Estimating the Temporal Distribution of Traffic Within the Peak Period

Final Report

John N. Ivan, Assistant Professor
Scott A. Allaire, Graduate Student
January, 2000

JHR 99-273 **Project 97-2**

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 Connecticut Transportation Institute of the University of Connecticut.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JHR 99-273	N/A	N/A

4. Title and Subtitle Estimating the Temporal Distribution of Traffic Within the Peak Period	5. Report Date January, 2000
	6. Performing Organization Code N/A
7. Author(s) John N. Ivan and Scott A. Allaire	8. Performing Organization Report No. JHR 99-273
9. Performing Organization Name and Address University of Connecticut Connecticut Transportation Institute Storrs, CT 06269-5202	10 Work Unit No. (TRAIS) N/A
	11. Contract or Grant No. N/A
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
	14. Sponsoring Agency Code N/A

15 Supplementary Notes

N/A

16. Abstract

Predicting accurate peak-hour traffic volumes has become increasingly important as they are being used more and more for quantitative analysis both for determining benefits of highway improvements and for estimating pollutant emission levels. A traditional and commonly used procedure predicts the daily traffic demand for all network links, converting these volumes to peak-hour quantities using "K" factors. The accuracy at the link level is highly questionable, as these factors do not capture the variability in the characteristics of the trip or trip-maker across links, which might influence the peak hour factors.

The research presented in this report describes a study of peak spreading at ten freeway locations in Connecticut. It considers both congestion and site-related variables for predicting the proportion of the peak-period volume on a highway link concentrated in the highest hour during the period. This effort is intended to enhance the existing traditional four-step transportation planning procedure. Log-linear regression is used to relate the congestion measure and site-related variables to the peak-hour proportion. Findings of the research are presented along with discussion of ongoing research activities in this area.

17. Key Words	18. Distribution Statement
Peak-spreading, peak-hour, peak-period, congestion, freeways	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	21. Price
Unclassified	Unclassified	78	N/A

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)		
^oF	Fahrenheit temperature	5(F-32)/9	Celcius temperature	^oC

* 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

°F 32 98.6 212 °F
-40 0 40 80 120 180 200
-40 -20 0 20 40 60 80 100
°C 37 °C

Table of Contents

List of Figures

List of Tables

Chapter 1: INTRODUCTION

Problem Statement

For several decades, transportation planners have estimated and created model systems to predict travel demand on urban networks, primarily to identify deficiencies in an existing highway network and to evaluate major infrastructure investments, such as a new transit rail line or a substantial increase in highway capacity. The sensitivity and accuracy of peak hour volume and speed predictions produced by these models are generally sufficient for evaluating such investments.

However, legislation such as the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, in combination with increasing traffic congestion and decreasing urban space for expanding highways, has resulted in transportation planners investigating other methods of increasing supply or reducing demand of current highways. These methods include transportation system management (TSM) and transportation control measures (TCM) such as improved public transit, trip reduction ordinances, traffic flow improvements, and employer-based programs, all aimed at making the most efficient use of existing transportation facilities [1].

Although these new strategies tend to have small effects on the network and travel patterns of a transportation system, potentially high benefit to cost ratios makes them desirable alternatives. These small effects are difficult to model in existing travel models, often resulting in predicted outputs that are either unrealistic or unfounded due to the insensitivity of these models to small impacts. Therefore, methods to increase the accuracy and sensitivity of quantitative predictions by travel demand prediction models have become increasingly important for evaluating travel time savings, air quality improvements, and safety in the face of such strategies.

The Clean Air Act Amendments (CAAA) of 1990 recognized vehicle emissions as one of the major factors contributing to air pollution. CAAA requirements created a need to increase the degree of interaction between travel demand models and air quality models which predict vehicular emissions and pollution dispersion [2]. In order to meet data requirements for air quality analysis, travel demand models need to produce accurate predictions of speeds by hour of the day and by vehicle type, due to the variation in emission levels produced by different types of vehicles and vehicle speeds. Since peak period and peak hour congestion levels help explain air quality problems, prediction of accurate volumes and speeds during these times is essential.

Transportation planners have created a number of enhancements to meet the required output for air quality analysis. These include revisions to the traffic assignment process to provide link volumes by vehicle operating mode, improvements in volume-speed functions, and assignment post-processors providing more accurate speed and link volume estimates. This document describes the estimation of a post-processor aimed at enhancing existing models by predicting more accurate peak hour volumes on highway links, and thus, more accurate peak hour speeds.

Background

Transportation planning is the process of making decisions regarding whether or not transportation alternatives such as new highways or transit routes, and less substantial alternatives such as those previously discussed, should be implemented. Travel demand models produce the necessary output to analyze the potential benefits and impacts on travelers, society, and the environment that would result from these alternatives. The traditional method of estimating a model for analysis of transportation alternatives and air quality consists of four major steps. These steps are trip generation, trip distribution, mode choice, and trip assignment, and are commonly referred to as the Four-Step Process. The Connecticut Department of Transportation (ConnDOT) Statewide Person Trip Forecasting Model (PERFORM), depicted in Figure 1, uses this traditional four-step modeling procedure with additional intermediate steps as indicated.

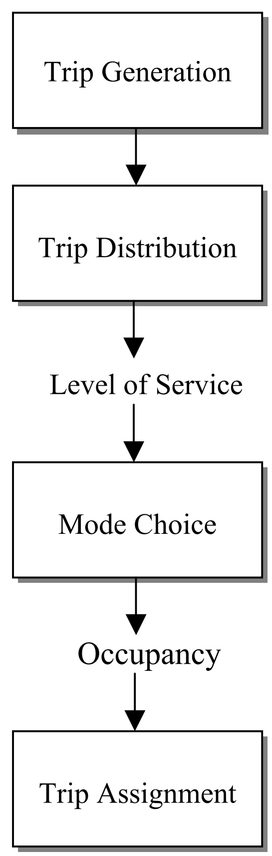


Figure 1. Four-step Process as Implemented in PERFORM [3]

PERFORM is a network-based computer model using the travel demand modeling software package TRANPLAN [4]. The system models highway and transit networks, allowing for proposed changes in highway or transit service to be modeled. PERFORM models five trip purposes: home-based work, home-based non-work, non-home based, truck, and through-trips, and four travel modes: auto drive alone, auto shared ride, bus, and rail. The generated output (trips) from PERFORM is for an average weekday. The four major steps and two intermediate steps to produce this output are reviewed below.

Trip generation predicts the number of person trips produced from and attracted to a traffic analysis zone (TAZ) on an average daily basis by trip purpose. The number of person trips are determined by estimating a relationship between socioeconomic and land use data and the number of trips produced and attracted for a base year [5]. These relationships are then used with land use projections to predict future year estimates of trip productions and attractions.

Trip distribution determines how the trip productions and attractions predicted in trip generation are distributed among the TAZ's. The process is based on the Gravity Model, where the number of trips between two areas (zones) is directly related to the activities in those areas and inversely related to the distance between those areas. The amount of activity in an area is the number of trip productions and attractions predicted in trip generation, while the distance is represented by the travel time between the two areas. The output is a series of trip matrices, depicting the number of average daily person trips produced and attracted between every pair of TAZ's, by trip purpose.

Level of service determination is an intermediate step in PERFORM, where service characteristics of the competing transportation modes for trip interchanges are input. The most common service characteristics of the modes used are travel time and cost. The mode choice step then allocates the interzonal person trips computed in the trip distribution step to each available travel mode. The output is average daily person trip matrices by purpose, broken down by travel mode.

The occupancy module is another intermediate step in PERFORM, in which the shared ride <u>person</u> trips computed in the mode choice step are converted to shared ride <u>vehicle</u> trips by trip purpose. The shared ride vehicle trips are then added to the single occupant vehicle (person) trips for assignment to the highway network.

Trip assignment in PERFORM employs an equilibrium algorithm producing average daily link volumes. Vehicle trips are iteratively assigned to the network to balance the travel for alternate paths between an origin and destination. Equilibrium is achieved when travel times for all used paths between an origin and destination are equal and no unused paths have travel times less than the used paths.

The products and uses of PERFORM are shown below:

- Highway assignments are used to design and evaluate alternative highway proposals.
- Vehicle miles of travel (VMT), the product of traffic volumes and travel distances, is the basic input to highway source emissions models.
- Trip tables are used to examine town-to-town movements and mass transit and carpooling potential.
- Level of Service is used to evaluate the quality of service on highway sections.
- Transit forecasts are used to evaluate transit systems and new transit alternatives.

As mentioned earlier, policy changes, increasing congestion and lack of urban space are requiring transportation improvement alternatives other than expansion of highways or major transit investments to be examined. Transportation planners are in need of more accurate peak period and peak hour estimates, since these are times when congestion and air quality problems are most prevalent. A traditional method (used by ConnDOT) for predicting peak hour link volumes is to first forecast link volumes on a daily basis and then apply constant, or "K", factors. While this procedure may provide representative average peak hour traffic volumes on a network basis, the accuracy at the link level is highly questionable. This view is supported by research by Loudon *et al*. [6], where the peak hour volume as a percentage of the daily volume was shown to vary widely throughout a network. This is because the likelihood of a particular trip being made during the peak hour is not identical for all trips in an area, or even for all trips on the same link, but varies according to characteristics of the trip and the individual making the trip. In general, peak period models have been developed by creating a peak period trip table from the percentage of each trip type that falls within the peak period. However, Loudon *et al*. note that a factor is still applied to the peak period to produce the peak hour volume, with no effort made to relate peaking characteristics to the anticipated level of congestion for the assignment.

Gaining an understanding of how travel behavior patterns in a congested transportation system change when capacity or congestion increases is important when creating methods to produce better peak period and peak hour forecasts [7]. A common response of travelers to increasing congestion (especially as workers are being allowed more flex time) is to depart for work earlier or later than normally desired (with a similar response for the trip home), to avoid the peak congestion. This results in a reduction in the proportion of peak hour to peak period volume, or a widening and flattening of the peak profile, a phenomenon known as "peak spreading." Hounsell [8] defined this form of peak spreading as "active" peak spreading, due to travelers retiming their journeys to avoid unacceptable levels of congestion during the peak. This in turn causes traffic volumes to grow more slowly in peak periods than adjacent off-peak periods. Hounsell also defines another form of peak spreading called "passive" peak spreading. This spreading can happen when increased congestion causes unserved trips in the most intense part of the peak period to shift into later time periods. Most research on peak spreading has not distinguished between active and passive peak spreading, but on capturing the overall effects of the combination.

The phenomenon of peak spreading is of great interest for several reasons. First, increasing capacity may not alleviate the low speeds associated with congestion. The increased capacity can entice travelers who had previously forgone their trips or begun their trips earlier or later than desired to shift their travel back into the heart of the peak period, causing the peak hour volume to rise back up to congestion levels. Second, the shift from funding of new highway construction to upgrading, rehabilitating, and incremental improvements to the existing infrastructure is requiring more accurate modeling of the gains that can be expected from these less substantial improvements. Third, by not capturing peak spreading, there may be an over-prediction of traffic volumes and an under-prediction of speeds for the peak hour assignment. These outputs would then cause higher estimates of pollutant emission levels.

To improve on the traditional method of predicting peak hour volumes, the effect of congestion on travel behavior in the peak period needs to be incorporated. A method is needed to account for individuals adjusting their start times to avoid the most congested hour of the peak period and for capturing the shifting of trips into later periods caused by insufficient capacity. A post-processor method can be used to reflect the capacity-restraining impact the highway network has on traffic volumes and the behavior of trip makers in a congested system. Relating the proportion of peak hour to peak period volume to a congestion measure is appropriate, with the idea that as this congestion measure increases, the proportion decreases as a result of individual trip time changes and trips being shifted into later time periods. Since other factors may influence a trip maker's decision whether or not to adjust his or her travel time, other characteristics of the trips using the links need to be incorporated, accounting for the differences between links in the likelihood of making a trip in the peak hour. The peak hour volumes predicted will then be more consistent with the capacity of individual links and the type of trips using the links, and allow more accurate evaluation of capacity increases and prediction of emission levels.

Objective and Scope

This research uses the Loudon *et al*. [6] peak spreading methodology to estimate models for predicting peak hour to peak period volume proportions. This peak spreading model would best be applied in the form of a post-processor following a peak period trip assignment, but it could also be applied to a factored daily assignment.

The peak spreading model uses a volume to capacity ratio (v/c) computed by dividing the peak period link volumes output from trip assignment by the peak period link capacity to predict the proportion of the peak period volume that occurs in the highest hour. The relationship between the proportion and v/c ratio is shown below:

$$\frac{\text{Peak Hour Volume}}{\text{Peak Period Volume}} = f\left(\frac{\text{Peak Period Volume}}{\text{Peak Period Capacity}}\right) \tag{1}$$

The effects of increasing congestion were modeled with the v/c ratio in a decreasing non-linear function, resulting in a reduction in the proportion as congestion increases.

These models were estimated using data from ten interstate freeway locations in Connecticut where congestion was prevalent. Investigation of the ten sites clearly showed the peak spreading phenomenon to be present. Estimated coefficients varied from site to site even when controlling for congestion level and travel direction, indicating that influences other than congestion are affecting the spreading of the peak. Results from models with additional variables showed site attribute, site location, and trip length distribution variables to be significant in explaining differences in the degree of spreading between sites.

Chapter 2 of this thesis contains a review of related research on capturing the peak spreading phenomenon. Chapter 3 presents the general methodology used to estimate our preliminary peak spreading models, along with discussion of the results from these estimations. Chapter 4 discusses the creation of trip variables and presents the results of estimations with these variables. Chapter 5 discusses the creation of site attribute and location variables and presents the results of estimations with these variables. Chapter 6 discusses creation of area type and regional variables and presents the results of estimations with these variables. Chapter 6 also offers more discussion on a particular model that is viewed to be the most promising and how that model could be applied following a peak period assignment. Chapter 7 offers conclusions and suggestions for future research.

Chapter 2: RELATED RESEARCH

Techniques to Account for Over-assignment and Peak Spreading

With the passing of Federal legislature requiring more sensitive and accurate travel demand modeling has come an increased need to close the gap between “state of the practice” and “state of the art”. How fast that gap is closed depends on the availability of personnel, technical “know-how”, and financial resources, which restrict possibilities for improving current models used by Metropolitan Planning Organizations (MPO) and State agencies. Techniques aimed at improving the accuracy of model outputs are continually being researched to allow improvements to the state of the practice, while not making demands that can't be met by these planning agencies. Such techniques include methods to capture the effects of congestion on travel behavior. When demand levels rise and cause unacceptable levels of service, drivers may choose not to travel, change the time they travel, or change mode. AL-Azzawi describes three methods that may be used to simulate drivers' decisions in a congested network: shadow networks, matrix capping, and elastic assignment [9].

According to AL-Azzawi, a shadow network is a duplicate of the modeled real highway network, connected to the real network at origin and destination zones only, essentially creating an alternative path for trips between an origin and destination. A simple example of a shadow network is depicted in Figure 2, with the shadow links indicated with dashed lines. The number of links in the network is approximately doubled. All shadow links have infinite capacities and fixed speeds with lengths equal to those in the real network. As traffic is assigned to the real network and travel time between an origin and destination increases beyond the fixed travel time of the corresponding path on the shadow network, the shadow network becomes the more attractive alternative. When this occurs, those trips that have not yet been assigned between that origin and destination are diverted to the shadow network, representing trips that have been suppressed or canceled due to unacceptable levels of congestion.

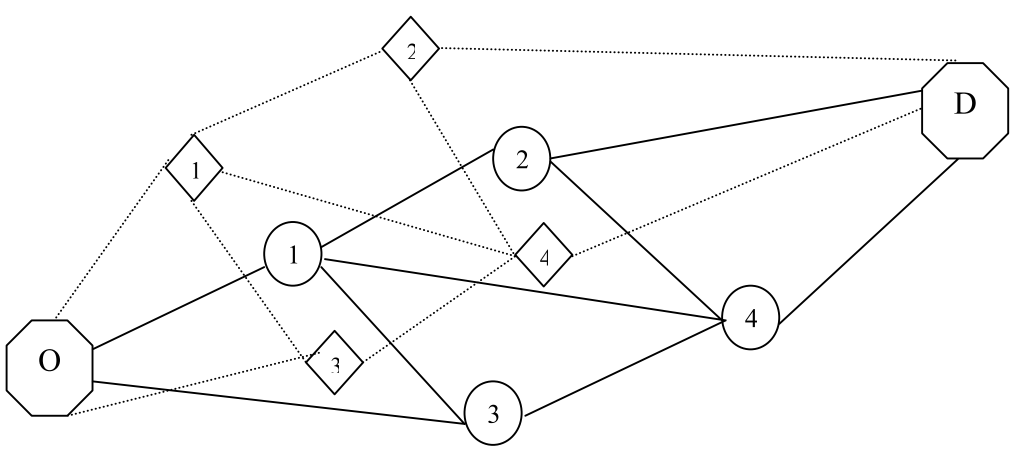


Figure 2. Example of Real Modeled Network With Shadow Network

The shadow network contains more links than is necessary to connect all origin-destination pairs. A skeletal representation of the alternative network can be used, but care must be taken in not using too dense a skeletal network with unnecessary links, or too sparse a network that cannot properly reflect the real network travel cost associated with traveling between an origin and destination. Link lengths in the skeletal network must be increased to reflect the trip portion along any omitted coding of the real network. The main problem with the shadow network is that when the "suppression" or threshold speed on the real network is reached, all further growth in demand diverts to the shadow network. AL-Azzawi admits there is doubt to the realism of this assumption. I tend to agree, as the suppression speed is not likely to be the same for all individual trip makers and all types of trips. Shadow networks will tend to suppress short trips, as congestion on a section of a road will have greater effects on average speeds of shorter trips.

Simulation of peak spreading is one of the possible applications of the shadow network technique, where the true costs of changing trip departure time from peak to off-peak periods are more accurately represented. Rather than fixing speeds on the shadow network, an off-peak demand matrix can be assigned to the shadow network producing off-peak travel conditions (volumes, speeds, and travel times). The 'cost' of changing a trip from peak to off-peak periods is then represented by applying a penalty to the connectors of the real network. The amount of peak spreading is measured by the number of trips being diverted to the shadow network as the peak demand matrix is assigned.

Another method of modeling peak spreading described by AL-Azzawi is called matrix capping. While the shadow network method is applied during assignment, the matrix capping method is performed afterward. Matrix capping constrains traffic to match network capacity by retaining only those trips that can be made on the network in the time period being modeled. This technique can be viewed as a way to capture the "passive" spreading that occurs in a congested network. The technique first identifies links in the network where the demand exceeds capacity. Select link analysis is used to determine the origin-destination matrix of trips loading these links. The identified trips on overloaded links are then reduced proportionally to reduce the demand matrix to near supply and the revised matrix is then reassigned. These steps are repeated until all links have volumes that are less than or equal to their capacity, or some acceptable level of network overloading is met. In a multi-stage transportation model, the trips not served or "lost" in the modeled period would be reallocated by the assignment, modal split and/or distribution models.

This process has an advantage over the shadow network because there is no need to increase network representation or to specify speed diversion criteria. Unlike the shadow network, the capping technique does not discriminate between short and long distance trips. The main disadvantage is that depending on the amount of overloaded links and size of the network, the whole process can take a considerable amount of computing time.

The third method of modeling peak spreading described by AL-Azzawi is elastic assignment, which considers that demand for travel for any origin-destination pair is not fixed, but rather a decreasing function of the cost of making that trip. The function is as follows:

$$T_{ij} = T_{ij}^{\max} \frac{e^{-bC_{ij}^{*}}}{e^{-bC_{ij}^{*}} + e^{-bC_{ij}^{o}}} \tag{2}$$

where T_{ij} = constrained demand for travel between zones *i* and *j*,
$T_{ij}^{\max}$ = unconstrained demand for travel between zones *i* and *j*,
C^{*}_{ij} = minimum cost of travel from zones *i* to *j*,
C^{o}_{ij} = acceptable fixed cost of travel from zones *i* to *j*, and
b = an elasticity parameter.

Each origin-destination pair is connected with a pseudo-link allowing for diversion of trips with increasing congestion. At the minimum cost of travel, T_{ij} trips are assigned to the real network and $T_{ij}^{\max} - T_{ij}$ trips are assigned to the pseudo-link from *i* to *j*. The equations can be solved as a fixed demand equilibrium assignment problem. As the highway network becomes more congested, and the cost C_{ij}^{*} increases, more trips are diverted to the pseudo links, representing suppressed trips, mode changes, or changes in departure times to avoid the congestion. The number of links in an N zone network is increased by N(N-1). This method has the advantage of allowing suppression functions to be defined on a zone to zone basis. The main problem with elastic assignment is the choice of the demand function. A logit form (shown above), has a tendency to suppress long trips, while the power function formulation shown in Equation (3) tends to suppress short trips. The parameters *a* and *b* in the functions determine the overall level of suppression. The greater *a* and *b* are, the greater the level of suppression.

$$T_{ij} = T_{ij}^{\max} \left(\frac{C_{ij}^{*}}{C_{ij}^{o}} \right)^{-a} \tag{3}$$

Like the shadow network, the elastic assignment could be used to simulate peak spreading as a result of unacceptable levels of congestion.

The three techniques by AL-Azzawi that account for over-assignment and peak spreading are intended to aid transportation planners in traffic modeling studies. The shadow network and elastic assignment techniques allow suppression of trips, based on the concept that increasing costs to the trip maker will reduce the number of trips being made on the highway network. The matrix capping approach accounts for overestimation by iteratively adjusting the demand matrix to near supply level, retaining only those movements that could be made during the modeled period on the network. While these techniques seem to do a good job of just that, they do not capture any of the factors other than congestion that cause trip makers to change their time of travel or suppress their trip. Characteristics of the site, trip, or trip maker are not used in determining the extent of peak spreading on a link or differences between links. For example, a work trip may not be as easily adjusted and certainly not suppressed, as a non-work trip might be. Therefore, the type of trips on a link may determine at what level of congestion (cost) trip makers adjust their time of travel. The following research aims at identifying and capturing additional factors that influence the degree of spreading that occurs.

Peak Spreading Models

Research by Allen presents a methodology for forecasting future flattening or shifting of the peak hour on a link-specific basis as congestion increases [10]. The technique uses a modified Poisson distribution to describe the spread of four-hour volumes (6-10 AM) across each 15-minute period within the four hours. The data used in the study came from Interstate 80 in northern New Jersey, and includes four-hour volumes by 15-minute periods and a comprehensive ramp survey providing information about trip origin, destination, purpose, vehicle occupancy, and other roadways used. The technique consists of tabulating and graphing the 15-minute traffic counts for each link in the corridor as a proportion of the total four-hour volume. The modified Poisson curve was then hand-fit to each of the graphs by adjusting the Poisson coefficients until the best fit was determined. These curves are then used as the observed data, to be fitted to one Poisson model for all links. A calibration file is built containing the Poisson coefficients and all available independent variables for each link. These variables were then used to estimate the Poisson coefficients using regression analysis. The independent variables used to estimate the Poisson coefficients included a speed difference variable equal to the free-flow minus congested speed, a delay variable, a dummy variable representing the link location, and a volume variable. Allen admits the research effort is an awkward attempt to quantify and forecast peaking and the results may be difficult to generalize for use elsewhere. However, his technique was able to identify and use variables other than a single congestion measure. This is an important part of estimating a peak spreading model that is transferable to all links.

Research involving the "peak-spreading" phenomenon under congested conditions has also included the development of a post mode choice procedure. Allen *et al.*, using an O-D survey and highway networks with peak and off-peak speeds, estimated models by trip purpose that predict the flattening of the AM peak hour as congestion and trip length increase [11]. The method estimates peak hour vehicle trips by first estimating the share of the peak hour travel occurring in the three hour peak period and then applying this share to the estimated peak period trips.

A file was assembled for the calibration of the peak spreading model containing auto trips by purpose. Only those trips with valid production and attraction zones, valid start and end times, and were in progress between 6:30 and 9:30 AM, were kept. Network peak and off-peak travel times and distances were attached to each record. For each trip record, the vehicle hours of travel (VHT) spent between 6:30 and 9:30 AM and between 7:30 and 8:30 AM were calculated. The ratio of peak hour VHT to peak period VHT was used as the dependent variable in estimation. The independent variables used in the model are the trip distance and a measure of congestion, defined as the difference in minutes between the AM peak one-hour travel time and the off-peak travel time. No *a priori* assumptions were made about the model form, and initial data investigation found a great deal of scatter. The trip distance and congested time difference data were then grouped into ranges to accommodate the variation in data. The final model structure was a series of stratified curves.

The main findings from this research were that trip purpose and trip distance, in addition to a congestion measure, are important variables for predicting the share of peak hour travel within a peak period. One of the good things about this model is that it incorporates a zone-to-zone congestion measure. This is a good travel time measure for determining if an individual will adjust his or her travel time in order to avoid unacceptable levels of congestion because it considers the entire trip.

Another congestion measure that can be used for capturing travel behavior of an individual when faced with unacceptable levels of congestion is the v/c ratio on individual links. The following research utilized this congestion measure in predicting the degree of peak spreading that occurs on a link.

Loudon *et al*. conducted research for the Arizona Department of Transportation (ADOT) on the phenomenon of peak spreading on congested roadways [6]. The research used data from forty-nine freeway and arterial corridors in Arizona, Texas, and California. The corridors were chosen because they had historical hourly count information covering at least a five-year period. Examination of the data showed the ratio of maximum one-hour volume counts to daily volume counts across sites varied widely from the most often assumed value of 0.10, suggesting the need for more accurate modeling of peak hour volumes. The research was designed to result in recommended changes to the UTPS-based forecasting system used by the Maricopa Association of Governments (MAG) Transportation Planning Office, allowing future year forecasting to reflect the peak spreading phenomenon.

The first step in producing peak hour assignments was dividing total daily travel by trip purpose into three periods: 6-9 AM, 3-6 PM, and off-peak (all other hours). These periods were chosen based on the feeling that there was some degree of stability within each period, that is, travelers would not tend to shift out of these peak periods to avoid congestion. With no trips shifting out of these peak periods, the percentage of travel predicted for each peak period should remain constant with the level of congestion. This hypothesis was tested using least squares regression between the ratio of the three-hour peak period volume to twenty-four-hour volume (dependent variable) and the three-hour peak period v/c ratio (independent variable). The hypothesis is rejected if the coefficient estimated for the independent variable is significant, indicating a relationship between the independent and dependent variables. Results from thirty-six regressions showed some tendency for peak spreading to affect the three-hour peak period as twenty-eight of the estimated coefficients had the correct sign (negative), but most were not significantly different from zero at the 95 percent confidence level.

Using the historical count data, a functional relationship was estimated between two quantities; the ratio of the peak hour volume to the peak period volume, and the overall v/c ratio during the peak period. Using ordinary least-squares regression, the parameter estimates were obtained. Analysis of the signs and significance of the coefficients on the v/c variable clearly demonstrated the presence of peak-spreading, as eighteen of the nineteen corridors in the analysis had a negative v/c coefficient (indicating a decrease in the proportion of the peak hour to peak period volume as the v/c increases), and more than half had slopes that were significantly different from zero with 95 percent confidence.

The average value of the slope for each facility type (freeway and arterial) was estimated by aggregating the data from all the individual corridors. Only observations with a v/c ratio of

0.50 and above were used, since the peak spreading phenomenon occurs only at higher congestion levels. The freeway sites were also broken down into two groupings, one containing freeways that have two or three lanes, and another with freeways having four or five lanes. Regression results using these two groupings produced slope values significantly different from zero, and a trend of decreasing slope (more negative) with an increase in the number of lanes. Aggregate analysis on the arterial corridor sites resulted in an estimated slope coefficient that was not significant at even a 90 percent confidence level. This is attributed to the lack of high v/c ratios in these corridors. Because of this, the results of a single arterial site having a significant slope coefficient were chosen to represent arterial corridors in the MAG models.

In order to reflect current conditions in Phoenix more closely, the model was then recalibrated for different facility type and area type combinations in the Phoenix network using current observed data. The observed data allowed computation of an average peaking factor (or the dependent variable in the estimations) and an average v/c ratio for each area type-facility type combination. These values were used with the slope coefficients from the previous estimations to calibrate the constant to a specific link type.

A test of the peak-spreading procedures was conducted by comparing the new assignments resulting from use of a three-hour peak period and the peak spreading model with observed data on the links where volume counts and speeds were available. A significant increase in the accuracy of peak hour link volumes and speeds was found when the new assignments were compared with results previously obtained from MAG for the Phoenix area using a twenty-four hour assignment with a constant peak hour factor.

Although the results of the model estimations showed a clear pattern of peak spreading as congestion arises, and implementation of the modeling procedures demonstrated a great improvement in accuracy, there are limitations to this procedure. First, two adjacent links having different v/c ratios in the peak period could result in a different amount of peak spreading predicted for each, and thus, potentially, inconsistent peak hour volumes. The impact of this is likely to be small, since the authors point out that calibration of the peak spreading model is performed on a facility type rather than a link-specific basis, thereby averaging the effects over a facility.

Second, the peak-spreading model is applied at the link level, while the peak-spreading on a link could be attributed to congestion at another location in the network, or because the link is in a corridor where the level of congestion is perceived to be high. To the extent that links in a corridor have capacities that are generally proportional to the peak period volume flow on the link and are fairly homogenous, the results of this limitation will not be serious.

Third, the procedure recommended does not reflect spreading of the peak outside of a three-hour period. Because the procedure depends on the assumption of a stable peak period, a larger peak period may be needed in other locations.

Several factors were examined in determining the direction and method of the research presented in this thesis, including the availability of data, results from previous research, and the results that are needed to accomplish our objective. These factors led to the decision to extend the Loudon *et al.* procedure just described by incorporating site variables likely to influence peak spreading at a given site, thus making the model transferable statewide. A four-hour peak period (3-7 PM) is used instead of the commonly used two to three hour peak period to allow for stability within the peak period. This peak period will also allow the peak spreading model to be applicable to peak period link forecasts, which could be estimated using a peak period model previously estimated for ConnDOT's statewide model [12]. Since trip purpose has been shown in previous research to be important in determining the degree of peak spreading that occurs, and air quality impacts are associated with higher volumes, the PM peak was chosen, as it generally contains a more diverse set of trip purposes and higher volumes than the AM peak.

Chapter 3: INDIVIDUAL SITE MODELS

Methodology

Because this study defined a four-hour peak period, the basic functional form for the peak-spreading model is:

$$P = \tfrac{1}{4} + ae^{bX} \qquad (4)$$

where

P = the ratio of the peak-hour volume to the four-hour peak-period volume,
X = the volume/capacity ratio for the four-hour period, and
a,b = parameters to be estimated.

The functional form has several advantageous characteristics:

1. The lowest value it can have is 0.25, representing a totally flat peak period.
2. For large values of X, P approaches 0.25.
3. Valid values of P can exist for values of X greater than 1.0. This precludes any discontinuities in applying the function.

Figure 3, a plot of the function, shows how the parameters a and b adjust the shape of the curve. The parameter a adjusts the curve by shifting the intercept up or down (indicating how the peak period traffic is distributed at a low congestion level), while the parameter b adjusts the convexity of the curve (indicating how sensitive the peak spreading is to increasing congestion).

Function Transformation

By transforming the equation, ordinary least squares regression can be used to estimate the model parameters. The transformation consists of moving the constant to the dependent variable side of the equation and then taking the natural log of both sides of the equation, resulting in Equation (5), which resembles the linear regression model formulation.

$$\ln(P - \tfrac{1}{4}) = \ln a + bX \qquad (5)$$

Equation (6) simplifies further, where $C = \ln a$

$$\ln(P - \tfrac{1}{4}) = C + bX \qquad (6)$$

ATR Data

Automatic traffic recorder (ATR) station data at sixteen locations providing hourly volume counts over a five-year period were obtained from ConnDOT. The raw ATR data in ASCII file format contains functional classification, station number, town number, month, day and year, day of the week and directional code, and traffic volumes by hour. The format of the

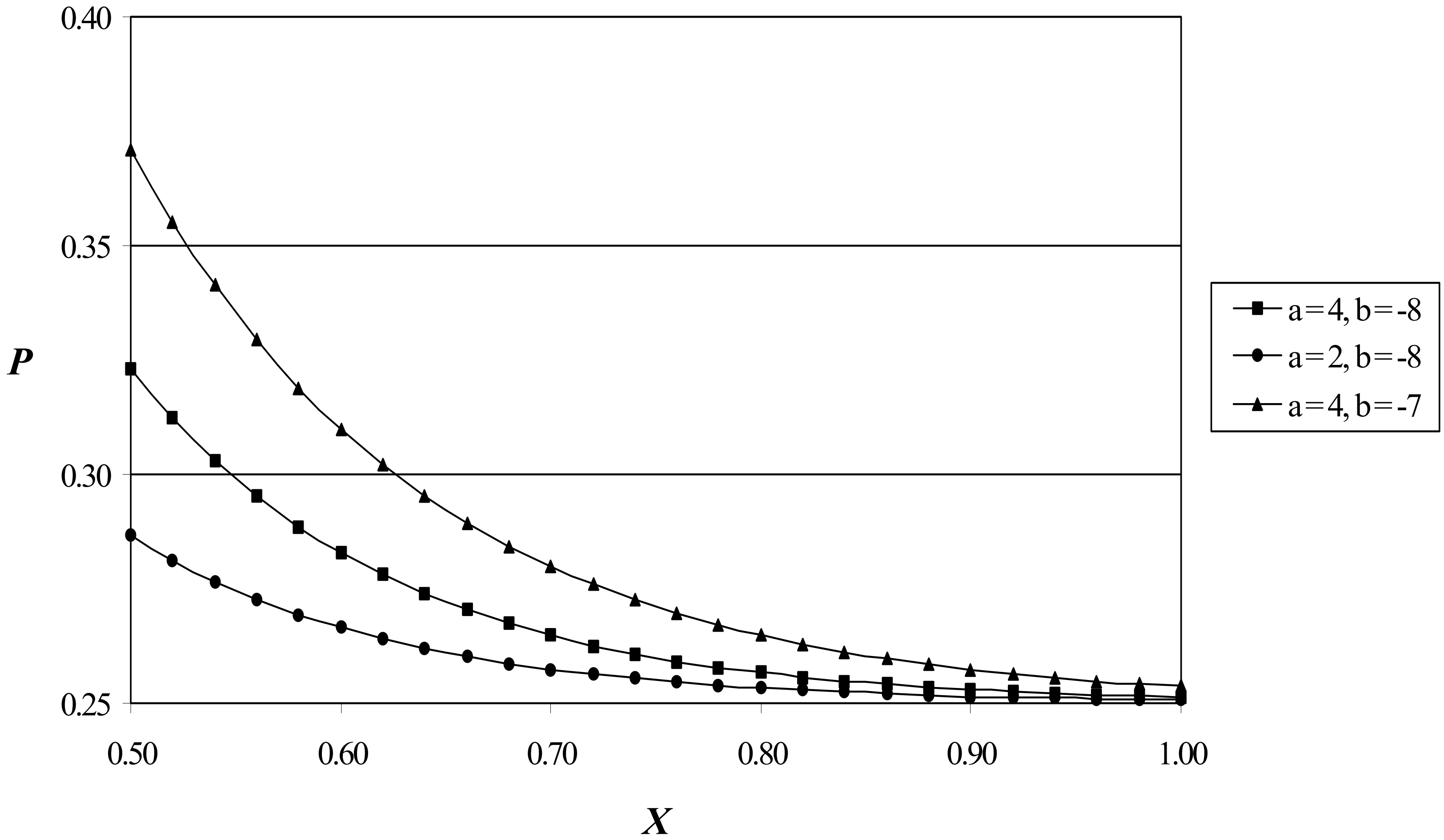


Figure 3. Functional Form Displaying Effects of Parameters *a* and *b*

ASCII files is shown in Table 1. Ten sites were selected for analysis, all with extensive congestion along the corridors during the peak period. The selected sites are located along freeways I-95, I-91, and I-84 and are shown in Figure 4. Four of the sites are located near Hartford, four are in the southwest region of Connecticut, and two are in the southeast region of Connecticut.

The ASCII site data were imported into the statistical software package SPSS for processing and analysis. For each site and day, the PM peak period volume (*PPV*) was calculated by totaling the hourly volumes from 3:00 to 7:00 PM. The ratio (*P*) of the peak hour volume (*PHV*), or the highest of the four-hour volumes, to the four-hour peak period volume (*PPV*) was calculated and used as the dependent variable in the regression analysis.

Other variables that were created include the capacity of the roadway at each site for the peak period and the volume to capacity ratio (*X*) during the four-hour period. Capacity was based on the number of lanes and the geometric characteristics at each site, with a reduction imposed for heavy vehicle percentages obtained from Highway Performance Monitoring System (HPMS) station data.

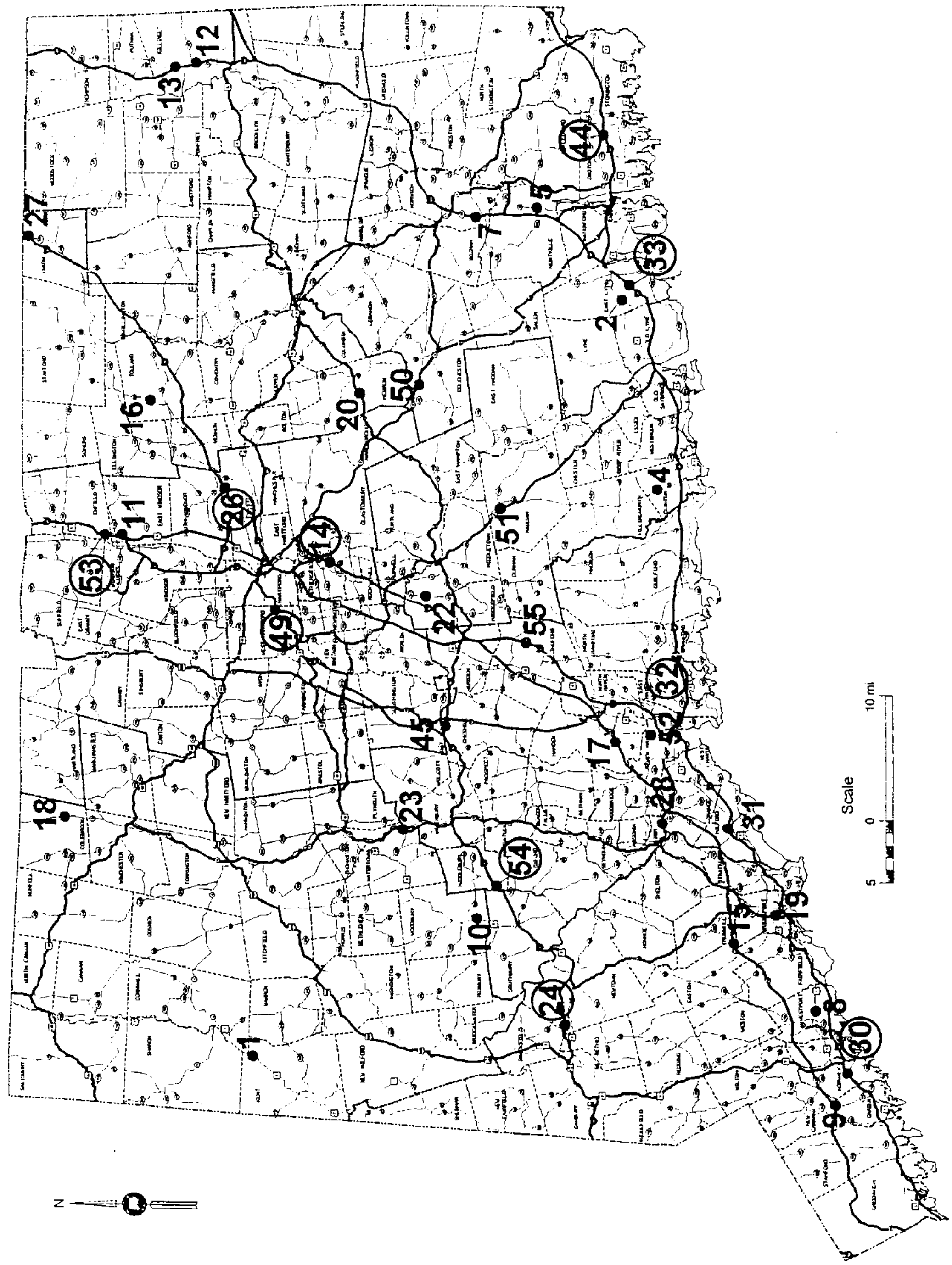
The number of lanes at each site was counted using a ConnDOT photolog station, a laser video disc system containing images of the entire 6300 centerline kilometers (3,900 miles) of the Connecticut state-maintained highway network. Two photolog vans gather photographic images of the roadway and its immediate surroundings at 16-meter (0.01-mile) intervals. Highway geometric data, location and direction are also simultaneously collected and recorded.

Data Set Exploration

Frequencies and summary statistics were run on the variables discussed above and are shown in Table 2. The variable *PHV* contains the volume of the highest hour within the peak period. The values for the dependent variable *P*, the ratio of *PHV* over *PPV*, range from a minimum of 0.25 to a maximum of 0.43. The value of 0.25 is the lowest value the proportion can have, and represents the *PPV* being equally distributed among the four hours (or a totally flat peak period). The extreme case here where 43 percent of the *PPV* is in the peak hour indicates a rather high peaking of traffic volumes. The values for the main independent variable *X* range from 0.50 to 0.94. The value of 0.50 was used as the criterion for selecting the study sites, as the peak spreading phenomenon is most valid for congested conditions. All observations with a v/c ratio above 0.50 were kept. The maximum value of 0.94 suggests extreme congestion over the four-hour period. The mean is the average value of all cases and the standard deviation is a measure of how widely values are dispersed from the mean. Overall, the summary statistics show a wide range of values for each variable.

Table 1. ASCII File Format of ATR Station Data

Column	Field Length	Alpha/Numeric	Description
1	1	N	Record Type
2-3	2	N	FIPS State Code
4-5	2	N	Functional Classification
6-11	6	A	Station Identification
12	1	N	Direction of Travel
13	1	N	Lane of Travel
14-15	2	N	Year of Data
16-17	2	N	Month of Data
18-19	2	N	Day of Data
20	1	N	Day of Week
21-25	5	N	Traffic Volume Count, 00:01-01:00
26-30	5	N	Traffic Volume Count, 01:01-02:00
31-35	5	N	Traffic Volume Count, 02:01-03:00
36-40	5	N	Traffic Volume Count, 03:01-04:00
41-45	5	N	Traffic Volume Count, 04:01-05:00
46-50	5	N	Traffic Volume Count, 05:01-06:00
51-55	5	N	Traffic Volume Count, 06:01-07:00
56-60	5	N	Traffic Volume Count, 07:01-08:00
61-65	5	N	Traffic Volume Count, 08:01-09:00
66-70	5	N	Traffic Volume Count, 09:01-10:00
71-75	5	N	Traffic Volume Count, 10:01-11:00
76-80	5	N	Traffic Volume Count, 11:01-12:00
81-85	5	N	Traffic Volume Count, 12:01-13:00
86-90	5	N	Traffic Volume Count, 13:01-14:00
91-95	5	N	Traffic Volume Count, 14:01-15:00
96-100	5	N	Traffic Volume Count, 15:01-16:00
101-105	5	N	Traffic Volume Count, 16:01-17:00
106-110	5	N	Traffic Volume Count, 17:01-18:00
111-115	5	N	Traffic Volume Count, 18:01-19:00
116-120	5	N	Traffic Volume Count, 19:01-20:00
121-125	5	N	Traffic Volume Count, 20:01-21:00
126-130	5	N	Traffic Volume Count, 21:01-22:00
131-135	5	N	Traffic Volume Count, 22:01-23:00
136-140	5	N	Traffic Volume Count, 23:01-24:00
141	1	N	Footnotes



N
Scale
0
5
10 mi

Table 2. Summary Statistics of Major Variables

	Number of Cases	Minimum	Maximum	Mean	Standard Deviation
PHV	14,894	2,177	7,070	4169.01	1150.49
PPV	14,894	8,531	24,796	14650.32	3903.68
P	14,894	0.25	0.43	0.2842	0.01519
X	14,894	0.50	0.94	0.6336	0.08885

Graphing *P* versus *X* for each travel direction at each site demonstrated the difference in the peak spreading phenomenon between travel directions. For example, Figure 5 shows the relationship between *P* and *X* in the commute direction (or the direction that typically has heavier volumes) for Site 30 in Norwalk. *P* decreases fairly rapidly as *X* increases (for values above 0.5). The rate of decrease in *P* gradually diminishes asymptotically as the level of congestion increases. This may be because as the four-hour volume approaches saturation, the benefit of moving out of the peak also diminishes. This would appear to validate the choice of an exponential model form.

The reverse-commute direction shown in Figure 6 (also for site 30 in Norwalk) has smaller P values, indicating traffic is fairly well spread through the peak period for all congestion levels. In other words, peak spreading is less sensitive to increases in congestion levels than in the commute direction. This is likely due to the different distribution of trip purposes between the two directions, as a higher percentage of discretionary trips in the reverse-commute direction would tend toward peak period trips being less concentrated in one hour.

These observations led to the creation of two additional variables, which account for the directional differences at a site. First, a directional dummy variable D_R was created, equal to 1 if the direction of flow was in the reverse-commute direction. This allows estimation of different constants for each flow direction. Second, a directional v/c variable X_R was created, equal to *X* if the direction of flow was in the reverse-commute direction. This allows estimation of different slopes for each flow direction, as well as direct statistical testing of the significance of their difference. Essentially, an individual model is estimated for each direction. A sample of the data set containing these two variables, the main independent variable *X*, and the dependent variable ln(*P* - 1/4) used in base model estimation is shown in Table 3.

Each study site has a unique atr code defining the direction of travel at the site. For example, 9014_ 1 is the northbound direction of travel on I-91 in Wethersfield, while 9014_5 is the southbound direction of travel on I-91 in Wethersfield. 9024_3 is the eastbound direction of travel on I-84 in Newtown, while 9024_7 is the westbound direction of travel on I-84 in Newtown. Notice that the northbound direction of flow in Wethersfield was defined as the reverse-commute direction, and has a value of 1 for D_R and the value of *X* for X_R.

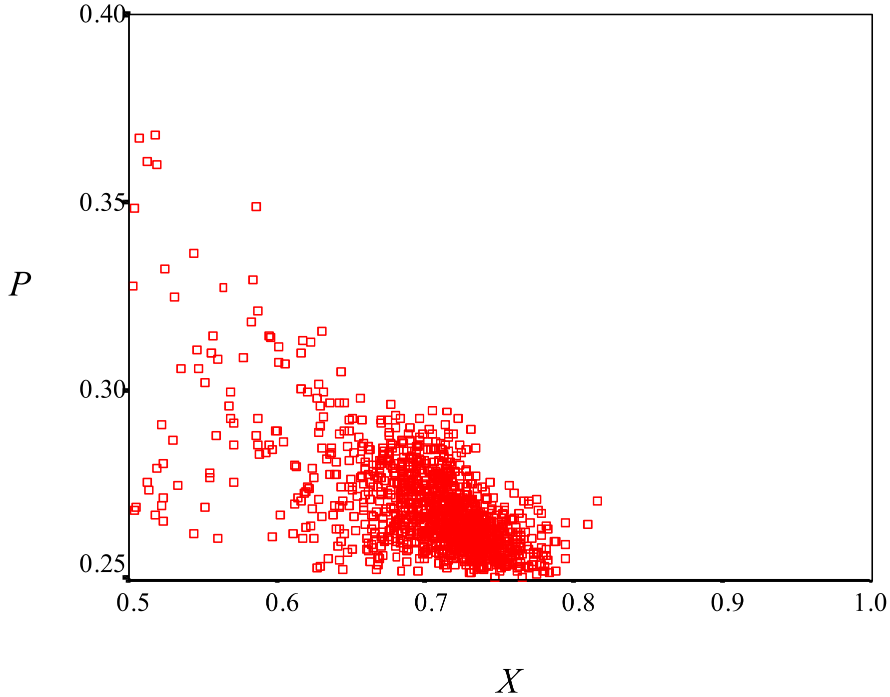


Figure 5. Observed Peak Spreading on I-95 in Norwalk Commute Direction

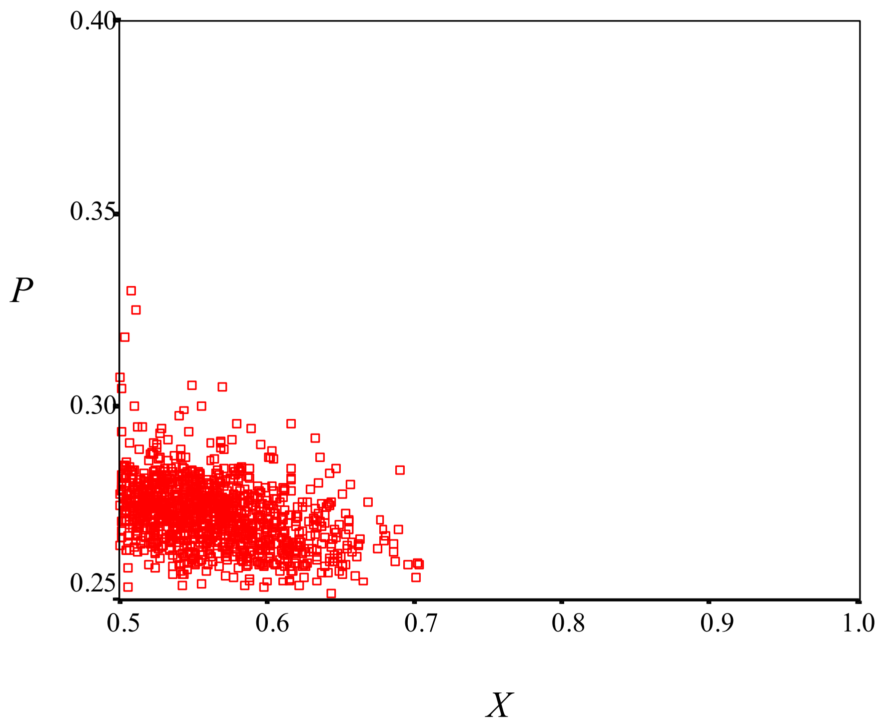


Figure 6. Observed Peak Spreading on I-95 in Norwalk Reverse-Commute Direction

Table 3. Sample Data for Estimation of Base Model

ATR	*PPV*	*PHV*	*P*	ln(*P* - 1/4)	D_R	*X*	X_R
9014_1	19370	5370	.277	-3.60	1	.549	.549
9014_1	17735	4832	.272	-3.80	1	.503	.503
9014_1	18306	5083	.278	-3.59	1	.519	.519
9014_1	17824	5003	.281	-3.48	1	.505	.505
9014_1	17834	4668	.262	-4.44	1	.505	.505
9014_5	16750	5031	.300	-2.99	0	.633	.000
9014_5	17119	5080	.297	-3.06	0	.647	.000
9014_5	17300	5103	.295	-3.10	0	.654	.000
9014_5	17099	5012	.293	-3.14	0	.646	.000
9014_5	18958	5090	.268	-3.99	0	.716	.000
9024_3	10368	3008	.290	-3.22	0	.588	.000
9024_3	10399	3180	.306	-2.89	0	.589	.000
9024_3	11941	3508	.294	-3.13	0	.677	.000
9024_3	10211	2992	.293	-3.15	0	.579	.000
9024_3	10317	2982	.289	-3.24	0	.585	.000
9024_7	10191	2713	.266	-4.12	1	.577	.577
9024_7	9819	2573	.262	-4.42	1	.556	.556
9024_7	9391	2427	.258	-4.77	1	.532	.532
9024_7	9700	2548	.263	-4.37	1	.550	.550
9024_7	9627	2545	.264	-4.24	1	.546	.546

Hypotheses

The congestion measure, or v/c ratio, of the link is most likely the best variable for capturing the peak-spreading phenomenon, but other variables are likely to be significant as well. As previously mentioned, the likelihood of a trip being made during the peak-hour is not the same for all trips in a given area or even for all trips on the same link, but varies according to characteristics of the trip and the individual making the trip. For example, Allen showed that trip distance and purpose were both significant when predicting the peak hour to peak period proportion [11]. Trip purpose is likely to be significant because work trips are less discretionary than non-work trips, meaning a higher percentage of non-work trips don't have to be made during the peak-hour. Therefore, different sites will most likely have different levels of peak spreading, influenced by characteristics such as the distribution of trip purposes and trip lengths at a site, as well as a site's location and physical attributes.

To test this hypothesis the model formulation was adjusted to incorporate site-specific dummy variables and site-specific v/c variables, representing those factors which may influence differing degrees of peak spreading among sites. Because the reverse-commute direction tends to have fewer work trips, and thus, more discretionary trips, the directional dummy variable and directional v/c variable were included. These variables may be somewhat representative of the general effect of trip purposes. The transformed model formulation used for estimation is now of the form:

$$\ln(P - \tfrac{1}{4}) = C_0 + \boldsymbol{C_1 S} + C_2 D_R + b_0 X + \boldsymbol{b_1 X_S} + b_2 X_R \qquad (7)$$

where

C_0, C_2, b_0, b_2 = scalar parameters to be estimated,
$\boldsymbol{C_1}$, $\boldsymbol{b_1}$ **=** vector parameters to be estimated, each with one value for all but one of the analysis sites,
$\boldsymbol{S}$ = vector of site-specific dummy variables having values of 1 or 0, with one value for all but one of the sites,
D_R = directional dummy variable having a value of 1 if the flow direction is in the reverse-commute direction, 0 otherwise,
X = the volume/capacity ratio for the four-hour period,
$\boldsymbol{X_S}$ = vector of site-specific variables equal to the volume/capacity ratio for the four-hour period X multiplied by the site-specific dummy vector $\boldsymbol{S}$, with one value for all but one of the sites, and
X_R = directional site-specific variable having the value of the volume/capacity ratio for the four-hour period if the flow direction is in the reverse-commute direction, 0 otherwise.

Results and Discussion

Before estimating a model using the transformed equation, a more generic model was run without any site-specific variables. The results are shown in Table 4 as the generic model. Then Equation (7) was used to estimate a regression model with variables distinguishing the 10 sites. The results are shown in Table 4 as the site-specific model. East Lyme has coefficients equal to zero as it was chosen to be the base site.

The total effect on the intercept and slope of the model can be determined by adding the like terms, which results in a function resembling Equation (6): $\ln(P - 1/4) = C + bX$. The constant C for a given site i is calculated by adding the general model constant C_0, the site-specific constant C_{1i}, and the directional constant C_2 if the flow is in the reverse-commute direction. Similarly, the slope b of a given site is calculated by adding the generic v/c coefficient b_0, the site-specific v/c coefficient b_{1i} that acts as a correction to the generic coefficient, and the v/c directional coefficient b_2 if the flow is in the reverse-commute direction. Essentially, an individual model is estimated for each site. For example, for I-95 in Norwalk in the reverse-commute direction, C and b are computed as follows:

$$C = -2.364 + (-2.743) + 3.733 = -1.374$$
$$b = -2.476 + 3.375 + (-5.382) = -4.483$$

The peak-spreading phenomenon is clearly seen in both models, but the discussion will focus on the site-specific model. All factor categories explain variation among sites and direction, and are significant. The coefficient estimated for the base v/c ratio is negative and significant, indicating that, overall, the ratio of peak-hour to peak-period volume decreases as the v/c ratio on the link increases; this is the basic definition of peak spreading. The nine site-specific differential v/c coefficients act as corrections to the generic v/c coefficient by increasing or decreasing the slope relative to the East Lyme base site. Seven of these coefficients are significantly different from 0 at the 95 percent confidence level. This variation from the base case and among each other indicates that other variables besides congestion are affecting the amount of peak spreading that occurs. The coefficient on the directional v/c variable is positive and highly significant, indicating peak spreading is affected by the orientation of the trip as well as the location, possibly representing the trip purpose or characteristics of the trip-makers using the link. The positive coefficient flattens the slope, decreasing sensitivity of spreading to the v/c ratio relative to the commute direction.

The constants are of interest in that eight of the nine site-specific constants have t-statistics that are significantly different from 0 at the 95 percent confidence level. Seven of the nine are positive, indicating there are characteristics at these sites that resist the spreading of the peak relative to the base case. These constants could capture effects attributable to the characteristics of the trip and trip-maker, as well as characteristics of the site. The coefficient on the directional dummy variable is negative and significant, indicating that the reverse-commute direction does not have a highly concentrated peak even for low v/c levels. The distribution of the peak-period volume in the reverse-commute direction is flatter, probably because of the higher percentage of discretionary trips being made in the reverse-commute direction.

Overall model fit is very good as the R-squared value indicates. The F-statistic is large, rejecting the null hypothesis that all coefficients are equal to zero. Comparison between the generic and site-specific models is done by using the F-test, which tests whether the restricted, or generic model, provides almost as good a fit as the unrestricted, or site-specific model. The hypotheses are formulated as follows:

> H_0: Unrestricted (Site-Specific) Model does not explain more variation than the Restricted (Generic) Model.
> H_a: Unrestricted (Site-Specific) Model does explain more variation than the Restricted (Generic) Model.

The test statistic depends on the reduction in unexplained variance, and is computed with the following equation:

$$f = \frac{(SSE_l - SSE_k)/(k-l)}{SSE_k/(n-k)} \tag{8}$$

where

SSE_k = sum of square errors or unexplained variation for the unrestricted model,
SSE_l = sum of square errors or unexplained variation for the restricted model,
k = number of parameters in unrestricted model,
l = number of parameters in restricted model, and
n = sample size.

Table 4. Regression Results

	Generic Model		Site-Specific Model	
	Coefficient	*t-statistic*	*Coefficient*	t-statistic
Constants (*C*)				
Base	-2.450	-66.246	-2.364	-16.657
D_R	-0.464	-4.035	-2.743	-29.002
I-91, Wethersfield	-	-	0.963	5.748
I-84, Newtown	-	-	0.170	0.915
I-84, Manchester	-	-	-0.989	-6.117
I-95, Norwalk	-	-	3.733	23.166
I-95, Branford	-	-	2.649	15.842
I-95, East Lyme	-	-	0	-
I-95, Groton	-	-	0.607	3.761
I-84, W. Hartford	-	-	0.480	3.121
I-91, Enfield	-	-	-0.473	-2.631
I-84, Middlebury	-	-	0.749	3.952
Slopes (*b*)				
X	-1.402	-25.193	-2.476	-10.031
X_R	-0.351	-1.734	3.375	20.837
I-91, Wethersfield	-	-	0.080	0.285
I-84, Newtown	-	-	0.665	2.115
I-84, Manchester	-	-	2.752	10.103
I-95, Norwalk	-	-	-5.382	-19.782
I-95, Branford	-	-	-3.443	-12.182
I-95, East Lyme	-	-	0	-
I-95, Groton	-	-	0.353	1.288
I-84, W. Hartford	-	-	0.630	2.424
I-91, Enfield	-	-	1.994	6.499
I-84, Middlebury	-	-	-0.734	-2.206
Statistics				
R-squared	0.198		0.559	
SSR	934.305		2643.920	
SSE	3791.555		2081.940	
F-statistic	1223.051 critical F = 2.60		899.353 critical F = 1.57	
Degrees of Freedom	14,890		14,872	

Shading indicates failed t-statistics at 95% confidence level

The rejection region for the test at a confidence level α is:

$$f \geq F_{\alpha,k-l,n-k} \tag{9}$$

The F-test for a 95% confidence level shown below rejects the null hypothesis, indicating the site-specific model to be the significantly better model:

$$\begin{aligned} &F_{.05,18,14872} = 1.6 \\ &f = \frac{(3791.555 - 2081.940)/18}{(2081.940/14872)} = 678.5 \\ &f > F_{.05,18,14872} \end{aligned} \tag{10}$$

Further statistical examination of the site-specific v/c coefficients and constants was done using paired t-tests to examine whether sites could be grouped with similar coefficients. It was hoped that by grouping sites, it would be easier to identify characteristics that influence the different levels of peak spreading between these groups. The hypothesis is that each pair of site-specific constants are equal to each other, and likewise for each pair of site-specific slope coefficients. Tests were evaluated for the constants and slope coefficients at the 95% confidence level, with the resulting t-statistics presented in Tables 5 and 6 respectively. The pairs that have t-statistics less than the critical 1.96 are shaded, indicating the hypothesis can't be rejected and we can't say that these sites are unique. Overall, Groton and West Hartford were the only grouping of sites that could not reject the null hypothesis for both the site-specific constants and the site-specific v/c coefficients, although other site groupings were not rejected for either one or the other.

The results show that an important factor for determining the extent of peak-spreading is differentiating between direction of flow, as the constant and v/c coefficient are significantly different for the commute and reverse-commute directions. The model results show the extent of peak-spreading at the various sites studied differs substantially, which might be attributed to the effect of characteristics such as trip purpose distributions, trip distances, and a site's location and physical attributes. The next three chapters discuss these variables and results from more specific hypothesis testing concerning these variables.

Table 5. Paired t-statistics for Constants

	I-95, East Lyme	I-91, Wethersfield	I-84, Newtown	I-84, Manchester	I-95, Norwalk	I-95, Branford	I-95, Groton	I-84, W. Hartford	I-91, Enfield
I-91, Wethersfield	5.748								
I-84, Newtown	0.915	5.62							
I-84, Manchester	-6.117	17.50	8.35						
I-95, Norwalk	23.166	27.56	27.01	47.08					
I-95, Branford	15.842	15.17	17.71	32.96	11.03				
I-95, Groton	3.761	3.04	3.07	14.69	29.06	17.53			
I-84, W. Hartford	3.121	5.23	2.48	16.34	42.49	24.04	1.32		
I-91, Enfield	-2.631	10.18	3.95	3.84	31.61	22.23	8.04	7.67	
I-84, Middlebury	3.952	1.42	3.37	11.91	20.86	12.67	0.97	1.99	7.31

Shading indicates failed t-statistics at 95% confidence level

Table 6. Paired t-statistics for Slopes

	I-95, East Lyme	I-91, Wethersfield	I-84, Newtown	I-84, Manchester	I-95, Norwalk	I-95, Branford	I-95, Groton	I-84, W. Hartford	I-91, Enfield
I-91, Wethersfield	0.285								
I-84, Newtown	2.115	2.60							
I-84, Manchester	10.103	15.84	9.46						
I-95, Norwalk	-19.782	35.35	28.53	53.72					
I-95, Branford	-12.182	20.34	18.19	36.42	12.51				
I-95, Groton	1.288	1.52	1.37	14.51	34.87	20.92			
I-84, W. Hartford	2.424	3.94	0.17	16.08	51.53	28.95	1.93		
I-91, Enfield	6.499	8.49	5.00	3.53	34.54	23.96	7.54	6.88	
I-84, Middlebury	-2.206	3.17	4.77	13.96	18.85	10.49	4.29	5.82	9.46

Shading indicates failed t-statistics at 95% confidence level

Chapter 4: TRIP VARIABLES

Approach

The results of the individual site model estimations showed that factors other than congestion are influencing the spreading of the peak. This chapter focuses on the creation and testing of trip variables which are intended to capture the variability in peak spreading from site to site, and thus, improve the transferability of the model results to other highway links in Connecticut or elsewhere.

The addition of these new variables to the generic, or base model will allow testing of hypotheses regarding these variables. To distinguish between direction of flow (commute vs. reverse-commute), the base model in addition to the main independent variable X contains the directional dummy and directional v/c variables for the reverse-commute direction. Originally it was hypothesized that a trip purpose distribution may help explain the differences in peak spreading between directions. A daily trip purpose distribution for each link was calculated from select link analysis files from PERFORM, but examination of the distributions showed them to be of no significant use. The daily trip purpose distributions did not represent the expected trip purpose distributions within the peak period, and thus, would not help in explaining the differences in peak spreading between directions.

Variables

Characteristics of the trips using links in a network may have an effect on the likelihood of a trip being made in the peak hour. Trip characteristics such as total trip length and elapsed trip time (time spent traveling from origin to study site) are thought to be possible factors influencing the degree of peak spreading at our study sites. To test this assumption, trip length and elapsed trip time distributions at each site were created. Moreover, different distributions among sites may help explain the differences in the degree of spreading among sites. Origin/destination matrices for each site, generated by TRANPLAN select link analysis from PERFORM, along with congested and uncongested skim matrices from ConnDOT's model provided the data with which to create these distributions.

Trip length frequencies were created for each site to test the hypothesis that total trip length is a factor in whether trip makers will change their time of travel. It is hypothesized that it may be easier for an individual making a short trip to adjust his or her time of travel to avoid the highest congestion level than an individual making a longer trip. By calculating percentages for ranges of trip lengths for each link, it was hoped that differences among sites would prove to be a good indicator of the level of peak spreading which will occur at each site.

In order to produce trip length frequencies, select link files have to be used in conjunction with the congested skim matrix in a TRANPLAN function called TRIP

LENGTH FREQUENCY. Since travel times for external-external (E-E) and internal-external/external-internal (I-E/E-I) travel only include the portion of each trip inside the state, a decision was made to use only the internal to internal (I-I) zone trips when computing the frequencies. This was accomplished by modifying the select link files to contain only the I-I zone trips. In place of trip length frequencies for the E-E and I-E/E-I movements, variables were created by totaling the number of trips made by each of these movements and dividing by the total number of trips traversing the study link.

The TRANPLAN code for calculating trip length frequencies for the I-I trips is shown in Appendix 2. The output contains the number of I-I trips associated with trip lengths in one-minute increments. Percentages of trips falling in specified ranges of trip lengths were calculated by dividing the number of I-I trips in that range by the total number of trips traversing the site. These percentages were then used as variables in model estimation. The I-I trip ranges and percentages for each range are shown in Table 7, along with the percentages of E-E and I-E/E-I trips and a statistical summary of the variables.

Elapsed trip time frequencies are similar to trip length frequencies, but only include the time from trip origins to the study link. This was thought to be a more powerful variable than the total trip length, since the time spent travelling before reaching the link may influence the time of departure more than the entire trip length might. Travelling only a short time before reaching the link may allow a traveler to adjust his or her start time to avoid the highest level of congestion, while a longer elapsed time from the start of the trip to the link may make it harder to avoid the peak hour on this link. The steps to calculate these elapsed trip time frequencies are shown below.

1) The highway network link file in PERFORM was modified by connecting an external centroid to the node at both ends of each study link and a new skim matrix was then created, that includes these relocated centroids.
2) TRANPLAN's MATRIX UPDATE function was then used to replace all skim entries with zeros except those with destinations at each relocated centroid. This function was run for each study link in both directions.
3) Travel times from all zones to the study link centroids were then reported using the REPORT MATRIX function in TRANPLAN.
4) The number of trips from each origin zone to each destination zone were then reported using REPORT MATRIX on the select link files for each study link.
5) Origin zone numbers were then matched between the travel time files and the trip files created in Step 3 and 4, creating an elapsed time frequency distribution for each study link.
6) Percentages of trips falling in specified ranges of elapsed trip lengths were calculated by dividing the number of trips in that range by the total number of trips traversing the site. These percentages were then used as variables in model estimation.

The corresponding external centroids connected to study link nodes and TRANPLAN code for executing MATRIX UPDATE and REPORT MATRIX are in Appendix 3a and

Table 7. Trip Length Variables

Site	Number of I-I Trips	Number of I-E / E-I Trips	Number of Thru Trips	Total Number of Trips	% I-I Trips ≤15 min	% I-I Trips >15-≤30 min	% I-I Trips >30-≤60 min	% I-I Trips >60 min	% I-E / E-I Trips	% Thru Trips
I-91 North, Wethersfield	34835	8982	5233	49050	7%	25%	31%	8%	18%	11%
I-91 South, Wethersfield*	35343	8403	5217	48963	8%	24%	31%	8%	17%	11%
I-84 East, Newtown*	12644	6580	7129	26353	2%	16%	23%	7%	25%	27%
I-84 West, Newtown	12137	9973	7107	29217	2%	14%	21%	5%	34%	24%
I-84 East, Manchester*	20456	8291	7747	36494	11%	21%	22%	2%	23%	21%
I-84 West, Manchester	21448	9014	7720	38182	11%	20%	23%	2%	24%	20%
I-95 North, Norwalk*	27594	29011	10814	67419	9%	18%	12%	2%	43%	16%
I-95 South, Norwalk	27962	42542	10781	81285	6%	15%	11%	2%	52%	13%
I-95 North, Branford*	25555	3420	5976	34951	26%	24%	17%	7%	10%	17%
I-95 South, Branford	26132	4044	5958	36134	23%	24%	18%	8%	11%	16%
I-95 North, East Lyme+	12810	7998	5976	26784	3%	15%	19%	10%	30%	22%
I-95 South, East Lyme+	13721	6944	5958	26623	3%	16%	21%	11%	26%	22%
I-95 North, Groton*	7760	10481	5836	24077	10%	17%	5%	1%	44%	24%
I-95 South, Groton	8367	15499	5817	29683	9%	15%	3%	1%	52%	20%
I-84 East, W. Hartford	39530	7965	6592	54087	15%	28%	25%	5%	15%	12%
I-84 West, W. Hartford*	39449	5977	6570	51996	18%	30%	24%	4%	11%	13%
I-91 North, Enfield*	11581	28011	4596	44188	5%	11%	9%	1%	63%	10%
I-91 South, Enfield	11794	34916	4584	51294	4%	10%	8%	1%	68%	9%
I-84 East, Middlebury*	14066	3146	6734	23946	0%	18%	31%	10%	13%	28%
I-84 West, Middlebury	13783	5141	6713	25637	0%	17%	29%	8%	20%	26%
Number of Cases	20	20	20	20	14894	14894	14894	14894	14894	14894
Minimum	7760	3146	4584	23946	0%	10%	3%	1%	10%	9%
Maximum	39530	42542	10814	81285	26%	30%	31%	11%	68%	28%
Mean	20848	12817	6653	40318	10.43%	20.23%	19.07%	4.76%	28.10%	17.29%
Standard Deviation	10505	11331	1670	15627	7.22%	5.32%	8.30%	3.20%	17.05%	5.73%

* indicates commute direction in the peak period.

+ no clear direction at East Lyme.

3b respectively. The elapsed trip time ranges and percentages for each range are shown in Table 8, as well as a statistical summary of the variables.

Hypotheses Testing with Trip Variables

Models were first run with the trip length distribution variables. Only the E-E or *THRU* variable was used in conjunction with the I-I variables due to colinearity problems associated with having all movements incorporated in estimation. Trip length variables for the I-I distribution were created by determining appropriate ranges of trip lengths and calculating the percentage of trips in each range. Initially, trip length ranges were set at 0-15 (*INT0_15*), 15-30 (*INT15_30*), 30-60 (*INT30_60*), and over 60 minutes (*INT60UP*) with the idea that these ranges were comparable to short, medium, medium-to-long, and long range trips. Alternative trip length ranges were then set by determining the 25th, 50th,and 75th percentiles of the overall trip length distribution. These ranges are 0-18 (*INT0_18*), 18-28 (*INT18_28*), 28-42 (*INT28_42*), and over 42 minutes (*INT42UP*) which are similar to the initial ranges, except with regard to the longer range trips. The model forms used in estimation for the first and second representations, respectively, are shown below.

$$\begin{aligned} Y &= C_0 + C_1 D_R + C_2(INT0\_15) + C_3(INT15\_30) \\ &+ C_4(INT30\_60) + C_5(INT60UP) + C_6(THRU) + b_0 X + b_1 X_R \end{aligned} \tag{11}$$

$$\begin{aligned} Y &= C_0 + C_1 D_R + C_2(INT0\_18) + C_3(INT18\_28) \\ &+ C_4(INT28\_42) + C_5(INT42UP) + C_6(THRU) + b_0 X + b_1 X_R \end{aligned} \tag{12}$$

The elapsed trip time variables were then used in model estimation. Again, only the E-E or *THRU* variable was used in conjunction with the I-I variables due to colinearity problems associated with having all movements incorporated in estimation. Elapsed trip time variables were created using 5-minute interval ranges. A total of seven variables were created with the first six having a range of 5-minutes. The last variable contains the percentage of trips with elapsed trip times greater than 30-minutes. The model form used in estimation is shown below.

$$\begin{aligned} Y &= C_0 + C_1 D_R + C_2(ELA0\_5) + C_3(ELA5\_10) + C_4(ELA10\_15) \\ &+ C_5(ELA15\_20) + C_6(ELA20\_25) + C_7(ELA25\_30) + C_8(ELA30UP) \\ &+ C_9(THRU) + b_0 X + b_1 X_R \end{aligned} \tag{13}$$

A listing and description of all variables used in model estimations covered in this chapter is shown in Table 9.

Table 8. Elapsed Trip Time Variables

Site	% of trips ≤5 min	% of trips >5-≤10 min	% of trips >10-≤15 min	% of trips >15-≤20 min	% of trips >20-≤25 min	% of trips >25-≤30 min	% of trips >30 min
I-91 North, Wethersfield	22%	2%	13%	8%	8%	4%	14%
I-91 South, Wethersfield*	5%	27%	17%	8%	5%	6%	5%
I-84 East, Newtown*	9%	25%	7%	2%	1%	4%	0%
I-84 West, Newtown	1%	7%	3%	7%	3%	7%	12%
I-84 East, Manchester*	12%	10%	9%	12%	5%	2%	5%
I-84 West, Manchester	6%	27%	9%	4%	5%	1%	3%
I-95 North, Norwalk*	7%	18%	10%	2%	3%	0%	0%
I-95 South, Norwalk	9%	10%	4%	3%	3%	1%	5%
I-95 North, Branford*	19%	12%	19%	7%	5%	2%	9%
I-95 South, Branford	26%	18%	9%	6%	3%	3%	8%
I-95 North, East Lyme+	4%	6%	10%	6%	6%	3%	14%
I-95 South, East Lyme+	8%	11%	13%	5%	4%	1%	10%
I-95 North, Groton*	2%	17%	5%	3%	1%	1%	2%
I-95 South, Groton	15%	6%	4%	3%	1%	0%	0%
I-84 East, W. Hartford	19%	12%	14%	7%	5%	7%	9%
I-84 West, W. Hartford*	27%	22%	9%	6%	6%	2%	4%
I-91 North, Enfield*	5%	5%	5%	3%	4%	1%	4%
I-91 South, Enfield	6%	14%	1%	0%	0%	0%	0%
I-84 East, Middlebury*	2%	16%	10%	9%	9%	7%	6%
I-84 West, Middlebury	0%	15%	16%	4%	4%	4%	10%
Number of Cases	14894	14894	14894	14894	14894	14894	14894
Minimum	0%	2%	1%	0%	0%	0%	0%
Maximum	27%	27%	19%	12%	9%	7%	14%
Mean	11.01%	15.21%	9.96%	5.80%	4.28%	2.82%	5.33%
Standard Deviation	8.38%	6.74%	4.59%	2.98%	2.01%	2.28%	3.45%

* indicates commute direction in the peak period.

+ no clear direction at East Lyme.

Table 9. List and Description of Variables in Model Estimations

Variable Name	Variable Description
$Y = \ln(P - 1/4)$	Dependent variable which is the natural log of the ratio of peak hour to peak period volume minus one-fourth
D_R	Dummy variable having the value of 1 if the direction of travel is in the reverse-commute direction
X	Volume to capacity ratio for the peak period
X_R	Volume to capacity ratio for the peak period if the direction of travel is in the reverse-commute direction
INT0_15	Pct. of I-I trips with travel times 0-15 minutes divided by one-hundred
INT15_30	Pct. of I-I trips with travel times 15-30 minutes divided by one-hundred
INT30_60	Pct. of I-I trips with travel times 30-60 minutes divided by one-hundred
INT60UP	Pct. of I-I trips with travel times >60 minutes divided by one-hundred
INT0_18	Pct. of I-I trips with travel times 0-18 minutes divided by one-hundred
INT18_28	Pct. of I-I trips with travel times 18-28 minutes divided by one-hundred
INT28_42	Pct. of I-I trips with travel times 28-42 minutes divided by one-hundred
INT42UP	Pct. of I-I trips with travel times >42 minutes divided by one-hundred
ELA0_5	Pct. of I-I trips with elapsed travel time of 0-5 divided by one-hundred
ELA5_10	Pct. of I-I trips with elapsed travel time of 5-10 divided by one-hundred
ELA10_15	Pct. of I-I trips with elapsed travel time of 10-15 divided by one-hundred
ELA15_20	Pct. of I-I trips with elapsed travel time of 15-20 divided by one-hundred
ELA20_25	Pct. of I-I trips with elapsed travel time of 20-25 divided by one-hundred
ELA25_30	Pct. of I-I trips with elapsed travel time of 25-30 divided by one-hundred
ELA30UP	Pct. of I-I trips with elapsed travel time of >30 divided by one-hundred
THRU	Pct. of E-E (thru) trips at each site divided by one-hundred

Results and Discussion

Results from using the trip length variables in model estimation are shown in Table 10. The addition of the trip length distribution variables produces a significantly better model than without these variables. This is shown with F-test comparisons to the base model. The first test is for the first set of ranges and the second is for the second set of ranges.

$$F_{.05,5,14885} = 2.21$$

$$f_1 = \frac{(3791.555 - 3343.941)/5}{(3343.941/14885)} = 398.50 \tag{14}$$

$$f_2 = \frac{(3791.555 - 3195.471)/5}{(3195.471/14885)} = 555.33 \tag{15}$$

$$f_1 > F_{.05,5,14885}$$
$$f_2 > F_{.05,5,14885}$$

Therefore, the null hypothesis is rejected in both cases.

It was hypothesized that a site with a greater number of short trips would tend to have a peak profile that was more spread. The signs of the coefficients on the trip length variables determine if this hypothesis is accepted. The coefficients for the first range in both models is negative, indicating a site with a higher percentage of short trips will tend to have a peak period profile which is more spread than one with a lower percentage of short trips. The coefficient for the second range of trip lengths in both models is positive, indicating that a site with a higher percentage of these trip lengths will tend to have a peak period profile which is more peaked than one with a lower percentage of trips in this range. The coefficient for the third range of trip lengths is positive in the first set of ranges, but negative in the second set of ranges. The positive value is more consistent with our hypothesis, in that a majority of these trips are likely to occur in the peak hour, therefore contributing to a higher peaking of the profile. The coefficient for the last range of trip lengths is negative in both models, indicating an increase in spreading of the peak profile with a higher percentage of these long trip lengths. The coefficient for the *THRU* variable is significant and negative, indicating a site which has a high percentage of "thru" trips will tend have a peak profile that is more spread. This sign seems to be correct, as an individual making a thru trip would likely make an effort to avoid the peak hour congestion. The thru trip may be more discretionary in nature, allowing for advance planning of departure time in order to minimize the time traveling under congested conditions.

Overall, these two models tend to support the hypothesis that the peak period distribution at a site with a high percentage of short trips will tend to be more spread. The negative coefficient on the short trip range suggests a number of these trips are avoiding the peak hour, while the positive coefficient on the medium range trips suggests a number of these trips cannot avoid the peak hour. The possible inconsistency in the hypotheses lies in the change from the medium to longer trips, as the coefficient once again becomes negative. This change in sign may be appropriate, as a majority of these trips could be arriving at the site before or beyond the peak hour due to their length, depending on the distance between their departure location and the study link. For example, a trip maker who completes a majority of their trip before reaching the study site may arrive beyond the peak hour due to the time it takes to reach the site. Moreover, trip makers may be adjusting their departure time to avoid the peak hour at a site, as any increase in trip time associated with traveling in the peak hour at that site would make the trip length unacceptable. The elapsed trip time variables attempt to account for these possibilities by only considering the travel time from an origin to a study site.

Results from estimation with the elapsed trip time variables are shown in Table 11. The addition of the elapsed trip time variables produces a significantly better model than without these variables. This is shown with the following F-test:

$$F_{.05,8,14882} = 1.94$$

$$f = \frac{(3791.555 - 2762.600)/8}{(2762.600/14882)} = 692.87 \qquad (16)$$

$$f > F_{.05,8,14882}$$

Therefore, the null hypothesis is rejected.

Although all elapsed trip time coefficients are significant, their effects (signs) are not consistent with the hypothesis, as they are positive through the first two ranges and then alternate from range to range until the last range where the coefficient stays positive. There is no clear pattern that can be associated with the behavior of the trip makers from these coefficients.

The possible reason that these trip length distribution variables are giving mixed results (a better fit with inconsistent coefficient signs) is that the outputs from a daily assignment were used to create the variables, since a peak period assignment was not at our disposal. The daily trip length distribution may not be representative of the peak period distribution due to temporal variations in trip making patterns during the day.

The next chapter discusses the creation of site attribute and location variables and results from model estimations with these variables.

Table 10. Regression Results: Trip Length Distribution Variables

	Base Model		**Base Model with Trip Length Variables - Alt. 1**		**Base Model with Trip Length Variables - Alt. 2**	
	Coefficient	t-statistic	*Coefficient*	t-statistic	*Coefficient*	t-statistic
Constants (*C*)						
Base	-2.450	-66.246	-1.813	-41.419	-1.967	-46.950
D_R	-0.464	-4.035	-1.975	-17.431	-1.715	-15.707
Trip Variables						
THRU	-	-	-0.896	-11.623	-1.851	-23.876
INT0_15	-	-	-0.870	-7.243	-	-
INT15_30	-	-	2.996	14.672	-	-
INT30_60	-	-	1.203	9.740	-	-
INT60UP	-	-	-5.589	-26.461	-	-
INT0_18	-	-	-	-	-2.837	-33.139
INT18_28	-	-	-	-	11.893	41.356
INT28_42	-	-	-	-	-2.503	-10.346
INT42UP	-	-	-	-	-2.158	-19.034
Slopes (*b*)						
X	-1.402	-25.193	-2.843	-44.290	-2.502	-41.680
X_R	-0.351	-1.734	2.024	10.246	1.413	7.418
Statistics						
R-squared	0.198		0.292		0.324	
SSR	934.305		1381.919		1530.390	
SSE	3791.555		3343.941		3195.471	
F-statistic	1223.051		768.923		891.099	
Degrees of Freedom	14,890		14,885		14,885	

Shading indicates failed t-statistics at 95% confidence level

Table 11. Regression Results: Elapsed Trip Time Variables

	Base Model		**Base Model with Elapsed Trip Time Variables**	
	Coefficient	t-statistic	*Coefficient*	t-statistic
Constants (*C*)				
Base	-2.450	-66.246	-1.162	-27.142
D_R	-0.464	-4.035	-0.823	-8.268
Trip Variables				
THRU	-	-	-3.393	-43.983
ELA0_5	-	-	1.054	17.753
ELA5_10	-	-	0.929	9.717
ELA10_15	-	-	-4.931	-38.537
ELA15_20	-	-	8.331	45.094
ELA20_25	-	-	-9.771	-35.692
ELA25_30	-	-	6.690	24.608
ELA30UP	-	-	0.689	3.125
Slopes (*b*)				
X	-1.402	-25.193	-2.446	-42.196
X_R	-0.351	-1.734	-0.393	-2.250
Statistics				
R-squared	0.198		0.415	
SSR	934.305		1963.260	
SSE	3791.555		2762.600	
F-statistic	1223.051		961.454	
Degrees of Freedom	14,890		14,882	

Shading indicates failed t-statistics at 95% confidence level

Chapter 5: SITE VARIABLES

Variables

Like the trip characteristics discussed in Chapter 4, physical attribute and location characteristics of a site may influence the spreading of the peak. To test whether or not some site characteristics affect the spreading of the peak, several location and site attribute variables were created and included as independent variables in estimation.

A variable called "Distance From Central Business District (CBD)" was calculated for each site to test the hypothesis that the relative location of a site within the metropolitan area may help predict the amount of spreading that occurs there. CBDs for each link were selected by first identifying the commute direction of the PM peak period traffic flow at each study link. That flow direction was then associated with a CBD thought to be the economic center of each respective region in Connecticut. Three general regions were identified in Connecticut with regard to where the study sites are located: the Hartford region, the Southwest region, and the Southeast region. For these regions, the major CBDs identified were Hartford, New York City, and New London, respectively. Upon further consideration, other CBDs were identified for Southwest Region sites due to the multi-centric nature of that region. The elapsed trip time distribution and the number of I-I and I-E/E-I trips at each site helped in this review, by identifying travel times from origins to the study sites, the amount of thru travel, and the amount of out-of-state travel, all clues as to where a majority of the trips at a site could be originating.

The PM peak commute direction of flow for the four sites around Hartford indicate most of this travel is coming from Hartford, or the Capitol region surrounding Hartford. Therefore, Hartford as before was assigned to these sites as the CBD. Review of the elapsed trip time distribution for these sites seems to validate this CBD association to Hartford. Forty-nine percent of the trips traveling west at the West Hartford site have elapsed trip times of 10 minutes or less. Similarly, forty-nine percent of the trips traveling south at the Wethersfield site have elapsed trip times of 15 minutes or less. At the Manchester site, forty-three percent of the trips have elapsed trip times of 20 minutes or less. These elapsed trip times are consistent with the physical distances from Hartford to each study site. A similar direct observation cannot be made about the Enfield site, as only twenty-seven percent of its total travel is comprised of I-I trips. However, the Enfield site has sixty-three percent of its total travel being comprised of I-E/E-I trips, indicating that a great deal of commuting is being done from Massachusetts to Connecticut and then back to Massachusetts in the evening. It is believed the Capitol region is attracting a great deal of these trips.

Previously the Southeast sites were associated with New London as the CBD for this area. The elapsed trip times along with the direction of flow also seem to validate this assignment. Thirty-two percent of the commute direction of travel at the Groton site is comprised of I-I trips, with more than half having elapsed trip times of 10 minutes or less. The site at Groton also has forty-four percent of its total travel being comprised of I-E/E-I trips, indicating a great deal of travel from out of state, which is similar to the Enfield site situation. For the East Lyme site, a clear commute direction in the PM peak period could not be seen. However, there are shorter elapsed trip times in the southbound direction of travel, indicating trips are being produced somewhere not far east of East Lyme. Therefore, the association with New London for these two sites seems appropriate.

New York City was previously identified as being the major economic center for the Southwest region. This association to the Norwalk site is likely correct, as the commute direction has forty-three percent of its total travel being comprised of I-E/E-I trips. This observation, along with the physical location of the Norwalk site confirms New York City as the CBD to be measured from.

The remaining sites that comprise the rest of that Southwest region had also been previously associated with New York City. Review of the elapsed trip times at these sites indicates this may not have been the most appropriate assumption. The commute direction at the Middlebury site has the highest percentage of E-E trips among all sites. Forty-one percent of the travel at this site is comprised of either E-E or I-E/E-I trips. The remaining trips (I-I) have elapsed trip times that are fairly evenly distributed within our ranges. The commute direction at the Newtown site has thirty-four percent of its travel with elapsed trip times of 10 minutes or less. On the other hand, fifty-two percent of its total travel is comprised of either E-E or I-E/E-I trips. Danbury was the only other viable city identified that provides some consistency between the physical distance and the elapsed trip times. Therefore, Danbury was also chosen as a potential CBD for these two sites. The Branford site was also previously associated with New York City. Review of the elapsed trip times at the Branford site shows fifty percent of the trips in the commute direction have elapsed trip times of 15 minutes or less. The Branford site also has the least percentage of I-E/E-I trips among all sites. These observations led to New Haven and Bridgeport being identified as viable CBD's for the Branford site. Although New Haven is considered as a CBD here, the travel times may be indicative of commuters taking the train to New Haven and then beginning an auto trip.

For the most part, the elapsed trip time distributions and percentage of thru and out-of-state travel have helped confirm some of our previous CBD associations, while also identifying other CBD associations that should be considered. Our assumption which may or may not be true is that the daily travel patterns produced by PERFORM are representative of the PM peak period. Overall, three different CBD associations to the study sites were identified. The commute direction of flow, CBD associations, and distance between the CBD and the study sites for each association is shown in Table 12.

A variable containing the number of lanes at each site was used for testing the hypothesis that the number of lanes at a site may help predict the amount of spreading, and thus, explain some of the variation in peak spreading among sites. As discussed previously, the number of lanes at each site was counted using a Connecticut photolog station for the purpose of calculating capacity at each site. The number of lanes at each site is also shown in Table 12.

Hypotheses Testing with Site Variables

Models were first run with the three *DISTCBD* variables. The model forms are shown below:

$$Y = C_0 + C_1 D_R + C_2(DISTCBD1) + b_0 X + b_1 X_R \qquad (17)$$

$$Y = C_0 + C_1 D_R + C_2(DISTCBD2) + b_0 X + b_1 X_R \qquad (18)$$

$$Y = C_0 + C_1 D_R + C_2(DISTCBD3) + b_0 X + b_1 X_R \qquad (19)$$

It was hypothesized that a site with a greater value of *DISTCBD* would have a lower peak hour to peak period proportion at low congestion levels than a site with a smaller value for it. This is supported by the idea that a majority of the trips leaving the CBD in the PM peak period will concentrate within a relatively small range of time on a link that is close to the CBD, due to a high percentage of work trips. As these trips progress outward from the CBD, additional traffic comprised of other trip purposes joins in, causing a link that is farther away to have trips arriving over a less concentrated, but longer period of time. Thus, a site farther from the CBD for its area would tend to have a peak profile that is more evenly distributed at low levels of congestion than a site with a closer CBD.

It was also hypothesized that a site with a greater number of lanes would tend to have a higher peak hour to peak period proportion at low levels of congestion than a site with fewer lanes. This is supported by the same basic idea, that a site with more lanes tends to be located closer to the CBD for its area. In other words, a site that is farther from the CBD for its area tends to have fewer lanes, and thus, less peaking of traffic. This relationship between the *DISTCBD1* and *LANES* variables was identified by the high correlation between the two variables. The LANES model formulation is shown below:

$$Y = C_0 + C_1 D_R + C_2(LANES) + b_0 X + b_1 X_R \qquad (20)$$

A list and description of the variables used in model estimations covered in this chapter are shown in Table 13.

Table 12. Associated CBD's and Number of Lanes at Each Site

Site	3-7 PM Commute Direction of Flow	Associated CBD			Distance From CBD (mi)			
		Alt - 1	Alt - 2	Alt -3	Alt - 1	Alt - 2	Alt - 3	Number of Lanes
I-91, Wethersfield	South	Hartford	Hartford	Hartford	6	6	6	4 NB, 3 SB
I-84, W. Hartford	West	Hartford	Hartford	Hartford	4	4	4	3
I-84, Manchester	East	Hartford	Hartford	Hartford	12	12	12	3
I-91, Enfield	North	Hartford	Hartford	Hartford	19	19	19	3
I-95, Groton	North	New London	New London	New London	8	8	8	2 NB, 3 SB
I-95, East Lyme	No Clear Direction	New London	New London	New London	12	12	12	2
I-95, Norwalk	North	NYC	NYC	NYC	39	39	39	3
I-84, Middlebury	East	NYC	Danbury	Danbury	75	21	21	2
I-84, Newtown	East	NYC	Danbury	Danbury	59	8	8	2
I-95, Branford	North	NYC	New Haven	Bridgeport	83	8	28	2
Number of Cases					14,894	14,894	14,894	14,894
Minimum					4	4	4	2
Maximum					83	39	39	4
Mean					30.95	14.40	16.99	2.62
Standard Deviation					28.94	11.58	12.09	0.53

Table 13. List and Description of Variables in Model Estimations

Variable Name	Variable Description
$Y = \ln(P - 1/4)$	Dependent variable which is the natural log of the ratio of peak hour to peak period volume minus one-fourth
D_R	Dummy variable having the value of 1 if the direction of travel is in the reverse-commute direction
X	Volume to capacity ratio for the peak period
X_R	Volume to capacity ratio for the peak period if the direction of travel is in the reverse-commute direction
LANES	Number of lanes at each site
D_{LANES}	Dummy variable having the value of 1 if the number of lanes is ≥ 3
X_{LANES}	Volume to capacity ratio for the peak period if $D_{LANES} = 1$
DISTCBD1	Distance from CBD to study site for Alt - 1 divided by one-hundred
DISTCBD2	Distance from CBD to study site for Alt - 2 divided by one-hundred
DISTCBD3	Distance from CBD to study site for Alt - 3 divided by one-hundred
CBD1	Dummy variable having the value of 1 if *DISTCBD1* ≤ .25
CBD2	Dummy variable having the value of 1 if *DISTCBD1* > .25 and ≤ .50
X_{CBD1}	Volume to capacity ratio for the peak period if *CBD1* = 1
X_{CBD2}	Volume to capacity ratio for the peak period if *CBD2* = 1

Results and Discussion

Results from the model estimations using Equations 17, 18, 19, and 20 are shown in Table 14. Once again, the Base model is shown for comparison purposes. The addition of the *DISTCBD* variable produces significantly better models than without the variable. This is shown below with the F-test comparisons of the DISTCBD1, DISTCBD2, and DISTCBD3 models to the base model, respectively.

$$
\begin{aligned}
&F_{.05,1,14889} = 3.84 \\
&f_1 = \frac{(3791.555 - 3406.377)/1}{(3406.377/14889)} = 1683.58 \\
&f_2 = \frac{(3791.555 - 3320.169)/1}{(3320.169/14889)} = 2113.88 \\
&f_3 = \frac{(3791.555 - 3157.513)/1}{(3157.513/14889)} = 2989.77 \\
&f_1 > F_{.05,1,14889} \\
&f_2 > F_{.05,1,14889} \\
&f_3 > F_{.05,1,14889}
\end{aligned}
\tag{21}
$$

All three *DISTCBD* coefficients are negative and significant indicating that a site which is further from the CBD for its area will tend to have a peak period profile which is flatter at low levels of congestion than a site that is closer. This is consistent with the hypothesis. Although a direct F-test cannot be computed between the three CBD models, an increase in the explanatory value of the model is seen as we progress from the DISTCBD1 model to the DISTCBD3 model. This increase in explanatory value is attributed to the reordering of sites relative to CBD distance, which results in Norwalk having the largest CBD distance among all sites. Norwalk tends to have the most peak spreading among all sites. Having the largest CBD distance results in an increase in the correlation between the CBD distance and the dependent variable resulting in a better model fit, which is seen from the increase of the t-statistic on the *DISTCBD2* and *DISTCBD3* variables and the increase in R-squared. The DISTCBD1 and DISTCBD3 models are viewed as better than the DISTCBD2 model, since the latter has an adverse effect on the X_{REV} coefficient, causing it to become even more insignificant than without the *DISTCBD2* variable. The DISTCBD1 and DISTCBD3 models cause the X_{REV} coefficient to become significant and positive. This is may be due to the CBD for each site being identified with the commute direction of flow, causing the reverse-commute slope variable to capture the difference between directions. The positive sign is consistent with the reverse-commute direction being less sensitive to increasing congestion.

Table 14. Regression Results: DISTCBD and LANES Variables

	Base Model		DISTCBD1 Model		DISTCBD2 Model		DISTCBD3 Model		LANES Model	
	Coeff.	t-statistic	*Coeff.*	t-statistic	*Coeff.*	t-statistic	*Coeff.*	t-statistic	*Coeff.*	t-statistic
Constants (*C*)										
Base	-2.450	-66.246	-1.998	-54.397	-2.120	-59.997	-2.050	-59.356	-2.662	-71.272
D_R	-0.464	-4.035	-1.232	-11.148	-0.739	-6.864	-1.168	-11.054	-1.102	-9.515
Site-Specific Variables										
DISTCBD1	-	-	-0.579	-41.032	-	-	-	-	-	-
DISTCBD2	-	-	-	-	-1.553	-45.977	-	-	-	-
DISTCBD3	-	-	-	-	-	-	-1.760	-54.679	-	-
LANES	-	-	-	-	-	-	-	-	0.197	23.976
Slopes (*b*)										
X	-1.402	-25.193	-1.838	-34.168	-1.581	-30.286	-1.593	-31.292	-1.851	-32.064
X_R	-0.351	-1.734	1.058	5.434	0.200	1.056	1.053	5.653	0.646	3.187
Statistics										
R-squared	0.198		0.279		0.297		0.332		0.228	
SSR	934.305		1319.483		1405.691		1568.347		1075.254	
SSE	3791.555		3406.377		3320.169		3157.513		3650.606	
F-statistic	1223.051		1441.839		1575.924		1848.854		1096.357	
Degrees of Freedom	14,890		14,889		14,889		14,889		14,889	

Shading indicates failed t-statistics at 95% confidence level

The addition of the *LANES* variable to the base model also produced a significantly better model than without the variable. This is shown below with the F-test comparison to the base model.

$$F_{.05,1,14889} = 3.84$$
$$f = \frac{(3791.555 - 3650.606)/1}{(3650.606/14889)} = 574.86 \qquad (22)$$
$$f > F_{.05,1,14889}$$

The *LANES* coefficient is positive and significant indicating that a site with more lanes will tend to have a peak period profile that is more peaked at low levels of congestion than a site with less lanes. This is also consistent with the hypothesis. Another important observation made from the results is that the X_{REV} coefficient becomes positive and significant when this variable is included.

DISTCBD with LANES

The results of incorporating both the *DISTCBD* and *LANES* variables in the same model are shown in Table 15. The model formulation is shown below:

$$Y = C_0 + C_1 D_R + C_2(DISTCBD) + C_3(LANES) + b_0 X + b_1 X_R \qquad (23)$$

The DISTCBD1 and LANES model shows the results of a high correlation between the two variables. Because of the correlation between the two variables, the *LANES* coefficient becomes insignificant. Moreover, the model coefficients are essentially identical to the model estimated with the addition of the *DISTCBD1* variable only. Addition of the *LANES* variable to the *DISTCBD1* model provides no increase in the explanatory power of the model, which is shown below with the F-test comparison between this model and the model containing only the *DISTCBD1* variable. Therefore, the *LANES* variable should not be included with the *DISTCBD1* variable in estimation.

$$F_{.05,1,14888} = 3.84$$
$$f = \frac{(3406.377 - 3406.377)/1}{(3406.377/14888)} = 0 \qquad (24)$$
$$f < f_{.05,1,14888}$$

The *DISTCBD2* variable is not as correlated with the *LANES* variable, and therefore, no adverse effect on either variable is seen. Actually, the addition of the *LANES* variable to the DISTCBD2 model significantly increases the explanatory value of the model and causes the reverse commute slope correction coefficient to become significant. The F-test comparison between this model and the DISTCBD2 model is

Table 15. Regression Results: DISTCBD with LANES Variable

	Base Model		DISTCBD1 and LANES Model		DISTCBD2 and LANES Model		DISTCBD3 and LANES Model	
	Coeff.	t-statistic	*Coeff.*	t-statistic	*Coeff.*	t-statistic	*Coeff.*	t-statistic
Constants (*C*)								
Base	-2.450	-66.246	-1.999	-48.275	-2.378	-69.246	-2.246	-64.211
D_R	-0.464	-4.035	-1.232	-11.005	-1.756	-16.575	-1.714	-16.097
Site-Specific Variables								
DISTCBD1	-	-	-0.579	-32.672	-	-	-	-
DISTCBD2	-	-			-1.853	-55.940	-	-
DISTCBD3	-	-			-	-	-1.718	-54.231
LANES	-	-	2.382E-04	0.024	0.298	38.710	0.174	23.082
Slopes (*b*)								
X	-1.402	-25.193	-1.839	-32.975	-2.294	-43.225	-1.984	-37.569
X_R	-0.351	-1.734	1.059	5.395	1.811	9.764	1.898	10.166
Statistics								
R-squared	0.198		0.279		0.362		0.355	
SSR	934.305		1319.483		1709.303		1677.440	
SSE	3791.555		3406.377		3016.557		3048.420	
F-statistic	1223.051		1153.394		1687.228		1638.471	
Degrees of Freedom	14,890		14,888		14,888		14,888	

Shading indicates failed t-statistics at 95% confidence level

shown below:

$$F_{.05,1,14888} = 3.84$$
$$f = \frac{(3320.169 - 3016.557)/1}{(3016.557/14888)} = 1498.46 \qquad (25)$$
$$f > F_{.05,1,1488}$$

The *DISTCBD3* variable is the least correlated variable with the *LANES* variable among the DISTCBD variables. The results show a significant increase in the explanatory value of the model with the addition of the *LANES* variable to the DISTCBD3 model. The F-test comparison between this model and the DISTCBD3 model is shown below:

$$F_{.05,1,14888} = 3.84$$
$$f = \frac{(3157.513 - 3048.420)/1}{(3048.420/14888)} = 532.79 \qquad (26)$$
$$f > F_{.05,1,1488}$$

The next section analyses the CBD and LANES variables in categorical and dummy variable representations as an alternative to using the actual values of the variables which forces a linear relationship. This will help to determine how these variables are best represented.

Representations of LANES and DISTCBD

The *LANES* and *DISTCBD* variables were also represented using dummy and categorical variable representations. This allowed the testing of different representations of the variables and creation of slope variables, allowing for non-linear relationships. A variable D_{LANES} was created and coded as 1 if the number of lanes was greater than or equal to 3 and 0 otherwise. A slope effect for the number of lanes at a site was incorporated by creating a variable X_{LANES}, equal to X if the D_{LANES} variable is equal to 1. Three categories of distance were used for the creation of two new variables *CBD1* and *CBD2* using the *DISTCBD1* variable. *CBD1* was coded as 1 for CBD distances less than or equal to 25 miles. *CBD2* was coded as 1 for CBD distances greater than 25 and less than or equal to 50 miles. To incorporate the effect of the CBD distance on the slope, two additional variables were created. The first variable X_{CBD1} is equal to X if the *CBD1* variable is equal to 1, otherwise the value is zero. The second variable X_{CBD2} is equal to X if the *CBD2* variable is equal to 1, otherwise the value is zero. The model formulations are shown below:

$$Y = C_0 + C_1 D_R + C_2(D_{LANES}) + b_0 X + b_1 X_R + b_2(X_{LANES}) \qquad (27)$$

$$Y = C_0 + C_1 D_R + C_2(CBD1) + C_3(CBD2) + b_0 X + b_1 X_R + b_2(X_{CBD1}) + b_3(X_{CBD2}) \qquad (28)$$

Results and Discussion

Results from the model estimations using Equation 27 and 28 are shown in Table 16. The addition of the new LANES variables to the base model produces a significantly better model than without these variables which is shown below with the F-test comparison to the base model.

$$F_{.05,2,14888} = 3.00$$
$$f = \frac{(3791.555 - 3682.380)/2}{(3682.380/14888)} = 220.70 \qquad (29)$$
$$f > F_{.05,2,14888}$$

However, the slope coefficient for the X_{LANES} variable is not significant and should not be included. This indicates that sites differing only by their number of lanes are equally sensitive to increasing congestion. A direct F-test comparison to the LANES model cannot be made, but model statistics such as the t-statistics on the *LANES* variables and R-squared values show the linear relationship to be stronger and that using the actual number of lanes at a site for the variable seems to be better. The addition of the new categorical CBD variables allowing for non-linear relationships produces a significantly better model than without these variables which is shown below with the F-test comparison to the base model.

$$F_{.05,4,14886} = 2.37$$
$$f = \frac{(3791.555 - 2606.651)/4}{(2606.651/14886)} = 1691.68 \qquad (30)$$
$$f > F_{.05,4,14886}$$

A direct comparison to the DISTCBD1 model cannot be made, but a large increase in the R-squared value and F-statistic shows this categorical representation is the better model formulation for the *DISTCBD1* variable. This representation of the CBD variable is interesting in that by not forcing a linear relationship and incorporating both intercept and slope effects a better model fit is achieved over the DISTCBD1 model. The intercept is adjusted by the *CBD1* and *CBD2* coefficients. The results indicate that the study sites in the *CBD1* category tend to have the flattest peak period profiles at low congestion levels, while Norwalk (*CBD2*) tends to have a peak period profile that is the most peaked. With regard to the slope effects, the sites in the *CBD1* category tend to be the least sensitive to increasing congestion, while Norwalk (*CBD2*) tends to be the most sensitive to increasing congestion. The ordering of the intercept and slope coefficients are important in that they show these categories are "out of order", meaning they do not line up in their respective order of increasing distance from CBD, validating the non-linear representation. Furthermore, these categories were chosen fairly arbitrarily and might be chosen in such a fashion as to result in a different outcome. Because of this, the categories were examined further to examine other possibilities of groupings that may be more theoretically sound and stable. The results of this re-examination proved to be useful, as groupings of sites by region and area type resulted in models that are statistically and theoretically sound. The next chapter discusses the results from using these regional and area type variables in model estimation.

Table 16. Regression Results: Categorical DISTCBD and LANES Variables

	Base Model		LANES Dummy and LANES Slope Variables		Categorical CBD and CBD Slope Variables Based on DISTCBD1	
	Coeff.	t-statistic	*Coeff.*	t-statistic	*Coeff.*	t-statistic
Constants (*C*)						
Base	-2.450	-66.246	-2.258	-40.801	-1.587	-25.374
D_R	-0.464	-4.035	-0.948	-8.200	-1.776	-18.147
D_{LANES}	-	-	0.157	2.385	-	-
CBD1	-	-	-	-	-1.240	-18.631
CBD2	-	-	-	-	2.165	23.764
Slopes (*b*)						
X	-1.402	-25.193	-1.860	-20.790	-3.079	-30.296
X_R	-0.351	-1.734	0.410	2.025	1.989	11.577
X_{LANES}	-	-	0.049	0.466	-	-
X_{CBD1}	-	-	-	-	2.522	23.347
X_{CBD2}	-	-	-	-	-3.676	-25.408
Statistics						
R-squared	0.198		0.221		0.448	
SSR	934.305		1043.480		2119.210	
SSE	3791.555		3682.380		2606.651	
F-statistic	1223.051		843.766		1728.905	
Degrees of Freedom	14,890		14,888		14,886	

Shading indicates failed t-statistics at 95% confidence level

Chapter 6: REGIONAL AND AREA TYPE MODELS

Approach

Model results of linear and categorical representations of the DISTCBD variable show significant improvements for each over the base model. The results of the categorical representation also show the scale adjustments for the ranges to be "out-of-order", and therefore, non-linear with increasing CBD distance. This indicates the factor influencing peak spreading is probably not CBD distance, but something else that these ranges help to capture. The categorical representation is also viewed to be stronger than the linear, because it allows for different effects on both the intercept and slope of the curve, rather than just adjusting the intercept.

However, the ranges of CBD distance used to create the categorical CBD variables don't necessarily separate the sites into groups with common attributes. Consequently, this chapter presents the results from a re-examination of the sites with regard to the CBD distances of variables *DISTCBD1* and *DISTCBD3*, resulting in slightly different categorical groups defined by similarities such as geographical location and area type characteristics, respectively. Differences among these groups with regard to the amount of peak spreading that occurs is shown by the estimated constants and coefficients for these groups. These differences may then be attributed to specific differences among these groups with respect to regional or area type characteristics of each group.

Regional Model

Table 17 shows the groupings that resulted from the previous categorical breakdown of the DISTCBD1 variable. What is noticed from the sites within these categories is that they are representative of regions within the state. For instance, Range 1 contains all of the Capitol regional sites as well as the Southeast regional sites. The Norwalk site (Range 2) is located in a region that vastly different from the other site

Table 17. Sites in *DISTCBD1* Categories

Range 1	Range 2	Remaining Sites
Wethersfield Manchester W. Hartford Enfield East Lyme Groton	Norwalk	Newtown Branford Middlebury

locations in Connecticut. This was shown by the estimated constants and slope coefficients in the Categorical CBD Model, which indicate Norwalk has significantly higher peak hour to peak period proportions at low levels of congestion and more sensitivity to increasing congestion than the other categories. The third category of sites contains the sites in the Southwest region of Connecticut. With this observation a new set of regional variables was created, separating the Southeast sites from the Capitol region sites, resulting in the four regional categories shown in Table 18. The model formulation for estimation is shown below:

$$\begin{aligned} Y &= C_0 + C_1 D_R + C_2(CAPITOL) + C_3(SOUTHEAST) \\ &+ C_4(SOUTHWEST) + b_0 X + b_1 X_R + b_2(X_{CAP}) + b_3(X_{SE}) + b_4(X_{SW}) \end{aligned} \tag{31}$$

Table 18. Sites in Regional Categories

Capitol	Southeast	Southwest	Norwalk
Wethersfield Manchester W. Hartford Enfield	East Lyme Groton	Newtown Branford Middlebury	Norwalk

A statistical comparison of the estimated constants and coefficients for the Capitol and Southeast sites will indicate whether or not the Southeast sites should be separated from the Capitol sites. A list and description of all variables used in model estimations covered in this chapter are shown in Table 19.

The results of model estimation using the formulation in Equation 31 are shown in Table 20 along with the Categorical CBD model for comparison purposes. The results show that there are clear differences between regions with respect to the amount of peak spreading that occurs. Examination of the coefficients affecting scale indicates the base site, or Norwalk, tends to have a peak period profile that is more peaked at low levels of congestion than any other region. The Southeast region tends to have the flattest peak period profile at low congestion levels, followed by the Capitol and the Southwest regions, respectively. Examination of the coefficients affecting slope indicates the Southeast region to be the least sensitive region with regard to increasing congestion, followed by the Capitol region. The Southwest is also less sensitive to increasing congestion with respect to Norwalk, but more sensitive than the Capitol and Southeast regions.

An F-test comparison between this model and the Categorical CBD Model can be performed, as the latter is just a restricted version of the Regional Model. The only difference is that in the Regional model we are not forcing the Southeast and Capitol regions to have equal coefficients. The F-test is shown below.

$$\begin{aligned} &F_{.05,2,14884} = 3.00 \\ &f = \frac{(2606.651 - 2529.437)/2}{(2529.437/14884)} = 227.18 \\ &\quad f > F_{.05,2,14884} \end{aligned} \tag{32}$$

Table 19. List and Description of Variables in Model Estimations

Variable Name	Variable Description
$Y = \ln(P - 1/4)$	Dependent variable which is the natural log of the ratio of peak hour to peak period volume minus one-fourth
D_R	Dummy variable having the value of 1 if the direction of travel is in the reverse-commute direction
X	Volume to capacity ratio for the peak period
X_R	Volume to capacity ratio for the peak period if the direction of travel is in the reverse-commute direction
CAPITOL	Dummy variable having the value of 1 for Capitol region sites
$CAPTITOL_{REV}$	Dummy variable equal to D_R multiplied by *CAPITOL*
SOUTHEAST	Dummy variable having the value of 1 for the Southeast region sites
$SOUTHEAST_{REV}$	Dummy variable equal to D_R multiplied by *SOUTHEAST*
SOUTHWEST	Dummy variable having the value of 1 for the Southwest region sites
WEST	Dummy variable having the value of 1 for the West region sites
$WEST_{REV}$	Dummy variable equal to D_R multiplied by *WEST*
BRANFORD or GROUP4	Dummy variable having the value of 1 for the Branford site
$BRANFORD_{REV}$	Dummy variable equal to D_R multiplied by *BRANFORD*
$NORWALK_{REV}$	Dummy variable having the value of 1 for the Norwalk site in the reverse-commute direction of flow
GROUP1	Dummy variable having the value of 1 for the urban area sites
GROUP2	Dummy variable having the value of 1 for the semi-rural sites
GROUP3	Dummy variable having the value of 1 for the suburban sites
X_{CAP}	Volume to capacity ratio for the peak period if $CAPITOL = 1$
$X_{REV\text{-}CAP}$	Volume to capacity ratio for the peak period if $CAPITOL_{REV} = 1$
X_{SE}	Volume to capacity ratio for the peak period if $SOUTHEAST = 1$
$X_{REV\text{-}SE}$	Volume to capacity ratio for the peak period if $SOUTHEAST_{REV} = 1$
X_{SW}	Volume to capacity ratio for the peak period if $SOUTHWEST = 1$
X_W	Volume to capacity ratio for the peak period if $WEST = 1$
$X_{REV\text{-}W}$	Volume to capacity ratio for the peak period if $WEST_{REV} = 1$
X_{BR} *or* X_{GROUP4}	Volume to capacity ratio for the peak period if *BRANFORD* or $GROUP4 = 1$
$X_{REV\text{-}BR}$	Volume to capacity ratio for the peak period if $BRANFORD_{REV} = 1$
$X_{REV\text{-}NORW}$	Volume to capacity ratio for the peak period if $NORWALK_{REV} = 1$
X_{GROUP1}	Volume to capacity ratio for the peak period if $GROUP1 = 1$
X_{GROUP2}	Volume to capacity ratio for the peak period if $GROUP2 = 1$
X_{GROUP3}	Volume to capacity ratio for the peak period if $GROUP3 = 1$

Table 20. Regression Results: Categorical CBD and Regional Variables

	Categorical CBD Model		**Regional Model**	
	Coefficient	t-statistic	*Coefficient*	t-statistic
Constants (C)				
Base	-1.587	-25.374	0.857	10.970
D_R	-1.776	-18.147	-2.116	-21.535
CBD1	-1.240	-18.631	-	-
CBD2	2.165	23.764	-	-
CAPITOL	-	-	-3.338	-43.057
SOUTHEAST	-	-	-3.900	-37.694
SOUTHWEST	-	-	-2.289	-25.437
Slopes (b)				
X	-3.079	-30.296	-7.144	-60.316
X_R	1.989	11.577	2.477	14.439
X_{CBD1}	2.522	23.347	-	-
X_{CBD2}	-3.676	-25.408	-	-
X_{CAP}	-	-	6.157	51.529
X_{SE}	-	-	6.688	41.522
X_{SW}	-	-	3.842	26.909
Statistics				
R-squared	0.448		0.465	
SSR	2119.210		2196.423	
SSE	2606.651		2529.437	
F-statistic	1728.905		1436.049	
Degrees of Freedom	14,886		14,884	

Norwalk is base case

This test indicates we should reject the null hypothesis that there is no difference between the two models. The *Base* constant is also much smaller for the Regional Model, meaning this representation is capturing more of the behavior associated with peak spreading. Therefore, the Regional Model, or unrestricted model with the Southeast region being considered separately from the Capitol region, is the better model. A paired t-test was done between each region for both the constant and slope coefficient. This was done because the Capitol and Southeast regions have similar estimated constants and slope coefficients. The results of these tests shown in Table 21 and 22 reject the null hypothesis that the constant or slope coefficient of any region is equal to the constant or slope coefficient of any other region.

Table 21. Paired t-tests between Regional Constants

	Norwalk	Capitol	Southeast
Capitol	-43.06		
Southeast	-37.69	7.07	
Southwest	-25.44	15.57	17.48

Table 22. Paired t-tests between Regional Slope Coefficients

	Norwalk	Capitol	Southeast
Capitol	51.53		
Southeast	41.52	4.25	
Southwest	26.91	21.30	19.13

Area type model

Similar to the regional categories discussed above, the CBD distances in the DISTCBD3 model were analyzed and grouped based on area type groupings rather than an arbitrary selection of distances. These area type groupings and the ranges of CBD distances are shown in Table 23. These groups are viewed to have different area type characteristics or traits, rather than regional type characteristics. The speculated area or effects for each group are shown in parentheses below each group. The v/c range (min. and max. values) and mean for each group were calculated to examine any distinct differences that might help to explain the area type differences. The results are also shown in Table 23. The first three groups we categorized as urban, semi-rural, and suburban. Group 4 is categorized as having an I-95 effect and Group 5 as having a New York City (NYC) effect. A model was run incorporating group specific dummy and v/c

Results and Discussion

The results from model estimation with the formulation in Equation 33 are shown in Table 24. Again, by not forcing a linear relationship between the variables, a better fit is obtained. For example, Group 2 (semi-rural) tends to have a peak profile that is the most spread at low levels of congestion, meaning it has lower peak hour to peak period

Table 23. Groupings Based on *DISTCBD3* Variable

Group	DISTCBD3 Range	Sites	X		
			Minimum	Maximum	Mean
1 (urban)	0 to 8	W. Hartford Wethersfield	0.50	0.90	0.7231
2 (semi-rural)	8 to 12	Newtown Groton Manchester East Lyme	0.50	0.94	0.6408
3 (suburban)	12 to 21	Enfield Middlebury	0.50	0.81	0.5870
4 (I-95 effect)	21 to 28	Branford	0.50	0.84	0.6617
5 (NYC effect)	28 to 39	Norwalk	0.50	0.82	0.7020

proportions at lower levels of congestion than the rest of the groupings. The Group 3 (suburban) area type constant is very similar to Group 2, indicating a fairly spread peak period profile at a low level of congestion. Group 1 (urban), which has the shortest CBD distances, has a peak period profile that is more peaked at lower levels of congestion than Group 2 and 3. Group 4's area type coefficient is much greater (less negative) than the other three groupings, indicating that the area is more closely associated with Group 5 or the Norwalk site. With regard to the slope coefficients for the groupings, Group 2 tends to be the least sensitive to increasing congestion, followed by Group 3 and Group 1 respectively. Group 4 is much more sensitive to increasing congestion than these groupings, but is less sensitive than Group 5 or the Norwalk site.

From these observations, the groupings are shown to be "out of order", meaning they do not line up in their respective order of 1 to 5, validating the non-linear representation. The coefficients for Group 1 place it after Group 2 and 3. Furthermore, the results indicate that the Branford site should not be included in the Southwest region of sites but should be viewed as an area of its own. Paired t-tests shown in Table 25 and 26 were done between each grouping for both the constant and slope coefficient. This was done because Group 2 and Group 3 have similar estimated constants and slope coefficients. The results of these tests rejected the null hypothesis that the constant or slope coefficient of any area is equal to the constant or slope coefficient of any other area.

Regional model revisited

With the observation that the Branford site is significantly different than the Southwest Sites, the Regional model formulation in Equation (31) was modified by pulling the Branford site out of the Southwest region and renaming the Southwest region as the West region. The resulting model formulation is shown below:

$$
\begin{aligned}
Y &= C_0 + C_1 D_R + C_2(CAPITOL) + C_3(SOUTHEAST) + C_4(WEST) \\
&+ C_5(BRANFORD) + b_0 X + b_1 X_R + b_2(X_{CAP}) + b_3(X_{SE}) + b_4(X_W) \\
&+ b_5(X_{BR})
\end{aligned} \tag{34}
$$

Table 24. Regression Results: Area Type Variables

	Base Model		Area Type Model	
	Coeff.	t-statistic	*Coeff.*	t-statistic
Constants (C)				
Base	-2.450	-66.246	1.427	17.809
D_R	-0.464	-4.035	-2.720	-27.259
GROUP1	-	-	-3.068	-38.768
GROUP2	-	-	-4.617	-52.590
GROUP3	-	-	-4.384	-38.973
GROUP4	-	-	-1.096	-10.246
Slopes (b)				
X	-1.402	-25.193	-7.936	-65.709
X_{REV}	-0.351	-1.734	3.303	19.141
X_{GROUP1}	-	-	5.819	47.852
X_{GROUP2}	-	-	7.818	58.009
X_{GROUP3}	-	-	7.417	40.606
X_{GROUP4}	-	-	1.956	11.591
Statistics				
R-squared	0.198		0.477	
SSR	934.305		2255.275	
SSE	3791.555		2470.585	
F-statistic	1223.051		1235.004	
Degrees of Freedom	14,890		14,882	

Shading indicates failed t-statistics at 95% confidence level.
Norwalk is base case.

Table 25. Paired t-tests between Group Constants

	Group 1	Group2	Group 3	Group 4
Group 2	23.12			
Group 3	13.57	2.44		
Group 4	20.84	34.80	26.72	
Group 5	38.77	52.59	38.97	10.25

Table 26. Paired t-tests between Group Slope Coefficients

	Group 1	Group2	Group 3	Group 4
Group 2	19.88			
Group 3	10.05	2.50		
Group 4	25.97	36.86	27.15	
Group 5	47.85	58.01	40.61	11.59

The results of the re-estimation are shown in Table 27. Separation of the Branford site produces a significantly better model than including the site in the Southwest region. This is shown with the F-test comparison between the two Regional models below:

$$F_{.05,2,14882} = 3.00$$
$$f = \frac{(2529.437 - 2457.523)/2}{(2457.523/14882)} = 217.74 \qquad (35)$$
$$f > F_{.05,2,14882}$$

The coefficients for the Branford site and West regions are significantly different from each other. Results of paired t-tests (Table 28 and 29) between the regions showed that the null hypothesis could not be rejected for the constants between the Capitol region and West region. Their slope coefficients however, were shown to be significantly different. Attributing the differences in the amount of peak spreading that occurs among these regions to specific characteristics within each region is difficult. However, each of these regions of the state in general have different labor forces or specific attributes that may be affecting the amount of peak spreading that occurs. For example, a number of the jobs in the Capitol region are comprised of Government and Insurance related jobs. The Southeast has a large number of jobs relating to the support of the U.S. military. The Southwest and West regions are the most densely populated parts of the state. The Branford and Norwalk sites along the I-95 corridor possess unique attributes in an area of the state that is highly congested and highly influenced by the New York City metropolitan area. Clearly, different travel patterns and socioeconomic characteristics of the travelers in these regions must exist from these regional differences alone. Although it is beyond the scope of this project to clearly identify the distinguishing differences in these regions, these observable differences and others like them are likely contributing to the amount of peak spreading that occurs.

Best Model

These regional models are viewed as a better alternative to the Area type models, which may not be as easily defined and supported as different regions of the state may be. Therefore, the regional model was chosen as the "best" representation, and a new regional model was run with separate variables for the reverse-commute direction. This allowed each region's reverse-commute constant and slope coefficient to be estimated relative to the commute direction in that region, rather than an aggregated difference estimated for all sites. Essentially, a unique model for each region in the commute and reverse-commute direction of flow was estimated.

Results of the final model estimation are shown in Table 30. The estimation showed that the reverse-commute constant for the Southeast and reverse-commute slope correction coefficients for the Southeast and West regions were insignificant. Essentially no difference is seen between the reverse-commute and commute directions of travel in the Southeast region in terms of peak spreading. These variables were then excluded from estimation and the results of final estimation are shown as the Final Model in Table 30.

Table 27. Regression Results: Final Regional Variables

	Regional Model		**Final Regional Model**	
	Coeff.	t-statistic	*Coeff.*	t-statistic
Constants (*C*)				
Base	0.857	10.970	1.048	13.453
D_R	-2.116	-21.535	-2.331	-23.929
CAPITOL	-3.338	-43.057	-3.428	-44.741
SOUTHEAST	-3.900	-37.694	-4.089	-39.831
SOUTHWEST	-2.289	-25.437	-	-
WEST	-	-	-3.554	-32.328
BRANFORD	-	-	-1.059	-9.928
Slopes (*b*)				
X	-7.144	-60.316	-7.409	-62.870
X_R	2.477	14.439	2.777	16.355
X_{CAP}	6.157	51.529	6.287	53.250
X_{SE}	6.688	41.522	6.950	43.555
X_{SW}	3.842	26.909	-	-
X_W			5.911	32.882
X_{BR}			1.918	11.396
Statistics				
R-squared	0.465		0.480	
SSR	2196.423		2268.337	
SSE	2529.437		2457.523	
F-statistic	1436.049		1248.759	
Degrees of Freedom	14,884		14,882	

Shading indicates failed t-statistics at 95% confidence level.
Norwalk is base case.

Table 28. Paired t-tests between Final Regional Model Constants

	Norwalk	Capitol	Southeast	West
Capitol	44.74			
Southeast	39.83	8.41		
West	32.33	1.38	4.92	
Branford	9.93	25.71	26.55	20.63

Table 29. Paired t-tests between Final Regional Model Slope Coefficients

	Norwalk	Capitol	Southeast	West
Capitol	53.25			
Southeast	43.56	5.37		
West	32.88	2.46	5.74	
Branford	11.40	30.02	27.89	20.12

Model Application

In explaining how one of the estimated models might be applied, it was decided to use the Regional model since it is viewed to be the "best". The best model is defined not only by the amount of variation it explains, but how theoretically sound the variables are, and the application effort that would be needed for production of peak hour link volumes.

The results from the Final Regional model can be simplified by adding the like terms of each region to produce the constant *C* and slope coefficient *b* in the equation shown below:

$$\ln(P - \tfrac{1}{4}) = C + bX \qquad (36)$$

For example, the calculation of the Capitol region's constant and slope coefficient in the commute direction using the values in Table 30 is shown below:

$$C = 1.205 - 3.656 = -2.451$$
$$b = -7.639 + 6.618 = -1.021$$

The Capitol region's constant and slope coefficient in the reverse commute direction is then calculated by adding the reverse-commute constant and reverse-commute slope coefficient to the commute direction. Simplifying for each region results in the equations shown below:

$$
\begin{aligned}
&\ln(P-\tfrac{1}{4})_{NORWALK} = 1.205 - 2.895D_R - 7.639X + 3.737X_R \\
&\ln(P-\tfrac{1}{4})_{BRANFORD} = -0.486 - 2.351D_R - 4.845X + 3.103X_R \\
&\ln(P-\tfrac{1}{4})_{WEST} = -2.373 - 1.003D_R - 1.652X \\
&\ln(P-\tfrac{1}{4})_{CAPITOL} = -2.451 - 1.233D_R - 1.021X + 0.801X_R \\
&\ln(P-\tfrac{1}{4})_{SOUTHEAST} = -3.080 - 0.403X
\end{aligned}
\tag{37}
$$

where
D_R = 1, if the direction of flow is in the reverse-commute direction, otherwise 0 and,
$X_R = X$, if the direction of flow is in the reverse-commute direction, otherwise 0.

Note that because the reverse-commute constant and slope coefficient for the Southeast sites are not significant, the variables D_R and X_R are excluded from the Southeast region equation. Similarly, the reverse-commute slope coefficient for the West region is not significant and is therefore excluded from the West region equation.

Since we are interested in obtaining the actual parameters to calculate the peak factors, a transformation back to the original form of the equation is needed. This entails taking the inverse log of both sides of the equation and moving the 1/4 back to the right hand side of the equation:

$$P = \tfrac{1}{4} + ae^{bX} \tag{38}$$

where

$$a = e^{C}$$

The different model parameters for each region in the commute and reverse-commute direction of flow are shown below in Tables 31 and 32, respectively.

The next step would be to calculate the factors for application after a peak period (3-7 PM) equilibrium traffic assignment or after a factored daily assignment. Highway link-specific peak factors would be computed based on the four-hour X. For example, a link in the Capitol region (commute direction) with an X of 0.75 would have a peaking factor calculated as:

$$P = 0.25 + 0.0862e^{-1.021(0.75)} = 0.290 \tag{39}$$

This factor would then be applied to the peak period volume on that link to produce the peak hour volume estimate.

To demonstrate the application, the parameter estimates for each site were applied to an aggregate peak period volume/capacity ratio to predict the highest hourly volume within the peak period. The aggregate peak period volume/capacity ratio for each site in the commute and reverse-commute direction was computed using all observed weekday traffic volume data for each site. The predicted ratio of peak hour to peak period volume (P), as well as the predicted peak hour volumes (PHV) are shown in Table 33, along with

Table 30. Regression Results: Final Model

	Final Regional Model with Separate Directional Variables for Each Region		**Final Model with only Significant Variables**	
	Coeff.	t-statistic	*Coeff.*	t-statistic
Constants (*C*)				
Base	1.205	7.243	1.205	7.231
CAPITOL	-3.656	-21.218	-3.656	-21.183
SOUTHEAST	-4.244	-23.669	-4.285	-23.870
WEST	-3.595	-18.957	-3.578	-18.989
BRANFORD	-1.691	-7.819	-1.691	-7.806
$CAPITOL_{REV}$	-1.233	-6.975	-1.233	-6.964
$SOUTHEAST_{REV}$	-3.265	-1.064	-	-
$WEST_{REV}$	-0.776	-2.422	-1.003	-43.192
$BRANFORD_{REV}$	-2.351	-11.127	-2.351	-11.108
$NORWALK_{REV}$	-2.895	-12.271	-2.895	-12.251
Slopes (*b*)				
X	-7.639	-32.321	-7.639	-32.268
X_{CAP}	6.618	26.992	6.618	26.947
X_{SE}	7.178	27.640	7.236	27.835
X_{W}	6.015	21.366	5.987	21.452
X_{BR}	2.794	8.870	2.794	8.855
$X_{REV\text{-}CAP}$	0.801	2.508	0.801	2.504
$X_{REV\text{-}SE}$	4.489	0.780	-	-
$X_{REV\text{-}W}$	-0.410	-0.710	-	-
$X_{REV\text{-}BR}$	3.103	8.975	3.103	8.960
$X_{REV\text{-}NORW}$	3.737	9.860	3.737	9.844
Statistics				
R-squared	0.487		0.486	
SSR	2303.582		2295.078	
SSE	2422.278		2430.782	
F-statistic	744.482		877.904	
Degrees of Freedom	14,874		14,877	

Shading indicates failed t-statistics at 95% confidence level.
Norwalk is base case.

Table 31. Commute Direction Parameters

Region	*a*	*b*
Norwalk	3.3368	-7.639
Branford	0.6151	-4.845
West (Newtown, Middlebury)	0.0932	-1.652
Capitol (W. Hartford, Wethersfield, Manchester, Enfield)	0.0862	-1.021
Southeast (East Lyme, Groton)	0.0460	-0.403

Table 32. Reverse-Commute Direction Parameters

Region	*a*	*b*
Norwalk	0.1845	-3.902
Branford	0.0586	-1.742
West (Newtown, Middlebury)	0.0342	-1.652
Capitol (W. Hartford, Wethersfield, Manchester, Enfield)	0.0251	-0.220
Southeast (East Lyme, Groton)	0.0460	-0.403

the observed average weekday daily traffic (*AWDT*), peak period volume (*PPV*), peak period volume/capacity ratio (*PPV/C*), and parameters *a* and *b*. For comparison purposes the observed peak hour volume and percent difference between the predicted and observed peak hour volumes are shown.

For statewide application in general, regional boundaries would have to be established along with a method to extract the peak period v/c ratios of the freeway links. Factors are then calculated for each highway link and applied to the PM peak period volume on that link, producing peak hour volume estimates of the highest hour within the peak period. The final chapter offers conclusions from the research, along with ideas for future variable testing.

Table 33. Calculation of Peak Hour Volume from Aggregate Peak Period Volume Using Estimated Parameters

Site	AWDT	PPV	PPV/C	Parameter *a*	Parameter *b*	P	Predicted PHV	Observed PHV	Percent Difference
I-91 North, Wethersfield	56717	15944	0.45	0.0251	-0.220	0.273	4348	4546	4.3%
I-91 South, Wethersfield	55241	17610	0.67	0.0862	-1.021	0.293	5168	5256	1.7%
I-84 East, Newtown	31299	10888	0.62	0.0932	-1.652	0.283	3086	3177	2.9%
I-84 West, Newtown	31268	7963	0.45	0.0342	-1.652	0.266	2120	2174	2.5%
I-84 East, Manchester	45418	17578	0.66	0.0862	-1.021	0.294	5167	5203	0.7%
I-84 West, Manchester	45667	10662	0.40	0.0251	-0.220	0.273	2911	2838	2.6%
I-95 North, Norwalk	59934	18223	0.69	3.3368	-7.639	0.267	4868	4917	1.0%
I-95 South, Norwalk	34159	13988	0.53	0.1845	-3.902	0.273	3823	3829	0.1%
I-95 North, Branford	35382	11241	0.65	0.6151	-4.845	0.276	3107	3124	0.6%
I-95 South, Branford	34893	9163	0.53	0.0586	-1.742	0.273	2504	2521	0.7%
I-95 North, East Lyme	27520	7477	0.43	0.0460	-0.403	0.289	2158	2132	1.2%
I-95 South, East Lyme	27247	7895	0.45	0.0460	-0.403	0.288	2277	2225	2.3%
I-95 North, Groton	32181	10949	0.62	0.0460	-0.403	0.286	3130	3238	3.3%
I-95 South, Groton	30321	7475	0.28	0.0460	-0.403	0.291	2176	2065	5.4%
I-84 East, W. Hartford	54932	13657	0.52	0.0251	-0.220	0.272	3720	3778	1.5%
I-84 West, W. Hartford	56502	20101	0.76	0.0862	-1.021	0.290	5823	5781	0.7%
I-91 North, Enfield	39662	15234	0.58	0.0862	-1.021	0.298	4535	4491	1.0%
I-91 South, Enfield	40496	9859	0.38	0.0251	-0.220	0.273	2692	2685	0.3%
I-84 East, Middlebury	26755	8944	0.52	0.0932	-1.652	0.289	2589	2555	1.3%
I-84 West, Middlebury	27091	7019	0.41	0.0342	-1.652	0.267	1877	1941	3.3%

Chapter 7: CONCLUSIONS

Increasing the accuracy of peak hour forecasts is becoming ever more important, as model outputs are commonly being used for quantitative analysis in a time of limited highway budgets and increasing congestion. The research of peak spreading in Connecticut presented here shows this is an important phenomenon that needs to be addressed when predicting peak hour volumes. It is directed at enhancing existing traditional four-step models that often predict peak hour flow on links as a fixed percentage of the daily assignment.

The exponential model form chosen for estimation has several inherent characteristics that are ideal for capturing the effects of increasing congestion. For one, the rate of decrease in P gradually diminishes asymptotically as the level of congestion increases. This resembles the diminishing benefit of moving out of the peak as the four-hour volume approaches saturation.

An important factor for determining the extent of peak spreading is differentiating between direction of flow, as the constant and v/c coefficient are significantly different for the commute and reverse-commute directions. This is most likely due to the different distributions of trip purposes between the commute and reverse commute direction, as the commute direction probably consists of a higher percentage of work trips than the reverse commute direction, which consists of more discretionary trips.

The extent of peak spreading among the study sites was shown to differ substantially, which again emphasizes that the likelihood of a particular trip being made during the peak hour is not identical for all trips, but varies according to the characteristics of the trip and trip maker. Adding trip and site variables to the model was intended to increase the model's ability to capture the variation in the likelihood of making a trip in the peak hour.

Related research [11] has shown trip distance to be an important variable in predicting the amount of peak spreading that occurs with increasing congestion. In this research, trip variable testing included using internal to internal zone trip length and elapsed trip time distributions at each site. Accurate trip length distributions for thru trips and internal-external trips could not be calculated, resulting in these trips being represented as a percentage of the overall trip making at a site. In addition to not being able to create a trip length distribution for the internal-external trips, the overall percentage of these trips at each site means something different among sites. For instance, a large percentage of trips at the Enfield site are comprised of internal-external trips. A large percentage of these trips are likely work trips and may be very similar to those that are started and ended within the state. Therefore, separating the internal-external trips from the internal-internal trips at the Enfield site is not the same as doing so at the Wethersfield site where there is a much smaller percentage of these trips.

Future research could focus on this dilemma by obtaining more accurate locations for trip ends outside the state. This would allow for an overall trip length distribution to be calculated without differentiating between the type of trip. A trip purpose variable would be the primary source of capturing the differences in the type of trips at a site. This is a desirable thing to do as the trip length distribution at a site was shown to be an important factor in determining the amount of peak spreading that occurs. A site that has a higher percentage of short trip lengths tends to have smaller peak hour to peak period proportions at low levels of congestion than a site with a lower percentage of short trip lengths.

Although the elapsed trip time distributions produced a better model fit, the coefficients were inconsistent with respect to their signs and did not provide any conclusions as to their effect on travel behavior. This inconsistency may be due to the fact that these variables were created from a daily trip assignment, and may not be representative of the peak period trip length distribution. For example, a majority of the trips in the peak period are likely to be work trips; including the trip length distribution on a 24 hour basis may change that trip length distribution depending on the trip length characteristics of those trip purposes we are now including. Therefore, confidence in using these variables is not as high as it would be had a peak period trip length distribution been available.

Site variable testing showed that the number of lanes at a site and the relative distance from a CBD for a given area to a site are good indicators of how the peak period profile is distributed at low levels of congestion. A site that has more lanes will tend to have higher peak hour to peak period proportions at low levels of congestion than a site with fewer lanes. The distance from CBD variable is seen as a possible indicator of the type of trips at a site and was shown to be significant in explaining scale and shape. The LANES and CBD variables together also proved to be a good formulation with the exception of the *DISTCBD1* and *LANES* combination, which demonstrated the high correlation effect between the two. Other representations of the CBD variable were explored with great success. The CBD variable was expressed categorically allowing effects on both the intercept and slope, resulting in a much better model than when forcing the variable to be in a linear association with the dependent variable. One problem with the CBD variables is assigning links to their respective CBD. By definition of the CBD and observation of the elapsed trip times at each site in a region, several CBD associations were tested. Another problem with the CBD variable was defining the ranges of CBD distance for which categorical variables could be tested.

However, the CBD variable proved to be important in formulating the Area type and Regional Models. Area type and Regional variables were found to be the best in explaining scale and shape as they tend to consider overall differences between areas or regions rather than a single specific characteristic of the site or trip. In addition to be explaining a great deal of variation, the area type and regional variables tend to be more theoretically sound and more easily applicable.

The Final Regional Model incorporates directional and regional factors that were shown to influence the spreading of the peak. Including these variables improves the transferability of model results to other highway links in Connecticut, so the model can be applied in conjunction with the statewide travel forecasting model.

The transferability of the Area type model to other regions of the country is more easily applicable than the regional model, which is more specific to application to the Connecticut statewide forecasting model. A further study with additional data from other states would likely focus on the Area type variables, as common area types such as rural, suburban, and urban among states are more easily identified than regions. Finally, further testing is suggested using a trip purpose distribution variable, preferably from a PM peak period or factored daily assignment from PERFORM. This would also enable the trip length and elapsed trip time variables to be revisited to determine if the daily distributions used in this research are representative of the peak period.

REFERENCES

1. *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.

2. *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.

3. ConnDOT's Modeling Process and Flowchart of Modeling Steps. Technical paper provided by the Forecasting office, Connecticut Department of Transportation. Newington, Connecticut 06131, February, 1997.

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

5. Ortuzar, J. de D. and Willumsen, L. G. Modelling Transport. Chichester, England: John Wiley & Sons Ltd., 1994.

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

7. Stopher, P. R. Deficiencies of Travel-Forecasting Methods Relative to Mobile Emissions. *Journal of Transportation Engineering*, Vol. 119, No. 5, September/October, 1993, pp. 723-741.

8. Hounsell, N. B. Understanding the Effects of Congestion: Peak Spreading and Congestion. *Transportation Planning Systems*, Vol. 1, No. 3, 1991, pp. 39-46.

9. AL-Azzawi, M. An Overview of Three Techniques Designed to Aid Planners with Over-Assignment and Peak Spreading in Traffic Modelling Studies. *Traffic Engineering and Control*, Vol. 38, No. 11, 1997, pp. 604-606.

10. Allen, W. G. An Analysis of Corridor Traffic Peaking. *Transportation Research Record 1305*, TRB, National Research Council, Washington, D.C., 1991, pp. 50-60.

11. Allen, W. G., and G. W. Schultz. Congestion-Based Peak spreading Model. *Transportation Research Record 1556*, TRB, National Research Council, Washington, D.C., 1996, pp. 8-15.

12. U. Jha and J. Ivan. Predicting Peak Period Trips within the Four-Step Transportation Planning Process. JHR 97-258, Project 96-3, September 1997.

Appendix 1. Select Link File Names for Study Sites

ConnDOT ran select links for different trip purposes as well as with all purposes combined. The “???” in “MEM???” is replaced with HBW, HBO, NHB, THRU, GM, and TOT, representing the purposes of home-based work, home-based other, non-home-based, thru trips, trucks, and the total trip table which is combination of these purposes. The select links were also run by zone and town and can be identified by their extensions of “130” and “197” respectively.

Interstate	Town	Link	Direction	File	Trip Table
84	Middlebury	2556-2992	East	MEM???1	1
95	Branford	6774-6778	North	MEM???1	4
91	Enfield	7406-7482	North	MEM???1	6
84	Manchester	7201-7300	West	MEM???1	7
84	W. Hartford	5947-5952	East	MEM???1	8
91	Wethersfield	6562-6532	North	MEM???2	1
95	Groton	8668-8669	North	MEM???2	2
95	East Lyme	7852-7837	North	MEM???2	3
84	Newtown	1920-2510	East	MEM???2	5
95	Norwalk	1613-1683	North	MEM???2	6
84	Middlebury	2992-2556	West	MEM???3	1
95	Branford	6778-6774	South	MEM???3	4
91	Enfield	7482-7406	South	MEM???3	6
84	Manchester	7300-7201	East	MEM???3	7
84	W. Hartford	5952-5947	West	MEM???3	8
91	Wethersfield	6532-6562	South	MEM???4	1
95	Groton	8669-8668	South	MEM???4	2
95	East Lyme	7837-7852	South	MEM???4	3
84	Newtown	2510-1920	West	MEM???4	5
95	Norwalk	1683-1613	South	MEM???4	6

Appendix 2. Computation of Internal-Internal (I-I) Trip Table and Trip Length Frequency Distribution

The file "MEMTOT1.130" is a select link analysis trip table produced by CONNDOT, which contains the number of trips by origin-destination which transverse over the study link. There are three other trips tables beside this one that contain links used in this research. They are named MEMTOT2.130, MEMTOT3.130, and MEMTOT4.130. The output matrix "INTERNL1.130 contains only I-I trips and is used as an input file with the congested skim matrix to produce the trip length frequency distribution.

```
$MATRIX UPDATE
$FILES
    INPUT FILE = UPDIN, USER ID = $MEMTOT1.130$
    OUTPUT FILE = UPDOUT, USER ID = $INTERNL1.130$
$HEADERS
    EXTRACTION OF INTERNAL TRIPS
    REMOVAL OF EE, EI, AND IE TRIPS
$OPTIONS
    PRINT TRIP ENDS
$PARAMETERS
$DATA
    ORIGIN, T1, 1249-1300, R 0
    ORIGIN, T2, 1249-1300, R 0
    ORIGIN, T3, 1249-1300, R 0
    ORIGIN, T4, 1249-1300, R 0
    ORIGIN, T5, 1249-1300, R 0
    ORIGIN, T6, 1249-1300, R 0
    ORIGIN, T7, 1249-1300, R 0
    ORIGIN, T8, 1249-1300, R 0
    DESTINATION, T1, 1249-1300, R 0
    DESTINATION, T2, 1249-1300, R 0
    DESTINATION, T3, 1249-1300, R 0
    DESTINATION, T4, 1249-1300, R 0
    DESTINATION, T5, 1249-1300, R 0
    DESTINATION, T6, 1249-1300, R 0
    DESTINATION, T7, 1249-1300, R 0
    DESTINATION, T8, 1249-1300, R 0
$END TP FUNCTION
$REPORT TRIP LENGTH FREQUENCY
$FILE
    INPUT FILE = SKIM, USER ID = $SKIM90C.130$
    INPUT FILE = VOLUME, USER ID = $INTERNL1.130$
$HEADERS
```

```
      TRIP LENGTH FREQUENCIES FOR INTERNAL TRIPS
      MEMTOT1
$OPTIONS
      ZERO INTRAZONALS
$PARAMETERS
      IMPEDANCE = TIME 2
$END TP FUNCTION
```

Appendix 3a. External Centroids Connected to Link Nodes

Study Site	External Centroid	Connected to Node
I-84 East, Middlebury	1265	2556
I-84 West, Middlebury	1262	2992
I-95 North, Branford	1257	6774
I-95 South, Branford	1258	6778
I-91 North, Enfield	1283	7406
I-91 South, Enfield	1282	7482
I-84 West, Manchester	1284	7201
I-84 East, Manchester	1285	7300
I-84 East, W. Hartford	1280	5947
I-84 West, W. Hartford	1281	5952
I-91 North, Wethersfield	1278	6562
I-91 South, Wethersfield	1279	6532
I-95 North, Groton	1297	8668
I-95 South, Groton	1298	8669
I-95 North, East Lyme	1295	7852
I-95 South, East Lyme	1296	7837
I-84 East, Newtown	1254	1920
I-84 West, Newtown	1255	2510
I-95 North, Norwalk	1252	1613
I-95 South, Norwalk	1253	1683

Appendix 3b. Computation of Elapsed Trip Times to Study Links and Reporting of Trip Ends for use in Elapsed Trip Time Analysis

The file "UCONNPTH.96E" is the skim matrix produced with the external centroids connected to nodes at each study link. The MATRIX UPDATE function is run for each study link in each direction by zeroing out the proper origins. The output file is then reported.

```
$MATRIX UPDATE
$FILES
      INPUT FILE = UPDIN, USER ID = $UCONNPTH.96E$
      OUTPUT FILE = UPDOUT, USER ID = $LINKSKIM.96E$
$HEADERS
      ELAPSED TRIP TIME ANALYSIS
      NORWALK, I-95 SOUTH, 1683-1613
$OPTIONS
$PARAMETERS
$DATA
      DESTINATION, T3, 1-1252, R 0
      DESTINATION, T3, 1254-1300, R 0
$END TP FUNCTION
$REPORT MATRIX
$FILE
   INPUT FILE = RTABIN, USER ID = $LINKSKIM.96E$
$HEADERS
      REPORTING ROW SUMS
      NORWALK, I-95 SOUTH, 1683-1613
$OPTIONS
   PRINT TRIP ENDS
$PARAMETERS
   SELECTED PURPOSES = 3
   SELECTED IMPEDANCES = 3
   SELECTED ZONES = 1-1300
$END TP FUNCTION
```

The file "INTERNL1.130" is the file that was created for trip length distributions and contains only internal to internal trips.

```
$REPORT MATRIX
$FILE
   INPUT FILE = RTABIN, USER ID = $INTERNL1.130$
$HEADERS
   REPORTING I-I TRIPS FROM SELECT LINK FILES
$OPTIONS
```

```
   PRINT TRIP ENDS
$PARAMETERS
   SELECTED PURPOSES = 1,4,6,7,8
$END TP FUNCTION
```