DETERMINATION OF FUNCTIONAL SUBREGIONS WITHIN AN URBAN AREA FOR TRANSPORTATION PLANNING

Thomas F. Golob and Salvatore J. Hepper,
Transportation and Urban Analysis Department,
General Motors Research Laboratories; and
John J. Pershing, Jr.*, Fairfax County Planning Commission

It is axiomatic that large urban areas are not spatially homogeneous with respect to transportation demand, supply, and impact phenomena. This paper addresses this heterogeneity in terms of the transportation planning process. A technique for using areawide travel, land use, and population data to divide an urban area into a set of functional subregions is presented. Each subregion represents a planning area, and interregion planning is proposed on a different scale of analysis. The technique is based on the statistical decomposition of origin-destination flow matrices. The decomposition method can be considered a generalized type of factor analysis in which raw data observations are used as opposed to variable correlations. The units can be any spatial aggregation of people and activities, such as census tracts or minor civil divisions, and the travel can be trips for any specific purpose or a composite of all trips. Selection in both cases depends on the objectives of the planning process. Multiple discriminant and regression analyses are then used to define the subregions in terms of differences in population and land use characteristics. Results from an application of the technique in the Detroit area are presented as a case study. Six subregions, composed of groups of minor civil divisions and central city subcommunities, were found and successfully described for home-based work travel in this urban area. The results support urban economic theories of a central city core area, suburban industrial centers, and satellite cities.

•THIS PAPER presents a specific transportation planning technique as a possible addition to the set of planning tools available to decision-makers. This technique breaks down a metropolitan area or reasonably extensive planning area into functional subregions on the basis of people's trip-making behavior. The technique can be categorized as an inductive search for hypotheses of transportation and urban form through the identification of regularities in spatially aggregated data.

Analyzing urban transportation needs and evaluating alternatives on a scale smaller than an entire urban area are addressed. Studies of activity center distributions, feeder transit services (e.g., involving dial-a-bus alternatives), and local and collector roadway systems are just a few cases in which analysis could be accomplished on a subregional basis with appropriate aggregated subregional interactions. For example, in the case of the home-based work trip, a transportation system must be able to deal with the travel needs of persons who work outside of the central business district. This paper indicates how suburban areas can be organized into meaningful subsystems for transportation analysis and planning.

The technique is based on multivariate analyses of transportation flows and related data on the characteristics of the origin and destination spatial units. The analyses

*Mr. Pershing was with General Motors Research Laboratories when this work was developed.

are all well documented in theoretical textbooks, and computer programs are readily available in standard statistical packages.

STATISTICAL METHODS

The multivariate statistical method of singular decomposition is used to determine structural similarities in a data matrix comprised of travel variables. The data are organized in the form of origin-destination (O-D) matrices used in traditional transportation planning processes. Each cell of the matrix represents a measurement of the flow of people from an origin to a destination for a common travel purpose. In the terminology of traditional multivariate data matrices, each origin spatial unit is treated as a variable and each destination spatial unit is treated as an observation. For flow matrices the metric for the variables is the same for each variable; therefore it is appropriate to decompose the data without applying standardizing transformations, such as subtracting out the mean for each variable (column in the matrix) and dividing each variable by the standard deviation. This transformation characterizes the factor analysis version of singular decomposition (2). Such standardizations result in a loss of information (both the level and dispersion of the variables in the case of factor analysis), whereas singular decomposition preserves the level of information present in the raw data.

The Eckart-Young theorem (1) and Nash's extension (4) can be used to decompose an $n \times m$ data matrix, X, into three matrices such that

$$X \cong P \Lambda Q' \tag{1}$$

where

P = $n \times r$ matrix of orthonormal (independent and normalized) vectors,
Q = $m \times r$ matrix of orthonormal (independent and normalized) vectors,
Λ = $r \times r$ diagonal matrix of eigenvalues (latent roots of X), and
r = number of vectors extracted ($r \leq m \leq n$).

In flow matrices, $m = n$. In the terminology of traditional factor analysis, it is convenient to let $L = (Q\Lambda)$ represent the factor loading matrix and to treat the associated vectors of L as that linear combination of the original n variables (origin spatial units) that describes each new factor. Similarly, the vectors of P represent the factor scores or evaluations of the n observations (destination spatial units) on the new latent factors.

The number of independent latent factors extracted is determined by subjective evaluation of the associated eigenvalues Λ, the first differences in eigenvalues, and the cumulative fraction of trace accounted for by each factor or eigenvalue. The fraction of trace of the cross-product ($X'X$) matrix is analogous to the percentage of variance in factor analysis, and it represents the importance of a factor in describing the information in the original data. The trade-off addressed entails a sufficient explanation of the original data (in cross-product terms) within the minimum number of factors possible.

After the number of latent factors to be extracted is determined, a rotation to simple structure is achieved by application of the successive factor varimax rotation (3).

After rotation the latent factors can be expressed in terms of the original origin variables by identifying the origin spatial units most strongly related to each factor. The strength of this relationship is proportional to the absolute value of the j th term in the i th column of the rotated factor loadings matrix. Each origin spatial unit is then assigned to the factor to which it is most strongly related. The factors are then interpreted as functional subregions with relatively homogeneous travel patterns, each subregion being delineated by the origins that are assigned to that factor.

The next phase of regional decomposition is to relate the socioeconomic characteristics (referred to as "state" variables) to the subregions and associated latent origin factors. First, a correlation analysis is employed by forming an $[(r + s) \times n]$ data matrix, z, where r = factor loading vectors, s = socioeconomic characteristic vectors, and n = origin spatial units. All correlation coefficients that are significantly different from zero at the $\alpha = 0.05$ level of significance are explored. Correlations between r and s can be used to broaden the interpretation of the factors and thus to provide a better understanding of the differences between the subregions. Such understanding can be of use in initiating detailed planning analysis for specific transportation needs.

A logical extension of the correlation analysis is the application of linear regression, in which each factor is treated as a dependent variable and the relevant socioeconomic characteristics determined from the correlation analysis are treated as the independent or explanatory variables. However, to use this approach as anything other than a broad indicator of meaningful interrelationship requires that a variable or a linear combination of variables be found that maps the socioeconomic characteristics (variables) into each factor loading of the flow matrix.

A third and more satisfying test is consequently performed to determine whether the groupings can be significantly differentiated from each other on the basis of their mean value measurements on each state variable. The multivariate analysis of variance is used to determine the significance of the overall difference among several group means on a single variable by performing an F-test on the Mahalonobis D-statistic as suggested in Tatsuoka (5). A group mean that is significantly different from other group means of a particular state variable is identified at the $\alpha = 0.01$ level of significance. A summary of these results can yield penetrating descriptions of each mutually exclusive group of origins.

The results presented later are based on the difference between groups of origin spatial units with respect to a single state variable descriptor, whereas correlation analysis and linear regression are based on the similarities of the origin spatial units as evaluated on a latent factor and one or more state variables.

A CASE STUDY APPLICATION

Data

Investigation of the interrelations between transportation behavior and urban form was based on a case study of the Detroit, Michigan, urbanized area as defined by the Bureau of the Census for the base year, 1965. The data used in this analysis were collected in 1965 during the conduct of the Detroit Region Transportation and Land Use Study (TALUS). The Southeastern Michigan Council of Governments (SEMCOG) generously provided these data to the Transportation and Urban Analysis Department of the General Motors Research Laboratories in support of research into the costs and benefits of proposed new systems of public transportation and the development of improved methods of urban transportation planning and evaluation. Two types of data have been used: home-to-work trip travel flows, which are used to determine travel patterns, and state variables, which are used to explain the socioeconomic and demographic structure of a spatially defined area.

The data were supplied in the form of observations of 1,278 transportation analysis zones as defined by TALUS within the five-county Detroit region. These regions were converted to observations on 59 central city subcommunities (CCSs) and 61 minor civil divisions (MCDs) within the urbanized area. These areas represent portions of Wayne, Oakland, and Macomb Counties. Inasmuch as each MCD was made up of one or more analysis zones, the conversion entailed aggregation by code sequences of all zones within the urbanized area. The MCDs were defined consistently with the MCDs of the Bureau of the Census, which in Michigan are incorporated municipalities or townships. Each CCS follows boundaries determined by a local census committee and delineates relatively homogeneous neighborhoods of approximately the same population.

The flow variables represent home-to-work trip movements of people from an MCD or CCS to every MCD or CCS within the urbanized area. The focus on home-based work travel is for illustrative purposes only; the technique is considered relevant for a variety of transportation issues. Gross trip flows were divided by the total number of households at each origin; the resultant measure is the number of work trips per household made to each destination. This statistic can also be considered as the probability that a household at j will make a work trip to i. The state variables consist of socioeconomic characteristics of the population by place or residence, employment characteristics by place of work, and land use characteristics of each MCD or CCS. The socioeconomic characteristics that describe the spatial units are

Median family income,
Percentage of families with incomes under $3,000,
Percentage of families with incomes over $15,000,
Percentage of households with youngest over 18,
Households with residence less than 1 year,
Percentage of households owning no car,
Percentage of households owning two cars or more,
Percentage of renter-occupied dwellings,
Households per residential acre,
Persons per occupied dwelling unit,
Percentage of household heads with less than 12 years' education,
Percentage of female household heads,
Percentage of skilled or unskilled household heads,
Percentage of nonwhite population,
Percentage of population under 20 years of age, and
Percentage of population over 64 years of age.

The types of employment included in the employment characteristics are

Professional and related,
Public administration,
Service,
Finance, insurance, and real estate,
Retail trade,
Wholesale trade,
Transportation, communication, and utilities, and
Manufacturing.

Land use characteristics of interest are the percentages of commercial, public, industrial, and recreational land.

Subregion Determination

The original 120 × 120 matrix of home-based work travel per household was reduced to seven principal components that accounted for 76.40 percent of the trace in the data cross-product matrix. This number of independent factors was determined subjectively as described earlier. A graph of the cumulative percentage of trace accounted for by each level of reduction is shown in Figure 1 for up to 15 factors; doubling the number of factors selected would have accounted for only an additional 10 percent of the original variance.

By grouping each of the 120 MCD or CCS origins according to the factor on which it loaded most highly, we delineated six functional subregions of homogeneous travel patterns. The seventh factor was found to be much weaker and only represented an orientation slightly different from the others. Dimensionality was thus reduced in making the transition from factors to subregions, which is an effective check against accepting too many factors through the extraction cutoff procedure. The six subregions are shown in Figure 2. The factors are discussed in order of the fraction of trace (percentage of original variance) accounted for after rotation.

The first factor accounted for 22.1 percent of the original trace, and the subregion determined by the factor includes almost all of the central city of Detroit, as well as adjoining suburban areas along major radial transportation corridors leading into the CBD. (The subregion is labeled central in Fig. 2.) The CBD (MCD 8 in Fig. 2) is the destination with the highest score on this factor. This suggests that central city residents tend to work within the city and that there is much less out-commuting to suburban employment centers than there is in-commuting to the central city. The four major transportation corridors of in-commuting in 1965 that can be identified with this

Figure 1. Analysis of total home-based work origins.

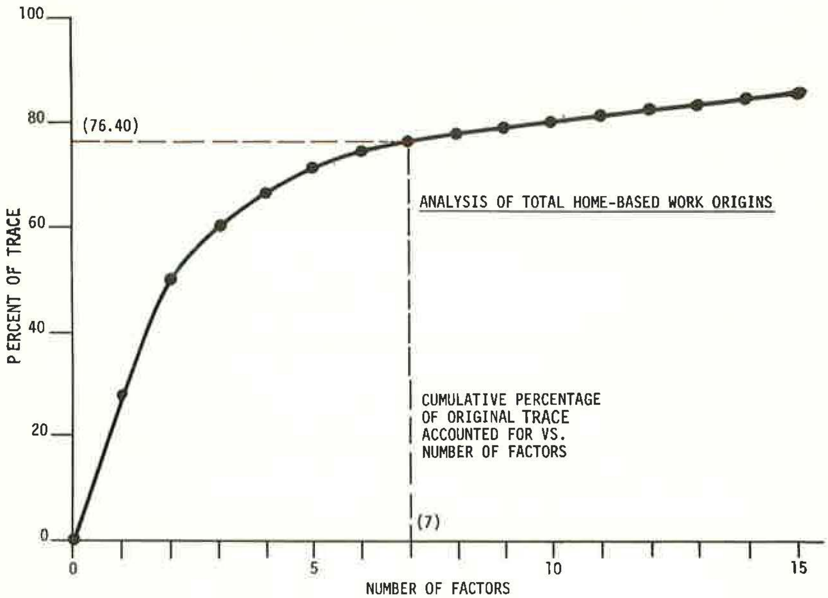
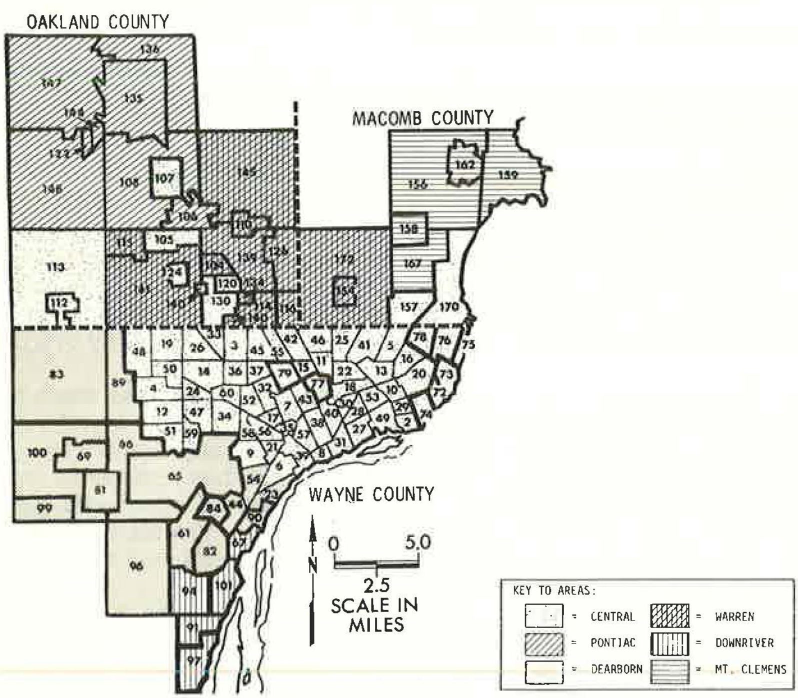


Figure 2. Groupings of MCDs by highest loadings (1965 Detroit urbanized area).



subregion were I-94/I-75 to the southwest, I-96/Grand River Avenue to the west northeast, I-75/Woodward Avenue to the north northwest, and I-94/Gratiot Avenue to the northeast.

The second factor accounted for 21.0 percent of the trace and determines the northwest portion of the urbanized area. The major destination is overwhelmingly Pontiac (MCD 135). The adjacent areas are primarily bedroom communities surrounding this industrial satellite city. The Pontiac subregion seems to be spatially separated from the central city.

Factor three accounted for 12.3 percent of the trace, and its subregion includes the southwest portion of Wayne County. Dearborn (MCD 65) is the major attractor, and Livonia (MCD 83), Ecorse (MCD 67), and Fort Wayne (MCD 23) are of secondary importance as employment attractors. Three adjacent central city subcommunities (23, 54, and 59) load highly on this factor, suggesting that the Dearborn industrial area is a more important employment center for residents of these areas than are other central city opportunities.

The fourth factor, accounting for 7.8 percent of the trace, centers on Warren (MCD 172), another major suburban industrial community. Royal Oak (139) is almost as important an attractor, whereas Ferndale (114) and Southfield (141) are somewhat lesser attractions. Again, out-commuting from the city of Detroit is not a primary characteristic in the formation of the subregion. Factor seven, which accounted for 3.6 percent of the trace, also centered on Warren but was oriented toward the east and south and did include Detroit communities; but this factor was insufficiently strong (in relation to the other six factors) to form a subregion via the origin factor loadings method used here.

Factor five accounted for only 4.9 percent of the trace, and its subregion is composed of the heavy industrial areas downriver from Detroit. Trenton (97), Ecorse (67), and Wyandotte (101) all are important suburban employment destinations represented by this factor.

The remaining sixth factor accounted for as little as 4.7 percent of the trace and centers around the satellite city of Mt. Clemens (162) to the northeast of Detroit. Warren (172), Roseville (167), and Clinton Township (156) have lesser destination scores and are thus secondary attractors for persons living in this subregion. Mt. Clemens is an established commercial center and the county seat of Macomb County.

When these six subregions are examined as separate homogeneous areas, interesting results are obtained. Between 38 and 74 percent of the home-based work trips originating within each of the six subregions have a destination also within the origin subregion. For the Detroit urbanized area as a whole, 65.3 percent of the home-based work trips remain within the origin subregion. This fact suggests that transportation planning should be oriented more toward the analysis of functional urban subregions. The within-subarea trips are given in Table 1.

Analytical Description of Subregions

The next phase of the case study dealt with the identification of the seven factors and associated six subregions in terms of the 28 state variables given earlier. It is obvious from Figure 2 that distance and travel time are major determinants of people's home/work location behavior. All of the subregional groupings are spatially contiguous with the exception of the few suburban communities with relatively large numbers of central city commuters. However, the nature of the seven factors can also be related to the characteristics of the people who live and work in each spatial area and related land use. Correlation, regression, and discrimination analyses were thus applied to the results of the decomposition.

Correlations between each factor loading and the 28 state variables with absolute values of 0.20 or more were considered in the analytical description. These results are given in Tables 2, 4, 6, 8, 10, and 11. Stepwise regression was then used to determine whether the factor loadings could be predicted from the socioeconomic characteristics in a linear manner. Functions in which all t-ratios were significant at the $\alpha = 0.05$ level were selected, up to a maximum of five variables. Problems of multicollinearity were not considered since the purpose of this portion of the study was

Table 1. Total work trips within origin subregion.

Subregion	Percentage Within Subregion
Central	74.3
Pontiac	71.5
Dearborn	56.7
Warren	47.8
Downriver	55.8
Mt. Clemens	38.0
Total urbanized area	65.3

Table 2. Significant correlations for the central subregion factor.

Variable	Correlation
Percentage of population over 64	0.47
Percentage of households with youngest over 18	0.36
Households per residential acre	0.33
Percentage of female household heads	0.33
Percentage of service employment	0.28
Percentage of renter households	0.24
Percentage of finance, insurance, and real estate employment	0.21
Percentage of family incomes under $3,000	0.20
Percentage of households with no car	0.20
Percentage of commercial land	0.20
Percentage of residence under 1 year	-0.20
Percentage of manufacturing employment	-0.22
Percentage of blue-collar household heads	-0.25
Percentage of households with two or more cars	-0.26
Persons per occupied dwelling unit	-0.35
Percentage of population under 20	-0.45

Table 3. Characteristics of central subregion.

Characteristic	Sub-region Mean	Urban Area Mean	Significantly Different Means				
			Pontiac	Dearborn	Warren	Downriver	Mt. Clemens
Persons per occupied dwelling unit	3.26	3.43		3.67	3.65	3.87	3.90
Percentage of family incomes under $3,000	17.30	14.12			7.47		
Percentage of households with youngest over 18	11.12	9.99		8.44		6.40	
Percentage of population under 20	37.51	40.15		44.28	42.64	46.94	48.24
Percentage of population over 64	10.47	8.53	5.27	5.96	6.43	4.21	4.47
Percentage of renter households	32.02	26.27			14.13		
Percentage of blue-collar household heads	33.14	36.73		46.81			53.73
Percentage of households with no car	23.42	19.53			10.06		
Percentage of households with two or more cars	24.92	28.93	40.09		36.18		
Households per residential acre	9.65	7.77	2.91		4.21		
Percentage of female household heads	19.79	16.54	11.11	12.61	10.76		
Percentage of professional employment	13.88	13.01			7.37		
Percentage of service employment	12.21	9.88		5.78			

Table 4. Significant correlations for the Pontiac subregion factor.

Variable	Correlation
Percentage of households with 2 or more cars	0.27
Households per residential acre	-0.20

Table 5. Characteristics of Pontiac subregion.

Characteristic	Sub-region Mean	Urban Area Mean	Significantly Different Means				
			Central	Dearborn	Warren	Downriver	Mt. Clemens
Percentage of family incomes over $15,000	28.73	12.60		5.23			
Percentage of blue-collar household heads	30.31	36.78					53.73

Table 6. Significant correlations for the Dearborn subregion factor.

Variable	Correlation
Percentage of blue-collar household heads	0.26
Percentage of transportation, communications, and utility employment	0.26
Percentage of population under 20	0.25
Percentage of industrial land	0.25
Persons per occupied dwelling unit	0.22
Percentage of service employment	-0.20
Percentage of family incomes over $15,000	-0.21
Percentage of population over 64	-0.26

basically exploratory. It was found that 13 of the 28 variables entered into the seven multiple regressions.

Finally, discriminant analysis was applied to the six subregional groupings of MCD/CCSs to determine whether they could be significantly discriminated from each other by the state variables. The F-ratio test was used to determine significant groups at a 1 percent level of significance. Variables that distinguish the six groupings most effectively in this manner are given in Tables 3, 5, 7, 9, and 12. The group means for each value are shown for each subregion that was significantly differentiated. Of the 28 socioeconomic variables, 16 were able to distinguish between at least one pair of the groups significantly. As expected, the central subregion proved to be the most complex. Significant correlations for the central subregion are given in Table 2. This central factor was related through multiple regression to three variables taken in combination: percentage of households with youngest child under 18 years of age (+), population over 64 years old (+), and manufacturing employment (-); a multiple R^2 of 0.55 was obtained.

The central subregion is the most easily discriminated group (Table 3). These results are generally in keeping with urban economic and sociological definitions of core areas of large metropolitan environments.

The Pontiac subregion factor loadings did not correlate very highly with the state variables. However, percentage of households with two or more cars was positively related and households per residential acre negatively related (Table 4).

Pontiac was related through multiple regression to residency for less than 1 year (+) and percentage of households with two or more cars (+), but this regression accounted for only 39 percent of the variance in this factor loading vector. Also, the Pontiac grouping was found to be distinguishable in skilled and unskilled household heads and family incomes over $15,000 (Table 5). The statistics relate to the sphere of influence of an industrial satellite city of a large metropolitan environment.

Positive correlations for the Dearborn subregion are given in Table 6. The Dearborn regression equation included population over 64 (-), percentage of industrial land (+), percentage of transportation, communications, and utilities employment (+), and percentage of professional employment (+). The regression accounted for 49 percent of the variance. The Dearborn grouping was distinguished from the other subregions by percentage of families with incomes over $15,000, percentage of professional employment, percentage of skilled and unskilled household heads, and percentage of service employment (Table 7). These results are consistent with theories of predominantly blue-collar workers with large, young families living close to an extensive suburban industrial center.

The Warren subregion factor was also rather complex, for it includes lower middle to upper class communities. Positive and negative correlations are given in Table 8.

The first Warren factor was related through regression to percentage of households with no car (-), percentage of industrial land (-), and percentage of manufacturing employment (+); the R^2 was 0.49. Important characteristics of the Warren grouping discovered through discriminant analysis were low values on families under $3,000, renter-occupied dwellings, households with no cars, and professional employment (Table 9). An important industrial suburban center similar to Dearborn is indicated, although the difference between the population characteristics of this subregion and the central area are accentuated in this case.

The Downriver subregion factor was correlated with two related indexes (Table 10). Downriver was regressed on percentage of family incomes under $3,000 (-), percentage with youngest child over 18 (-), and percentage of industrial land (+); the level of explanation was 48 percent. Discriminant analysis results pointed out the high mean value of 21.38 for the subregion grouping on percentage of industrial land. Knowledge of the area as a heavy lineal (i.e., riverfront) industrial development with adjacent blue-collar residential neighborhoods cross-validates these results.

The Mt. Clemens subregion factor was positively and negatively correlated with the variables given in Table 11. The factor was jointly related to family incomes under $3,000 (+), percentage of skilled and unskilled household heads (+), households with no car (-), and employment in public administration (+); the multiple R^2 was 0.47.

Table 7. Characteristics of Dearborn subregion.

Characteristic	Sub-region Mean	Urban Area Mean	Significantly Different Means				
			Central	Pontiac	Warren	Downriver	Mt. Clemens
Percentage of family incomes over $15,000	5.23	12.60		28.73			
Percentage of professional employment	15.71	13.01			7.37		
Percentage of service employment	5.78	9.88	12.21				

Table 8. Significant correlations for the Warren subregion factor.

Variable	Correlation
Percentage of households with two or more cars	0.36
Median family income	0.31
Persons per occupied dwelling unit	0.26
Percentage of industrial land	-0.21
Percentage of population over 64	-0.22
Percentage of residence under 1 year	-0.23
Percentage of education under 12 years	-0.30
Households per residential acre	-0.31
Percentage of nonwhite population	-0.31
Percentage of female household heads	-0.38
Percentage of family incomes under $3,000	-0.38
Percentage of renter households	-0.40
Percentage of households with no car	-0.42

Table 9. Characteristics of Warren subregion.

Characteristic	Sub-region Mean	Urban Area Mean	Significantly Different Means				
			Central	Pontiac	Dearborn	Downriver	Mt. Clemens
Percentage of family incomes under $3,000	7.47	14.12	17.30				
Percentage of renter households	14.13	26.27	32.02				
Percentage of households with no car	10.06	19.53	23.42				
Percentage of professional employment	7.37	13.01			15.71		

Table 10. Significant correlations for the Downriver subregion factor.

Variable	Correlation
Percentage of industrial land	0.35
Percentage of manufacturing employment	0.24

Table 11. Significant correlations for the Mt. Clemens subregion factor.

Variable	Correlation
Percentage of blue-collar household heads	0.27
Percentage of public administration employment	0.26
Percentage of population under 20	0.20
Percentage of female household heads	-0.20
Percentage of households with no car	-0.21

Table 12. Characteristics of Mt. Clemens subregion.

Characteristic	Sub-region Mean	Urban Area Mean	Significantly Different Means				
			Central	Pontiac	Dearborn	Warren	Downriver
Percentage of blue-collar household heads	53.73	36.78		30.31			
Percentage of public employment	14.19	4.41	3.85	4.04	3.52	6.03	2.03

Finally, the Mt. Clemens grouping was distinguished by skilled and unskilled household heads as well as employment in public administration (Table 12). These statistics indeed describe an established commercial and light-industrial satellite city.

CONCLUSIONS AND DIRECTIONS FOR RESEARCH

This paper has presented a method of reducing a large and complex urban region into smaller subregions for transportation planning purposes. It was shown that a large matrix of origin-destination flows can be reduced from 120 columns to seven without losing a great deal of the original information. Singular decomposition enables the reduction of complex matrices into smaller, more easily handled matrices and at the same time provides for the delineation of subregions exhibiting similar travel behavior analyzed—in this case home-based work trips.

These subregions were found to account for two-thirds of the home-based work trips for the total region and could be described in terms of differences in population, employment, and land use characteristics. More broadly, the subregions can be related to urban economic theories of activity distribution within metropolitan environments.

Ultimate tests of the validity of the outputs of the technique and its usefulness in urban transportation planning will have to be deferred until it can be tested in practice. One possible test might be to compare subregion results from application of the technique with results from consensus judgments of a panel of experts intimately involved with a particular planning question.

This study also makes explicit the importance of other factors besides distance or travel time in the analysis of trip-making behavior. The characteristics of the household, the employment centers, and the overall land use patterns also play an important part in travel behavior. Consequently, several areas of analysis opened up by this study should be pursued further.

The importance of distance and travel time in travel behavior needs to be considered in a subregional and regional context. What is the importance of multiemployment centers and their relationship to existing and potential transportation systems? How is time interrelated with the socioeconomic characteristics of the population, the distribution of jobs by types, and the patterns of land use?

The technique should be extended to the study of various motivations and means of trip-making. Likewise, the analysis should consider other trip purposes, such as shopping and social and recreational travel. Similarly, complex trips that involve modal splits and are multipurpose should eventually be handled. But these extensions can be deferred until the results of more simplistic applications, such as the one reported here, can be competently and objectively evaluated.

REFERENCES

1. Eckart, C., and Young, G. The Approximation of One Matrix by Another of Lower Rank. Psychometrika, Vol. 33, 1936, pp. 73-113.
2. Harman, H. H. Modern Factor Analysis. University of Chicago Press, 1967.
3. Horst, P. Factor Analysis of Data Matrices. Holt, Rinehart, and Winston, 1965.
4. Nash, D. The Generalized Eckart-Young Principal Axis Theorem. General Motors Research Laboratories, 1972.
5. Tatsuoka, M. M. Multivariate Analysis: Techniques for Educational and Psychological Research. John Wiley and Sons, 1971.