PREDICTING PEAK PERIOD TRIPS WITHIN THE FOUR-STEP TRANSPORTATION PLANNING PROCESS

Udesh Jha, Graduate Assistant
John N. Ivan, Assistant Professor

Final Report
September, 1997

JHR 97-258 Project 96-3

This research was sponsored by the Joint Highway Research Advisory Council (JHRAC) of the University of Connecticut and the Connecticut Department of Transportation and was carried out in the Civil and Environmental Department of the University of Connecticut.

Technical Report Documentation Page

1. Report No	2. Government Accession No.	3. Recipient's Catalog No.
JHR 97-258		

4. Title and Subtitle	5. Report Date
Predicting Peak Period Trips within the Four-Step Transportation Planning Process	September, 1997
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Jha, Udesh; Ivan, John N.	JHR 97-258
University of Connecticut Department of Civil and Environmental Engineering 191 Auditorium Road, Box U-37 Storrs, CT 06269	10 Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Connecticut Department of Transportation 280 West Street Rocky Hill, CT 06067-0207	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15 Supplementary Notes

16. Abstract

Transportation Planning in the United States has traditionally focused upon estimating significant transportation network deficiencies and evaluating the effectiveness of large transportation investments. However, the current legislative climate requires the ability to model more microscopic peak hour effects of much smaller scale strategies such as rideshare incentives and transit service improvements. Peak hour modeling has typically been implemented with simple proportionality factors either before or after the trip assignment stage.

This report documents research into augmenting the Connecticut Department of Transportation (ConnDOT) Person Forecasting Model (PERFORM) to estimate PM peak period link volumes. Several approaches were investigated including linear regression, category analysis, discrete choice estimation, and discriminant analysis. Travel behavior data for Connecticut households polled in the 1990 National Personal Transportation Survey (NPTS) were used in the analysis.

Models for estimating PM peak period trip proportion directly were not successful, so an alternate method of predicting daily and PM peak trips separately was used instead. The ratio of PM peak to daily trips was then applied to daily trip tables following the mode choice step to compute PM peak trip tables for each trip purpose. Socio-economic variables such as number of workers and children of various ages were the most significant variables for all purposes, with income being significant for nonhome-based trips. Estimation models for work trips and for stratifying other purposes by geographic region within the state were not successful. The implementation of the resulting models for home-based non-work and nonhome-based trips into PERFORM is demonstrated.

17. Key Words	18. Distribution Statement
Travel Demand Forecasting, Peak Period Trip Forecasting, Four-Step Process	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimetres	mm
ft	feet	0.305	metres	m
yd	yards	0.914	metres	m
mi	miles	1.61	kilometres	km
		AREA		
in^2	square inches	645.2	millimetres squared	mm^2
ft^2	square feet	0.093	metres squared	m^2
yd^2	square yards	0.836	metres squared	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	kilometres squared	km^2
		VOLUME		
fl oz	fluid ounces	29.57	mililitres	mL
gal	gallons	3.785	Litres	L
ft^3	cubic feet	0.028	metres cubed	m^3
yd^3	cubic yards	0.765	metres cubed	m^3

NOTE: Volumes greater than 1000 L shall be shown in m^3

Symbol	When You Know	Multiply By	To Find	Symbol
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9	Celcius temperature	°C

* SI is the symbol for the International System of Measurement

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimetres	0.039	inches	in
m	metres	3.28	feet	ft
m	metres	1.09	yards	yd
km	kilometres	0.621	miles	mi
		AREA		
mm^2	millimetres squared	0.0016	square inches	in^2
m^2	metres squared	10.764	square feet	ft^2
ha	hectares	2.47	acres	ac
km^2	kilometres squared	0.386	square miles	mi^2
		VOLUME		
mL	millilitres	0.034	fluid ounces	fl oz
L	litres	0.264	gallons	gal
m^3	metres cubed	35.315	cubic feet	ft^3
m^3	metres cubed	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celcius temperature	1.8C+32	Fahrenheit temperature	°F

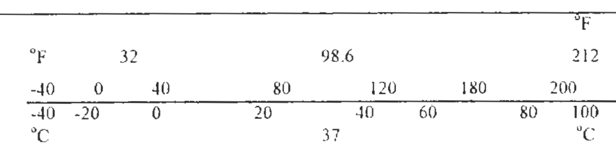

Table of Contents

	Page
List of Figures	v
List of Tables	vi
Chapter 1: Introduction	1
Chapter 2: Literature Review	5
Chapter 3: The CONNDOT PERFORM Model	12
Chapter 4: Data Set	16
Chapter 5: Investigation of Model Forms	27
Chapter 6: Model Estimation Results	53
Chapter 7: Implementation in PERFORM	68
Chapter 8: Conclusions	72
Appendix	75
References	82

List of Figures

	Page
Fig. 1: PERFORM Four-Step Process	13
Fig. 2: Pre-Trip Generation Approach	29
Fig. 3: Post-Trip Generation Approach	30
Fig. 4: Pre-Mode Choice Approach	31
Fig. 5: Post-Mode Choice Approach	32
Fig. 6a: PM Peak Trip Distribution: HBO	34
Fig. 6b: PM Peak Trip Distribution: NHB	34
Fig. 7a: Cross-classification Table for PM Peak Proportion for daily HBW	37
Fig. 7b: Analysis of Variance for Proportion of Daily HBW in PM Peak	37
Fig. 8a: Cross-classification Table for PM Peak Proportion for daily HBO	38
Fig. 8b: Analysis of Variance for Proportion of Daily HBO in PM Peak	39
Fig. 9a: Cross-classification Table for PM Peak Proportion for daily NHB	40
Fig. 9b: Analysis of Variance for Proportion of Daily NHB in PM Peak	41
Fig. 10: Discriminant Analysis Model Output: PM Peak HBW	46
Fig. 11: Discriminant Analysis Model Output: PM Peak HBO	48
Fig. 12: Discriminant Analysis Model Output: PM Peak NHB	50
Fig. 13a: Error in Aggregate Prediction of PM Peak Proportion: HBO	70
Fig. 13b: Error in Aggregate Prediction of PM Peak Proportion: NHB	70

List of Tables

Page

Table 1: List of Variables Used in Model Estimations 20

Table 2: Paired-t Test for Independent Variables: HBW peak/off-peak 23

Table 3: Paired-t Test for Independent Variables: HBO peak/off-peak 24

Table 4: Paired-t Test for Independent Variables: NHB peak/off-peak 25

Table 5: Direct Prediction of PM Peak Proportion by Purpose 36

Table 6: Individual Choice Model Results 43

Table 7: Linear Regression Parameters for Statewide Models by Purpose 55

Table 8: Summary of Land Use Data for Different Regions of the State 61

Table 9: Linear Regression Parameters for Home Based Work (HBW) 62

Table 10: Linear Regression Parameters for Home Based Other (HBO) 63

Table 11: Linear Regression Parameters for Non-Home Based (NHB) 64

Table 12: F-test Statistics for Models Compared for Different Regions 66

Table 13: Sample Application of the Model to Aggregate Data 69

Chapter 1: INTRODUCTION

Transportation planning in the United States has traditionally focused upon estimating significant transportation network deficiencies and evaluating the effectiveness of large transportation system investments. Recent changes in national transportation investment priorities have placed new demands on the traditional transportation planning process. Since the passage of the Clean Air Act Amendments (CAAA) of 1990, much higher levels of importance have been placed on the interface between travel models and air quality models which predict vehicular emissions and pollution dispersion [1].

The first transportation legislation that had specific mandates with regard to achieving the objectives of CAAA was the Intermodal Surface Transportation Efficiency Act (ISTEA), passed in 1991. ISTEA recognized changing transportation planning priorities and attached new stipulations on state eligibility for Federal transportation project funds. Some of these stipulations were the establishment of specific management systems and serious consideration of alternatives for reducing congestion and improving air quality that do not involve construction of new highway lanes. Thus, cost effective strategies which influence travel demand (rather than supply) are being encouraged.

Federal regulations have been enacted in recent years to control and reduce air pollution caused by vehicular traffic. In areas that do not meet air quality requirements set by the National Ambient Air Quality Standards (NAAQS), ISTEA calls for the adoption of Transportation Control Measures (TCM). Some of these measures are programs to limit or restrict vehicle use in downtown areas during peak periods, and employer based programs to reduce single occupancy vehicles and permit flexible work schedules. The effectiveness of these measures are extremely sensitive to predicted peak period travel patterns. In this context, research into improving methods for predicting peak period travel behavior assume immense significance.

The traditional travel demand forecasting process, as currently configured, is inadequate to meet the needs implied by these changes in the planning environment. Since the process was designed to support major investment decisions, forecast results are not sensitive to the relatively small changes in facility characteristics or transportation choice parameters caused by the less intensive and less costly alternatives listed above. In particular, most travel demand forecasting models, including the Connecticut PERson FORecasting Model (PERFORM, [2]), are not calibrated to estimate peak period travel volumes with sensitivity to peak period highway network congestion. Instead, they usually compute peak period trip tables by applying aggregate proportionality factors to daily trip totals. These factors are not sensitive to characteristics of the traffic network, or of the tripmakers themselves, which likely affect the trip-making time choice. In fact, daily trip generation models are often also insensitive to alternative trip-making patterns due to variation in the accessibility of traffic zones, and the land use characteristics of the zones.

Some important information required for vehicular emissions modeling are vehicle miles traveled (VMT) and average trip speed, both by hour of day and vehicular mode, particularly during the peak hour, as recommended by a U.S. Department of Transportation (DOT) report [3]. For demonstrating ozone attainment, the photochemical grid dispersion model requires VMT by hour of the day. Therefore, ISTEA requires the four-step transportation planning process to be modified to provide information on trip-making by time of day, and to be sensitive to transportation policies which may alter the time that trip-making occurs. To monitor congestion on roadways, this report suggests using percent person miles traveled (% PMT) by speed range (or increment), origin-destination travel time by purpose and time of day, and percent heavy vehicles by speed range. Most of these required measures cannot be accurately modeled within the traditional travel demand forecasting models which are insensitive to the impact of peak period congestion levels, household characteristics or travel demand reduction policy initiatives on person travel.

The Travel Model Improvement Program (TMIP), a joint initiative of several agencies in the United States Department of Transportation, the Environmental Protection Agency (EPA) and the Department of Energy, was formed to encourage development of enhancements to the travel demand forecasting process [3]. Another purpose of this forum was to better integrate forecasting techniques into the decision making process. This TMIP report focused upon identifying problems with current travel demand forecasting models and preparing a plan to address them. To improve the four-step process, this forum recommended development of time of day trip tables immediately after trip generation, prior to the first use of any network derived data. An alternative approach suggested was to add a time of day model as the fifth step in the process.

Another reason why aggregate travel demand forecasting models fail in their contexts is that they treat regional trip generation as completely independent of the network characteristics: that is, they do not account for how changes in highway level of service affect household trip-making habits. Stopher [4] points out that current land use and trip generation models do not predict accurately because congestion effects on the forecast distribution of population and employment as well as total trip-making are neglected. These models also do not predict how different household types might respond to specific governmental and organizational policies that cause changes in peak hour highway travel choices such as preferential parking of vanpools, employer transit subsidies, and flexible working hours. The ability to modify the estimated share of peak-hour travel based on congestion levels and trip distance will improve the ability of urban regions to estimate accurate congestion levels for future conditions [5].

The research described here was aimed at meeting some of these deficiencies in the present implementation of travel demand forecasting models with respect to peak period trip forecasting. It was expected that realizing the effects of socio-economic, land use and congestion variables, rather than ignoring them, would lead to better peak period forecasts. Several methods of predicting peak period trips and incorporating them into the four step process are proposed. One of the methods has been chosen to estimate peak period trips for the PERFORM model used by the Connecticut Department of Transportation (CONNDOT). In this method, models have been estimated to predict the proportion of daily trips made during the PM peak period, by trip purpose. The proportions are applied to the auto trip table, by trip purpose, to estimate a peak period auto trip table. The incorporation of these models into PERFORM is also described.

Chapter 2: LITERATURE REVIEW

Peak period travel demand models

The most common practice in the modeling of peak period travel is to produce a twenty-four hour assignment and predict peak hour trips as a constant percentage of the twenty-four hour volume [6]. In this method. no effort is made to relate peaking characteristics to the anticipated level of congestion. However. recent researchers have tried to improve peak period modeling by recognizing the effects of socio-economic characteristics. land use. and roadway congestion. These efforts have mainly been in one of three directions: (a) pre-distribution. which factor daily trip ends into peak/off-peak; (b) post-distribution. which split daily trip interchanges; and (c) post-assignment. which split daily link volumes or daily VMT.

In models developed for the Metropolitan Transportation Authority's Red Line East Side Extension Project in Los Angeles, a pre-distribution time-of-day model was used. In this technique. the trip ends were split by time period for each trip purpose [1]. The initial step in this study was the computation of a set of factors to compute trips by time of day. usually for multi-hour peak and off-peak periods. and by trip purpose. The factors were based on peaking characteristics such as trip purpose, jurisdiction. area type. and socio-economic stratification. These factors were applied to the trip ends from the trip generation model and produced trip-ends by peak and off-peak periods for each of the trip purposes.

In an another form of pre-distribution model. Kumar and Levinson [7] developed linear regression models to directly predict afternoon-peak period trip generation. For developing a model to estimate travel during a part of the day, they emphasized the criticality of explicitly accounting for each trip end. because the trips are balanced on a daily basis but not within the morning or afternoon peak. Thus. the home-based-work (HBW) trip in the morning is almost always a home-to-work trip, with home as the origin and the workplace as the destination. In the afternoon. it is usually a work-to-home trip. with the workplace as the origin and home as the destination. Similarly, the home-based-other (HBO) trip may involve going shopping or returning home.

As part of the SATURN [8] suite of travel models in London. a different method of estimating peak period trip tables was used [9]. In this study. a series of origin and destination surveys were conducted across a cordon around the proposed study area to collect information on travel patterns. Using the trip origins and destinations and the times in which the trips were made, trip matrices were computed for peak periods.

In the post-distribution models (which could be applied after trip distribution or mode choice), the number of daily trip interchanges are factored. Unlike trip-end factoring which can depend only on characteristics of the origin or the production zone. peak period factors for trip interchanges can depend on characteristics of the destination or attraction zone as well. Trip-interchange peak period factors have also been computed using link volumes from traffic count stations or from travel survey data. These link volumes are used to develop hourly distributions of traffic demand. Peak period percentages are computed using these distributions. In a recent study by Crevo and Virkud [10], nontruck peak hour trips were factored using peak hour percentages from

National Personal Transportation Survey (NPTS) data. For truck trips, they used the factors from a FHWA report [11] to extract peak hour trip tables from daily trip tables.

One major concern for pre- and post-distribution models is the conversion of the trip table from production-attraction (P/A) format to origin-destination (O/D) format for input into the trip assignment module. The trip table after mode choice is typically in P/A format and multiplied by diurnal factors to convert to O/D format. These factors split the trips between two directions; for example, daily HBW trips are split between home-to-work or work-to-home trips. These factors are best derived from home interview survey data [1]. For the peak period we would expect the factors to be skewed because most of the trips are not balanced within the same peak period.

Stopher [4] says that another problem related to the diurnal factors applied after trip distribution arises where transit modeling is required. In some cases, factors are applied to the trip tables after isolating the transit trips from the table. Therefore, diurnal factoring after mode choice and prior to assignment may result in the outright loss of transit trips or the need to undertake an involved and arbitrary accounting procedure to retain all peak period transit trips. This situation arises because most urban areas have certain transit services that operate in the peak period only, so these trips are not included in the peak period trip table. One remedy to this problem is to perform diurnal factoring after trip generation itself, i.e., estimate the directional trip productions and attractions separately and then perform trip distribution separately.

In the post-assignment models, the most popular method is to extract peak hour volumes as a fixed percentage of link volumes, generally for congested links. This type of factoring is insensitive to trip purpose since link volumes include trips of all purposes. In

PERFORM, peak period percentages are applied to daily vehicle miles traveled (VMT) to get peak period VMT which is later used in air quality modeling.

Spiess and Suter [12] developed a dynamic model to predict flows for every link in a network for every hour of the day. In their model, given hourly traffic counts on a subset of links in a network, a set of coefficients for each hour for each link is computed by combining the equilibrium assignment model with a multiple regression model. The hourly coefficients are used with demand matrices, which are defined for each hour, to compute hourly flows. This model was used in Basel, Switzerland as a basis for the computation of fuel consumption and vehicle emissions caused by the road traffic.

The peak period volumes obtained through these procedures are used in various post four step analyses, such as peak spreading studies, transportation control measure studies, and air quality monitoring. For peak spreading studies, the peak period volumes are used to develop a histogram of hourly flow during the peak period. Allen [13] describes a modified Poisson distribution model to predict the spread of 4-hr volumes across each 15-minute period on a portion of the I-80 corridor in New Jersey. He determined that as twenty-four hour volumes continue to increase, peak hour volumes do not increase at the same rate. Instead, roadway conditions force drivers to modify their departure times, and travel in the "shoulders" of the peak period. Hence, this is an evidence of how congestion and land use characteristics might influence peak period travel demand. To introduce roadway conditions into peak period travel modeling, Ruiter [14] gave a model for estimating the ratio of peak hour to peak period flow as a decreasing function of volume/capacity ratio.

Land use and congestion in travel demand modeling

Incorporating congestion and land use variables in travel demand models has become a major topic for discussion and debate over the past several years. An examination of trends in travel behavior shows increases in the number of workers per household, licensed drivers per household, vehicles per household, vehicle trips, and average trip length [15]. Both employment and housing are growing at faster rates in suburban than at center city locations, with the majority of employment growth now occurring at suburban employment centers. Auto occupancy rates are declining and the percentage of single-occupant commuting is increasing. Overall, vehicle miles of travel, in most areas, is growing at a higher rate than either employment or population. The result of these trends is an increasing level of traffic congestion in urban and suburban areas across the country, particularly during peak commuting periods.

A response to these increasing levels of congestion has been attempts to increase the efficiency of existing transportation resources. Experimentation with practical and cost- effective alternatives to building new lane miles of roadway is underway. These alternatives have frequently been transportation demand management measures, which attempt to influence travel demand behavior, especially reducing single occupancy vehicle trips. These measures have experienced mixed success. For example, it has been found that transit works best in areas with moderate to high densities at both the work and home ends of the trip; and ridesharing is more effective for long trips attracted to dense activity centers [15]. A better understanding of these travel patterns and methods of influencing travel behavior accordingly is needed. With the emerging importance of alternatives to the drive alone mode, travel time choices, and travel purposes on person travel, it is imperative that we improve the conventional travel demand models to include their effects. Two enhancements which might help are improvements in peak period modeling and inclusion of the effects of congestion and land use in travel modeling.

Researchers have reported mixed success in using land use and congestion variables in modeling travel demand. Purvis, et. al [16] have studied methods to incorporate workplace accessibility in regional travel forecasting models and to better analyze the issue of induced trip-making, and to provide a better understanding of the linkage between congestion and trip frequency choice behavior. They developed nonwork trip generation models incorporating work trip duration as a variable. They indicated that a 10 percent decrease in average work trip duration would yield a 1.2 percent increase in home based shopping trips and a 0.9 percent increase in home based social/recreation trips.

Kockelman [17] reported the relative significance of a variety of measures of urban form on household VMT, automobile ownership, and mode choice and proposed formulations for computing different variables like accessibility, land use mix, and land use balance. He found that, in many cases, accessibility and land use are more relevant than several other household and traveler characteristics (e.g., household size, auto ownership, age, sex, race, and household income) that often form a basis for travel behavior prediction. With the improved technology of geographical information systems and the increasing availability of land use datasets, he says, accessibility measures can be computed for a multitude of zones at relatively low cost. He has called for the inclusion of land use and congestion variables in travel demand modeling.

A study by McNally and Kulkarni [18] examined the effects of land use and network characteristics on travel behavior for twenty neighborhoods in Orange County. To control for household and socio-economic variables, the households were clustered by various neighborhood types based on their socio-economic characteristics. Models were developed for each cluster, thus eliminating the effect of other socio-economic and household characteristics on trip-making. The travel models were analyzed to see if land use and network characteristics had a significant effect on the travel behavior of the different neighborhoods. Their results seem to show a very weak relationship between land use and transportation systems and travel behavior, so they recommend exercising caution in interpreting the results. Further studies are needed to decide whether these results are unique to Southern California.

Ewing, et. al. [19] studied the independent effects of land use and congestion variables on household trip rates using data from a Florida travel survey. They found that, after controlling for socio-demographic variables, residential density, mixed use, and congestion do not have significant, independent effects on household trip rates. They say that while land use and congestion variables may have no direct effects, it is logical that they may effect household trip rates indirectly.

The dataset used in the research for this thesis provided socio-economic as well as land use data. It was therefore possible to investigate whether or not congestion and land use variables have any effect on daily and peak period trip-making. Model estimation results and their analyses are discussed in later chapters.

Chapter 3: The CONNDOT PERFORM MODEL

PERFORM (depicted in Figure 1) is a network-based computer model implemented in the TRANPLAN metropolitan planning software package and is the forecasting context for this research. This model consists of four basic steps: trip generation, trip distribution, modal split and trip assignment. It analyzes three different trip purposes: home based work (HBW), home based non-work (HBO), and non-home based (NHB) trips; and four travel modes: auto drive alone (SOV), auto shared ride (HOV), bus (local and express), and rail. The model estimates total vehicle link volumes on an average weekday.

Trip generation is performed separately for each trip purpose using cross-classification models, linear regression models or models based on trip rates. These models predict person trip productions and attractions for each traffic analysis zone (TAZ) from the existing socio-economic and land use data, by trip purpose. The trip purposes described earlier are further stratified by internal trips and external trips. Internal trips begin and end within Connecticut, while external trips have one end inside and the other end outside Connecticut. External trips comprise those trips which are produced outside the state and attracted into the state, or produced in the state and attracted outside the state. Thru trips (neither the origin nor the destination is in Connecticut) are estimated separately and are added to the trip table immediately before performing trip assignment. Predictor variables in these trip generation models are auto ownership, population, retail and non-retail employment, and shopping area.

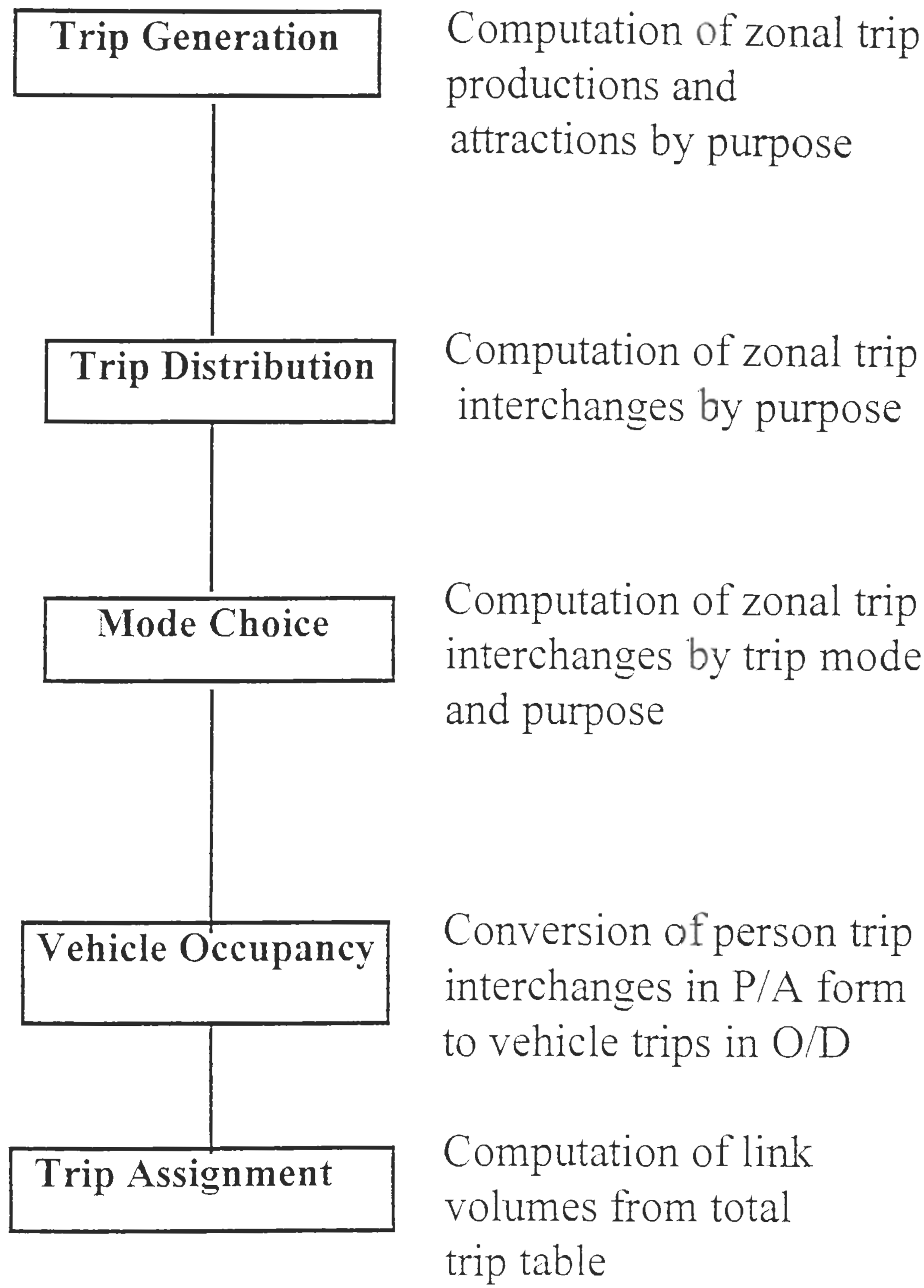

Fig 1. PERFORM four-step process

Trip distribution uses the Gravity Model, in which the estimated number of trips between two areas is directly related to the activities in the areas (represented by trip generation) and inversely proportional to the separation between the areas (represented as a function of the travel time). The trip distribution model produces a series of trip tables, which are matrices depicting the number of average daily person trips produced and attracted between every pair of TAZs, by trip purpose.

In the modal split portion of the model, the interzonal person trips distributed by the Gravity Model are allocated among the available travel modes. This model is also delineated by trip purpose. The HBW and HBO mode splits are performed by multinomial logit models, with the HBO splits also used for NHB trips.

The vehicle occupancy module of PERFORM, an intermediate step, converts the HOV person trips from the mode split process to HOV vehicle trips by trip purpose. Occupancy factors for HBW trips are calculated using a curve derived from observations of occupancy at various highway locations during the commuter traffic hours. Occupancy rates for non-work trip purposes are from ConnDOT's traffic accident database, in which statistics from highway accidents are used to determine average occupancies by time of day. After the conversion from person to vehicle trips, the vehicle trip tables for each mode and purpose are added together for input into the trip assignment model.

Highway trip assignment in PERFORM is done by an equilibrium assignment process on an average daily basis. In this step, vehicle trips are iteratively assigned to the network so as to equate the travel between alternate paths between the same origin and destination so that all used paths between the same origin and destination zones have the same travel times and path flow. Once the trips have been assigned to the links, daily

VMT is computed from the link volumes. The peak period VMT is then factored out of daily VMT using aggregate proportionality factors.

Chapter 4: DATA SET

Data Sources

The data used for this project came from three sources. The first source was the 1990 National Personal Transportation Survey (NPTS). NPTS contains information about the amount and nature of personal travel in the United States, as related to the demographics of persons and households (*9*). This is a unique source of data describing personal travel and provides information about household trips, their purposes, and the transportation modes used. NPTS collected information on all trips made by the respondent on a specific day (known as "travel day"), combined with longer trip data taken over a two-week period (known as "travel period"). The different tables in the database contained information on trips made by each responding household as well as trips made by each responding individual. In the data for Connecticut, there were 15,088 individual trip records associated with 2,266 reference households in the state. Early analyses of the nationwide NPTS dataset suggest changes in peak period travel behavior such as workers departing earlier for work in order to avoid the peak traffic congestion. The second source of data was land use inventories for towns in Connecticut, providing population, retail and non-retail employment, retail space, mean household size, and the number of households in each town. The third source of data was the congested and uncongested skim tables in the form of travel times between the different TAZ's in the state. These travel times, along with the land use data, were used for computing daily and peak accessibility indices for different purposes for all the TAZs in the state.

Data Set Preparation

Accessibility indices represent the relative ease of reaching activity opportunities from a single zone under different levels of congestion in the network. To test the relationship between congestion and peak period trip generation in this Connecticut dataset, these indices were computed for each production zone, by trip purpose. Kockelman [17] suggests a popular method of computing accessibility indices from the Gravity Model of the trip distribution step in the four-step process. This functional form for accessibility index assumes a direct proportionality with the number of opportunities and an inverse relation relative to the cost of accessing those opportunities. The cost of accessing those opportunities is assumed to be the PERFORM travel skims, related through calibrated F-factors. The opportunities for three different purposes were obtained from the land use inventories for the towns in Connecticut. O_j, the opportunities in a destination zone 'j' from an origin zone 'i' for each purpose was computed as:

1. HBW: Non-Retail employment in zone 'j'.
2. HBO: Population of zone 'j'.
3. NHB: Retail employment in zone 'j'.

The NPTS dataset has zip code rather than TAZ as the locational variable for the individual/household trip. This dataset for the state of Connecticut included 288 different zip codes. A centroid TAZ was established for each zip code using the PERFORM zone map; using these relationships a zip code to TAZ reference table was created. Using this reference table, the centroid TAZ and the corresponding town name were added as new variables to the NPTS dataset. Using the town name as the key variable between the

NPTS dataset and land use dataset, the land use variables were also copied to the NPTS dataset. These land use variables were used for computing accessibility indices.

The formulation for the computation of accessibility indices is:

$$A_j = \sum_j O_j F_{ij} K_{ij}$$

where

A_j = accessibility index for the origin zone 'j'.

O_j = opportunity available in the destination zone for the particular purpose. The way in which the opportunities were computed was discussed earlier in the section.

F_{ij} = empirically derived travel time (or friction) factors that express the average areawide effect of spatial separation on trip interchanges between zones. These factors had been calibrated in PERFORM.

K_{ij} = specific zone to zone adjustment factors to allow for the incorporation of the effect on travel patterns of defined social or economic linkages not otherwise accounted for in the Gravity Model formulation. These K-factors had been calibrated in PERFORM.

The Gravity Model module of TRANPLAN was used to compute the accessibility indices for all three PERFORM trip purposes. Using congested and uncongested travel times, accessibility indices were computed for PM peak period as well as daily trips for the three purposes. The ratio of the PM peak to the daily accessibility index was used as the independent variable in model estimations, for each purpose. These accessibility index ratios were added to the NPTS dataset using the origin zone TAZ as the key variable. Finally, two sets of data (individual trips and household trips) were obtained. The household trips dataset was an aggregate form of the individual trips dataset.

Table 1. List of variables used in model estimations

Variable Name	Variable Description
DRVRCNT	Number of drivers in the household
AUTOCNT	Number of autos in the household
HHFAMINC	Income of the household (Categorical variable)
LIFCYC	Family life cycle of the household (Categorical Variable)
FAMLRACE	Race of the household
NUMADULTS	Number of adults in the household
SCHOOL_AGE	Number of children in the household aged between 5 and 21
PRE_SCHOOL	Number of children in the household aged between 0 and 4
HHSIZE	Size of the household
POVERTY	Is the family below, near or above poverty level (Categorical Variable)
PUBTRANS	Public transit availability for the household
WKRCNT	Number of workers in the household
POPLN	Population of the town of the household
POPDENS	Population density of the town of the household
RET	Number of retail employment in the town of the household
NRET	Number of non-retail employment in the town of the household
SHSF	Shopping area in the town of the household
HHSZ	Mean household size of the town of the household
A_RATIO (HBW)	Ratio of peak to off-peak HBW accessibility index of the TAZ in which the household exists
A_RATIO (NHB)	Ratio of peak to off-peak NHB accessibility index of the TAZ in which the household exists in which the household exists
A_RATIO (HBO)	Ratio of peak to off-peak HBO accessibility index of the TAZ in which the household exists in which the household exists

Data Set Exploration

Apart from the trends described earlier regarding the nationwide NPTS dataset. there are other interesting results worth mentioning with regard to the changes in travel behavior patterns which could affect peak period trip-making. The 1990 NPTS shows a growth in person miles of travel of about 19 percent from 1983 that could be attributed to increases in per capita trip-making and average trip length. The factors. which were found to be significant in affecting the per capita trip-making, were household size and number of licensed drivers in the household. The stage in the household life cycle also had an important effect on trip-making because so many trips are created by the specific needs of different families' types. like the need to serve younger members of the household. or care for the household itself.

The two most significant factors in average trip length growth are population shifts to very large metropolitan areas. and the increasing number of driver's licenses. particularly among women. For example. work trips by women with driver's licenses are 50 percent longer than work trips by women without driver's licenses. Women also led the growth in trip-making with a major share of their increase being personal business trips. A third factor noticed was that some trip purposes have been growing in trip length (e.g., work and personal business). while others (like shopping), have been more stable.

Another issue to be considered in evaluating trip lengths is the start time of the trip. Increasing congestion on the roadways has pushed traffic into the "shoulder" periods either before or after the peak. Changing job patterns. particularly the shift from manufacturing to service sector employment have tended to move work trips away from traditional peak periods. The incorporation of these trends into travel demand models would significantly improve daily as well as peak period trip estimations.

Table 1 gives a list of important independent variables from the dataset investigated to estimate PM peak period trips and daily trips. Exploratory analyses on the dataset included frequencies, histograms, and cross-classifications for relevant independent variables. The inter-correlation among the independent variables were observed as well as their correlation with PM peak and daily trip-making. Towards the first step in variable selection, paired t-tests were done for important independent variables to find the variation between peak and off-peak trips for each trip purpose. Tables 2 through 4 show the results of the paired-t tests for different purposes. Only those independent variables, which showed significant variation from peak to off-peak trips, were selected as independent variables to develop peak period models.

For the HBW purpose, the variables that were significantly different from peak to off-peak are NRET, SHSF, A_RATIO (HBW), and WKRCNT. SHSF and NRET are indirectly related to labor force and employment in a zone and hence they could be important predictors for work trips. A_RATIO (HBW), which is computed using peak and off-peak travel times is significant, confirming that peak period congestion affects the timing of work trips.

For HBO trips, the independent variables that were significantly different from peak to off-peak were SHSF, PRE_SCHOOL, WKRCNT, and SCHOOL_AGE. The significance of WKRCNT could be due to the fact that many commuters are chaining their trips while on their way to work, which makes the number of trips differ significantly from peak to off- peak. The significance of PRE_SCHOOL and

Table 2. Paired-t test for independent variables : HBW peak/off-peak

Variables	Mean in Peak	SD in Peak	Mean in Off-Peak	SD in Off-Peak	Test Statistic	DOF	t (critical) at 95% confidence	Hypothesis Result
POP	43848.06	36462.40	43169.45	35040.18	0.9876	11220	1.96	
RET	3787.58	3011.80	3823.12	2986.42	-0.6183	11220	1.96	
NRET	19206.38	23052.73	18241.69	21048.11	2.2648	11220	1.96	Significant
SHSF	1033.28	993.73	1073.02	1037.29	-2.0503	11220	1.96	Significant
HHLD	16933.41	14163.21	16681.80	13671.96	0.9410	11220	1.96	
HHSZ	2.63	0.18	2.64	0.20	-0.4472	11220	1.96	
ARATIO(HBW)	6.30	3.50	6.15	3.36	2.2777	11220	1.96	Significant
ARATIO(HBO)	1.10	0.75	1.10	0.71	0.1711	11220	1.96	
ARATIO(NHB)	0.10	0.08	0.10	0.08	2.0872	11220	1.96	Significant
AUTOCNT	1.96	0.92	1.95	0.96	0.7477	11220	1.96	
DRVRCNT	2.04	0.86	2.04	0.87	0.3397	11220	1.96	
PRE_SCHOOL	0.22	0.52	0.21	0.51	1.0091	11220	1.96	
POPDENS	2055.98	2344.30	1980.31	2207.30	1.7266	11220	1.96	
WKRCNT	1.65	0.93	1.58	0.98	3.765	11220	1.96	Significant
SCHOOL_AGE	0.87	1.31	0.93	1.22	-1.2863	11220	1.96	Significant

- The variables which are significant in the above test are vary statistically in their mean value from peak to off-peak trips
- Number of cases in peak period = 4669; number of cases in off-peak period = 6553

Table 3. Paired-t test for independent variables : HBO peak/off-peak

Variables	Mean in Peak	SD in Peak	Mean in Off-Peak	SD in Off-Peak	Test Statistic	DOF	t (critical) at 95% confidence	Hypothesis Result
POP	43476.75	36318.49	42860.75	34463.70	0.7275	7932	1.96	
RET	3756.26	2951.53	3803.24	2941.55	-0.6719	7932	1.96	
NRET	18743.45	21985.35	17895.46	20128.27	1.6727	7932	1.96	
SHSF	1033.94	982.98	1081.86	1034.12	-2.019	7932	1.96	Significant
HHLD	16779.47	14085.02	16565.31	13459.00	0.6507	7932	1.96	
HHSZ	2.64	0.18	2.64	0.20	-0.4687	7932	1.96	
ARATIO(HBW)	6.37	3.58	6.13	3.33	2.9496	7932	1.96	Significant
ARATIO(HBO)	1.10	0.74	1.10	0.70	0.1860	7932	1.96	
ARATIO(NHB)	0.10	0.08	0.10	0.07	1.5473	7932	1.96	
AUTOCNT	1.92	0.91	1.93	0.96	-0.5919	7932	1.96	
DRVRCNT	1.98	0.81	2.02	0.87	-1.9198	7932	1.96	Significant
PRE_SCHOOL	0.25	0.52	0.22	0.51	2.1501	7932	1.96	Significant
POPDENS	2003.12	2329.99	1966.09	2163.58	0.6863	7932	1.96	
WKRCNT	1.47	0.92	1.51	1.00	-1.6947	7932	1.96	Significant
SCHOOL_AGE	0.88	1.31	0.96	1.23	-2.1131	7932	1.96	Significant

- The variables which are significant in the above test are vary statistically in their mean value from peak to off-peak trips
- Number of cases in peak period = 2689; number of cases in off-peak period = 5245

Table 4. Paired-t test for independent variables : NHB peak/off-peak

Variables	Mean in Peak	SD in Peak	Mean in Off-Peak	SD in Off-Peak	Test Statistic	DOF	t (critical) at 95% confidence	Hypothesis Result
POP	40878.11	35208.24	43302.20	35474.71	-1.8474	3299	1.96	
RET	3570.84	2859.76	3700.76	2917.09	-1.2140	3299	1.96	
NRET	17676.05	22402.49	18512.12	21986.84	-1.0099	3299	1.96	
SHSF	984.28	962.72	1039.69	982.50	-1.5377	3299	1.96	
HHLD	15774.07	13720.80	16732.35	13842.49	-1.8732	3299	1.96	
HHSZ	2.64	0.22	2.64	0.21	0.1724	3299	1.96	
ARATIO(HBW)	5.86	3.17	6.03	3.27	-1.4446	3299	1.96	
ARATIO(HBO)	1.08	0.70	1.13	0.75	-1.6186	3299	1.96	
ARATIO(NHB)	0.11	0.08	0.10	0.08	1.2121	3299	1.96	
AUTOCNT	1.96	0.93	1.98	0.96	-0.7546	3299	1.96	
DRVRCNT	1.98	0.86	1.95	0.79	0.8530	3299	1.96	
PRE_SCHOOL	0.22	0.53	0.24	0.51	-0.5354	3299	1.96	
POPDENS	1.66	0.90	1.58	0.91	2.3509	3299	1.96	Significant
WKRCNT	1881.74	2237.27	2001.00	2266.93	-1.4277	3299	1.96	
SCHOOL_AGE	0.92	1.14	0.83	1.13	-2.4915	3299	1.96	Significant

- The variables which are significant in the above test are vary statistically in their mean value from peak to off-peak trips
- Number of cases in peak period = 2689; number of cases in off-peak period = 5245

SCHOOL_AGE suggests that the presence of infants and children in the household affects decisions about when discretionary trips are made. For NHB trips, the variables that differed significantly from peak to off-peak are WKRCNT, SCHOOL_AGE, and POPDENS. WKRCNT appears to be significant again due to the fact that many peak period commuters and single member households are chaining trips.

Two other important issues regarding the selection of independent variables for model estimation were their ease in prediction as well as their inter-correlation. Variables like DRVRCNT, AUTOCNT, NUMADULTS were not used in model estimations because they were highly correlated with WKRCNT. A recent study by Allen, et. al. [22] suggests that stage in the household life cycle is an important predictive variable for trip generation. The NPTS report [21] also says that, according to 1983 NPTS, the family groups in which a population distributes itself has a very important effect on trip-making. The corresponding data show that person trip rates can double between a household consisting of a single adult with a child under six and a household with two children. Therefore, stage in the household life cycle has been identified as a significant variable in estimating travel demand. For this Connecticut dataset, too, life cycle was highly correlated with daily and PM peak period trip-making. However, it was not included in model developments in this research due to difficulties related to its forecasting within PERFORM; instead variables such as the number of children of different ages, number of workers, and income levels are used. Other variables like family race, poverty status, and mean household size also showed promising significance but were excluded because of expected forecasting difficulties.

Chapter 3: The CONNDOT PERFORM MODEL

PERFORM (depicted in Figure 1) is a network-based computer model implemented in the TRANPLAN metropolitan planning software package and is the forecasting context for this research. This model consists of four basic steps: trip generation, trip distribution, modal split and trip assignment. It analyzes three different trip purposes: home based work (HBW), home based non-work (HBO), and non-home based (NHB) trips; and four travel modes: auto drive alone (SOV), auto shared ride (HOV), bus (local and express), and rail. The model estimates total vehicle link volumes on an average weekday.

Trip generation is performed separately for each trip purpose using cross-classification models, linear regression models or models based on trip rates. These models predict person trip productions and attractions for each traffic analysis zone (TAZ) from the existing socio-economic and land use data, by trip purpose. The trip purposes described earlier are further stratified by internal trips and external trips. Internal trips begin and end within Connecticut, while external trips have one end inside and the other end outside Connecticut. External trips comprise those trips which are produced outside the state and attracted into the state, or produced in the state and attracted outside the state. Thru trips (neither the origin nor the destination is in Connecticut) are estimated separately and are added to the trip table immediately before performing trip assignment. Predictor variables in these trip generation models are auto ownership, population, retail and non-retail employment, and shopping area.

Approach 2: Post-Trip Generation (Figure 3)

1. Estimate daily trip ends by purpose.
2. Allocate daily trip ends between peak and off-peak periods.
3. Estimate trip distribution, mode choice and trip assignment on a peak period basis.

Approach 3: Pre-Mode Choice (Figure 4)

1. Estimate daily person trip interchanges by purpose.
2. Allocate daily person trip interchanges between peak and off-peak periods.
3. Estimate mode choice and trip assignment on a peak period basis.

Approach 4: Post-Mode Choice (Figure 5)

1. Estimate daily auto trip interchanges by purpose.
2. Allocate daily auto trip interchanges between peak and off-peak periods.
3. Estimate trip assignment on a peak-period basis.

Methods 1 and 2 are applied pre-distribution and factor the trip ends by peak and off-peak period. These methods are best suited if only the production zone characteristics of the trips are known and information about the other end of the trip is unknown. Suitable model forms for these methods include linear regression and cross-classification models that could predict peak and daily trip ends directly or predict the peak proportion. Methods 2 and 3 are applied post-distribution and factor the trip interchanges between the peak and off-peak periods. These methods are most effective if information about both ends of each trip are known. NPTS does not provide information about the attraction end of the trips, so for this estimation exercise, peak period proportion models following trip

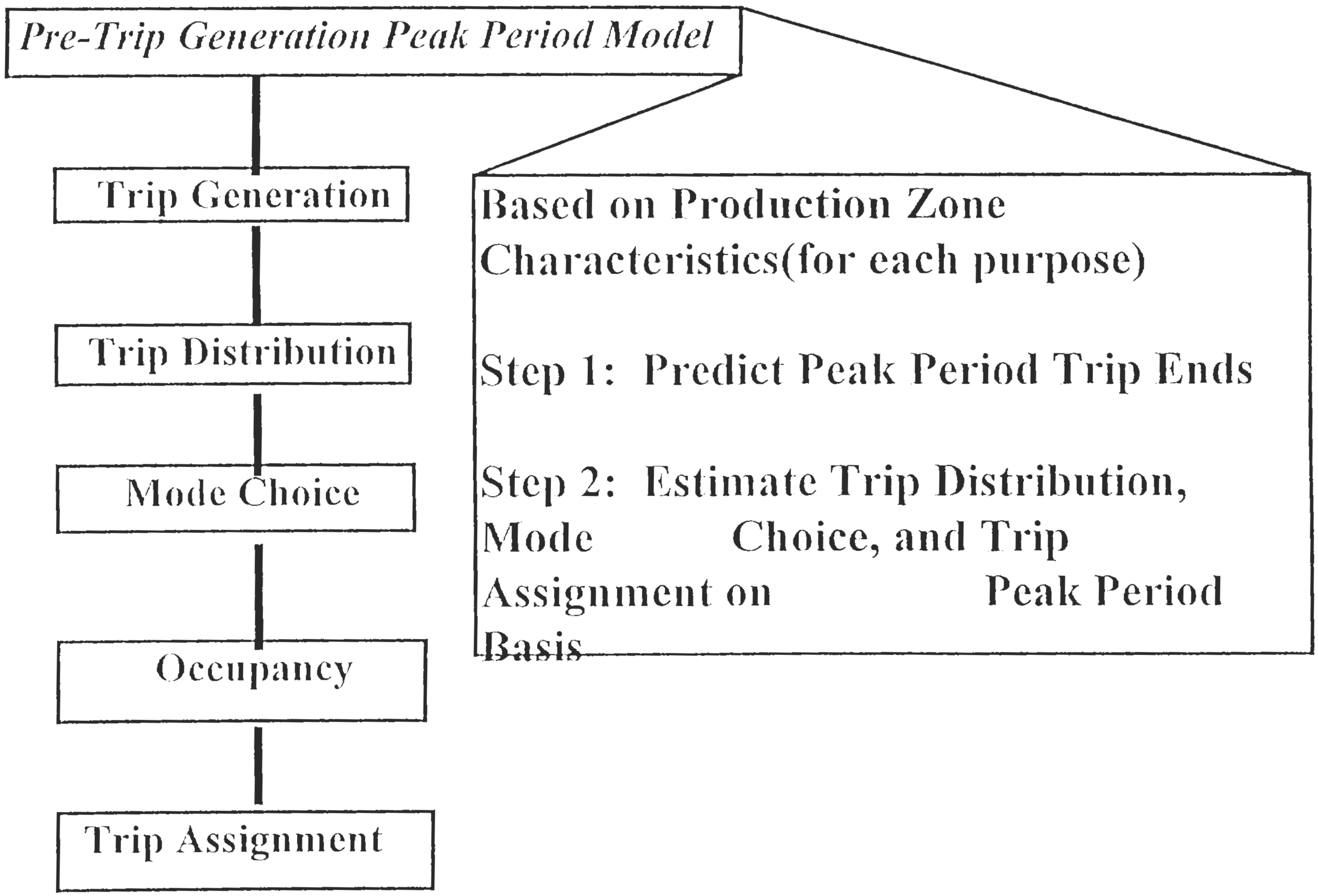

FIGURE 2. Pre-trip generation approach

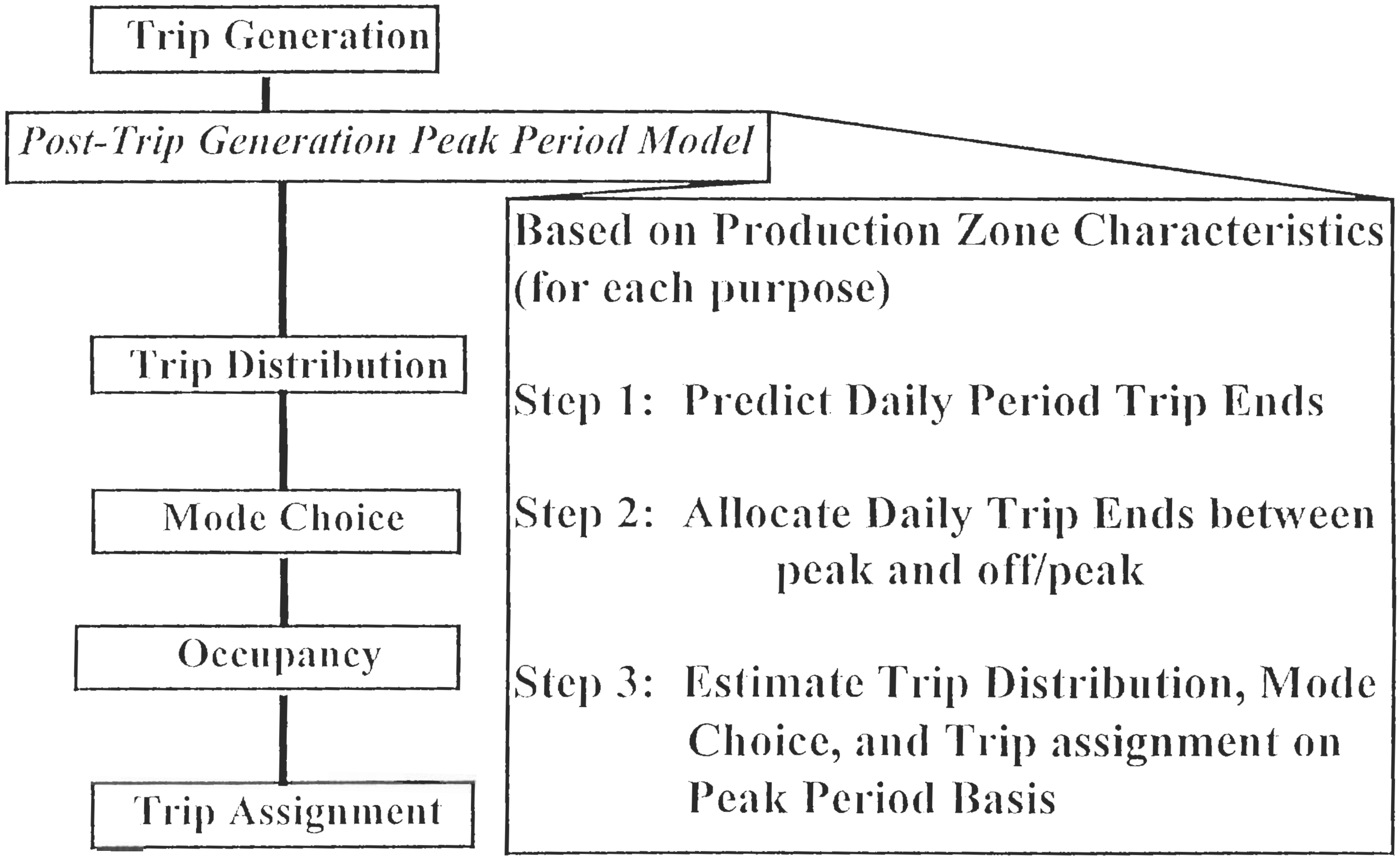

FIGURE 3. Post-Trip Generation Approach

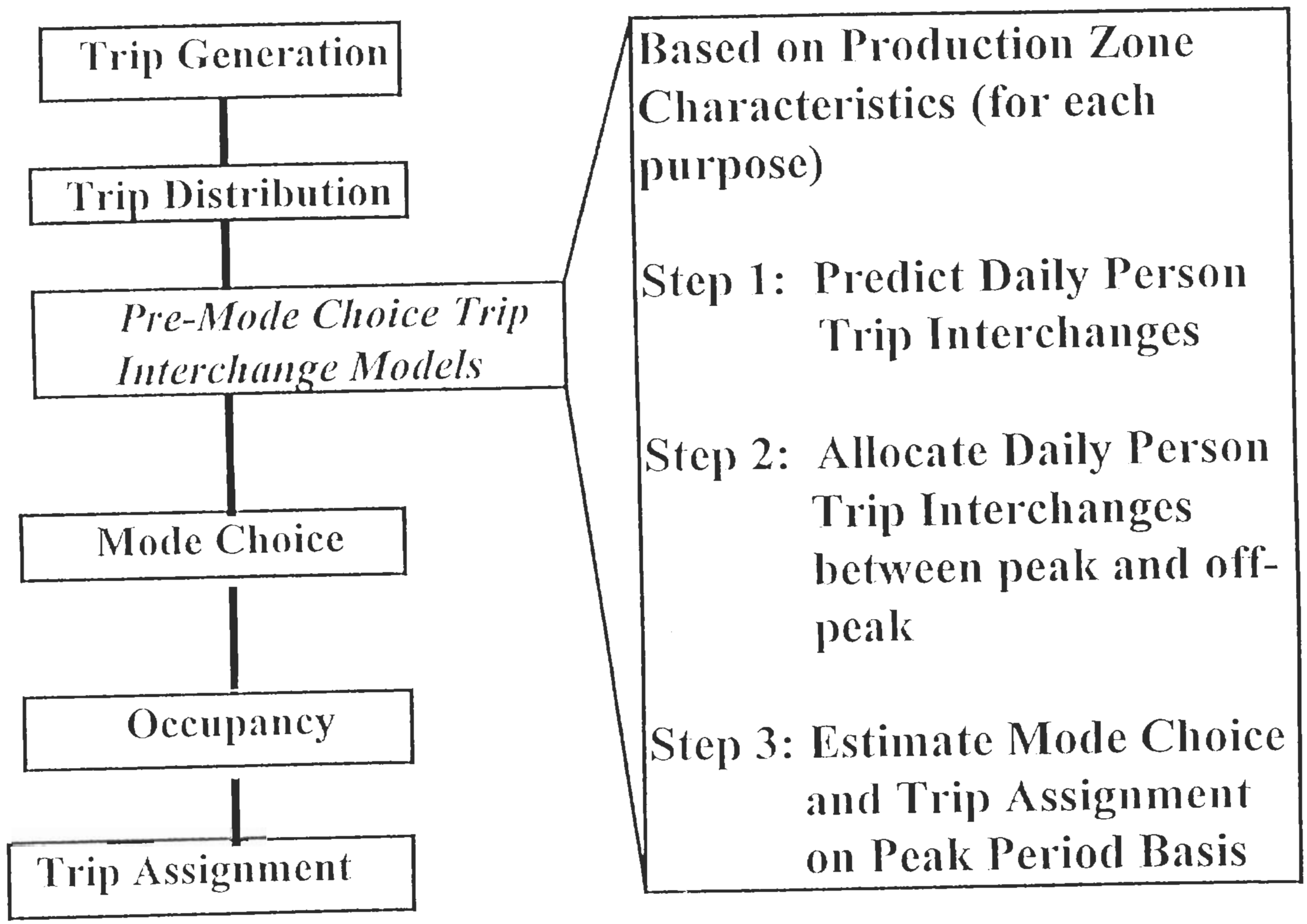

FIGURE 4. Pre-Mode Choice Method

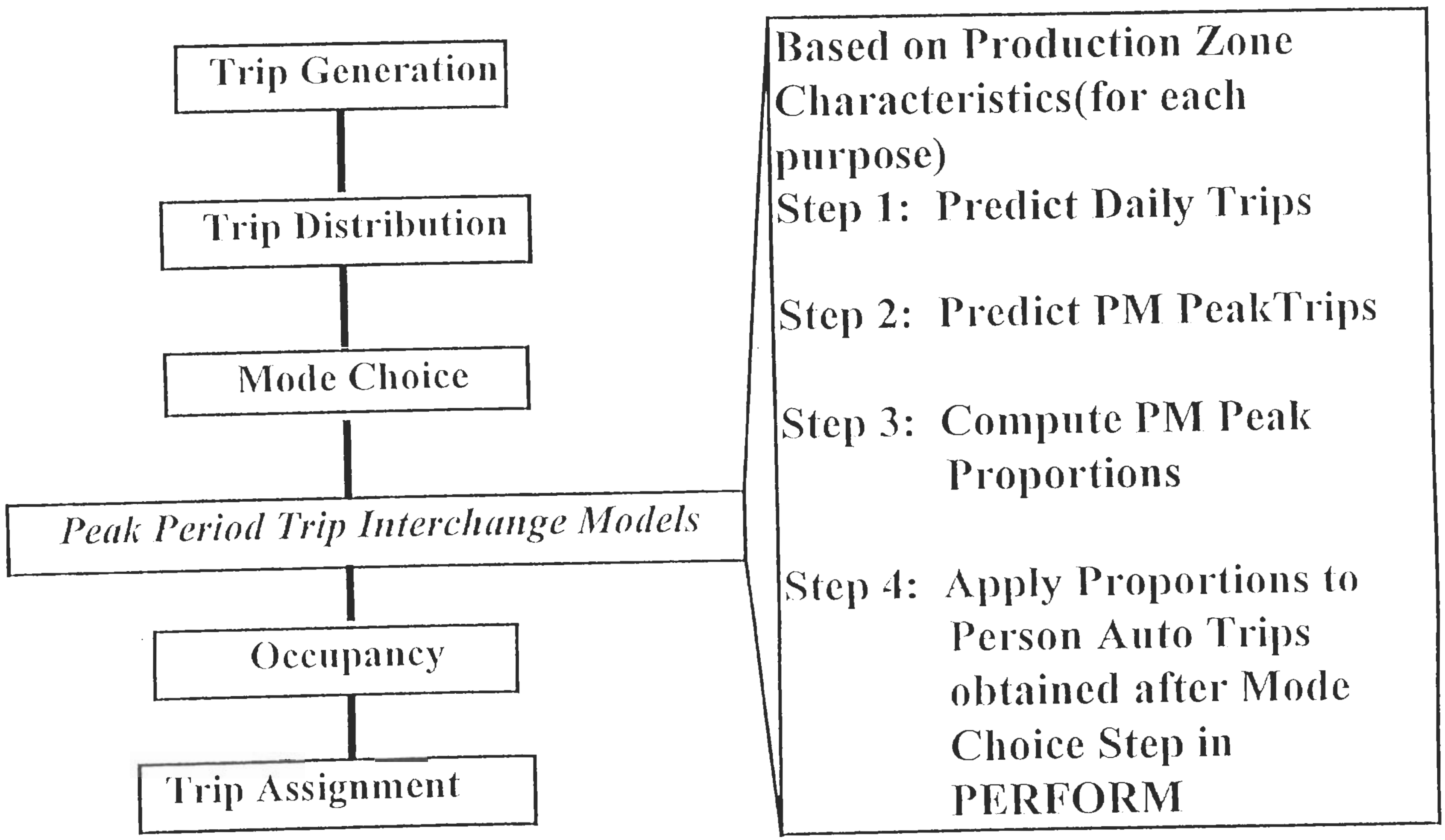

FIGURE 5. Post Mode Choice Method

distribution or mode choice may only consider production zone characteristics, and not characteristics of the trip or its attraction zone. In effect, these proportions could thus also be computed at the trip end stage. However, application at the trip interchange level incurs the least computational burden for PERFORM. This is because application at the trip end stage would result in roughly the same predictions but would require separate trip distribution and mode choice runs for peak as well as off-peak trips. In the fourth method, only trip assignment has to be repeated for peak and off-peak trips. Therefore, the fourth method (depicted on Figure 5) was selected for estimation of peak period models.

Direct prediction of PM peak period proportion

The statistical analysis software SPSS was used to estimate all models in this research. Initially, linear regression and cross classification models were investigated to directly predict the proportion of daily trips made during the PM peak period by trip purpose. However, this estimation effort was not successful. None of the independent variables were significant in the linear regression models, and an ANOVA table for evaluating cross-classification tables showed that even the best independent variables explained very little of the variation in the dependent variable (peak period proportion). This can be partially explained using Figures 6a and 6b, which show distributions of PM peak period proportion for HBO and NHB trips respectively, for each purpose. These figures show that PM peak proportion is not distributed evenly over the range from 0 to 1. In fact, the proportions are computed from discrete integers (number of PM peak trips) divided by other discrete integers (number of daily trips). This causes the "spikes" in the frequency plots at 0, 0.5, and 1.0 which may arise from numerous different combinations

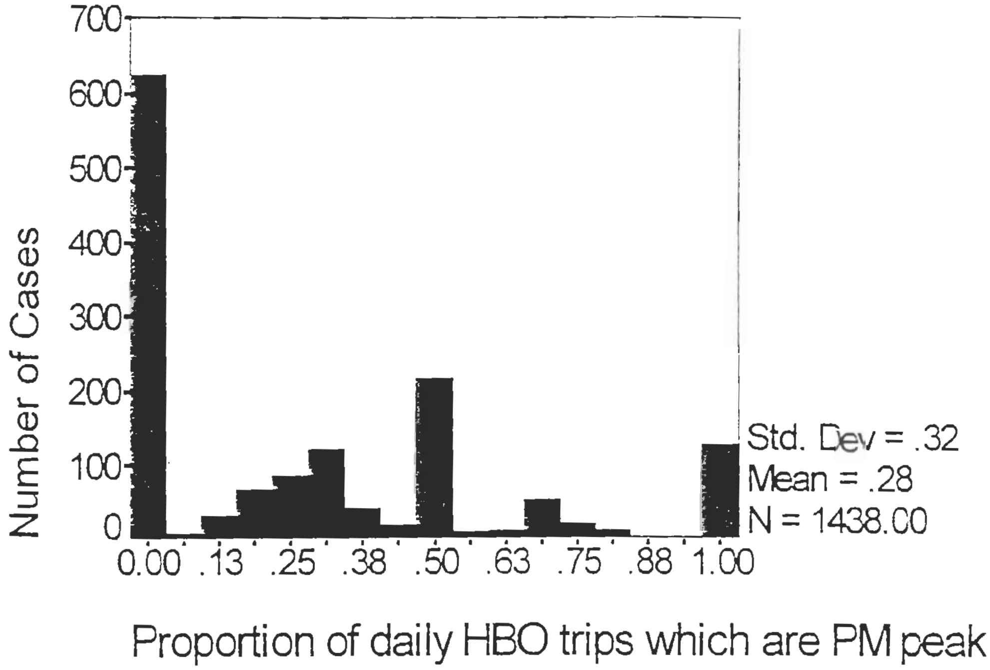

Figure 6a. PM peak trip distribution: HBO

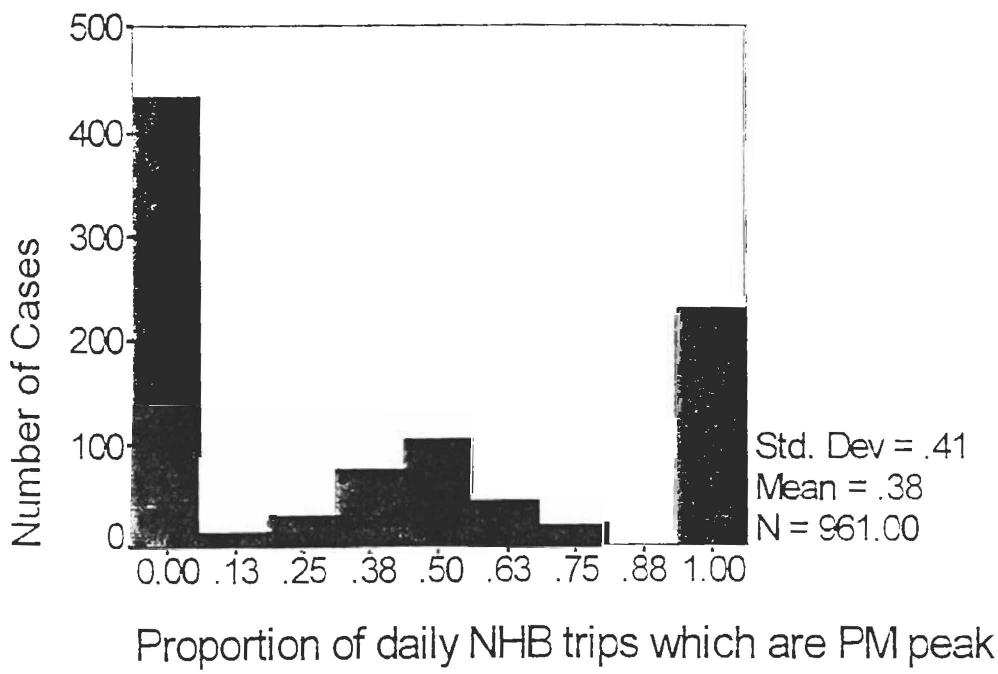

Figure 6b. PM peak trip distribution: NHB

of number of peak period and daily trips. As a result, it seemed that actual proportions were more related to the daily trip total than to the other explanatory variables.

Table 5 gives the results of the models estimated to directly predict the PM peak proportion for different purposes. For the HBW proportion, the most significant variables were SCHOOL_AGE and SHSF. However, the significance of t-statistics for these two variables were marginal. The F and R^2 statistics indicate poor explanation of sample variance. WKRCNT, which later proved to be significant in other work models, was very poorly correlated with HBW PM peak proportion. This could be attributed to the fact that observed PM peak proportions for HBW were almost constant at about 29 percent. Table 5 shows the same trend for the HBO PM peak proportion and could be due to the presence of "spikes" as discussed earlier. The models for NHB PM peak proportion were also similar to the other proposed models. Due to these pronounced defects, this method of predicting PM peak proportions was abandoned.

Figures 7a, 8a, and 9a show the results of the cross classification method applied to predict the proportion of PM peak trips for the three purposes. Different independent variables were tried in various one-way, two-way and three-way combinations to cross-classify the PM peak period proportion for different purposes. The best models among those experimented are reported here. For HBW the proportion of daily HBW trips which are PM peak are cross tabulated by different levels of WKRCNT. The analysis of variance statistics in Figure 7b for this model shows that WKRCNT is unable to explain the variation in the peak proportion (F-statistic is very high). Similarly, the proportion of HBO trips in the PM peak was two-way cross classified with WKRCNT and

Table 5. Direct prediction of PM peak proportion by purpose

Variables	PM Peak Proportion HBW	PM Peak Proportion HBO	PM Peak Proportion NHB
WKRCNT			-.06578
SCHOOL_AGE	-.00692	.00961	
PRE_SCHOOL			-.14049
HHSIZE			.14828
SHSF	4.04089e-06		
POPDENS			1.99562 e -06
NRET		2.31244 e-07	
CONSTANT	.29018	.17053	.08297
R^2	0.00163	.04874	.23022
F	0.92073 (sig. F = .3985)	1.95268 (sig. F = .1422)	15.76838 (sig. F = .0000)
$RMSE_Y^*$	.22630	.23907	.74703

PMP_HBW Proportion of Daily HBW trips which are PM Peak
by WKRCNT Number of Workers in the Household

Total Population

Proportion	.29
Number of Cases	(1129)

WKRCNT

0	1	2	3	4	5
.22	.29	.29	.28	.32	.26
(16)	(523)	(465)	(91)	(30)	(4)

Figure 7a. Cross classification table for PM peak proportion of daily HBW

PMP_HBW Proportion of Daily HBO trips which are PM Peak
by WKRCNT Number of Workers in the Household

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	.124	5	.025	.483	.789
WKRCNT	.124	5	.025	.483	.789
Explained	.124	5	.025	.483	.789
Residual	57.635	1123	.051		
Total	57.758	1128	.051		

2266 cases were processed.
1137 cases (50.2 pct) were missing.

Figure 7b. Analysis of variance for proportion of daily HBW in PM peak

PMP_HBO	Proportion of Daily HBO trips which are PM Peak
by WKRCNT	Number of Workers in the Household
SCHOOL_AGE	Number of Children in Household Age 5-21

Total Population

Proportion	.18
Number of Cases	(1641)

WKRCNT

	0	1	2	3	4	5
Proportion	.16	.19	.19	.22	.13	.09
Number of cases	(341)	(616)	(548)	(103)	(29)	(4)

SCHOOL_AGE

	0	1	2	3	4	5
Proportion	.17	.20	.21	.17	.18	.24
Number of cases	(902)	(269)	(318)	(121)	(27)	(4)

SCHOOL_AGE / WKRCNT	0	1	2	3	4	5
0	.15	.26	.18	.08	.00	.00
	(282)	(21)	(31)	(6)	(1)	(0)
1	.18	.19	.21	.21	.18	.00
	(310)	(105)	(133)	(50)	(17)	(1)
2	.18	.19	.22	.16	.23	.28
	(254)	(115)	(126)	(46)	(5)	(2)
3	.26	.19	.21	.15	.20	.00
	(45)	(23)	(20)	(13)	(2)	(0)
4	.06	.21	.21	.08	.11	.43
	(11)	(3)	(7)	(5)	(2)	(1)
5	.00	.18	.00	.00	.00	.00
	(0)	(2)	(1)	(1)	(0)	(0)

Figure 8a. Cross classification table for PM peak proportion of daily HBO

PMP_HBO Proportion of Daily HBO trips which are PM Peak
by WKRCNT Number of Workers in the Household
SCHOOL_AGE Number of Children in Household Aged 5-21

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	.818	10	.082	1.436	.158
WKRCNT	.454	5	.091	1.593	.159
SCHOOL_AGE	.264	5	.053	.926	.463
2-Way Interactions	.734	20	.037	.644	.881
WKRCNT SCHOOL_AGE					
Explained	1.552	30	.052	.908	.610
Residual	91.702	1610	.057		
Total	93.254	1640	.057		

2266 cases were processed.
625 cases (27.6 pct) were missing.

Figure 8b. Analysis of variance for proportion of daily HBO in PM peak

PMP_NHB	Proportion of Daily NHB trips which are PM Peak
by WKRCNT	Number of Workers in the Household
PRE_SCHOOL	Number of Persons in Household Age 0-4

Total Population
Proportion .43
Number of Cases (1132)

WKRCNT

0	1	2	3	4	5
.40	.41	.46	.45	.44	.55
(153)	(436)	(429)	(85)	(25)	(4)

PRE_SCHOOL

0	1	2	3	4	5
.43	.41	.56	.42	.00	.00
(948)	(142)	(36)	(6)	(0)	(0)

PRE_SCHOOL / WKRCNT	0	1	2	3	4	5
0	.41	.21	.00	.00	.00	.00
	(144)	(9)	(0)	(0)	(0)	(0)
1	.40	.44	.50	.30	.00	.00
	(345)	(65)	(21	(5)	(0)	(0)
2	.47	.40	.63	.00	.00	.00
	(355)	(59)	(15)	(0)	(0)	(0)
3	.43	.70	.00	1.00	.00	.00
	(78)	(6)	(0)	(1)	(0)	(0)
4	.46	.29	.00	.00	.00	.00
	(22)	(3)	(0)	(0)	(0)	(0)
5	.55	.00	.00	.00	.00	.00
	(4)	(0)	(0)	(0)	(0)	(0)

Figure 9a. Cross classification table for PM peak proportion of daily NHB

PMP_NHB Proportion of Daily NHB trips which are PM Peak
by WKRCNT Number of Workers in the Household
PRE_SCHOOL Number of Persons in Household Age 0-4

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	1.580	8	.198	.334	.953
WKRCNT	.974	5	.195	.329	.895
PRE_SCHOOL	.620	3	.207	.349	.790
2-Way Interactions	1.546	6	.258	.435	.855
WKRCNT PRE_SCHOOL					
Explained	3.127	14	.223	.377	.981
Residual	661.003	1117	.592		
Total	664.129	1131	.587		

2266 cases were processed.
1134 cases (50.0 pct) were missing.

Figure 9b. Analysis of variance for proportion of daily NHB in PM peak

SCHOOL_AGE. Again in this case, the variation in the dependent variable is insufficiently explained by the independent variables (Figure 8b). The same holds true for the NHB case where the independent variables are WKRCNT and PRE_SCHOOL (Figure 9b). The reason for poor performance of cross classification models is also due to the discontinuities and "spikes" in the values of the peak proportion.

Individual choice models

Apart from cross-classification and linear regression models, other statistical model forms estimated at the disaggregate level were individual choice and discriminant analysis models. The individual choice models were estimated using binomial logit models, which predicted the probability of making a particular trip in the peak versus the off-peak period given socio-economic, land use and congestion characteristics of the individual or household. The models developed using this model technique are depicted in Table 6.

The performance of these binomial logit models was unsatisfactory. In Table 6, it can be seen that the value of the statistic ρ^2 is very small for all three purposes. The statistic ρ^2 compares the log likelihood of the estimated model to the null model where all predictions are wrong. For the small number of observed cases which were in the PM peak, this model never predicted correctly. In fact, the model never predicts a PM peak trip for any case in the NPTS dataset. The reason for this is the skewness in the observed percentages of PM peak and non PM peak trips. Since most of the daily trips are observed in the non PM peak, the model predict all trips as non PM peak; and because the observed proportions of PM peak trip is very low, the model never predicts a PM peak trip.

Table 6. Individual choice model results

Variables	PM Peak HBW	PM Peak HBO	PM Peak NHB
A_RATIO (HBW)	-0.0046 (-0.0317)		
A_RATIO (HBO)		-0.0077 (-0.0194)	
A_RATIO (NHB)			0.9337 (0.0062)
SCHOOL_AGE	-0.0931 (-0.0494)		-0.897 (-0.0361)
NUMADULT	0.0068 (0.0451)		-0.0106 (-0.0412)
POVERTY	0.002 (0.0366)	-0.002 (-0.0246)	-0.0272 (-0.0298)
WKRCNT	0.0234 (0.0248)	-0.0733 (-0.0175)	0.189 (0.0217)
CONSTANT	0.0598 (0.0046)	-0.7286 (-0.0002)	0.039 (0.0043)
Log Likelihood (0)	1253.1275	3346.155	1610.564
ρ^2 Statistic	0.009	0.002	0.014
χ^2- Statistic	11.755	8.347	22.593
Degree of Freedom	1814	5525	2479

In a slight modification to the modeling technique, the dataset was curtailed to include only the PM trips. Now the choice was between PM peak and PM off-peak, rather than PM peak and daily non PM peak. In these models, the observed percentage of PM peak trips is high, and the model predicts substantial cases which are PM peak. However, the application of this model on an aggregate basis would be problematic since PERFORM does not predict PM trips. Due to these drawbacks observed in binomial logit models, further investigation into this modeling technique was not undertaken.

Discriminant choice models

Discriminant analysis, first introduced by Sir Ronald Fisher, is the statistical technique most commonly used to distinguish among several mutually exclusive groups based on a collection of variables. Linear combinations of the independent, or predictor, variables are formed and serve as the basis for classifying cases into one of the groups. A linear discriminant function of independent variables is estimated by maximizing the distance between the different groups. For each case, a discriminant score is then obtained which determines the class in which the case belongs. The technique used was based on Bayes's rule of conditional probability. The probability that a case with a discriminant score of 'D' belongs to group 'i' is estimated by:

$$P(G_i|D) = \frac{P(D|G_i)P(G_i)}{\sum_{i=1}^{g} P(D|G_i)P(G_i)}$$

where $P(G_i|D)$ = probability of the case being in group 'i' given the discriminant score 'D'.

$P(D|G_i)$ = conditional probability of 'D' given the group 'i'.

$P(G_i)$ = prior probability of a case being in group 'i' when no information is given.

g = number of groups 'i'

To compute the conditional probability $P(D|G_i)$, a case is assumed to belong to a particular group, and the probability of the observed discriminant score given membership in the group is estimated. This provides an idea of how likely the score is for members of a particular group. The prior probability $P(G_i)$, on the other hand, is a general estimate of the likelihood that a case belongs to a particular group. Most commonly, the observed proportions of cases in each group can serve as estimates of the prior probabilities.

In this research, the aim was to distinguish between two classes, PM peak trip and non PM peak trip, for each purpose. The prior probabilities were assumed to be the observed proportions of PM peak and non PM peak for a trip purpose. The conditional probabilities were internally computed in SPSS after formulating the discriminant function using the NPTS dataset. Figures 10, 11, 12 show the results for models estimated using this technique for HBW, HBO and NHB respectively. A look at the cross-classification table shows that the models are unsatisfactory. Again, as in the case of individual choice models, these models are also predicting all trips to be non PM peak. The reason for this poor performance is due to the values of the prior probabilities. Since these are taken as the observed proportions which are very small for PM peak non-work trips, the probability of making a PM peak trip is never greater than that of not making the trip. Hence, the model predicts no PM peak trips for any combination of independent variables.

On groups defined by PMP_HBW Flag Variable for Daily HBW trip being PM PEAK Number of cases by group

	Number of cases		
PMP_HBW	Unweighted	Weighted	Label
0	14119	14119.0	
1	969	969.0	
Total	15088	15088.0	

Canonical Discriminant Functions
Maximum number of functions............. 1
Minimum cumulative percent of variance... 100.00
Maximum significance of Wilks' Lambda.... 1.0000

Prior probabilities

Group	Prior	Label
0	.93578	
1	.06422	
Total	1.00000	

Standardized canonical discriminant function coefficients

	Func 1
PRE_SCHOOL	.23127
NRET	.00423
SCHOOL_AGE	-.72237
POPDENS	.07992
WKRCNT	.86802

Canonical discriminant functions evaluated at group means (group centroids)

Group	Func 1
0	-.02940
1	.42844

Figure 10. Discriminant analysis model output: PM peak HBW

Classification results -

Actual Group	No. of Cases	Predicted Group Membership 0	1
Group 0	14119	14192	0
		100.0%	.0%
Group 1	969	896	0
		100.0%	.0%

Percent of "grouped" cases correctly classified: 93.57%

Figure 10. Discriminant analysis model output: PM peak HBW (contd.)

On groups defined by PMP_HBO Flag Variable for Daily HBO trip being PM Peak

Number of cases by group

	Number of cases		
PMP_HBO	Unweighted	Weighted	Label
0	13464	13464.0	
1	1624	1624.0	
Total	15088	15088.0	

Canonical Discriminant Functions

Maximum number of functions............. 1
Minimum cumulative percent of variance... 100.00
Maximum significance of Wilks' Lambda.... 1.0000

Prior probabilities

Group	Prior	Label
0	.89236	
1	.10764	
Total	1.00000	

Standardized canonical discriminant function coefficients

	Func 1
PRE_SCHOOL	.29528
SCHOOL_AGE	-.88924
WKRCNT	.73149

Canonical discriminant functions evaluated at group means (group centroids)

Group	Func 1
0	.02367
1	-.19624

Figure 11. Discriminant analysis model output: PM peak HBO

Classification results -

Actual Group		No. of Cases	Predicted Group Membership 0	1
Group	0	13464	13464 100.0%	0 .0%
Group	1	1624	1624 100.0%	0 .0%

Percent of "grouped" cases correctly classified: 89.23%

Figure 11. Discriminant analysis model output: PM peak HBW (contd.)

On groups defined by PMP_NHB   Flag Variable for Daily NHB trip being PM Peak

Number of cases by group

	Number of cases		
PMP_NHB	Unweighted	Weighted	Label
0	14192	14192.0	
1	896	896.0	
Total	15088	15088.0	

Canonical Discriminant Functions

Maximum number of functions............. 1
Minimum cumulative percent of variance... 100.00
Maximum significance of Wilks' Lambda.... 1.0000

Prior probabilities

Group	Prior	Label
0	.94062	
1	.05938	
Total	1.00000	

Standardized canonical discriminant function coefficients

	Func 1
SCHOOL_AGE	.78483
POPDENS	.21137
WKRCNT	-.71807

Canonical discriminant functions evaluated at group means (group centroids)

Group	Func 1
0	.00995
1	-.15756

Figure 12. Discriminant analysis model output: PM peak NHB

Classification results -

Actual Group		No. of Cases	Predicted Group Membership 0	1
Group	0	14192	14192	0
			100.0%	.0%
Group	1	896	896	0
			100.0%	.0%

Percent of "grouped" cases correctly classified: 94.06%

Figure 12. Discriminant analysis model output: PM peak NHB (contd.)

Proposed Alternate Methodology

Owing to the difficulties in linear regression, cross classification and choice models, a two step method was developed for estimating PM peak proportion. The first step was to estimate parameters for predicting daily trips using socio-economic variables; the second to estimate parameters for predicting PM peak period trips using the same or other variables. The PM peak proportion is then computed by dividing the PM peak trip estimates by the daily trip estimates. This proportion is then applied to the daily auto person trip interchanges produced by the mode choice model to produce a PM peak auto person trip table for assignment. This is the procedure described earlier as "Post Mode Choice Model." Thus, the peak period proportion is defined as a ratio of two functions of socio-economic and congestion variables. Linear regression equations are calibrated for PM peak and daily trips using the NPTS dataset. The next chapter details the estimation of these models.

Chapter 6. MODEL ESTIMATION RESULTS

Statewide Models

Table 7 gives the parameters of the statewide models for different purposes which were calibrated using the NPTS dataset. The t-ratios for congestion variables (PUBTRANS, and A_RATIO) were not found to be significant in daily or the PM peak period models at ninety-five percent confidence. Other land use variables like NRET, RET, SHSF, POPLN, and POPDENS were also insignificant in the prediction of daily or PM peak trips for different purposes.

WKRCNT was the only variable significant in the daily and PM peak models for HBW. Except for households without any workers, the values obtained for PM peak proportion for households was almost constant at 29 percent as indicated below:

PM Peak Proportion of daily HBW trips by WKRCNT

WKRCNT	Estimated PM Peak Proportion	Observed PM Peak Proportion
0	0.156	0.22
1	0.287	0.29
2	0.291	0.29
3	0.292	0.28
4	0.293	0.32
5	0.293	0.26

Thus, for the HBW model, the PM peak trips are simply computed using the constant factors.

The proportions are almost the same except for households with no workers. The model correctly predicts the mean value of the proportion but the variation in the observed proportion is not captured well. A histogram plot of PM peak HBW proportion shows spikes at 0.0, 0.5 and 1.0 although the mean is around 0.29. Deviations from these values could then be attributed to the fact that some HBW trips are chained with NHB trips and therefore affect the proportion. Another factor in modeling work trips is that some workers work second or third shifts. In households with multiple workers, the likelihood of at least one working the second or third shift is greater. Therefore, their PM peak proportions for HBW are expected to be different from households with traditional workers (working at regular hours), and therefore the models fail to capture their peak period trip-making behavior. It was also seen in the above table that households with more workers have larger errors in their predictions.

Also, the choice of making a work trip in peak versus off-peak depends heavily on the flexibility in the work schedule of an individual. More and more employers are encouraging alternative commuting arrangements such as rideshare matching, telecommuting, and alternate work shifts to avoid travel in the peak period. Therefore, for peak period HBW trips, a dataset with other land use variables like labor force, employment classification, and work schedules could give better estimates of peak period work trips.

In the models estimated for HBO trip-making, the number of school age and pre-school children (SCHOOL_AGE and PRE_SCHOOL, respectively) and the number of workers (WKRCNT) were all found to be significant variables at ninety five percent confidence. WKRCNT and SCHOOL_AGE both have positive coefficients for both dailyand PM peak trip-making, but the coefficient on PRE_SCHOOL is negative. These values make intuitive sense due to the discretionary nature of HBO trips. A household

Table 7. Linear regression parameters for statewide models for different purposes

	HBW		HBO		NHB		HBO + NHB (Non Work)	
Variables	Daily	PM Peak	Daily	PM Peak	Daily	PM Peak	Daily	PM Peak
WKRCNT	1.19605	.353473	0.6988	0.1328	0.5560	0.1802	1.2317	.3229
SCHOOL_AGE			2.0651	0.4323	0.3742	0.0617	2.4712	.4092
PRE_SCHOOL			-1.7179	-0.3492			-1.8516	
FAMINC16					1.5573	1.037	3.0680	1.1540
FAMINC17					0.5539	0.3544	1.5872	.6034
CONSTANT	0.070076	.010911	1.9269	0.3885	0.6401	0.0871	2.0812	.3621
R^2	.35365	.22969	0.29532	0.1256	0.1273	0.08633	.29031	.14229
F	1238.763	675.01	236.88	81.17	65.94	53.41	189.42	93.77
$RMSE_Y^*$			3.4039	1.2169	2.2137	0.9043	4.5709	1.6169
$RMSE_{Peak\ Ratio}^*$			0.2395		0.3776		0.3129	

with multiple workers has more resources at its disposal and will therefore do more shopping and recreating. Similarly, the presence of school age children increases the maintenance requirements for the household. This coefficient dominates the other with respect to size. It is noted, however, that pre-school age children reduce the level of trip-making by the household, presumably because they do not generate any trips on their own, and also require constant supervision by an adult caretaker, making discretionary trips for that individual less convenient. Coefficients estimated for the PM peak model are all about one-fifth of the daily coefficients (including the constant), with the ratio for WKRCNT slightly smaller than the others. It would thus appear that large differences in peak trip proportions are explained by factors not available in the dataset. Both models also pass the F-test for explained variation at a significance level of ninety five percent confidence.

It is also interesting to observe in Figures 6a and 6b (presented earlier) that a significant number of proportions are different from 0, 0.5 and 1.0, indicating division by an odd number. This means that people making the HBO trips are chaining at least some of these trips with NHB trips in the peak periods. Several studies in the past decade have analyzed commuters' trip chaining behavior in which a stop for nonwork activities is introduced to the HBW pattern [23]. A study by Schultz and Allen [5] studied the temporal distribution of NHB trips and found out that the NHB trips peak between 4:30 and 6:30 PM, thus indicating that a majority of such trips are actually made on the way home from work. However, as explained earlier in the paragraph, results from current research support the idea that NHB trips are also chained to HBO trips. Many research studies have also concluded that increased participation of women in the labor force and the suburbanization of households are primary causes for increases in chained trips. With women making more and more non-work trips, it is not unreasonable to believe that they are chaining NHB trips with their HBO trips.

Apart from chained trips, many researchers have noticed changes in the total number of nonwork trips during the peak period over the last two decades. Gordon, et. al. [25] have noticed that increase in nonwork trips was higher in the peak compared to the off-peak periods. They also observed that the share of work trips in total trips declined between 1977 and 1983 in both AM and PM peak periods while the share of nonwork trips increased, both in AM and PM peak and off-peak periods. They say that the evidence of increasing congestion due to nonwork trips supports implementing congestion fees, because this is expected to have more effect on nonwork trips than commuting trips. Hence, detailed study of nonwork trip-making behavior in the peak period is very important in the present context.

For the NHB models, indicators for family incomes greater than seventy-five and eighty thousand dollars (FAMINC16 and FAMINC17) replace PRE_SCHOOL. WKRCNT and SCHOOL_AGE remain in the model, but in different proportions to one another, and between daily and PM peak. The income indicator coefficients are much greater in the PM peak model relative to the daily coefficients than are the coefficients on WKRCNT and SCHOOL_AGE. This suggests that the income effect is stronger in the PM peak period, i.e., higher income increases peak period trip chaining more than having more school children or workers in the family.

Also, due to the nature of a NHB trip <u>not</u> beginning or ending at the home of the individual making the trip, characteristics of neither the origin nor the destination zone may be used to estimate the proportion of trips made in the peak period. Therefore, it was decided to have a constant proportion for the urban regions which would be the average of the proportions estimated for all zones in those region. For the rural regions of the state, in which the zones are considerably larger in area, different proportions were used for each zone. These urban and rural regions are classified in detail in the next section.

Models were also estimated for a combined prediction of discretionary trips (HBO + NHB), which we called nonwork (NW) trips. The aim of this test was to see whether there are significant statistical differences in characteristics that determine HBO trips and NHB trips. These tests were done separately for PM peak and daily to capture any differences attributed to the timing of the trips. The formulation of the F-statistic for the daily models was as follows (some of the values are from HBO and NHB models in Table 7):

Restricted Model (Predicting HBO + NHB together):-

$$Y_{RES} = 2.0812 + 1.2317*WKRCNT + 2.4712*SCHOOL\_AGE - 1.8516*PRE\_SCHOOL + 3.0679*FAMINC16 + 1.5872*FAMINC17$$

Unrestricted Model (Predicting HBO and NHB Separately)

$$Y_{UNR} = 1.9269 + 0.6988*WKRCNT + 2.0651*SCHOOL\_AGE - 1.7179*PRE\_SCHOOL + 0.6401 + 0.5560*WKRCNT + 0.3742*SCHOOL\_AGE + 1.5573*FAMINC16 + 0.5539*FAMINC17.$$

The F-statistic for the test was:

$$F^* = \frac{(SSE_{RES} - SSE_{UNR}) / r}{SSE_{UNR} / (N - K)}$$

where r = number of restrictions in the test

K = number of parameters

The results of the F-tests between NW and separate models (HBO & NHB) suggested significant differences among models for HBO and NHB. Therefore, separate models for HBO and NHB were selected rather than a combined model.

Regional Models

Connecticut is a diverse state with the prevailing land use densities varying widely from one area to another. It was thus felt that peak period trip-making could conceivably vary by region as well as by household characteristics. This hypothesis was tested by dividing the state into three regions: (a) Capital Region, comprising towns in the Hartford metropolitan planning area: (b) Southwest Connecticut Region, comprising towns in Fairfield and New Haven counties along the I-95 corridor from New Haven to the border with New York State; and (c) Rural region, comprising the rest of the state which is generally rural or very low density suburban with several small cities.

The basis for the segregation of the dataset was the differences in land use characteristics among the different areas. Table 8 gives the average values of land use characteristics in the three areas. The mean household size is almost the same for all three regions. However, average values for population, population density, retail and non-retail employment, and retail space are significantly different from one region to the other. The Southwest region has the largest concentration of population and the rural region the least dense. Regarding retail and non-retail employment and retail space, the densest region is the Capital region, while the rural region is the least. There are significant differences in the accessibility index ratios, too, for all purposes. These differences warranted investigation into the need to have segregated peak proportion models for different areas in the state.

To test this hypothesis, models to predict PM peak period proportion were estimated for these three segregated regions for each purpose. These models were compared to the models estimated using data from the entire state, by trip purpose. These results are given in Tables 9, 10 and 11.

Table 8. Summary of land use data for different regions for the state

	Capital Region		Southwest Region		Rural Region	
Variable	Mean Value	S.D	Mean Value	S.D	Mean Value	S.D
A_RATIO(NHB)	0.19	0.14	0.09	0.04	0.08	0.06
A_RATIO(HBO)	0.92	0.49	1.19	0.56	1.05	0.93
HHSZ	2.60	0.18	2.65	0.23	2.62	0.14
A_RATIO(HBW)	3.01	0.54	5.3	2.27	8.80	3.72
SHSF	1432.4	1337.24	860.65	978.42	1087.93	777.17
POPDENS	2400.21	2805.42	1477.8	1964.46	2914.07	2551.03
RET	4709.95	3398.71	2746.64	2840.62	4656.85	2593.93
HHLD	18342.72	14400.68	11500.85	11111.30	23424.46	15592.62
NRET	28227.85	38132.64	12630.26	11938.84	25691.42	21088.68
POPLN	47300.58	38100.56	32523.71	29509.86	60765.26	40699.29

Table 9. Linear regression parameters for home based work (HBW) models by region

	Entire State		Capital Region		Southwest Region		Rural Region	
Variables	Daily	PM Peak	Daily	PM Peak	Daily	PM Peak	Daily	PM Peak
WKRCNT	1.196047	.353473	1.147276	.321412	1.291511	.385783	1.138104	.338632
SCHOOL_AGE								
PRE_SCHOOL								
FAMINC16								
FAMINC17								
CONSTANT	0.070076	.010911	.083095	.026286	.013345	-.013985	.11316	.025704
R^2	.35365	.22969	.37752	.20959	.39491	.25914	.33151	.21432
F	1238.763	675.01	216.239	117.73544	563.591	301.564	473.099	260.229
$RMSE_Y$*	2.0795	0.8169	1.968	0.7605	1.7001	0.7803	1.8933	0.8131
$RMSE_{Peak\ Ratio}$*	0.2395		0.2831		0.2406		0.2275	

* Two pairs of RMSE have been estimated. The pair on the top represents RMSE for the corresponding models. The pair on the bottom represents RMSE computed when the entire state model was applied to that region

Table 10. Linear regression parameters for home based other (HBO) models by region

	Entire State		Capital Region		Southwest Region		Rural Region	
Variables	**Daily**	**PM Peak**	**Daily**	**PM Peak**	**Daily**	**PM Peak**	**Daily**	**PM Peak**
WKRCNT	0.6988	0.1328	0.8072		0.8008	0.1704	0.6768	0.1374
SCHOOL_AGE	2.0651	0.4323	1.423	0.18656	2.1382	0.5161	2.0625	0.4356
PRE_SCHOOL	-1.7179	-0.3492			-1.625	-0.42217	-1.7329	-0.2942
FAMINC16								
FAMINC17								
CONSTANT	1.9269	0.3885	1.753	0.6213	1.999	0.3794	1.6351	0.2700
R^2	0.29532	0.1256	0.255	0.044	0.2979	0.1469	0.2966	0.14412
F	236.88	81.17	50.55	10.254	91.02	36.98	133.82	53.433
$RMSE_Y$*	3.4039	1.2169	3.268	1.1605	3.6001	1.3403	3.2933	1.1131
$RMSE_{Peak\ Ratio}$*	0.2395		0.2831		0.2406		0.2275	

* Two pairs of RMSE have been estimated. The pair on the top represents RMSE for the corresponding models. The pair on the bottom represents RMSE computed when the entire state model was applied to that region

Table 11. Linear regression parameters for non home based (NHB) models by region

	Entire State		Capital Region		Southwest Region		Rural Region	
Variables	**Daily**	**PM Peak**	**Daily**	**PM Peak**	**Daily**	**PM Peak**	**Daily**	**PM Peak**
WKRCNT	0.5560	0.1802	0.6832	0.2608	0.6132	0.1602	0.4788	0.1867
SCHOOL_AGE	0.3742	0.0617	0.3432		0.4013	0.077	0.3642	
PRE_SCHOOL								
FAMINC16	1.5573	1.037			1.9043	1.084	2.0587	1.3886
FAMINC17	0.5539	0.3544			0.6252	0.423	0.7073	0.2620
CONSTANT	0.6401	0.0871	0.3254	0.0117	0.4542	0.0853	0.6343	0.1534
R^2	0.1273	0.08633	0.1393	0.1127	0.1376	0.096	0.1083	0.0769
F	65.94	53.41	35.84	56.44	34.28	22.78	28.86	26.43
$RMSE_y$*	2.2137	0.9043	2.017	0.6923	2.259	0.924	2.271	0.969
$RMSE_{Peak\ Ratio}$*	0.3776		0.3141		0.3797		0.3732	

* Two pairs of RMSE have been estimated. The pair on the top represents RMSE for the corresponding models. The pair on the bottom represents RMSE computed when the entire state model was applied to that region

The F-test was formulated according to the following example (testing differences in PM Peak HBO trips between Capital Region and Southwest Region, Table 9):

Restricted Model (Predicting same HBO for both regions and computed with cases from these two regions only):-

$$Y_{RES} = 0.319808 + 0.148931*WKRCNT + .428872*SCHOOL\_AGE - 0.36504*PRE\_SCHOOL$$

Unrestricted Model (Predicting HBO separately for each region using dummy variables; computed with cases from these regions only):-

$$Y_{UNR} = 0.381237*REGION1 + 0.274808*REGION2 + 0.112686*WKRCNT1 + 0.181584*WKRCNT2 + 0.238247*SCHOOL\_AGE1 + 0.51625*SCHOOL\_AGE2 - 0.275212*PRE\_SCHOOL1 - 0.40635*PRE\_SCHOOL2$$

The F-statistic for the test was:

$$F^* = \frac{(SSE_{RES} - SSE_{UNR}) / r}{SSE_{UNR} / (N - K)}$$

where r = number of restrictions in the test

k = number of parameters

Table 12 gives the values for F-tests for all possible 2-way and 3-way comparisons between models for different regions. For daily trips, the models for different regions are not statistically different for any trip purpose. For the PM peak period, the models for

Table 12. F-test statistics for models compared for different regions

Regions Compared	F(PM Peak HBO)	F (PM Peak NHB)	F (PM Peak NW)
Capital vs. Southwest	**3.73 (2.21)**	1.71 (2.1)	**2.177 (2.01)**
Capital vs. Rural	**2.436 (2.21)**	**20.86 (2.1)**	1.77 (2.01)
Southwest vs. Rural	1.1845 (2.21)	**31.35 (2.1)**	1.16 (2.01
Capital vs. (Southwest + Rural)	**3.56 (2.21)**	4.35 (2.1)	**2.02 (2.01)**
Southwest vs. (Capital + Rural)	**2.543 (2.21)**	**2.397 (2.1)**	1.623 (2.01)
Rural vs. (Capital + Southwest)	0.437 (2.21)	**3.438 (2.1)**	0.963 (2.01)
Capital vs. Southwest vs. Rural	**2.543 (2.21)**	**2.483 (1.75)**	1.615 (1.7)

Notes:

1. F-statistics in **boldface** suggest significant difference between models estimated for indicated regions (or region groups).
2. The numbers in parentheses are the 95th percentile F-values that were compared to the F-statistics for testing significant differences amongst the segmented models.

different regions show significant statistical differences. The Capital region differs from the rural region as well as the South-West region for both PM peak HBO and PM peak NHB models. The models for the Southwest and the Rural regions do not seem to differ statistically for either NHB or HBO. This result seemed contradictory since it was expected that the Southwest region of the state should differ from the rest of the state owing to its proximity to New York City which has a significant effect on this region's trip-making. However, the comparisons of the root mean square error (for peak proportion predictions) for the segmented models (see Tables 9, 10, 11) do not show improvement over the model for the entire state. Nevertheless, there was no definite conclusion as to whether the segregated models would perform better than a statewide model. In consultation with project sponsors, it was decided to use statewide models rather than segmented models to minimize the computational burden.

Chapter 7: IMPLEMENTATION IN PERFORM

Figure 5 illustrates the method for incorporating the proposed peak period proportion models into PERFORM. It is again noted that the PM peak period proportions have been estimated using NPTS and land use data but they are applied to daily models in PERFORM which are estimated using data from a household survey conducted in 1976. The peak period models estimated using the NPTS dataset are disaggregate models; however PERFORM uses aggregate variables for its model forecasts. NPTS data are not coded with traffic analysis zones, so the dataset was aggregated by zip code to test the accuracy of the PM peak proportion models for aggregate predictions.

The "average household" aggregation method was used, where the disaggregate model coefficients were multiplied by the number of households in the zip code and then applied to the average values for each variable for that zip code. Table 13 shows a sample comparison of aggregate prediction with observed values. While observed PM peak proportions have zero values, the predicted proportions can never assume zero values. This is due to the non-zero constants in the estimated models for daily and PM peak. Otherwise, the proportions seem to follow the observed trend to a reasonable accuracy. Histograms of the errors in PM peak proportions for HBO and NHB (Figures 13a, 13b) were heavily concentrated at zero showing that the models predict well for most cases.

The PM peak period trip tables generated after applying peak proportions are in production/attraction (P/A) format. However, we need to transpose the P/A matrices into

Table 13. Sample application of the model to aggregate data

Zip Code	Number of Households	Daily Home Based Other Trips		PM Peak Home Based Other Trips		PM Peak Proportion of Daily Home Based Other Trips	
		Observed	Predicted	Observed	Predicted	Observed	Predicted
06001	10	41.00	31.10	5.00	5.91	.12	.19
06002	14	40.00	40.23	6.00	7.75	.15	.19
06010	39	166.00	154.75	25.00	30.40	.15	.20
06013	5	28.00	20.74	5.00	4.02	.18	.19
06015	1	.00	1.14	.00	.17	.	.15
06016	6	20.00	27.24	3.00	5.37	.15	.20
06018	1	2.00	2.60	.00	.51	.00	.20
06019	3	12.00	18.83	3.00	3.84	.25	.20
06022	1	1.00	3.24	1.00	.57	1.00	.18
06023	1	5.00	5.37	1.00	1.08	.20	.20
06026	1	.00	2.60	.00	.51	.	.20
06027	2	4.00	10.73	.00	2.16	.00	.20
06029	7	25.00	32.32	4.00	6.51	.16	.20
06030	1	.00	1.14	.00	.17	.	.15
06032	9	32.00	27.53	6.00	5.26	.19	.19
06033	14	26.00	54.31	5.00	10.61	.19	.20
06035	7	10.00	22.58	1.00	4.43	.10	.20

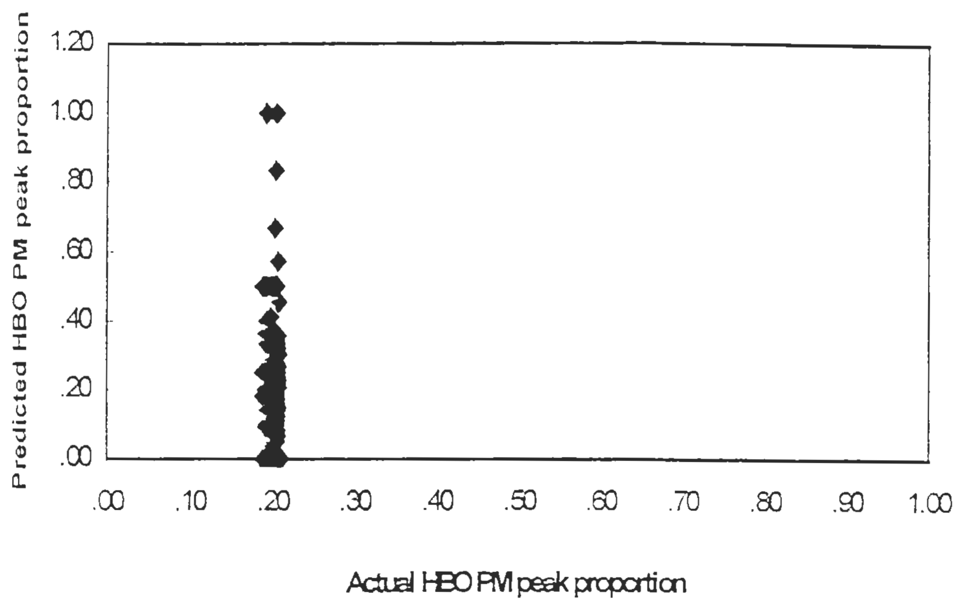

Figure 13a. Error in aggregate prediction of PM Peak proportion: HBO

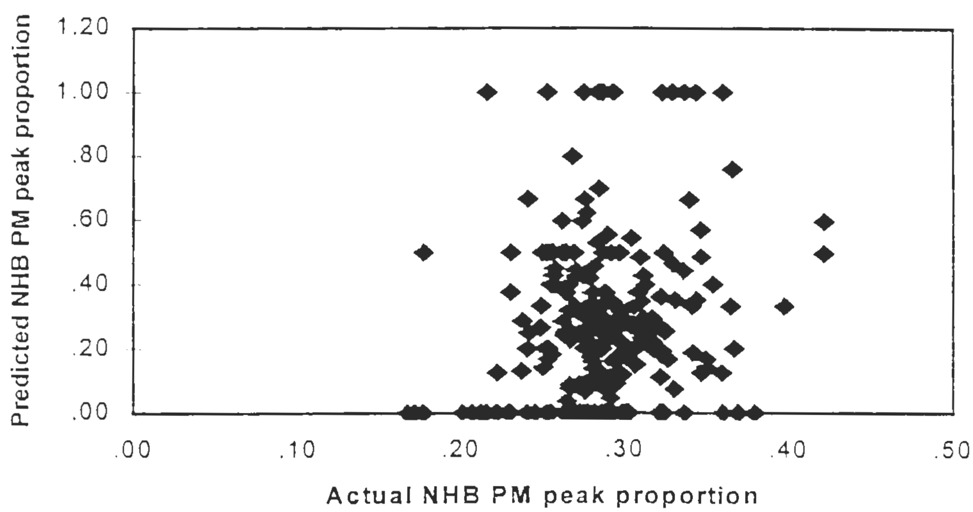

Figure 13b. Error in aggregate prediction of PM peak proportion: NHB

origin/destination (O/D) matrices before they can be used for trip assignment in PERFORM. In PERFORM, the diurnal splits used are 0.5 and 0.5 for converting daily trip tables in P/A format to O/D format. Most daily trips can be considered round trips, so this assumption is valid. However, in a peak period, the trips are not expected to be balanced by direction, due to the nature of afternoon versus morning trips. PM peak splits were computed from the NPTS data using information regarding the origin zone of each trip, the time of the trip, and the purpose of the trip. These splits were computed for the three different purposes. For HBW, the diurnal splits for PM peak is expected to be skewed towards the inbound (work to home) direction, since most of the people return home during this peak period, rather than going to work. In the case of HBO, the splits are expected to be more balanced because many HBO trips are expected to be round trips within the peak period. Some people who chain their work trips while coming home are making an inbound HBO trip, while others who are going for shopping, social recreation, or sports during the PM peak period are making an outbound HBO trip.

Individual trip information in NPTS was used to compute percentage directional splits for HBW and HBO trips in the PM Peak period. In the case of NHB trips the percentage splits were not computed because the NHB peak proportion model is applied to a daily NHB trip table already in O/D format. The computed directional factors are 0.07 from home and 0.93 to home for HBW trips, and 0.41 from home and 0.59 to home for HBO trips. These factors are used in the TRANPLAN matrix inversion program as row and column factors.

Chapter 7: CONCLUSIONS

This research was conducted to study methods to improve the modeling of peak period trip-making. Over the last few years, congestion and air quality problems have caught the attention of many and federal regulations have been enacted to reduce their magnitude. For example, the adoption of CAAA in 1990 and ISTEA in 1991 call for improvements in travel demand modeling to include changes in travel trends due to congestion. This context increases the value of peak period trip modeling improvements.

The significance of land use and congestion variables in daily and peak period travel modeling was explored using Connecticut NPTS and landuse data. Models for predicting PM peak period trips were estimated using household socio-economic data from Connecticut NPTS records. Socio-economic variables were found to be quite significant, while congestion and land-use variables for the location of the household were not. Incorporation of these models into the Connecticut statewide travel demand forecasting model has been demonstrated using a procedure that can be easily implemented in any four- step travel demand forecasting process. The resulting procedure is expected to significantly improve the accuracy of peak period trip estimation over procedures that use fixed percentages to extract peak period trip tables or peak period link volumes.

Linear regression models and cross-classification models to directly predict the proportion of peak trips did not perform well. Hence, a two step methodology was selected. Peak period models were developed using the proposed methodology for different purposes and results were examined in detail. For HBW models, the peak proportion did not vary significantly with any of the variables available in the destination dataset. Hence, the conventional method of extracting a fixed percentage from daily trips for peak period trips was adopted for HBW trips. The choice of making a work trip in peak versus off-peak under congested conditions depends heavily on the flexibility in the work schedule of an individual. More and more employers are moving to different work strategies like rideshare matching, telecommuting, alternate work shifts, etc. to avoid travel in the peak period. Therefore, for peak period HBW trips by different categories, a dataset with other land use variables like labor force, employment classifications, and work schedules could give even better estimates of peak period work trips.

For non-home based (NHB) and home-based-other (HBO) models, the PM peak period proportions depend significantly on other socio-economic characteristics of the household/individual besides WKRCNT. Variables like SCHOOL_AGE, PRE_SCHOOL and some family income categories are also significant for these models. These proportions vary by household characteristics and we cannot use constant values. The results also provide evidence of more and more discretionary trips in the peak period being chained. Moreover, there are significant statistical differences between NHB and HBO models and these ought to be modeled separately.

PM peak proportion models were estimated for three different Connecticut regions with varying land use intensities. Some of the estimated models were statistically different in these three areas suggesting different trip-making behavior, but not in a manner that could be easily explained. It was difficult to generalize, based on these models, that segregated models would perform better than a statewide model. Therefore, in the absence of clear understanding of this phenomenon, and in order to save computational time, a single model was adopted for each purpose for the entire state.

Due to the lack of information regarding the destination end of trips in the NPTS data, the models are restricted to the characteristics of the production zone. Since proportions are applied to trip interchanges, models which have information about both ends of the trip could improve the prediction of proportions and hence peak period forecasting.

These results also shed light onto the applicability of different statistical modeling techniques. It was very difficult to directly predict the peak period proportions and therefore a two step methodology was used. With the NPTS dataset used, linear regression models gave the closest estimates to the observed values. The complexities in peak period travel modeling were difficult to be captured by one-way or two-way cross classification tables. Individual choice and discriminant analysis methods performed badly because the prior probabilities for peak period trips were negligible.

The results arising from this approach pave the way towards new improvements in peak spreading estimation. It is expected that an improvement in peak period modeling would significantly improve the modeling of trips in the peak hour within the peak period. Some other areas bound to benefit from this research are modeling air quality and studies on congestion mitigation since these models are based on peak period forecasts. In future, the models can be greatly enhanced by considering discretionary effects on work trips, and by using datasets which provides information on both ends of the trips.

Appendix 1. Computation of Peak Proportions

This contains the FORTRAN code for the computation of PM peak proportions for the three purposes using socio-economic data. The coefficients for daily and PM Peak models are provided in the file "COEFFS.DAT" and the socio-economic variables in "SOCIO.DAT." The program outputs the proportion in a matrix form which is later used by the MATRIX UPDATE procedure in TRANPLAN to extract the percentages from the daily trip tables.

```
PROGRAM PROP
    REAL COEFFS(6,6)
    REAL PMPHBO(1248), PMPNHB(1248), DLYHBO(1248)
    REAL PMPHBW(1248), DLYHBW(1248)
    REAL DLYNHB(1248), PHBO(1248), PHBW(1248), PNHB(1248)
    INTEGER ZONE(1248)
    REAL ZLF(1248), ZP521(1248), ZP04(1248)
    REAL ZI7580(1248), ZIGT80(1248)
    OPEN (UNIT=9,FILE='COEFFS.DAT',STATUS='OLD')
    READ (9,*) COEFFS
    OPEN (UNIT=10,FILE='SOCIO.DAT',STATUS='OLD')
    OPEN (UNIT=20,FILE='MUPDATA.HBO',STATUS='OLD')
    OPEN (UNIT=25,FILE='MUPDATA.HBO',STATUS='OLD')
    OPEN (UNIT=30,FILE='MUPDATA.NHB',STATUS='OLD')
    DO 20 I=1,1248
      READ (10,*) ZONE(I), ZLF(I), ZP521(I), ZP04(I), ZI7580(I),
   C          ZIGT80(I)
      PMPHBW(I) = COEFFS(1,1)+COEFFS(2,1)*ZLF(I)+
   C          COEFFS(3,1)*ZP521(I)+COEFFS(4,1)*ZP04(I)+
   C          COEFFS(5,1)*ZI7580(I)+COEFFS(6,1)*ZIGT80(I)
      DLYHBW(I) = COEFFS(1,2) + COEFFS(2,2)*ZLF(I) +
   C          COEFFS(3,2)*ZP521(I) + COEFFS(4,2)*ZP04(I) +
   C          COEFFS(5,2)*ZI7580(I) + COEFFS(6,2)*ZIGT80(I)
      PMPHBO(I) = COEFFS(1,3) + COEFFS(2,3)*ZLF(I) +
   C          COEFFS(3,3)*ZP521(I) + COEFFS(4,3)*ZP04(I) +
   C          COEFFS(6,3)*ZI7580(I) + COEFFS(5,3)*ZIGT80(I)
      DLYHBO(I) = COEFFS(1,4) + COEFFS(2,4)*ZLF(I) +
   C          COEFFS(3,4)*ZP521(I) + COEFFS(4,4)*ZP04(I) +
   C          COEFFS(5,4)*ZI7580(I) + COEFFS(6,4)*ZIGT80(I)
```

```
C           COEFFS(3,5)*ZP521(I) + COEFFS(4,5)*ZP04(I) +
C           COEFFS(5,5)*ZI7580(I) + COEFFS(6,5)*ZIGT80(I)
   DLYNHB(I) = COEFFS(1,6) + COEFFS(2,6)*ZLF(I) +
C           COEFFS(3,6)*ZP521(I) + COEFFS(4,6)*ZP04(I) +
C           COEFFS(5,6)*ZI7580(I) + COEFFS(6,6)*ZIGT80(I)

   PHBO(I) = PMPHBO(I)/DLYHBO(I)
   PHBW(I) = 0.29
   PNHB(I) = PMPNHB(I)/DLYNHB(I)
   WRITE (20,60) I, PHBW(I)
   WRITE (25,60) I, PHBO(I)
   WRITE (30,60) I, PNHB(I)
60 FORMAT (2X,'T1, ',I4,', 1-1248, *',F10.2)
20 CONTINUE
   END
```

Appendix 2a. Extraction of PM Peak Trip Table: HBW

The file "95HBW.130" is the 1300 zone daily HBW trip table in P/A format for Connecticut for the year 1995. The output file is "PMPHBW.PA" which is the PM Peak HBW trip table in P/A format

```
$MATRIX UPDATE
$FILES
   INPUT FILE = UPDIN, USER ID = $95HBW.T1$ , UNLOAD
   INPUT FILE = MUPDATA, USER ID = $MUPDATA.HBW$
   OUTPUT FILE = UPDOUT , USERID = $95HBO.PA$
$HEADERS
   MATRIX EXRACTION FOR HBW PM PEAK
   EXTRACT PM PEAK HBW TRIPS FROM DAILY HBW
$OPTIONS
   MUPDATA
   PRINT TRIP ENDS
$END TP FUNCTION
```

Appendix 2b. Extraction of PM Peak Trip Table: HBO

First, the daily trip table, which is converted from purpose 2 to purpose 1 for performing the MATRIX UPDATE. The file "95HBO.130" is the 1300 zone daily HBO trip table in P/A format for Connecticut for the year 1995. The output file is "PMPHBO.PA" which is the PM Peak HBO trip table in P/A format.

```
$MATRIX MANIPULATE
$FILES
INPUT FILE =  TMAN1, USER ID =  $95HBO.130$
OUTPUT FILE =  TMAN2, USER ID =  $95HBO.T1$
$HEADERS
          CONVERTING FROM PURPOSE 2 TO 1
$DATA
  TMAN2.T1 =  TMAN1,T2
$END TP FUNCTION
$MATRIX UPDATE
$FILES
   INPUT FILE = UPDIN, USER ID = $95HBO.T1$ , UNLOAD
   INPUT FILE = MUPDATA. USER ID = $MUPDATA.HBO$
   OUTPUT FILE = UPDOUT , USERID = $PMPHBO.PA$
$HEADERS
   MATRIX EXRACTION FOR HBO PM PEAK
   EXTRACT PM PEAK HBW TRIPS FROM DAILY HBW
$OPTIONS
   MUPDATA
   PRINT TRIP ENDS
$END TP FUNCTION
```

Appendix 2c. Extraction of PM Peak Trip Table: NHB

For the case of NHB, the daily trip table is already in the O/D format. Hence average values of the PM Peak Proportions are multiplied by the daily trips to get PM peak trip table. The file "95NHB.130" is the 1300 zone daily trip table in O/D format corresponding to the base year for Connecticut

```
$MATRIX MANIPULATE
$FILES
INPUT FILE =  TMAN1, USER ID =  $95NHB.130$
OUTPUT FILE =  TMAN2, USER ID =  $95NHB.T1$
$HEADERS
          CONVERTING FROM PURPOSE 3 TO 1
$DATA
  TMAN2,T1 =  TMAN1,T3
$END TP FUNCTION
$MATRIX UPDATE
$FILES
   INPUT FILE = UPDIN, USER ID = $95NHB.T1$ , UNLOAD
   INPUT FILE = MUPDATA, USER ID = $MUPDATA.NHB$
   OUTPUT FILE = UPDOUT , USERID = $PMPNHB.OD$
$HEADERS
   MATRIX EXRACTION FOR NHB PM PEAK
   EXTRACT PM PEAK HBW TRIPS FROM DAILY NHB
$OPTIONS
   MUPDATA
   PRINT TRIP ENDS
$END TP FUNCTION
```

Appendix 3a: Conversion from P/A to O/D: HBW

This TRANPLAN code helps to convert the PM peak trip table of HBW from P/A to O/D format. The input file is "PMPHBW.PA" which was the output from MATRIX UPDATE and the final output file is "PMPHBW.OD"

```
$MATRIX TRANSPOSE
$FILES
  INPUT FILE = TRNSPIN, USER ID = $PMPHBW.PA
  OUTPUT FILE = ODVOL, USER ID = $PMPHBW.OD
$HEADERS
  MATRIX TRANSPOSE RUN
  CONVERTS PM PEAK HBW TRIPS FROM P/A TO O/D FORMAT
$OPTIONS
  PRINT TRIP ENDS
  NO TRANSPOSED FILE
$PARAMETERS
  PA FACTOR = 0.07
  AP FACTOR = 0.93
```

Appendix 3b: Conversion from P/A to O/D: HBO

This TRANPLAN code helps to convert the PM peak trip table of HBO from P/A to O/D format. The input file is "PMPHBO.PA" which was the output from MATRIX UPDATE and the final output file is "PMPHBO.OD"

```
$MATRIX TRANSPOSE
$FILES
  INPUT FILE = TRNSPIN, USER ID = $PMPHBO.PA
  OUTPUT FILE = ODVOL, USER ID = $PMPHBO.OD
$HEADERS
  MATRIX TRANSPOSE RUN
  CONVERTS PM PEAK HBO TRIPS FROM P/A TO O/D FORMAT
$OPTIONS
  PRINT TRIP ENDS
  NO TRANSPOSED FILE
$PARAMETERS
  PA FACTOR = 0.07
  AP FACTOR = 0.93
```

REFERENCES

1. *Short Term Travel Model Improvements.* Report No. DOT-T-95-05. Technology Sharing Program, U.S. Department of Transportation, Washington, D.C. 20590, October, 1994.

2. *PERson FORecasting Model.* Staff Paper 93-1, Connecticut Department of Transportation, Bureau of Policy and Planning, Office of Inventory and Forecasting, July 1993.

3. *Identification of Transportation Planning Data Requirements in Federal Legislation.* Report No. DOT-T-94-21. Technology Sharing Program, U.S. Department of Transportation, Washington, D.C. 20590, October, 1994.

4. Stoupher, P. R. Deficiencies of Travel Forecasting Methods Relative to Mobile Emissions. *Journal of Transportation Engineering*, Vol. 119, No. 5, Sept/Oct 1993.

5. Allen, W. G., Jr., and Schultz, G. W. Congestion Based Peak Spreading Model. In *Transportation Research Record 1556*. National Research Council, Washington, D.C., 1996, pp. 8-15.

6. Loudon, W. R., Ruiter, E. R., and Schlappi, M.L. Predicting Peak Spreading Under Congested Conditions. In *Transportation Research Record 1203*. National Research Council, Washington, D.C., 1988, pp. 1-9.

7. Kumar, A., and Levinson, D. Specifying, Estimating, and Validating a New Trip Generation Model: Case Study in Montgomery County, Maryland. In *Transportation*

8. Johnson, K. R., and Hodson, J. D. Development of a SATURN NETWORK for the South Circular Assessment Study - A Two Tier Approach. In *Proceedings of the Seminar H held at the PTC Transport and Planning Summer Annual Meeting, University of Sussex.* Vol. P335, September 1990, pp. 269-279.

9. Coombe, D., Bamford, J., and Annesley, T. Recent Advances in Traffic Modeling in London. In *Proceedings of the Seminar H held at the PTC Transport and Planning Summer Annual Meeting, University of Sussex.* Vol. P335, September 1990, pp. 281-300.

10. Crevo, C. C., and Virkud, U. Practical Approach to Deriving Peak-Hour Estimates from 24-Hour Travel Demand Models. In *Transportation Research Record 1443.* National Research Council, Washington, D.C., 1994, pp. 30-37.

11. *Analysis of Urban Area Travel by Time of Day.* Table I-6, Report No. FM-11-7519, FHWA, U.S. Department of Transportation, Washington, D.C. 20590, 1972.

12. Spiess, H., and Suter, D. Modeling the Daily Traffic Flows on an Hourly Basis. In *Proceedings of the Seminar H held at the PTC Transport and Planning Summer Annual Meeting, University of Sussex.* Vol. P335, September 1990, pp. 301-310.

13. Allen, W.G., Jr. Analysis of Corridor Traffic Peaking. In *Transportation Research Record 1305,* National Research Council, Washington, D.C., 1990, pp. 50-60.

14. Ruiter, E. R. Microcomputer Capabilities for Trip Table Estimation and Peak Hour Assignments which Consider Peak Spreading. In *Proceedings of the International Conference on Microcomputers in Transportation, San Francisco, California.* June 1989, pp. 279-289.

15. *The Effects of Land Use and TDM Strategies on Commuting Behavior.* Report No. DOT-T-95-06. U.S. Department of Transportation, Washington, D.C. 20590, November, 1994.

16. Purvis, C. S., Iglesias, M., and Eisen, V. A. Incorporating Work Trip Accessibility in Nonwork Trip Generation Models in San Francisco Bay Area. In *Transportation Research Record 1556*, National Research Council, Washington, D.C., 1996, pp. 37-45.

17. Kockelman, K. M. Travel Behavior as a Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from the San Francisco Bay Area. *Paper No. 97-0048. Presented at the 76th Annual Meeting,* Transportation Research Board, Washington, D.C., January 1997.

18. McNally, M. G., and Kulkarni, A. An Assessment of the Land Use - Transportation System and Travel Behavior. *Paper No. 97-1120. Presented at the 76th Annual Meeting,* Transportation Research Board, Washington, D.C., January 1997.

19. Ewing, R., DeAnna, M., and Li, S. Land Use Impacts on Trip Generation Rates. In *Transportation Research Record* 1518, National Research Council, Washington, D.C., 1996, pp. 1-6.

20. Hu, P.S., and Young, J. *Summary of Travel Trends: 1990 Nationwide Personal Transportation Survey.* Report FHWA-PL-92-027, Office of Highway Information Management, FHWA, HPM-40, Washington, D.C. 20590, March 1992.

21. Pisarski, A. E. *Travel Behavior Issues in the 90's.* Office of Highway Information Management, FHWA, HPM-40, Washington, D.C. 20590, July 1992.

22. Allen, W. G., Jr., and Curley, J. D. Using "Life Cycle" in Trip Generation. *Paper No. 97-0758. Presented at the 76th Annual Meeting,* Transportation Research Board, Washington, D.C., January 1997.

23. Kumar, A., and Levinson, D. M. Chained trips in Montgomery County, Maryland. *ITE Journal.* May 1995, pp. 27-32.

24. Kumar, A., Richardson, H. W., and Gordon, P. Beyond the Journey to Work. *Transportation Research - A.* Vol. 22A, No. 6, 1988, pp. 419-426.

25. Ben-Akiva, M., Cyna, M., and Palma, A., D. Dynamic Model of Peak Period Congestion. In *Transportation Research-B,* Vol. 18B, No. 4/5., 1984, pp. 339-355.

26. Ben-Akiva, M., Palma, A., D., and Kanaroglou, P. Effects of Capacity Constraints on Peak-Period Traffic Congestion. In *Transportation Research Record 1085*, National Research Council, Washington, D.C., 1986, pp. 16-26.

27. *Tranplan (Ver. 7) and NIS (Ver. 3) User Manuals.* Distributed by The Urban Analysis Group, 375 Diablo Group, Suite 110, Danville, CA 94526-3430.