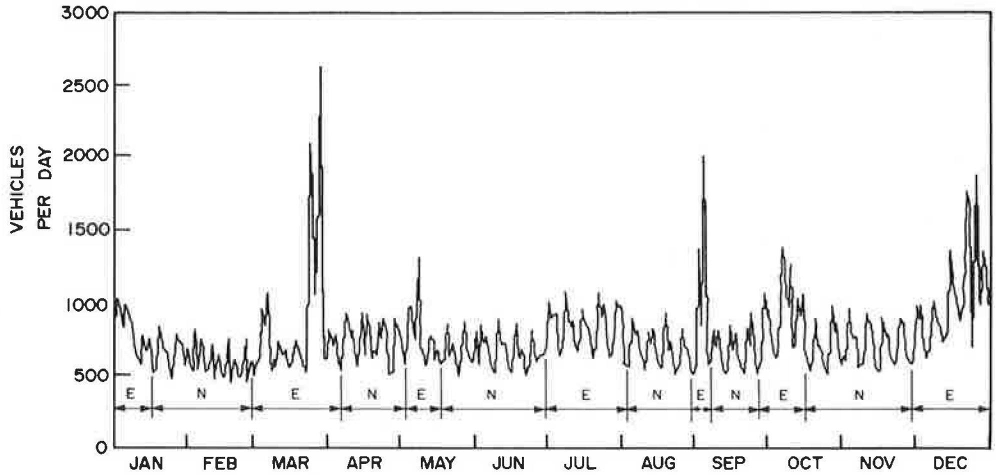


FIGURE 1 *E*-days and *N*-days for a specific road.

PILOT STUDY

The pilot study took place between April 1, 1983, and March 31, 1984. The road network with its 6,000 links was classified in accordance with the framework given in the previous section. This was done by officials of the Roads Department of the Provincial Administration of the Orange Free State, who used their knowledge of the trip purpose on the various links.

For the study it was decided to concentrate on Classes B, C, and D. Twelve positions were randomly selected for the location of the PTCs, which were allocated to the classes as follows:

Class	*Counters*
B2	1, 2, 3, and 4
C1	5 and 6
C2	7 and 8
D1	9 and 10
D2	11 and 12

The distribution of the counters throughout the province is shown in Figure 2.

After the data had been collected it was obvious that the year could be divided into five distinct counting seasons:

Season	*From*	*To*
1	April 14, 1983	May 10, 1983
2	May 17, 1983	June 21, 1983
3	August 3, 1983	September 20, 1983
4	October 13, 1983	December 1, 1983
5	January 20, 1984	March 21, 1984

In spite of a number of interruptions while data from the PTCs were being collected, the original goal of the pilot study was achieved.

Because of the low traffic volumes at Counters 11 and 12 (12 and 19 vehicles per day, respectively) and high daily variations, no meaningful analysis could be done for Class D2. It is in any case doubtful whether roads in this class would ever be included in a formal traffic-counting program.

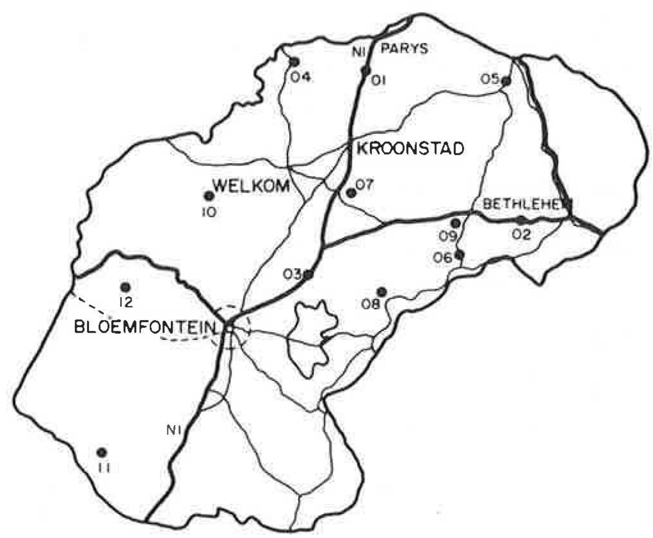


FIGURE 2 Location of counters.

TRAFFIC PATTERNS

Traffic patterns can be described as regular variations in traffic volumes and are sometimes quantified by means of monthly expansion factors *(3)*. Regular daily, weekly, or seasonal relationships can, however, also be used to quantify the traffic patterns for the different classes of road, and these relationships were specifically analyzed.

Daily Traffic Variation

Because short-term manual counts are usually undertaken for less than 24 hr, it is important to know what is happening to traffic volumes during the rest of the day. In the pilot study 16-hr counts (from 6 a.m. to 10 p.m.) were used. The data from the PTCs were used to determine the 24-hr-to-16-hr relationship (24/16) for each counter and class of road. This relationship gives an indication of the proportion of nighttime traffic on a road section. In Figure 3 the nighttime traffic is shown as a percentage of the daytime traffic for the different counters and classes. The differences between the classes is statistically significant at the 99 percent confidence level.

Weekly Traffic Variation

A typical weekly traffic pattern is shown in Figure 4. One way of quantifying this pattern is to calculate the ratio of the traffic on a

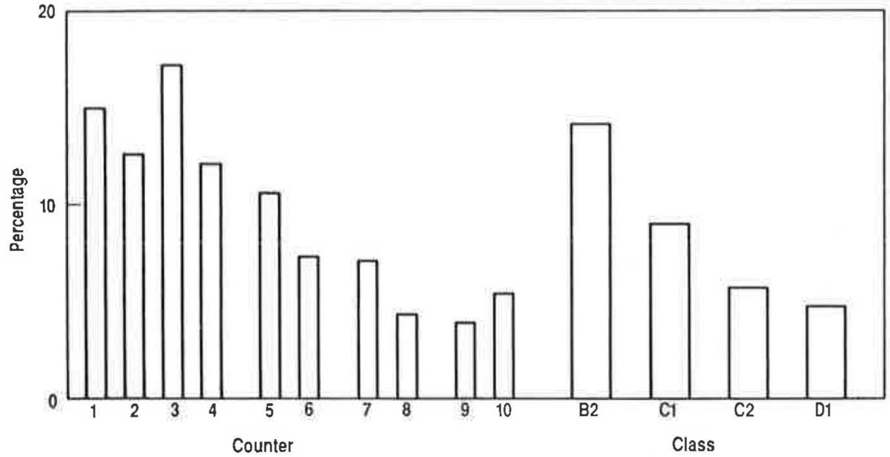


FIGURE 3 Nighttime traffic as a percentage of daytime traffic.

Friday to the weekly average daily traffic (*F/WD*). This is indicative of the amount of weekend traffic. The *F/WD* ratio for the different counters and classes is shown in Figure 5. This ratio differs very little for Classes B2 and C1.

Seasonal Traffic Variation

Long-term traffic variation on a road can be expressed in various ways, for example as

- Seasonal expansion factors,
- The coefficient of variation of the daily traffic volumes on a road, and
- The ratio between *E*-day and *N*-day traffic volumes (*E/N*).

Because *E*-days are not used for short-term counts and because of the exceptional traffic volumes on *E*-days, it is clear that the *E/N*-day ratio will be reflected in both the expansion factors and the coefficients of variation. It was therefore decided to use only the *E/N*-day ratio to quantify the seasonal traffic variations.

For this purpose 13 specific *E*-days were identified. They were mostly public holidays or days on which the provincial schools were closed. For each of these days the traffic volume was divided by the average *N*-day traffic. The *E/N*-day ratio for each counter and class is shown in Figure 6. It is interesting to note that on Class-D roads the traffic volume on *E*-days is on average less than on *N*-days. This was also found to be the case for roads carrying mostly commuter traffic, such as Class-A roads *(7)*.

EFFECT OF TRIP LENGTH ON TRAFFIC PATTERNS

It is well known that trip length has an effect on traffic patterns *(3)*. However, to determine the average trip length or trip-length distribution, an origin-destination survey is necessary. The cost of this is prohibitive when a large number of road sections must be considered.

At the end of 1983 the South African Rural Traffic Model *(5)* became available. This model was developed to predict future demand of traffic on the rural road network of South Africa. The model network, which consists of 1,324 links, covers 85 percent of the surfaced national and provincial roads in the country as well as some 5000 km of unsurfaced roads. By using the calibrated origin-destination matrix together with the assignment routine of the model, it was possible to calculate the average length of through trips for each link. Unfortunately, the links on which Counters

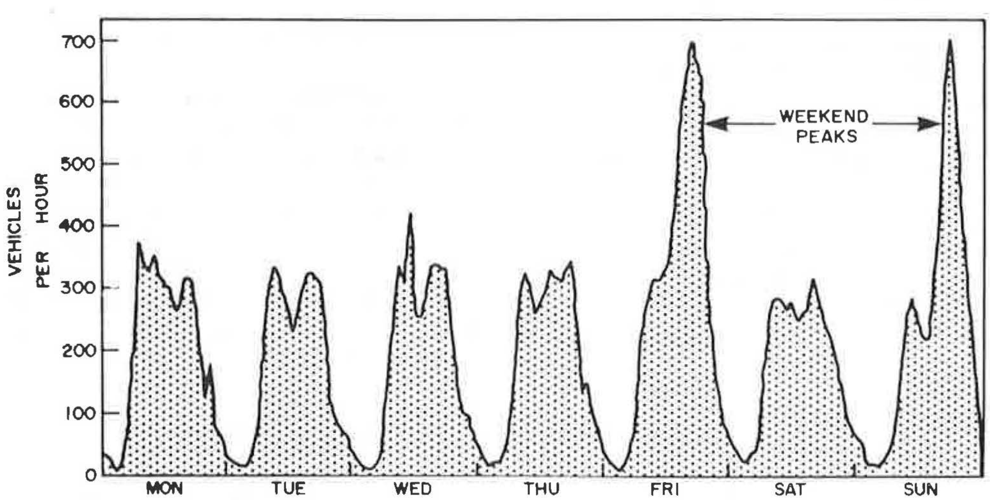


FIGURE 4 A typical weekly traffic pattern.

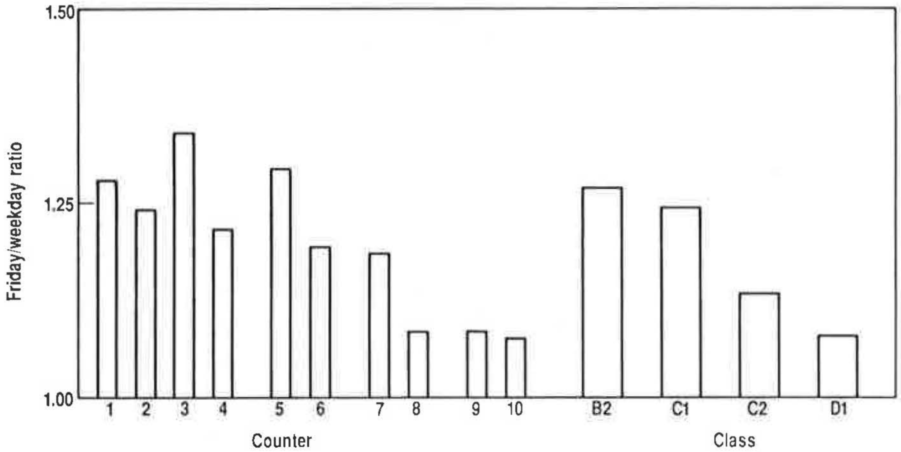


FIGURE 5 Friday-to-average weekday traffic.

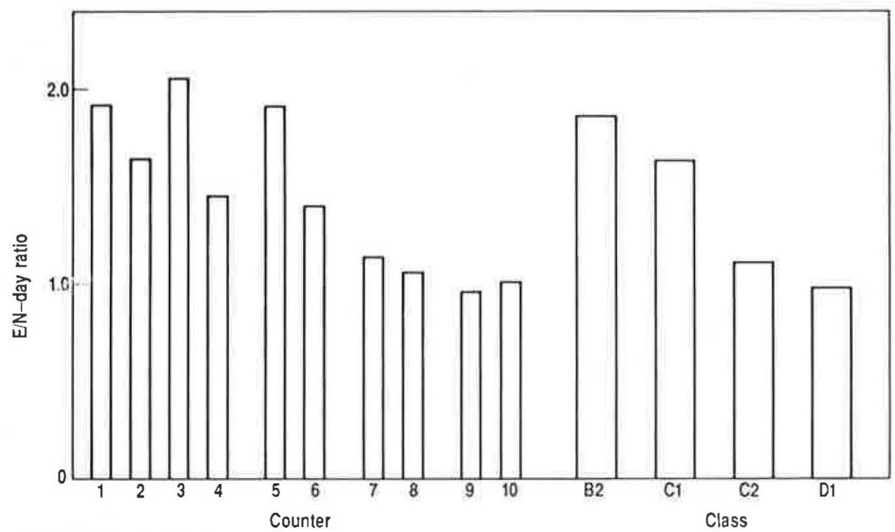


FIGURE 6 *E/N*-day ratio.

7–10 were located were not part of the model network. Therefore the following assumptions were made:

- For Class-C2 roads (counters 7 and 8) the average length of through trips is equal to the distance between the two neighboring towns (this follows from the definition) and
- For Class-D1 roads (Counters 9 and 10) the average length of through trips is equal to the length of the link on which the counter is located.

In Table 1 the average lengths of through trips (L) at each counter are given, together with the various traffic patterns as quantified in accordance with the methods described in the previous section. The correlation between the trip lengths and the traffic patterns can best be illustrated by the correlation matrix given in Table 2. From this it is evident that a good correlation exists not only between trip length and traffic patterns but also between the daily, weekly, and seasonal patterns themselves. The effect of trip length on the *E/N*-day ratio is shown in Figure 7.

With an estimate of through trip length available for the most important links in the South African rural road network, it is now possible to classify these links for the purpose of traffic counting. For the other, less important, links the assumptions made about trip length appear to be adequate.

Another way to determine trip length on a specific road is to do a license plate survey. In three of the four provinces in South Africa it is easy to identify the town of registration of a vehicle from the license plate. Where a manual short-term counting program is used, a sample of license plates can be recorded and used for the classification of the specific road section.

TABLE 1 TRIP LENGTHS AND TRAFFIC PATTERNS

Counter No.	Trip Length (km)	Traffic Patterns Daily (24/16)	Weekly (*F/WD*)	Seasonal (*E/N*)
1	463	1.150	1.279	1.92
2	368	1.126	1.241	1.64
3	557	1.172	1.340	2.05
4	176	1.121	1.216	1.46
5	358	1.106	1.294	1.91
6	159	1.073	1.193	1.40
7	45	1.071	1.184	1.14
8	53	1.043	1.084	1.06
9	40	1.039	1.085	0.96
10	35	1.054	1.076	1.01

TABLE 2 CORRELATION MATRIX

	L	24/16	F/WD	E/N
L		0.935	0.916	0.969
24/16	0.935		0.921	0.922
F/WD	0.916	0.921		0.966
E/N	0.969	0.922	0.966	

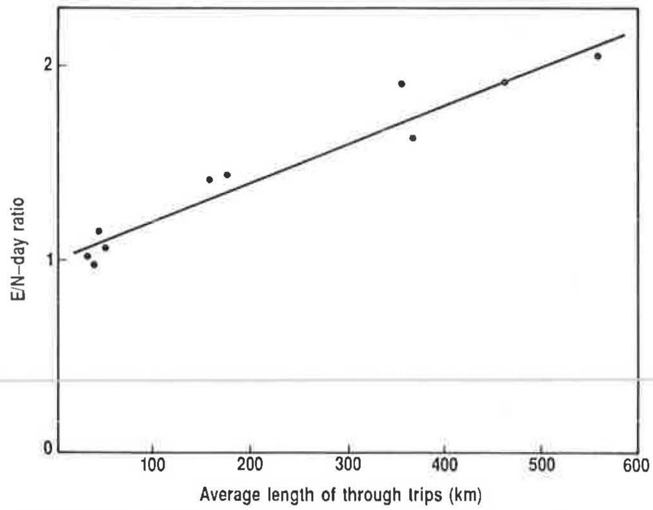


FIGURE 7 Effect of trip length on the E/N-day ratio.

CONCLUSIONS AND RECOMMENDATIONS

The results of the pilot study on traffic counting show that the subjective judgment of officials of a roads department can lead to a reasonable classification of the road network. When no data from PTCs or data on trip distance and trip purpose are available, subjective judgment is still the only way to classify the road network.

An interesting aspect of the results is the good correlation among the short-, medium-, and long-term traffic patterns, which indicates that roads that carry a high proportion of nighttime traffic also have relatively high traffic volumes during weekends and on holidays (E-days).

The most important finding of the study is that the average length of through trips on a link, as estimated by the South African Rural Traffic Model *(5)*, can be used for a uniform countrywide classification of the road network. It is therefore recommended that the following classification, based on trip lengths, be used:

Class	*Description*	*Trip Length*
A	Urban commuter roads	0–100 km
B	Intermetropolitan roads	>350 km
C	Interregional roads	100–350 km
D	Rural access roads	<100 km
E	Roads with exceptional traffic patterns	Not applicable

ACKNOWLEDGMENTS

This paper is presented with the permission of the Chief Director of the National Institute for Transport and Road Research, CSIR, Pretoria, South Africa. The contribution of the Roads Department of the Provincial Administration of the Orange Free State to the pilot study is gratefully acknowledged.

REFERENCES

1. *Guide for Traffic Volume Counting Manual.* Bureau of Public Roads, U.S. Department of Commerce, 1965.
2. P. H. Bellamy. *Seasonal Variation in Traffic Flows.* Supplementary Report 437. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1978.
3. S. C. 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.
4. *RSA National Rural Roads Needs Study.* South African Department of Transport, Pretoria, 1980.
5. P. W. Jordaan, W. W. Crous, and C. A. van Tonder. The South African Rural Traffic Model: Modelling Concepts and Forecasts. *Annual Transportation Convention*, Vol. H, Pretoria, Republic of South Africa, 1984.
6. D. T. Hartgen and J. H. Lemmerman. *Streamlining the Collection and Processing of Traffic Count Statistics.* Transportation Analysis Report 19. Planning Division, New York State Department of Transportation, Albany, 1982.
7. R. J. Sweet and D. N. Lockwood. *The Accuracy and Estimation of Annual Traffic Flow.* Council for Scientific and Industrial Research Manual K70. National Institute for Transport and Road Research, Pretoria, Republic of South Africa, 1983.

Evaluation of a Statewide Highway Data Collection Program

STEPHEN G. RITCHIE AND MARK E. HALLENBECK

This paper is a discussion of an in-depth evaluation study of highway data development and analysis activities of the Washington State Department of Transportation. Statistically based procedures and recommendations that were developed to streamline the highway data collection program are described. Opportunities to reduce manpower and equipment costs, streamline work activities, improve the quality of data collected, and provide accurate and timely data for the various users were identified. Given the focus on highway data, a major effort was devoted to the department's traffic-counting program. However, many data items and programs were considered, and the following items received particular attention: traffic volume counting, including estimation of annual average daily traffic at any location throughout the state highway system; associated seasonal, axle, and growth factors; vehicle classification; truck weight; and the relationships between the statistical sampling requirements recommended for these items and those associated with the FHWA Highway Performance Monitoring System (HPMS) in the state. Employing statistical sampling methods that complement the HPMS sample offers a strong potential for significantly improving the cost-effectiveness of a statewide highway data collection program.

In 1981, as a result of major budget cutbacks, the Washington State Department of Transportation (WSDOT) created a high-level committee to review the amount of highway data collected. The committee recommended a sharp reduction in the level of traffic counting. This decision was based primarily on stated data needs by upper-level management. The committee did not, however, address the statistical validity and quality of the data collected. Neither did the committee attempt to integrate the remaining data collection effort.

Thus, in recent years, considerable concern has existed about the appropriate level of resources to be allocated to various data collection activities and about the statistical basis for these activities. The shifting emphasis in WSDOT'S highway program from construction to maintenance and rehabilitation is another important factor. These issues are of concern to many state DOTs.

In this paper are presented the results of a research study that was undertaken to evaluate WSDOT's data collection and analysis activities. The statistically based procedures and recommendations that were developed to streamline these activities are described. The primary purpose of this program was to satisfy the internal needs of WSDOT, although all major users and uses were identified. A rigorous statistical approach to program design and data collection was necessary to permit estimation of data accuracy and to provide a rational basis to assist in allocating limited resources among the various possible data collection activities. Thus the study results should also be of interest to many other state DOT officials, particularly in evaluating their own programs and in complying with requests of the FHWA to integrate statewide traffic-counting activities with the Highway Performance Monitoring System (HPMS) *(1)*. In addition, the issues identified were of special significance to WSDOT given the development of a new Transportation Information and Planning Support (TRIPS) system. TRIPS is essentially a computerized, on-line, data base management system for assembling, maintaining, and reporting information about the state's highway network *(2)*.

S. G. Ritchie, Department of Civil Engineering and Institute of Transportation Studies, University of California, Irvine, Calif. 92717. M. E. Hallenbeck, Washington State Transportation Center, University of Washington, Seattle, Wash. 98195.

BACKGROUND

Overview of Previous Work

Historically, highway data and specifically traffic count data have been collected by state transportation agencies to support a wide range of programs and needs. These have included the use of traffic count data to develop estimates of annual average daily traffic (AADT), vehicle miles of travel (VMT) and design hour volume (DHV) for individual highway sections, and functional classifications of highways and regional or other divisions of the state highway system. In addition, the FHWA has required submission of various traffic and truck data and estimates for use by FHWA and other federal agencies. These have been required in order to establish national travel trends, prepare reports requested by Congress, plan for future transportation needs, and assess the overall efficiency of various programs and policies.

Several studies have been reported in recent years that relate to general efforts to develop more cost-effective approaches to statewide highway data collection. These include the work of Hallenbeck and Bowman *(3)*, which proposed a general statewide traffic-counting program based on the HPMS *(1)*; the study by Wright Forssen Associates *(4)*, which evaluated and developed improvement recommendations for the highway traffic data program of the Alaska Department of Transportation and Public Facilities; and work by the New York State Department of Transportation to streamline and reduce the cost of its traffic-counting program *(5)*. Although each of these studies provided useful background and guidance for this project, the conceptual basis of Hallenbeck and Bowman *(3)*, using the HPMS framework for purposes of statewide highway data collection, appeared particularly promising.

The HPMS was introduced by FHWA in 1978 to consolidate many previous federal data requirements and to strengthen the methods used by the states for collecting, estimating, and reporting traffic count data. It involves a sample of highway sections that provide a basic set of traffic count locations for which geometric, operational, and traffic volume data are to be available on a continuing basis. Employing statistical sampling methods that complement the HPMs sample appeared to offer a strong potential for significantly improving the efficiency of a highway data collection program by coordinating the collection of traffic count data, vehicle classification data, and truck weight data. This approach was explored in this study as a possible basis for overall program design.

There are a number of other relevant and useful works in the general area *(6–13)*.

Study Approach

Given the focus of this study on highway-related data, a major effort was devoted to the department's traffic-counting program. A number of data items and programs were considered. Data uses and users and their needs were determined by building on work previously performed within the department as a result of the 1981 budget cutbacks. This also involved reviewing available literature on the subject of statewide traffic data collection. The two primary literature sources were

- FHWA's *Draft Traffic Counting Guide (14)* and
- The technical basis for that guide, *Development of a Statewide Traffic Monitoring Guide Based on the Highway Performance Monitoring System (3)*.

A total of 45 major uses of traffic information were identified. These uses were broken into the categories given in Table 1. In all, 14 types of traffic information were identified and could be further categorized as belonging to five groups:

1. Volume
   - Average annual daily traffic (AADT)
   - Design hourly volume (DHV)
   - Peak-hour traffic percentage (K)
   - Directional split (D)
   - Peak-hour volume turning movements
   - Vehicle miles of travel (VMT)
2. Vehicle classification
   - Average daily truck traffic (ADTT)
   - Percentage of trucks in peak (T)
   - Percentage by vehicle class
3. Truck weight
   - Truck weights
   - Equivalent 18-kip axle loads (EAL)
4. Speed data, percentage of vehicles by speed range
5. Accident data, state highway patrol accident reports

Vehicle speed information was dropped from the analysis when it was determined that the department was performing speed studies as mandated by federal regulation and had no desire to refine or expand this data collection process. Further, the department does not perform the field data collection for accident analyses. This information is supplied by the state patrol on computer tapes. It was concluded that existing procedures and data were sufficient to meet the department's needs.

The specific data items to be addressed in the study were determined to be the "system" traffic data estimates (not project-level estimates) collected by the Data Office of the department's Planning, Research and Public Transportation Division. Roadway information and pavement condition data were excluded from the scope of the project.

One of the most difficult tasks in the study was the attempt to establish appropriate statistical levels of confidence and precision to serve as objectives in the sample design process. The study team went back to all identified users of traffic information to elicit their data quality needs. As a result of this effort, it was soon realized that the vast majority of the data users could not articulate a need for a specific level of data precision for their analyses. All available literature was then reviewed in an attempt to learn if statistical standards had been suggested by other researchers. To a large extent, this also proved to be fruitless.

Because such sources failed to provide the needed guidance, a selected number of sensitivity analyses and statistical derivations were undertaken to examine the effect of data quality on the results of particularly important analyses. Among the analyses examined were

- Priority array determination (a complex set of ranking procedures used by WSDOT to objectively establish the need for highway system improvements),
- Pavement overlay calculations,
- New pavement design,
- Bridge design,
- Pavement management system, and
- Determination of level of development.

This information was supplemented by the small amount of guidance available from data users and published literature and a large amount of professional judgment on the part of project staff and a WSDOT technical committee.

While the investigation of data needs proceeded, the project team reviewed the current activities of the Data Office. Included were data being collected, methods for determining locations of data collection, and manipulations performed on the data collected before they are provided to users. This information was later compared with the data needs determined at the beginning of the project to investigate the strengths and weaknesses of the existing data collection procedures.

Information was also obtained from both the department and FHWA to assist in assessing the variability of data (i.e., the variation in traffic volumes, truck travel, etc. among days, locations, and seasons). Current costs of data collection were also gathered. This information was used to estimate the sample sizes needed to meet the department's needs for accuracy (precision) and to determine the approximate cost of meeting those needs.

After this information was gathered, several alternatives were developed to meet the identified needs of the department. This information was presented to the study's steering and technical committees for review.

EXISTING DATA COLLECTION AND MANIPULATION PROCESS

Volume Data

Permanent Traffic Recorders

Currently, about 80 pieces of permanent traffic recorder (PTR) equipment collect data year round at approximately 65 locations (more than one counter is necessary at some locations to handle multiple lanes or several different legs of intersecting roadways).

The PTR data provide information for calculating

- Seasonal adjustment factors for converting short-duration counts to AADT estimates,

TABLE 1 USERS' DATA NEEDS

Use	AADT	Average Daily Truck Travel	DHV	Peak-Hour Fraction	Directional Split	Truck Percentage	Peak-Hour Volume	Equivalent Axle Loading	Truck Weight	Turning Movement	Vehicle Class (%)	VMT	Speed	Accidents
Project-level traffic forecast	1		1	1	1	1	1			1	1			
Highway geometric design	2		2	2	2	2	2			2				2
Highway pavement design	3					3		3	3		3			
Project-level bridge design	4	4												
Signal warrants	5									5	5			5
Intersection design	6									6				
Traffic engineering control and operation	7												7	7
Speed study analysis	8												8	
Vehicle weight enforcement	9					9			9					
System-level traffic forecasting	10											10		
System-level bridge design	11													
Long-range transportation system planning	12											12		
Capacity needs analysis	13			13	13	13	13							
Highway performance monitoring system	14			14	14	14					14	14	14	14
Pavement management system	15					15								
Model calibration and validation	16		16											
Survey control	17													
Freight analysis movement	18			18	18	18	18							
VMT determination	19											19		
Flow maps	20													
Priority array	21			21	21	21								21
Project-level investment analysis	22						22							22
Maintenance programming	23													
Maintenance scheduling	24													
Accident analysis	25											25		25
Safety studies	26											26		26
Air quality	27					27					27		27	
Water quality	28													
Noise quality	29						29				29		29	
Economic impact of development	30													
Energy consumption	31										31			
Economic studies and analyses	32										32		32	32
Revenue	33										33			
State patrol	34											34	34	34
Traffic safety commission	35											35	35	35
Commerce and economic development	36													
AAA	37													
Motel chains	38													
Service station chains	39													
Chamber of commerce	40													
Outdoor advertising	41													
Litigation of tort claims	42		42	42	42	42	42			42	42		42	42
Damage recovery	43													43
Construction manpower planning	44													
Maintenance manpower planning	45													

- Estimated design hour (and other design) factors for non-PTR locations, and
- Growth trends.

PTRs also provide volume information (in terms of vehicles, not axles) for the sections of highway on which they are located.

Currently, these data are collected using a telemetry system, to which the department recently converted. This conversion has reduced the amount of manpower needed to collect and manipulate PTR data.

Short-Duration Counts

The majority of traffic volume counts taken by the department falls into this category. Short-duration axle counts usually last 72 hr but may also be of 48- and 24-hr lengths. Because of budget cutbacks in 1981, short-period counts are currently collected only when requested for specific projects or when manpower is available to place and retrieve counters while other tasks are being performed. In 1983, 2,281 such counts were made.

The data collected from these counts are seasonally adjusted and entered into the existing traffic volume data base for future reference. Volume data already in the data base, and not replaced by a new volume count, are adjusted annually to reflect VMT growth in the state. The seasonal factors applied to each raw count are derived from available PTR data. A transportation data office engineer or technician determines the particular PTR or PTRs to be used for the factor on the basis of his knowledge of the road being counted, the roads that contain PTR stations, and a book containing previous estimates of seasonal factors for various road sections (based on old PTRs, old control counts, and professional judgment).

In most cases, axle correction factors have not been applied to the raw axle counts, which results in systematic overestimation of vehicular traffic on state highways.

Vehicle Classification Counts

Vehicle classification data are collected at both project and PTR locations. For project-specific counts, vehicle classification counts are either 6-hr manual counts or part of 4-hr manual intersection counts.

At PTR stations counts are now performed on a quarterly basis to better understand the vehicle mix present on the state highway system. Consideration is also being given to automatic vehicle classification, based on vehicle lengths, at PTRs. However, the department's usable vehicle classification data base is currently insufficient for estimating seasonal or locational variation in truck travel for most of the state highway system.

The principal vehicle classification need of the department was judged to be the number (or percentage) of trucks in each of the following categories:

- Two-axle trucks (not including pickups),
- Three-axle trucks,
- Four-axle trucks,
- Five-axle trucks, and
- Trucks with six or more axles.

These categories are more aggregate than those now requested by FHWA for use in data submittals and the manual classification categories actually collected by department field crews.

Truck Weight Data

Currently, truck weight data for purposes other than enforcement are only collected as part of the FHWA's long-term pavement monitoring (LTPM) program. These weighings are being used in lieu of truck weighings that would normally be performed as part of the federal biennial truck weight survey. This program has been temporarily suspended by FHWA pending the outcome of ongoing research on various weigh-in-motion (WIM) strategies. Data are collected using low-speed WIM scales at specific sites selected for the LTPM study.

Data resulting from this effort are sent to FHWA. After analyzing the data, FHWA provides vehicle weight, average equivalent axle load (EAL), and equivalent wheel load (EWL) data to the department for use in construction and pavement management functions.

PROGRAM EVALUATION

Overview

In a limited sense, the existing data collection program fulfills the majority of the department's current needs. The program can be characterized as the lowest possible level of data collection permissible to meet immediate data needs with resources concentrated in those areas that most significantly affect departmental finances. Although this low level of data collection results in the lowest short-run cost to the department, it also causes some data deficiencies that quite possibly could cost more money than is being saved. A summary of findings follows.

- The department generally has relatively good project-level data but an old and increasingly obsolete base traffic data file;
- The department does little traffic counting other than at project locations;
- An axle correction factor is not currently applied to raw axle counts (although this is being changed);
- Ad hoc seasonal factors are applied manually, as opposed to statistically derived factors and an automated approach;
- No HPMS data are collected by WSDOT off the state highway system;
- The state currently lacks an adequate vehicle classification data base, and existing programs are insufficient to significantly improve that data base;
- The only vehicle weighings being performed for planning purposes are part of the federal LTPM study and are inadequate for cost-effective pavement design; and
- It is unclear how statistically valid the data from these efforts are when used for analyses covering the entire state because the data are not being collected in a statistically rigorous manner.

Volume Counting

As was stated earlier, volume counting consists primarily of project-related traffic counting. This means that non-project-related counts tend to cluster around project locations because field crews do not have the time or travel allowance to move away from the project area when collecting these counts.

Although any one nonproject data need might not appear that "important," the combined impact of these analyses can be significant. Further, because traffic counting is centered around project sites, those parts of the state not involved in major projects will have little or no traffic counting performed. As the counts in these areas grow older, users of those data start to question (sometimes rightly) the validity of the available traffic estimates. Considering that these estimates are included in such analyses as the priority array, the HPMS submittal (which includes the information used to apportion Interstate 4R funds), and other non-site-specific analyses, the state has a need to maintain the quality of traffic information on road sections that are not project locations. In addition, system-level data have, in recent years, been used for pavement overlay design purposes when location-specific data could not be collected in time. This represents a very significant use of system data.

Factoring and Data Manipulation

Currently, most of the factoring and data manipulation performed by the department is done manually. The department supplies traffic estimates in terms of automobile equivalents and an estimate of the percentage of truck travel, but it does not automatically apply an axle correction factor.

The current seasonal factor process also requires a considerable amount of judgmental intervention. This can lead to inconsistencies because two different engineers or technicians using the same volume counts might develop considerably different AADT estimates based on their individual perceptions of what the "correct" seasonal factor should be.

Thus a consistent, statistically valid seasonal factoring procedure is required. TRIPS *(2)* provides an ideal tool for automatically performing all necessary factoring procedures for converting raw data into useful traffic estimates. The data for calculating the necessary factors are already collected as part of the ongoing traffic-counting program. Therefore such factors could be stored and utilized as a series of tables created within TRIPS. These tables could then be used on a look-up basis for application to any raw traffic count.

Vehicle Classification

The department collects few vehicle classification data, and the majority of these are stored in a manner that makes them unavailable to most users of departmental data.

The biggest difficulty with the existing data collection effort, however, is that the department has no knowledge of how truck travel changes seasonally, from month to month, or from day to day on its highway system. Because of this, short-duration counts (e.g., 6-hr manual counts at project sites) cannot be expanded to an average annual total at that location with any degree of accuracy or confidence. Designs based on the collected data are therefore not likely to be as precise as they should be.

A further problem with the current data collection method is that no statistically valid estimate can be made of truck travel in the state or on the state highway system. This becomes a serious problem when viewed in conjunction with traffic forecasting for pavement design. The pavement design process allows for the changing of truck travel percentages over time (e.g., if truck travel is expected to grow, more EALs will be applied to the pavement over its design life, and the pavement will need to be correspondingly thicker). At this time, the department has no knowledge of how those percentages have changed and, consequently, has little basis for forecasting such travel.

Truck Weighing

The department's truck weighing consists of the LTPM data collection described previously. This data collection is probably insufficient for the department's needs, but it is appropriate given the equipment and resources currently available to the department.

The biggest problem with this data collection procedure is that it cannot account for biases that are apparent in every above-ground weighing system. Heavy and overweight trucks tend to avoid scales, even when those scales are not used for enforcement purposes. As a result, the weights that are obtained tend to underestimate the average weight of trucks on the highway system.

To collect the data that are really needed, the department will need to acquire a weighing system that is unobtrusive to truckers so that avoidance of the scales is not a problem. When such equipment is available, the state can expand on the LTPM sample for weighing. The LTPM sites are a good start for an appropriately sized sample, but the existing sample size is relatively small for estimating statewide averages.

RECOMMENDATIONS

Overview

If the department wished to collect all of the data requested by users, it would need to collect volume counts at 0.1-mi intervals on all state highways (requested as an input into the priority array) as well as similar amounts of vehicle classification data and lesser amounts of vehicle weight data. This is obviously an impossible task for a state highway system approximately 7,000 mi in length. The recommended program therefore consists of two data collection tiers:

- Project-specific data collection and
- Statistically based statewide sampling.

The intent of this program structure is to ensure the minimum base of information necessary to supply system estimates, maintain the quality of the most important department analyses, and minimize the total cost of the program.

The statewide element consists of taking counts at a limited number of locations on a routine basis to provide the department with statistically valid estimates of statewide vehicle travel. The detailed statistical basis of this program is described in the paper by Ritchie in this Record. Direct uses of this statistical sample include estimating

- Statewide VMT,
- Average percentage of travel by truck versus automobile,
- Statewide axle correction factors, and
- Truck weights.

These data are needed as the best alternative to site-specific data. Nowhere is the use of these system averages more prevalent than for estimating truck travel for pavement overlay purposes, one of the major tasks of the department (approximately $100 million is spent annually on pavement resurfacing).

Statewide data collection, in particular, needs a statistically valid sample. This provides the department with a rational means for understanding the quality of the data it is using for factors and defaults in all of its analyses. The department's sample is most appropriately taken as part of the FHWA's HPMS data base.

Although the HPMS sample has limitations, it provides the most cost-effective basis for choosing samples for statistically valid data collection.

Unlike the first tier of project-specific data collection, in which only volume and vehicle classification data are collected, the second or statewide tier should collect volume, vehicle classification, truck weight, and speed data. The department's volume-counting locations already exist in the form of the HPMS volume sample. The vehicle classification locations should be taken as a subset of the volume count locations. The truck weight sample should in turn be taken as a subsample of the vehicle classification sample.

It is recommended that the statistically valid sample be taken on a 3-year cycle. That is, only one-third of the total number of sample locations should be counted in any given year. This cycle length is recommended by FHWA because

- Traffic changes (on a systemwide level) occur relatively slowly and
- The 3-year cycle is reasonable in terms of the amount of data that needs to be collected in any given year.

This recommendation applies to all HPMS counts (volume, vehicle classification, and truck weights) but does not include the speed survey, which is based on a 1-year sampling cycle.

The department needs to review the HPMS sample count locations it collects data for and divide those sections into three roughly equivalent count groups, for counting over the 3-year cycle. The department then needs to institutionalize a yearly review of proposed project count locations and HPMS count needs. This should be done at the time project counts are being scheduled. The review simply entails the comparison of proposed project count locations and those HPMS locations that are scheduled for counts that year. The HPMS sections not scheduled for project counts will then need to be added to the yearly count schedule as most appropriately fits the department's manpower scheduling.

Finally, all traffic data collected by either the WSDOT Data Office or the districts should be input into the new TRIPS system, which would make these data available for other departmental uses. In this manner systemwide data collection will be supplemented by the more extensive counts taken at project locations. The result will be a more up-to-date traffic-counting base file.

Volume Counting

HPMS Needs

The data collected for the HPMS submittal meet the needs of the department and the FHWA for Interstate 4R appropriations and priority array calculations.

The current FHWA request for HPMS volume data consists of yearly traffic counts on all Interstate sample sections and new counts on one-third of all other sample sections. New volume counts are requested for 48 hr at one time at each location.

This annual level of traffic counting represents a need for 483 short-duration count locations (or 781 traffic counter settings):

- 222 sample sections on Interstates (444 traffic counter settings) and
- An average of about 261 locations (337 traffic counter settings) on other state roads.

Each year some of these locations will be counted via project counts and existing PTRs. The department does not directly collect information on HPMS sections off the state highway system. If FHWA were to insist on the department collecting this information as well, the second of the previous estimates would increase to approximately 700 locations or 1,050 counter settings per year.

Project Counts

In fiscal years 1984 and 1985, the Data Office provided project counts at roughly 110 and 100 separate locations, respectively. These numbers are similar to expected levels of project counting for the near future.

Counting for the average project includes roughly

- Ten 72-hr machine axle counts,
- One 6-hr manual vehicle classification count, and
- Two 4-hr manual intersection counts.

This process requires 1 man-week of field crew effort, including travel time but not including supervision or data reduction.

Manpower Needs

It is estimated that the Data Office needs about 3.5 full-time-equivalent (FTE) employees to perform the field data collection for the HPMS and project-specific counts described. This estimate is based on the following considerations:

- Between 100 and 130 projects per year will require project-specific information (i.e., approximately 1,300 counts);
- For each project, 1 man-week of field effort is required to provide the necessary data, for a total of 130 person-weeks; and
- For HPMS, roughly 600 counter settings not included in the project counts will be necessary; conservatively, these HPMS counts will require 45 person-weeks of field data collection to perform.

This proposed reorganization represents a total of 175 man-weeks of effort, or 3.5 FTEs, which is roughly equivalent to current levels. However, in addition to these 3.5 FTEs, personnel time will be needed for office support, data reduction, and supervision of field crews, as is now the case.

Permanent Traffic Recorders

One of the most important uses of PTRs is for estimating seasonal factors.

The factor process currently used by the department makes extensive use of subjective selection of seasonal factors. The recommended factor process (see paper by Ritchie in this Record) places PTRs in the following groups for the estimation of seasonal factors:

- Rural Interstates,
- Urban roads,
- Other rural roads in the northeastern part of the state,
- Other rural roads in the southeastern part of the state,
- Other rural roads in the southwestern part of the state,
- Other rural roads in the northwestern part of the state, and
- Central mountain passes.

Each of the counties in the state is assigned to one of the four "other rural" factor groups.

To supply the data necessary for estimating seasonal factors, the department needs between three and eight PTRs in each of the seven factor groups (*13* and paper by Ritchie in this Record). Strictly on the basis of need for seasonal factors, the department could eliminate at least 10 PTR locations. This would result in savings of roughly $300 per month ($3,600 per year). This is a fairly small sum given the amount of data the counters generate and their potential for providing other useful vehicle classification information to the department.

Vehicle Classification Counts

Like volume counting, vehicle classification information needs to be provided on both a systemwide and a site-specific basis. The existing program element provides a limited amount of project data and very little systemwide information.

The recommended vehicle classification program is similar to the volume count program. The HPMS is used as the basis for providing a statistically valid estimate of travel by vehicle type, and project-specific counting is performed as necessary for individual analyses. The use of permanent vehicle classifying counters (i.e., 365-day-per-year counts by vehicle type) at PTR locations is also recommended to provide the state with knowledge of the seasonal variation of truck travel throughout the year. Existing PTRs have the capability of collecting vehicle length information, but they cannot yet be interrogated by the telemetry system to provide classification information. An interim recommendation was made for 20 existing PTR locations to be upgraded to further investigate seasonality.

It was recommended that the department collect a statistically valid statewide sample of 452 vehicle classification counts on six strata:

- Rural Interstates,
- Urban Interstates and other freeways and expressways,
- Rural principal arterials,
- Urban principal arterials,
- Rural minor arterials and collectors, and
- Urban minor arterials and collectors.

The recommended counts and levels of precision for each of these strata are given in Table 2. For rural Interstates this level of precision means that the percentage of travel by five-axle trucks on rural Interstates can be estimated within 15 percent with 90 percent confidence. These levels of precision were chosen primarily on the basis of

- Similarity to suggested levels of precision expressed by FHWA in the draft counting guide *(14)*,
- The importance of each stratum of highways to the department, and
- The cost-to-benefit ratio of collecting better or worse information for each stratum.

TABLE 2 RECOMMENDED NUMBER OF VEHICLE CLASSIFICATION COUNTS AND LEVEL OF PRECISION FOR THE MEAN PERCENTAGE OF TRAVEL BY FIVE-AXLE VEHICLES

Roadway Category	No. of Counts	Relative Precision[a] (%)	Level of Confidence (%)
Rural Interstates	104	±15	90
Urban Interstates	99	±15	90
Rural principal arterials	99	±20	80
Rural minor arterials and collectors	83	±20	80
Urban principal and minor arterials and collectors	67	±20	80

[a]In estimating the average percentage of travel by five-axle combination trucks on the stated roadway category.

The counts in Table 2 would be taken during the 3-year counting cycle. These counts involve a single count day at a given location, randomly selected from all days in the count year including weekend days as well as weekdays.

These counts would be taken at HPMS volume sample locations. They would preferably be 48-hour, automatic (i.e., machine as opposed to manual) counts. The purchase of 10 additional vehicle classifiers was recommended for this purpose. These counts would also meet the need for volume counting at those locations to meet the systemwide needs described in the previous section. It was estimated that 0.75 FTE would be required for this counting element, an increase of 0.4 FTE over current manpower. Until the PTR classification program is in place, the department should probably use 6-hr manual counts in conjunction with its 48-hr HPMS volume counts. Although the longer count duration is preferable, the benefits to be gained by taking vehicle classification counts for 24 to 48 hr in place of 6 hr do not exceed the costs of performing that counting manually.

It was also recommended that the department update its vehicle classification categories and use FHWA's 13-category classification *(14)*.

Truck Weighing

The truck-weighing program element has a slightly different structure than the volume and vehicle classification elements. Currently, the department does not collect project-specific truck weights. As a result, the recommended program structure is for a statistically valid sample of truck weighings to be carried out at HPMS vehicle classification count locations, including the FHWA LTPM sites. Further research is warranted to determine the feasibility, desirability, and cost of collecting project-specific vehicle weights. Results from current in-state testing of bridge-WIM and piezoelectric cable weighing systems should assist in this analysis.

The interim recommended truck-weighing program is therefore to weigh at least 200 vehicles with five or six or more axles at each of five locations on each of three strata. Thus 15 annual surveys would be involved. It was estimated that this program would save 0.4 FTE over current levels. The three strata for weighing are

- Rural Interstates,
- Urban Interstates, and
- Rural principal arterials.

This means the department will need two new rural Interstate and four new urban weighing locations. Average weights per vehicle type for urban Interstates would be used for all urban road designs, whereas average weights per vehicle type for rural principal arterials would be used for all non-Interstate rural highways. The department may choose to sample from lower functional class roads as well.

The recommended weighing element also differs from the volume and vehicle classification elements in that the sampling framework is not based on the number of days counting should take place but on the number of trucks that should be weighed at each location. This sampling scheme is currently used by Wisconsin DOT. This scheme was chosen because it is the only method for which data were available to estimate required sample sizes. The recommended weighing program is given in Table 3.

This sampling program involves several basic assumptions:

- Truck weights by vehicle type do not change over the course of the year (i.e., the average 3 S2 truck weighs the same in July as it does in February);
- Truck weights do not vary between weekdays and weekends;
- Truck weights do not change with the time of day;
- Truck weights by vehicle type are not different on high-volume roads than on low-volume roads (i.e., an average 3 S2 on a low-volume rural principal arterial weighs the same as an average 3 S2 on a high-volume principal arterial); and
- The act of weighing does not bias the data being collected (i.e., trucks do not intentionally bypass the weighing location).

The most significant impact of this interim data collection scheme is evident in the amount of field crew time spent at each truck weight location. For high-volume roads, the time needed to weigh the appropriate number of trucks will be fairly small, certainly less than 24 hr. In the case of Interstate highways, one standard shift of the field crew may be sufficient. For low-volume roads, the field crew may need several days to collect the desired number of truck weighings.

TABLE 3 RECOMMENDED TRUCK-WEIGHING PROGRAM

Vehicle Type	No. of Vehicles to be Weighed at Each Location[a]	Resulting Precision[b] (%)	Confidence Limits (%)
Two-axle, four-tire, single units	200	35	80
Two-axle, six-tire, single units	200	16	80
Single units with three or more axles	200	20	80
Three-axle combinations	200	19	80
Four-axle combinations	200	10	80
Combinations with five or more axles	200	10	95
Five-axle doubles	200	11	95
Doubles with six or more axles	200	14	95

Note: Strata are rural Interstates, rural primary arterials, and urban Interstates. Weighing is done at five locations per stratum.

[a]The controlling vehicles should be five- or six-or-more-axle doubles on the Interstate system, and five-or-more-axle combinations on the rural primary system. All trucks for all other categories should be weighed. If more than 200 are weighed per location, the precision estimates should be better than those indicated here. If fewer than 200 are weighed, precision may be worse than indicated here.

[b]Of estimated mean weight per vehicle type.

Calculated Factors

There were three primary areas in which changes were recommended to the existing departmental process for estimating the various factors applied to raw traffic counts:

- Seasonal factors,
- Axle correction factors, and
- Growth factors.

The raw data needed to estimate these factors are already collected as part of the counting strategies. The statistical framework for deriving and applying these factors, particularly the seasonal factors, is described by Ritchie elsewhere in this Record.

CONCLUDING COMMENTS

The statistically based procedures and recommendations that were developed as a result of an in-depth evaluation of the WSDOT highway data collection program have been described. The evaluation framework used (which would be applicable to other state DOTs) focused on data requirements of users; sampling plans for the various components; data collection, count processing, and data management and storage procedures; count and processing equipment requirements; staffing requirements; and procedures for implementation of the recommendations. Opportunities were identified for streamlining work activities, improving the quality of the data collected, and providing accurate and timely data for the various users. Although the overall recommended level of volume counting and total manpower for field collection of those counts are not significantly different from current levels, the recommended program serves the department's needs much more effectively, not only in the short run but, perhaps more important, in the medium and long run. Also, a statistically based approach permits a rational determination of the quality of data being used in important analyses. When resources are limited and insufficient for the desired sample size, the trade-offs among sample size, precision, and level of confidence are explicit. If statistical sampling methods that complement the HPMS sample are employed, a strong potential exists in many states to significantly improve the cost-effectiveness of a statewide highway data collection program.

ACKNOWLEDGMENT

Many individuals from WSDOT, the Washington State Transportation Center (TRAC), and FHWA contributed significantly to this study, and their input is gratefully acknowledged. This paper is based on a research project conducted by the University of Washington and TRAC for the Washington State Transportation Commission and Department of Transportation in cooperation with the FHWA, U.S. Department of Transportation.

REFERENCES

1. *Highway Performance Monitoring System Field Manual for the Continuing Analytical and Statistical Data Base.* Office of Highway Planning, FHWA, U.S. Department of Transportation, 1984.
2. Arthur Andersen & Co. *TRIPS System.* Management Summary Report. Washington State Department of Transportation, Olympia, May 1984.
3. M. E. Hallenbeck and L. A. Bowman. *Development of a Statewide Traffic Counting Program Based on the Highway Performance Monitoring System.* Final Report. FHWA, U.S. Department of Transportation, 1984.
4. Wright Forssen Associates, Comsis Corporation, and TRANSPO

Group. *Highway Traffic Data Study,* Vol. I: *Final Report.* Alaska Department of Transportation and Public Facilities, Planning and Programming Division, Southeast Region, 1983.
5. D. T. Hartgen and J. H. Lemmerman. "Streamlining Collection and Processing of Traffic Count Statistics." In *Transportation Research Record 928,* TRB, National Research Council, Washington, D.C., 1983, pp. 11–20.
6. Peat, Marwick and Mitchell. *Guide to Urban Traffic Volume Counting.* Final Report. FHWA, U.S. Department of Transportation, 1980.
7. J. DiRenzo, L. Bowman, and M. Hallenbeck. *Review of the Department's Traffic Count Program.* Pennsylvania Department of Transportation, Harrisburg, 1983.
8. John Hamburg & Associates. *Improved Methods for Vehicle Counting and Determining Vehicle-Miles of Travel—Final Report on the State of the Art.* NCHRP, TRB, National Research Council, Washington, D.C., 1982.
9. John Hamburg & Associates. *Improved Methods for Vehicle Counting and Determining Vehicle-Miles of Travel—Traffic Counting Program Design.* NCHRP, TRB, National Research Council, Washington, D.C., 1982.
10. L. T. Hoang and V. P. Poteat. "Estimating Vehicle Miles of Travel by Using Random Sampling Techniques." In *Transportation Research Record 779,* TRB, National Research Council, Washington, D.C., pp. 6–10.
11. L. M. Rudman. "Vehicle Kilometers Traveled: Evaluation of Existing Data Sources." In *Transportation Research Record 726,* TRB, National Research Council, Washington, D.C., 1979, pp. 19–28.
12. D. L. Greene and A. S. Loebl. *Vehicle-Miles of Travel Statistics, Lifetime Vehicle-Miles of Travel and Current Methods of Estimating Vehicle-Miles of Travel.* Oak Ridge National Laboratory, Oak Ridge, Tenn., 1979.
13. J. P. Mahoney. *Traffic Data Needs for Pavement Rehabilitation.* Washington State Transportation Commission, Olympia, 1983.
14. *Draft Traffic Monitoring Guide.* Office of Highway Planning, FHWA, U.S. Department of Transportation, July 1984.

The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Transportation Commission, the Washington State Department of Transportation, or the FHWA. This paper does not constitute a standard, specification, or regulation.

Estimating Sampling Error for Cluster Sample Travel Surveys by Replicated Subsampling

DON L. OCHOA AND GEORGE M. RAMSEY

The California Department of Transportation conducted in-person home interview travel surveys in six counties of the state before converting to the telephone survey technique in 1979. The surveys updated existing data bases that support the development of regional travel forecasting models. During the survey period cluster sampling was employed to minimize travel time for survey interviewers and facilitate call-back procedures. Because cluster sampling was used, the simple random sample model often cited [$s/(n^{1/2})$] was not appropriate for estimating sampling error because that formula tends to underestimate actual standard errors. Estimates of sampling error for the surveys were thus made using the method of "replicated subsampling," which takes sample clustering into account and yields a higher total standard error than does the conventional method. This paper is intended to illustrate application of replicated subsampling in estimating sampling error for cluster sample travel surveys. Comparisons of standard errors derived using the method of replicated subsampling are made with standard errors derived by the conventional formula, which assumes a simple random sampling design. Replicated subsampling provides an unbiased, reliable, and generally applicable framework for estimating sampling error.

The California Department of Transportation (Caltrans) conducted in-person home interview travel surveys in the counties of Fresno, Kern, Sacramento, San Diego, San Joaquin, and Stanislaus before converting to the telephone survey technique in 1979. The six regional travel surveys, conducted in 1977 and 1978, updated data bases that support the development of regional travel forecasting models and augmented the data base of California's more extensive 1976–1980 Statewide Travel Survey. Travel survey findings that were previously reported *(1)* will not be discussed; rather, application of W. Edwards Deming's method (*2*, pp.87–101) of "replicated subsampling" for estimating sampling error, particularly for cluster sample travel surveys, will be demonstrated. The conventional standard error formula [$s/(n^{1/2})$] is not appropriate to use on cluster samples because it assumes a simple random sampling design and usually underestimates actual standard errors.

SAMPLE SELECTION

Because of budget and time constraints for conducting the surveys, only 500 households were sampled in each of the survey regions, except in the San Diego area where 1,000 households were sampled. The larger sample size in the San Diego area was in recognition of the complexity of transportation problems and the need for more highly stratified information in that region. A brief discussion of the sample selection process follows.

Sample selection for each of the surveys involved a three-stage process. In the first stage, 25 census tracts were systematically selected (except in the San Diego area where 50 tracts were selected). From a random start, the census tract containing every *n*th housing unit was selected. [Because the skip interval *(n)* was based on the number of housing units in a region, the skip interval varied by region.]

In the second stage, five census blocks within each selected census tract were systematically chosen. From a random start, the block with every *k*th housing unit was selected. This skip interval was based on the number of housing units in a census tract and varied by census tract.

Finally, at the third stage, 16 housing units within each of the blocks selected in the second stage were enumerated in the field by employing a uniform enumeration and systematic sample selection procedure. From a random start, every fourth housing unit among the 16 listed in the block was systematically selected to be interviewed.

The uniform enumeration procedure involved starting at the northwestern corner of the block selected, proceeding in a clockwise direction around the block, and listing housing units on the right side of the street in the direction of travel until 16 housing units were listed (Figures 1 and 2). Note that the housing units

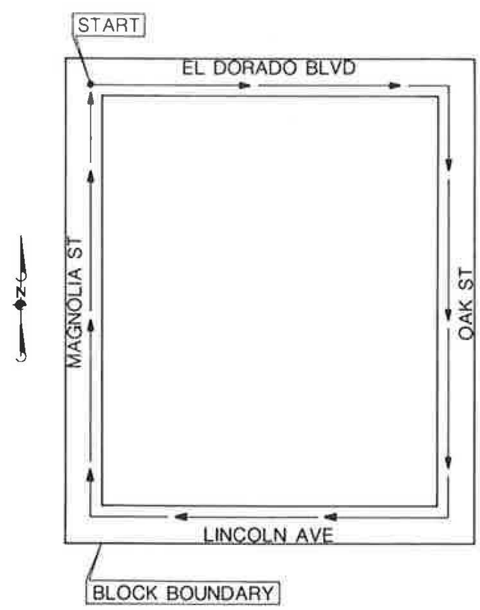


FIGURE 1 Field enumeration procedure showing both sides of the street and indicating the lister's starting place and direction of travel within a selected rectangular census block. The lister starts at the northeastern corner of the block and proceeds clockwise tallying housing units on the right side.

Division of Transportation Planning, California Department of Transportation, 1120 N Street, Sacramento, Calif. 95814.

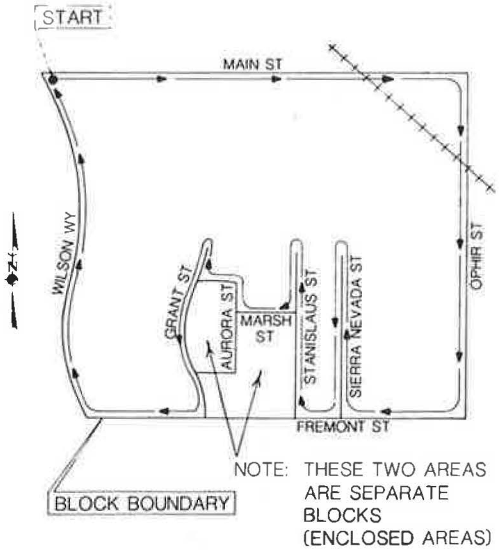


FIGURE 2 Field enumeration procedure showing right side of the street only and indicating the lister's starting place and direction of travel within a nonrectangular census block.

selected do not represent an equal proportion of the units by block; they do, however, meet the requirements of attaining a minimum of 500 samples for each survey region. (Statistical weights were applied to compensate for nonproportional samples when needed for survey data summaries.)

CLUSTER SAMPLING

During the survey period, cluster sampling was employed to minimize travel time for survey interviewers and facilitate calling back. The conventional standard error formula $[s/(n^{1/2})]$ is not appropriate in cluster sampling situations because variances of estimates derived from cluster samples tend to be greater than those derived from simple random samples (or systematic random samples) of the same size. As pointed out by Hubert M. Blalock, "[For cluster sampling] . . . the simple random sample formula will underestimate the true error" (*3*, p.527).

Guidelines for Designing Travel Surveys for Statewide Transportation Planning (*4*, p.5.12) suggests the use of the "design effect" for estimating standard errors of statistics acquired through cluster sample travel surveys. Leslie Kish who initially described the design effect as a means of accounting for the effects of clustering is quoted: "[T]he ratio of the actual variance [of a cluster or other complex sample] to the variance of a simple random sample of the same number of elements" (*4*, p.5.12). However, the design effect factors for the 1977–1978 surveys could not be determined because of lack of comparable data from simple random sampling.

Because multistage cluster sampling was employed in the six regional surveys, estimates of sampling error for those surveys were made using Deming's method of replicated subsamples. This method takes sample clustering into consideration and usually yields a higher (more conservative or safer) total standard error than does the conventional method.

Herbert Arkin and Raymond Colton define "standard error" as follows: "The standard deviation of a sampling distribution of means, or any other statistical measure computed from samples, is termed the standard error of the mean . . . or the standard error of the other statistical measure" (*5*, p.144).

Deming points out (*2*, p.87), "The distinguishing feature of the [replicated subsampling] design is . . . subsamples, drawn and processed completely independent of each other. The chief advantage of replication is ease in the estimation of the standard errors."

STATISTICAL RELIABILITY OF KEY SURVEY ESTIMATES

It should be noted that the particular variables presented in this paper are not intended to be all-inclusive. They are provided simply to illustrate application of the method of replicated subsampling to determine standard errors from cluster sample travel surveys.

Reliability estimates for the cluster sample surveys are presented for three variables by survey region and type of housing unit—persons per household, vehicles per household, and weekday person trips per household. The confidence intervals given in Tables 1–6 represent ranges of estimated sampling error at both the 90 percent and 95 percent confidence levels. (It should be kept in mind that errors occur whether a sample or a complete enumeration is used and that nonsampling errors are not taken into account when presenting statistical reliability estimates. Strict quality control procedures, of course, are required to minimize errors.)

TABLE 1 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, FRESNO REGION.

	SINGLE HOUSING UNITS					
	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	3.12		1.91		9.73	
Standard Error of the Mean[a]	0.088		0.062		0.481	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.145	±0.172	±0.102	±0.122	±0.794	±0.943

TABLE 1 *continued*

MULTIPLE HOUSING UNITS

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
Mean	2.09		0.92		5.30	
Standard Error of the Mean[a]	0.143		0.121		0.784	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.236	±0.280	±0.200	±0.237	±1.294	±1.537

TOTAL HOUSING UNITS

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
Mean	2.83		1.63		8.49	
Standard Error of the Mean[a]	0.126		0.102		0.553	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.208	±0.247	±0.168	±0.200	±0.912	±1.084

Note: Derived by replicated subsamples.

[a]± 1.65 times the standard error of the mean.

[b]± 1.96 times the standard error of the mean.

TABLE 2 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, KERN REGION

SINGLE HOUSING UNITS

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
Mean	3.11		1.85		10.33	
Standard Error of the Mean[a]	0.242		0.045		0.484	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.399	±0.474	±0.074	±0.088	±0.799	±0.949

MULTIPLE HOUSING UNITS

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
Mean	2.46		1.32		8.12	
Standard Error of the Mean[a]	0.119		0.118		0.657	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.196	±0.233	±0.195	±0.231	±1.084	±1.288

TOTAL HOUSING UNITS

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
Mean	2.94		1.72		9.77	
Standard Error of the Mean[a]	0.077		0.053		0.525	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.127	±0.151	±0.087	±0.104	±0.866	±1.029

Note: Derived by replicated subsamples.

[a]± 1.65 times the standard error of the mean.

[b]± 1.96 times the standard error of the mean.

TABLE 3 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, SACRAMENTO REGION

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
SINGLE HOUSING UNITS						
Mean	3.02		1.86		11.40	
Standard Error of the Mean[a]	0.106		0.095		0.739	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.175	±0.208	±0.157	±0.186	±1.219	±1.448
MULTIPLE HOUSING UNITS						
Mean	2.01		1.05		5.82	
Standard Error of the Mean[a]	0.176		0.090		0.834	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.290	±0.345	±0.148	±0.176	±1.376	±1.635
TOTAL HOUSING UNITS						
Mean	2.65		1.56		9.34	
Standard Error of the Mean[a]	0.071		0.063		0.648	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.117	±0.139	±0.104	±0.123	±0.069	±1.270

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

TABLE 4 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, SAN DIEGO REGION

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
SINGLE HOUSING UNITS						
Mean	3.29		2.00		11.88	
Standard Error of the Mean[a]	0.041		0.037		0.267	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.068	±0.080	±0.061	±0.073	±0.441	±0.523
MULTIPLE HOUSING UNITS						
Mean	2.10		1.19		6.87	
Standard Error of the Mean[a]	0.092		0.053		0.274	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.152	±0.180	±0.087	±0.104	±0.452	±0.537

TABLE 4 *continued*

TOTAL HOUSING UNITS	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.84		1.69		9.97	
Standard Error of the Mean[a]	0.049		0.038		0.226	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.081	±0.096	±0.063	±0.074	±0.373	±0.443

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

METHOD OF REPLICATED SUBSAMPLING

Briefly, the method of replicated subsampling (henceforth subsampling) is applied by examining estimates of a particular statistic derived from subsamples designed into the original survey sample. To find the standard error of a statistic derived from cluster samples, the lowest subsample mean value is subtracted from the highest mean value and divided by the number of subsamples compared (Table 7). The resulting number is an unbiased and reliable estimate of the standard error of the sample. To find its 90 percent or 95 percent confidence interval, the standard error is multiplied by the confidence factor 1.65 or 1.96, respectively.

TABLE 5 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, SAN JOAQUIN REGION

SINGLE HOUSING UNITS	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.96		1.88		9.82	
Standard Error of the Mean[a]	0.056		0.076		0.187	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.092	±0.110	±0.125	±0.149	±0.309	±0.367
MULTIPLE HOUSING UNITS	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.25		1.08		6.02	
Standard Error of the Mean[a]	0.275		0.116		1.400	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.454	±0.539	±0.191	±0.227	±2.310	±2.744
TOTAL HOUSING UNITS	Persons/ Household		Vehicles/ Household		Weekday Person Trips/ Household	
Mean	2.75		1.64		8.70	
Standard Error of the Mean[a]	0.092		0.101		0.500	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.152	±0.180	±0.167	±0.198	±0.825	±0.980

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

TABLE 6 RELIABILITY ESTIMATES OF PERSONS PER HOUSEHOLD, VEHICLES PER HOUSEHOLD, AND WEEKDAY PERSON TRIPS PER HOUSEHOLD AT THE 90 AND 95 PERCENT CONFIDENCE LEVELS, STANISLAUS REGION

	Persons/Household		Vehicles/Household		Weekday Person Trips/Household	
SINGLE HOUSING UNITS						
Mean	3.11		1.81		9.40	
Standard Error of the Mean[a]	0.140		0.068		0.406	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.231	±0.274	±0.112	±0.133	±0.670	±0.796
MULTIPLE HOUSING UNITS						
Mean	2.03		1.07		5.79	
Standard Error of the Mean[a]	0.211		0.111		0.981	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.348	±0.414	±0.183	±0.218	±1.619	±1.923
TOTAL HOUSING UNITS						
Mean	2.85		1.63		8.53	
Standard Error of the Mean[a]	0.106		0.071		0.346	
Confidence Interval	90%[a]	95%[b]	90%[a]	95%[b]	90%[a]	95%[b]
	±0.175	±0.208	±0.117	±0.139	±0.571	±0.678

Note: Derived by replicated subsamples.
[a]± 1.65 times the standard error of the mean.
[b]± 1.96 times the standard error of the mean.

TABLE 7 WEEKDAY MEAN PERSON TRIPS PER HOUSEHOLD BY CENSUS TRACT IN SACRAMENTO REGION

Census Tract Subsample	Weekday Person Trips per Household	Census Tract Subsample	Weekday Person Trips per Household
1	7.60	14	9.58
2	2.41	15	5.59
3	8.25	16	10.60
4	14.73	17	18.61
5	6.30	18	8.41
6	8.10	19	12.01
7	8.15	20	9.32
8	9.17	21	11.99
9	6.96	22	16.31
10	9.84	23	11.10
11	9.87	24	10.96
12	7.95	25	9.62
13	10.20		

Note: Scored numbers are lowest and highest mean values.

For these surveys, the subsamples to be considered are the census tracts selected for sampling. So, to estimate the standard error of a survey statistic, the mean value of the statistic was computed for each census tract subsample. Examination of the census tract means yields the range of the statistic.

For each survey region, except the San Diego region, 20 households (five blocks per census tract and four housing units per block) were sampled in each of the 25 census tracts. In the case of San Diego, 20 households were sampled in each of 50 census tracts.

To estimate the standard error of weekday person trips per household, means were obtained for each of the census tracts in a region. For example, census tract means for the Sacramento region were as given in Table 7. The range of means was found to be 18.61 − 2.41 = 16.20. therefore dividing the range by the number of subsamples yields the estimate of the standard error of person trips per household (16.20/25 = 0.648).

Subsample means were computed for each census tract within each of the six surveys. Table 8 gives a comparison of the standard

TABLE 8 COMPARISON OF STANDARD ERRORS DERIVED BY REPLICATED SUBSAMPLING AND BY THE CONVENTIONAL STANDARD ERROR FORMULA $[s/(n^{1/2})]$

		Standard Errors for		
Region	Method	Persons per Household	Vehicles per Household	Person Trips per Household
Fresno	Replicated subsampling	0.126	0.102	0.553
	Conventional	0.076	0.050	0.417
Kern	Replicated subsampling	0.077	0.053	0.525
	Conventional	0.069	0.044	0.400
Sacramento	Replicated subsampling	0.071	0.063	0.648
	Conventional	0.069	0.046	0.525
San Diego	Replicated subsampling	0.049	0.038	0.266
	Conventional	0.047	0.032	0.273
San Joaquin	Replicated subsampling	0.092	0.101	0.500
	Conventional	0.071	0.050	0.375
Stanislaus	Replicated subsampling	0.106	0.071	0.346
	Conventional	0.079	0.048	0.342

errors obtained by subsampling with those derived by the conventional standard error formula $[s/(n^{1/2})]$.

As the data in Table 8 indicate, subsampling almost always provided higher estimates of standard errors than did the conventional method for the variables measured. In only one case (for the variable "Person Trips per Household" in the San Diego region) did the standard error acquired from the conventional formula exceed that derived from subsampling. This was a rare situation in which variances within sample clusters were greater than the variance of cluster means.

SUMMARY AND CONCLUSIONS

Because multistage cluster sampling was employed for six regional home interview travel surveys conducted in California, the conventional standard error formula $[s/(n^{1/2})]$ underestimated actual standard errors in the survey regions of concern. It was possible, however, to estimate standard errors for the regions using Deming's method of replicated subsampling, which takes into account sample clustering.

Application of replicated subsampling yielded higher and more defensible estimates of total sample error than did the conventional standard error formula, which assumes a simple random sampling design. Leslie Kish's method for calculating standard errors for statistics obtained by cluster sampling is another available technique, but it does not have general application because appropriate design effect factors are not always determinable. Replicated subsampling for large data sets can now, of course, be done quite easily and expeditiously with the use of modern high-speed computers. In brief, replicated subsampling provides an appropriate, unbiased, reliable, and generally applicable framework for estimating sampling error for cluster sample travel surveys.

REFERENCES

1. D. L. Ochoa and G. M. Ramsey. *The 1976–1980 Statewide Travel Survey.* Division of Transportation Planning, California Department of Transportation, Sacramento, Dec. 1981.
2. W. E. Deming. *Sampling Designs in Business Research.* John Wiley and Sons Inc., New York, 1960.
3. H. M. Blalock. *Social Statistics,* 2nd ed. McGraw Hill Book Company, New York, 1978.
4. Peat, Marwick, Mitchell and Co. *Guidelines for Designing Travel Surveys for Statewide Transportation Planning.* FHWA, U.S. Department of Transportation, May 1976.
5. H. Arkin and R. R. Colton. *Statistical Methods.* Barnes and Noble Books, New York, 1970.

California's regional travel surveys were financed through the FHWA, U.S. Department of Transportation. The authors are responsible, however, for the facts and accuracy of the information provided herein. The contents reflect the views of the authors and not necessarily those of the California Department of Transportation or the U.S. Department of Transportation. This paper does not constitute a standard, specification, or regulation.

A Comparison of Census Journey-to-Work and Model-Generated Transportation Data in the Puget Sound Region

RAYMOND G. DEARDORF AND JERRY B. SCHNEIDER

Journey-to-work trip data from the 1980 Census and output from the Urban Transportation Planning System are compared for the Puget Sound Region in Washington State. The purpose of this comparison is twofold: to identify where regional transportation models may need adjustment and to determine whether census journey-to-work data are a valid substitute for large-scale origin-destination surveys. Home-based total work trip tables and home-based transit work trip tables from the census and the model are compared using two methods: a trip length frequency distribution comparison and mapping the differences between the two sets of trip tables using the FLOWMAP mapping program. The trip length frequency distribution comparison shows that census work trips averaged slightly longer than model trips. FLOWMAP analysis, which maps the differences between the two sets of trip tables, reveals that, for total trips, census trips exceeded those from the model for trips attracted to the central business district of three major cities in the region. A second significant finding in the FLOWMAP analysis is that the model shows a few more longer transit trips than do the census data. An evaluation of census journey-to-work data is undertaken.

This research is focused on a comparison of the model output from the Urban Transportation Planning System (UTPS) with transportation data developed from the 1980 Census, known as census journey-to-work data or the Urban Transportation Planning package (UTPP). Home-based work trip tables from both of these data sources are compared in terms of total daily trips and transit daily trips for the most heavily urbanized parts of three counties of the central Puget Sound region in Washington State. Differences with respect to the outputs of trip generation, trip distribution, and mode-split model are discussed. Also, the way the questions on the census form were worded and the way the census data were factored uniformly, with no regard for varying trip generation rates, are examined as possible causes for discrepancies between the two trip tables.

Conclusions, in the form of potential causes of these differences, are discussed; recommendations are made regarding the need to recalibrate the Puget Sound Council of Government's (PSCOG) trip generation, distribution, and mode-split models, and recommendations regarding the use of census data are made. In general, the analysis assumed that census data provide the base with which model output should be compared because they are observed data and the model output is considered estimated data. Caveats to the use of census data will be discussed later.

R. G. Deardorf, Parsons Brinckerhoff Quade & Douglas, Inc., 710 Second Avenue, Suite 960, Seattle, Wash. 98104. J. B. Schneider, Department of Civil Engineering, University of Washington, Seattle, Wash. 98195.

PURPOSE OF STUDY

Transportation planners, including transit planners, use several methods to plan future transportation facilities and services. In the past, origin-destination surveys were an excellent source of valid information on which to structure a planning study. Unfortunately, such surveys have become expensive to conduct and consequently are done rarely. Another method involves the use of large-scale transportation models, which frequently rely on origin-destination surveys for calibration. UTPS models generate trips at places of employment and residences; distribute them among appropriate zones; and apply a mode split to determine how many of these trips are by automobile, carpool, transit, and so forth. Models, however, may contain inaccuracies that are hard to detect if an origin-destination survey has not been conducted for quite a while. Most urban regions have not had large-scale origin-destination surveys since the 1960s. Since then, land use changes have generated changes in trip patterns, especially in metropolitan areas that have experienced high growth. Demographics, such as household size and the number of workers in the work force, have also changed, and these changes affect trip-making characteristics. Ground checks of actual traffic or transit volumes can be used to check some aspects of the model runs.

Census journey-to-work data compiled in 1980 are the most recent data source available to transportation planners for comparison with UTPS model output to determine if additional calibration work is needed. It is hoped by many transportation planners that the census journey-to-work data can be used as an inexpensive partial substitute for origin-destination surveys. This would involve using the census data to recalibrate the UTPS models. The goal of this research, however, was not to prove that one or the other is incorrect but to compare them and hypothesize about the causes of differences between them.

STEPS IN ANALYSIS

The zone-to-zone trip interchanges from the census journey-to-work data and the UTPS model are compared. The types of trips to be compared include total home-based work trips and transit home-based work trips. Total home-based work trips are compared because that will help identify any problems with the UTPS trip generation and distribution process. The transit home-based work trip comparison shows if there are problems associated with the mode split model.

The two trip categories from the two data sources are compared at two different levels of aggregation. The first is a 16-zone aggregation of the three most heavily urbanized Puget Sound counties: Snohomish, King, and Pierce. The next level of analysis is a 40-zone aggregation of the same area, which focuses on areas in more detail.

When the trip tables are analyzed at these separate levels, broad, general trends become apparent (e.g., intra-Seattle transit trips are underestimated and suburban transit trips are overestimated) if the larger zones are examined.

The trip tables from both UTPP and UTPS were originally in a 295 x 295 matrix [at the Transportation Analysis Zone (TAZ) level, which is too small for this analysis]. These large-scale trip matrices were combined into the desired 16- and 40-zone trip tables. The UTPP trip tables were multiplied by a factor so that they could be compared with the UTPS model trip table.

The first phase of comparison involved the use of the origin-destination data-mapping program, FLOWMAP. The capabilities of FLOWMAP are such that it can graphically display trip tables; that is, it plots the flow of anything (in this case, trips) between geographic zones using an arrow proportionate in width to the number of trips between zones. The utility of FLOWMAP for this particular exercise is that it can subtract one trip table from another; this ability enables it to work with a table of differences between the two trip tables. The FLOWMAP program can also plot positive and negative differences. In this manner differences in the two data sets can be shown on a map and significant spatial patterns (if any) of differences can be readily spotted *(1)*.

In addition to the FLOWMAP analysis, trip length frequency distribution comparisons with implications for the trip distribution model have been run for both data sets.

The objectives and conclusions focus on interpreting spatial differences (revealed from the use of FLOWMAP) and aspatial differences (revealed by plots of the trip length frequency distributions). Findings should help identify problems that are isolated with the mode-split model for transit, with the trip generation and distribution models, or by using census data as a measure of transportation model accuracy. Again, this exercise is not intended to use one as a basis for judging the other but to point out the differences and, if possible, provide reasons for them.

Conclusions from this analysis concerning the transportation models and their calibrations can be considered directly applicable to the Puget Sound region because these models have been tailored to fit that region only. Conclusions reached concerning the census journey-to-work data should be of interest to UTPP users around the nation.

BACKGROUND INFORMATION

In this section is provided background information on three important components of this analysis: the travel-forecasting process used to produce simulated trips, the 1980 Census journey-to-work data and the UTPP, and the basis for adjusting the census journey-to-work data into a form comparable with trip model output.

Trip Forecasting

The travel-forecasting process currently used by the PSCOG documents the basis for some of these models grounded in previous surveys. Assumptions underlying the trip generation, trip distribution, and mode-split models (the results of which will be tested against the UTPP data) are examined.

The transportation-modeling process, simplified here somewhat, uses population and employment data for particular geographic zones to generate trips from population and employment data and to distribute these trips among zones (according to their proximity to one another and relative sizes in addition to other factors). A mode split is then applied to those zone-to-zone interchanges to arrive at trip tables by mode of transportation.

Trip generation actually consists of two models, a trip production model and a trip attraction model. These models estimate trip productions and attractions from demographic and economic data developed for each analysis zone. These productions and attractions are calculated using trip production and attraction rates for each of several land use categories. Attraction rates relate to trips attracted to destinations such as places of employment and other activity centers. Employment data used by PSCOG were obtained from a 1980 employment inventory. Trip generation rates are assigned to attractions on the basis of land use, which is usually divided into the retail, office employment, manufacturing, and educational categories *(2)*.

Trip productions are generated by applying rates to total population, households, and income categories in a particular analysis zone. As with attractions, productions looked at in this exercise are only home-based work trips.

The trip generation models were calibrated in 1976 using as a basis the 1970–1971 origin-destination survey conducted by the PSCOG (then known as the Puget Sound Governmental Conference). This was a survey of 2,339 households in the region. The trip generation model has been further validated using 1980 Census data for population. It has also been checked with actual ground counts.

After trips are generated they are distributed between zones by the trip distribution model. The model that is used by PSCOG is a gravity-type model programmed for UTPS operation. The gravity model assigns the number of trips between any two zones as directly proportional to the number of productions and attractions in each zone and inversely proportional to the travel time between them. Like the trip generation models, the trip distribution model has been calibrated with data from the 1970–1971 origin-destination survey conducted in the region.

The mode split model uses a multinomial logit framework to determine the mode of travel for each zone-to-zone trip interchange. The model used during the period when this analysis was conducted was originally developed for the Minneapolis–St. Paul region, but the model had been calibrated and validated for the Puget Sound region using passenger screenline counts from 1980 and 1977 on-board bus surveys and vehicle screenline counts. The trip assignment model, used to assign trips along actual paths (roads, bus routes), is not examined in this paper.

For this exercise, total work trips from the UTPP are compared with total work trips from the model to see if there might be a problem with trip generation and distribution models. If a difference that is unlike what was observed in the total trip comparison occurs with the transit trip comparisons, the mode split model is a likely source of the problem *(2)*.

Census Journey-to-Work Data

In the past, census data have played an integral part in transportation modeling. Population data are a critical part of the trip generation step of the transportation-planning process. In recent decades, the Census Bureau has included questions on its form to obtain journey-to-work information. In 1960 the Census Bureau provided journey-to-work information coded to the county or place level. These large zonal areas were not small enough for use by transportation planners. Also, zonal definitions were ususally not compatible with zones used by transportation planners.

For the 1970 Census, the development of the Geographic Base File/Dual Independent Map Encoding (GBF/DIME) capability enabled census data to be coded by block; these data could then be translated into a local area's transportation analysis zone (TAZ) structure through the use of an equivalency table.

The 1980 Census provided an expanded questionnaire for journey-to-work items, which made possible observation of travel times and mode choice and elicited answers to detailed questions about carpooling. On the basis of these data, a table of work trips, split by mode, from TAZ to TAZ was developed *(3)*.

The journey-to-work information is contained in the UTPP. This information is available to metropolitan planning organizations (MPOs), coded specifically for the MPO's regional TAZs. The PSCOG purchased this information from the Census Bureau in early 1984. These data were coded to the PSCOG's TAZ structure by the Census Bureau.

The UTPP, as received by the PSCOG, consisted of six parts:

1. Demographic data at the residence end, at the TAZ level;
2. Demographic data at the residence end, at the large-area level (major city, county, standard metropolitan statistical area);
3. Demographic data at the workplace end, at the TAZ level;
4. Trips from residence to work, at the TAZ level;
5. Demographic data from the worker end, at the block group level; and
6. Trips from residence to work, at the county or city level.

All of these categories are broken down into travel times and modes of travel and are grouped in such demographic categories as income, race, sex, and age.

This research was concerned exclusively with data from Part 4—trips from the residence end to employment—coded by the Census Bureau to the PSCOG TAZs. These trips were split into 12 different mode categories. The long form of the census asked the following questions to obtain journey-to-work data:

- Did this person work at any time last week? (yes or no)
- How many hours did this person work last week?
- At what location did this person work last week? (address)
- Last week, how long did it usually take this person to get from home to work (one way)? (minutes)
- How did this person usually get to work last week? (check mode of travel)
- When going to work last week did this person usually drive alone, share driving, drive others only, or ride as a passenger only?
- How many people, including this person, usually rode to work in the car, truck, or van last week? (2, 3, 4, 5, 6, 7, or more)
- Was this person temporarily absent or on layoff from a job last week? (layoff, vacation/illness, no)

Although the long form of the census was sent to one of every six households, the place of work responses of workers in about only one in twelve or about 8 percent of all households were actually coded due to budget constraints. Trips were coded from the residence census block to the work end census block and then inflated to 100 percent *(3)*.

Adjusting the Census Journey-to-Work Data

Adjustments must be made to census journey-to-work data, which are in trip table form, before any comparison can be made with the results from the PSCOG's transportation models. The census data must be adjusted to compensate for several shortcomings in the 1980 Census long-form questionnaire and the fact that it is not in the trip production and attraction format that is typically used by MPOs such as PSCOG in their trip-modeling process. The major shortcoming of the census form is that the question asks how a person usually got to work in the previous week not how the person traveled to work on an average day.

Most MPO transportation model calibrations are based on transportation surveys that ask questions about work trips on the previous day. This statistical sampling accounts for occasional sickness, occasional change in mode of travel, and people who do not work a full workweek. By asking how an individual usually traveled to work in the previous week, the census cannot account for these occasional changes in travel characteristics during the workweek.

One approach to correcting the problem created by the inconsistent wording underlying the two sources of work trip data is discussed by William Mann of the Metropolitan Washington, D.C., Council of Governments (WASHCOG). WASHCOG staff derived several factors to apply to the UTPP trip table in Part 4 of the census journey-to-work data *(4)*.

The first factor is designed to account for work trips made by people who worked in a standard metropolitan statistical area (SMSA) yet lived outside it. This factor is calculated by dividing the total number of workers residing inside the SMSA by the sum of the number of workers reporting work sites inside the SMSA and the number of workers reporting work sites outside the SMSA:

$$
\begin{aligned}
\text{Factor} &= \text{Total}/(\text{In} + \text{Out}) \\
\text{Total} &= \text{Total number of workers residing inside SMSA} \\
\text{In} &= \text{Number of workers reporting work sites inside SMSA} \\
\text{Out} &= \text{Number of workers reporting work sites outside SMSA}
\end{aligned}
$$

The second factor is designed to account for those people who do not make it to work on the average day. This is to correct the overreporting of work trips in response to the question, "how did you usually get to work last week?" This question does not account for workdays missed during the week due to absenteeism, a reduced workweek, or other related reasons that would have shown up had the question been worded, "how did you get to work yesterday?"

Factor three is designed to convert census trip data to the MPO trip production and attraction format. This factor would be 2 if everyone who traveled from home to work also traveled directly back from work to home. However, that is not the case because there is a tendency to go somewhere else after work instead of directly home. This factor can be calculated by dividing a region's total work trips, from and to home, by the number of trips from home to work.

The fourth factor is designed to account for occasional mode shift during the week. This would apply to situations in which the census respondent would report his usual mode of work trip for last week (e.g., drive alone) yet take the bus or carpool once or twice a week, or vice versa. The way the census question was worded, there is no way the occasional mode change can be detected.

These four factors are multiplied to form one factor to apply to the UTPP trip table so that it becomes more comparable with the

PSCOG's transportation model output. The factors for the Puget Sound region were calculated in the following manner.

The data necessary to calculate the first factor were derived from the UTPP. The data used were derived from a cross-classification by mode of SMSA totals for workers. Table 1 gives the calculations resulting in a factor of 1.080 applied to the transit trip table and 1.081 to the total trip table.

TABLE 1 ADJUSTMENT FACTOR FOR WORKERS INSIDE AND OUTSIDE THE PUGET SOUND REGION (Snohomish, King, and Pierce counties)

Place of Work	Transit	Total
Inside SMSA	73,034	812,441
Outside SMSA	3,054	53,754
Not reported	6,056	72,983
No fixed place	0	5,069
Total	82,144	944,249

Note: Mode of travel factor for transit is 82,144/(73,034 + 3,054) = 1.080 and for the total is 944,249/(812,441 + 53,754) = 1.081.

Subsequent analysis has revealed that this first factor need not have been applied. Mann's application of this factor was to the 1977 Travel-to-Work Supplement to the Bureau of the Census Annual Housing Survey. However, the 1980 UTPP did not need to be factored in such a manner to make it comparable to MPO transportation model output. The implication of this, with regard to the findings of this analysis, is discussed under Summary and Conclusions.

The second factor, to account for non-travel to work on an average day, was determined to be 0.85 on the basis of a 1968 home-interview survey conducted in the Washington, D.C., area *(5)*. This conclusion is based on the fact that 15 percent of the working population does not make a work trip on a given day for the reasons previously discussed. This number was also used by Mann *(4)*.

However, the census data do take into account people who did not work the entire previous week. The factor of 0.85 may be somewhat low because part of the 15 percent of the working population not making a work trip is already accounted for in the census journey-to-work data. The effect of this on the 0.85 factor is difficult to objectively calculate because of insufficient data, especially for the Puget Sound region. However, this figure may be derived using certain assumptions:

Number of weekdays in a year, excluding holidays = 250
Days of vacation per year = 10
Percentage of absentees that has already been accounted for in UTPP = 10/250 = 0.04

A figure of 0.89, which is the 0.85 factor adjusted to accommodate the fact that the UTPP already accounts for weekly vacations, can be derived:

0.04 + 0.85 = 0.89

Assuming that 10 days of vacation per year is average and dividing that figure by 250 gives a factor of 0.89 for Factor 2 by adding 4 percent to 0.85.

The third factor, converting the census trip data into the PSCOG trip production and attraction format, is calculated in the following manner. Total work trips in the region, to and from home, are divided by trips from home to work to arrive at a factor for the Puget Sound region. These trip data come from PSCOG's 1970–1971 origin-destination survey results *(6)*.

As the following calculation shows, the UTPP trip table is multiplied by 1.889 to make it comparable with the PSCOG trip table in production and attraction format.

Trips to work 452,642
Trips home 402,522
Total 855,164
855,164/452,642 = 1.889

The fourth factor is designed to compensate for occasional mode shifts during the week. It is difficult to quantify because no data exist for the Puget Sound region that can be used to calculate this factor. It is likely that shifts between modes throughout the week could offset one another. A behavioral change like this is hard to speculate on without data. For the Puget Sound region, it is left at 1 (i.e., no adjustment is made).

Multiplying these four factors, an overall adjustment factor is obtained for application to the UTPP transit and total trip tables to make them comparable to the PSCOG's transportation model outputs:

For transit trips, 1.080 x 0.89 x 1.889 x 1.0 = 1.81.
For total trips, 1.081 x 0.89 x 1.889 x 1.0 = 1.81.

For purposes of this study, both census total trips and transit trip tables were multiplied by 1.81 to make them comparable to the UTPS model output.

Description of TAZs

Both the model-generated and the census trip tables were originally in a 295 × 295 zonal matrix. This corresponds to the 295 TAZs into which the PSCOG has divided the urban areas of Snohomish, King, and Pierce counties for transportation-modeling purposes. The census data, normally coded to census blocks within census tracts, were coded to these TAZs by the Census Bureau as part of the UTPP using a TAZ/census block equivalency table supplied to them by the PSCOG. It should be noted that external trips outside the three-county urban area were not included in the analysis. This is because this information coded to the zone level was not provided as part of the UTPP.

Trip Length Frequency Distribution

The first step of the analysis, using the 295 x 295 TAZ trip table matrices, is to compare the two data sets using a trip length frequency distribution. The trip length frequencies of both model and census data sets (for their total or transit trips) were plotted on the same graph. Zone-to-zone travel times are an additional input to this program.

Travel times between the 295 zones were obtained for the regional street and highway network and are used for the total trip length frequency distribution comparison. The regional transit network, with much slower intrazonal travel times, is used for the transit trip length frequency distribution comparison.

The first comparison was of trip lengths for the total trip tables plotted using over-highway travel times. In analyzing this plot, it was apparent that the model produced slightly shorter trips than were obtained from the census.

The second comparison was of transit trips, which are plotted using interzonal transit travel times. Results were similar to those observed with the total trip length frequency distribution comparisons. The census travel times are slightly longer than those of the model for both the total and the transit trip tables.

The second step involves the comparison of the two data sets on a much larger scale. The 295 zones are compressed into 16 very large zones. This 16-zone aggregation is shown in Figure 1. The reason for the aggregation into bigger zones is so major differences between the data sets can be shown and analyzed as being a certain type of work trip.

Areas of large differences in the trip tables have been examined in greater detail by splitting the 16-zone trip tables into 40-zone trip tables, as shown in Figure 2. This enables further analysis of areas of significant differences and a more specific pinpointing of possible reasons for any discrepancies. Another reason for a smaller level of analysis is that some interzonal differences may only be apparent at the 40-zone level; in the 16-zone analysis they may disappear because they were aggregated into internal (intrazonal) trips.

FIGURE 1 16-zone division of Seattle-Everett-Tacoma urban region.

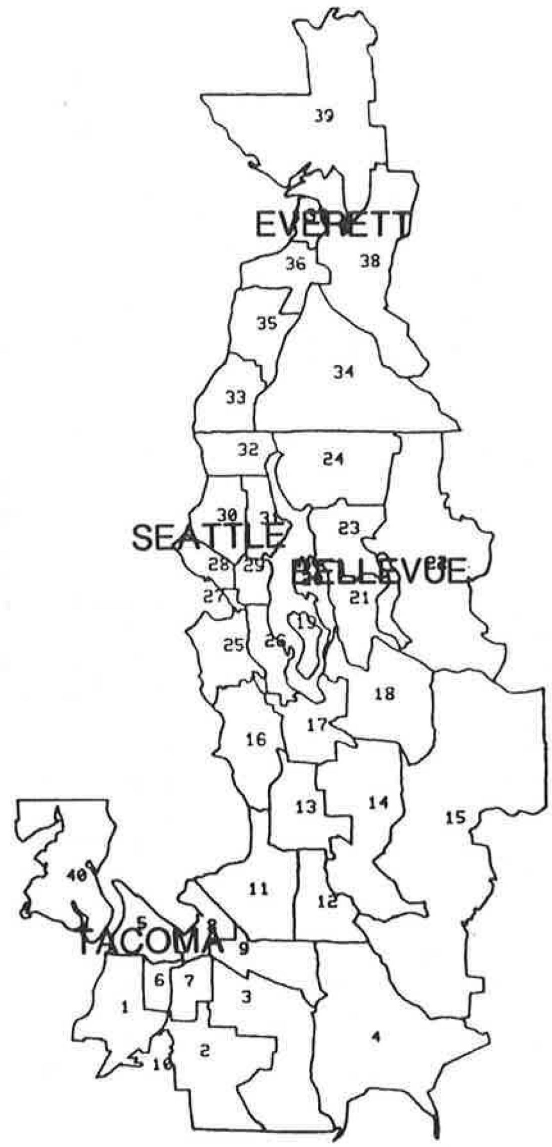


FIGURE 2 40-zone division of Seattle-Everett-Tacoma urban region.

The census trip tables were subtracted from the trip tables generated from the model. The reason it was done in this order and not vice versa was that the model produced a total of 1,535,767 trips whereas the census had only 1,482,715. Therefore it was known that census trips, even after having been factored upward, are still fewer than those produced by the model. Part of this discrepancy may be explained by home-based college trips being included in the model and not in the census.

The end result of the matrix subtraction is four matrices of differences—transit trip differences for 16- and 40-zone trip tables and total trip differences for 16- and 40-zone trip tables. Most of these flows are positive, although a significant number of negative flows exist (where the volume produced by the census exceeded that of the model).

The differences in the flow of trips between zones is displayed using the origin-destination mapping program FLOWMAP.

Description of FLOWMAP

FLOWMAP was developed in 1979 by Bob Evatt, Jr., under the guidance of Jerry B. Schneider of the Urban Transportation Program at the University of Washington and has been extensively revised by Harvey Greenberg since its development. The FLOWMAP program can map origin-destination data interactively on a Tektronix 4014 graphics terminal. Several types of flow maps are possible. This exercise primarily uses one type of display: interzonal flows displayed as variable width arrows the widths of which are proportional to the volume of flow. All zones are considered origins or destinations. Another type of display used in this report is for intrazonal flows; this display takes the form of circles proportional in diameter to the intrazonal flow.

Needed inputs to FLOWMAP are a geographic feature file and a data file. The geographic feature file is a set of gridded zonal coordinates. In this case, universal transverse mercator (UTM) coordinates are used for both the 16-zone and the 40-zone division of the Puget Sound region. Also, centroid points loaded in approximately the geographic center of each zone are included as part of the geographic feature file. These centroid points mark the origins and destinations of the arrows *(1)*.

The data files are the actual matrices of zone-to-zone trip differences created by subtracting the census trip tables from the corresponding model trip tables and modified for use by FLOWMAP. In this case, the four matrices (total and transit trip differences for 16 and 40 zones) contain both negative and positive numbers. The positive numbers occur when the model-generated trips exceed those of the census, and negative numbers occur when the census exceeds the model. On the maps generated by FLOWMAP, positive flows are shaded and negative flows are not shaded.

After the geographic and data files were set up, the analysis proceeded as follows. First, a histogram of each difference table was produced. This showed a distribution of the flow volumes. Following that, the 16-zone trip tables (actually, trip difference tables) for transit and total trips were mapped. One of the features of the FLOWMAP program is that the number of arrows shown on a particular map can be screened for minimum and maximum absolute values of flows to be shown. This way, only the few maximum difference flows can be shown on a particular map, so that the significant pattern of the largest flows can be seen. Figure 3 shows the distribution of difference flows for total trips at the 16 x 16 zone level.

Several maps were produced for each trip difference table. These ranged from the maximum value of the flow difference down to a point where lowering the minimum flow would have produced too many arrows, rendering the map overly complex and incomprehensible.

The 40-zone analysis using FLOWMAP was conducted in the same way, although the difference flows are somewhat smaller because there are more zones and consequently fewer trips between them. Previous intrazonal trip table differences that occurred in the 16-zone analysis started to show up as intrazonal differences in the 40-zone analysis.

Figure 3 shows that the vast majority of difference flows have an absolute value less than 1,000. The FLOWMAP analysis looked at the major flows toward the ends of the distribution spectrum. The distributions of the other three trip table differences were similar.

Examples of the FLOWMAP analysis are shown in Figures 4 and 5. Figure 4 shows the difference flows between 1,000 and 1,500 in value for the transit trip tables aggregated into a 16 x 16 matrix. This figure shows two shaded flows, which indicate that the model transit trips exceed the census transit trips for those particular zone-to-zone interchanges. The lower arrow indicates model transit trips that exceed census transit trips produced in the suburban area known as Federal Way and attracted to the central business district (CBD) of Seattle. The other arrow indicates a situation in which model transit trips exceed census transit trips produced in the zone representing South Seattle and attracted to North Seattle.

Figure 5 shows the transit trip differences in the 40 x 40 zone comparison. This figure shows model transit trips exceeding census transit trips (shaded flows) attracted to Northeast Seattle

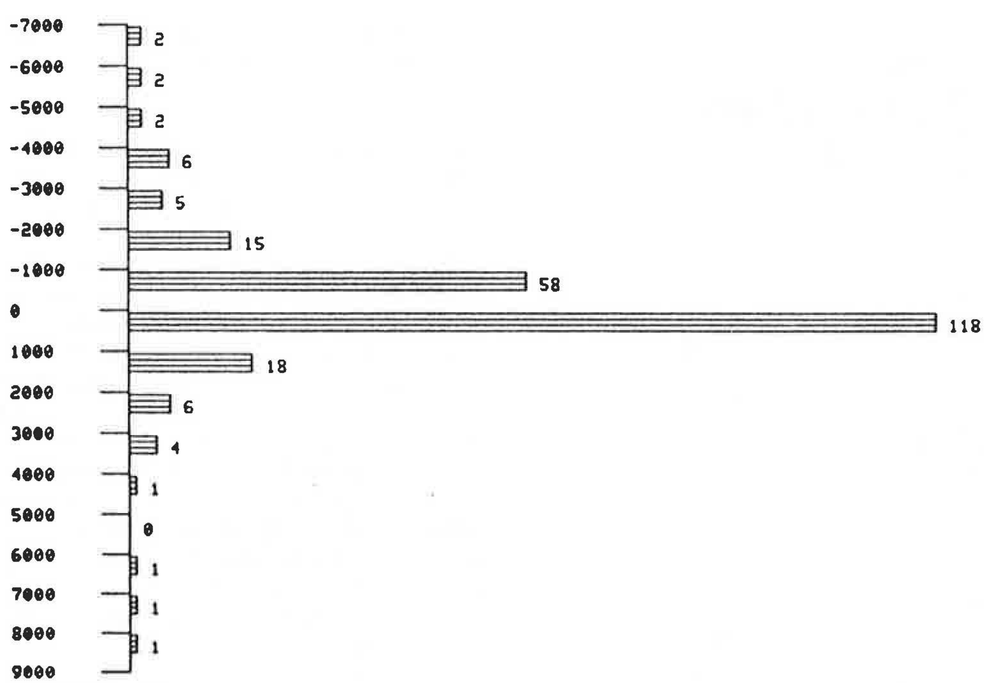


FIGURE 3 Distribution of 16-zone total trip differences.

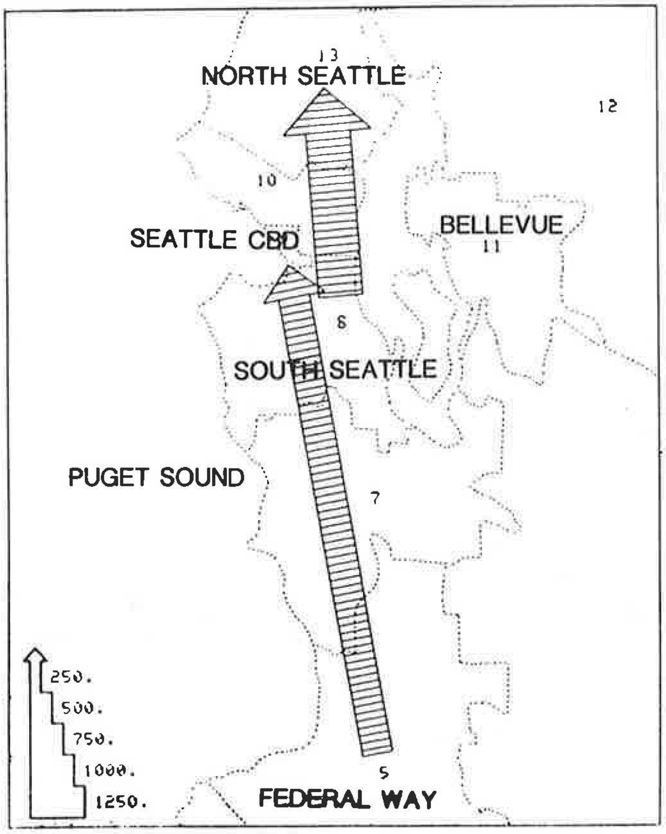


FIGURE 4 Example of 16 x 16 zone matrix transit trip difference flows.

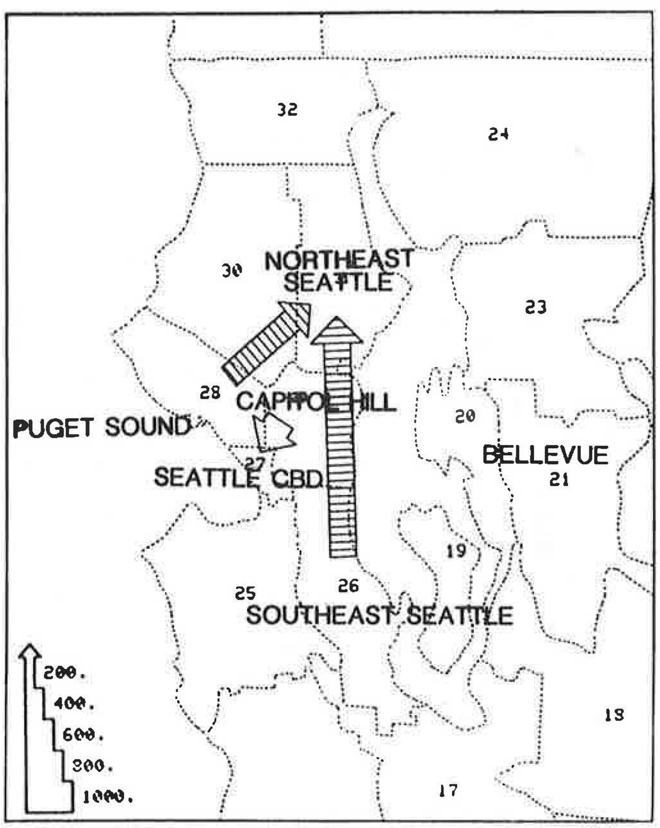


FIGURE 5 Example of 40 x 40 zone matrix transit trip difference flows.

from nearby areas. Also, it shows census transit trips exceeding model transit trips attracted to the Seattle CBD from the adjacent Capitol Hill residential area.

Comparison of 16-Zone and 40-Zone FLOWMAP Analysis

The 16-zone and 40-zone analyses using FLOWMAP to display differences between the model-generated and the census-generated trip data yielded similar results in both the total trip comparison and the transit trip comparison.

Similarities observed in the total trip comparison for both sets of zones include model trips exceeding census trips bound for the zone representing the North Seattle area from several other zones. In the 16-zone analysis, a large number of model transit trips exceeding census transit trips within this zone were noted. In the 40-zone analysis, it becomes clear that this zone is better defined as Northeast Seattle, which includes the University of Washington.

Similarities with regard to census total trips exceeding model total trips in both the 16-zone and the 40-zone analyses can be summarized as follows. There appears to be a pattern for this difference for trips attracted to the CBDs of three major cities from close-by residential zones.

For transit trips, similarities in the 16-zone and 40-zone analyses that show the model exceeding the census include transit trips attracted to North Seattle (as was noted in the total trip comparison) and transit trips attracted to the Seattle CBD from zones medium to far away in distance from the Seattle CBD (from North Seattle, the suburban eastside communities, and southern King County). This latter observation is exactly the opposite of that made in the total trip analysis, where the census generally exceeded the model for trips attracted to the Seattle CBD.

Similarities in the 16-zone and 40-zone analyses with regard to census transit trips exceeding model transit trips show that this occurs for the transit trips attracted to the Seattle CBD from the neighborhoods immediately to the north and east. This corresponds to what was observed in the total trip analysis. Intrazonal differences that did not appear in the 16-zone analysis became interzonal differences.

EVALUATION OF FINDINGS WITH RESPECT TO MODELS

The total trips comparison is examined with regard to the trip generation and the trip distribution models. The trip length frequency distribution comparison and implications for the trip distribution model are also discussed.

The transit trip comparison is used to make judgments about the mode split model. If differences in the volumes of transit trips correspond to differences observed in the total trip comparison, that reflects the trip generation and trip distribution models and not the mode split model. When differences in the transit trip comparison do not match those in the total trip comparison, the mode split process is isolated as the source of the discrepancy.

Total Trip Comparison—Trip Generation and Distribution

Some areas of spatial trip differences between the two data sets, documented in the previous section, can be explained by acknowledging certain basic differences in the two data sets. For example, the home-based work trip table generated by the model also contained college trips; the census trip table did not. The indicates that it may be assumed that zones with large colleges and universities may have more model trips attracted to them than census trips. The FLOWMAP analysis confirmed this. The University of Washington, the largest university in the region, is the main attraction of the large shaded flows into the zone in which it is located. Other zones with significant model attractions that can be explained by the omission of college trips in the census data include the Capitol Hill and Queen Anne zones, which have universities of significant size, although not as large as the University of Washington.

After the difference arrows caused by basic differences in the two data bases are discarded, one predominant type of discrepancy is left. That is where the census trips exceed those from the model for trips attracted to the CBDs of Everett, Seattle, and Tacoma.

These discrepancies may be explained by reviewing the trip generation attraction rates. The CBDs of Everett, Seattle, and Tacoma consist mainly of large numbers of places of financial, insurance, real estate, other services and wholesale, transportation, and communication and utilities employment. The distribution of retail employment is relatively more evenly spread throughout the region (suburban shopping centers, etc.). The significance of this is that the types of employment that are concentrated in these CBDs have a lower trip attraction rate than does retail employment; this gives rise to a lower-than-average trip attraction rate for the entire region. However, the methodology used in factoring the census data into a form comparable to the transportation model output applied a uniform factor for the whole region, with no regard to employment types and the various trip attraction rates associated with them.

This leaves the possibility that the average rate applied across the board to the census was higher than that associated with certain types of employment concentrated in specific geographic areas, namely the CBDs of Everett, Seattle, and Tacoma. This would explain the excess of census total trips attracted to these areas. This also leaves open the possibility, in analyzing this objectively, that it is the trip generation attraction rates of the model that may be too low, or a combination of the two.

There is also a possibility that this problem can be traced to the trip distribution model. The trip distribution model is a gravity-type model used to calculate the number of trips between Zones i and j:

$$T_{(ij)} = (P_i \times A_j \times F_{ij} \times K_{ij}) / \sum_{j=1}^{n} (A_j \times F_{ij} \times K_{ij})$$

P_i and A_j are the production and attraction inputs from the trip generation process described earlier. If they are not correct, the trip distribution process will be affected. If they are correct, however, that leaves the time distance friction factor (F_{ij}) and the socioeconomic adjustment factor (K_{ij}) as sources of error.

The trip length frequency distribution comparisons described earlier help with the analysis of the F_{ij} factor. The distribution comparisons showed that the census trips averaged slightly longer in length than those trips estimated by the model. The next step is to explore why this difference is occurring. It appears that the observed data (census) are showing longer travel times to work than the estimated travel times (model). Travel times from the census are longer than the model estimates. It can be concluded that the journey-to-work travel time increased in this region between the 1960s when the origin-destination surveys on which the travel forecasting model was based were conducted and 1980 when the census journey-to-work data were gathered.

There are many possible reasons or combinations of reasons for the increase in travel time to work in this region during the past 10 to 20 years. The main reason is that the urban area has expanded: vacant land has been urbanized along the freeway network that was just opening in the 1960s. Along with that came the growth of not only suburban areas, but also rural areas, with many people seeing a slower-paced way of life or cheaper land while still working in the urbanized area. The increase in transit trip length can be attributed to the longer bus service routes. Countywide bus service was established in King County in 1974 and in Pierce and Snohomish counties in the late 1970s. The use of commuter park-and-ride lots in suburbs and rural areas has also contributed to the increase in transit trip length.

The adjustment to be made to the trip distribution model centers on the F_{ij} factor. The F_{ij} factor, as described earlier, is a travel time friction factor. There is a different F factor for each minute of travel time. For a particular travel time, the friction factor

$$F_{ij} = 1/t_{ij}^n$$

where T_{ij} is the travel time between Zones i and j and is an exponent that can vary among travel time increments between zones *(2)*. Because the propensity to take longer trips is shown by the trip length frequency comparison, the n factor must be adjusted on both the short end and the long end of the distribution curve.

Transit Trip Analysis—Mode Split

The observed differences in transit trips between the model and the census can be separated into two categories. The first category consists of those differences that are similar to those observed between the total model and the total census trip tables. This would appear to indicate that the mode split model was not at fault; it was just reflecting those differences caused by basic differences in the two data bases or the trip generation or distribution process, or both.

The second category is those differences between the model-generated and the census-generated transit trip tables that are different from those observed for the total trip comparison. It is these differences that show where the mode split model is overestimating (there were no cases of underestimating).

The FLOWMAP analysis of the transit trip tables showed an excess of model transit trips attracted to the zones containing the major colleges in Seattle; this too was reflected in the total trip comparisons. Also, the transit trip comparison shows an excess of census trips attracted to the Seattle CBD from the close-in neighborhoods of Queen Anne and Capitol Hill; this, again, was reflected in the total trip comparisons.

However, the excess of model transit trips attracted to the Seattle CBD from suburban areas is not reflected and is in some cases contradicted by the results of the total trip analysis. The conclusion that can be drawn from this analysis is that the mode split model overestimates transit trips bound for the Seattle CBD from suburban areas.

SUMMARY AND CONCLUSIONS

Some findings that resulted from this comparison were not related to the questions of whether the transportation models needed calibrating and whether the census data were inaccurate. That college trips were included in one data set and not the other for both sets of trip tables complicated the analysis somewhat. However, when these differences had been recognized and accounted for, a clearer picture emerged of the real differences between the two trip table sets and what explained these differences.

The real differences discovered between the two sets of trip tables that relate to trip generation, distribution, and mode split are as follows. The FLOWMAP analysis of total trips revealed an excess of census trips attracted to the CBD areas of Everett, Seattle, and Tacoma. The trip generation process was examined with respect to this pattern of total trip difference and a plausible explanation, that of varying trip generation rates, was discussed.

It is possible, however, that the reason for the excess census trips attracted to the CBDs of Everett, Seattle, and Tacoma rested not with the production and attraction inputs to the trip distribution model but with the friction of time-distance factor. The trip length frequency distribution comparison showed that the distribution of travel time was slightly different, which indicates that census trips averaged slightly longer in travel time than model trips. This would appear to indicate that the average trip length is longer than the models are estimating, which means that the friction of time-distance factor could be adjusted so that time-distance is not as much of as an inhibitor as it is in the present trip distribution model.

The mode split model was found to overestimate transit trips from suburban residential areas to Seattle's CBD. The mode split model, using travel time and monetary cost variables to determine mode choice, clearly is making the transit mode more attractive than do census data for these particular types of transit trips. On the basis of this comparison, some adjustment to the travel time or monetary cost variables, or both, may be in order.

In this analysis, most of the findings and conclusions are based on the assumption that the factors used to adjust the census journey-to-work trip tables to a form comparable to that of the model trip tables are accurate. These factors, discussed earlier, can be adjusted either up or down depending on the methodology used in calculating them and the subjective judgment of the individual determining these factors. The methodology set forth by WASH-COG, modified somewhat because of the lack of availability of certain types of data for the Puget Sound region, was used.

The later realization, after the analysis had been performed, that the first of the four factors need not have been applied to the 1980 Census package serves to emphasize the statements made in the previous paragraphs. Leaving out the factor of 1.08 would have resulted in an overall adjustment factor of 1.68 instead of 1.81 for both the total and the transit trip categories.

If the factor used to adjust total census trips had been smaller, the overestimation of census trips attracted to the CBDs of Everett, Seattle, and Tacoma would not have occurred or been as large. In this example, other areas would show an overestimation of model trips where none existed in this analysis. If the factor used to adjust transit trips had been smaller, the model's apparent overestimation of long-distance transit trips into the Seattle CBD would have been even greater. If, on the other hand, the factor used to adjust transit census trips had been larger, the excess of model transit trips from some suburban areas to Seattle's CBD would have been less significant or disappeared completely, and instances in which census transit trips would have exceeded model transit trips would have occurred. It is the researchers' position, however, that the factor of 1.81 provided reasonable results. Conclusions from this analysis also reflect what travel forecasters at the PSCOG suspected about their trip models. Perhaps with some additional information, particularly about occasional mode shifts during the week for the Puget Sound region, these factors could have been tuned more finely.

A significant portion of this factoring process, and therefore a significant portion of the room for error, could be eliminated by making the wording on the long form of the census questionnaire match that used by transportation surveys. This would involve asking where people worked and how they traveled to work on the previous day instead of how they usually traveled to work the previous week.

ARE CENSUS DATA GOOD VALIDATORS OF TRANSPORTATION MODELS?

The answer to this question is yes, with certain important qualifiers. The first and foremost would be to change the wording on the census questionnaire to match that typically used in transportation surveys, as noted previously. This would reduce the room for error in factoring the census trip table to the transportation models' production and attraction format.

The second change would be to separate college trips from work trips in the transportation model's generation and distribution process. Including college trips with work trips is not a requirement of the UTPS modeling process.

Supplemental surveys concerned with trip generation production and attraction rates would make up for the shortcoming in this particular area, in which large concentrations of land use areas that have different-from-average trip generation rates are located.

In summary, with these qualifiers, the data from the census journey-to-work questions provide a great opportunity for transportation planners at all levels to obtain a good picture of the actual condition of the transportation system. For those engaged in transportation modeling at the MPO level, the chance to compare these data with transportation model output is a much less costly alternative than a full-scale origin-destination survey.

The 1990 Census will provide an opportunity to change the wording of the census form to conform to those questions typically asked in transportation surveys. This would eliminate a good deal (but not quite all) of the "apple and orange" comparisons that cause this kind of analysis to be subject to skeptical scrutiny. With that kind of "fine tuning" in the production of the UTPP, it appears that census journey-to-work data can fill an important role in supplementing and supplanting large-scale regional origin-destination surveys.

REFERENCES

1. B. Evatt, Jr., J. Schneider, and H. Greenberg. "Mapping Origin-destination Patterns in Space and Time by Using Interactive Graphics." In *Transportation Research Record 866*, TRB, National Research Council, Washington, D.C., 1982, pp. 25–35.
2. *Travel Forecasting Procedures to Support the Puget Sound Transportation Alternatives Analysis—North Corridor.* Puget Sound Council of Governments, Seattle, Wash., Jan. 1983.
3. *Transportation Planner's Guide to Using the 1980 Census.* U.S. Department of Transportation, Jan. 1983.
4. W. Mann. Converting Census Journey-to-Work Data to MPO Trip Data. *Journal of the Institute of Transportation Engineers,* Vol. 54, No. 2, Feb. 1984, pp. 48–51.
5. G. V. Wickstrom. "Comparison of Census Journey-to-Work Data with Metropolitan Planning Organization Data." Presented at 60th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1981.
6. *Travel Demand Analysis—Trip Generation Update.* Puget Sound Council of Governments, Seattle, Wash., 1976.

New Algorithm for Grouping Observations from a Large Transportation Data Base

RUDI HAMERSLAG AND WIM H. SCHELTES

In this paper is presented a new cluster-segmentation algorithm. Its distance measure, derived by using Fisher's likelihood theory, depends on the probability density function (frequency function) of the observations. The resulting measure of similarity or dissimilarity is consistent with the likelihood theory. It shows attractive features: *(a)* curtailment of cluster-segmentation techniques; each probability density function has its own optimal measure of similarity or dissimilarity; *(b)* detection of dependencies between variables; and *(c)* all the advantages of hierarchical divisive techniques, which makes it suitable for analysis of large transportation surveys. The use of the new algorithm is illustrated by using a large data base, the Netherlands National Travel Survey. The goal of this research is to analyze mobility (expressed in daily mileage) by constructing homogeneous population groups. This example clearly demonstrates that the methodology can satisfactorily deal with numerous observations.

Policy making, decision making, designing, research, and so forth require knowledge. Experience is normally one of the major sources of knowledge. Information from surveyed data are also often used. Nowadays a wide spectrum of information about trips and the persons making them is generally available, mostly in the form of data bases. Increasing computer usage makes it possible for an increasing number of people to incorporate these data in their research.

In general, people are only able to focus on a limited amount of information. Therefore it is recommended that the data be simplifed with respect to the specified problem. Often this is done by selection and aggregation of data into groups so as to obtain a manageable number of observations.

The selection of data is linked to the object of study (also named phenomenon or entity) that is to be analyzed. The object of study, in turn, depends on the research goal. Some examples are car ownership, (daily) mileage traveled by chosen mode of transport, and number of trips (per person). Aggregation means that separate observations are put into groups depending on their characteristics ("attributes") and categories.

Aggregating observations that form the object of study into separate groups causes loss of information *(1, 2)*. Different data grouping results in different losses. Unskillful aggregation may therefore lead to erroneous clarification of observations, which induces imprecise management decisions or ineffective infrastructural design.

Characteristics, or variables, are often chosen on the basis of personal experience. Sometimes the scientific background of the researcher plays a role: economists tend to favor a person's income, whereas sociologists show preference for educational level. Each choice will lead to different data groupings, so a "multiple trueness" exists, which results in a reduced, instead of enlarged, insight into the analyzed phenomenon. Policy makers and decision makers are not likely to use such information. Therefore it is advisable to give more attention to the grouping of observations. This can be done, for example, by means of clustering and segmentation. However, several techniques exist and each one will generally lead to a, more or less, different group composition. Thus the problem of "multiple trueness" still remains, as was demonstrated at the Transportation Planning and Research Colloquium in The Netherlands *(3)*. Several reverse-clustering methods also have small disadvantages; for example,

- Limitations on the number of characteristics, which makes an a priori selection necessary;
- Limitations on the size of the data base *(4)*; to solve this, observations are intuitively aggregated; and
- Dependencies between characteristics cannot be detected.

In this paper a new method is presented that overcomes the previously mentioned objections to the traditional methods. It was derived using Fisher's likelihood theory. Proof will be given that the dissimilarity measure is determined by its underlying probability density function. This leads to a curtailment of cluster-segmentation techniques. Detection of dependencies between variables is also now possible.

TRADITIONAL CLUSTER-SEGMENTATION METHODS

The variables used to characterize data groups can be selected by means of cluster-segmentation techniques. Several algorithms exist. Only the agglomerative and divisive hierarchical methods are discussed in this paper, because they are the most commonly used. Further information about cluster-segmentation algorithms is available elsewhere *(5–8)*.

The agglomerative hierarchical techniques use the bottom-up approach ("clustering"). They represent an attempt to minimize the information loss caused by clustering observations. They are popular and also the oldest known hierarchical techniques (e.g., nearest neighbor and farthest neighbor algorithm). Clustering is done by means of a distance matrix. Thus a large number of observations (say, more than 1,000) results in an enormous matrix, which is practically impossible for most computers to solve *(4)*.

In the case of a large number of observations, divisive hierarchical techniques are preferable. They use the top-down approach ("segmentation"); the aim is to maximize the information gain that results from splitting the data base into two subordinant data bases.

R. Hamerslag, Department of Civil Engineering, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands. W. H. Scheltes, Netherlands Institute of Transport, Polakweg 13, 2288 GG Rijswijk, The Netherlands.

The difference between groups of observations is measured by similarities and dissimilarities. One of the most favored methods is the Euclidian distance, a special case of the Minkowski metric. The Minkowski distance *(D)* between two observations ($P_{k,1}$ and $P_{k,2}$) is

$$D(1,2) = \left[\sum_{k=1}^{h} (|P_{k,1} - P_{k,2}|^{r}) \right]^{1/r}$$

where $D(1,2)$ is the dissimilarity value between first and second observation and $P_{k,j}$ is the value of the *j*th observation in the *k*th dimension. Notice that for $r = 2$ this dissimilarity measure is equivalent to the Euclidian distance in an *h*-dimensional space. For $r = 1$ the measure transforms into the so-called city-block (Manhattan) metric.

Dissimilarities can also be qualified by complicated statistical measures, as is done, for example, in the well-known automatic interaction detection (AID) analysis. Its distance measure between groups can be written as follows:

$$D(1,2) = (N1*MEAN1^2 + N2*MEAN2^2 - N12*MEAN12^2)/S$$

where

$D(1,2)$ = dissimilarity value,
$MEANi$ = observed average value in the *i*th group,
$MEAN12$ = observed average value in the combined group,
Ni = size of the *i*th group,
$N12$ = size of the combined group ($N1 + N2$), and
S = standard deviation in the combined group.

NEW METHODOLOGY

This section deals with a recently developed grouping technique. The new dissimilarity measures, the algorithm, and the interaction of data variables are discussed.

The most interesting aspects of the method are

- It is derived by using Fisher's likelihood theory,
- Its distance measure depends on the probability density function of the observations,
- There are no restrictions with respect to the size of the data base and the number of characteristics to be analyzed, and
- Detection of dependencies between characteristics is possible by using the "likelihood ratio test."

Development of New Dissimilarity Measures

The new dissimilarity measures depend on the probability density function ("frequency function") of the observations. They are developed by using the likelihood estimation theory.

Consider a large set of observations of a certain phenomenon, for example, the daily number of trips (phenomenon) of several interviewees (observations). Suppose the probability density function *(f)* of the data base is known. When all observations ($X1$, $X2, \ldots, Xn$) are stochastically independent, then the likelihood *(L)* follows from

$$L = f(X1) * f(X2) * f(X3) * \ldots * f(Xn)$$

and the logarithm of the likelihood *(LN L)* is

$$LN\ L = LN[f(X1)] + LN[f(X2)] + LN[f(X3)] + \ldots + LN\ [f(Xn)]$$
$$= \sum_{j=1}^{n} LN[f(Xj)]$$

The dissimilarity measure *(D)* between two groups of observations (Group 1 and Group 2) is defined as the difference in log-likelihood before and after clustering them:

$$D(1,2) = LN\ L1 + LN\ L2 - LN\ L12$$

where

$D(1,2)$ = dissimilarity value between the groups (the difference in log-likelihood before and after grouping),
Li = likelihood value of the *i*th group, and
$L12$ = likelihood value of the combined group.

New dissimilarity measures can be calculated for every data base; they optimize the gain of information that results from segmentation. Each probability density function leads to its own characteristic dissimilarity. In the Appendix the derivations are given for

- Binominal (Bernoulli) distribution, a discrete function with true/false or yes/no values (e.g., car ownership, driver's licence);
- Normal distribution, a continuous function; and
- Poisson distribution, a discrete function with nonnegative integers (e.g., daily number of trips per person).

Application of the log-likelihood difference as a new dissimilarity measure will lead to the formulas given hereafter, which are valid for multidimensional *(k)* space. The following abbreviations will be used:

$D(1,2)$ = dissimilarity value between Group 1 and Group 2,
$DIFF_{k,i}$ = auxiliary variable = $1 - MEAN_{k,i}$,
$MEAN_{k,i}$ = observed average value in the *i*th group,
$MEAN_{k,12}$ = observed overall average,
Ni = number of observations in the *i*th group,
$N12$ = total number of observations ($N1 + N2$),
$RTOT_{k,i}$ = auxiliary variable = $Ni - TOT_{k,i}$,
S_k = standard deviation, and
$TOT_{k,i}$ = total observed value in the *i*th group.

For a binominal function (with $MEAN > 0$):

$$D(1,2) = \sum_k [TOT_{k,1}*LN(MEAN_{k,1}) + RTOT_{k,1}*LN(DIFF_{k,1})$$
$$+ TOT_{k,2}*LN(MEAN_{k,2}) + RTOT_{k,2}*LN(DIFF_{k,2})$$
$$-TOT_{k,12}*LN(MEAN_{k,12})$$
$$- RTOT_{k,12}*LN(DIFF_{k,12})]$$

A normal distribution leads to the following dissimilarity measure:

$$D(1,2) = \sum_k [N1^*(MEAN_{k,1}^2) + N2^*(MEAN_{k,2}^2) - N12^*(MEAN_{k,12}^2)]/2S_k^2$$

Notice that this formula largely corresponds with the measure used by Ward *(9)* and in AID analysis *(5)*.

For a Poisson distribution (with $MEAN > 0$):

$$D(1,2) = \sum_k [TOT_{k,1}^*LN(MEAN_{k,1}) + TOT_{k,2}^*LN(MEAN_{k,2}) - (TOT_{k,1} + TOT_{k,2})^*LN(MEAN_{k,12})]$$

For the mathematical derivations of the previous formulas, see the Appendix. General information can be found elsewhere *(10, 11)*.

Grouping Algorithm

In this subsection the new dissimilarity measures in a divisive algorithm (segmentation or reverse clustering) are illustrated. With similar ease an agglomerative algorithm (clustering) could be used. Both methods have advantages and disadvantages *(12, 13)*.

In practice the new dissimilarity measures are used as follows. The data base that is to be analyzed contains observations about several variables. For each class ("category") of a variable, the size (number of interviewees) and average observed value (object of study) are noted. The distance formula is derived from the probability density function. This is used in an internal clustering process: all classes of a variable are grouped, and the dissimilarity is calculated for each variable. The variable with the largest value is, under normal circumstances, the most discriminating one; therefore the data base is split up into its classes. This results in several subordinant data bases. Each of these will be analyzed using a similar process. The final outcome is a hierarchical list of discriminating variables.

Dependency of Group Characteristics

There is a possibility that group characteristics (variables in the data base) are mutually dependent. It is essential to know of these dependencies, especially when the use of proxies is considered or the results need to be interpreted, or both.

Notice that the previously presented dissimilarity measures represent the difference in log-likelihood before and after combining the observations. Those results can be applied directly in the formula for the likelihood ratio test statistic.

For two stochastically independent variables (A and B) the following relationship is valid because there is no "overlap":

$$PROB(A \text{ and } B)/PROB(A) * PROB(B) = 1$$

or, using logarithms:

$$LN\ PROB(A \text{ and } B) - LN\ PROB(A) - LN\ PROB(B) = 0$$

This relationship is hidden in the so-called likelihood ratio test, also known as the G^2-statistic, an easy-to-use method for analyzing dependency or independency between variables *(14, 15)*. The formula for the test statistic is

$$G^2 = -2\ LN[L(A \text{ and } B)/L(A) * L(B)] = -2\ [LN\ L(A \text{ and } B) - LN\ L(A) - LN\ L(B)]$$

where G^2 is the test statistic, which has a χ^2 distribution and $L(i)$ is the likelihood value of the ith group.

Dependencies between variables are often observed. For example, personal income is, under normal circumstances, strongly related to educational level and age. With each of these variables, the possibility of creating nearly equally homogeneous groups exists. Experience, theoretical knowledge, and insight with respect to the phenomenon under analysis can be usable expedients in making a prudent choice.

ILLUSTRATING THE NEW METHODOLOGY

The new dissimilarity measures have been applied in

- Analyzing mobility *(16–18)*,
- Analyzing travel performance in home-work traffic *(19)*,
- Analyzing the mobility of elderly people *(20)*,
- Predicting the development of public transport usage,
- Analyzing differences in trip generation *(21)*,
- Analyzing car ownership *(22)*, and
- Predicting the development of car population and mobility *(23)*.

In this subsection the use of the new dissimilarity measures with a large data base is demonstrated.

Suppose the research goal is to analyze mobility of the population (expressed in daily mileage) by means of constructing homogenous population groups. A data base is available: the Netherlands National Travel Survey, which contains extensive information about households, the persons belonging to them, and the trips they make. Since 1978 about 23,000 persons have been interviewed annually. All potentially relevant characteristics (demographic, socioeconomic, etc.) were selected for analysis. This resulted in the following 14 variables (the number of distinct classes is shown in parentheses):

- Age (5)
- Car availability (3)
- Children in household (6)
- Citizenship (2)
- City size (3)
- Educational level (7)
- Employment status (5)
- Gender (2)
- Household income (6)
- Income per adult in household (6)
- Income of interviewee (6)
- Marital status (4)
- Position in household (5)
- Railway station nearby (2)

The object of study (daily mileage per person per travel mode) is assumed to have a Poisson-like distribution. Its dissimilarity formula can be found in the subsection on Development of New Dissimilarity Measures. Each travel mode can be viewed as a dimension in multidimensional space; analysis of all modes is done simultaneously.

Data from the Netherlands National Travel Survey result in the

TABLE 1 CALCULATED DISSIMILARITY VALUES (000)

Variable	No. of Classes in Variable	Loss of Information (dissimilarity) Caused by Internal Clustering of Variables (000)	
		All Classes	Final Step
Age	5	80	53
Car availability	3	201	178
Children in household	6	10	2
Citizenship	2	0	0
City size	3	8	4
Educational level	7	51	25
Employment status	5	49	49
Gender	2	74	74
Household income	6	14	8
Income per adult in household	6	13	12
Income of interviewee	6	105	74
Marital status	4	30	26
Position in household	5	101	80
Railway station nearby	2	5	4

scheme of Table 1, which indicates that car availability is the most significant characteristic; much less significant are personal income and position in household, followed by all other analyzed characteristics.

The data base is segmented into the classes of the most discriminating variable. Each subordinate data base is analyzed in a similar way. The final results are given in Table 2 and shown in Figure 1, which shows homogenous Dutch population groups with respect to daily mileage. Similar research, using data from years in the same time range, showed that these groups are fairly time stable *(17)*.

The example demonstrates clearly that this methodology can tackle large data bases without any problems. The only required input information is, for Poisson-distributed data, size and average value of each variable class.

The data in Table 3 make it possible to analyze dependencies. Two variables are independent when

$$G^2 > -2[LN\ L(A \text{ and } B) - LN\ L(A) - LN\ L(B)]$$

The value of $LN\ L(A \text{ and } B) - LN\ L(A)$ is given in the last column of Table 3 and shows the additional information gained in the second segmentation step. The first column gives the value of $LN\ L(B)$, the increase in information had the segmentation process been started with that variable. When used, the G^2-statistic will demonstrate that (in this case) only a few variables are independent of car availability.

Researching the stability of these variables with respect to geographic area or time, or both, might give additional insight; and experience, theoretical knowledge, and the like with respect to the phenomenon under analysis can be usable expedients in making a prudent selection.

CLOSING REMARKS

A new methodology for clustering and segmentation has been presented. Its main advantages are that

TABLE 2 HOMOGENEOUS POPULATION GROUPS IN THE NETHERLANDS WITH RESPECT TO TRAVEL PERFORMANCE (1980)

Group	Size of Group	Daily Kilometrage: Car Driver	Car Passenger	Train	Bus, Subway, Streetcar	Bicycle	Moped	Pedestrian	Total (includes other modes)
Car not available									
Under 18 years[a]	7,407		7.4	1.2	2.2	6.3	1.5	0.8	19.7
Housewife in non-car-owning household	1,027	0.1	4.4	2.2	2.0	2.2	0.3	1.1	12.6
Nonhousewife in non-car-owning household	2,580	0.9	4.8	3.1	2.5	3.0	0.5	1.1	16.5
Housewife in car-owning household	2,022	0.3	13.0	0.8	1.2	1.6	0.1	1.0	18.2
Nonhousewife in car-owning household	1,196	1.2	9.1	3.5	2.8	3.7	1.7	0.9	23.7
Subtotal	14,232	0.5	7.7	2.0	2.1	3.7	0.8	1.0	18.2
Car Sometimes Available									
Housewife	2,188	6.9	12.5	0.5	0.7	1.7	0.1	0.8	23.4
Nonhousewife	487	14.0	9.5	5.5	2.2	3.4	0.6	0.9	37.6
Subtotal	2,675	8.0	12.0	1.3	1.0	2.0	0.1	0.8	25.7
Car Available									
No Income	469	16.7	9.6	0.2	0.8	1.0	0.0	0.7	29.4
Under DFL 8,000 (net)	175	22.0	6.1	0.8	0.4	1.0	0.1	0.7	31.8
DFL 8,000-17,000 (net)	846	23.6	4.4	0.5	0.4	1.2	0.0	0.6	30.9
DFL 17,000-24,000 (net)	1,705	27.0	4.9	0.8	0.9	1.2	0.1	0.6	36.3
DFL 24,000-38,000 (net)	1,730	33.0	4.6	1.4	0.9	1.5	0.0	0.7	42.9
Over DFL 38,000 (net)	925	42.9	3.5	2.4	0.6	1.4	0.0	0.8	52.4
Subtotal	5,850	29.6	4.8	1.2	0.7	1.3	0.0	0.7	38.9
Total	22,757	12.3	7.3	1.6	1.5	2.5	0.4	0.9	26.9

Note: 1 km ≈ 0.62 mi.

[a]Minimum age for car drivers in The Netherlands is 18 years.

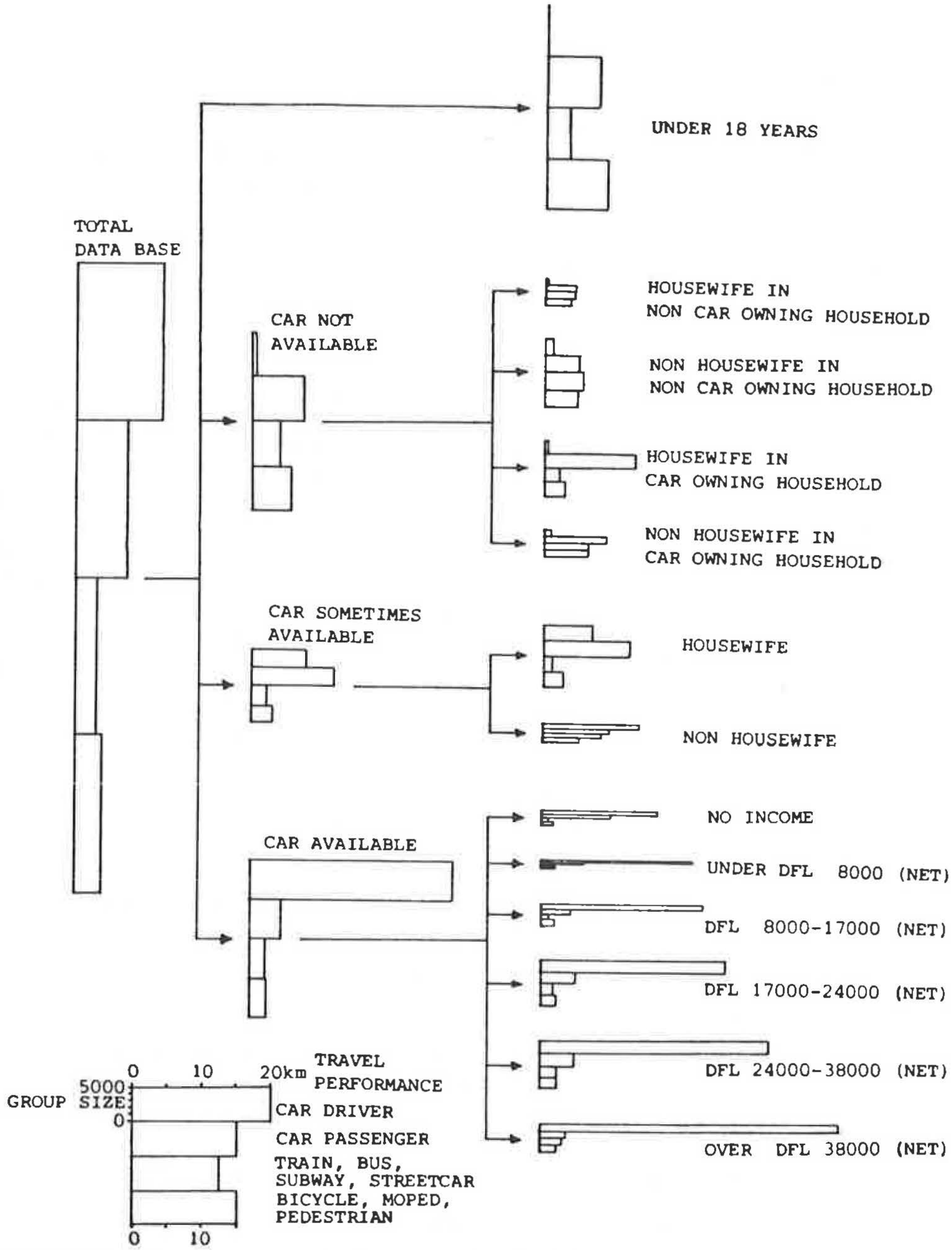


FIGURE 1 Homogenous population groups in The Netherlands with respect to travel performance (1980).

- There are no restrictions on the number of observations. There is also no limit on the quantity of data base variables. An a priori grouping of observations is therefore never necessary. This advantage is present because the technique belongs to the divisive hierarchical algorithms. Agglomerative hierarchical algorithms use distance matrices to calculate the differences between each pair of observations. However, the number of elements in such a matrix is limited by the memory of the computer used.
- The new methodology is consistent with the likelihood theory. It is therefore easier to justify its use than that of other cluster-segmentation methods because each probability density function will have its own specific measure of similarity or dissimilarity.
- Its dissimilarities can be multidimensional; for example, a measure based on daily mileage by several modes of transport.
- Dependencies between variables can be detected by using the likelihood ratio test.

The new method has been applied to various kinds of transportation research. Not only travel performance and trip generation, but also analysis of home-work trips and car ownership were objects of study. The results were in general accordance with expectations.

ACKNOWLEDGMENTS

The authors wish to express their thanks to DHV consulting engineers for providing the necessary facilities during the development of the derived cluster-segmentation algorithm. Also

TABLE 3 ANALYZING DEPENDENCIES

Variable	LN L(B)	Car Availabilty Categories[a] A	B	C	D	LN L(A and B) – LN L(A)
Car availabilty	201					
Personal income	105	0	9	2	12	23
Position in household	101	0	19	3	8	30
Age	80	0	15	2	3	20
Gender	74	1	12	3	7	23
City size	8	1	5	1	4	11
Railway station nearby	5	1	3	0	2	6
Citizenship	0	0	0	0	0	0

Note: Only the most extreme results are given. A = under 18 years, B = car not available, C = car sometimes available, and D = car available.

acknowledged are Steve Ashworth, Caroline Balhuizen, and Eric Kroes of NVI; Ben Immers of Delft University; and Ryuichi Kitamura of the University of California at Davis for their kind assistance during the preparation of this paper.

REFERENCES

1. H. Kassoff and H. D. Deutschmann. "Trip Generation: A Critical Appraisal." In *Highway Research Record 297*, HRB, National Research Council, Washington, D.C., 1969, pp. 15–30.
2. M. G. Richards. *The Use of Disaggregate Models in Travel Demand Analysis*. BGC Consulting Engineers, The Netherlands, 1975.
3. *Proc., Transportation Planning and Research Colloquium* (P. H. L. Bovy, ed.). Zandvoort, The Netherlands, 1983.
4. D. Wishart. *Clustan User Manual (Cluster Analysis Package)*. Edinburgh University, Scotland. 1978.
5. B. Everitt. *Cluster Analysis*. Heinemann Editions, London, England, 1974.
6. J. A. Hartigan. *Clustering Algorithms*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1975.
7. H. Späth. *Cluster Analysis Algorithms*. Univeristy of London, England, 1980.
8. D. Steinhausen and K. Langer. *Clusteranalyse, ein Einfuhrung in Methoden und Verfahren der Automatischen Klassifikation*. Federal Republic of Germany, 1977.
9. J. M. Ward. Application of a Hierarchical Grouping Procedure to a Problem of Grouping People. *Educational and Psychological Measurement*, No. 23, 1963.
10. R. Hamerslag. "Afstandsmaat voor de Samenstelling van Probleemgerichte Homogene Bevolkingsgroepen." In *Colloquium Vervoersplanologisch Speurwerk 1980* (P. H. L. Bovy et al., eds.), The Netherlands, 1980.
11. R. Hamerslag and W. H. Scheltes. "Het Gebruik van Cluster/Segmentatiemethoden in de Verkeerskunde." In *Verkeerskundige Werkdagen 1985* (T. De Wit, ed.), The Netherlands, 1985.
12. R. Dubes and A. K. Jain. *Clustering Methodologies in Exploratory Data Analysis*. MSF Grant ENG76-11936A01.1981.
13. B. Everitt. Unresolved Problems in Cluster Analysis. *Biometrics*, No. 35, 1979.
14. Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete Multivariate Analysis, Theory and Practice*. MIT Press, Cambridge, Mass., 1975.
15. M. G. Kendall and W. R. Buckland. *A Dictionary of Statistical Terms*. Longman, Edinburgh, Scotland. 1971.
16. R. Hamerslag. *Population Groups with a Homogeneous Travel Performance by Car, Public Transport and Bicycle*. Department of Civil Engineering, Delft University of Technology, The Netherlands, 1982.
17. R. Hamerslag, L. H. Immers, and J. M. Jager. "The Development of Mobility in The Netherlands." In *Tenth Transportation Planning and Research Colloquium*. (P. H. L. Bovy, ed.), Zandvoort, The Netherlands, 1983.
18. *Maatschappelijke Ontwikkelingen en Mobiliteit*. Projectbureau IVVS. Netherlands Institute of Transport, Rijswijk, The Netherlands, 1983.
19. J. M. Jager. *Onderzoek naar de Factoren die de Hoeveelheid Woon-Werkverkeer Bepalen*. Department of Civil Engineering, Delft University of Technology, The Netherlands, 1984.
20. J. J. Hooning et al. *Een Verkenning naar het Mobiliteitspatroon en de Mobileitsbehoefte van ouderen*. DHV Consulting Engineers, The Netherlands, 1985.
21. J. Termorshuizen, A. W. Dersjant, and T. Grijpma. *Studie in Ruimte en Tijd van Verplaatsingsproductie (Casestudie)*. Municipality of The Hague, The Netherlands, 1983.
22. J. M. Jager and W. H. Scheltes. "Analysis of Auto-Ownership by Using a Divisive Hierarchical Technique." In *Transportation Research Record 1037*, TRB, National Research Council, Washington, D.C., 1985, pp. 66–72.
23. R. Hamerslag and W. H. Scheltes. Impact of Fuel Price and Income Changes on Transportation. *Proc., Transportation Research Forum Annual Meeting*, Amelia Island, Fla., 1985, pp. 184–193.

APPENDIX

The derivation of the distance measure was as follows. Starting with the probability density function of the data base, the log-likelihood (*LN L*) is

$$LN\ L = LN[f(X1)] + LN[f(X2)] + LN[f(X3)] + \ldots LN[f(Xn)]$$

$$= \sum_{j=1}^{n} LN[f(Xj)]$$

where

f = probability density function,
L = likelihood value of the cluster, and
Xj = the jth observation in the cluster.

The dissimilarity measure between two observation clusters is defined as the difference in log-likelihood before and after clustering:

$$D(1,2) = (LN\ L1 + LN\ L2) - LN\ L12$$

where

$D(1,2)$ = dissimilarity measure between Group 1 and Group 2,
Li = likelihood value of the ith group, and
$L12$ = likelihood value of the combined group.

This appendix will demonstrate the derivation of dissimilarity formulas for some widely used density functions. Included are

- Binominal distribution,
- Normal distribution, and
- Poisson distribution.

Suppose there are two groups of data ($G1$ and $G2$) of known size ($N1$ and $N2$, respectively). All observations are stochastically independent. Let $X_{k,i,p}$ be the pth observation (pth person) of the ith group in the kth dimension of this multidimensional space. The mean value *(MEAN)* for the ith group in the kth dimension and the total value *(TOTAL)* follow from

$$MEAN_{k,i} = \sum_{p=1}^{Ni} (X_{k,i,p}/Ni)$$

and

$$TOTAL_{k,i} = \sum_{p=1}^{Ni} X_{k,i,p}$$

where

$MEAN_{k,i}$ = mean value for the ith group in the kth dimension,
Ni = number of observations in the ith group,
$TOTAL_{k,i}$ = total value of the ith group in the kth dimension, and
$X_{k,i,p}$ = the pth observation of the ith group in the kth dimension.

Symbolic names used in this appendix are given in Table A-1

Binominal Distribution

The binominal (Bernoulli) distribution is only defined for the values true/false (or 1/0, yes/no, etc.). Its mathematical form is

$$\begin{aligned} f(X) &= MEAN && \text{for } X = 1 \\ &= 1 - MEAN && \text{for } X = 0 \end{aligned}$$

Its likelihood *(L)* of N observations is

$$L = MEAN^{TOTAL} * (1 - MEAN)^{N - TOTAL}$$

and the log likelihood *(LN)* is

$$\begin{aligned} LN\ L &= LN(MEAN^{TOTAL}) + LN[(1 - MEAN)^{N - TOTAL}] \\ &= TOTAL*LN(MEAN) + (N - TOTAL)*LN(1 - MEAN) \end{aligned}$$

For Group $G1$ this results in

$$LN\ L(G1) = \sum_k [TOTAL_{k,1}*LN(MEAN_{k,1}) + (N1 - TOTAL_{k,1})*LN(1 - MEAN_{k,1})]$$

where

L = likelihood,
$MEAN$ = average observed value,
$N1$ = size of Cluster $G1$, and
$TOTAL$ = total observed value in cluster.

Groups $G2$ and $G12$ lead to similar formulas. Calculation of the dissimilarity *(D)* follows from

$$\begin{aligned} D(1,2) &= LN\ L(G1) + LN\ L(G2) - LN\ L(G12) \\ &= \sum_k [TOTAL_{k,1}*LN(MEAN_{k,1}) \\ &\quad + TOTAL_{k,2}*LN(MEAN_{k,2}) \\ &\quad - TOTAL_{k,12}*LN(MEAN_{k,12}) \\ &\quad + (N1 - TOTAL_{k,1})*LN(1 - MEAN_{k,1}) \\ &\quad + (N2 - TOTAL_{k,2})*LN(1 - MEAN_{k,2}) \\ &\quad - (N12 - TOTAL_{k,12})*LN(1 - MEAN_{k,12})] \end{aligned}$$

where

$MEAN$ = average observed value,
Ni = size of the ith cluster, and
$TOTAL$ = total observed value.

Normal Distribution with Constant Variance

In general, a normal distribution will be chosen when the object of study contains both positive and negative observations. It can also be used in case its mean value differs significantly from zero. The probability density function is

$$f(X) = [S(2\pi)^{1/2}]^{-1} * \exp\{-(1/2)[(X - MEAN)/S]^2\}$$

where

f = probability density function,
$MEAN$ = average observed value,
S = standard deviation,
X = stochastic variable, and
π = mathematical constant (about 3.14)

TABLE A-1 SYMBOLIC NAMES

Group Identification	Size of Group	Observations	Mean Value	Total Value
$G1$	$N1$	$X_{k,01,1} \cdots X_{k,01,N1}$	$MEAN_{k,01}$	$TOTAL_{k,01}$
$G2$	$N2$	$X_{k,02,1} \cdots X_{k,02,N2}$	$MEAN_{k,02}$	$TOTAL_{k,02}$
$G12$ ($G1$ and $G2$)	$N12 = N1 + N2$	$X_{k,12,1} \cdots X_{k,12,N12}$	$MEAN_{k,12}$	$TOTAL_{k,12}$

The calculation of likelihood (L) and log-likelihood $(LN\ L)$ results in

$$L = [S(2\pi)^{1/2}]^{-1} * (\exp\{-(1/2)[(X1 - MEAN)/S]^2\}) * [S(2\pi)^{1/2}]^{-1} * (\exp\{-(1/2)[(X2 - MEAN)/S]^2\}) * [S(2\pi)^{1/2}]^{-1} * (\exp\{-(1/2)[(X3 - MEAN)/S]^2\}) * \ldots$$

and

$$LN\ L = LN([S(2\pi)^{1/2}]^{-1} * \exp\{-(1/2)[(X1 - MEAN)/S]^2\}) + LN([S(2\pi)^{1/2}]^{-1} * \exp\{-(1/2)[(X2 - MEAN)/S]^2\}) + LN([S(2\pi)^{1/2}]^{-1} * \exp\{-(1/2)[(X3 - MEAN)/S]^2\}) + \ldots$$

where

L = likelihood,
$MEAN$ = average observed value,
S = standard deviation,
Xj = the jth observed value of a stochastic variable, and
π = pi, mathematical constant (about 3.14)

For group $G1$ (size $N1$ and observations $X_{k,1,1}$ through $X_{k,1,N1}$) this leads to

$$LN\ L(G1) = N1*LN[S(2\pi)^{1/2}]^{-1} + \left\{-(1/2)*\sum_k \sum_{p=1}^{N1} [(X_{k,1,p} - MEAN_{k,1})/S]^2\right\}$$

Similar formulas are found for Groups $G1$ and $G12$. The dissimilarity measure (D) can be calculated from

$$D(1,2) = LN\ L(G1) + LN\ L(G2) - LN\ L(G12)$$

which results in

$$D = N1*LN[S(2\pi)^{1/2}]^{-1} - (1/2)*\sum_k \left\{\sum_{p=1}^{N1} [(X_{k,1,p} - MEAN_{k,1})/S]^2\right\} + N2*LN[S(2\pi)^{1/2}]^{-1} - (1/2)*\sum_k \left\{\sum_{p=1}^{N2} [(X_{k,2,p} - MEAN_{k,2})/S]^2\right\} + N12*LN[S(2\pi)^{1/2}]^{-1} - (1/2)*\sum_k \left\{\sum_{p=1}^{N12} [(X_{k,12,p} - MEAN_{k,12})/S]^2\right\}$$

Simplified,

$$D = -\sum_k \left\{\sum_p [MEAN_{k,1}^2 - 2*(MEAN_{k,1}*N1) * MEAN_{k,1}]\right\}/2S^2 - \sum_k \left\{\sum_p [MEAN_{k,2}^2 - 2*(MEAN_{k,2}*N2) * MEAN_{k,2}]\right\}/2S^2 + \sum_k \left\{\sum_p [MEAN_{k,2}^2 - 2*(MEAN_{k,12}*N12) * MEAN_{k,12}]\right\}/2S^2$$

Finally the formula results in

$$D = \sum_k (N1*MEAN_{k,1}^2 + N2*MEAN_{k,2}^2 - N12 * MEAN_{k,12}^2)/2S^2$$

Poisson Distribution

A Poisson distribution is characterized by exclusively nonnegative integer observations. For large mean observed values, the function approaches a normal distribution. The mathematical form of the Poisson probability density function (f) is

$$f(X) = [MEAN^X * exp(-MEAN)]/X!$$

where

f = probability density function,
$MEAN$ = average observed value, and
X = observation.

Likelihood (L) and log-likelihood $(LN\ L)$ follow from

$$L = [MEAN^{X1} * \exp(-MEAN)]/X1! * [MEAN^{X2} * \exp(-MEAN)]/X2! * \ldots$$

and

$$LN\ L = X1 * LN(MEAN) - MEAN - LN(X1!) + X2 * LN(MEAN) - MEAN - LN(X2!) + \ldots$$

where

L = likelihood,
$MEAN$ = average observed value, and
Xj = the jth observed value of a stochastic variable.

For Group $G1$ (size $N1$ and observations $X_{k,1,1}$ through $X_{k,1,p}$) this results in

$$LN\ L(G1) = \sum_k \sum_{p=1}^{N1} [X_{k,1,p} * LN(MEAN_{k,1}) - MEAN_{k,1} - LN(X_{k,1,p}\,!)]$$

Similar formulas are found for Groups $G2$ and $G12$. The dissimilarity (D) follows from

$$D(1,2) = LN\ L(G1) + LN\ L(G2) - LN\ L(G12)$$

This leads to

$$D = \sum_k \sum_{p=1}^{N1} [X_{k,1,p} * LN(MEAN_{k,1}) - MEAN_{k,1} - LN(X_{k,1,p}\,!)]$$

$$+ \sum_k \sum_{p=1}^{N2} [X_{k,2,p} * LN(MEAN_{k,2}) - MEAN_{k,1} - LN(X_{k,2,p}\,!)]$$

$$- \sum_k \sum_{p=1}^{N12} [X_{k,12,p} * LN(MEAN_{k,12}) - MEAN_{k,12} - LN(X_{k,12,p}\,!)]$$

because

$$\sum_{p=1}^{N1} LN(X_{k,1,p}!) + \sum_{p=1}^{N2} LN(X_{k,2,p}!) = \sum_{p=1}^{N12} LN(X_{k,12,p}!)$$

and

$$N1*MEAN_{k,1} + N2*MEAN_{k,2} = N12*MEAN_{k,12}$$

Thus the formula can be simplified to

$$D = \sum_k [TOTAL_{k,1}*LN(MEAN_{k,1}) + TOTAL_{k,2}*LN(MEAN_{k,2}) - TOTAL_{k,12}*LN(MEAN_{k,12})]$$

Notice that this formula is also usable where all observations are zero ($MEAN = 0$) because

$$LIM\ TOTAL_{k,i} * LN(MEAN_{k,i}/Ni) =$$

$$LIM\ TOTAL_{k,i} * LN(TOTAL_{k,i}/Ni) = 0 \quad \text{for } TOTAL_{k,i} \downarrow 0$$

Road Classification According to Driver Population

SATISH C. SHARMA, PAWAN J. LINGRAS, MASOOD U. HASSAN, AND NARASIMHA A. S. MURTHY

A proposed model for classification of rural roads according to driver population characteristics is described. The driver population is distinguished by such traffic stream characteristics as trip purpose and trip length distribution, and the basic assumption made in the analysis is that the different traffic flow patterns observed at road sites result from different mixes of these characteristics. The highway systems of the provinces of Alberta and Saskatchewan are investigated for the purpose of developing and testing the model. The model suggests the application of some standard computational and statistical techniques to develop master patterns of traffic flow in order to recognize the driver population of a given road site. The proposed model is simple to apply and its data requirements can easily be satisfied by the types of data collection programs normally undertaken by highway agencies. Also, it is believed to be more objective and comprehensive than the existing methods used for the same purposes. The road classification resulting from the model could be used as an important criterion for many highway planning and design functions. Some examples of its application are *(a)* rationalization of provincewide traffic-counting programs, *(b)* design hourly volume considerations, *(c)* highway improvement programming, and *(d)* highway capacity analysis.

Road classification is important for administration, planning, design, and operation of facilities. Systems for classifying roads are numerous, and the class definitions vary depending on the purpose of classification. Some examples of classification systems used are *(a)* according to jurisdiction and funding—federal, provincial, and municipal roads—and *(b)* according to type of road function—freeways and expressways, arterials, collectors, and local roads. Several of the common classification systems are presented in the Institute of Transportation Engineers handbook (*1*, pp. 599–604) and Roads and Transportation Association of Canada (RTAC) *(2, 3)* publications.

Many of the provincial or state highway agencies also classify roads according to volume characteristics, such as temporal volume variations and other traffic stream characteristics. This type of road classification is frequently used in traffic volume estimation, design hourly volume (DHV), and peak-hour traffic considerations. There exists a diversity of definitions, systems, and procedures for this type of road classification both in Canada and abroad. Consequently a considerable input of subjective judgment is used in various planning and designing activities that pertain to different types of roads.

Recently, the variable "driver population" or user's perspective has been considered a significantly important factor in an increasing number of highway planning and design functions. An interesting example of this is the new (1985) edition of the Highway Capacity Manual (HCM) (*4*, p. 3–17), which suggests the use of an adjustment factor (f_p) to reflect the influence of driver population on highway capacity calculations. Some other examples of areas in which the road use (the terms "road use" and "driver population" are used interchangeably to reflect the traffic stream characteristics) variable has appeared to be an important factor are *(a)* design hourly volume considerations *(5–7)*, *(b)* cost-effective sizing and upgrading of two-lane rural highways *(7)*, and *(c)* rationalization of traffic monitoring *(8–12)*.

S. C. Sharma and P. J. Lingras, Faculty of Engineering, University of Regina, Regina, Saskatchewan S4S 0A2, Canada. M. U. Hassan, Saskatchewan Highways and Transportation, 1855 Victoria Avenue, Regina, Saskatchewan S4P 3V5, Canada. N. A. S. Murthy, Department of Civil Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada.

The main objective of this paper is to offer an improved method of classification, based on trip purpose and trip length distribution and temporal volume variations, that would lead to a better understanding of the road user's perspective and hence provide further insight into planning and design of road facilities from the users' point of view.

The proposed method suggests the use of standard computational and statistical techniques and therefore is expected to yield more objective and statistically credible groupings of road sites than do existing methods. Another objective of this paper is to recommend more specific values (compared with the 1985 HCM) of the factor f_p to be used in the capacity analysis for different types of driver populations. The aim here is to provide a sound basis for engineering judgment that must be exercised by the analyst in selecting an exact value of f_p from the wide range of choice presented by the new HCM.

BRIEF REVIEW OF THE STATE OF THE ART

Before an alternative method of road classification is presented, it is worthwhile to review various procedures that are presently used in Canada and abroad and to describe briefly matters that are interrelated with classification of road sites according to driver population.

Traffic Monitoring and Classification of Road Sites

Grouping of permanent traffic counter (PTC) sites is required for such traffic-monitoring purposes as the estimation of average annual daily traffic (AADT) from sample counts. The most commonly used method of grouping PTCs is that recommended by the Bureau of Public Roads (BPR) *(13)*. In this method, the counters are grouped on the basis of monthly traffic factors, which are defined as the ratio of the AADT to the average weekday traffic of the month. The BPR method utilizes a manual ranking system in which the PTCs are listed in ascending order of monthly factors.

For each month, a group of counters is determined so that the difference between the smallest and the largest factor does not exceed a range of 0.20 in the values of the factors. In other words, the criterion of grouping is a subjectively and arbitrarily chosen value of ±0.10 from the assumed mean. The final grouping of counters in this method is supposed to be such that all or as many as possible of the same counters fall into the same group each month.

The provincial highway agencies in Canada use a variety of methods for grouping their PTC sites *(14,15)*. Some of them use classifications that are similar to the BPR method. Others employ different grouping criteria and class definitions that are largely arbitrary and subjective in nature. One agency has more than 150 PTCs but it has not carried out any specific classification of these sites.

Bellamy *(16)* used a cluster analysis technique in an attempt to classify temporal variations in traffic flows at the 50-point PTC sites in Britain. From the results of the cluster analysis, the study found that it was difficult to decide what grouping of the sites was most appropriate. Because the cluster analysis grouping could not be regarded as conclusive, the study was reconciled by using a largely subjective method that was quite similar to the BPR method. Four descriptive classes of road sites suggested in the study were *(a)* urban and commuter, *(b)* nonrecreational low flow, *(c)* rural long distance, and *(d)* recreational.

DHV Considerations and Classification of Road Sites

Highway authorities and researchers recognize the importance of road use in estimating the DHV for a new facility or upgrading an existing facility *(1;5–7;17,* pp. 170–175). In this regard, the ITE handbook *(1)* identifies different highway routes by such types as "urban through route," "suburban through route," "main rural route," "secondary rural route," partially recreational route," and "highly recreational route." Figure 4.9 of the handbook shows the conventional presentation of highest hourly volumes for different types of highway routes. No systematic and clear definitions of such route classes are available in the literature.

Highway Capacity Analysis and Driver Population

Research work on highway capacity analysis conducted in the past few years indicates that capacity as well as service flow rates for other levels of service are significantly affected by traffic stream characteristics. The new HCM *(4)* recognizes this factor and uses an adjustment factor called f_p in the calculation of highway capacity. An adjustment factor (f_p) of 1.0 is assigned to "weekday or commuter" traffic. A range of f_p of between 0.75 and 0.90 is suggested for use in the cases of "other" types of traffic streams.

The range of adjustment factor (f_p) suggested by the HCM is substantial. The use of such a wide range (i.e., 0.75 to 0.90 or 0.75 to 0.95) for "other" types of facilities will require great caution on the part of the analyst. Classification of road sites according to traffic stream characteristics would be useful for conducting comparative field studies and selecting an appropriate value of the factor for accurate capacity analysis.

Limitations of the Existing Classification Methods

It is apparent that the existing methods of grouping road sites are generally arbitrary, and, consequently, opinions differ on the number and best definition of classes. A considerable input of subjective judgment is used not only in factoring a sample count to estimate the AADT but also in various planning and design activities pertaining to different types of roads. As a result of this, significant differences exist among Canadian provinces in the way they carry out certain transportation functions. An example of this is the use of more than 150 PTCs by the province of Quebec compared with only 21 PTCs used by Ontario.

The previously mentioned limitations of the existing classification and the need to consider driver population in capacity analysis suggest that more effort is needed to investigate the classification of road sites. The central theme of this research is to develop a systematic and objective method of road classification according to type of road uses. The analysis presented here includes consideration of temporal variations in flows at given roadway locations and information, such as trip purpose and trip length distribution, to explain these variations.

STUDY DATA

The primary highway system of the province of Alberta was investigated for the purpose of this study. The permanent traffic counters located on the system provided seasonal, daily, and hourly patterns of traffic. An analysis of monthly and daily variations in traffic flows was carried out over a number of years. The analysis appeared to show that at a number of PTC sites there were some differences between the patterns for different years, but these differences did not show an overall systematic trend with time. As a result of the analysis, it was considered preferable to use the monthly and daily traffic data average over a period of 5 years. Considering the reliability of the available Alberta PTC data, a total of 52 counter sites were selected for the study.

From the past origin-destination surveys conducted by Alberta Transportation, trip purpose and journey length information corresponding to some of the counter locations in the province was available. This type of information was used to rationalize the proposed method of classification of the road sites.

For the purpose of testing the proposed method of road classification, a total of 28 PTC sites located on the highway system of the province of Saskatchewan were also investigated on the basis of their seasonal, daily, and hourly variations in traffic flows.

MODEL OF ROAD CLASSIFICATION ACCORDING TO DRIVER POPULATION

The parameter f_p, driver population, used in the new HCM refers to trip purpose characteristics, such as commuter and recreational. Another traffic stream characteristic that could be important from the point of view of capacity calculation or other transportation functions is the trip length distribution aspect of the driver population. This study utilizes both trip purpose (e.g., commuter, recreational) and trip length distribution (e.g., urban, regional, interprovincial) as the descriptors of the driver population.

The basic assumption made in the development of the proposed model, called the DRIPOP model hereafter, is that the difference in traffic flow patterns observed at road sites results from different mixes of trip characteristics. The high morning and afternoon peaks are due to the high number of home-to-work and work-to-home trips in urban areas. High weekend traffic is associated with high weekend social and recreational trips. Also, the higher traffic volumes in certain months, such as July and August, can be related to summer holidays and long-distance tourist recreational trips.

Development of Master Traffic Patterns

The objective of this step of the DRIPOP model is to group the volume distributions of road sites and obtain the typical or "master" patterns of traffic flow. The patterns that result from this step provide input to the final classification of roads according to driver population.

Hierarchical Grouping

The hierarchical grouping method is used mostly in behavioral research. The purpose of this method is to compare a set of N objects (e.g., 52 road sites in this study) each measured on K different variables (e.g., 12 monthly traffic factors) and group them in such a manner that groups are similar in their values of the K variables.

In this paper no attempt is made to explain the hierarchical grouping method in detail. Instead, the basic premise and the main criteria for this method are briefly described with reference to the classification of roads. There are several sources of information on the hierarchical grouping procedure, and particular reference may be made to Veldman (*18*, pp. 308–317) and Ward (*19*). A FORTRAN computer program of the hierarchical grouping is provided in Veldman (*18*). The procedure carried out in this step for grouping the sample road sites is adopted from a paper by Sharma and Werner (*20*).

The hierarchical grouping of the road sites is based on the premise that the maximum amount of information is available when the N sites are ungrouped. Hence the grouping process begins by defining each of the N sites as a "group." The first step in grouping reduces by one the number of groups by selecting two groups that produce the least amount of within-group error. The remaining $N - 1$ groups are then reduced in number by a series of step decisions until all the sites are put in a single group. Each step of the process systematically reduces the number of groups by one.

The errors associated with successive stages of the grouping process indicate the marginal "cost" of reducing the number of groups by one. The error at a particular stage of grouping is greater than or equal to the error associated with the previous stage of grouping.

It has to be emphasized here that this method is primarily descriptive and does not indicate specifically what the optimum number of groups is for the study objectives. However, the errors associated with the successive stages of the grouping process will usually reveal a "knee-of-curve" range of grouping stages that is especially worthy of study. By applying this procedure and plotting the results (Figure 1), the errors associated with the groupings of the 52 road sites on the basis of the 12 monthly factors of this study can be shown. It appears from this figure that the optimum number of groups lies somewhere between 20 and 5 because a substantially large increase in error is observed in this range and beyond.

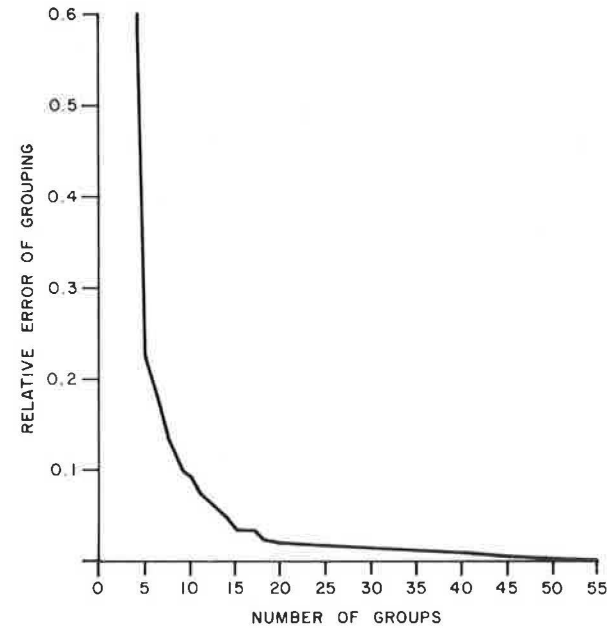


FIGURE 1 Incremental errors associated with the hierarchical grouping of monthly traffic factors.

Seasonal Traffic Patterns

One important advantage of the hierarchical grouping technique is that it provides the analyst with the opportunity to analyze the resulting groups at any chosen stage of the process. Referring to Figure 1, for example, an analyst might consider analyzing subjectively the patterns at the stage of 20 groups, and after a critical examination he might decide, even without performing any statistical analysis, to stop the hierarchical grouping at that stage. Such a large number (i.e., 20) might be considered quite appropriate by some agencies for grouping PTCs for obtaining monthly factors for estimating AADT. But if the objective is to obtain a smaller and more manageable number of road classes, the analyst might need to explore the entire range of grouping encompassing the knee-of-curve portions of the grouping errors such as shown in Figure 1.

Statistical comparisons, such as Scheffe's S-method of multiple comparisons of group means (*21*, pp. 53–73) or simple t-tests can help in the determination of the most appropriate number of master patterns of traffic distribution. For this purpose the study sites can be assumed to have been selected at random from the point of view of statistical theory (*16,22*).

The significance of differences among the group means can be established on the basis of group comparisons for each month of the year in the case of seasonal patterns. However, there is a need to exercise caution in that the F-tests computed for the S-method or t-tests used in comparing the mean monthly factors can be artificially and unreliably significant. The experience gained in this study indicates that the patterns should be considered different when the tests are significant for 5 or more months at a 95 percent confidence level.

The application of hierarchical grouping and statistical comparisons for the purpose of Alberta road classification resulted in four distinct major patterns at the level of seven groups (Figure 1) of the hierarchical process. These four major patterns, shown in Figure 2, account for nearly 95 percent of the study sites. The rest of the study sites included in the remaining three groups appear to represent special patterns, such as Site C 162 that exhibits a winter peak due to a large number of ski trips in winter months. The plots of patterns for the special cases are excluded from Figure 2 to avoid overcrowding the figure.

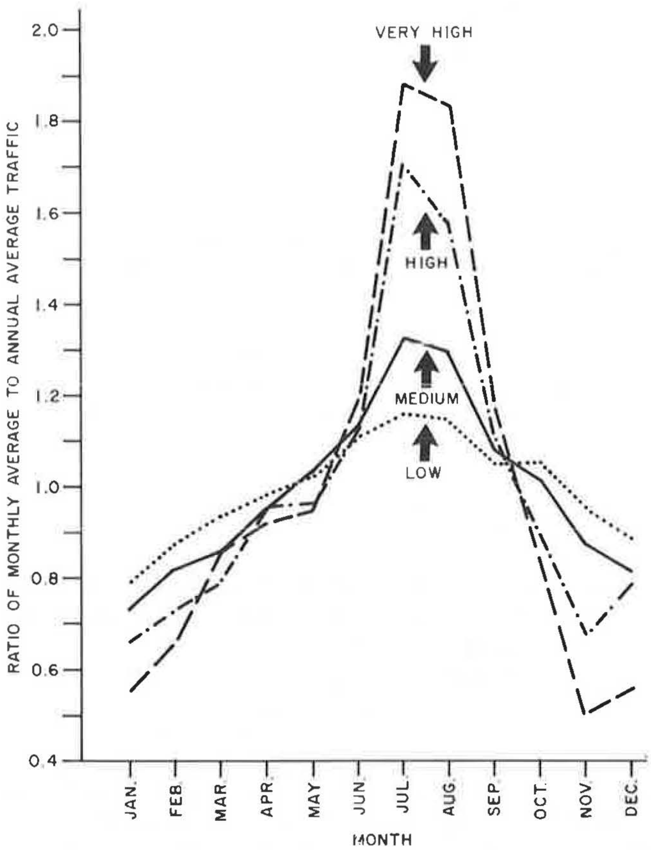


FIGURE 2 Average monthly traffic factors for the four major patterns of seasonal variation.

For the sake of simplicity of presentation, the four major patterns of seasonal variation are called "low," "medium," "high," and "very high." The magnitude of the seasonal variation exhibited by the master patterns of Figure 2 can be explained by the variation in trip purpose at road sites. However, because these master patterns should be used along with the daily and hourly variations and the trip length distribution, their explanation is postponed until a road classification is proposed.

Daily and Hourly Patterns

The different seasonal traffic patterns that resulted from the previous step were further analyzed systematically in terms of their ability to represent a more specific categorization of the Alberta roads. The study sites assigned to a particular master seasonal pattern indeed exhibited discernible and consistent patterns of daily and hourly variations in traffic flows. It became evident from this that the temporal variation patterns could be reliably and systematically related to different types of road uses.

The experience of working with the study data indicated that it is best to analyze daily and hourly characteristics during the months, such as May to August, inclusive, when the volumes are expected to be most stable *(9,23)* and they represent average to heaviest traffic conditions. Another reason to select such months is that most of the rural traffic data collection programs (e.g., seasonal and short-period sample counts) that can be helpful in the proposed road classification are undertaken during these months.

The relative magnitudes of traffic volume when averaged over all 7 days were carefully investigated to distinguish various types of roads. But the observations made in this study clearly indicated that it would be equally effective and simpler to use the Sunday volume factor for July—the month of the heaviest traffic volume. The Sunday factor, used as an indicator of the master daily pattern for the road classification, was defined as the ratio of average Sunday volume to the average weekday volume where the weekdays included Tuesday, Wednesday, and Thursday. The hierarchical grouping technique was used to group all of the study sites into three groups, which are defined as follows.

Daily Variation Pattern	*Sunday Factor in July*
Low	Lower than 1.1
Medium	1.1–1.4
High	Higher than 1.4

The relative Sunday volume of traffic can be taken as an indicator of weekend social-recreational trips. A high Sunday factor would result when Sunday traffic is much heavier than average weekday traffic volume. The road sites that carry large volumes of commuter and business trips on weekdays and do not serve an appreciable number of weekend social-recreational trips would be expected to exhibit a low Sunday factor. But it may be noted that a low (or medium) Sunday factor may also be exhibited by those roads that carry large numbers of weekend recreational trips. An example of such a case is when the road carries a large number of summer holiday or tourist trips that continue throughout the week. The increased level of weekday traffic due to the tourist trips tends to lower the Sunday factor.

The distributions of traffic volume by hour of the day often describe the peak demands for service. The morning and afternoon peak-hour periods on weekdays represent home-to-work and work-to-home trips, respectively. Consideration of such peaks during the weekdays could help to better understand the classification of roads.

Figure 3 shows three typical patterns of hourly volumes identified in this study for summer weekdays in July. These patterns are *(a)* commuter pattern, in which the morning and afternoon peaks are quite clearly visible; *(b)* partly commuter pattern, in which only moderate increase in traffic is experienced during the peaks; and *(c)* noncommuter pattern, in which the morning and afternoon peaks are not visible.

Considerations of Trip Characteristics and Classification of Roads

The basic assumption made in the analysis for the proposed DRIPOP model is that the difference in overall flow patterns

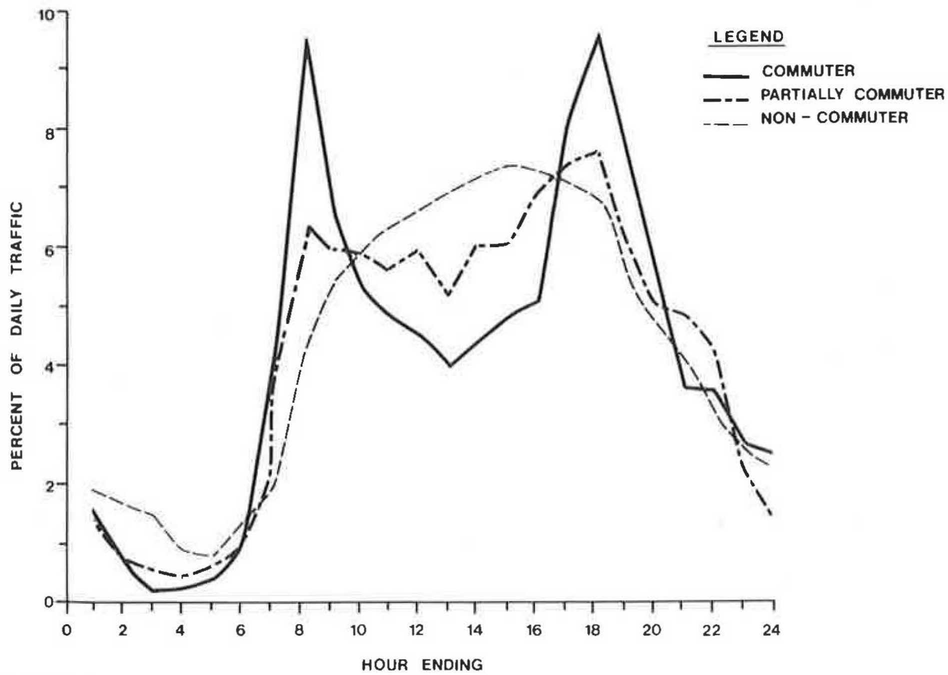


FIGURE 3 Typical patterns of hourly traffic volume during summer weekdays.

observed at road sites results from different mixes of trip characteristics. As mentioned earlier, trip purpose and trip length information was available from past origin-destination surveys by Alberta Transportation for a limited number of study sites. Trip purpose data were used to verify the temporal volume variations that provide the basis of the classification of roads, and trip length distribution information was used to determine whether the road uses were mainly local, regional, or interregional and long distance in nature.

The various trip purposes were grouped in two broad categories: work-business purposes, the number of trips for which is not considered to vary much throughout the year, and social-recreational trips, for which the amount of travel obviously increases during certain seasons of the year, such as the summer months. Table 1 gives the available data in percentages of work-business and social-recreational trips for a limited number of study sites of different seasonal groups.

The seasonal variation in traffic volumes, as shown in Figure 2, can be easily explained by the corresponding variation of trip purposes given in Table 1. The lowest seasonal variation in the case of the low group is due to a high proportion of weekday work-business trips and the associated low or medium Sunday factor. The progressively high seasonal variation for the other sites is due to the higher proportion of social-recreational trips and the associated weekend (Sunday) traffic volumes. The very high seasonal variation in the case of Site C 114 can be attributed to the high number of both weekend and holiday recreational trips during the summer months. Note that the presence of holiday or tourist recreational trips that continue during the weekdays will tend to decrease the relative magnitude of weekend traffic that is represented by the Sunday factor in this study.

The cumulative trip length distribution provided further insight into the classification of roads. Figure 4 shows typical patterns that may be used to describe the trip length characteristics of road sites. These patterns can be grouped into three broad types: *(a)* regional road sites, *(b)* interregional road sites, and *(c)* long-distance road sites. A majority of regional trips would take less than 60 min travel time. The interregional and long-distance patterns would exhibit longer trip length distributions.

TABLE 1 TRIP PURPOSE AT SOME STUDY SITES DURING SUMMER WEEKDAYS

Road Site	Seasonal Variation Group[a]	Daily Variation (July Sunday) Pattern[b]	Trip Purpose (%) Work-Business	Social-Recreational
C 9	Low	Low	83	17
C 66	Low	Low	74	26
C 72	Low	Low	81	19
C 75	Low	Low	82	18
C 138	Low	Low	77	23
C 144	Low	Low	81	19
C 42	Low	Medium	62	38
C 93	Low	Medium	68	32
C 15	Medium	Low	62	38
C 18	Medium	Low	37	63
C 39	Medium	Medium	64	36
C 57	Medium	High	56	44
C 63	Medium	Medium	60	40
C 36	High	High	39	61
C 114	Very high	Medium	28	72
C 165	Very high	High	27	73

[a]Based on master patterns as shown in Figure 2.
[b]Based on the July Sunday factor.

The data in Table 2 put in perspective the resulting classification of the Alberta road sites based on the temporal volume patterns that represent the trip purpose variable and the trip length distribution. This classification is mainly a function of traffic stream characteristics during the summer months when most of the predominant road uses, such as work-business trips, social-recreational trips, and tourist trips, are present in the traffic stream. It may be noted that the eight road classes listed in Table 2 are significantly different from each other in at least one consideration.

Regional commuter routes are located in the commutershed areas of major urban centers and serve predominantly work-busi-

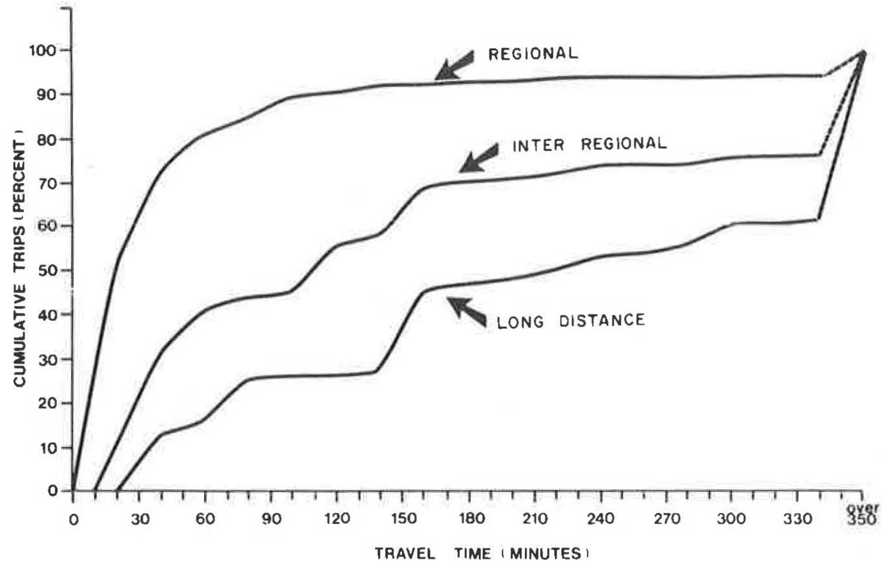


FIGURE 4 Typical trip length distributions for the different types of road sites.

ness trips. This group also serves medical, shopping, and social-recreational trips. The distinguishing road uses for the regional recreational and commuter routes are social-recreational trips on weekends and commuter and other types of regional trips on weekdays. These roads are also found in the commutershed areas.

Interregional and long-distance roads are located outside the major commutershed areas and carry a large proportion of tourist or holiday recreational trips during the summer months. A large amount of weekend social-recreational traffic is carried by long-distance and recreational routes in addition to the usual long-distance trips. Those major highways are included in the highly recreational group, which carries large numbers of both weekend and tourist recreational trips during summer months.

The rural commuter and business group includes low-volume (AADT in the range of from 1,000 to 1,500) rural roads that do not carry many social-recreational trips. Farm-to-market roads are included in this group.

These patterns of roads would be expected to result from the application of the DRIPOP model to any provincial highway system. But there will be some special routes that serve the needs of a specific community or region. Some examples of Alberta's special routes identified in this study are *(a)* winter resource supply route (e.g., site C 162); *(b)* regional resource development commuter route (e.g., Site C 153); and *(c)* summer-winter recreational route (e.g., Site C 162). The temporal volume patterns of this group are expected to be variable.

The urban commuter group, patterns of which are well known to traffic analysts, is not included in the description because the study data obtained from Alberta Transportation did not contain any sites in urban centers. However, urban commuter routes can be easily recognized by low seasonal variation (lower than regional commuter), low Sunday, and very prominent morning and afternoon peak periods.

Another point that can be made here is that the roads identified in this classification are major facilities that carry traffic volume in an AADT range of 1,000 or higher. Roads with lower traffic volumes tend to show generally unstable flow patterns and therefore are not recommended for classification by this method.

TABLE 2 ROAD CLASS DEFINITIONS BY TRAFFIC STREAM CHARACTERISTICS

	Temporal Volume Characteristics			
Road Class	Seasonal Variation Group[a]	Daily Variation (July Sunday) Pattern[b]	Hourly Variation Pattern[c]	Trip Length Distribution[d]
Regional commuter	Low	Low	Commuter	Regional
Regional recreational and commuter	Low, medium	Medium	Partly commuter	Regional and interregional
Interregional	Medium	Low, medium	Noncommuter	Interregional
Long distance	Medium	Low	Noncommuter	Long distance
Long distance and recreational	High	High	Noncommuter	Long distance
Highly recreational	Very high	Low, medium	Noncommuter	Long distance
Rural commuter and business	Low	Low	Partly to noncommuter	Variable
Special	Variable	Variable	Variable	Variable

[a]As shown in Figure 2.
[b]The Sunday factor as defined previously.
[c]As shown in Figure 3.
[d]As shown in Figure 4.

Classification of Roads Using Seasonal Traffic Counts

The basic requirements of the DRIPOP model are the master patterns of temporal variations in traffic volume. The trip purpose information in the preceding development of the Alberta road classes was used only to rationalize some of the volume patterns. It is not necessarily a required type of data input. Similarly, the actual data on trip length distribution are also not essential because highway authorities generally know whether road uses of their systems are local, regional, or provincial and interprovincial in nature.

The PTCs that are employed by nearly all highway agencies are the sources of the information needed to develop the master patterns of volume variations. A highway agency can also classify its roads by taking seasonal traffic counts because, when carefully programmed, such sample counts can provide a reasonably good estimate of monthly, daily, and hourly patterns. The variation patterns that result from the sample counts can be matched with the master patterns to classify the study sites into a road group.

TESTING AND APPLICATIONS OF THE DRIPOP MODEL

Application of the DRIPOP Model to Saskatchewan Highways

At present, the Traffic Analysis Section of Saskatchewan Highways and Transportation classifies the provincial roads into *(a)* Trans-Canada Highway, *(b)* rural highways, *(c)* resort highways, *(d)* urban streets, and *(e)* municipal roads. The rural highways class is further divided into two groups according to the volume of traffic: greater than 600 AADT and less than 600 AADT. The resort roads are separated into two AADT ranges: more than 2,000 AADT and under 2,000 AADT. The context and objectives of such classification are the same as those of the DRIPOP model.

The master patterns of seasonal, daily, and hourly volume patterns were developed for the 24 PTC sites in Saskatchewan, and the road sites were reclassified using the DRIPOP model. The resulting classes were found to be quite similar to the classes for Alberta highways. From the reclassification of roads it became evident that the existing method of classifying count sites on Saskatchewan highways is arbitrary and subjective in nature. The DRIPOP model showed one major discrepancy in the grouping of count sites in Saskatchewan. The four PTC sites on the Trans-Canada Highway should be in two groups instead of one. Two of the sites exhibit traffic patterns that are related to the regional commuter group, and the other two sites fall into the long-distance group. The highway agency has already rectified this discrepancy by regrouping the road sites.

Examples of Some Recent Applications of the DRIPOP Model

The experience gained from the development of the DRIPOP model and the increased understanding of various road classes according to driver population have provided a better comprehension of several aspects of highway planning and design functions. Sharma *(24)* showed that the traffic stream characteristics of the road site being surveyed are among the most important considerations for rationalization of short-period manual counts on rural highways. He concluded that the duration and schedule of counting could vary significantly from one type of road to another and still achieve the same accuracy of counts. For example, a 4-hour afternoon (2:00 p.m. to 6:00 p.m.) count at a regional commuter site would yield approximately the same accuracy of counts as a 12-hr (7:00 a.m. to 7:00 p.m.) count at a long-distance and recreational site.

Sharma et al. *(7)* studied the design hourly volume concept and upgrading of two-lane rural highways from the perspective of driver population. It was concluded in the study that *(a)* the driver population is a significant variable that must be considered for appropriate sizing of roads from both economic and users' perspectives; *(b)* to provide a more uniform service to the users of various road facilities it would be more appropriate to use different highest volume hours for designing different types of roads (e.g., 15th to 20th highest hour for a recreational site to 50th highest hour for a regional commuter site); and *(c)* the AADT values at which typical two-lane rural roads would need upgrading can vary from 1,750 to 2,500 for highly recreational routes and from 6,500 to 8,500 for commuter routes.

Yet another example in which the understanding of roads according to driver population proved beneficial was the prediction of design hourly volume as specified by the 30th highest hourly volume. One commonly used model *(6)* in which the 30th hour volume *(Y)* is estimated as a function of AADT (e.g., $Y = 41.0 + 0.12$ AADT for Alberta highways) produces systematic errors of prediction *(25)*. It grossly underestimates the dependent variable for recreational and highly recreational routes and overestimates it for regional commuter and rural business and commuter facilities. As an alternative, Westermann *(25)* suggests a significantly improved method of prediction that is derived from a clear understanding of the road classes proposed in this paper. This model for Alberta highways is $Y = 45 + 0.087X$, where X is the average Sunday volume of traffic in July.

Driver Population Factor in the New HCM

The most interesting application of road classification according to traffic stream characteristics is its potential use in highway capacity analysis. The driver population factor (f_p) is an important adjustment factor that should be applied in capacity calculations. The new HCM *(4)* recommends a value of f_p equal to 1.0 for commuter sites and a range of f_p of between 0.75 and 0.90 for other types of traffic streams.

It is believed that the proposed classification of road sites can provide an excellent basis for making engineering judgments about the selection of a particular f_p-value from the range suggested by the new HCM. The drivers who commute every weekday for regular work-business trips can be assumed to be familiar with the subject facility and its environs. Weekend or tourist recreational trip makers are expected to be much less familiar with the regional or long-distance routes on which such trips are made. Thus, both the trip purpose and the trip length distribution would be expected to influence the selection of an appropriate f_p-value. The value of f_p to be selected can be assumed to decrease when trip purpose changes from commuter to recreational. It can also be assumed to decrease when the trip length distribution changes from urban to regional and long distance. Table 3 gives the suggested values of f_p, which it is hoped can be used as a guide in exercising engineering judgments in selecting a specific f_p-value. It may also be added

TABLE 3 RECOMMENDED CAPACITY ADJUSTMENT FACTOR FOR THE CHARACTER OF THE TRAFFIC STREAM

Traffic Stream Type	Factor (f_p)
Urban commuter	1.0[a]
Regional commuter	0.95
Regional recreational and commuter	0.90
Interregional	0.85
Long distance	0.85
Long distance and recreational	0.80
Highly recreational	0.75

[a]Value recommended by the new HCM.

here that the proposed classification can help in selecting sample highway segments to further study the effect of the driver population on highway capacity analysis.

SUMMARY AND CONCLUSIONS

In this paper a model, DRIPOP, for classification of roads according to traffic stream characteristics is proposed. The basic assumption made in the analysis for the DRIPOP model is that the differences in overall flow patterns observed at road sites result from different mixes of trip purpose and trip length characteristics.

The DRIPOP model includes the application of the standard computational technique of hierarchical grouping, and it also suggests the use of standard statistical methods of group comparisons. The use of such computational and statistical techniques helps to develop the master patterns of temporal volume variation, which are the main criteria of road classification according to driver population. Note that the model is intended to be used for classifying major highways that carry traffic volumes in excess of 1,000 AADT.

The proposed method of grouping road sites on the basis of temporal variations and road use characteristics is more objective and comprehensive than the conventional methods. It can enable highway agencies to group roads into distinct classes that are significantly different from each other. Also, the model is simple to apply and its data requirements can easily be satisfied by the regular types of traffic data collection programs undertaken by highway agencies. Road sites can be classified by seasonal traffic counts where traffic is counted a few times a year for periods of from 48 hr to several weeks in length.

The analysis and classification of count sites using the DRIPOP model can help to rationalize and economize the permanent traffic counting program. The DRIPOP model can be used for monitoring and reviewing count site classifications with respect to time. If, over a period of time, a particular road is suspected to have undergone a significant change in traffic flow characteristics, the DRIPOP model can help to reassign the count site to a proper class. Also the DRIPOP model can help to identify count sites that are not required or to identify areas where additional count sites are required. Seasonal and short-period traffic counts can also be rationalized with the understanding and application of this model.

In addition to its application for rationalizing the various traffic-counting programs, the road classification that results from the DRIPOP model could be an important criterion in many other highway planning and design functions. Some examples of such functions are *(a)* design hourly volume considerations, *(b)* highway capacity analysis, and *(c)* highway improvement programming.

Another important finding is that the DRIPOP model produces similar road classes for Alberta and Saskatchewan highways on the basis of temporal volume fluctuations and trip characteristics. The similarity of road classes for the two provinces has implications for a standard classification of provincial highways according to traffic stream characteristics.

ACKNOWLEDGMENTS

The authors are grateful to Alberta Transportation and Saskatchewan Highways and Transportation for providing the necessary data. The financial assistance of the Natural Sciences and Engineering Council of Canada is also appreciated. The observations and views presented in this paper are those of the authors and do not represent policies or opinions of the highway agencies.

REFERENCES

1. *Transportation and Traffic Engineering Handbook.* Institute of Transportation Engineers; Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.
2. *Road Classification, Rural and Urban—1964.* Technical Publication 26. Roads and Transportation Association of Canada, Ottawa, Ontario, Canada, 1965.
3. *Highway Statistics Project Committee Final Report.* Roads and Transportation Association of Canada, Ottawa, Ontario, Canada, 1982.
4. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
5. Reexamination of Design Hour Volume Concepts. *ITE Journal,* Vol. 49, No. 9, 1979, pp. 45–49.
6. M. J. Bailey, M. Eng, and P. Eng. Estimation of Design Hour Volume. *ITE Journal,* Vol. 51, No, 8, 1981, pp. 50–56.
7. S. C. Sharma, A. A. Tayebali, and A. Werner. "Cost-Effective and User-Oriented Sizing of Roads: Alberta Example." In *Transportation Research Record 1009,* TRB, National Research Council, Washington, D.C., 1985, pp. 15–23.
8. G. Phillips. *Accuracy of Annual Traffic Flow Estimates from Automatic Counts.* Supplementary Report 515. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
9. G. Phillips. When to Mount a Traffic Count. *Traffic Engineering and Control,* Vol. 10, No. 3, 1980, pp. 4–6.
10. I. Harrison, and M. Roberts. A Model for Predicting Seasonal Traffic. *Traffic Engineering and Control,* Vol. 22, No. 3, 1981, pp. 122–126.
11. G. Phillip, P. Blake, and D. Resson. *Estimation of Annual Flow from Short Period Traffic Counts.* Supplementary Report 802. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
12. G. Phillips, and D. Resson. *Representing the Distribution of Hourly Flows Throughout the Year.* Supplementary Report 804. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
13. *Guide for Traffic Volume Counting Manual.* Bureau of Public Roads, U.S. Department of Commerce, 1965.
14. "Discussion by Planning, Design and Traffic Group." Presented at Annual Conference of Western Association of Canadian Highway Officials, Regina, Sakatchewan, Canada, 1980.
15. S. C. Sharma, and A. Werner. Grouping of Permanent Traffic Counters and Classification of Rural Highways According to User Characteristics. *RTAC Forum Journal,* Road and Transportation Association of Canada, Vol. 5, No. 2, 1982.
16. P. H. Bellamy. *Seasonal Variation in Traffic Flows.* Supplementary Report 437. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1978.
17. M. Wohl, and B. Martin. *Traffic System Analysis for Engineers and Planners.* McGraw-Hill Book Company, New York, 1967.
18. D. J. Veldman. *Fortran Programming for Behavioral Sciences.* Holt, Rinehart and Winston, Inc., New York, 1967.
19. J. H. Ward. Hierarchical Grouping to Optimize an Objective Function. *American Statistical Association Journal,* Vol. 58, 1963, pp. 236–244.

20. S. C. 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.
21. H. Scheffe. *The Analysis of Variance*. John Wiley & Sons, Inc., New York, 1964.
22. B. B. Petroff. Experience in Application of Statistical Methods to Traffic Counting. *Public Roads*, Vol. 29, No. 5, 1956, pp. 110–117.
23. M. Morris. "Standard Deviation and Coefficient of Variation of Automatic Traffic Recorder Counts." *Proc. of the Highway Research Board*, HRB, National Research Council, Washington, D.C., Vol. 30, 1951, pp. 336–346.
24. S. C. Sharma. "Minimizing Cost of Manual Traffic Counts: Canadian Example." In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp.1–7.
25. H. Westermann. *Determination of Design Hour Volume from Seasonal Traffic Counts*. Faculty of Engineering, University of Regina, Regina, Saskatchewan, Canada, 1985.

Developing Link Performance Functions Using Highway Performance Monitoring System Data Files

HOW-MING SHIEH AND C. MICHAEL WALTON

The principal objective of this paper is to propose a practical approach to developing highway performance functions and estimating annual truck activities using Highway Performance Monitoring System (HPMS) data files. This approach was applied to the Interstate highways. A highway performance function was defined as the relationship of average travel speed (or average travel time, operating speed) versus volume-to-capacity (V/C) ratio. A preliminary analysis was initially performed to summarize collected HPMS data and to revise data inconsistencies in HPMS data records. The revised HPMS data records were further analyzed and used to test underlying assumptions of the proposed approach. Because of substantial variation of traffic flow patterns in different time periods, especially the difference between the nighttime period and peak-hour and off-peak-hour periods, it was proposed to average traffic conditions for developing link performance functions in a 16-hr period, including both peak-hour and off-peak-hour periods. An expansion factor was then used to provide an estimate of average daily traffic volume by vehicle type. The relationship of average travel speed versus V/C ratio by average highway speed and total number of through lanes, as reported in the 1985 Highway Capacity Manual, was used as a basis for developing specific highway performance functions. It was found that in the year 2000 Texas would have 10 billion vehicle miles traveled (VMT) by local and intercity truck traffic on the existing Interstate highways. It was also predicted that there would be 89 billion VMT of truck traffic on the Interstate highways in the United States in the year 2000.

The impetus for this research was the need to develop highway performance functions for the highway segments (links) on a national designated highway network to facilitate the operations of large combination vehicles. This research stems from a research project at the Center for Transportation Research, the University of Texas at Austin *(1)*. Hence, the principal objective of this paper is to propose a practical approach to developing highway performance functions and, for the purpose of providing a preliminary analysis of national truck activities on the designated network, estimating annual truck activities by using, in part, the Highway Performance Monitoring System (HPMS) data files *(2)*.

Department of Civil Engineering, The University of Texas at Austin, Austin, Tex. 78712.

The format of an HPMS data record includes variables needed to characterize the roadway segment: identification, operation, travel, geometry and configuration, traffic/capacity, environmental, and so forth. As indicated, each record characterizes one highway segment. Some variables, such as traffic and operations, may vary over time and some may not be uniform within an HPMS highway section. For instance, the geometric design of a highway section may be made up of several different highway grades and curvatures in the HPMS record. In this research the variables for highway infrastructure are assumed consistent over time; however, annual average daily traffic (AADT), operating speed, and volume-to-capacity (V/C) ratio vary over time.

In general, the average time taken for a driver to traverse a highway section depends on several factors. These factors may include, for example, section length, access control, longitudinal and vertical alignment, speed limits, traffic volume, traffic components, highway capacity, driver's behavior, vehicle characteristics, weather conditions, and traffic control devices. Most, if not all, of these factors may vary over time. In this research, the following factors were considered: traffic volume, traffic components, highway capacity, number of through lanes, topology, average highway speed, speed limit control, and time. These factors were considered important in deriving highway performance functions (i.e., the relationship of operating speed versus V/C ratio) using collected HPMS data. The basis for this derivation is the 1965 and the 1985 Highway Capacity Manuals (HCMs) *(3,4)*. The newly published 1985 Highway Capacity Manual updates speed-flow relationship by replacing operating speed with average travel speed as the dependent variable and takes into consideration the 55-mph speed limit. However, hourly traffic volume by vehicle type (i.e., for passenger cars and trucks) was not updated in the 1985 HCM.

Because of data inconsistencies, primarily resulting from missing data or coding errors, or both, it was necessary to verify and revise collected HPMS data. The procedures used to revise the data are presented in the next section. To reflect average level of service for each link, an investigation of hourly traffic flow patterns was conducted. The findings suggest that there were significant fluctuations in traffic flow at different time intervals, which led to an averaging of hourly traffic flow pattern in a 16-hr period *(3,5)*. Because the majority of highway segments (83 percent) that constitute the candidate network for the operation of large combination vehicle (LCVs) are Interstate highways, this average would facilitate an estimate of annual vehicle miles traveled (VMT) by truck traffic on the Interstate system. A discussion of the formulation of highway performance functions follows along with the final discussion of the conclusions.

PRELIMINARY ANALYSIS OF COLLECTED HPMS DATA FOR INTERSTATE HIGHWAYS

Until recently not all states submitted annually collected HPMS data to the FHWA. In the development of this study, HPMS data for the years 1978, 1980, and 1981 were used. Three states, Oklahoma, Mississippi, and Rhode Island, had not submitted HPMS data by the end of 1982. In summary, HPMS data for the states of Arkansas and South Carolina were collected in 1978; HPMS data for 21 states were collected in 1980; data for 5 states were collected in both 1980 and 1981; and data for 17 states were collected in 1981 (Figure 1).

For the Interstate highways, 8,441 HPMS records were collected for 45 states. Those states for which HPMS records were not collected were Mississippi, Oklahoma, Rhode Island, Hawaii, and Alaska (Table 1). To complete the required data for the development of highway performance functions for the Interstate highways of the 48 contiguous states, 742 HPMS records from neighboring states of Mississippi, Oklahoma, and Rhode Island were assigned by route number to the Interstate highways of these three states, respectively. The process used to locate HPMS data records on the highways of the candidate network was developed and published in a report by the Center for Transportation Research, the University of Texas at Austin *(1)*.

The items of an HPMS record used to develop highway perfor-

TABLE 1 TOTAL NUMBER OF COLLECTED HPMS RECORDS BY STATE FOR INTERSTATE HIGHWAYS

State	No. of Records	State	No. of Records
Alabama	123	Nebraska	69
Arizona	199	Nevada	135
Arkansas	81	New Hampshire	35
California	343	New Jersey	176
Colorado	234	New Mexico	141
Connecticut	182	New York	239
Delaware	29	North Carolina	85
Florida	317	North Dakota	101
Georgia	203	Ohio	349
Idaho	161	Oklahoma	345
Illinois	258	Oregon	167
Indiana	228	Pennsylvania	330
Iowa	557	Rhode Island	145
Kansas	209	South Carolina	149
Kentucky	262	South Dakota	101
Louisiana	127	Tennessee	222
Maine	103	Texas	281
Maryland	161	Utah	181
Massachusetts	244	Vermont	91
Michigan	313	Virginia	231
Minnesota	129	Washington	199
Mississippi	252	West Virginia	145
Missouri	155	Wisconsin	148
Montana	173	Wyoming	75

Note: Total number of records collected = 9,183.

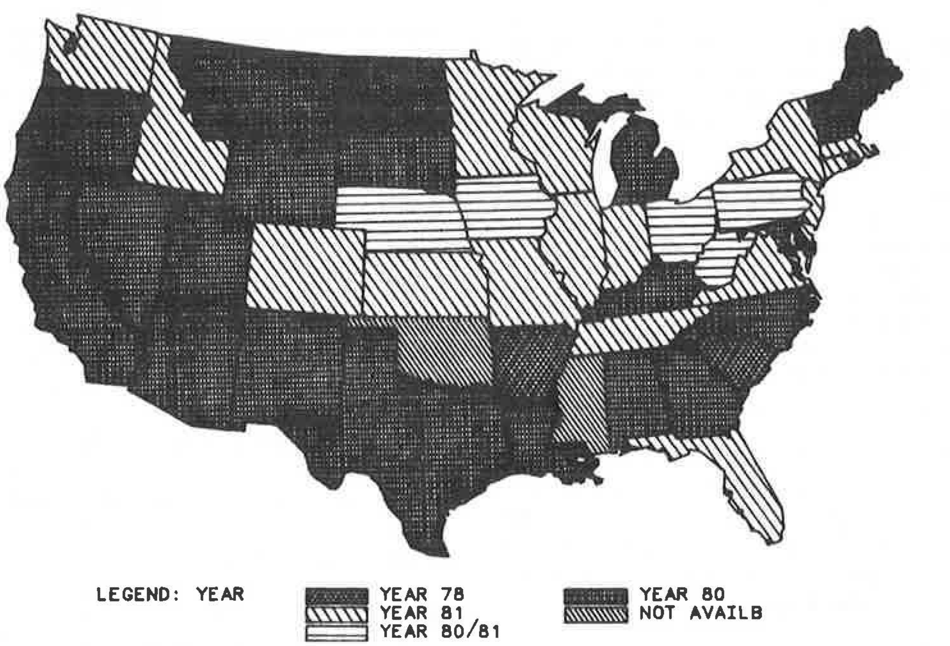


FIGURE 1 HPMS data collected by state and year.

mance functions include the following elements associated with the highway infrastructure, traffic condition, environment, and identification specification:

- Annual average daily traffic (AADT),
- Average highway speed (mph),
- County code,
- Estimated future AADT (year 2000 AADT),
- *K*-factor (the 30th highest hourly traffic volume that constitutes the AADT),
- Percentage of truck traffic in peak hours,
- Estimated operating speed (mph),
- Route number,
- Section length (in thousandth mile),
- Federal Information Processing Standards (FIPS) state code,
- Type of terrain,
- Number of through lanes,
- Computed V/C ratio (percentage),
- Feasibility of widening, and
- Year data were collected.

Some items of the HPMS records were found to be incomplete or inconsistent, primarily because of missing data or coding errors, or both. In general, the identification variables, including the year the data were collected, FIPS state code, and county code, were found to be complete. Other items required special treatment to complete or revise. Some items were found to be unreliable. For instance, some records indicated a speed limit of less than 20 mph or a *K*-factor of less than 5 percent of AADT for an Interstate highway section. Such inconsistencies were treated in the following manner:

- Items that were considered unrealistic (i.e., too low or too high) were eliminated. Default values were substituted. These default values could be either the averages or the upper bounds of corresponding items.
- Missing data were justified by other available items in the same record. For instance, some missing terrain data types were justified by highway grades and their corresponding lengths if applicable.
- Missing data files were computed by substituting the values given for the same items of neighboring observations.
- Because collected HPMS data on a coded highway link do not always match the distance measured from state maps, the length of an HPMS record was proportionally adjusted by this constraint.
- Remaining unjustifiable items were replaced with default values.

The revised HPMS data were then preliminarily analyzed for the year 2000, assuming that the Interstate highways would be adequately maintained to perpetuate existing road conditions and that some of the traffic-related characteristics would remain consistent over time. These traffic-related consistency items include *K*-factor and percentage of truck traffic in both peak-hour and off-peak-hour periods, assuming that current state regulations on truck size and weight limits and dispatching strategies of truck carriers remain essentially static or unchanged. The objective of this assumption is to provide a basis for assessing impacts resulting from the introduction of large combination vehicles on the candidate network. The items selected for preliminary analysis included 1977 and year 2000 AADTs, *K*-factor, percentage of truck traffic in peak hours, percentage of truck traffic in off-peak hours, terrain type, number of through lanes, speed limit, average highway speed, computed V/C ratio, and operating speed. The 1977 AADTs were obtained by extrapolating linearly the AADT of the year the HPMS record was collected and the year 2000 AADT supplied by each of the state highway departments. In general, each selected item was categorized in one of four groups on the basis of its quartiles. The first quartile (less than 25 percent) was categorized as the low-value group; the second and the third quartiles (between 25 percent and 75 percent) as the moderate-value group; and the other two groups as the high-value group (ranging from 75 percent to 95 percent) and the extremely high-value group (higher than 95 percent). Extreme values for each selected item were also discussed.

1977 AADT

On average, 25,000 vehicles per day per mile (vpdm) traveled on the Interstate highways. Of the overall Interstate highway system (41,560 mi), about 3 percent (1,229 mi) of total mileage accommodated more than 76,000 vehicles per day (vpd) (Figure 2). The following three highways were found to have vpd of more than 222,000 in some highway sections:

- Between the city of Chicago and loop I-90,
- I-405 between Los Angeles and Anaheim, and
- I-10 near Los Angeles.

California was found to have 283 mi of Interstate highways with extremely high daily traffic volume. Illinois, Texas, Ohio, and New Jersey had more than 70 mi of Interstate highway segments with heavy daily traffic volumes.

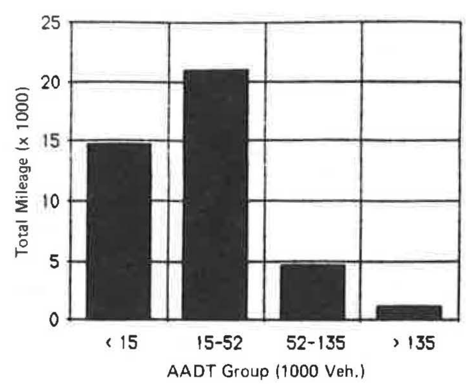


FIGURE 2 Total 1977 Interstate highway mileage by AADT.

Year 2000 AADT

The average daily vehicles per mile of the Interstate highways in the year 2000 was computed as 1.7 times that of 1977 (i.e., approximately 42,400 vpdm). Of the Interstate highways, 1,176 mi were predicted to have more than 135,000 vehicles per day (i.e., V/C ratio of about 90 percent in a 16-hr period for four-lane highways). These highly utilized Interstate highways were found distributed in Texas (203 mi), California (199 mi), and Florida (154 mi). Specifically, parts of three highways were predicted to have more than 300,000 vpd in the year 2000:

- I-95 between Miami and West Palm Beach;
- I-635 loop of Dallas; and
- I-5 between Everett, Washington, and Lynnwood, Washington.

Average Highway Speed

The 1985 HCM *(3)* defines average highway speed as "the weighted average of the design speeds within a highway section, when each subsection within the section is considered to have an individual speed." As expected, most of the HPMS records for the Interstate highways indicated an average highway speed of 70 mph. However, there existed 17 mi of Interstate highways with average highway speeds of less than 60 mph, and 897 mi with speeds between 60 and 69 mph. Kentucky, Montana, Pennsylvania, and Arizona were the four states that had average Interstate highway speeds of less than 60 mph.

K-Factor

It was estimated that the average *K*-factor for the Interstate highways was 12.79 percent based on the collected HPMS data. There were 355 mi of Interstate highways (0.9 percent) that had *K*-factors greater than 17 percent. Texas (200 mi), Kansas (67 mi), Illinois (59 mi), and California (21 mi) had very high *K*-factors on sections of their Interstate highways. Most Interstate highways (95.3 percent) had *K*-factors of less than 15 percent.

Percentage of Off-Peak Truck Traffic

For the Interstate highways, the average off-peak truck traffic was estimated to be 16.31 percent. It was found that about 30 percent of the Interstate highways (12,880 mi) had off-peak truck traffic greater than 24 percent (Figure 3). The following three sections of I-80 were found to have extreme values for this item:

- Between Pine Bluffs, Nebraska, and Big Springs, Nebraska, (46 percent);
- Between Wheatland, Pennsylvania, and Mercer, Pennsylvania, (45 percent); and
- Between Columbus, Ohio, and Toledo, Ohio, (45 percent).

Texas, Pennsylvania, Ohio, and Wyoming were found to have more than 300 mi of Interstate highways with a heavy truck traffic component (more than 32 percent). Heavy trucks made up more than 32 percent of the traffic stream on 240 mi of Interstate highways in California.

Percentage of Truck Traffic in Peak Hour

In general, the percentage of truck traffic on Interstate highway sections during peak hours was found to be less than that in off-peak hours. Maybe this is because of the significant increase of passenger car traffic volume. On average, peak-hour truck traffic for Interstate highways was 13.60 percent. The extremes of this item were similar to those found for off-peak hours.

Speed Limit Controlled

About 97 percent of Interstate highway mileage (40,327 mi) is controlled by a posted maximum speed limit of 55 mph; however, 578 mi of Interstate highways are controlled by a posted maximum speed limit of less than 40 mph. Most of these lower speed limit sections (384 mi) were in Florida. Texas and Maryland had 50 mi of Interstate highways controlled by a speed limit equal to or less than 40 mph.

Type of Terrain

After necessary verification and revision, it was estimated that 2,194 mi of Interstate highways can be classified as being in mountainous terrain. Most of these sections were located in Montana, West Virginia, Virginia, Pennsylvania, California, Maryland, Utah, Alabama, and Colorado.

Number of Through Lanes

For Interstate highways, 88 percent of total mileage was recorded as four-lane divided highways, 3,979 mi of Interstate highways were found to be six-lane divided highways, and 3.11 percent of all Interstate highways (1,294 mi) were constructed to be more than six lanes (Figure 4). Of those Interstate highways with more than six through lanes, more than half were located in California (663 mi). Texas, Maryland, and Georgia were found to have 96, 87, and

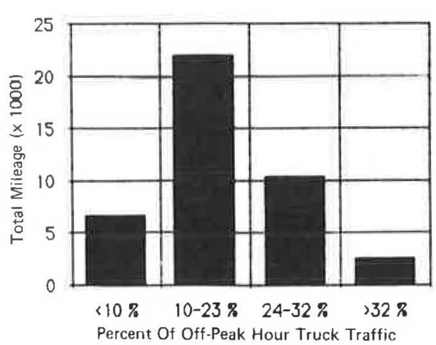


FIGURE 3 Total Interstate highway mileage by percentage of off-peak truck traffic.

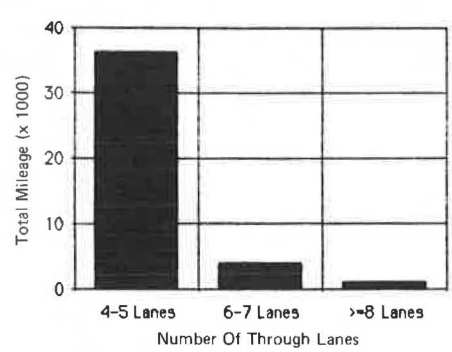


FIGURE 4 Total Interstate highway mileage by number of through lanes.

78 mi with more than six through lanes, respectively. Four locations were recorded as having more than 14 through lanes:

- I-94 between Wheeling, Wisconsin, and Wisconsin border (16 through lanes);
- I-5 between Anaheim, California, and San Diego, California, (15 through lanes);
- I-90 between Park Ridge, Illinois, and Chicago, Illinois, (15 through lanes); and
- I-5 between Everett, Washington, and Lynnwood, Washington, (15 through lanes).

V/C RATIO

V/C ratio and operating speed were computed by the FHWA using the HPMS data submitted by the states. Traffic during the peak-hour period was used to compute V/C ratios. Some of the ratios computed from HPMS records may be greater than 100 percent because a penalty was assigned to those records that had an assigned operating speed below the operating speed where the ratio was given as 1.0. For Interstate highways, 1,388 mi were classified as highly utilized during the peak-hour periods (i.e., V/C >1) (Figure 5). Most of these Interstate highways were located in California, Ohio, and Texas (175, 149, and 141 mi, respectively).

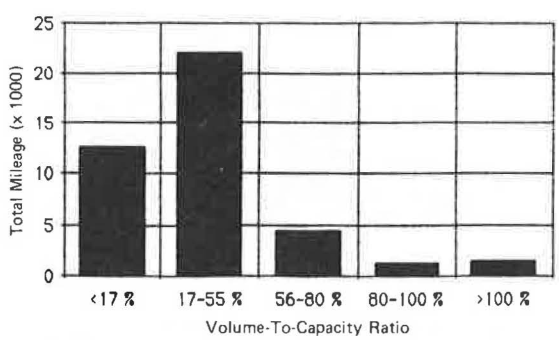


FIGURE 5 Total Interstate highway mileage by V/C Ratio.

The extreme values for the V/C ratios were found to be as high as 437 percent. Some of these extreme values include part of the following four Interstate highways:

- I-95 between Rye, New York, and New York City (437 percent of V/C);
- I-495 between New York City and Harrison, New York, (350 percent of V/C);
- I-95 between Byran, New York, and Rye, New York, (326 percent of V/C); and
- I-87 between Champlain, New York, and Albany, New York, (306 percent of V/C).

Operating Speed

Operating speed was divided into five groups: below 20 mph, 21 to 30 mph, 31 to 45 mph, 46 to 55 mph, and more than 55 mph. About 80 percent of Interstate highway mileage was found to be at operating speeds greater than 55 mph; however, there were 1,363 mi of Interstate highways that had operating speeds of less than 30

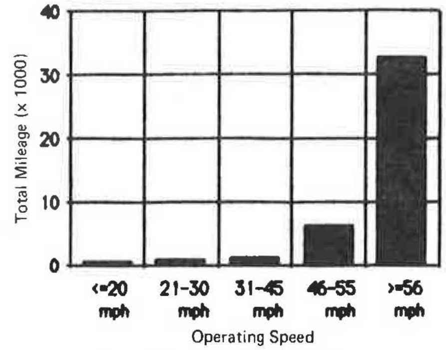


FIGURE 6 Total Interstate highway mileage by operating speed.

mph (Figure 6). Most of these sections were located in California, Ohio, Texas, and Pennsylvania (175, 145, 140, and 98 mi, respectively). California and Texas had more than 100 mi of Interstate highways with operating speeds of less than 20 mph (109 and 105 mi, respectively).

DEFAULT VALUES AND AVERAGE TRAFFIC CONDITIONS

The preliminary analysis provided insight into the steps required to correct data inconsistencies, particularly time dimensions of the data. The percentage of truck traffic was reported on an hourly basis for both peak-hour and off-peak-hour periods; and the AADT was expressed as a daily equivalent. The *K*-factor represents the 30th highest hourly traffic volume in a year. To cope with these discrepancies in time units and to obtain the average of traffic conditions corresponding to truck traffic in various time periods, default values were computed based on the 1965 HCM *(3)* and the 1975–1979 National Truck Characteristic Report *(5)*. Two assumptions were made to facilitate the analysis:

- The ratio of the *K*-factor and average peak-hour factor remains constant for reported HPMS data records and
- Computed default values are transferable and remain stable over time.

As illustrated in Figure 7, the 30th highest hourly traffic volumes in a year (*K*-factor) for rural expresssways and urban expressways were 13 percent for the 1985 HCM (or 14 percent for the 1965 HCM) and 11 percent of AADT, respectively. The computed average *K*-factor was 12.8 percent of AADT for Interstate highways. This computed average *K*-factor was comparable to that reported in the Highway Capacity Manuals; therefore, the data used to illustrate the relationship of hourly traffic volumes and AADT in the 1965 HCM were considered transferable.

Some default values were derived from two diagrams reported in the 1965 HCM *(3)* (Figures 8 and 9). The data used to draw the two diagrams were collected at 49 rural stations in Wisconsin's truck highway system in 1961. Reported hourly variations of traffic on rural highways for an average weekday (Figure 8) provided insight into three time periods that could be identified on the basis of percentage of ADT in the peak-hour period (2 p.m. to 6 p.m.), off-peak-hour period (6 a.m. to 2 p.m.), and nighttime period (10 p.m. to 6 a.m.). It was found that traffic volumes varied significantly in different time periods for the two simplified vehicle

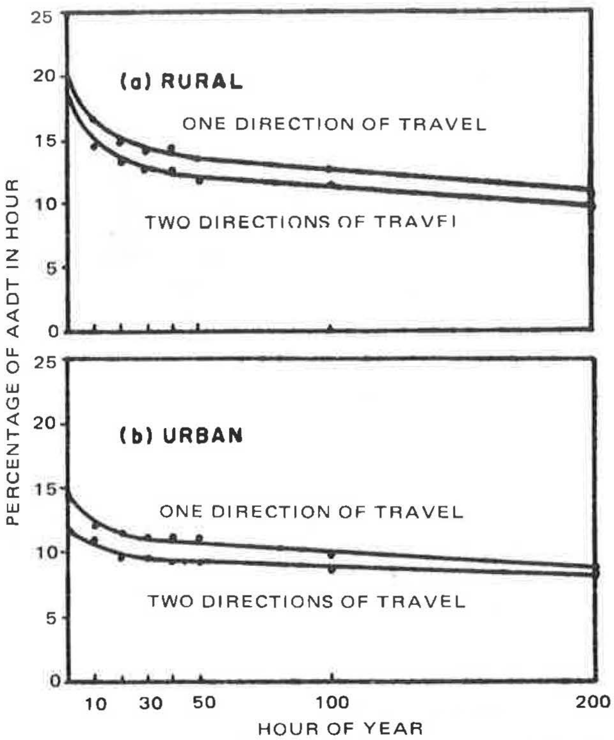


FIGURE 7 Relation of hourly traffic volume and AADT on expressways.

classes, passenger cars and trucks (Figures 10 and 11). The total number of passenger vehicles per hour during the peak-hour period was approximately eight times that observed during the nighttime period. Even during the off-peak period, the total number of vehicles per hour was about five times that of the nighttime period. Therefore, it is assumed that, on average, traffic congestion does not occur during the nighttime period. Averaging the traffic conditions of peak-hour and off-peak-hour periods over a 16-hr rather than a 24-hr period was done to expedite traffic assignment on a daily basis. Two factors were used to expand traffic volume from a 16-hr to a 24-hr period. The two expansion factors were computed as 113 percent and 133 percent of AADT for passenger cars and trucks, respectively.

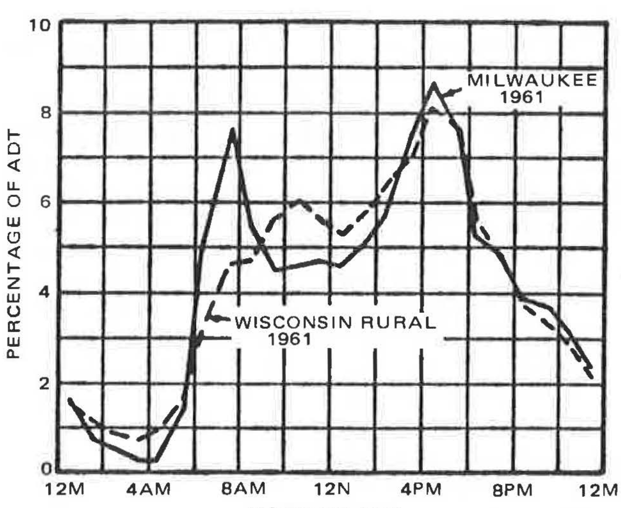


FIGURE 8 Hourly variations of traffic for average weekday.

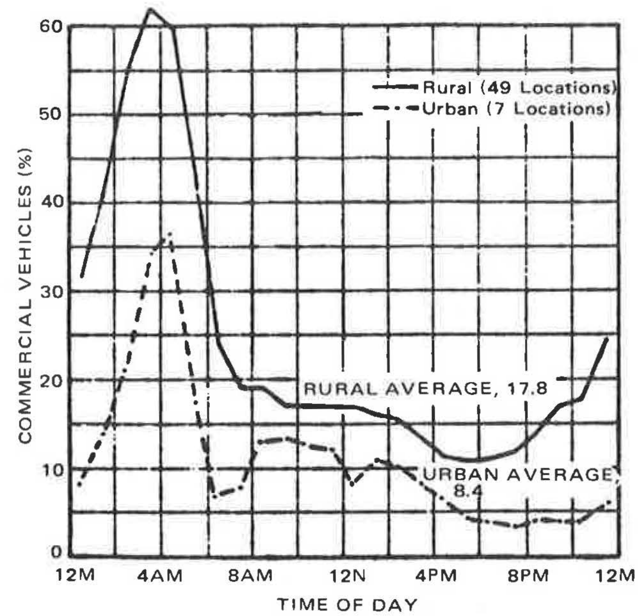


FIGURE 9 Traffic composition by time of day, Wisconsin highways, weekdays, 1961.

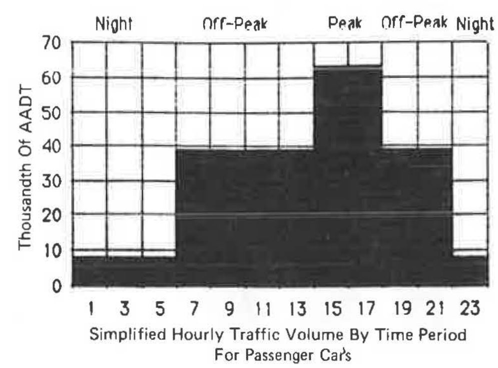


FIGURE 10 Representative hourly traffic volume by time period for passenger cars.

Referring to Figure 8 and the expansion factors just discussed, the total number of vehicles per hour during the peak hours was estimated as 7.25 percent of AADT. Combined with a default *K*-factor (13 percent of AADT) and the assumption that the ratio of *K*-factor and average peak-hour factor remain constant, the total number of vehicles per hour during the peak-hour period (7.25

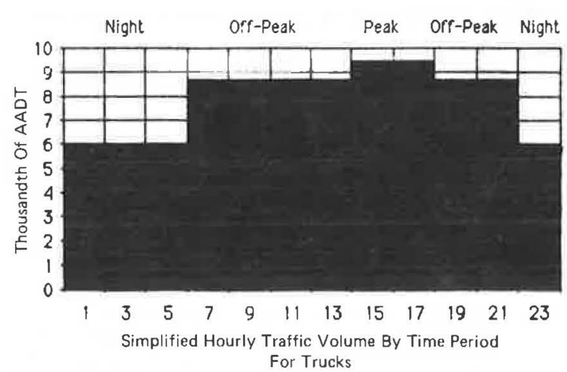


FIGURE 11 Representative hourly traffic volume by time period for trucks.

percent of AADT) was used to find the average peak-hour factor *(PF)* and average off-peak-hour factor *(OPF)* for a given highway section. Equations 1 and 2 illustrate this.

$$PF = 0.5578\, K \quad (1)$$

$$OPF = 0.07378 - (PF/3) \quad (2)$$

where

K = K-factor of a highway section,
PF = average number of vehicles in peak-hour period in terms of AADT, and
OPF = average number of vehicles in off-peak-hour period in terms of AADT.

From these two equations, total annual VMT for trucks (including local and intercity truck traffic) operating on Interstate highways were estimated to be 89 billion in the year 2000. Of these, most were predicted to occur in Texas, Florida, Pennsylvania, California, and Indiana (10.1, 6.1, 5.0, 4.9, and 4.8 billion VMT, respectively).

On a linkwise basis, 76 percent of Interstate highway mileage (31,694 mi) was estimated to have truck traffic of less than 8,000 vpd; 20 percent was between 8,000 and 17,000 vehicles; and 1,538 mi of Interstate highways (3.7 percent) were predicted to have more than 17,000 trucks per day (Figure 12). Florida, Indiana, Texas, and North Carolina were predicted to have more than 100 mi of Interstate highways with daily truck traffic of more than 17,000 vehicles (362, 250, 183, and 132 mi, respectively).

DEVELOPMENT OF HIGHWAY PERFORMANCE FUNCTIONS

A highway performance function was formulated as a function of traffic volume *(V)*, traffic component *(P)*, highway capacity *(C)*, number of through lanes *(N)*, topology *(R)*, average highway speed *(H)*, and speed limit controlled *(L)* at specific time period *(T)*:

$$v \text{ (or } t) = f\,(V, P, C, N, R, H, L, T) \quad (3)$$

The dependent variable could be either average travel speed (or operating speed) *(v)* or average travel time per mile *(t)*. The time period employed was the 16-hr period discussed previously. Highway capacity *(C)* was defined as total number of passenger car equivalents (PCEs) of ideal capacity. For multilane, divided, fully accessed-controlled highways, 2,000 PCEs per lane was used. The categories of topology were simplied to be level, rolling, and mountainous terrain for extended highways. The traffic component was classified into two classes of vehicles, passenger cars and trucks.

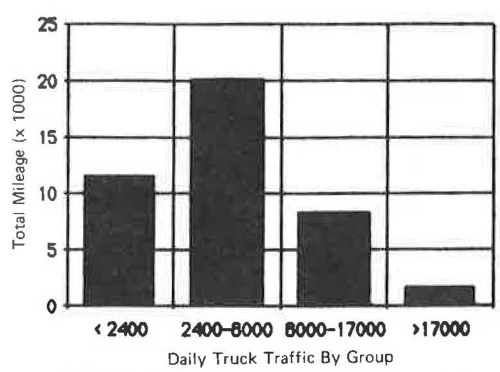


FIGURE 12 Total highway mileage by daily truck traffic volume group for Interstate highways in the year 2000.

The 1985 Highway Capacity Manual reported the relationship between average travel speed (or operating speed in 1965 HCM, see Figure 13) and V/C ratio categorized by total number of through lanes and average highway speed (Figure 14). On the basis of this relationship, a highway performance function could be formulated:

$$v(\text{or } t) = f\,(V, P, C, R \mid H, T, N) \quad (4)$$

subject to v less than L.

The constraint used in Equation 4 has to be replaced by v (operating speed), which is 5 less than L when the speed-flow relationship in Figure 13 is employed. In this formulation, it was assumed that the operating speed could be 5 mph higher than the controlled speed limit. The constraint used in Equation 4 has to be replaced by v (average travel speed) less than L when the speed-flow relationship in Figure 14 is employed. To compute the V/C ratio, it was necessary to convert traffic volume into PCEs. This conversion was the mutual effect of total number of vehicles, traffic component, and topology for extended highways and particular highway sections *(3,4,6,7)*. The output from this conversion was the total number of passenger car equivalents for a highway section or an extended highway (V_{pce}). Equation 4, therefore, could be further simplified to

$$v \text{ (or } t) = f\,(V_{pce}/C \mid V, P, C, N, R, H, T) \quad (5)$$

subject to v less than L (or t greater than $60/L$).

The conversion of traffic volume to PCEs, however, did not consider the effects of vehicle width, vehicle weight, vehicle length, and the like. Therefore the introduction of larger and heavier trucks into the traffic mix of a highway section may change the total number of converted PCEs. Vehicle length was proposed for modifying PCE numbers for large combination vehicles on an extended highway section *(1)*.

From the formulation of highway performance functions and conversion of PCEs, each HPMS highway section could be characterized. This approach could therefore reflect performance characteristics of a highway section more accurately than the approach that calibrated a unique performance function for a global network *(8,9)*.

The relationship of average travel time (or operating speed), travel time per mile, and V/C ratio could be approximated either by a mathematical formulation or by a set of data points. For instance, mathematical formulation of the highway performance functions for four-lane freeways or expressways designed with an average highway speed of 70 mph could be expressed by the speed-flow relationship based on the 1965 HCM:

$$v \text{ (operating speed)} = 25 + [1470 - 14.7\,(V_{pce}/C)]^{1/2} \quad (6)$$

subject to (V_{pce}/C) in percent less than 100 percent and $v - 5$ less than L, or by

$$t = 1.0154 - 0.0052\,(V_{pce}/C) + 0.00015\,(V_{pce}/C)^2 \quad (7)$$

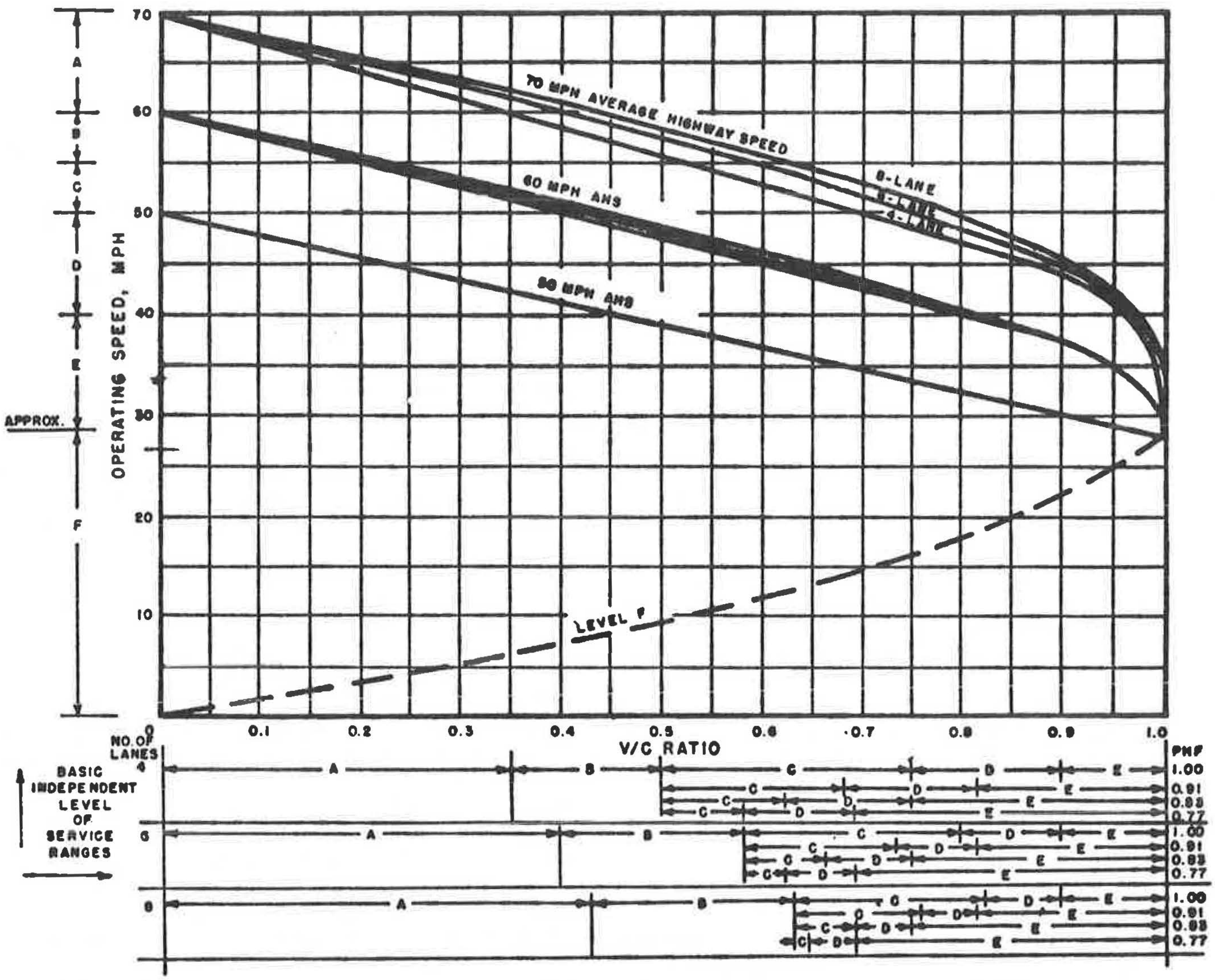


FIGURE 13 Relationships between V/C ratio and operating speed in one direction of travel on freeways and expressways under uninterrupted flow conditions.

subject to t greater than $60/L$ where L is in mph and t is in minutes per mile.

This formulation could also be expressed by the speed-flow relationship based on the 1985 HCM:

$$V_{pce}/C = 1{,}342.71 - 92.15\ v + 2.23\ v^2 - 0.0178\ v^3 \quad (8)$$

subject to v less than L where v is average travel speed in mph, or by

$$t = 0.9514 + 0.01307\ (V_{pce}/C) - 0.00041\ (V_{pce}/C)^2 + 0.0000038\ (V_{pce}/C)^3 \quad (9)$$

The relationship of average travel time versus V/C ratio by number of through lanes and average highway speed, based on the 1985 HCM, is given in Table 2.

All of the formulations were restricted to (V_{pce}/C) not greater than 100 percent. However, for the purpose of an all-or-nothing traffic assignment in an iterative process, the (V_{pce}/C) ratio could be greater than 100 percent. Penalty functions, therefore, were introduced in cases in which the (V_{pce}/C) ratio was greater than 100 percent.

SUMMARY AND CONCLUSIONS

In this paper, collected HPMS data for the Interstate highways were verified, revised where deemed appropriate, and preliminarily analyzed. Some default values were derived on the basis of the investigation of representative hourly traffic flow patterns for both passenger cars and trucks reported in the 1965 and 1985 Highway Capacity Manuals. It was proposed to average traffic conditions of peak-hour and off-peak-hour periods into a 16-hr period. The proposed approach was applied to estimate total annual VMT by truck traffic in the year 2000. It was estimated that there would be 89 billion VMT on the existing Interstate highways in the year 2000. The approach proposed to assess highway performance functions on a section-by-section basis using the HPMS

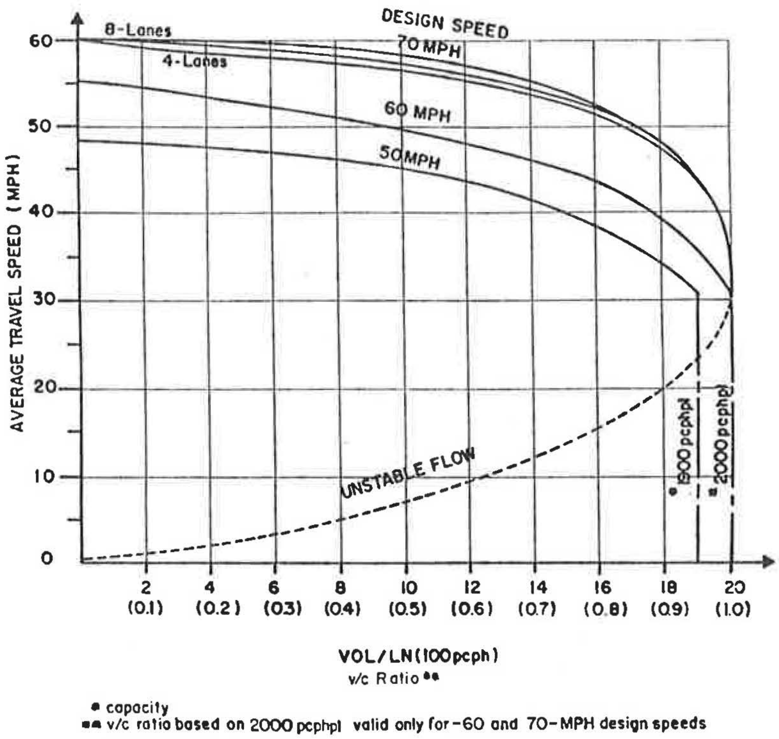


FIGURE 14 Average travel speed-flow relationships under ideal conditions for freeways.

TABLE 2 RELATIONSHIP OF AVERAGE TRAVEL TIME VERSUS *V/C* RATIO BY NUMBER OF THROUGH LANES AND AVERAGE HIGHWAY SPEED

Category	Relationship of Average Travel Time Versus *V/C* [*t* in minutes per mile and (V_{pce}/C) in percent]
70 mph, four lane	$t = 0.95140 + 0.01307\ (V_{pce}/C) - 0.00041\ (V_{pce}/C)^2 + 0.00000348\ (V_{pce}/C)^3$
70 mph, six lane	$t = 0.94228 + 0.01379\ (V_{pce}/C) - 0.00045\ (V_{pce}/C)^2 + 0.00000384\ (V_{pce}/C)^3$
70 mph, eight lane	$t = 0.94501 + 0.01312\ (V_{pce}/C) - 0.00044\ (V_{pce}/C)^2 + 0.00000387\ (V_{pce}/C)^3$
60 mph	$t = 1.05589 + 0.00938\ (V_{pce}/C) - 0.00027\ (V_{pce}/C)^2 + 0.00000253\ (V_{pce}/C)^3$
50 mph	$t = 1.2274 + 0.00481\ (V_{pce}/C) - 0.00016\ (V_{pce}/C)^2 + 0.00000175\ (V_{pce}/C)^3$

data was based on the relationship of average travel time (or average travel speed) versus V/C ratio reported in the 1985 Highway Capacity Manual (or the relationship of operating speed versus V/C ratio reported in the 1965 HCM). It is hoped that the proposed approach would foster the understanding and use of developing highway performance functions from collected HPMS data.

The following points are recommended for further improvement and research:

1. A reliable data base is the foundation for generating plausible results. It is recommended that the quality of HPMS data be improved. Missing data or coding errors, or both, in HPMS data records should be avoided. A uniform sample and identification scheme and systematic arrangement of data records are important for applying collected HPMS data to practice.
2. Understanding of the interaction between various vehicle classes and different highway infrastructure, traffic conditions, and the like has not been well developed. The introduction of large combination vehicles (Rocky Mountain double trailers, turnpike double trailers, and triple trailers) makes this interaction more complicated. An understanding of this interaction would provide a basis for reliable assessment of highway performance.
3. Most traffic assignment algorithms use an iterative process to assign traffic by an all-or-nothing method. The justification of a penalty function for highway performance on overcapacity highways may be worthy of further research.

REFERENCES

1. C. M. Walton et al. *Feasibility of a National Intercity Truck Route Network.* Center for Transportation Research, the University of Texas at Austin, 1986.
2. *Highway Performance Monitoring System Field Manual.* FHWA, U.S. Department of Transportation, Sept. 1980.
3. *Special Report 87: 1965 Highway Capacity Manual.* HRB, National Research Council, Washington, D.C., 1965, 397 pp.
4. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.

5. *1975–1979 National Truck Characteristics Report*. Highway Statistics Division, FHWA, U.S. Department of Transportation, June 1981.
6. *Transportation Research Circular 212: Interim Materials on Highway Capacity*. TRB, National Research Council, Washington, D.C., Jan. 1980, 272 pp.
7. R. Roess et al. *Passenger Car Equivalents for Uninterrupted Flow: Revision of the Circular 212 Values*. Transportation Training and Research Center, Polytechnic Institute of New York, Brooklyn, Nov. 1983.
8. K.B. Davidson. "A Flow Travel Time Relationship for Use in Transportation Planning". *Proc., Australian Road Research Board*, Melbourne, Vol. 3, 1966, pp. 183–194.
9. D. Branston. Link Capacity Function: A Review. *Transportion Research*, Vol. 10, 1976, pp. 223–230.