

**ESTIMATING LINK TRAFFIC VOLUMES BY
MONTH, DAY OF WEEK AND TIME OF DAY**

September 2002

JHR 02-287

Project 99-3

Principal Investigators:

John N. Ivan

Wael M. ElDessouki

Graduate Assistants:

Ming Zhao

Feng Guo

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 at the Connecticut Transportation Institute of the University of Connecticut.

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

Technical Report Documentation Page

1. Report No. JHR 02-287	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A	
4. Title and Subtitle Estimating Link Traffic Volumes by Month, Day of Week and Time of Day		5. Report Date September 2002	
		6. Performing Organization Code N/A	
7. Author(s) John N. Ivan, Wael M. ElDessouki Ming Zhao, Feng Guo		8. Performing Organization Report No. 02-287	
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 Accurate estimation of hourly traffic volumes on transportation networks is vital for transportation planning, operations, and analysis. For transportation planning, hourly traffic volumes, among other factors, dictate priorities in highway improvement plans and allocation of funds. For traffic operations, hourly traffic volumes on links affect signal timing plans, air quality estimation, and traveler information systems. For safety analysis, an accurate estimation for hourly traffic volumes will help in assessing safety of different locations in the transportation networks as well as risk exposure levels. There are many factors that affect the hourly proportion distributions. In general, these factors can be divided into two groups. Factors in the first group include geometric and operational features, socio-economic characteristics, and land use patterns associated with the highway network. Any changes in these factors, over time and/or location, can affect the hourly proportion distributions. Factors in the second group include hour, day of week and month. They differ from the factors in the first group in that they are temporal in nature and their effects on the hourly proportions are distinctly cyclical. Hence, the objective of this work is to investigate whether or not hour, day of week and month have interaction effects on the hourly volume proportions at freeway count stations, and further establish procedures to group these factors into manageable categories. In addition, this research also estimates hourly proportion models using land use data upstream and downstream trafficsheds of counting stations.			
17. Key Words Hourly traffic proportions, traffic volume, congestion, AADT, land use, trafficshed, hourly proportion distribution, transportation network, network topology, network connectivity		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service Springfield, Virginia 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 111	22. Price N/A

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
<u>LENGTH</u>				
in	inches	25.4	millimetres	mm
ft	feet	0.305	metres	m
yd	yards	0.914	metres	m
mi	miles	1.61	kilometres	km
<u>AREA</u>				
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
<u>VOLUME</u>				
fl oz	fluid ounces	29.57	millilitres	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
<u>MASS</u>				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
<u>TEMPERATURE (exact)</u>				
°F	Fahrenheit temperature	5(°F-32)/9	Celsius temperature	°C

NOTE: Volumes greater than 1000 L shall be shown in m³

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
<u>LENGTH</u>				
mm	millimetres	0.039	inches	in
m	metres	3.28	feet	ft
m	metres	1.09	yards	yd
km	kilometres	0.621	miles	mi
<u>AREA</u>				
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
<u>VOLUME</u>				
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
<u>MASS</u>				
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	short tons (2000 lb)	T
<u>TEMPERATURE (exact)</u>				
°C	Celsius temperature	1.8C+32	Fahrenheit temperature	°F

°F	32	98.6	120	180	200	212
°C	-40	0	40	80	120	160
°C	-40	-20	0	20	40	60
°C						80
°C						100

* SI is the symbol for the International System of Measurement

TABLE OF CONTENTS

TECHNICAL REPORT DOCUMENTATION PAGE	ii
MODERN METRIC CONVERSION FACTORS	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	vi
LIST OF TABLES.....	viii
1 INTRODUCTION	1
1.1 Problem Statement	1
1.2 Objectives and Scope of Report.....	2
2 LITERATURE REVIEW	4
2.1 Peak-Period Volume Models	4
2.2 Daily Volume Models.....	4
2.3 Discussion.....	5
3 ESTIMATING HOURLY PROPORTION MODELS	7
3.1 Introduction.....	7
3.1.1 Problem Statement	7
3.1.2 Objectives	7
3.1.3 Study Data.....	8
3.1.4 Preliminary Investigation.....	12
3.2 Methodology.....	13
3.2.1 Analysis of Variance.....	13
3.2.2 Null Hypothesis and Model	15
3.2.3 Test Statistics	15
3.2.4 Data Transformation	16
3.2.5 Model Parameter Estimation.....	16
3.3 Results.....	17
3.3.1 Factor Effect Significance.....	17
3.3.2 Hourly Proportion Models	22
3.4 Discussion.....	33
3.5 Conclusions.....	33
4 GROUPING MONTHS AND WEEKDAYS	35
4.1 Introduction.....	35
4.1.1 Problem Statement	35
4.1.2 Objective and Scope	35
4.2 Methodology.....	36
4.2.1 Overview.....	36

4.2.2	Comparison of Means	36
4.2.3	Group Pattern Generation	40
4.2.4	Grouping Selection	43
4.3	Results and Discussion	49
4.4	Conclusions.....	50
5	GROUPING HOURS AND ESTIMATING HOURLY PROPORTIONS	53
5.1	Introduction.....	53
5.2	Methodology.....	53
5.2.1	Hour Groups.....	53
5.2.2	Month and Weekday Grouping Selection.....	54
5.2.3	Hourly Proportion Models	55
5.2.4	Hourly Proportion Calculation.....	56
5.2.5	Accuracy Checking.....	56
5.3	Results.....	58
5.4	Discussion.....	59
5.5	Conclusions.....	73
6	A TRAFFIC SHED APPROACH FOR LOCATION-SPECIFIC ESTIMATION OF HOURLY TRAFFIC VOLUMES	74
6.1	Introduction and Background	74
6.2	Model Hypothesis	75
6.2.1	Demographic and Land Use Patterns.....	75
6.2.2	Network Topology and Structure.....	76
6.3	Methodology and Mathematical Model.....	77
6.3.1	Determination of Hourly Proportions Matrix	80
6.3.2	Weights Estimation Model	84
6.4	Results and Discussion	85
6.4.1	Estimation of Trip Purpose Weights.....	85
6.4.2	Estimation of Weight Functions	86
6.5	Summary and Conclusion	90
7	SUMMARY AND FURTHER STUDIES.....	91
7.1	Methodology and Results	91
7.2	Application and Future Work	93
	REFERENCES	95
	APPENDIX A: MODEL ADEQUACY CHECKING	99

LIST OF FIGURES

Figure 3-1 Distribution of ATR Stations (ADD).....	11
Figure 3-2 Hourly Proportions by Hour for Study Year.....	13
Figure 3-3 Hourly Proportions by Day of Week for Study Stations (A.M. Peak Hour).....	14
Figure 3-4 Hourly Proportions by Month for Study Stations (A.M. Peak Hour).....	14
Figure 3-5 Normal Probability Plot of Residuals (Full Model for Station 9027-3)	18
Figure 3-6 Residual vs. Fitted Value Plot (Full Model for Station 9027-3)	18
Figure 3-7 Time Series Plot of Residuals (Full Model for Station 9027-3)	19
Figure 3-8 Estimated Hourly Proportions by Hour and Month for Sunday for Station 9027-3	26
Figure 3-9 Estimated Hourly Proportions by Hour and Month for Monday for Station 9027-3 ...	27
Figure 3-10 Estimated Hourly Proportions by Hour and Month for Tuesday for Station 9027-3 .	28
Figure 3-11 Estimated Hourly Proportions by Hour and Month for Wednesday for Station 9027-3	29
Figure 3-12 Estimated Hourly Proportions by Hour and Month for Thursday for Station 9027-3	30
Figure 3-13 Estimated Hourly Proportions by Hour and Month for Friday for Station 9027-3	31
Figure 3-14 Estimated Hourly Proportions by Hour and Month for Saturday for Station 9027-3.	32
Figure 4-1 Convert	41
Figure 4-2 Matrix for Month Example	42
Figure 4-3 Group Pattern Generation Example	42
Figure 4-4 Month Group Pattern Example	43
Figure 4-5 Another Month Group Pattern Example	44
Figure 5-1 Hourly Proportions (after grouping) by Hour and Month for Sunday at Station 9027-3	65
Figure 5-2 Hourly Proportions (after grouping) by Hour and Month for Monday at Station 9027-3	66
Figure 5-3 Hourly Proportions (after grouping) by Hour and Month for Tuesday Saturday at Station 9027-3	67
Figure 5-4 Hourly Proportions (after grouping) by Hour and Month for Wednesday at Station 9027-3.....	68
Figure 5-5 Hourly Proportions (after grouping) by Hour and Month for Thursday at Station 9027-3	69
Figure 5-6 Hourly Proportions (after grouping) by Hour and Month for Friday at Station 9027-3	70
Figure 5-7 Hourly Proportions (after grouping) by Hour and Month for Saturday at Station 9027-3	71
Figure 6-1 Example for Observed Hourly Traffic Volumes.....	76
Figure 6-2 Conceptual Example for Estimating Hourly Volume Proportions.....	79
Figure 6-3 Trafficshed Determination Method.....	81
Figure 6-4 Cumulative Distribution for Home-Based Trip Duration in Connecticut.....	82
Figure 6-5 a & b Departure Profiles for home/work trips (NPTS 1996)	83
Figure 6-6 Departure profiles for other trip purposes (NPTS 1996)	83
Figure 6-7 Distributions for Estimated Weights by Trip Purpose.....	86
Figure 6-8 Estimated Weights for Work to Home Trips.....	87
Figure 6-9 Estimated Weights for Home to Work Trips.....	87
Figure 6-10 Estimated Weights for Other Trips.....	88
Figure 6-11 Error Distributions for Prediction Using the Estimated Weights.....	89

Figure A 0-1 Normal Probability Plot of Residuals (Hourly Model I for Station 9027-3)	100
Figure A 0-2 Time Series Plot of Residuals (Hourly Model I for Station 9027-3)	100
Figure A 0-3 Residual vs. Fitted Value Plot (Hourly Model I for 9027-3)	101
Figure A 0-4 Normal Probability Plot of Residuals (Hourly Model II for Station 9027-3)	102
Figure A 0-5 Time Series Plot of Residuals (Hourly Model II for Station 9027-3)	102
Figure A 0-6 Residual vs. Fitted Value Plot (Hourly Model II for 9027-3)	103

LIST OF TABLES

Table 3-1 ATR Station List	9
Table 3-2 Data Sample	12
Table 3-3 Number of Hours with Significant Effects (Weekdays)	21
Table 3-4a Model Parameters by Hour for Station 9027-3 (Intercept and Weekday)	23
Table 3-5 Estimated Hourly Proportions for Selected Hours for Station 9027-3	25
Table 4-1 Tukey's Comparisons and Engineering Criterion Incorporation	39
Table 4-2 Samples of Possible Adjacent Month Groupings	46
Table 4-3 Possible Adjacent Weekday Groupings	47
Table 4-4 Fitted Month Grouping Samples	48
Table 4-5 Grouping Selection	49
Table 4-6 Month Groupings by Hour for Station 9027-3	51
Table 4-7 Weekday Groupings by Hour for Station 9027-3	52
Table 5-1 Month Groupings for All Hours in the Day for Station 9027-3	54
Table 5-2 Weekday Groupings for All Hours in the Day for Station 9027-3	54
Table 5-3 Model Parameters by Hour Group for Weekdays for Station 9027-3	56
Table 5-4 Estimated Hourly Proportions by Hour Group for Station 9027-3	57
Table 5-5 RMSE and MAPE by Hour Group for Station 9027-3	58
Table 5-6 Month and Weekday Groupings for All Hour Groups by Station	60
Table 5-7 Month and Weekday Groupings for Morning Peak Hours by Station	61
Table 5-8 Month and Weekday Groupings for Afternoon Peak Hours by Station	62
Table 5-9 Month Groupings by Grouping Index	63
Table 5-10 Final Grouping Evaluation	72

1 INTRODUCTION

1.1 PROBLEM STATEMENT

Traffic volumes by time of day for highway links are important as input for models in air quality estimation, vehicle crash prediction, and transportation planning. For example, models for estimating mobile source emissions require traffic volumes by time of day to estimate important input quantities, such as vehicle-miles-traveled and speed by hour of the day (FHWA 1994). In addition, research into investigating the effect of actual traffic volumes on crash frequencies and rates reveals a distinct relationship between observed crashes and traffic volumes (Gwynn 1967; Zhou and Sisiopiku 1997; and Ivan *et al.* 1999). However, traffic volumes by time of day are normally not available for the vast majority of highway links because traffic volumes are not routinely measured in such detail due to the intractable cost of instrumenting the links with continuous counting stations. A solution to the problem is to estimate hourly (or time of day) traffic volumes for any location of interest.

The state-of-the-art procedure used to estimate hourly volumes for a highway network is to allocate daily volumes for highway links among the hours of interest using hourly proportions (or time-of-day factors). The daily volumes, normally annual average daily traffic (AADT) or annual average weekday traffic (AAWT), are generated from travel demand models (such as the ubiquitous four-step process) or traffic monitoring programs maintained in most states (Robertson *et al.* 1994). These daily volumes are the only consistent and readily available source for the statewide road network. The critical link in this procedure is estimating the hourly proportions adequately.

There are many factors that affect the hourly proportion distributions. In general, these factors can be divided into two groups. Factors in the first group include geometric and operational features, socio-economic characteristics, and land use patterns associated with the highway network. Any changes in these factors, over time and/or location, can affect the hourly proportion distributions. Factors in the second group include hour, day of week and month. They differ from the factors in the first group in that they are temporal in nature and their effects on the hourly proportions are distinctly cyclical. All these factors should be taken into account in an hourly proportion model.

However, previous research efforts in this area were all devoted to the first group of factors. Factors in the second group are either not considered or considered only approximately. Consequently, only annual average hourly proportions can be estimated using these models, which in turn result in annual average hourly volume estimates. In many cases, the accuracy of the estimated hourly volumes cannot satisfy the rapidly increasing demands of more detailed and accurate vehicle emissions, accident prediction, and transportation planning models.

This problem can be solved by incorporating the factors in the second group (*i.e.*, hour, day of week and month) into the hourly proportion models. This is actually a very challenging task, because if these factors are considered in complete detail the size of an

hourly proportion model (or in other words, the number of model parameters) can become so large that it is beyond the capacities of existing analytical software and computers. For example, an hourly proportion model with hour, day of week and month as independent variables may involve more than 2,016 ($24 \times 7 \times 12$) parameters, if all interactions among them are considered. If you consider a few other factors in the first group, such as highway capacity, truck percentage, urban/rural designation (i.e., rural, suburb, small urban, urbanized areas), or distance to central business district (CBD), the number of model parameters becomes overwhelmingly large, which make the model very difficult to estimate and infeasible to use in practice.

The model size can be significantly reduced if hour, day of week and month do not interact with the hourly proportions. In fact, omitting the interactions reduces the number of model parameters from 2,016 to 43 ($24 + 7 + 12$). Consequently, an issue here is to identify whether or not these factors significantly interact with each other.

In addition, if predictive covariates are included in the prediction process, the 43 remaining parameters may be still too many to use from a practical point of view, because the number of model parameters can still go over several hundred, if the covariates interact with the temporal factors. It is therefore desirable to further reduce the model size by grouping hours, days of the week and months. For example, if these factors can be grouped into five time periods in a day, weekday and weekend, and four seasons in a year, the number of parameters for models involving only these factors reduces to 11 ($5 + 2 + 4$). This reduction permits site related factors to be included without unduly increasing model complexity. This calls for investigation into how to group these factors, if possible.

Another way to improve the accuracy of the hourly proportions may be estimating prediction models specific to a location that can therefore consider hour, day of week and month only. The model size is still a problem if these factors interact with one another; if they do not, reasonably sized models for specific locations can be estimated to provide quite accurate and precise hourly volume estimates. These models would, of course, only be directly useful for the locations for which they were estimated, requiring further investigation into how to transfer the resulting parameter values to other highway locations where the hourly volumes need to be estimated. Nevertheless, this could be useful and appropriate alternative approach to solve the hourly volume estimation problem.

1.2 OBJECTIVES AND SCOPE OF REPORT

Consequently, the objective of this work is to investigate whether or not hour, day of week and month have interaction effects on the hourly volume proportions at freeway count stations, and further establish procedures to group these factors into manageable categories. In addition, this research also estimates hourly proportion models considering these factors.

The Analysis of Variance (ANOVA) statistical procedure is used to test if the interactions between hour, day of week and month are significant. Hourly proportion models are then

estimated based on the ANOVA results. Procedures are established to group the factors. This includes the use of Comparison of Means statistical procedure (along with an engineering criterion) and a grouping algorithm developed *ad hoc*. We also include a preliminary investigation into estimation of prediction models that account for location-oriented variables.

This document is divided into seven chapters. The first chapter (*i.e.*, this introduction) introduces the research problem, the objective and the organization of the report. Chapter 2 gives an overall literature review. Chapter 3 documents the ANOVA results and the estimated hourly proportion models. Chapters 4 and 5 deal with the issue of grouping the factors; specifically, Chapter 4 discusses the Comparison of Means statistical procedures (as well as an engineering criterion) and reports the procedure established to group months and weekdays at each hour, and Chapter 5 reports the grouping of hours and the final groupings involving month, weekday and hour. Chapter 6 introduces a procedure for estimating the hourly proportions as a function of the distribution of population and employment in the vicinity of the highway location. Finally, Chapter 7 provides a summary of the research findings and discusses some further studies that may be done to improve the accuracy of the hourly proportion estimates, and how the results and findings of this work may help in these further studies.

2 LITERATURE REVIEW

Previous studies in estimating traffic volumes by time of day for highway network shared a common base: the 24-hr travel demand models. Generally, they were all concerned with how to estimate hourly proportion (or time-of-day factors) models, where the hourly proportions represent the percentage of peak period (three or four hours) trips or daily trips made during the hours of interest. In other words, these models are developed using either peak period or daily volumes as the base; thus they can be categorized into peak-period volume models and daily volume models. The following sections give a review of these models, respectively.

2.1 PEAK-PERIOD VOLUME MODELS

One approach in estimating peak hour volumes based on peak period volumes was developed by Loudon (1988) for the Arizona Department of Transportation. This approach focused on modeling the peak spreading in a three-hour morning peak period. It assumed that travel during this peak period consisted of a fixed percentage of daily trips, but allowed the peak hour volume (as a percentage of the three-hour volume) to vary according to congestion levels measured by the ratio of volume to capacity (v/c). A link-specific peak spreading model that represented the effect of peak period congestion to the temporal distribution of travel during that period was estimated using data from 45 corridors in Arizona, Texas and California. This model was incorporated into the network equilibrium traffic assignment process, and the results were link-level peak hour and peak-period traffic volumes.

Allen and Schultz (1996) established another approach for the Washington, D.C., region in estimating peak hour traffic volumes based on peak period volumes. A peak spreading model was developed as a post-mode choice procedure, considering congestion, trip purpose, and trip distance as independent variables. Similar to the peak spreading model developed by Loudon, this study also assumed that travel during a three-hour peak period consisted of a fixed percentage of daily trips, but allowed the peak hour volume (as a percentage of the three-hour volume) to vary with the level of congestion and trip length. The final result of the study was a series of stratified curves of peak one-hour proportions by trip purposes. An origin-destination survey consisting of more than 45,000 trip records was used in the development of the stratified curves. The peak-hour traffic volumes were estimated by first determining the proportions of the peak-hour travel occurring in the peak three-hour period, and then applying the proportions to the estimated peak-period traffic volumes.

2.2 DAILY VOLUME MODELS

Several studies have estimated the proportions of daily volumes occurring during the peak hour(s). Daly *et al.* (1990) estimated models for predicting the proportion of trips falling within a two-hour peak period in the Netherlands. They estimated the peak two-hour proportion by modeling traveler's choice of time of day to travel, considering the congestion level and a time-of-day-dependent road pricing policy. The input data were

from a survey including stated preference questions of the trade-off between changes in travel time and congestion delays. They also proposed a procedure to incorporate the model into the existing travel demand forecast models in the Netherlands.

In an approach established by Crevo and Virkud (1994) for the Delaware Department of Transportation, hourly proportions for a two-hour evening peak period were separately developed for each system movement (internal-internal, external-internal, and external-external) and each trip purpose (work, shop, school, other, non-home-based and truck). The hourly proportions were either estimated using the Nationwide Personal Transportation Survey (NPTS) data or the permanent traffic count data, or borrowed from a Federal Highway Administration (FHWA) report. They were then applied to the 24-hour trip tables created by the trip distribution process to estimate two-hour peak period traffic volumes for each of the movements and purposes.

In addition, Gunawardena *et al.* (1996) estimated morning and afternoon peak hour factors for the Indiana Department of Transportation. Their study investigated the effects of location, year, month, season, and day of week on the peak hour factors using the Analysis of Variance (ANOVA) statistical method. The main effects and some selected interaction effects of these factors are tested with the traffic count data collected at ATR stations in the state of Indiana. The final results were sets of peak-hour volume (and direction) factors recommended based on the ANOVA results.

2.3 DISCUSSION

These studies represent the state of the art and significant advancements toward the estimation of network hourly volumes. However, the effects of hour, day of week and month, which directly result in cyclical patterns of traffic, are either not considered or considered in a highly approximate way. Kumar and Levinson (1995) pointed out that people's activity patterns (which directly result in variations of traffic volumes) vary significantly across the natural and cultural cycles reflected in the calendar and the clock. The effects of these factors deserve a thorough examination, and they should be considered in estimating hourly proportion models.

Gunawardena *et al.* (1996) attempted to address this issue in their study. However, the study still falls short in two aspects. First, the study in essence pooled together data at all study locations in the analysis of the effects of hour, day of week and month. As a result, the conclusions of the factor effects drawn from the analysis are only valid for an area-wide average situation. As pointed out in a previous study (Deakin, Harvey, Skabardonis, Inc. 1993), there is generally little reason to expect specific facilities to exhibit the same peaking patterns or characteristics as "region averages," and application of an area-wide average time-of-day factor may be a significant source of error.

Second, the study focused on only the peak one-hour in the morning and afternoon. Sometimes, it is important to estimate hourly volumes for other hours in the day. Also pointed out by Deakin, Harvey, Skabardonis, Inc. (1993), highway networks in many metropolitan areas experience congestion for 3-6 hours a day, and air quality issues require traffic volumes for hours other than the peak hour(s). For example, CO

concentrations are typically higher in the afternoon and evening hours, and an area with a CO problem needs to estimate traffic volumes of these hours. In addition, accident analysis may also need traffic volumes for hours other than the peak hours. For example, an accident analysis investigating the effect of night vision may require traffic volumes at night.

3 ESTIMATING HOURLY PROPORTION MODELS

3.1 INTRODUCTION

3.1.1 Problem Statement

Traffic volumes by time of day for highway links are needed as input for models in air quality estimation, vehicle crash prediction, and transportation planning. However, they are often not available and need to be estimated. The state of the art procedure in estimating traffic volume by time of day is to allocate the daily volume for a highway link among the hours of interest using hourly proportions. The key to the success of this procedure is getting good estimates of the hourly proportions.

Many factors should be taken into account in estimating the hourly proportions. In general, these factors include: 1) geometric and operational features, socio-economic characteristics, and land use patterns associated with the highway network, and 2) temporal factors, including hour, day of week and month of observations. The temporal factors in the second group contrast from those in the first group further in that they have distinct cyclic effects. However, previous research in this area is almost all devoted to the first group of factors (e.g., Loudon 1988, Dale *et al* 1990, Crevo and Virkud 1994, and Allen and Schultz 1996). Our literature search revealed only one research work that attempted to study the effects of the factors in the second group on the hourly proportions (Gunawardena *et al.* 1996). This work investigated the effects of location, year, month, season, and day of week on peak hour factors using the Analysis of Variance (ANOVA) statistical method. The main effects and some selected interaction effects of these factors were tested, with the traffic count data collected at the Automatic Traffic Recorder (ATR) stations in the state of Indiana as input.

However, this work still left two issues to be addressed. First, it in essence pooled data at all study locations together in the analysis of the effects of hour, day of week and month. As a result, the conclusions of the factor effects drawn from the analysis were only valid for an area-wide average situation. As pointed out by Deakin, Harvey, Skabardonis, Inc. (1993), there is generally little reason to expect specific facilities to exhibit the same peaking patterns or characteristics as "region averages," and application of an area-wide average time-of-day factor may be a significant source of error. Second, the study focused on only the peak one-hour in the morning and afternoon. In many cases, hourly volumes for other hours in the day are also very important and need to be estimated.

3.1.2 Objectives

This chapter describes our research into how to obtain more accurate and reliable hourly proportions. This is carried out by first thoroughly investigating the effects of hour, day of week and month on the hourly proportions at key highway locations. Hourly proportion models with these factors as independent variables are then estimated at each of the locations based on the investigation results. The Analysis of Variance (ANOVA)

statistical procedures are used; with traffic counts collected at the ATR stations on Connecticut freeways as input.

It is noted that an hourly proportion model should include the temporal factors and the other site related variables (*i.e.*, those in the first group as discussed earlier), because they all contribute to the hourly proportion distribution. Specifically, the effects of the temporal factors result in the cyclical variations in the hourly proportions, while the effects of the other variables result in the variations by station and year. However, if they are all considered, the model becomes so complex that it is infeasible to estimate.

One way to solve this problem is to reduce the number of the temporal factor categories, or in other words, to group these factors. For example, if hour, day of week and month can be grouped into five time periods in a day, weekday and weekend, and four seasons in a year, the number of parameters for an hourly proportion model involving only these factors may be reduced to 11 ($5 + 2 + 4$). This reduction permits site related factors to be included without unduly increasing model complexity. In order to group the factors appropriately, we need to investigate the effects of hour, day of week and month on the hourly proportions.

Alternatively, this problem can be solved by estimating hourly proportion models specific to a location that can therefore consider hour, day of week and month only. The model size is still a problem if these factors interact with each other; if they do not, reasonably sized models for specific locations can be estimated to provide quite accurate and precise hourly volume estimates. These models would, of course, only be directly useful for the locations for which they were estimated, requiring further investigation of how to transfer their applications to other highway locations where the hourly volumes need to be estimated. Nevertheless, this could be a useful and appropriate alternative approach to solve the hourly volume estimation problem.

Consequently, this chapter focuses on investigating the effects of the temporal factors and estimating hourly proportion models considering these factors. The findings obtained from the investigation are used to determine proper forms of the models and further to help in grouping the factors. Procedures established to group the factors are discussed in the next two chapters. The groupings of the factors are intended to be used in further work on this project that will be directed to estimate hourly proportion models considering both the temporal factors (after the grouping) and other site related variables.

3.1.3 Study Data

The data used here are generated from continuous hourly traffic counts recorded from 1991 through 1996 by the Connecticut Department of Transportation (ConnDOT) at the 15 Automatic Traffic Recorder (ATR) stations on Connecticut freeways. Table 3.1 gives a list of the ATR stations, including the station numbers assigned by ConnDOT, the town, the route, the location and the station IDs of each station. Note that there are two station IDs for each ATR station in the table. This is because the station ID is assigned separately for each direction of traffic, which in this case is denoted by 1, 3, 5 or 7 (1-

Table 3-1 ATR Station List

Station Number	Town	Route	Location	Station ID
7	Norwich	I-395	0.4 Mile North of Exit 80 on I 395	9007-1, 9007-5
12	Killingly	I-395	0.2 Mile South of Exit 93 on I-395	9012-1, 9012-5
14	Wethersfield	I-91	0.2 Mile North of Rocky Hill Town Line	9014-1, 9014-5
24	Newtown	I-84	0.4 Mile East of Brookfield Town Line	9024-3, 9024-7
26	Manchester	I-84	0.8 Mile West of Exit 63 on I-84	9026-3, 9026-7
27	Union	I-84	0.2 Mile East of Exit 74 on I-84	9027-3, 9027-7
30	Norwalk	I-95	1.25 Mile North of Darien Town Line	9030-1, 9030-5
32	Branford	I-95	0.8 Mile South of Exit 55 on I-95	9032-1, 9032-5
33	East Lyme	I-95	0.1 Mile South of Exit 73 on I-95	9033-1, 9033-5
44	Groton	I-95	0.6 Mile South of Exit 89 on I-95	9044-1, 9044-5
45	Cheshire	I-691	0.5 Mile West of Exit 3 on I-691	9045-3, 9045-7
49	West Hartford	I-84	0.1 Mile West of Exit 44 on I-84	9049-3, 9049-7
53	Enfield	I-91	0.5 Mile North of East Windsor	9053-1, 9053-5
54	Middlebury	I-84	0.5 Mile East of Southbury Town Line	9054-3, 9054-7
55	Wallingford	I-91	0.7 Mile South of Exit 5 on I-91	9055-1, 9055-5

North, 3-East, 5-South, and 7-West). In addition, Figure 3.1 provides a map that shows the geographic distribution of these stations throughout the state.

Using the hourly traffic counts, at each station the observed hourly proportions are calculated as

$$P_{ijkm} = \frac{V_{ijkm}}{ADT_{ij}} \quad (3.1)$$

where P_{ijkm} and V_{ijkm} are the hourly proportions and hourly traffic counts, respectively, for the m^{th} day in month i , day of week j , hour k . ADT_{ij} is the average daily traffic for month i and day of week j , and computed as

$$ADT_{ij} = \frac{\sum_{m=1}^{N_{ij}} \sum_{k=1}^{24} V_{ijkm}}{N_{ij}} \quad (3.2)$$

where N_{ij} is the number of days that fall in month i and day of week j . These observed hourly proportions are calculated separately for each year.

Table 3.2 shows a small sample of the analysis data. Column 1 in the table gives the station ID, which indicates where the traffic counts shown in Column 7 are collected. Columns 2 through 6 give the year, month, day of the month, day of the week and hour corresponding to each traffic count. Here the day of the week is coded by 1 to 7 (1-Sunday, 2-Monday, ..., 7-Saturday).

Column 8 in the table shows the holiday, "special day" and outlier indicators (0-normal day, 1-holiday, 2-special day, and 3-outlier). The holiday and "special day" indicators are assigned to the traffic counts by the Connecticut Department of Transportation. Holidays include official Federal and State holidays. Special days account for special events such as accidents, construction, or inclement weather conditions, and days immediately before or after major holidays. The outliers refer to traffic counts that deviate significantly from the normal patterns, *e.g.*, a very low value during a peak traffic period. Traffic counts for holidays and special days and traffic counts considered to be outliers are excluded from the study, because the hourly proportions are likely to be tainted by the unusual trip-making on these days and thus should not be mixed in with other normal days.

Columns 9 and 10 in Table 3.2 show the hourly proportions calculated based on the traffic counts and their Logit transformation (which will be discussed later in this chapter). Note that in these two columns, missing observations and data excluded from the analysis are indicated by dots (.). In fact, the entire data set contains about 12 percent missing observations, due to counting equipment malfunctions or the counting station being out of service (due to road construction or maintenance, for example). These hourly proportions (in fact, their Logit transformations) are used to study the effects of hour, day of week and month and estimate the hourly proportion models.

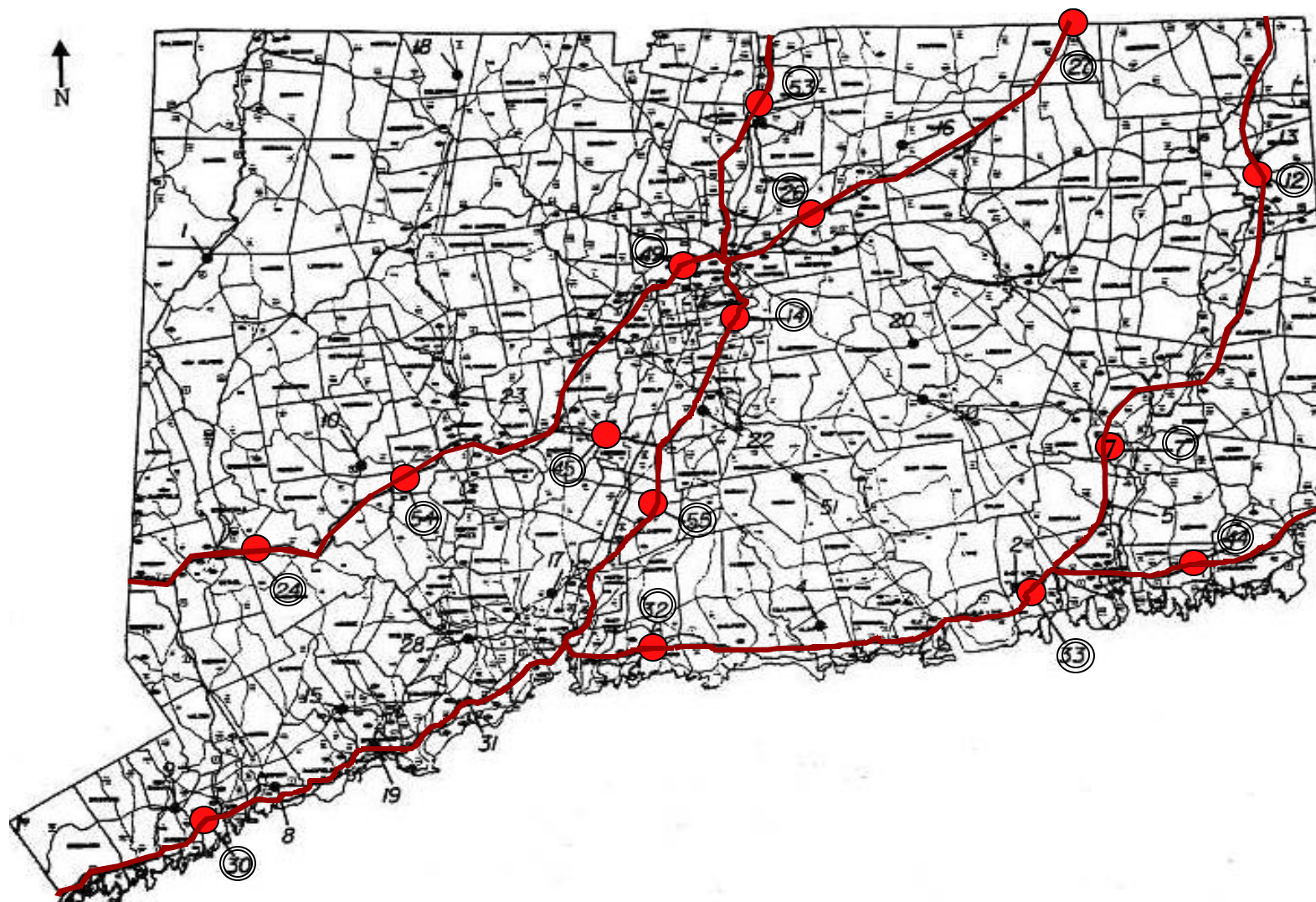


Figure 3-1 Distribution of ATR Stations (ADD)

Table 3-2 Data Sample

Station ID	Year	Month	Day	Day of Week	Hour	Hourly Count	Special Day	Hourly Prop.	Logit Trans.
9030-1	96	8	31	7	12:00	4776	0	.0688	-2.6042
9030-1	96	8	31	7	13:00	4790	0	.0690	-2.6011
9030-1	96	8	31	7	14:00	4343	0	.0626	-2.7059
9030-1	96	8	31	7	15:00	4132	0	.0595	-2.7590
9030-1	96	8	31	7	16:00	3857	0	.0556	-2.8320
9030-1	96	8	31	7	17:00	3719	0	.0536	-2.8706
9030-1	96	8	31	7	18:00	3484	0	.0502	-2.9394
9030-1	96	8	31	7	19:00	3002	0	.0432	-3.0956
9030-1	96	8	31	7	20:00	2821	0	.0406	-3.1605
9030-1	96	8	31	7	21:00	2531	0	.0364	-3.2734
9030-1	96	8	31	7	22:00	2315	0	.0333	-3.3658
9030-1	96	8	31	7	23:00	2018	0	.0290	-3.5075
9030-1	96	9	1	1	0:00	1535	2	.	.
9030-1	96	9	1	1	1:00	1036	2	.	.
9030-1	96	9	1	1	2:00	818	2	.	.
9030-1	96	9	1	1	3:00	504	2	.	.

3.1.4 Preliminary Investigation

The effects of hour, day of week and month should normally be studied separately at each year to eliminate the variations by year in the study data. At each year, however, only a very limited number of observations are available for analyzing the effects, especially the interaction effects. Alternatively, the data for several years can be pooled together, which would provide adequate observations to do the analysis. This is only valid when the variations by year are not significant; otherwise, the use of the resulting models may cause significant errors, because the models represent only the average situation. To determine whether or not the data for the study years can be pooled together, we examined the variations in hourly proportions by year.

Figure 3.2 shows the hourly proportions by hour for each study year, where the hourly proportions are calculated by averaging the observed hourly proportions over all weekdays, months and stations at each hour. This figure indicates that the variations by year are not quite so great. For example, the hourly proportions range from 0.075 to 0.077 for the a.m. peak hour (starting at 07:00). This would result in estimated hourly volumes ranging from 2,550 to 2,618 (a difference of only 68 vph) for a daily volume of 34,000 vehicles, although these locations obviously do not have the same daily volumes as one another. Therefore, the data for all the study years (five years total) are pooled together at each station in this study.

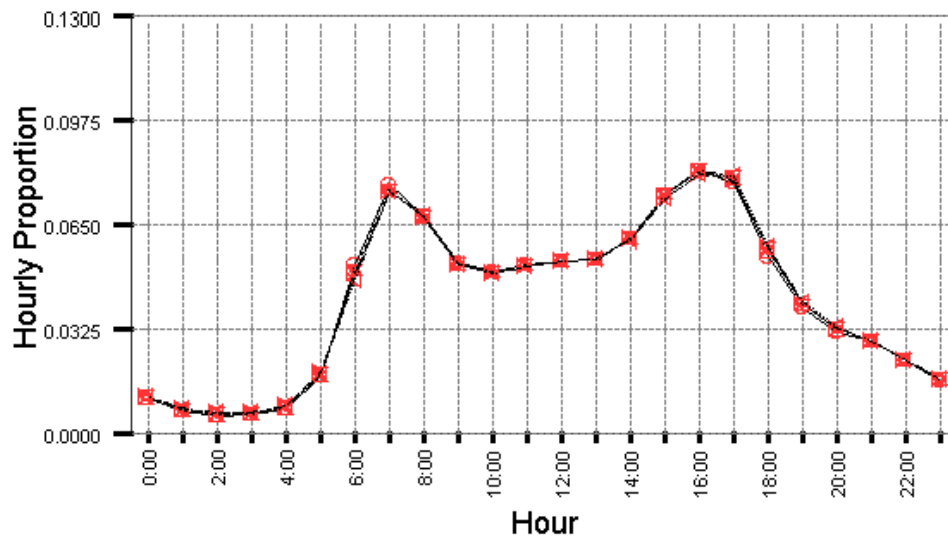


Figure 3-2 Hourly Proportions by Hour for Study Year

In addition, the variations in hourly proportions by the study factors (*i.e.*, hour, day of week and month) are also investigated in this preliminary study. A general understanding of the variations in hourly proportion by these factors is needed later in our more thorough study of the effects of these factors. In fact, Figure 3.2 also shows the variations in the hourly proportions by hour. As can be seen, the hourly proportions vary significantly through the day, with two predominant a.m. and p.m. peaks and relatively low values at other time periods. Figures 3.3 and 3.4 show the hourly proportions by day of week and month, respectively, for all study stations at the a.m. peak hour (starting at 07:00). Figure 3.3 indicates that the hourly proportions are also obviously significantly different by weekday, Friday, Saturday and Sunday, but the variations within weekdays are relatively small. Figure 3.4 shows that the variations by month are also relatively small (as opposed to the variations by hour and by weekday and weekend).

3.2 METHODOLOGY

3.2.1 Analysis of Variance

The ANOVA statistical procedure is used here to test whether hour, day of week and month have significant effects, especially interaction effects, to the hourly proportions. As mentioned earlier, an hourly proportion model with hour, day of week and month as independent variables may involve more than 2,016 ($24 \times 7 \times 12$) parameters, considering that they are all categorical with many levels and potentially interact with each other. A model with so many parameters is very difficult to estimate and of not much use in practice. The model size can be significantly reduced, if it is proved that the interactions

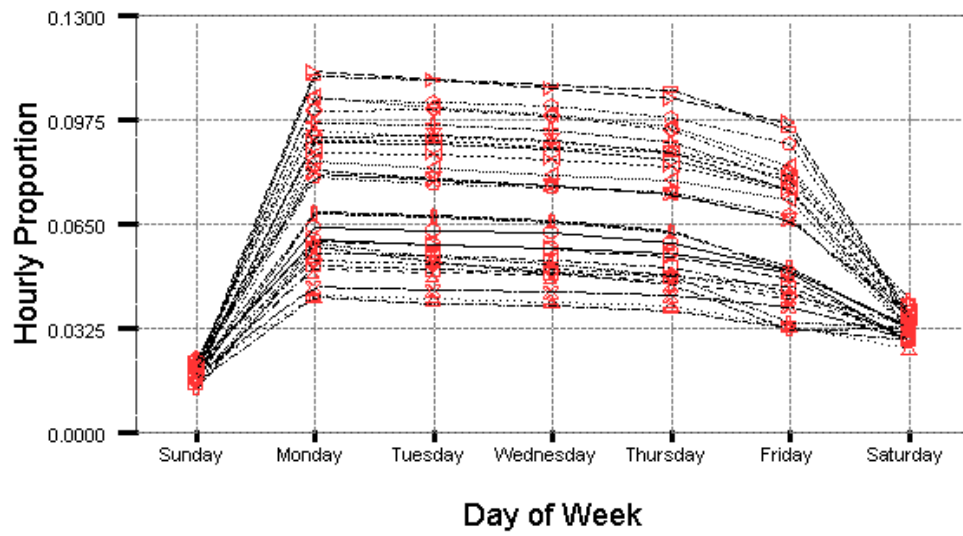


Figure 3-3 Hourly Proportions by Day of Week for Study Stations (A.M. Peak Hour)

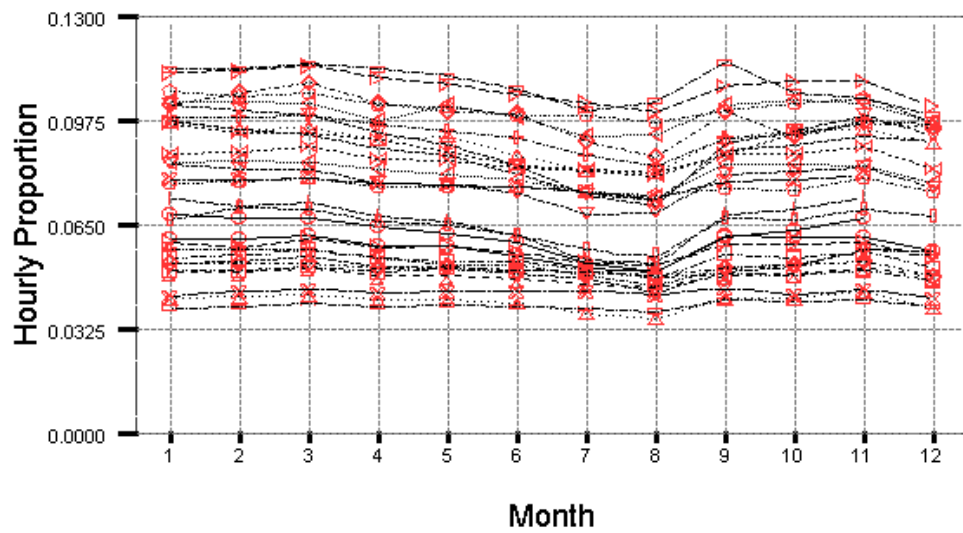


Figure 3-4 Hourly Proportions by Month for Study Stations (A.M. Peak Hour)

of the factors are not significant and can be omitted in the model. In fact, the number of model parameters becomes about 43 ($24 + 7 + 12$), if the interactions of the factors are not considered. The ANOVA procedure is designed to test null hypotheses of the effects of categorical factors (in this case, hour, day of week and month) to the response (in this case, the hourly proportions). The factor effects are defined to be the changes in the response produced by the changes in the levels of the factor(s). These changes are usually called main effects because they refer to the primary factor effects of interest. When two or more factors are involved, the effects also include the changes in the response produced by the changes in the levels of one factor at each level of the other factor(s). These changes are normally referred to as interaction effects because they reflect the interaction between the factors (Montgomery 1991).

3.2.2 Null Hypothesis and Model

Here the null hypotheses corresponding to the interaction effects of month i , day of week j and hour k , can be expressed as

$$\begin{aligned} H_0^1 : MD_{ij} &= 0 \quad i = (1, \dots, 12), j = (1, \dots, 7) \\ H_0^2 : MH_{ik} &= 0 \quad i = (1, \dots, 12), k = (1, \dots, 24) \\ H_0^3 : DH_{jk} &= 0 \quad j = (1, \dots, 7), k = (1, \dots, 24) \\ H_0^4 : MDH_{ijk} &= 0 \quad i = (1, \dots, 12), j = (1, \dots, 7); k = (1, \dots, 24) \end{aligned} \quad (3.3)$$

These null hypotheses can be tested using an ANOVA model:

$$P_{ijkm} = \mu + M_i + D_j + H_k + MD_{ij} + MH_{ik} + DH_{jk} + MDH_{ijk} + e_{m(ijk)} \quad (3.4)$$

where P_{ijkm} denotes the observed hourly proportions for the m^{th} day in month i , weekday j , and hour k ; μ is the unknown grand mean of the data estimated by the procedure; and $e_{m(ijk)}$ is a random error component, which is assumed to be normally and independently distributed with zero mean and constant but unknown variance. In addition to the interaction effects, this model can also be used to test the null hypotheses of the main effects of these factors - M_i , D_j and H_k .

3.2.3 Test Statistics

The test statistic for the null hypotheses is the ratio

$$F_0 = \frac{SS_{Treatment} / a}{SS_{Error} / b} \quad (3.5)$$

where $SS_{Treatment}$, SS_{Error} and a , b are, respectively, the sums of squares and degrees of freedom of the treatment effects and error. The appropriate reference distribution for F_0

is the F distribution with the treatment degrees of freedom, a , as the numerator degrees of freedom, and the error degrees of freedom, b , as the denominator degrees of freedom. The null hypothesis would be rejected at level of significance α if F_0 is greater than $F_{\alpha,a,b}$, where $F_{\alpha,a,b}$ denotes the upper one-tail critical region of the F distribution with a and b degrees of freedom and α is the pre-specified level of significance.

3.2.4 Data Transformation

One of the assumptions of the ANOVA procedure is that the error component (and the response) is normally distributed; however, the study data (*i.e.*, observed hourly proportions) by nature lie between 0 and 1, which is a direct violation of the assumption. This problem is solved here by a Logit transformation

$$Q_{ijkm} = \ln \left(\frac{P_{ijkm}}{1 - P_{ijkm}} \right) \quad (3.6)$$

where P_{ijkm} and Q_{ijkm} are the observed hourly proportion and its Logit transformation, respectively. This transformation brings data between (0, 1) into a real scale, *i.e.*, between $(-\infty, +\infty)$. Using these Logit transformed data, the ANOVA model in Equation 3.4 becomes

$$Q_{ijkm} = \mu + M_i + D_j + H_k + MD_{ij} + MH_{ik} + DH_{jk} + MDH_{ijk} + e_{m(ijk)} \quad (3.7)$$

Note that even though the Logit transformed data are used as the samples of the response here, the conclusions drawn from the test results also apply to the hourly proportions. This is because it is expected that the errors (if any) in the conclusions caused by using the transformation would be very limited. The range of the hourly proportions is actually very small at a particular hour (0.05 on average); as a result, the relationship between the hourly proportion and its transformation becomes almost linear. In fact, the ANOVA test results would be the same using either set of the data if a linear relationship exists between them.

3.2.5 Model Parameter Estimation

The parameters of the ANOVA model can be estimated using least squares regression (Montgomery 1991). Using the estimated model parameters, we can calculate the fitted values as

$$\hat{Q}_{ijk} = \hat{\mu} + \hat{M}_i + \hat{D}_j + \hat{H}_k + (\hat{MD})_{ij} + (\hat{MH})_{ik} + (\hat{DH})_{jk} + (\hat{MDH})_{ijk} \quad (3.8)$$

where $\hat{\mu}$ is the estimated intercept; $\hat{M}_i$, $\hat{D}_j$ and $\hat{H}_k$ are the estimated month, day of week and hour parameters, respectively; and $(\hat{MD})_{ij}$, $(\hat{MH})_{ik}$, $(\hat{DH})_{jk}$ and $(\hat{MDH})_{ijk}$ are the estimated interaction effect parameters.

Using the fitted values, model residuals (or errors) can be calculated as

$$\hat{e}_{ijkm} = Q_{ijkm} - \hat{Q}_{ijk} \quad (3.9)$$

These errors should be checked for model adequacy. In general, ANOVA results and models cannot be considered totally valid, if the model assumptions (*i.e.*, normally and independently distributed errors with zero mean and constant but unknown variance) are significantly violated. These assumptions can be checked using the normal probability, residual vs. fitted value and time series plots of the model residuals. If the model assumptions are valid, the normal probability plot should resemble a straight line, the residual vs. fitted value plot should be randomly scattered around zero, and time series plots of residuals should not show any tendency of runs of positive and negative values.

If the model is adequate, an estimated hourly proportion for month i , day of week j and hour k , $\hat{P}_{ijk}$ can be calculated using the reverse of the Logit transformation as

$$\hat{P}_{ijk} = \frac{e^{\hat{Q}_{ijk}}}{e^{\hat{Q}_{ijk}} + 1} \quad (3.10)$$

Note that there are a total of about 2,600 parameters to be estimated in this model, which makes it very difficult to use in practice. In fact, the model represents a full factorial design of the study factors. It may not be necessary to include all of the interaction terms in the model. By dispensing with the terms that are not significant, the model can be simplified. This is the primary reason for performing ANOVA here, *i.e.*, to test whether or not the interaction terms in the model are significant.

3.3 RESULTS

3.3.1 Factor Effect Significance

The ANOVA tables needed in testing the null hypotheses of the factor effects are produced using SAS statistical software (SAS Institute Inc. 1990). Here the factor effects are tested in three different levels using three sets of models. The null hypotheses, corresponding ANOVA models and test results are discussed individually in the following sections.

3.3.1.1 Full Model

The full factorial model in Equation 3.7 is used first to test whether or not the interaction effects of hour, day of week and month are significant, *i.e.*, whether or not the null hypotheses in Equation 3.3 can be rejected. The test results show that the interaction effects (as well as the main effects) are significant at over 95 percent confidence level.

To check whether or not this conclusion is valid, the model residuals are examined here. Figures 3.5 through 3.7 show the normal probability, residual vs. fitted value and time series plots, respectively, for the full model given in Equation 3.7 for ATR station 9027-3. As can be seen, the normal probability plot in Figure 3.5 roughly resembles a straight

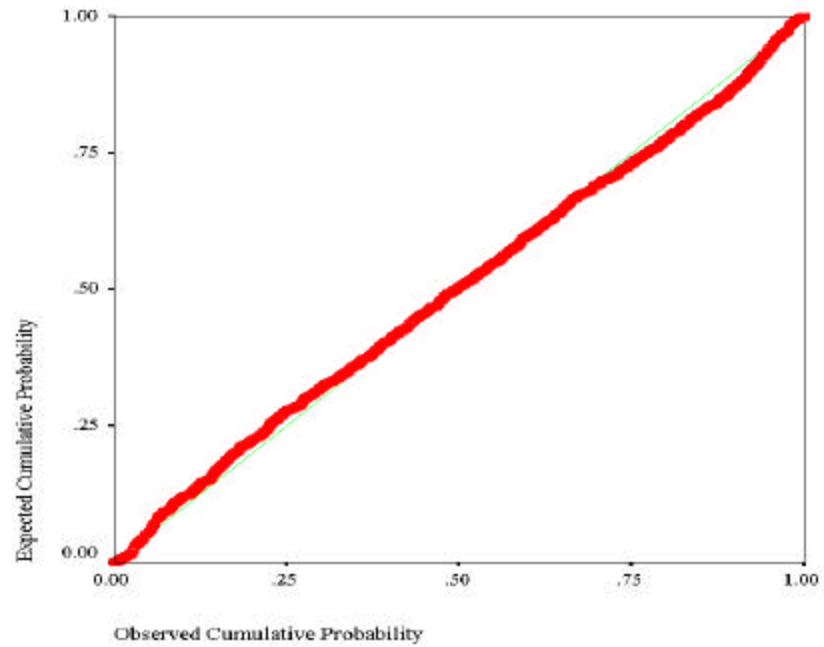


Figure 3-5 Normal Probability Plot of Residuals (Full Model for Station 9027-3)

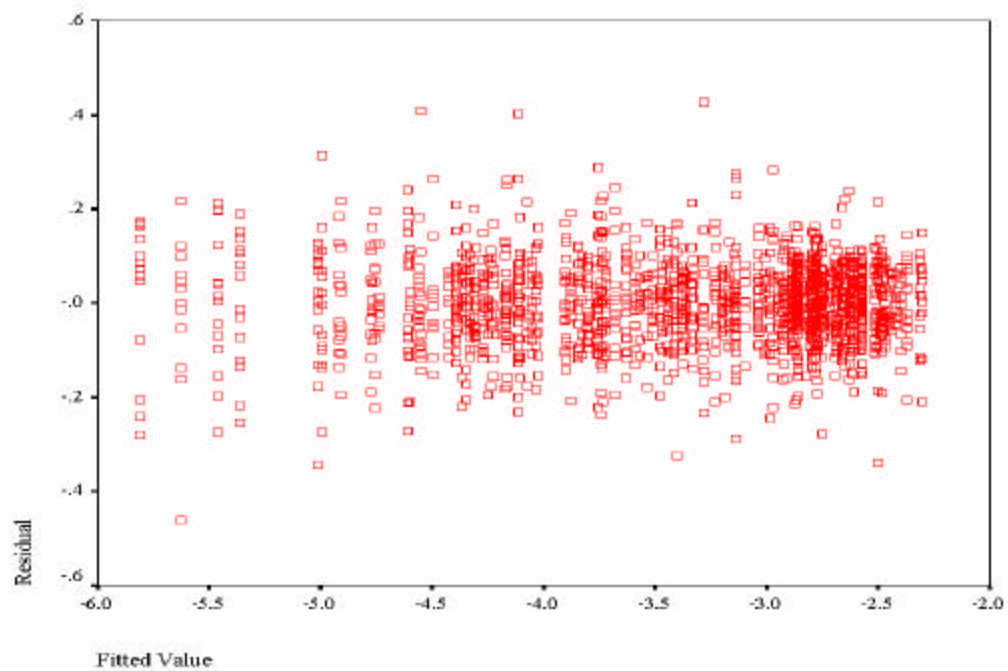


Figure 3-6 Residual vs. Fitted Value Plot (Full Model for Station 9027-3)

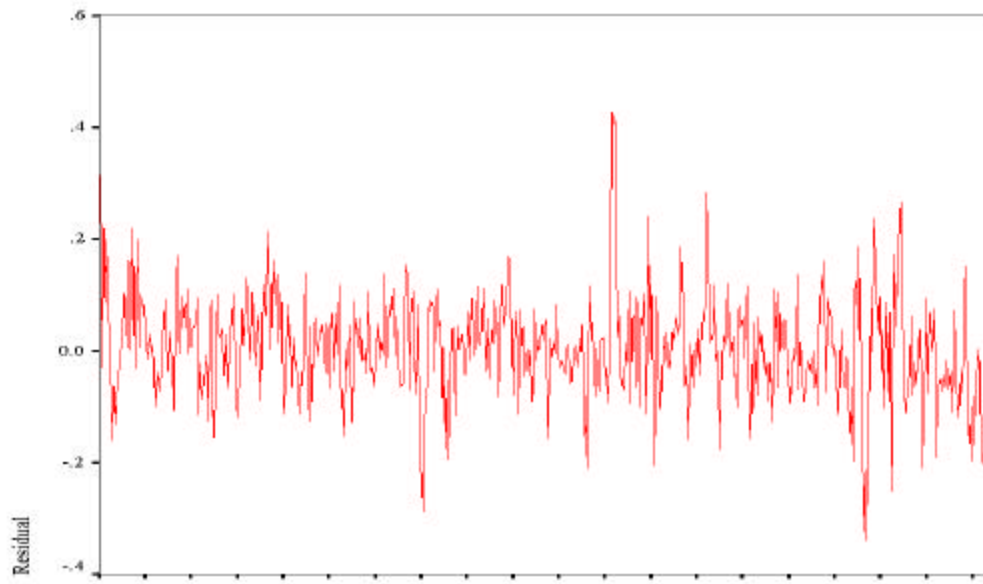


Figure 3-7 Time Series Plot of Residuals (Full Model for Station 9027-3)

line, and the residual vs. fitted value plot in Figure 3.6 does not show obvious patterns. This means that the normality and equal variance assumptions are basically valid in this case. However, the time series plot of the residuals given in Figure 3.7 shows some tendency of runs of positive and negative values, which implies that the residuals are not exactly independent. (Note that the residuals and fitted values shown in this figure are those for October in 1995. The plots for other time periods generally show similar patterns). As a result, the ANOVA conclusions drawn using the full model are somewhat questionable. This conclusion is generally applied to other ATR stations too.

Nonetheless, the test results indicate that, if the full factorial model is used to estimate hourly proportions, all terms in the model should be included, even though it may not be adequate. As discussed before, this model has over two thousand parameters, and thus it is not practical to use such a model. Consequently, we need to identify a way to reduce the model size.

3.3.1.2 Hourly Model I

Our first attempt to reduce the model size is to estimate hourly proportion models and factor effects for each hour separately. The hour is the primary factor affecting the hourly proportion distributions, as shown in the preliminary study. By confining analysis to a single hour at a time, we can focus on investigating the effects of month and day of week. In addition, it is noted that the underlying correlation and the distinct difference of traffic among hours in the day is the major contributor to the violation of assumptions in the full model. This simplification also removes this inherent correlation and non-normal nature of the data.

At each hour k , the null hypotheses in Equation 3.3 now become

$$H_0 : MD_{ij} = 0 \quad i = (1, \dots, 12); j = (1, \dots, 7) \quad (3.11)$$

Now our primary concern is to test whether or not the interaction effect of month and day of week are significant. The corresponding ANOVA model is

$$Q_{ijm} = \mu + M_i + D_j + MD_{ij} + e_{m(ij)} \quad (3.12)$$

This model can also be used to test the significance of the main effects of month and day of week.

The test results show that at each hour the interaction effect (as well as the main effects) is again significant at over 95 percent confidence. To check whether or not this conclusion is valid, plots similar to those shown in Figure 3.5 through 3.7 are produced and examined (see Appendix A for details). The results show only a slight violation of the model assumptions; thus this conclusion is generally valid, meaning that the interaction term should again be included in the model. As a result, the model still has over eighty parameters, which is again not desirable to be used in practice.

3.3.1.3 Hourly Model II

Our second attempt is to further confine the ANOVA and model estimation to weekday, Friday, Saturday and Sunday, individually, considering that the hourly proportions also vary significantly by weekday, Friday, Saturday and Sunday as shown in the preliminary study. This allows us to test whether or not the interaction effect of month and weekday is significant on weekdays. The null hypothesis is

$$H_0 : MD_{ij} = 0 \quad i = (1, \dots, 12); j = (2, \dots, 5) \quad (3.13)$$

The corresponding ANOVA model has the same form as in Equation 3.11. The only difference is that now day of week j is limited to weekdays (2 - Monday, 3 - Tuesday, 4 - Wednesday and 5 - Thursday).

The ANOVA test results are summarized in Table 3.3, where the numbers of hours with significant interaction and main effects of month and weekday at 95 percent confidence level are given for each study station. To further identify whether these numbers are different for peak and off-peak periods, the numbers are separately given for three time periods: a.m. peak (6:00 a.m. to 10:00 a.m.), p.m. peak (3:00 p.m. to 7:00 p.m.), and off-peak (all other hours). This results reveal that the interaction effect is significant at only a small number of hours, while the main effects of weekdays and months are significant at nearly all hours. This holds true for each of the three time periods.

Our examination of the model residuals shows that the model assumptions are generally satisfied (again, see Appendix A for details). This means that the conclusions drawn above are valid, or in other words, we can conclude that in general the interaction

Table 3-3 Number of Hours with Significant Effects (Weekdays)

Station ID	A.M. Peak (6 a.m. - 10 a.m.)			P.M. Peak (3 p.m. - 7 p.m.)			Off-Peak (other hours)		
	D _j	M _i	MD _{ij}	D _j	M _i	MD _{ij}	D _j	M _i	MD _{ij}
9007-1	4	4	0	4	4	0	15	16	2
9007-5	4	4	0	4	4	0	14	15	2
9012-1	4	4	0	4	4	0	15	16	0
9012-5	4	4	0	3	4	0	16	14	3
9014-1	4	4	0	2	4	0	15	16	5
9014-5	4	4	0	3	4	1	15	16	5
9024-3	4	4	0	3	4	0	15	16	1
9024-7	4	4	0	4	4	0	16	16	2
9026-3	4	4	0	3	4	1	15	16	4
9026-7	4	4	1	4	4	0	15	15	3
9027-3	4	3	0	4	3	0	15	15	3
9027-7	4	4	1	4	4	0	16	16	2
9030-1	4	4	0	3	4	0	14	15	4
9030-5	4	4	0	4	4	0	16	16	4
9032-1	4	4	0	4	4	0	16	16	6
9032-5	4	4	0	3	4	0	15	16	2
9033-1	4	4	1	4	4	0	15	16	6
9033-5	4	4	0	4	4	0	13	16	2
9044-1	4	4	1	4	4	0	16	16	7
9044-5	4	4	2	4	4	0	13	16	3
9045-3	4	4	0	2	4	1	14	16	1
9045-7	4	4	0	3	4	0	12	16	2
9049-3	4	4	0	2	4	0	14	16	4
9049-7	4	4	0	4	4	0	15	16	4
9053-1	4	4	0	4	4	0	14	16	6
9053-5	3	4	0	4	4	0	16	16	3
9054-3	4	4	1	4	4	0	15	16	1
9054-7	4	4	1	4	3	0	16	16	2
9055-1	4	4	0	4	4	0	16	16	4
9055-5	4	4	1	2	4	0	15	16	2
Average	4	4	0	4	4	0	15	16	3

effects of month and weekday are not significant, but separately, month and weekday are significant factors that affect the hourly proportions. Consequently, the interaction term of month and weekday can be omitted from the model, but the main effect terms should be kept in the model.

3.3.2 Hourly Proportion Models

These ANOVA conclusions lead to our final hourly proportion model for an hour at a study station. For weekdays, the final model is

$$Q_{ijm} = \mu + M_i + D_j + e_{m(ij)} \quad (3.14)$$

and for Friday, Saturday and Sunday, individually, the final model is

$$Q_{im} = \mu + M_i + e_{m(i)} \quad (3.15)$$

Here for each hour and ATR station, model parameters are produced separately for weekday and weekend again using SAS statistical software. For example, Tables 3.4a and 3.4b (see pages 23 and 24) give the model parameters for weekdays for each hour in the day at ATR station 9027-3, including the intercept $\hat{\mu}$, the weekday parameters $\hat{D}_j$, and the month parameters, $\hat{M}_i$.

These parameters can be used to estimate the hourly proportion for a specific hour, month and weekday for the study location. For example, for the morning peak hour (starting at 7:00 a.m.) on Monday in January, an estimate of the Logit transformation of the hourly proportion can be calculated as

$$\hat{Q}_{ij} = -2.93 + 0.234 - 0.0773 = -2.7733 \quad (3.16)$$

where -2.93 , 0.234 and -0.0773 are respectively the intercept, weekday parameter and month parameters as highlighted in Table 3.4 (a and b).

The corresponding hourly proportion can then be estimated using the reverse of the Logit transformation as

$$\hat{P}_{ij} = \frac{e^{-2.7733}}{e^{-2.7733} + 1} = 0.0588 \quad (3.17)$$

More examples of the estimated hourly proportions are given in Table 3.5 (page 25), which contains the hourly proportions by month and day of week for four selected hours in the day. These hourly proportions are again estimated based on the model parameters given in Tables 3.4a and 3.4b (pages 23 and 24). Further, Figures 3.8 through 3.14 (pages 26 through 32) show the estimated hourly proportions by hour and month for each day of the week (*i.e.*, Sunday, Monday, ... Saturday, respectively). Note that these hourly proportions are those for ATR Station

Table 3-4a Model Parameters by Hour for Station 9027-3 (Intercept and Weekday)

<i>Hour</i>	<i>Intercept</i>	<i>Monday</i>	<i>Tuesday</i>	<i>Wednesday</i>	<i>Thursday</i>
0:00	-4.3570	0.4490	0.0928	0.0421	0.0000
1:00	-4.3920	0.3220	0.0899	0.0502	0.0000
2:00	-4.4480	0.1770	0.0887	0.0764	0.0000
3:00	-4.3290	0.0971	0.0832	0.0673	0.0000
4:00	-4.1500	0.0497	0.0649	0.0446	0.0000
5:00	-3.8320	0.1240	0.0822	0.0430	0.0000
6:00	-3.3580	0.2470	0.1160	0.0718	0.0000
7:00	-2.9300	0.2340	0.1490	0.0796	0.0000
8:00	-2.8730	0.0951	0.1120	0.0684	0.0000
9:00	-2.9300	0.0979	0.0963	0.0689	0.0000
10:00	-2.9460	0.1050	0.0597	0.0362	0.0000
11:00	-2.8790	0.0890	0.0318	0.0057	0.0000
12:00	-2.8710	0.0792	0.0221	0.0021	0.0000
13:00	-2.7920	0.0171	-0.0039	-0.0130	0.0000
14:00	-2.6590	-0.0354	-0.0180	-0.0097	0.0000
15:00	-2.5540	-0.0895	-0.0320	-0.0019	0.0000
16:00	-2.4910	-0.1300	-0.0304	0.0134	0.0000
17:00	-2.5400	-0.1460	-0.0511	0.0053	0.0000
18:00	-2.7610	-0.1890	-0.1050	-0.0558	0.0000
19:00	-3.0340	-0.2370	-0.1600	-0.1210	0.0000
20:00	-3.2000	-0.2190	-0.1500	-0.1310	0.0000
21:00	-3.3500	-0.1860	-0.1340	-0.1090	0.0000
22:00	-3.4790	-0.1920	-0.1130	-0.0941	0.0000
23:00	-3.7970	-0.1840	-0.0938	-0.0758	0.0000

9027-3 only. Hourly proportions for other stations are estimated similarly; for brevity, they are not given here.

These hourly proportions can be used to estimate hourly volumes. For example, the hourly proportion given in Equation 3.17 would result in an estimated hourly volume of 1,140 ($19,400 \times 0.0588$) for the a.m. peak hour on Mondays in January at the study location, given that the average daily volume on Mondays in January at the location is 19,400. In addition, assuming that a highway link with a daily volume of 34,000 on Mondays in January is of interest, and that the link has a traffic pattern similar to this ATR station. This hourly proportion can be used to estimate an hourly volume for the a.m. peak hour for that location; specifically, the estimated hourly volume would be about 2,000 ($34,000 \times 0.0588$).

TABLE 3.4b: Model Parameters by Hour for Station 9027-3 (Month)

<i>Hour</i>	<i>January</i>	<i>February</i>	<i>April</i>	<i>March</i>	<i>May</i>	<i>June</i>	<i>July</i>	<i>August</i>	<i>September</i>	<i>October</i>	<i>November</i>	<i>December</i>
0:00	0.1240	0.1000	0.0857	0.1070	0.1150	0.1020	0.0990	0.0905	0.1490	0.1340	0.0648	0.0000
1:00	0.0073	-0.0055	-0.0467	-0.0469	-0.0664	-0.0675	-0.0712	-0.1190	-0.0236	-0.0003	-0.0284	0.0000
2:00	0.0181	-0.0269	-0.0295	-0.0711	-0.0751	-0.1240	-0.1200	-0.1480	-0.0327	-0.0363	0.0897	0.0000
3:00	-0.0376	-0.0082	-0.0303	-0.1050	-0.1290	-0.1660	-0.1890	-0.2570	-0.0551	-0.0174	0.0246	0.0000
4:00	0.0018	-0.0203	-0.0103	-0.0649	-0.0671	-0.1380	-0.1670	-0.2130	-0.0224	-0.0015	0.1030	0.0000
5:00	-0.0480	-0.0289	-0.0510	-0.1060	-0.0554	-0.1130	-0.1620	-0.2580	-0.0583	-0.0233	0.0684	0.0000
6:00	-0.0486	-0.0211	-0.0314	-0.0829	-0.0559	-0.0861	-0.1410	-0.2660	-0.0833	-0.0316	0.0081	0.0000
7:00	-0.0773	-0.0525	-0.0382	-0.0973	-0.1000	-0.1420	-0.1760	-0.2620	-0.1250	-0.0864	0.0485	0.0000
8:00	-0.0503	-0.0384	-0.0109	-0.0439	-0.0640	-0.0912	-0.0744	-0.1290	-0.0742	-0.0265	-0.0008	0.0000
9:00	0.0107	0.0142	0.0387	0.0597	0.0439	0.0456	0.0479	0.0488	0.0464	0.0456	0.0296	0.0000
10:00	-0.0024	-0.0120	-0.0178	0.0379	0.0373	0.0713	0.1300	0.1800	0.0510	0.0298	0.0156	0.0000
11:00	-0.0112	-0.0171	-0.0256	0.0131	0.0045	0.0504	0.1160	0.1600	0.0607	0.0081	0.0023	0.0000
12:00	-0.0092	-0.0194	-0.0041	0.0083	0.0089	0.0390	0.0763	0.0940	0.0269	-0.0097	-0.0309	0.0000
13:00	0.0114	-0.0168	-0.0020	0.0170	0.0271	0.0388	0.0572	0.0874	0.0359	0.0080	-0.0314	0.0000
14:00	-0.0072	-0.0178	-0.0102	0.0071	0.0053	0.0178	0.0256	0.0535	0.0010	-0.0177	-0.0329	0.0000
15:00	0.0077	-0.0142	-0.0050	0.0100	0.0091	0.0030	-0.0068	0.0051	-0.0202	-0.0185	-0.0213	0.0000
16:00	0.0010	0.0078	0.0156	-0.0120	-0.0205	-0.0246	-0.0629	-0.0526	0.0066	-0.0087	-0.0413	0.0000
17:00	-0.0264	0.0039	-0.0018	-0.0329	-0.0327	-0.0590	-0.0920	-0.0795	-0.0463	-0.0380	-0.0575	0.0000
18:00	0.0282	0.0192	0.0107	0.0423	0.0534	0.0231	0.0239	0.0043	-0.0153	0.0015	0.0028	0.0000
19:00	-0.0169	0.0065	-0.0021	0.0124	0.0461	0.0449	0.0593	0.0817	0.0233	0.0119	-0.0028	0.0000
20:00	-0.0422	-0.0118	-0.0420	-0.0287	0.0026	0.0120	0.0230	0.0149	-0.0287	-0.0350	-0.0231	0.0000
21:00	-0.0220	-0.0069	-0.0453	-0.0337	-0.0098	0.0142	0.0095	0.0125	-0.0353	-0.0331	-0.0419	0.0000
22:00	-0.0592	-0.0557	-0.0603	-0.0675	-0.0560	-0.0244	-0.0407	-0.0442	-0.0464	-0.0345	-0.1190	0.0000
23:00	0.0038	0.0018	-0.0413	-0.0471	-0.0529	-0.0116	-0.0080	-0.0257	-0.0299	0.0089	-0.0601	0.0000

Table 3-5 Estimated Hourly Proportions for Selected Hours for Station 9027-3

Hour	Day of Week	Month											
		Jan.	Feb.	April	March	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
7:00	Sun.	0.0101	0.0116	0.0118	0.0120	0.0136	0.0141	0.0166	0.0152	0.0146	0.0128	0.0146	0.0103
	Mon.	0.0587	0.0601	0.0609	0.0577	0.0575	0.0552	0.0535	0.0493	0.0561	0.0582	0.0661	0.0632
	Tue.	0.0542	0.0555	0.0563	0.0532	0.0531	0.0510	0.0494	0.0455	0.0518	0.0538	0.0611	0.0583
	Wed.	0.0508	0.0520	0.0527	0.0498	0.0497	0.0477	0.0462	0.0426	0.0485	0.0503	0.0572	0.0546
	Thr.	0.0471	0.0482	0.0489	0.0462	0.0461	0.0442	0.0428	0.0395	0.0450	0.0467	0.0531	0.0507
	Fri.	0.0345	0.0336	0.0337	0.0339	0.0313	0.0316	0.0298	0.0279	0.0305	0.0301	0.0373	0.0307
	Sat.	0.0253	0.0249	0.0295	0.0305	0.0337	0.0343	0.0403	0.0364	0.0328	0.0322	0.0358	0.0308
12:00	Sun.	0.0698	0.0683	0.0605	0.0649	0.0599	0.0607	0.0639	0.0674	0.0620	0.0654	0.0611	0.0765
	Mon.	0.0573	0.0567	0.0575	0.0582	0.0583	0.0599	0.0621	0.0631	0.0592	0.0572	0.0561	0.0578
	Tue.	0.0543	0.0537	0.0545	0.0552	0.0552	0.0568	0.0588	0.0598	0.0561	0.0542	0.0532	0.0547
	Wed.	0.0532	0.0527	0.0535	0.0541	0.0542	0.0557	0.0577	0.0587	0.0551	0.0532	0.0522	0.0537
	Thr.	0.0531	0.0526	0.0534	0.0540	0.0541	0.0556	0.0576	0.0586	0.0550	0.0531	0.0521	0.0536
	Fri.	0.0498	0.0526	0.0526	0.0521	0.0526	0.0540	0.0565	0.0562	0.0539	0.0520	0.0474	0.0598
	Sat.	0.0829	0.0843	0.0789	0.0744	0.0768	0.0767	0.0768	0.0792	0.0814	0.0745	0.0747	0.0755
17:00	Sun.	0.0852	0.0864	0.0919	0.0868	0.0872	0.0819	0.0777	0.0787	0.0804	0.0853	0.0837	0.0826
	Mon.	0.0622	0.0640	0.0637	0.0618	0.0618	0.0603	0.0585	0.0592	0.0611	0.0615	0.0604	0.0638
	Tue.	0.0680	0.0700	0.0696	0.0676	0.0676	0.0660	0.0640	0.0647	0.0668	0.0673	0.0661	0.0697
	Wed.	0.0717	0.0737	0.0733	0.0712	0.0713	0.0695	0.0674	0.0682	0.0704	0.0709	0.0696	0.0735
	Thr.	0.0713	0.0734	0.0730	0.0709	0.0709	0.0692	0.0671	0.0679	0.0700	0.0706	0.0693	0.0731
	Fri.	0.0845	0.0831	0.0831	0.0805	0.0777	0.0779	0.0689	0.0702	0.0761	0.0804	0.0693	0.0811
	Sat.	0.0597	0.0584	0.0596	0.0603	0.0552	0.0540	0.0492	0.0513	0.0513	0.0583	0.0553	0.0609
21:00	Sun.	0.0445	0.0447	0.0491	0.0495	0.0536	0.0553	0.0529	0.0503	0.0497	0.0459	0.0453	0.0423
	Mon.	0.0277	0.0281	0.0271	0.0274	0.0280	0.0287	0.0286	0.0286	0.0273	0.0274	0.0272	0.0283
	Tue.	0.0291	0.0296	0.0285	0.0288	0.0295	0.0302	0.0300	0.0301	0.0288	0.0288	0.0286	0.0298
	Wed.	0.0298	0.0303	0.0292	0.0295	0.0302	0.0309	0.0308	0.0309	0.0295	0.0295	0.0293	0.0305
	Thr.	0.0332	0.0337	0.0324	0.0328	0.0336	0.0344	0.0342	0.0343	0.0328	0.0328	0.0325	0.0339
	Fri.	0.0427	0.0403	0.0411	0.0420	0.0430	0.0424	0.0454	0.0456	0.0440	0.0484	0.0419	0.0364
	Sat.	0.0251	0.0250	0.0266	0.0274	0.0262	0.0285	0.0254	0.0251	0.0276	0.0260	0.0268	0.0287

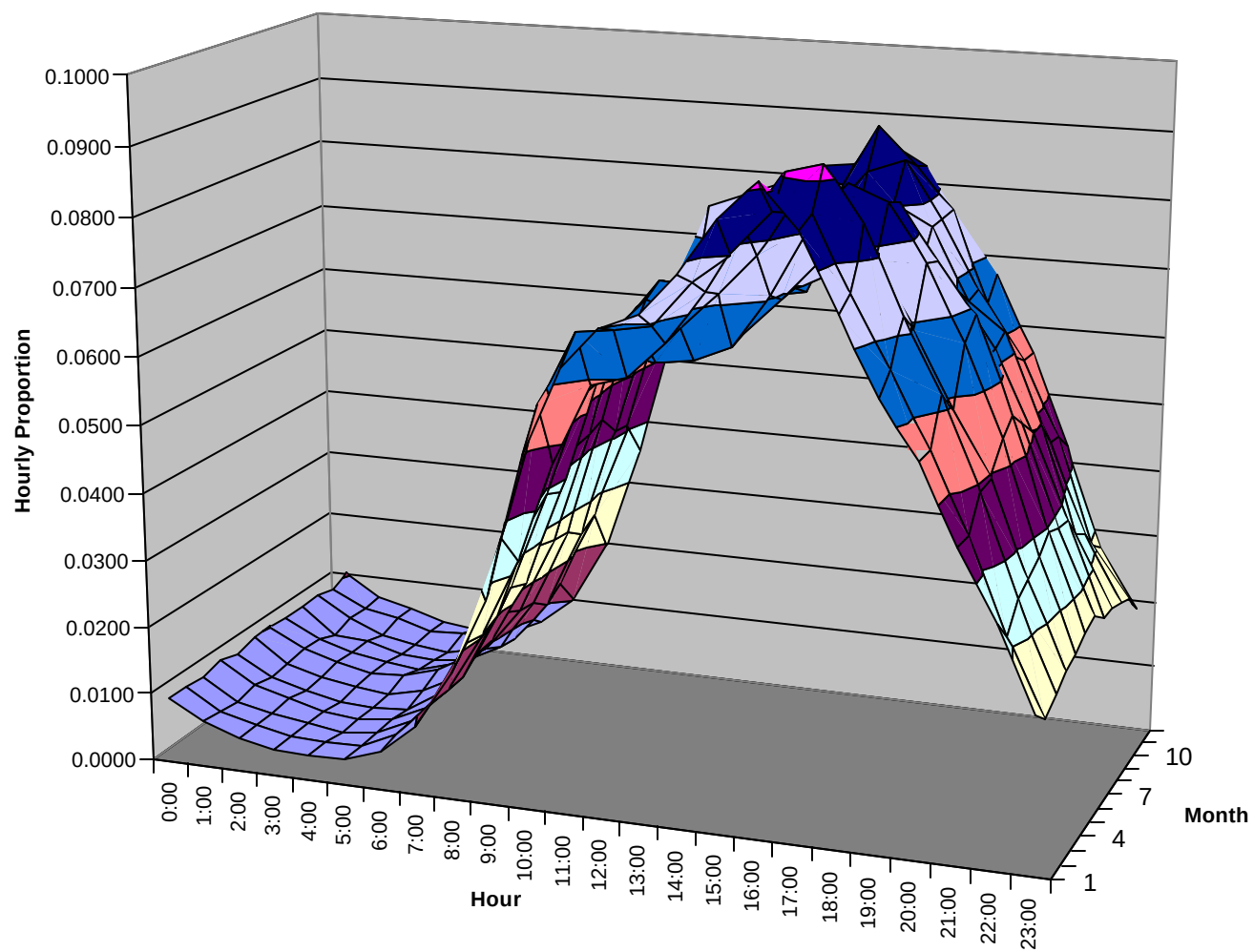


Figure 3-8 Estimated Hourly Proportions by Hour and Month for Sunday for Station 9027-3

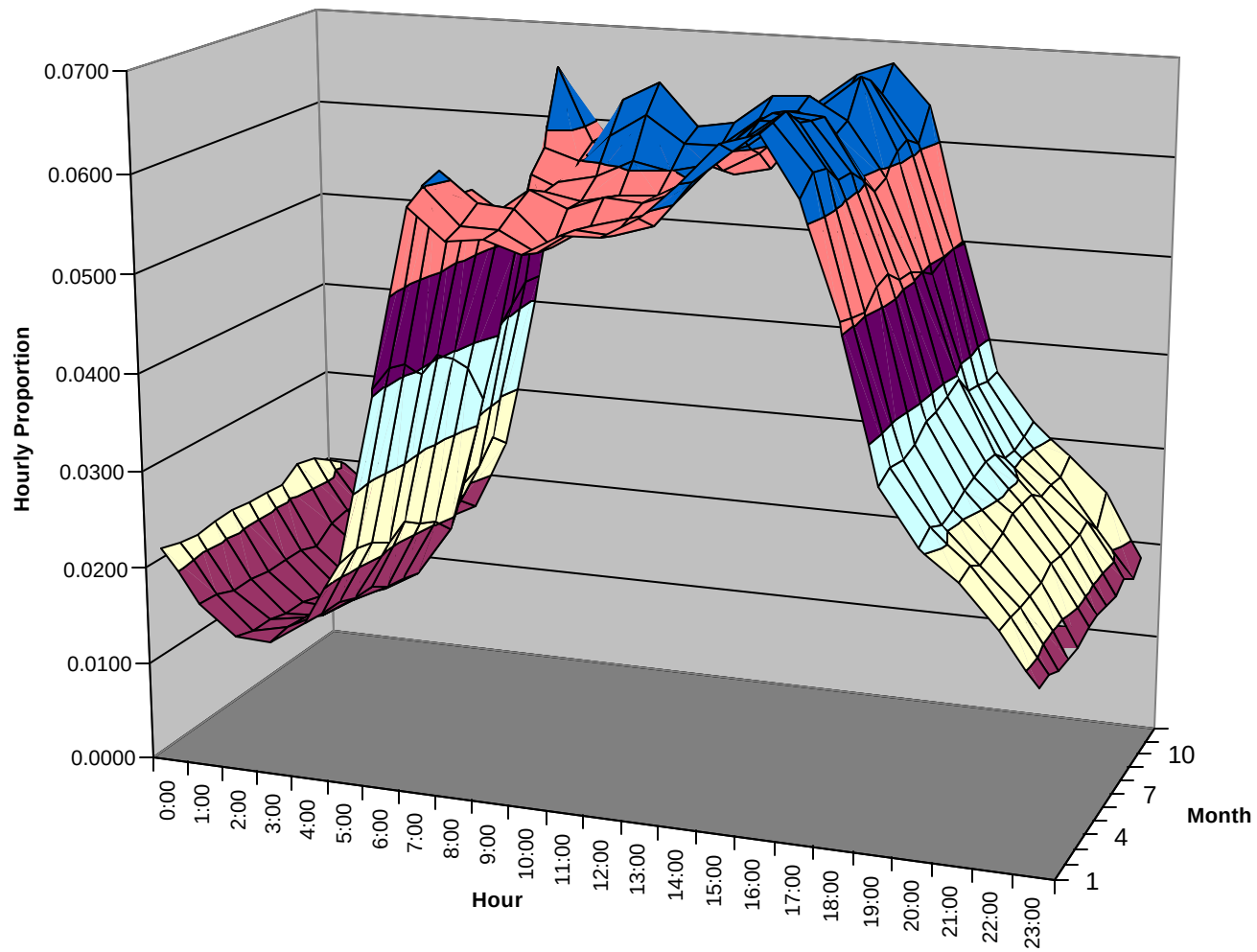


Figure 3-9 Estimated Hourly Proportions by Hour and Month for Monday for Station 9027-3

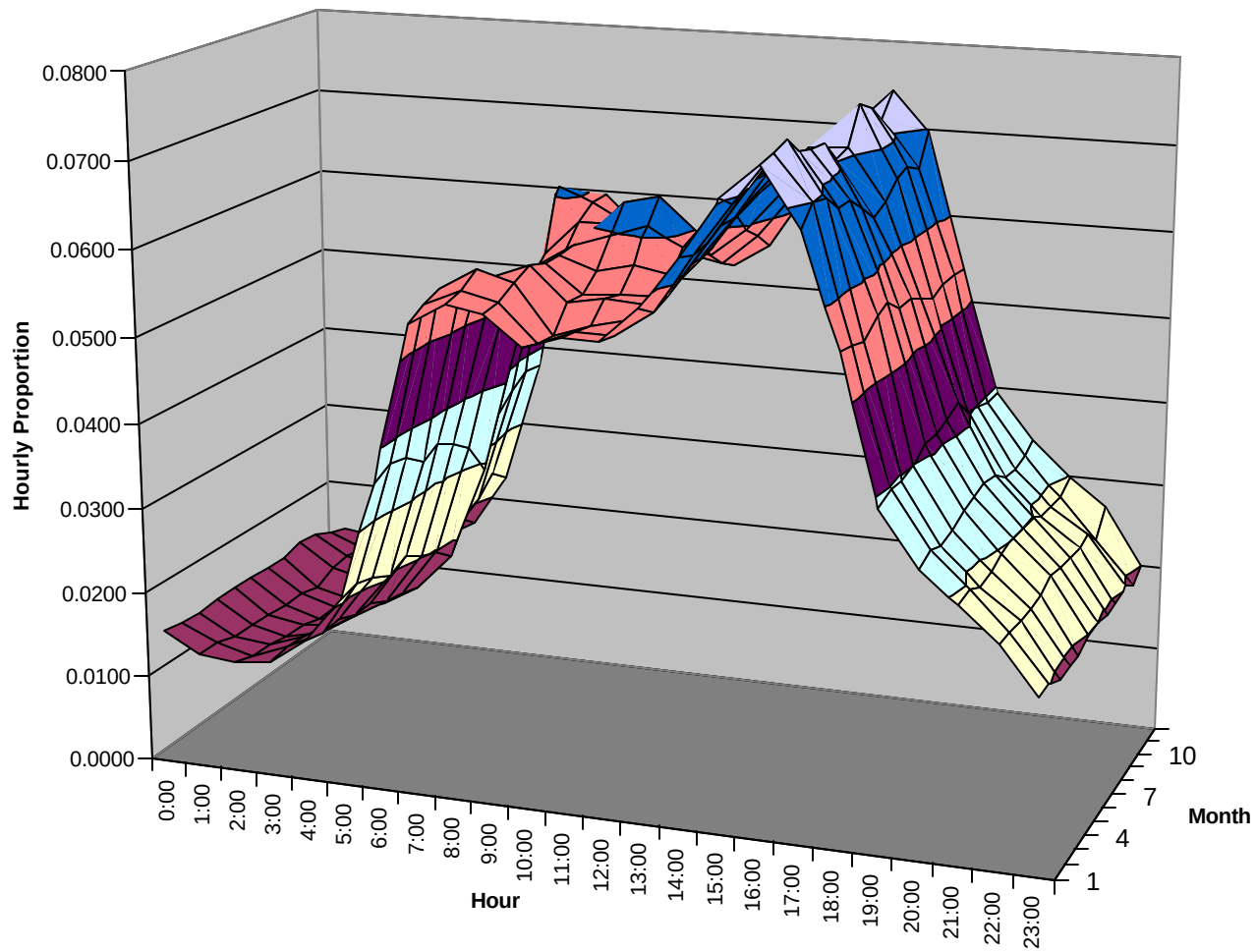


Figure 3-10 Estimated Hourly Proportions by Hour and Month for Tuesday for Station 9027-3

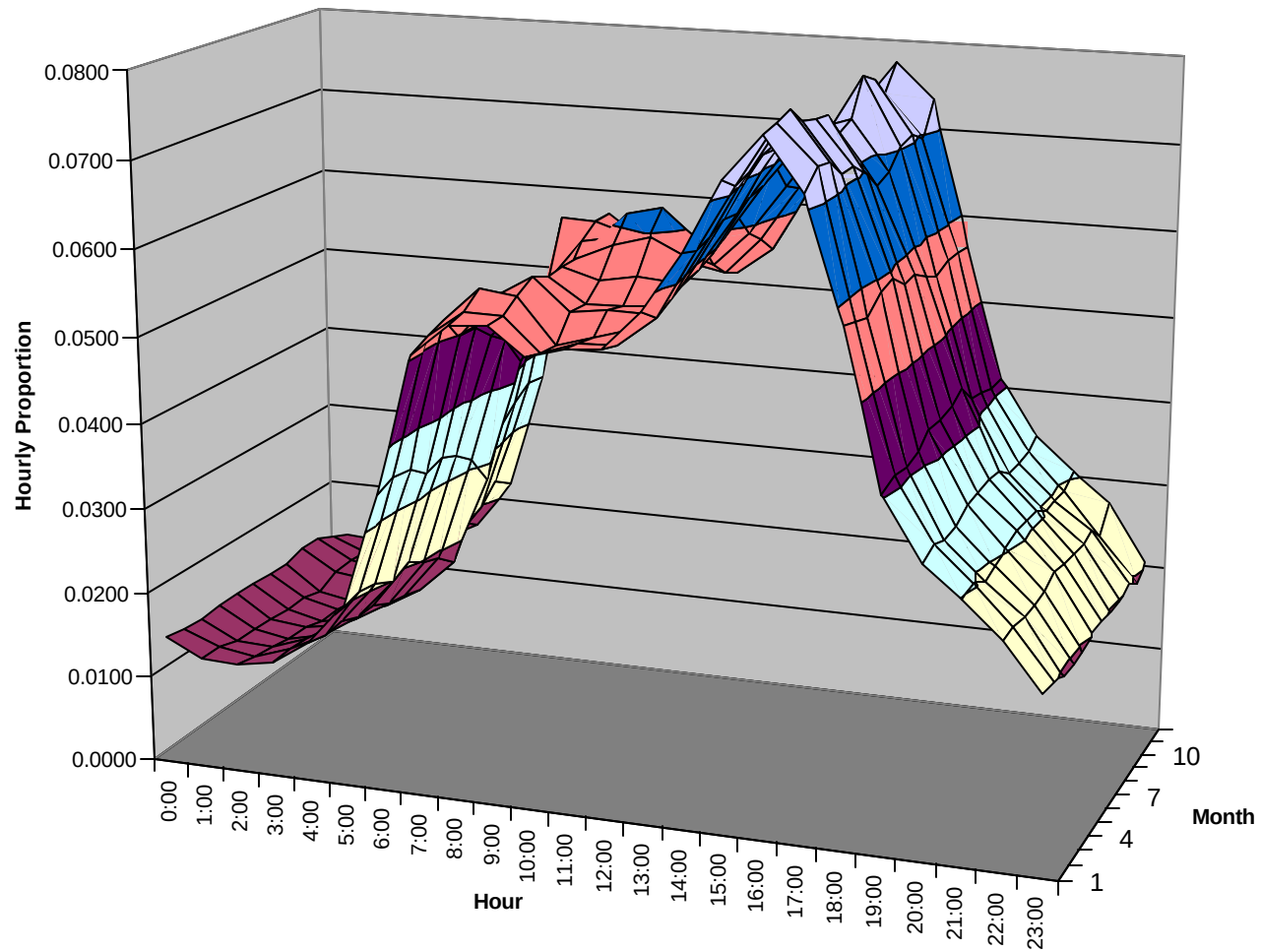


Figure 3-11 Estimated Hourly Proportions by Hour and Month for Wednesday for Station 9027-3

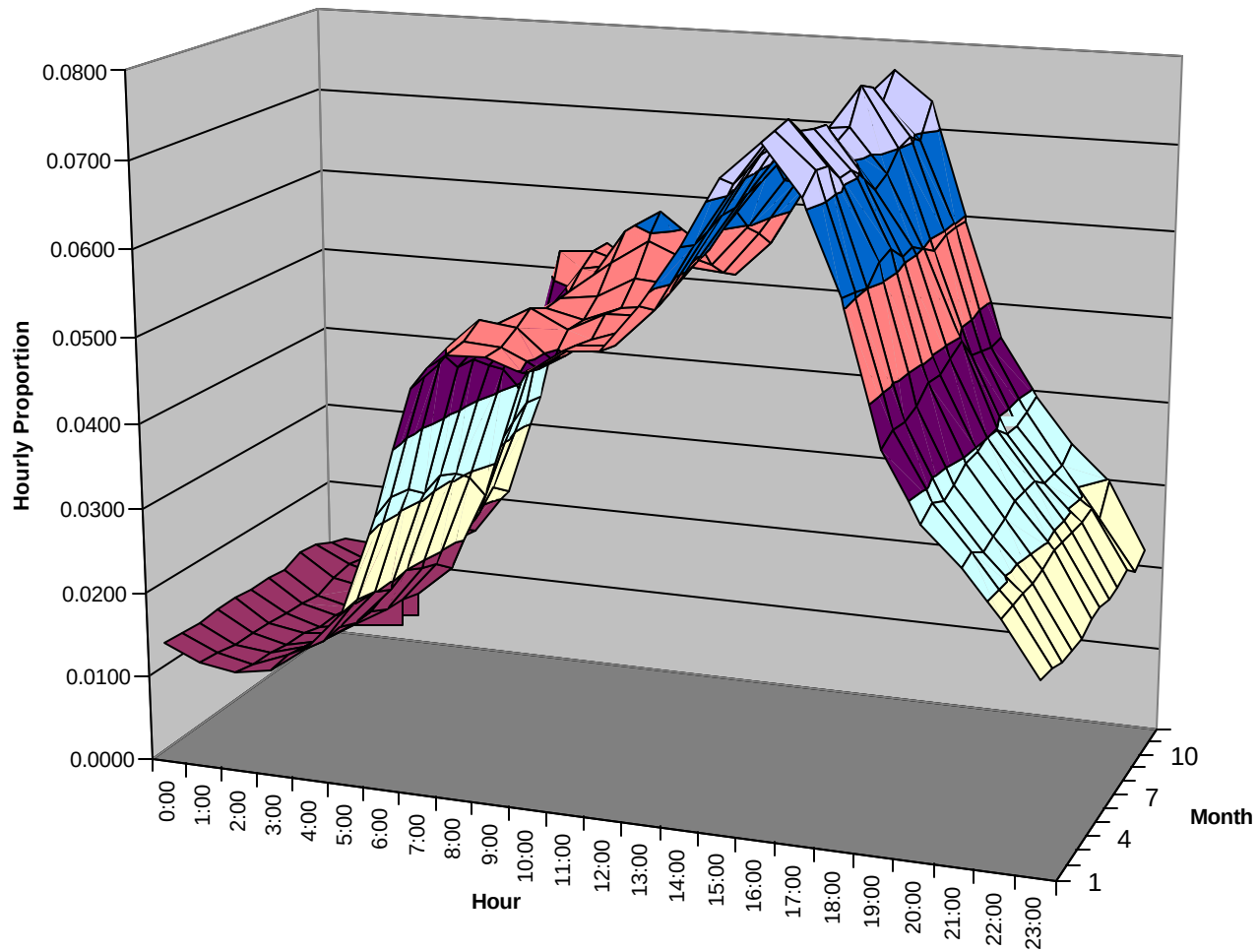


Figure 3-12 Estimated Hourly Proportions by Hour and Month for Thursday for Station 9027-3

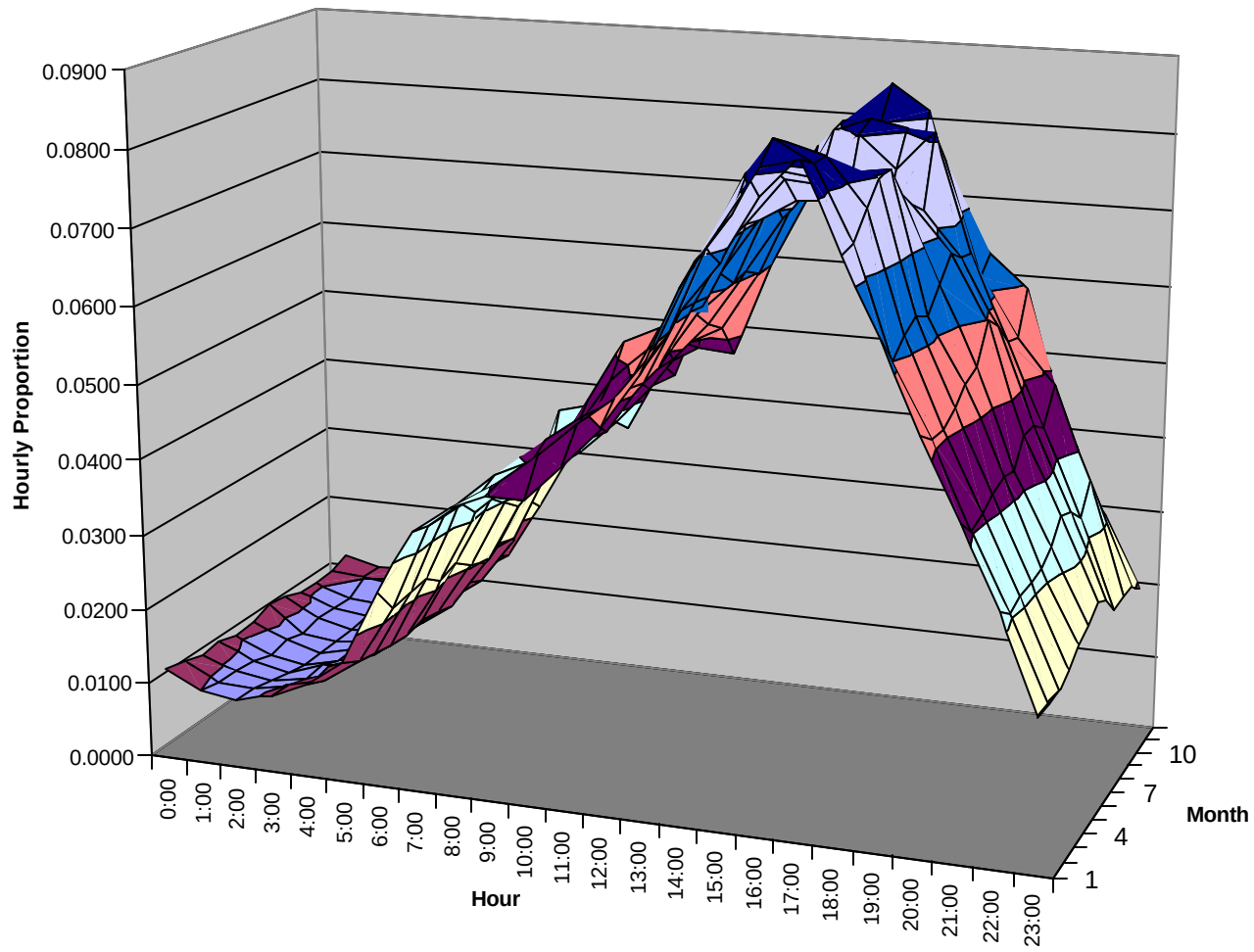


Figure 3-13 Estimated Hourly Proportions by Hour and Month for Friday for Station 9027-3

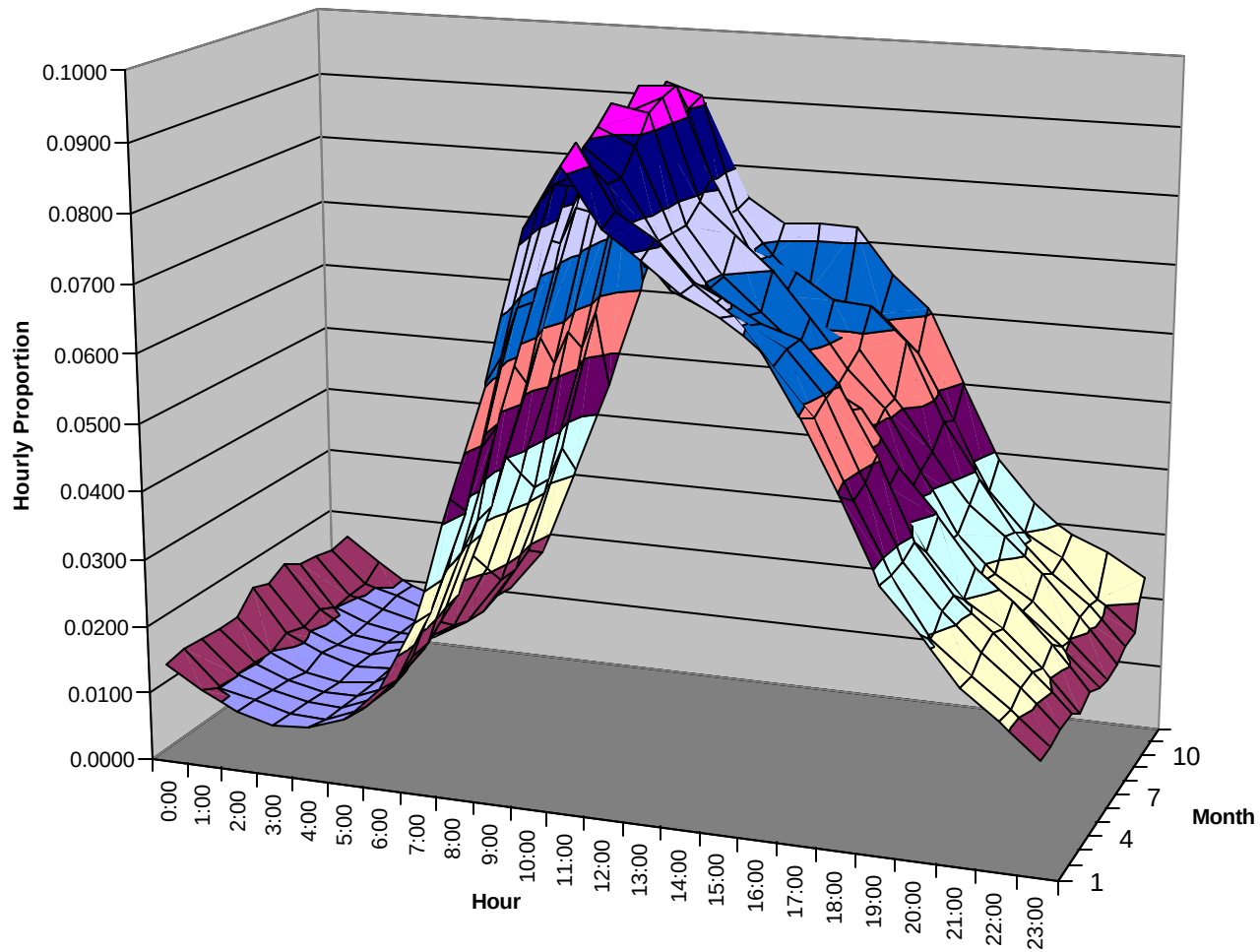


Figure 3-14 Estimated Hourly Proportions by Hour and Month for Saturday for Station 9027-3

3.4 DISCUSSION

Because the hourly proportion models are estimated separately at each study location and data for all study years are pooled together at each location, these models are only valid for each study location and represent the averages of the study years. Therefore, two issues need to be further addressed before these models can be used to estimate hourly proportions for a location and year other than the study locations and years.

One issue is to establish procedures for selecting a model to use for the location of interest. This procedure may involve only a simple comparison of the daily traffic pattern at the location of interest with the daily traffic patterns at the study locations. The daily traffic pattern for the location of interest may be obtained by short-term (*e.g.*, 24 hours) traffic counts. On the other hand, this procedure may involve extensive study of the variation in hourly proportions by locations, which may require identification of the factors that cause the variation and significant amount of data collection.

Another issue is to determine whether or not it is appropriate to use the selected model for the year of interest. This may not be of a great concern in most cases, because in practice it is common to assume that traffic for a specific year (commonly future year) has the same hourly proportion as the study years (or base years). Note that the hourly volumes may change from year to year, and this change is taken into account by the daily volumes. However, if significant changes occur at the location of interest over the years, this average may not be representative to the year of interest, and further study of the yearly variations may be needed.

3.5 CONCLUSIONS

The variations in the hourly proportions of daily volumes by hour, day of week and month are investigated here. It is noted that there are many factors that affect the hourly proportion distributions. These three factors are selected to study here, because they share a common feature, *i.e.*, temporal in nature with distinct cyclical effects to the hourly proportions. In addition, their effects on the hourly proportions are not thoroughly investigated in previous studies. The effects of these factors are examined using ANOVA statistical procedure, with the hourly traffic counts collected at the ART stations on Connecticut freeways as input.

The ANOVA procedure is used to test whether or not the interaction effects (as well as the main effects) of hour, day of week and month are significant. The test results show that the interaction effects of these factors are all significant with over 95 percent confidence. However, at each hour, the interaction effects of month and weekday are in general not significant, even though the main effects of them are still significant with 95 percent confidence.

These findings lead to our hourly proportion models. Specifically, these models are estimated separately at each hour and for weekday, Friday, Saturday and Sunday, individually. For weekdays, the models include the main effect terms of month and weekday, but not their interactions. For Friday, Saturday and Sunday, the models include

only the main effect term for month. The model parameters are estimated using least squares methods. Because these models are developed separately at each study location, and data at each location are pooled together for all study years, these models are only valid for the study locations and represent the average situation of the study years. Consequently, further procedures or models need to be developed before these models can be used to estimate hourly proportions for a specific location and year.

Nevertheless, with some additional research efforts, these models are expected to provide us with more accurate and reliable hourly proportion estimates, which in turn will give us better hourly volume estimates. In addition, the research findings and the models estimated here are helpful in further studies that are devoted to study the effects of other factors to the hourly proportion distributions.

4 GROUPING MONTHS AND WEEKDAYS

4.1 INTRODUCTION

4.1.1 Problem Statement

Hourly proportions (or time of day factors) are commonly needed for estimating highway link hourly volumes, which are important for models in air quality estimation, vehicle crash prediction and transportation planning. Consequently, it is often necessary to estimate hourly proportion models. Many factors should be taken into account in an hourly proportion model, including 1) geometric and operational features, socio-economic characteristics, and land use patterns associated with the highway network, and 2) temporal factors, such as hour, day of week and month. However, it is generally infeasible to include all these factors, because the model can become too complex to estimate.

For example, an hourly proportion model with hour, day of week and month as independent variables may involve more than 2,016 ($24 \times 7 \times 12$) parameters, since they are all categorical with many levels and potentially interact with each other. Adding other factors can make the model even more complex, and thus very difficult to estimate and use.

The model can be significantly simplified, if hour, day of week and month do not have significant interaction effects to the hourly proportions and further they can be grouped concerning the hourly proportions. Hence, in Chapter 3, the significance of hour, day of week and month interactions was studied using Analysis of Variance (ANOVA) statistical procedure. The results showed that at each hour the interactions of month and weekday are in general not significant. This means that, if an hourly proportion model is estimated at an hour, the interaction term of month and weekday can be omitted, which significantly simplify the model.

4.1.2 Objective and Scope

This chapter focuses on procedures to group months and weekdays for further simplifying the hourly proportion models. This grouping of months and weekdays is carried out at each hour and study station here. This is because at each hour the variations in the hourly proportions by months and weekdays are relatively small as shown in Chapter 3, and thus, it is potentially possible to group these two factors. In the next chapter, the 24 hours in the day are further grouped and overall groupings of hours, months and weekdays are produced. Those overall groupings are intended to be used in hourly proportion models that involve hour, month and day of week and other site-specific variables. Using the hour, month and weekday groups as factors, the hourly proportion models can be significantly simplified.

4.2 METHODOLOGY

4.2.1 Overview

At each hour and station, the grouping of months and weekdays is carried out by following three steps:

1. Compare the means of the hourly proportions corresponding to each weekday or month one by one;
2. Group the months or weekdays with means that are not significantly different. This results in multiple choices of month and weekday groupings.
3. Select a proper case from the multiple choices as the final grouping of months or weekdays.

The purpose of the first step is to identify the months or weekdays for which the means of the hourly proportions are not significantly different. It is considered that these months or weekdays can be grouped. Tukey's Comparison of Means statistical method, along with an engineering criterion, is used here to compare the means.

Based on the results of the first step, the second step groups the months or weekdays with means that are not significantly different. Most statistical software packages, such as SAS (SAS Institute Inc. 1990) and SPSS (SPSS Inc. 1998), have a function that provides such groupings as an option of the Comparison of Means procedure. However, these software packages cannot be directly used here because of the incorporation of the engineering criterion in the first step; therefore, an algorithm is developed in this study.

The last step produces groupings of months and weekdays. In fact, based on the comparison results in the first step, one month or weekday may be associated with multiple groups in the second step. Therefore, this step is to establish procedures and criteria to produce mutual exclusive grouping of months or weekdays to be used in an hourly proportion model.

The following sections discuss the procedures used in these three steps individually.

4.2.2 Comparison of Means

4.2.2.1 Tukey's Method

The means of the hourly proportions compared here are

$$\bar{P}_{i..} = \frac{\sum_{j=2}^5 \sum_{m=1}^{N_{ij}} P_{ijm}}{\sum_{j=2}^5 N_{ij}} \quad i = 1, 2, \dots, 12 \quad (4.1)$$

$$\bar{P}_{.j} = \frac{\sum_{i=1}^{12} \sum_{m=1}^{N_{ij}} P_{ijm}}{\sum_{i=1}^{12} N_{ij}} \quad j = 2, 3, 4, 5 \quad (4.2)$$

where $\bar{P}_{i.}$ and $\bar{P}_{.j}$ denote the means for months and weekdays, respectively; P_{ijm} is the hourly proportions for the m^{th} day in month i , weekday j , and N_{ij} is the number of weekdays that fall in month i and weekday j .

Note that the interactions of the factors normally need to be considered when two or more factors are involved as in this case. However, it is proved that the interaction effects of these two factors are in general not significant in Chapter 3, as a result, the interactions do not need to be considered here.

To further identify whether or not the months on Friday, Saturday and Sunday can be grouped, the means for months on each of these days are also compared. These means are

$$\bar{P}_{i.} = \frac{\sum_{m=1}^{N_{ij}} P_{im}}{N_i} \quad i = 1, 2, \dots, 12 \quad (4.3)$$

where $\bar{P}_{i.}$ denotes the means for month i ; P_{im} is the hourly proportions for the m^{th} day in month i , and N_i is the number of weekends that fall in the month.

Tukey's method is used here to compare these means. Two means are declared significantly different in Tukey's method, if the absolute value of their sample difference exceeds a single critical value T_a , defined as

$$T_a = q_a(a, f) S_{\bar{y}_{i.}} \quad (4.4)$$

where $q_a(a, f)$ is the upper α percentage point of the studentized range for the group of means of size a and f error degrees of freedom; and $S_{\bar{y}_{i.}}$ is the standard error of the means defined as

$$S_{\bar{y}_{i.}} = \sqrt{\frac{MS_E}{n_h}} \quad (4.5)$$

where MS_E is the mean square of errors, and n_h is the harmonic mean of the number of observations for each mean (Montgomery 1991).

In using Tukey's method, it is assumed that the observed hourly proportions are normally and independently distributed with constant but unknown variance σ^2 . The observed

hourly proportions by nature lie between 0 and 1, which is a direct violation of the normal distribution assumption. This problem is solved by a Logit transformation

$$Q_{ijkm} = \ln \left(\frac{P_{ijkm}}{1 - P_{ijkm}} \right) \quad (4.6)$$

where P_{ijkm} and Q_{ijkm} are the observed hourly proportion and its Logit transformation, respectively, for the m^{th} day in month i , day of week j , and hour k . This transformation brings data between (0, 1) into a real scale, i.e., between $(-\infty, +\infty)$.

Corresponding to these Logit transformed data, the means to be compared become

$$\bar{Q}_{i..} = \frac{\sum_{j=2}^5 \sum_{m=1}^{N_{ij}} Q_{ijm}}{\sum_{j=2}^5 N_{ij}} \quad i = 1, 2, \dots, 12 \quad (4.7)$$

$$\bar{Q}_{.j.} = \frac{\sum_{i=1}^{12} \sum_{m=1}^{N_{ij}} Q_{ijm}}{\sum_{i=1}^{12} N_{ij}} \quad j = 2, 3, 4, 5 \quad (4.8)$$

$$\bar{Q}_{i.} = \frac{\sum_{m=1}^{N_{ij}} Q_{im}}{N_i} \quad i = 1, 2, \dots, 12 \quad (4.9)$$

Note that it is proved that the assumptions stated earlier are generally valid for the transformed data in Chapter 3. In addition, even though the Logit transformed data are used here, the conclusions drawn using the transformed data are applied to the hourly proportions. This is because it is expected that the errors (if any) in the conclusions caused by using the transformation would be very limited. The range of the hourly proportions is actually very small at a particular hour (0.05 on average); consequently, the relationship between the hourly proportion and its transformation becomes almost linear. In fact, the comparison results would be the same using either set of the data if a linear relationship exists between them.

As an example, Tukey's comparison for hourly proportions corresponding to weekdays at the morning peak hour (starting at 7:00) at ATR station 9030-1 (located on I-95 northbound in Norwalk, Connecticut) is shown in Table 4.1, where the first column shows the weekday pairs. These pairs are formed by first ranking the weekdays in the descending order of their corresponding hourly proportions (after the Logit transformation). Then the ranked weekdays are paired, beginning with the first versus the second, and then the first versus the third, until the first weekday has been paired with all other weekdays. This process is continued with the second weekday until all

Table 4-1 Tukey's Comparisons and Engineering Criterion Incorporation

Weekday Pair	Mean Diff.	Tukey's Comp.	Vol.-1	Vol.-2	Vol. Diff.	Comp. Result
3 – 2	0.0093	-	-	-	-	-
3 – 4	0.0113	0	3376	3340	36	
3 – 5	0.0415	0	3376	3246	130	0
2 – 3	-0.0093	-	-	-	-	-
2 – 4	0.0020	-	-	-	-	-
2 – 5	0.0322	0	3347	3246	101	0
4 – 3	-0.0113	0	3340	3376	36	-
4 – 2	-0.0020	-	-	-	-	-
4 – 5	0.0301	0	3340	3246	94	0
5 – 3	-0.0415	0	3246	3376	130	0
5 – 2	-0.0322	0	3246	3347	101	0
5 – 4	-0.0301	0	3246	3340	94	0

Note: boldface indicates the final comparison results that are different from Tukey's comparison results.

weekdays are paired. Column 2 gives the differences of the hourly proportions corresponding to the weekday pairs. These differences are then compared with the critical value T_α calculated for a pre-specified confidence level (90% in this case). If the difference is larger than the critical value, the two weekdays in a pair are declared to have significantly different hourly proportions, as indicated in Column 3 with '0's.

4.2.2.2 Engineering Criterion

There is a drawback in using Tukey's procedure only to compare the means in this case. The actual difference of two means could be very small and yet declared to be significant in Tukey's procedure. Extremely small (yet statistically significant) differences in the means are not of much importance from the transportation engineering point of view. To overcome this drawback, an engineering criterion is incorporated into the Tukey's comparison results.

Here the engineering criterion is established towards estimated hourly volumes; specifically, the difference between two hourly volume estimates is considered to be negligible if it is smaller than a tolerance level t (in this case, a value of 50.0 is used for the tolerance level following the convention in traffic forecasting practice). This criterion can be expressed as

$$|ADT \times P_1 - ADT \times P_2| \leq t \quad (4.10)$$

where P_1 and P_2 denote means calculated using the observed hourly proportions, and ADT denotes an average daily volume corresponding to the means. Note that the

products of the means by the daily volume are hourly volume estimates, since the means represent estimated hourly proportions.

An example of the incorporation of the engineering criterion is also shown in Table 4.1 (Columns 4 through 7). For the pairs of weekdays with significantly different hourly proportions according to Tukey's comparisons, the hourly proportions and their corresponding average daily volumes (ADTs) are first calculated. The products of the hourly proportions and the ADTs are shown in Columns 4 and 5, and the differences of these products are given in Column 6. These differences are compared with the tolerance value (50 in this case) and the comparison results based on this engineering criterion are shown in Column 7, with the means that are significant different indicated again by '0's. These comparison results are taken as the final comparison results for the pairs of weekdays with significantly different hourly proportions according to Tukey's comparisons. In the table the final comparison results that are different from Tukey's comparison results are indicated in boldface.

4.2.3 Group Pattern Generation

As mentioned earlier, the purpose of performing the comparison of means is to find the months or weekdays with the means of hourly proportions that are not significantly different and, in turn, to group the months or weekdays. Hence, it is desirable to convert the one by one comparison results (*e.g.*, Table 4.1) to groups of months or weekdays, or group patterns, with the months or weekdays in each group having the means that are not significantly different. The following sections describe the algorithm developed to produce such group patterns.

4.2.3.1 Convert Comparison Results to a Matrix

In this first step, the comparison results (*e.g.*, Column 7 in Table 4.1) are translated into a matrix format for easy manipulating using a computer program. Here the rows and columns of a matrix are used to indicate the month or weekday pairs, and '0's or '1's are assigned to the elements of the matrix to indicate the comparison results. For example, the comparison results shown in Column 7 of Table 4.1 are translated into a matrix as shown in Figure 4.1 (a), where the rows and columns correspond to the weekday pairs and the '0's or '1's are corresponding to the comparison results. Specifically, the '0's indicate the comparisons that are not significantly different (corresponding to the '✓' in Column 7 of Table 4.1), and the '1's indicate that the comparisons that are significantly different (corresponding to the '-' in Column 7 of Table 4.1). Note that here the

	3	2	4	5
3	ÿ	1	1	0
2	1	ÿ	1	0
4	1	1	ÿ	0
5	0	0	0	ÿ

(a)

	3	2	4	5
3	1	1	1	0
2	1	1	1	0
4	1	1	1	0
5	0	0	0	1

(b)

Figure 4-1 Convert Comparison Results to Matrix Example

weekdays indicated by the rows and columns of the matrix are in the descending order of the corresponding means instead of the natural order of the weekdays.

4.2.3.2 Assign Empty Elements with '1's

Because of the setup in the comparison of means procedure, the elements in the diagonal of the matrix are left empty. These elements should be filled with '1's as shown in Figure 4.1 (b), because these elements indicate the comparisons of the means corresponding to each individual weekday itself. It is obvious that a mean cannot be significantly different from itself; therefore, these elements at the diagonal should be all '1's.

4.2.3.3 Take the Lower Triangle of the Matrix

This and the next steps are to convert the matrix to groups of the months and weekdays. To better illustrate how this conversion is conducted here and in the next step, another example in larger scale is given in Figure 4.2. This matrix is produced for the 12 months in the year for the morning peak hour for Station 9727-3, based on the comparison results.

As can be seen, the matrices shown in Figure 4.1 (b) and Figure 4.2 are both symmetric over the diagonal. This symmetric property is due to the way that the matrixes are produced in the first two steps. Consequently, we can use only the lower triangle of the matrix to produce the groupings of months and weekdays. Figure 4.3 shows the lower triangle of the matrix taken from the matrix shown in Figure 4.2.

4.2.3.4 Produce Group Pattern

The columns in the lower triangle of the matrix can be divided into separate sections based on the patterns of the '1's in the matrix. For example, the columns in the matrix shown in Figure 4.3 can be divided into seven column sections based on the horizontal sections of the zigzag line in the figure. Taking the first column in each column section, the months in Figure 4.3 can be grouped as shown in Figure 4.4.

	2	1	11	3	4	9	10	5	12	7	6	8
2	1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0	0	0	0
11	1	1	1	1	1	0	0	0	0	0	0	0
3	0	1	1	1	1	1	1	1	1	0	0	0
4	0	0	1	1	1	1	1	1	1	0	0	0
9	0	0	0	1	1	1	1	1	1	1	0	0
10	0	0	0	1	1	1	1	1	1	1	1	0
5	0	0	0	1	1	1	1	1	1	1	1	0
12	0	0	0	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	1	1	1	1	1	1	1
6	0	0	0	0	0	0	1	1	1	1	1	1
8	0	0	0	0	0	0	0	0	1	1	1	1

Figure 4-2 Matrix for Month Example

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	1						
1	1	1					
11	1	1	1				
3	0	1	1	1			
4	0	0	1	1	1		
9	0	0	0	1	1	1	
10	0	0	0	1	1	1	1
5	0	0	0	1	1	1	1
12	0	0	0	1	1	1	1
7	0	0	0	0	0	1	1
6	0	0	0	0	0	0	1
8	0	0	0	0	0	0	1

Figure 4-3 Group Pattern Generation Example

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2	1						
1	1	1					
11	1	1	1				
3	0	1	1	1			
4	0	0	1	1			
9	0	0	0	1	1		
10	0	0	0	1	1	1	
5	0	0	0	1	1	1	
12	0	0	0	1	1	1	1
7	0	0	0	0	1	1	1
6	0	0	0	0	0	1	1
8	0	0	0	0	0	0	1

Figure 4-4 Month Group Pattern Example

This grouping is valid because of the reasons as follows. An examination of the matrix shown in Figure 4.3 reveals that, starting from the diagonal, all rows or columns begin with '1's and continue until a '0' occurs, and then continue with '0's. In other words, the '1's and '0's in the matrix can be divided into two separate segments using the zigzag line shown in the figure with the '1's on top of the '0's. Also, the '1's can be divided into column sections based on the horizontal sections of the zigzag line. Because the string of '1's in each column indicates that their corresponding months (indicated by the row number) can be grouped together, the first column in each column section contains the information in the other columns in the same column section, and hence, the first column in each column section represents the group patterns in that column section. For the same reason, each column section contains a unique group of months. Therefore, the group pattern formed using the first columns represent the overall group pattern.

4.2.4 Grouping Selection

As shown in Figure 4.4, there are overlaps of the '1's in some rows of the matrix. This means that a month may fall in two or more groups; as a result, multiple choices of month groupings can be produced based on the group pattern. To better explain the choices and the grouping selection procedure to be discussed a little later, another example of the group pattern is given in Figure 4.5 (a). This group pattern is introduced here because it is simpler and thus more suitable for this discussion. As mentioned earlier, the months in Figure 4.5a (and Figure 4.4) are in the order of their corresponding means rather than their natural order. Considering that it does not make much sense to group non-adjacent months from a practical viewpoint, the months are rearranged in their natural order. To make the following discussion easy to understand, the group pattern in Figure 4.5 (a) is further transposed as shown in Figure 4.5 (b).

Based on the group pattern in Figure 4.5, the months can be grouped as:

- 1) 1- January through June, 2 - July through December;

2) 1- January through July, 2- August through December;

3) 1- January through August, 2 - September through December.

These are only a few of the many choices. To use the month (and weekday) groupings in an hourly proportion model, it is desirable to produce mutually exclusive groupings of months or weekdays. Therefore, a procedure needs to be established to identify the choices and to select one case from the choices to represent the final grouping of months or weekdays.

The choices are identified here by first obtaining the permutations (*i.e.*, groupings) of the months and weekdays that involve only adjacent months or weekdays. The groupings are then fitted to a group pattern, and the groupings that fit the group pattern are the choices. This is done for each of the group patterns for different hours and stations. Note that the resulting groupings involve only adjacent months or weekdays and are statistically valid. One grouping is then selected from the choices to be the final grouping of months or

	(1)	(2)
5	1	
4	1	1
2	1	1
1	1	1
6	1	1
7	1	1
10	1	1
3	1	1
8	1	1
12	1	1
11	0	1
9	0	1

(a)

Month											
1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1		1		1
1	1	1	1		1	1	1	1	1	1	1

(b)

Figure 4-5 Another Month Group Pattern Example

weekday for the group pattern. Specifically, the groupings with smallest number of groups of months or weekdays are first selected. Among the selected groupings, the one that fits the maximum number of group patterns across the hours and stations is then further selected as the final grouping.

The following sections further explain this procedure using the group pattern given in Figure 4.5 (b) as an example.

4.2.4.1 Choices Identification

The possible groupings of adjacent months or weekdays are produced by identifying all possible combinations of months or weekdays that include only adjacent months or weekdays. Here the twelve months in the year are considered cyclical, and thus December and January are considered adjacent. Consequently, for months, this case is analogous to a round table scenario with twelve persons and each with a fixed seat. For weekdays, it is slightly different in that the weekdays are not considered cyclical. Nonetheless, this is a permutation and combination problem that can be solved easily using a computer program.

Table 4.2 gives samples of such groupings for months, where the twelve numbers in Columns 1 through 12 in each row together represent a month grouping; specifically, the months with the same number belong to a group. For identification purpose, here a grouping index consisting of two numbers is assigned to each grouping as shown in the

Table 4-2 Samples of Possible Adjacent Month Groupings

Month												Grouping Index
1	2	3	4	5	6	7	8	9	10	11	12	
1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
1	1	1	1	1	1	1	1	1	1	1	2	2 - 1
1	1	1	1	1	1	1	1	1	1	2	1	2 - 2
...												...
1	2	2	2	2	2	2	2	2	2	2	2	2 - 66
1	1	1	1	1	1	1	1	1	1	2	3	3 - 1
1	1	1	1	1	1	1	1	1	2	2	3	3 - 2
...												...
1	2	3	3	3	3	3	3	3	3	3	3	3 - 220
1	1	1	1	1	1	1	1	1	2	3	4	4 - 1
1	1	1	1	1	1	1	1	2	2	3	4	4 - 2
...												...
1	2	3	4	4	4	4	4	4	4	4	4	4 - 495
1	1	1	1	1	1	1	1	2	3	4	5	5 - 1
1	1	1	1	1	1	1	2	2	3	4	5	5 - 2
...												...
1	2	3	4	5	5	5	5	5	5	5	5	5 - 792
1	1	1	1	1	1	1	2	3	4	5	6	6 - 1
1	1	1	1	1	1	2	2	3	4	5	6	6 - 2
...												...
1	2	3	4	5	6	6	6	6	6	6	6	6 - 924
1	1	1	1	1	1	2	3	4	5	6	7	7 - 1
1	1	1	1	1	2	2	3	4	5	6	7	7 - 2
...												...
1	2	3	4	5	6	7	7	7	7	7	7	7 - 792
1	1	1	1	1	2	3	4	5	6	7	8	8 - 1
1	1	1	1	2	2	3	4	5	6	7	8	8 - 2
...												...
1	2	3	4	5	6	7	8	8	8	8	8	8 - 495
1	1	1	1	2	3	4	5	6	7	8	9	9 - 1
1	1	1	2	2	3	4	5	6	7	8	9	9 - 2
...												...
1	2	3	4	5	6	7	8	9	9	9	9	9 - 220
1	1	1	2	3	4	5	6	7	8	9	10	10 - 1
1	1	2	2	3	4	5	6	7	8	9	10	10 - 2
...												...
1	2	3	4	5	6	7	8	9	10	10	10	10 - 66
1	1	2	3	4	5	6	7	8	9	10	11	11 - 1
1	2	2	3	4	5	6	7	8	9	10	11	11 - 2
...												...
1	2	3	4	5	6	7	8	9	10	11	11	11 - 12
1	2	3	4	5	6	7	8	9	10	11	12	12 - 1

table. The first number is in fact the number of month (or weekday) groups in the grouping, and the second number is a sequential index number assigned to each particular grouping. This assignment is done separately for each set of groupings with the same number of month (or weekday) groups. Note that in Table 4.2 only the first one or two and the last groupings for each number of groups are given. This is because there are actually 4,084 possible grouping patterns and it is cumbersome to list all of them. In addition, Table 4.3 lists the weekday groupings; all are listed here due to the much smaller number of possible groupings.

Table 4-3 Possible Adjacent Weekday Groupings

Weekday				Grouping Index
2	3	4	5	
1	1	1	1	1-1
1	1	1	2	2-1
1	1	2	2	2-2
1	2	2	2	2-3
1	1	2	3	3-1
1	2	2	3	3-2
1	2	3	3	3-3
1	2	3	4	4-1

These groupings are then fitted to the group patterns by matching each of the month or weekday groups (indicated by the same numbers) in the grouping to the '1's in the group patterns (*e.g.*, Figure 4.5b). Specifically, if a contiguous string of '1's can be found in the group pattern for each month or weekday group in a grouping, the grouping is considered to fit the group pattern. For example, Table 4.4 gives samples of the fitted groupings for the group pattern shown in Figure 4.5b. Specifically, the first grouping (2 - 9) in the table fits the group pattern, because contiguous strings of '1's exist in the first row in the figure for Month Group 1 - December through August, and contiguous strings of '1's exist in the second row in the figure for Month Group 2 - September through November. These fitted groupings are the choices stated earlier.

4.2.4.2 Month/Weekday Grouping Selection

Now we need to select one from the choices to be the final grouping for the group pattern. First, the groupings with smallest first number of the grouping index is selected, since this number indicates the number of month or weekday groups for the grouping. For example, the groupings highlighted in Table 4.4 are selected for this case.

There are still multiple choices of groupings. To further narrow down to one, the grouping with largest number of occurrences over the 24 hours in the day and the study stations is further selected. Table 4.5 gives the numbers of occurrences over 24 hours at the study location and over all study locations for the groupings highlighted in Table 4.4. First, the two groupings with largest numbers of occurrences (4) as shown in Column 2 in Table 4.5 are selected. Then, between these two groupings, the one with larger number

Table 4-4 Fitted Month Grouping Samples

Month												Grouping Index
1	2	3	4	5	6	7	8	9	10	11	12	
1	1	1	1	1	1	1	1	2	2	2	1	2 - 9
1	1	1	1	1	1	1	1	2	2	2	2	2 - 10
1	1	1	1	1	1	1	2	2	2	2	1	2 - 14
1	1	1	1	1	1	1	2	2	2	2	2	2 - 15
1	1	1	1	1	1	2	2	2	2	2	1	2 - 20
1	1	1	1	1	1	2	2	2	2	2	2	2 - 21
1	1	1	1	1	2	2	2	2	2	2	1	2 - 27
1	1	1	1	1	2	2	2	2	2	2	2	2 - 28
1	1	1	1	2	1	1	1	1	1	1	1	2 - 29
1	1	1	1	2	2	1	1	1	1	1	1	2 - 30
1	1	1	1	2	2	2	1	1	1	1	1	2 - 31
1	1	1	1	2	2	2	2	1	1	1	1	2 - 32
1	1	1	2	2	1	1	1	1	1	1	1	2 - 38
1	1	1	2	2	2	1	1	1	1	1	1	2 - 39
1	1	1	2	2	2	2	1	1	1	1	1	2 - 40
1	1	1	2	2	2	2	2	1	1	1	1	2 - 41
1	1	2	2	2	1	1	1	1	1	1	1	2 - 48
1	1	2	2	2	2	1	1	1	1	1	1	2 - 49
1	1	2	2	2	2	2	1	1	1	1	1	2 - 50
1	1	2	2	2	2	2	2	1	1	1	1	2 - 51
1	2	2	2	2	1	1	1	1	1	1	1	2 - 59
1	2	2	2	2	2	1	1	1	1	1	1	2 - 60
1	2	2	2	2	2	2	1	1	1	1	1	2 - 61
1	2	2	2	2	2	2	2	1	1	1	1	2 - 62
1	1	1	1	2	3	1	1	1	1	1	1	3 - 78
1	2	2	2	2	2	3	3	3	1	1	1	3 - 183
1	1	1	1	1	1	1	1	2	3	4	1	4 - 4
1	2	2	2	2	3	4	4	4	4	4	1	4 - 385
1	1	1	1	1	2	2	2	3	4	4	5	5 - 24
1	2	3	4	4	5	5	5	5	1	1	1	5 - 781
1	1	1	1	1	2	3	3	4	4	5	6	6 - 15
1	2	3	4	5	6	6	6	6	6	1	1	6 - 922
1	1	1	1	2	2	3	4	5	6	7	1	7 - 14
1	2	3	4	5	6	6	6	6	7	1	1	7 - 775
1	1	1	2	3	4	4	5	6	7	8	1	8 - 29
1	2	3	4	4	5	6	6	6	7	8	1	8 - 408
1	2	2	3	4	5	5	6	7	8	8	9	9 - 83
1	2	3	4	5	5	6	7	8	9	1	1	9 - 183
1	2	3	4	4	5	6	7	8	8	9	10	10 - 35
1	2	3	4	5	6	6	7	7	8	9	10	10 - 47
1	1	2	3	4	5	6	7	8	9	10	11	11 - 1
1	2	3	4	5	6	7	8	9	10	11	11	11 - 12
1	2	3	4	5	6	7	8	9	10	11	12	12 - 1

Table 4-5 Grouping Selection

Grouping Index	No. of Occurrences (24 hours)	No. of Occurrences (all study location)
2 - 9	2	-
2 - 10	2	-
2 - 14	1	-
2 - 15	1	-
2 - 20	1	-
2 - 21	1	-
2 - 27	1	-
2 - 28	1	-
2 - 29	1	-
2 - 30	1	-
2 - 31	1	-
2 - 32	4	772
2 - 38	1	-
2 - 39	1	-
2 - 40	1	-
2 - 41	4	719
2 - 48	1	-
2 - 49	1	-
2 - 50	1	-
2 - 51	2	-
2 - 59	1	-
2 - 60	1	-
2 - 61	1	-
2 - 62	2	-

of occurrences over all study locations, *i.e.*, the grouping with index 2-32 is selected as the final grouping of months for this case.

4.3 RESULTS AND DISCUSSION

The procedures established here are implemented using computer programs written in C++ and MatLab (The Mathworks, Inc 1998). The final results of this implementation are the month and weekday groupings at each hour for each of the study stations, with the month groupings produced separately for weekday, Friday, Saturday and Sunday. For example, Tables 4.6 and 4.7 show the month groupings on weekdays and weekday groupings, respectively, for all hours for Station 9027-3. For brevity, the groupings for other stations are not given here.

An examination of the groupings in Tables 4.6 and 4.7, particularly the numbers of month or weekday groups indicated by the first numbers of the grouping indexes, reveals that the months and weekdays can be grouped to a considerable level. Specifically, the months and weekdays can be grouped into less than two groups for nearly 80 percent of the groupings (19 out of 24 for both month and weekday Tables 4.6 and 4.7 also reveal that the groupings vary by hour, with the a.m. and p.m. peak hours having larger numbers of month and weekday groups than the other hours in the day. This variation means that the groupings cannot be directly used in an hourly proportion model that involves hour, month and day of week as factors. In other words, hourly proportion models may have to be estimated separately for each hour, if these groupings are used. As a result, there still will be a significantly number of hourly proportion models to estimate. To simplify the model estimation, it is desirable to further group the 24 hours in the day. This is the issue to be discussed in the next chapter.

4.4 CONCLUSIONS

Procedures are established to group months and weekdays for the purpose of simplifying hourly proportion models that involve these factors and other site-specific variables. Specifically, Tukey's method along with an engineering criterion is used to compare the means of the hourly proportions for each month and weekday at each hour and study location. Then, a procedure is developed to produce mutual exclusive groupings of months and weekdays based on the comparison results.

An examination of the groupings reveals that the months and weekdays can be grouped to a considerable level. This means that, when these groupings are used, the hourly proportion models can be significantly simplified. This makes the hourly proportion models easier to estimate and more desirable to use in practice.

Our examination of the groupings also reveals that they vary by the hours. This means that the groupings may not be suitable for using in an hourly proportion model that involves hour, month and day of week as factors. In other words, hourly proportion models may have to be estimated separately for each hour, if these groupings are used. As a result, there still will be a significantly number of hourly proportion models to estimate. To simplify the model estimation, it is desirable to further group the 24 hours in the day. This issue will be addressed in the next chapter.

Table 4-6 Month Groupings by Hour for Station 9027-3

Hour	Month												Grouping Index
	1	2	3	4	5	6	7	8	9	10	11	12	
0:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
1:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
2:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
3:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
4:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
5:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
6:00	1	1	1	1	1	2	2	3	1	1	1	1	3 - 46
7:00	1	1	1	1	2	2	2	3	4	4	5	5	5 - 68
8:00	1	1	1	1	1	2	2	2	2	3	3	3	3 - 41
9:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
10:00	1	1	1	2	2	2	3	3	4	4	4	1	4 - 155
11:00	1	1	1	2	2	2	3	3	4	4	4	1	4 - 155
12:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
13:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
14:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
15:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
16:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
17:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
18:00	1	1	1	1	2	2	2	2	1	1	1	1	2 - 32
19:00	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
20:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
21:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
22:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
23:00	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1

Table 4-7 Weekday Groupings by Hour for Station 9027-3

Hour	Weekday				Grouping Index
	2	3	4	5	
0:00	1	2	2	2	2 - 3
1:00	1	2	2	2	2 - 3
2:00	1	1	1	1	1 - 1
3:00	1	1	1	1	1 - 1
4:00	1	1	1	1	1 - 1
5:00	1	1	1	1	1 - 1
6:00	1	2	2	3	3 - 2
7:00	1	2	2	3	3 - 2
8:00	1	1	1	2	2 - 1
9:00	1	1	1	2	2 - 1
10:00	1	2	2	2	2 - 3
11:00	1	2	2	2	2 - 3
12:00	1	2	2	2	2 - 3
13:00	1	1	1	1	1 - 1
14:00	1	1	1	1	1 - 1
15:00	1	2	2	2	2 - 3
16:00	1	2	2	2	2 - 3
17:00	1	2	3	3	3 - 3
18:00	1	2	2	3	3 - 2
19:00	1	2	2	3	3 - 2
20:00	1	1	1	2	2 - 1
21:00	1	1	1	2	2 - 1
22:00	1	2	2	2	2 - 3
23:00	1	2	2	2	2 - 3

5 GROUPING HOURS AND ESTIMATING HOURLY PROPORTIONS

5.1 INTRODUCTION

As discussed in Chapter 4, many factors should be taken into account in an hourly proportion model; however, it is generally infeasible to do so, because the model can become too complex to estimate. Consequently, procedures were established in Chapter 4 to group months and weekdays at each hour and study station for simplifying the hourly proportion model. Examination of the resulting groupings of months and weekdays revealed that they vary by hour, meaning that the groupings may not be suitable for using in an hourly proportion model that involves hour, month and day of week as factors. In other words, hourly proportion models may have to be estimated separately for each hour when the groupings are used. This results in a significant number of hourly proportion models to estimate.

This problem can be solved by further grouping the 24 hours in the day. Therefore, this chapter is devoted to the following issues:

- Grouping the 24 hours in the day, and further
- Producing groupings of months and weekdays for each hour group.

In addition, this chapter is also devoted to

- Estimating hourly proportion models based on the groupings, and
- Calculating corresponding hourly proportions

The hourly proportions are calculated here, because they not only provide us with an overall understanding of the distribution of the hourly proportions but also can be used to estimate hourly volumes for a specific highway location.

5.2 METHODOLOGY

5.2.1 Hour Groups

Here the 24 hours are grouped as follows:

- 1 - 0:00 to 5:00 (early morning),
- 2 - 5:00 to 9:00 (a.m. peak),
- 3 - 9:00 to 15:00 (mid-day),
- 4 - 15:00 to 19:00 (p.m. peak),

5 - 19:00 to 0:00 (evening).

This grouping is produced for the following reasons. First, it is believed that hourly volumes are more likely needed in these time periods (compare to others), for models in air quality estimation, vehicle crash prediction and transportation planning. Also, in most cases, the hourly proportions averaged within each of these time periods, especially early morning, mid-day and evening time periods, are considered adequate. In addition, our studies showed that the hourly proportions in general exhibit distinct different patterns over these time periods.

5.2.2 Month and Weekday Grouping Selection

Now we face the issue of identifying month and weekday groupings for each hour group. This identification carried out here is based on the month and weekday groupings produced in Chapter 4. Specifically, among the groupings for the hours in an hour group, the one with the smallest number of month/weekday groups is selected for that hour group.

Tables 5.1 and 5.2 show respectively the month and weekday groupings selected for each hour group for Station 9027-3. Note that the groupings from which these are selected were given in Figures 4.6 and 4.7 in Chapter 4. For example, for the p.m. peak hour group (*i.e.*, Hour Group 4), month grouping (1-1) is selected, because this grouping has the smallest number of month groups (as reflected by the first number in the grouping index) among the four hours in that group (highlighted in Figure 4.6).

Table 5-1 Month Groupings for All Hours in the Day for Station 9027-3

Hour Group	Month												Grouping Index
	1	2	3	4	5	6	7	8	9	10	11	12	
1	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
2	1	1	1	1	1	2	2	2	1	1	1	1	2 - 24
3	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
4	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1
5	1	1	1	1	1	1	1	1	1	1	1	1	1 - 1

Table 5-2 Weekday Groupings for All Hours in the Day for Station 9027-3

Hour Group	Weekday				Grouping Index
	2	3	4	5	
1	1	1	1	1	1 - 1
2	1	1	1	1	1 - 1
3	1	1	1	1	1 - 1
4	1	2	2	2	2 - 3
5	1	2	2	2	2 - 3

This selection results in a maximum level of grouping of months and weekdays, at the expense of the accuracy of the hourly proportion models when these groupings are used. To determine whether or not these groupings are acceptable in terms of model accuracy, we need to estimate hourly proportion models with the weekday and month groupings as independent variables, and further to examine the accuracy of these models.

The hourly proportion models used here and the hourly proportions calculated based on these models are discussed in the next sections. The procedure used to investigate the accuracy of these models is described a little later.

5.2.3 Hourly Proportion Models

For each hour group, the hourly proportion models used here are

$$Q_{ijm} = \mathbf{p} + M_i + D_j + \mathbf{e}_{m(ij)} \quad (5.1)$$

for weekday, and

$$Q_{im} = \mathbf{p} + M_i + \mathbf{e}_{m(i)} \quad (5.2)$$

for Friday, Saturday and Sunday, individually. In the models, Q_{ijm} and Q_{im} denote the Logit transformed observed hourly proportions for the m^{th} day in month group i and weekday group j , and month group i only, respectively; $\mathbf{p}$ is the unknown grand mean of the data estimated by the procedure; and $\mathbf{e}_{m(ij)}$ and $\mathbf{e}_{m(i)}$ are the random error components, which are assumed to be normally and independently distributed with zero mean and constant but unknown variance. Note that these models are similar to those given in Equations 3.14 and 3.15 in Chapter 3; the only difference is that these models are associated with each hour groups instead of each hour.

Similar to the model estimation discussed in Chapter 3, here for each hour group, model parameters are produced separately for weekday and weekend. For example, Table 5.3 gives the model parameters by hour groups on weekdays for ATR station 9027-3, including the intercept $\hat{\mathbf{p}}$, the parameters for weekday groups, $\hat{D}_j$, and the parameters for month groups, $\hat{M}_i$. Comparing these model parameters with those given in Tables 3.4a and 3.4b shows that the hourly proportion models are significantly simplified using the groupings of month and weekday.

Table 5-3 Model Parameters by Hour Group for Weekdays for Station 9027-3

Hour Group	Month Group			Weekday Group	
	Intercept	1	2	1	2
1	-4.2827	.0000	-	.0000	-
2	-3.3110	.0975	.0000	.0000	-
3	-2.7937	.0000	-	.0000	-
4	-2.6155	.0000	-	-.1188	.0000
5	-3.4647	.0000	-	-.1248	.0000

5.2.4 Hourly Proportion Calculation

These model parameters can be used to estimate the hourly proportion for a specific hour, month and weekday group for the study location. For example, for the a.m. peak hour group (*i.e.*, 5:00am through 9:00am), the first month group (including January through May and September through December) and the first weekday group (including all four weekdays), an estimate of the Logit transformation of the hourly proportion for Station 9027-3 can be calculated as

$$\hat{Q}_{ij} = -3.3110 + 0.0975 + 0.0000 = -3.2125 \quad (5.3)$$

where -3.3110 , 0.0975 and 0.0000 are respectively the intercept, month group parameter and weekday group parameter highlighted in Table 5.3.

The corresponding hourly proportion can then be estimated using the reverse of the Logit transformation as

$$\hat{P}_{ij} = \frac{e^{-3.2125}}{e^{-3.2125} + 1} = 0.0387 \quad (5.4)$$

Table 5.4 gives the estimated hourly proportions by hour and day of week groups for each month group. Note that in Table 5.4 day of week group numbers 5, 6 and 7 are assigned to Friday, Saturday and Sunday, respectively, and hourly proportions are calculated for each month group separately for Friday, Saturday and Sunday. By comparing this table with Table 3.5, it is shown that the grouping of month and weekday significantly reduces the number of hourly proportion needed. In fact, the number of needed hourly proportions reduces from 2,106 to 31 in this case.

5.2.5 Accuracy Checking

To determine whether or not the hourly proportions and, in turn, the month and weekday groupings selected are acceptable in terms of accuracy, summary statistics of the model residuals, such as root mean square error (RMSE) and mean absolute percent

Table 5-4 Estimated Hourly Proportions by Hour Group for Station 9027-3

Hour Group	Day of Week Group	Month Group	
		1	2
1	1	0.0136	-
	5	0.0095	-
	6	0.0094	-
	7	0.0054	-
2	1	0.0387	0.0352
	5	0.0243	-
	6	0.0228	0.0283
	7	0.0091	0.0117
3	1	0.0577	-
	5	0.0524	-
	6	0.0776	-
	7	0.0589	0.0588
4	1	0.0610	-
	2	0.0681	-
	5	0.0791	0.0745
	6	0.0606	0.0540
	7	0.0824	0.0751
5	1	0.0269	-
	2	0.0303	-
	5	0.0375	0.0416
	6	0.0257	-
	7	0.0414	0.0458

error (MAPE), are examined here. These summary statistics measure how close the estimated hourly proportions are to the observed hourly proportions; smaller values indicate that they are close to each other.

If we denote the residuals by e_i , and the observed hourly proportions by z_i , the RMSE and MAPE can be expressed as (Abraham and Ledolter, 1983)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}} \quad (5.5)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{e_i}{z_i} \right| \quad (5.6)$$

where n is the number of cases in the comparison.

For example, Table 5.5 shows the RMSE and MAPE for Station 9027-3. These summary statistics are calculated by hour group, and separately for weekday (corresponding to the model in Equation 5.1) and, Friday, Saturday and Sunday (corresponding to the models in Equation 5.2). As can be seen, the RMSEs and MAPEs are both quite small, which means that the estimated hourly proportions based on the groupings are acceptable with regard to accuracy, which in turn indicates that the groupings are appropriate for Station 9027-3.

5.3 RESULTS

The final results of the work in this chapter include the month and weekday groupings and the estimated hourly proportions for each month and day of week group, for each hour group and each study station. In addition, summary statistics (*i.e.*, the RMSEs and MAPEs) are produced for each study station for checking whether or not the groupings and the hourly proportions estimated are acceptable in terms of accuracy.

Table 5.6 gives the final month groupings on weekdays and the final weekday groupings by hour group and study station. Tables 5.7 and 5.8 further show these groupings for

Table 5-5 RMSE and MAPE by Hour Group for Station 9027-3

Hour Group	Day of Week Group	RMSE	MAPE
1	Weekday	0.0014	8.0393
	Friday	0.0007	6.0223
	Saturday	0.0005	4.0331
	Sunday	0.0003	3.7309
2	Weekday	0.0029	5.7451
	Friday	0.0014	4.3442
	Saturday	0.0027	8.4802
	Sunday	0.0012	8.4379
3	Weekday	0.0022	2.9933
	Friday	0.0020	2.8918
	Saturday	0.0021	1.9495
	Sunday	0.0030	3.8611
4	Weekday	0.0017	2.0341
	Friday	0.0019	1.9465
	Saturday	0.0015	2.0405
	Sunday	0.0021	2.1543
	Weekday	0.0017	4.5338
	Friday	0.0018	3.3940
	Saturday	0.0015	4.1906
	Sunday	0.0022	3.8638
Average		0.0021	4.6700

each individual hour in the a.m. and p.m. peak periods. These latter groupings are given here, because it is considered that the hourly proportions for the hours in peak periods are usually more important than those for the other hours in the day. Note that in these tables the groupings are indicated by the grouping indexes. Table 5.9 shows the month groupings corresponding to these grouping indexes. (Note that an entire list of the weekday grouping indexes has been given in Table 4.3).

Figures 5.1 through 5.7 show the hourly proportions estimated for the hour, month and day of week groups for Station 9027-3. In these figures, the hourly proportions are plotted for each day of the week by hour and month rather than hour and month groups. It is believed that plotting the hourly proportions this way makes it easier to understand the groupings and to visualize the distributions of the estimated hourly proportions.

Figures 5.1 through 5.7 give a clear depiction of the effects of the grouping. This may be better explained by comparing these figures with those given in Figures 3.8 through 3.14 in Chapter 3, which show the hourly proportions before the grouping for the same study station. As can be seen, the grouping significantly reduces the number of hourly proportions needed, which also means that hourly proportion models can be significantly simplified, when these groupings are used. (Note that Figures 5.1 through 5.7 give the hourly proportions for Station 9027-3 only. Hourly proportions for other stations are also produced and plotted; for brevity, they are not given here).

Table 5.10 shows the RMSEs and MAPEs of the hourly proportion models corresponding to the month and weekday groupings. They are the averages for all hour groups at each station. As can be seen, the RMSEs and MAPEs are both quite small. This means that the estimated hourly proportions based on the groupings are acceptable with regard to accuracy, which in turn indicates that the groupings are appropriate.

5.4 DISCUSSION

An examination of the groupings in Table 5.6, particularly the numbers of month and weekday groups indicated by the first numbers of the grouping indexes, reveals that the months and weekdays can be grouped to a considerable level. Specifically, the months can be grouped into less than three groups for nearly 80 percent of the groupings, and the weekdays can be grouped into less than two groups for almost all groupings. This means that these groupings can significantly simplify an hourly proportion model size when they are used.

Table 5.6 also reveals that the groupings vary by hour group, with the a.m. and p.m. peak hour groups having larger numbers of groups. For the early morning hour group, the months and weekdays can be put into one group for all but only one of the study stations. This means that at this time period the month and weekday generally do not have significantly different effects on the hourly proportions, and thus, they do not need to be considered in an hourly proportion model. Table 5.6 also reveals that the groupings vary by study station. Specifically, the stations with small average daily volumes tend to have fewer groups; for example, the stations with smaller annual average daily volumes (9012-1, 9012-5, 9045-3 and 9045-7) have fewer numbers of groups.

Table 5-6 Month and Weekday Groupings for All Hour Groups by Station

Station ID	Early Morning		A.M. Peak		Mid-Day		P.M. Peak		Evening	
	Month	Weekday	Month	Weekday	Month	Weekday	Month	Weekday	Month	Weekday
9007-1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	3 - 57	1 - 1	2 - 24	1 - 1
9007-5	1 - 1	1 - 1	2 - 33	2 - 1	2 - 17	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1
9012-1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1
9012-5	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	2 - 24	1 - 1	1 - 1	1 - 1
9014-1	1 - 1	1 - 1	4 - 99	2 - 3	3 - 5	1 - 1	3 - 66	1 - 1	3 - 108	3 - 2
9014-5	1 - 1	1 - 1	4 - 156	2 - 3	2 - 10	1 - 1	3 - 101	1 - 1	2 - 53	3 - 2
9024-3	1 - 1	1 - 1	3 - 70	2 - 3	2 - 43	1 - 1	4 - 240	1 - 1	4 - 40	2 - 3
9024-7	1 - 1	1 - 1	2 - 17	2 - 3	3 - 182	1 - 1	2 - 24	1 - 1	2 - 24	2 - 3
9026-3	1 - 1	1 - 1	2 - 35	2 - 3	3 - 107	2 - 3	5 - 168	1 - 1	2 - 24	3 - 2
9026-7	1 - 1	1 - 1	6 - 140	2 - 3	3 - 74	2 - 3	2 - 53	2 - 3	2 - 24	2 - 3
9027-3	1 - 1	1 - 1	2 - 24	1 - 1	1 - 1	1 - 1	1 - 1	2 - 3	1 - 1	2 - 3
9027-7	1 - 1	1 - 1	2 - 17	2 - 3	2 - 17	1 - 1	1 - 1	2 - 3	1 - 1	1 - 1
9030-1	2 - 35	1 - 1	3 - 65	2 - 3	2 - 35	1 - 1	4 - 171	1 - 1	2 - 35	3 - 3
9030-5	1 - 1	1 - 1	4 - 65	2 - 1	4 - 20	1 - 1	1 - 1	1 - 1	6 - 125	3 - 1
9032-1	1 - 1	1 - 1	3 - 30	1 - 1	3 - 11	1 - 1	2 - 28	1 - 1	4 - 254	2 - 3
9032-5	1 - 1	1 - 1	2 - 11	2 - 3	2 - 17	1 - 1	2 - 36	1 - 1	3 - 45	2 - 3
9033-1	1 - 1	1 - 1	3 - 29	1 - 1	3 - 30	1 - 1	3 - 44	1 - 1	1 - 1	2 - 3
9033-5	1 - 1	1 - 1	2 - 17	2 - 3	2 - 17	1 - 1	2 - 25	2 - 3	2 - 17	2 - 3
9044-1	1 - 1	1 - 1	3 - 65	1 - 1	3 - 57	1 - 1	7 - 51	1 - 1	2 - 24	2 - 3
9044-5	1 - 1	1 - 1	7 - 40	2 - 3	3 - 36	1 - 1	4 - 155	2 - 3	2 - 24	2 - 3
9045-3	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	2 - 24	1 - 1	1 - 1	1 - 1
9045-7	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	1 - 1	2 - 24	1 - 1	2 - 24	1 - 1
9049-3	1 - 1	1 - 1	5 - 146	2 - 3	4 - 75	1 - 1	2 - 35	1 - 1	3 - 162	3 - 2
9049-7	1 - 1	1 - 1	3 - 70	2 - 3	2 - 43	1 - 1	3 - 52	2 - 3	3 - 126	3 - 2
9053-1	1 - 1	1 - 1	2 - 24	1 - 1	2 - 1	1 - 1	4 - 81	2 - 1	2 - 8	2 - 3
9053-5	1 - 1	1 - 1	4 - 45	2 - 3	4 - 72	1 - 1	3 - 133	1 - 1	5 - 146	2 - 3
9054-3	1 - 1	1 - 1	1 - 1	2 - 3	2 - 32	1 - 1	2 - 25	2 - 3	3 - 73	2 - 3
9054-7	1 - 1	1 - 1	2 - 6	2 - 3	2 - 17	1 - 1	2 - 24	1 - 1	2 - 24	1 - 1
9055-1	1 - 1	1 - 1	2 - 35	1 - 1	1 - 1	1 - 1	2 - 35	1 - 1	2 - 41	2 - 1
9055-5	1 - 1	1 - 1	2 - 33	2 - 3	2 - 24	1 - 1	3 - 184	1 - 1	2 - 24	1 - 1

Table 5-7 Month and Weekday Groupings for Morning Peak Hours by Station

Station ID	5:00 to 6:00		6:00 to 7:00		7:00 to 8:00		8:00 to 9:00	
	Month	Weekday	Month	Weekday	Month	Weekday	Month	Weekday
9007-1	1 - 1	1 - 1	3 - 44	2 - 3	6 - 125	2 - 3	5 - 76	2 - 3
9007-5	2 - 33	2 - 1	5 - 146	3 - 2	7 - 52	3 - 2	4 - 156	2 - 1
9012-1	1 - 1	1 - 1	4 - 20	2 - 3	5 - 146	2 - 3	2 - 32	1 - 1
9012-5	1 - 1	1 - 1	1 - 1	2 - 3	4 - 254	2 - 3	2 - 51	1 - 1
9014-1	4 - 99	2 - 3	7 - 40	3 - 1	7 - 105	3 - 1	6 - 286	3 - 1
9014-5	5 - 64	2 - 3	4 - 222	3 - 3	5 - 142	3 - 1	4 - 156	2 - 1
9024-3	3 - 70	2 - 3	8 - 85	3 - 2	7 - 430	3 - 2	5 - 166	2 - 1
9024-7	2 - 17	2 - 3	7 - 221	3 - 2	7 - 435	2 - 1	5 - 352	2 - 1
9026-3	2 - 35	2 - 3	5 - 351	3 - 2	4 - 89	3 - 2	3 - 129	3 - 2
9026-7	6 - 140	2 - 3	7 - 40	2 - 1	9 - 69	3 - 1	7 - 374	3 - 1
9027-3	2 - 24	1 - 1	3 - 46	3 - 2	5 - 68	3 - 2	3 - 41	2 - 1
9027-7	2 - 17	2 - 3	5 - 67	3 - 2	4 - 88	3 - 2	4 - 154	2 - 1
9030-1	3 - 65	2 - 3	4 - 271	2 - 3	7 - 424	3 - 2	6 - 233	2 - 1
9030-5	4 - 65	3 - 2	5 - 483	3 - 1	6 - 108	2 - 1	6 - 108	2 - 1
9032-1	3 - 30	1 - 1	5 - 273	3 - 2	6 - 92	3 - 2	5 - 23	2 - 1
9032-5	2 - 11	2 - 3	8 - 85	3 - 2	8 - 226	3 - 1	6 - 308	2 - 1
9033-1	3 - 29	1 - 1	5 - 147	3 - 2	9 - 67	3 - 2	6 - 127	2 - 1
9033-5	2 - 17	2 - 3	6 - 6	3 - 2	10 - 1	3 - 2	8 - 229	2 - 3
9044-1	3 - 65	1 - 1	5 - 273	2 - 3	6 - 36	2 - 3	4 - 20	2 - 1
9044-5	7 - 40	2 - 3	9 - 2	3 - 2	11 - 2	2 - 1	11 - 12	2 - 1
9045-3	1 - 1	1 - 1	4 - 20	2 - 3	6 - 125	3 - 1	6 - 608	2 - 3
9045-7	1 - 1	1 - 1	3 - 65	2 - 3	5 - 76	2 - 3	5 - 146	1 - 1
9049-3	5 - 544	2 - 3	7 - 124	3 - 1	5 - 146	3 - 2	6 - 628	3 - 2
9049-7	3 - 70	2 - 3	6 - 272	3 - 2	4 - 20	3 - 3	3 - 57	3 - 1
9053-1	2 - 24	1 - 1	7 - 490	3 - 2	7 - 51	3 - 1	5 - 96	3 - 1
9053-5	6 - 91	2 - 3	5 - 8	3 - 2	4 - 45	2 - 3	5 - 146	2 - 1
9054-3	1 - 1	2 - 3	4 - 239	3 - 2	8 - 225	3 - 2	5 - 191	2 - 1
9054-7	2 - 6	2 - 3	6 - 227	3 - 2	8 - 42	3 - 2	7 - 159	3 - 2
9055-1	2 - 35	1 - 1	4 - 215	3 - 2	6 - 10	3 - 2	7 - 46	2 - 1
9055-5	2 - 33	2 - 3	7 - 151	3 - 2	8 - 216	3 - 2	7 - 171	2 - 1

Table 5-8 Month and Weekday Groupings for Afternoon Peak Hours by Station

	15:00 to 16:00		16:00 to 17:00		17:00 to 18:00		18:00 to 19:00	
Station ID	Month	Weekday	Month	Weekday	Month	Weekday	Month	Weekday
9007-1	4 - 272	2 - 3	6 - 605	2 - 3	3 - 57	1 - 1	4 - 394	2 - 3
9007-5	3 - 52	1 - 1	2 - 33	2 - 1	1 - 1	2 - 1	4 - 127	2 - 3
9012-1	2 - 42	1 - 1	1 - 1	2 - 3	2 - 32	1 - 1	2 - 32	1 - 1
9012-5	3 - 73	1 - 1	2 - 24	1 - 1	3 - 45	1 - 1	2 - 24	2 - 3
9014-1	4 - 152	2 - 3	4 - 20	1 - 1	3 - 57	1 - 1	3 - 66	3 - 2
9014-5	5 - 148	1 - 1	3 - 101	2 - 1	4 - 131	3 - 2	4 - 83	2 - 3
9024-3	4 - 237	1 - 1	7 - 492	3 - 2	6 - 162	3 - 2	4 - 240	2 - 3
9024-7	3 - 48	1 - 1	2 - 24	2 - 3	2 - 24	1 - 1	4 - 86	2 - 3
9026-3	6 - 255	1 - 1	8 - 217	3 - 1	6 - 653	3 - 1	5 - 168	2 - 3
9026-7	6 - 31	2 - 3	3 - 99	2 - 3	2 - 53	3 - 2	6 - 636	3 - 2
9027-3	1 - 1	2 - 3	2 - 24	2 - 3	2 - 24	3 - 3	2 - 32	3 - 2
9027-7	1 - 1	2 - 3	1 - 1	2 - 3	2 - 52	2 - 3	2 - 41	3 - 2
9030-1	5 - 561	2 - 3	5 - 168	3 - 1	4 - 171	3 - 1	5 - 81	1 - 1
9030-5	5 - 31	1 - 1	3 - 46	1 - 1	1 - 1	3 - 1	3 - 59	2 - 3
9032-1	5 - 188	1 - 1	7 - 460	2 - 1	8 - 217	2 - 1	2 - 28	1 - 1
9032-5	2 - 36	1 - 1	3 - 45	2 - 3	4 - 147	2 - 3	6 - 600	2 - 3
9033-1	3 - 44	1 - 1	5 - 315	2 - 3	4 - 155	2 - 3	4 - 127	2 - 3
9033-5	4 - 373	2 - 3	2 - 25	2 - 3	3 - 21	2 - 3	4 - 358	2 - 3
9044-1	7 - 457	1 - 1	8 - 217	2 - 1	8 - 150	2 - 3	7 - 51	3 - 2
9044-5	4 - 155	2 - 3	6 - 162	2 - 3	6 - 299	2 - 3	7 - 432	3 - 2
9045-3	3 - 45	1 - 1	2 - 24	1 - 1	2 - 27	1 - 1	4 - 349	2 - 3
9045-7	2 - 24	1 - 1	3 - 41	2 - 3	3 - 156	2 - 3	2 - 36	1 - 1
9049-3	3 - 44	1 - 1	2 - 35	1 - 1	4 - 167	2 - 3	6 - 758	3 - 1
9049-7	3 - 52	2 - 3	4 - 75	3 - 2	4 - 231	3 - 2	6 - 51	2 - 3
9053-1	6 - 117	2 - 1	5 - 146	3 - 2	5 - 146	3 - 2	4 - 81	3 - 2
9053-5	4 - 72	1 - 1	3 - 11	2 - 3	3 - 133	1 - 1	4 - 149	2 - 3
9054-3	2 - 25	2 - 1	4 - 153	2 - 3	6 - 165	2 - 3	4 - 93	2 - 3
9054-7	2 - 23	2 - 1	2 - 24	2 - 3	2 - 6	1 - 1	2 - 63	2 - 1
9055-1	2 - 17	1 - 1	6 - 92	2 - 1	4 - 131	2 - 1	2 - 35	2 - 1
9055-5	4 - 40	1 - 1	5 - 150	1 - 1	4 - 40	1 - 1	3 - 184	2 - 3

Table 5-9 Month Groupings by Grouping Index

<i>Group</i>	<i>Month</i>												<i>Group</i>	<i>Month</i>											
<i>Index</i>	1	2	3	4	5	6	7	8	9	10	11	12	<i>Index</i>	1	2	3	4	5	6	7	8	9	10	11	12
1-1	1	1	1	1	1	1	1	1	1	1	1	1	3-101	1	1	1	2	2	2	3	3	1	1	1	1
2-1	1	1	1	1	1	1	1	1	1	1	1	2	3-107	1	1	1	2	2	3	3	1	1	1	1	1
2-6	1	1	1	1	1	1	1	1	1	2	2	2	3-108	1	1	1	2	2	3	3	3	1	1	1	1
2-8	1	1	1	1	1	1	1	1	2	2	1	1	3-126	1	1	2	2	2	2	2	2	3	3	3	3
2-10	1	1	1	1	1	1	1	1	2	2	2	2	3-129	1	1	2	2	2	2	2	2	3	3	3	1
2-11	1	1	1	1	1	1	1	2	1	1	1	1	3-133	1	1	2	2	2	2	2	3	3	3	1	1
2-17	1	1	1	1	1	1	2	2	1	1	1	1	3-156	1	1	2	2	3	3	3	3	3	3	3	3
2-23	1	1	1	1	1	2	2	1	1	1	1	1	3-162	1	1	2	3	3	3	3	3	3	1	1	1
2-24	1	1	1	1	1	2	2	2	1	1	1	1	3-182	1	2	2	2	2	2	3	3	1	1	1	1
2-25	1	1	1	1	1	2	2	2	2	1	1	1	3-184	1	2	2	2	2	2	3	3	3	3	1	1
2-27	1	1	1	1	1	2	2	2	2	2	2	1	4-20	1	1	1	1	1	1	2	2	3	3	3	4
2-28	1	1	1	1	1	2	2	2	2	2	2	2	4-40	1	1	1	1	1	2	2	2	3	3	3	4
2-32	1	1	1	1	2	2	2	2	1	1	1	1	4-45	1	1	1	1	1	2	2	2	3	4	4	4
2-33	1	1	1	1	2	2	2	2	2	1	1	1	4-65	1	1	1	1	1	2	3	3	4	4	4	4
2-35	1	1	1	1	2	2	2	2	2	2	2	1	4-72	1	1	1	1	2	2	2	2	2	3	3	4
2-36	1	1	1	1	2	2	2	2	2	2	2	2	4-75	1	1	1	1	2	2	2	2	3	3	3	4
2-41	1	1	1	2	2	2	2	2	1	1	1	1	4-81	1	1	1	1	2	2	2	3	3	3	3	4
2-42	1	1	1	2	2	2	2	2	2	1	1	1	4-83	1	1	1	1	2	2	2	3	3	3	4	4
2-43	1	1	1	2	2	2	2	2	2	2	1	1	4-86	1	1	1	1	2	2	2	3	3	4	4	4
2-51	1	1	2	2	2	2	2	2	1	1	1	1	4-88	1	1	1	1	2	2	2	3	4	4	1	1
2-52	1	1	2	2	2	2	2	2	2	1	1	1	4-89	1	1	1	1	2	2	2	3	4	4	4	1
2-53	1	1	2	2	2	2	2	2	2	2	1	1	4-93	1	1	1	1	2	2	3	3	3	3	4	4
2-63	1	2	2	2	2	2	2	2	2	1	1	1	4-99	1	1	1	1	2	2	3	3	4	4	4	1
3-5	1	1	1	1	1	1	1	1	2	2	2	3	4-127	1	1	1	2	2	2	2	2	2	2	3	4
3-11	1	1	1	1	1	1	1	2	2	2	2	3	4-131	1	1	1	2	2	2	2	2	3	3	3	4
3-21	1	1	1	1	1	1	2	2	2	2	2	3	4-147	1	1	1	2	2	2	3	3	3	3	3	4
3-29	1	1	1	1	1	1	2	2	3	3	3	1	4-149	1	1	1	2	2	2	3	3	3	3	4	4
3-30	1	1	1	1	1	1	2	2	3	3	3	3	4-152	1	1	1	2	2	2	3	3	3	4	4	4
3-36	1	1	1	1	1	2	2	2	2	2	2	3	4-153	1	1	1	2	2	2	3	3	4	1	1	1
3-41	1	1	1	1	1	2	2	2	2	3	3	3	4-154	1	1	1	2	2	2	3	3	4	4	1	1
3-44	1	1	1	1	1	2	2	2	3	3	3	1	4-155	1	1	1	2	2	2	3	3	4	4	4	1
3-45	1	1	1	1	1	2	2	2	3	3	3	3	4-156	1	1	1	2	2	2	3	3	4	4	4	4
3-46	1	1	1	1	1	2	2	3	1	1	1	1	4-167	1	1	1	2	2	3	3	3	3	4	4	4
3-48	1	1	1	1	1	2	2	3	3	3	1	1	4-171	1	1	1	2	2	3	3	3	4	4	4	4
3-52	1	1	1	1	1	2	3	3	1	1	1	1	4-215	1	1	2	2	2	2	2	2	3	3	3	4
3-57	1	1	1	1	2	2	2	2	2	2	2	3	4-222	1	1	2	2	2	2	2	3	3	3	4	1
3-59	1	1	1	1	2	2	2	2	2	2	3	3	4-231	1	1	2	2	2	2	3	3	3	3	3	4
3-65	1	1	1	1	2	2	2	2	3	3	3	1	4-237	1	1	2	2	2	2	3	3	4	1	1	1
3-66	1	1	1	1	2	2	2	2	3	3	3	3	4-239	1	1	2	2	2	2	3	3	4	4	4	1
3-70	1	1	1	1	2	2	2	3	3	3	3	1	4-240	1	1	2	2	2	2	3	3	4	4	4	4
3-73	1	1	1	1	2	2	3	3	1	1	1	1	4-254	1	1	2	2	2	3	3	3	4	4	4	1
3-74	1	1	1	1	2	2	3	3	3	1	1	1	4-271	1	1	2	2	3	3	3	3	3	4	4	1
3-99	1	1	1	2	2	2	2	3	3	3	3	3	4-272	1	1	2	2	3	3	3	3	3	4	4	4

Table 5-9 Month Groupings by Grouping Index (Cont'd)

<i>Group</i>	<i>Month</i>												<i>Group</i>	<i>Month</i>											
<i>Index</i>	1	2	3	4	5	6	7	8	9	10	11	12	<i>Index</i>	1	2	3	4	5	6	7	8	9	10	11	12
4-349	1	2	2	2	2	2	2	3	4	4	4	1	6-255	1	1	2	2	2	3	4	4	5	6	6	1
4-358	1	2	2	2	2	2	3	3	4	4	1	1	6-272	1	1	2	2	3	3	3	4	4	4	5	6
4-373	1	2	2	2	2	3	3	3	4	4	1	1	6-286	1	1	2	2	3	3	4	4	5	5	5	6
4-394	1	2	2	2	3	3	3	3	4	4	1	1	6-299	1	1	2	2	3	3	4	5	6	6	1	1
5-8	1	1	1	1	1	1	2	2	3	3	4	5	6-308	1	1	2	2	3	4	4	4	5	5	6	6
5-23	1	1	1	1	1	2	2	2	3	3	4	5	6-600	1	2	2	3	3	3	3	4	5	5	6	6
5-31	1	1	1	1	1	2	2	3	4	4	4	5	6-605	1	2	2	3	3	3	4	4	4	5	5	6
5-64	1	1	1	1	2	2	2	3	3	4	5	1	6-608	1	2	2	3	3	3	4	4	5	5	5	6
5-67	1	1	1	1	2	2	2	3	4	4	5	1	6-628	1	2	2	3	3	4	4	4	5	5	5	6
5-68	1	1	1	1	2	2	2	3	4	4	5	5	6-636	1	2	2	3	3	4	4	5	5	5	6	6
5-76	1	1	1	1	2	2	3	3	4	4	4	5	6-653	1	2	2	3	3	4	5	5	6	6	6	6
5-81	1	1	1	1	2	2	3	3	4	5	5	5	6-758	1	2	3	3	3	4	4	4	5	6	6	1
5-96	1	1	1	1	2	3	3	3	4	4	4	5	7-40	1	1	1	2	2	2	3	4	5	5	6	7
5-142	1	1	1	2	2	2	3	3	3	3	4	5	7-46	1	1	1	2	2	3	3	4	5	5	6	7
5-146	1	1	1	2	2	2	3	3	4	4	4	5	7-51	1	1	1	2	2	3	4	4	5	5	6	7
5-147	1	1	1	2	2	2	3	3	4	4	5	1	7-52	1	1	1	2	2	3	4	4	5	6	6	7
5-148	1	1	1	2	2	2	3	3	4	4	5	5	7-105	1	1	1	2	3	4	5	5	6	6	6	7
5-150	1	1	1	2	2	2	3	3	4	5	5	1	7-124	1	1	2	2	2	2	3	4	5	5	6	7
5-166	1	1	1	2	2	3	3	3	4	4	4	5	7-151	1	1	2	2	3	3	3	4	5	5	6	7
5-168	1	1	1	2	2	3	3	3	4	4	5	5	7-159	1	1	2	2	3	3	4	4	5	6	7	7
5-188	1	1	1	2	2	3	4	4	5	1	1	1	7-171	1	1	2	2	3	4	4	4	5	5	6	7
5-191	1	1	1	2	2	3	4	4	5	5	5	5	7-221	1	1	2	3	3	3	4	5	6	6	7	1
5-273	1	1	2	2	2	2	3	3	4	4	5	1	7-374	1	2	2	2	3	3	4	5	6	6	6	7
5-315	1	1	2	2	2	3	4	4	5	5	1	1	7-424	1	2	2	3	3	3	4	4	5	6	7	1
5-351	1	1	2	2	3	3	4	4	5	5	5	1	7-430	1	2	2	3	3	3	4	5	6	6	6	7
5-352	1	1	2	2	3	3	4	4	5	5	5	5	7-432	1	2	2	3	3	3	4	5	6	6	7	7
5-483	1	2	2	2	2	2	3	3	4	4	5	1	7-435	1	2	2	3	3	3	4	5	6	7	7	7
5-544	1	2	2	2	3	3	3	4	4	4	5	1	7-457	1	2	2	3	3	4	5	5	6	6	7	7
5-561	1	2	2	2	3	3	4	4	5	5	5	1	7-460	1	2	2	3	3	4	5	5	6	7	7	7
6-6	1	1	1	1	1	1	2	3	4	5	6	1	7-490	1	2	2	3	4	4	5	5	6	6	6	7
6-10	1	1	1	1	1	2	2	3	4	4	5	6	7-492	1	2	2	3	4	4	5	5	6	6	7	7
6-31	1	1	1	1	2	2	2	3	4	4	5	6	8-42	1	1	1	2	3	4	5	6	7	7	8	8
6-36	1	1	1	1	2	2	3	3	4	4	5	6	8-85	1	1	2	3	3	3	4	5	6	6	7	8
6-51	1	1	1	1	2	3	3	3	4	4	5	6	8-150	1	1	2	3	4	5	6	6	7	7	7	8
6-91	1	1	1	2	2	2	3	3	3	4	5	6	8-216	1	2	2	3	3	4	5	5	6	6	7	8
6-92	1	1	1	2	2	2	3	3	4	4	5	6	8-217	1	2	2	3	3	4	5	5	6	7	7	8
6-108	1	1	1	2	2	3	3	3	4	5	5	6	8-225	1	2	2	3	3	4	5	6	7	7	8	1
6-117	1	1	1	2	2	3	3	4	5	5	6	6	8-226	1	2	2	3	3	4	5	6	7	7	8	8
6-125	1	1	1	2	2	3	4	4	5	5	5	6	8-229	1	2	2	3	3	4	5	6	7	8	8	8
6-127	1	1	1	2	2	3	4	4	5	5	6	6	9-2	1	1	1	2	2	3	4	5	6	7	8	9
6-140	1	1	1	2	2	3	4	5	6	6	6	6	9-67	1	2	2	3	3	4	5	5	6	7	8	9
6-162	1	1	1	2	3	3	4	4	5	5	6	6	9-69	1	2	2	3	3	4	5	6	7	7	8	9
6-165	1	1	1	2	3	3	4	4	5	6	6	6	10-1	1	1	1	2	3	4	5	6	7	8	9	10
6-227	1	1	2	2	2	2	3	4	5	5	6	1	11-2	1	2	2	3	4	5	6	7	8	9	10	11

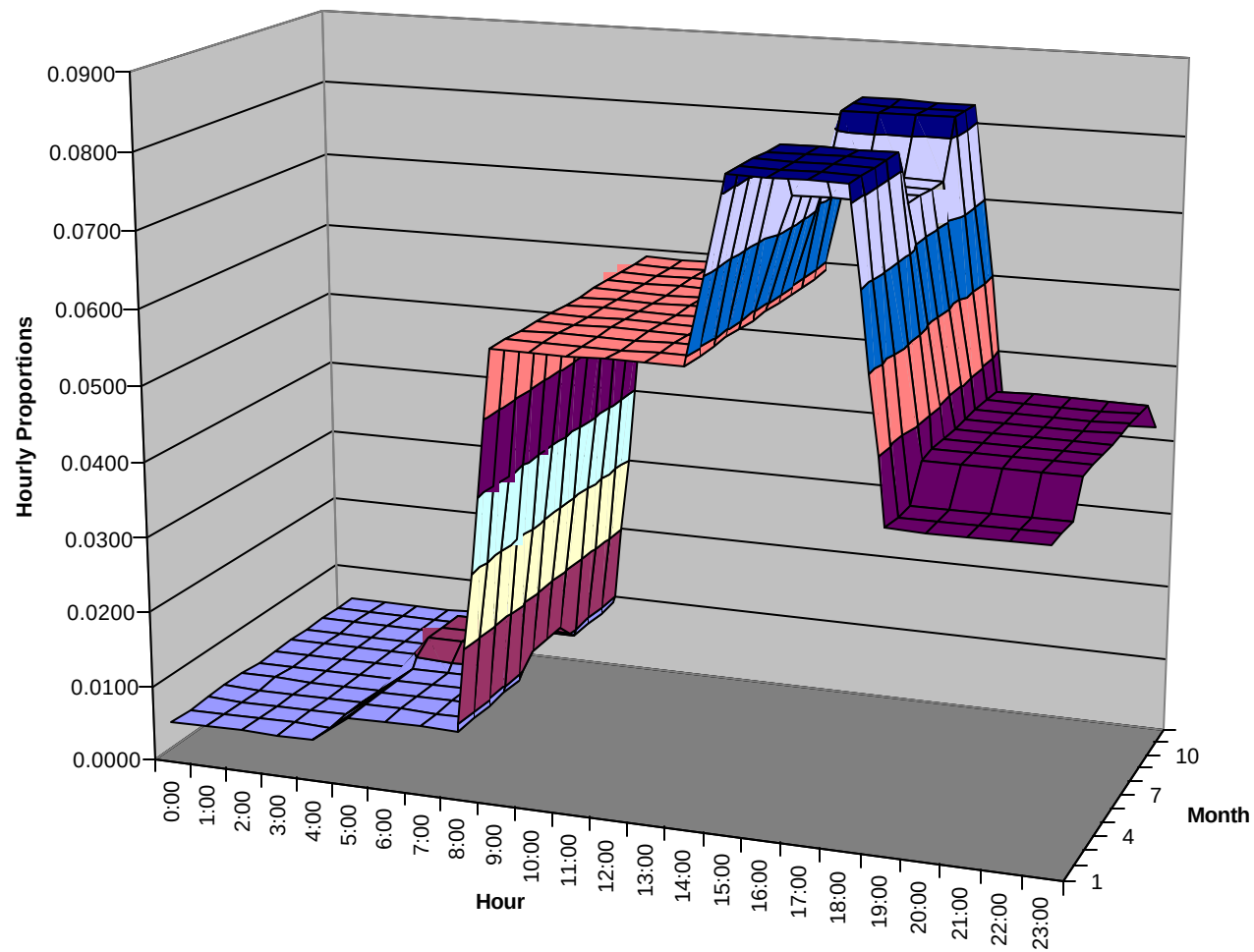


Figure 5-1 Hourly Proportions (after grouping) by Hour and Month for Sunday at Station 9027-3

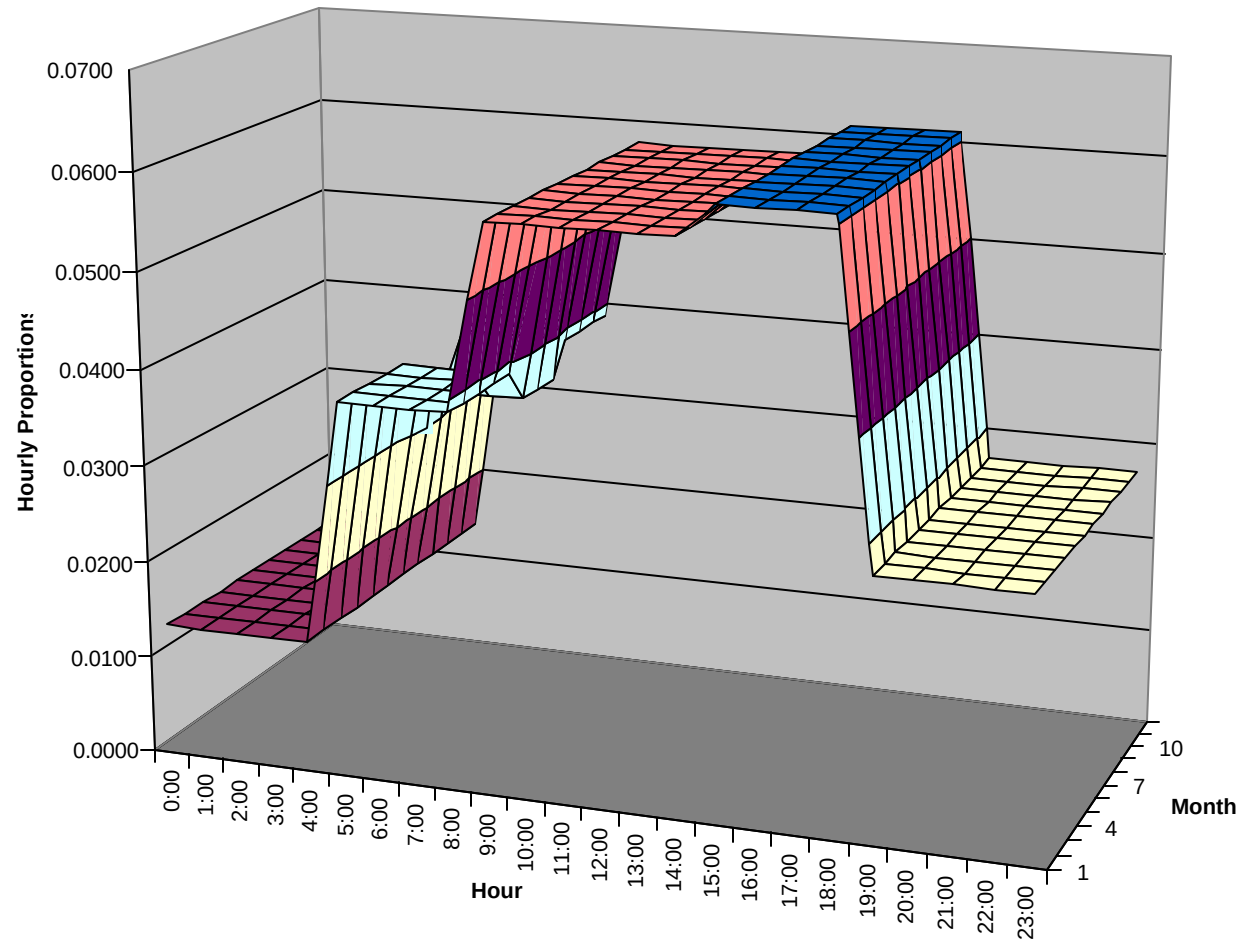


Figure 5-2 Hourly Proportions (after grouping) by Hour and Month for Monday at Station 9027-3

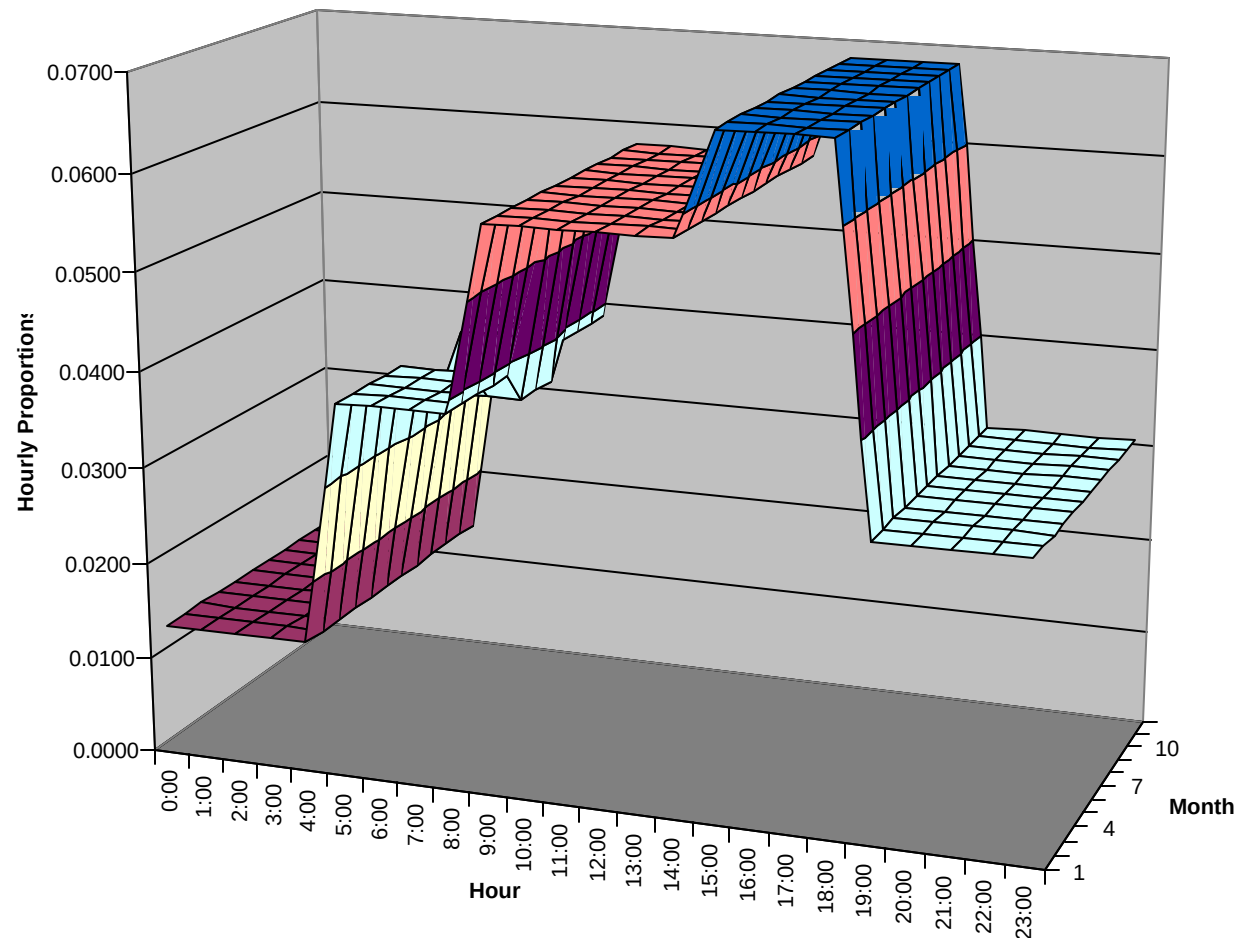


Figure 5-3 Hourly Proportions (after grouping) by Hour and Month for Tuesday at Station 9027-3

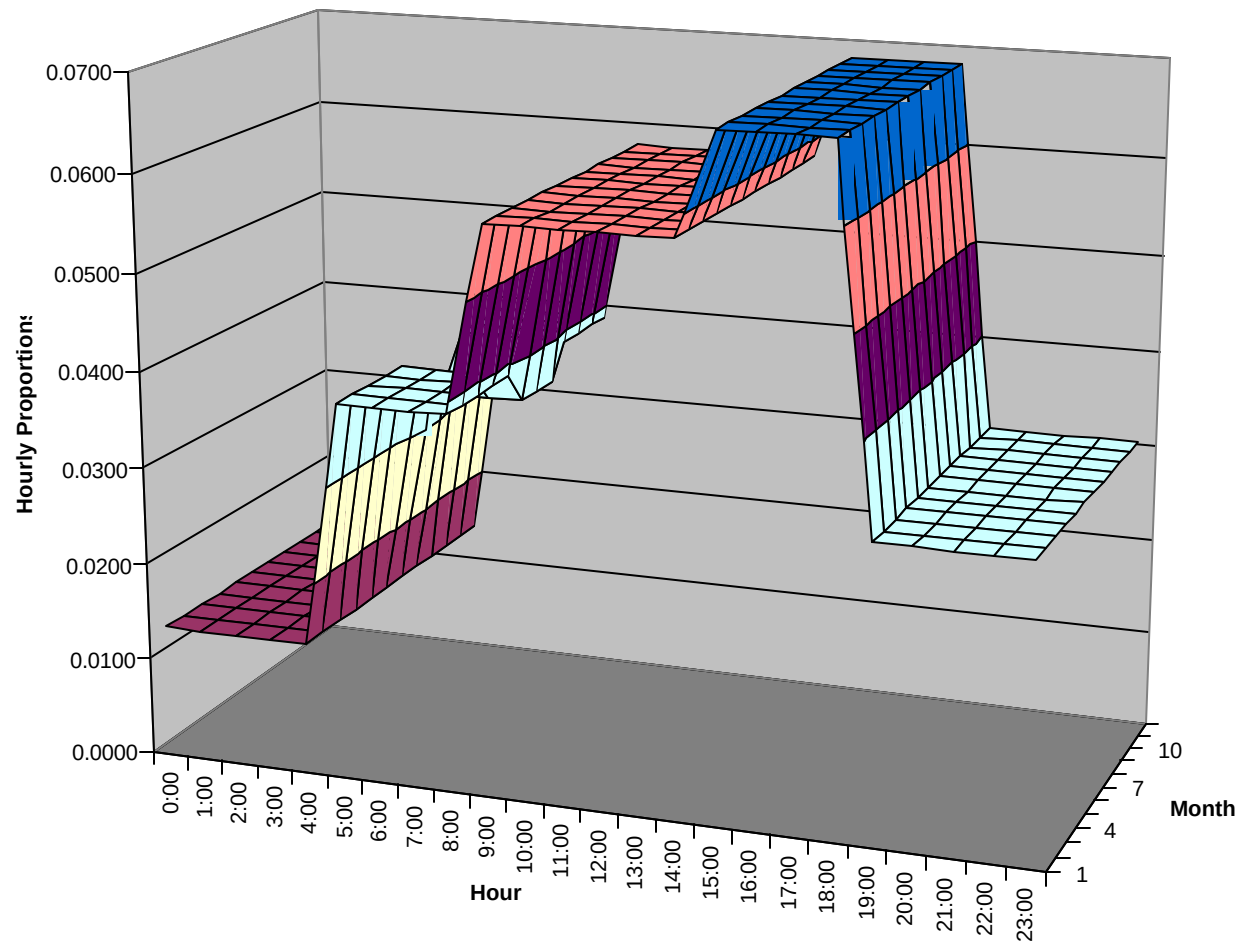


Figure 5-4 Hourly Proportions (after grouping) by Hour and Month for Wednesday at Station 9027-3

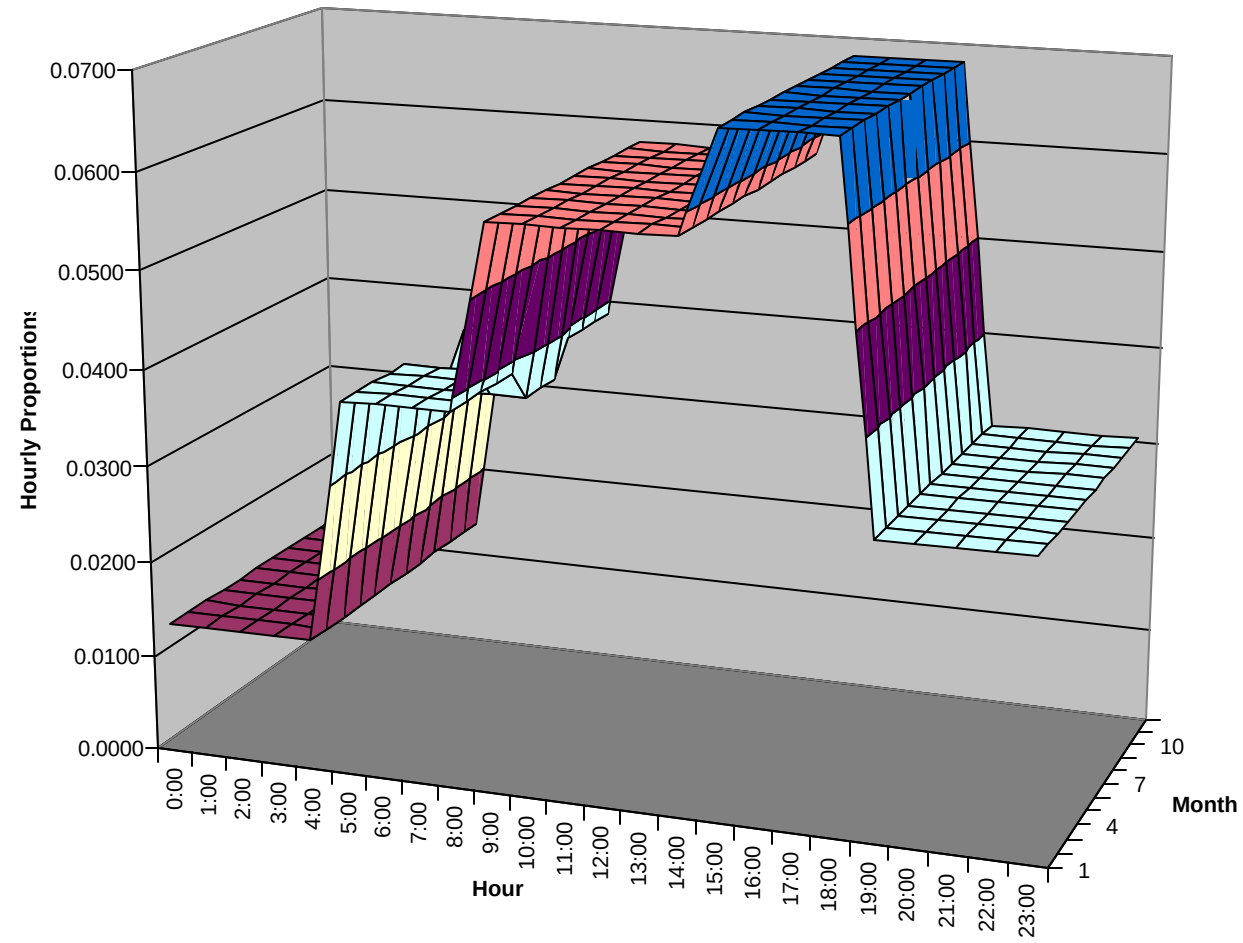


Figure 5-5 Hourly Proportions (after grouping) by Hour and Month for Thursday at Station 9027-3

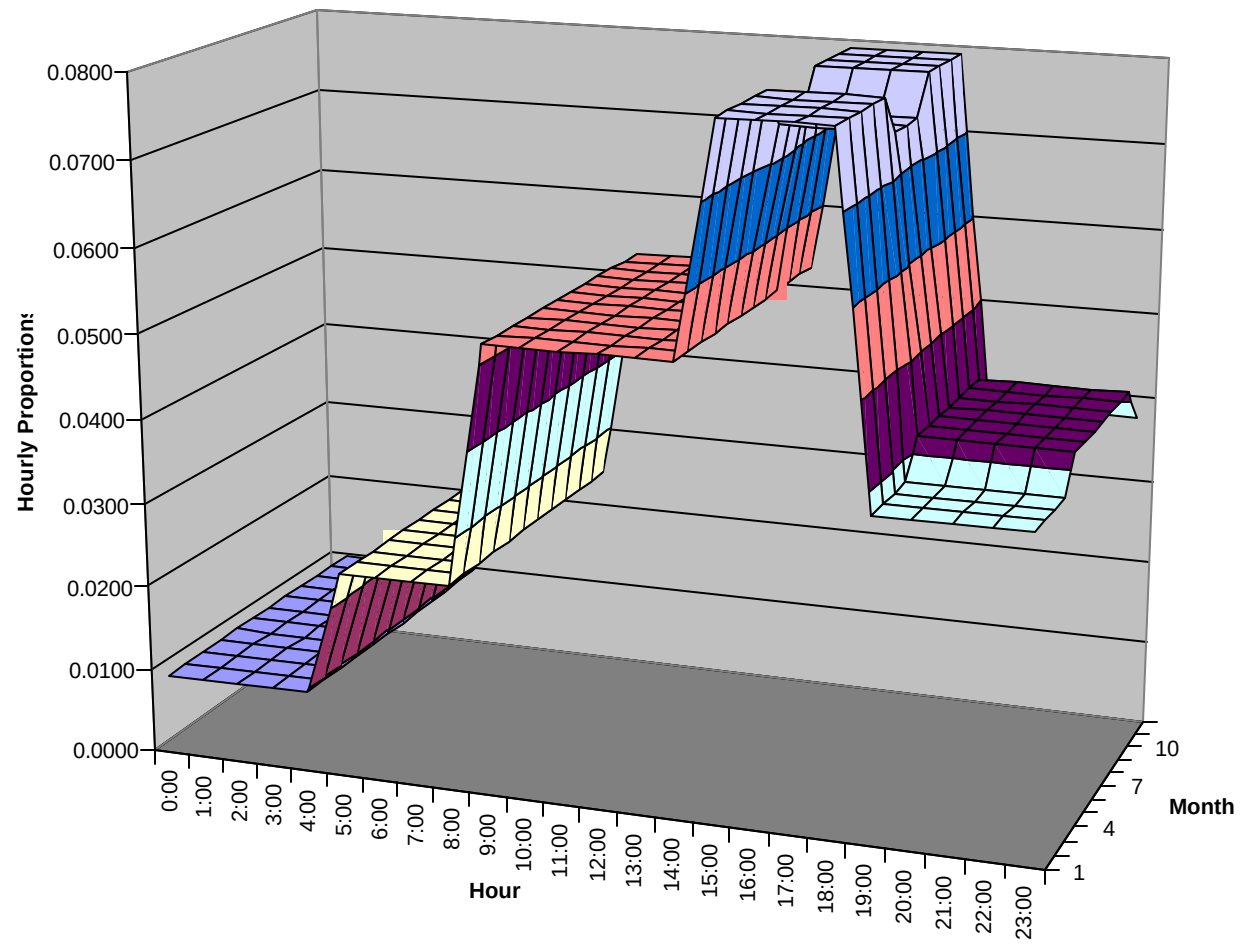


Figure 5-6 Hourly Proportions (after grouping) by Hour and Month for Friday at Station 9027-3

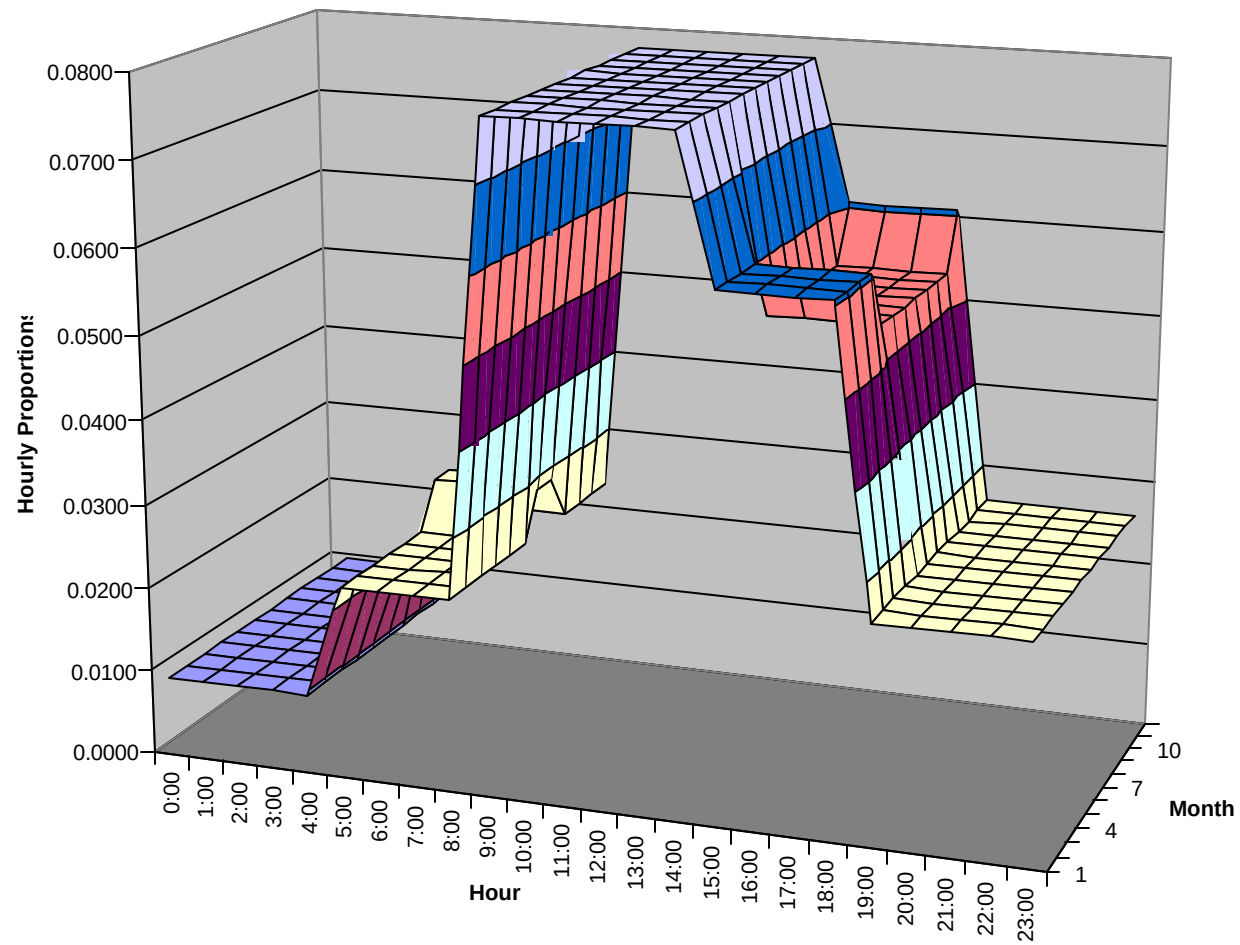


Figure 5-7 Hourly Proportions (after grouping) by Hour and Month for Saturday at Station 9027-3

Table 5-10 Final Grouping Evaluation

Location	RMSE	MAPE	
9007-1	.0014	4.17	
9007-5	.0013	2.52	
9012-1	.0019	4.19	
9012-5	.0015	4.05	
9014-1	.0009	1.90	
9014-5	.0008	1.99	
9024-3	.0012	3.46	
9024-7	.0014	3.03	
9026-3	.0009	2.10	
9026-7	.0010	2.39	
9027-3	.0021	4.67	
9027-7	.0017	4.11	
9030-1	.0009	2.09	
9030-5	.0009	1.86	
9032-1	.0016	2.92	
9032-5	.0016	2.90	
9033-1	.0015	3.31	
9033-5	.0011	2.16	
9044-1	.0015	2.78	
9044-5	.0012	2.44	
9045-3	.0015	3.28	
9045-7	.0014	3.97	
9049-3	.0009	2.41	
9049-7	.0010	2.79	
9053-1	.0012	3.46	
9053-5	.0011	1.81	
9054-3	.0013	3.30	
9054-7	.0015	2.99	
9055-1	.0009	2.05	
9055-5	.0010	2.19	
Average	.0013	2.90	

The variations in the groupings by hour and station means that hourly proportion models may have to be estimated separately for each hour group and study station if these groupings are used. It is generally acceptable to estimate hourly proportion models separately for each hour group. In fact, it is very common to estimate hourly proportions separately for these time periods in practice. However, it is not desirable to estimate hourly proportion models separately for each station, because such models basically have little or no use in practice. The next chapter describes a preliminary investigation into how this may be addressed.

5.5 CONCLUSIONS

The 24 hours in the day are first grouped into five time periods for the purpose of further simplifying hourly proportion models that involve hour, month and day of week and other site-specific variables. Then, groupings of months and weekdays are further identified for each hour group for each study station based on the grouping produced in Chapter 4. Furthermore, hourly proportions corresponding to the groupings are estimated. Finally, the accuracy of these hourly proportions and in turn the groupings are examined using summary statistics, such as root mean square error (RMSE) and mean absolute percent error (MAPE).

Our examination of the groupings reveals that they vary by the hour group and study station. This means that hourly proportion models may have to be estimated separately for each hour group and study stations if the groupings are used. While it is considered acceptable to estimate the models separately for each hour group, it is not desirable to estimate the models separately for each station, because such models basically have little or no use in practice. This issue needs to be addressed in further studies.

Nonetheless, an examination of the accuracy of the hourly proportion models estimated at each hour group and study station based on the groupings indicates that these groupings are appropriate to be used in hourly proportion models. In addition, closer examination of the groupings reveals that the months and weekdays can be grouped to a considerable level. This means that when these groupings are used, the hourly proportion models can be significantly simplified, which in turn makes the models easier to estimate and more desirable to use in practice.

6 A TRAFFIC SHED APPROACH FOR LOCATION-SPECIFIC ESTIMATION OF HOURLY TRAFFIC VOLUMES

6.1 INTRODUCTION AND BACKGROUND

Accurate estimation of hourly traffic volumes on transportation networks is vital for transportation planning, operations, and analysis. For transportation planning, hourly traffic volumes, among other factors, dictate priorities in highway improvement plans and allocation of funds. For traffic operations, hourly traffic volumes on links affect signal timing plans, air quality estimation, and traveler information systems. For safety analysis, an accurate estimation for hourly traffic volumes will help in assessing safety of different locations in the transportation networks as well as risk exposure levels.

It is evident that accurate estimation of link hourly volumes has challenged transportation systems analysts for decades. In the available literature, analysts developed the four-step transportation planning process, which is an aggregate model that eventually leads to estimating link volumes. In the first step in this process, land use, demographics and employment information are used to determine the number of trips generated by a transportation analysis zone (TAZ) and the number of trips attracted to a TAZ. Then, in the second step, these trips are combined with interzonal costs to derive an origin destination (OD) trip matrix. In the third step, OD trips are distributed among the available travel modes. Then, finally, for each mode of transportation, trips are assigned to the corresponding network to compute traffic volumes for each link.

Over the years, the steps in this process have been the subject of extensive research and refinement. Some of the efforts focused exclusively on developing superior models for one specific step. For example, the trip distribution step was first implemented using gravity and Fratar models, with logit-based models introduced later, and more recently discrete choice models (Sheffi 1984, Ben Akiva *et al.* 1985). Also, the traffic assignment step was originally carried out using Wardrop (Wardrop 1952) assignment criteria. Subsequently, researchers investigated other assignment criteria such as stochastic user equilibrium (Dial 1971, Daganzo and Sheffi 1977, Fisk 1980) and dynamic traffic assignment (Janson 1991, Florian and Hearn 1995, Daganzo 1994, 1995a, 1995b). Meanwhile, other research realized dependencies and feedback relationships among different steps and investigated models for solving combined steps, such as the Evans (1973, 1976) Model for solving the combined trip distribution and assignment problem. Other researchers reexamined the four-step process and suggested an activity-based transportation planning process. This high fidelity approach is based on modeling activity patterns for individual households and uses microsimulation to estimate daily household trip-making using activity patterns (TRANSIMS).

Indeed, all of these efforts significantly improved the quality of the transportation planning process, but at the same time made the process complex and data intensive. Furthermore, most of these approaches can be feasible for modeling urban areas with coherent networks and high population density. However, for suburban areas with heterogeneous network resolutions, sparse population densities and regional traffic, using

the traditional transportation planning process in estimating hourly traffic volumes becomes prohibitively data intensive and infeasible. Hence, there exists a need for developing a simple yet accurate method for estimating hourly traffic volumes in suburban/regional areas.

Given the annual average daily traffic (AADT), the conventional approach to estimate the peak hourly traffic volume is to use a K factor, defined as the ratio of the two-way design hour volume to the two-way AADT, and a D factor, defined as the ratio of the design hour volume in the major direction to the two-way design hour volume (May 1990). Allaire and Ivan (2001) estimated functions for predicting a peak hour factor to predict the peak hour volume as a proportion of the four-hour peak period traffic volume.

These approaches can only provide traffic distribution characteristics for one or several hour periods rather than portray a complete profile of daily traffic. This chapter introduces a compositional method for estimating hourly traffic volumes at a specific location using AADT and location characteristics. Traffic flow patterns at 15 continuous traffic count stations (ATR) on freeways in the state of Connecticut are explored. We present a methodology for estimating hourly traffic volumes on transportation links using annual average daily traffic volumes and land use characteristics in the “trafficheds” upstream and downstream of the location of interest. We define a traffiched area around each station representing the geographic areas from and to which trips passing the station are likely to originate and end. These traffic sheds are then used to extract spatial characteristics using Geographic Information Systems (GIS) to use as predictor variables in statistical models for predicting the daily flow patterns.

This chapter is organized as following: first we discuss the hypothesis of the approach and define our concept of traffiched. Following that, we present the mathematical model and the statistical approach we followed in determining model parameters. Then, we demonstrate some of the model results and conclude the chapter with summary and recommendations for future work.

6.2 MODEL HYPOTHESIS

The hypothesis for the proposed model is that the observed traffic volume at a specific station on a highway link depends primarily on two factors: A) demographics and land Use Patterns and B) network topology and structure. This section details the expected relationship of each of these factors.

6.2.1 Demographic and Land Use Patterns

Logically, the observed hourly traffic volume is affected by the demographic and land use characteristics upstream and downstream of the observation station. For example, densely populated areas are expected to have more trips than areas with low population densities. In addition, locations that are situated between predominantly residential areas and commercial areas are expected to experience more commuting trips than locations where upstream and downstream land use patterns are homogenous. For this reason, the

ratio between the number of employment opportunities and population in the upstream and downstream will affect the number of trips made across the observation station.

The factors that could be considered are, but not limited to, the following:

- a) Population: Number of people living up or down stream of the area of interest will affect the estimation of home-based trips. For example, a rural area will have low home based travel in comparison with an urban area.
- b) Employment: Number of job opportunities up and down stream of the area of interest will affect the estimated weights for trips between work and home locations,

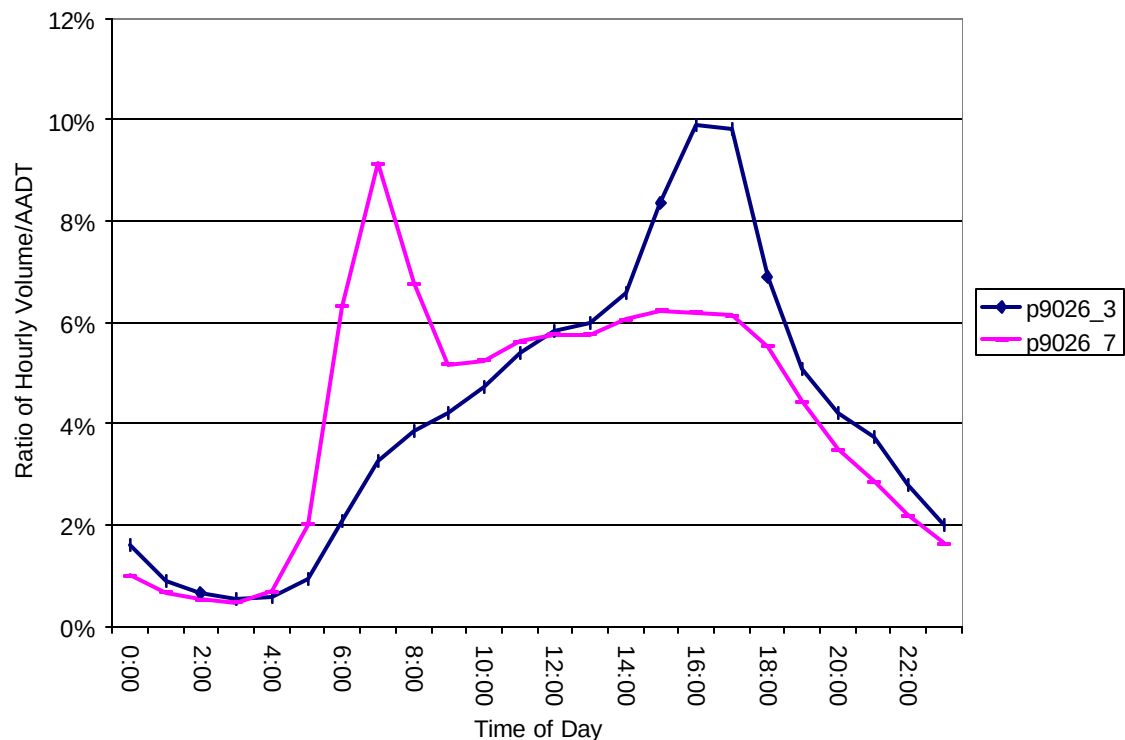


Figure 6-1 Example for Observed Hourly Traffic Volumes

6.2.2 Network Topology and Structure

Network topology and structure affect connectivity between different parts of the study area. Highly connected areas indicate the availability of alternative routes between origins and destinations. Hence, the size of trafficsheds up and down stream and the number of trips observed at a counting station will be highly dependant on the network

topology at that location. The factors that could be considered are (but not limited to) the following:

- a) Connectivity: network connectivity here refers to the quantity of routes crossing the upstream and downstream traffic sheds, in other words, the “longitudinal extent.”
- b) Network density: Network density refers to the quantity of routes parallel to the highway on which the count station is located, in other words, the lateral extent of the trafficshed.
- c) Regional Location: The relative location of the point of interest with respect to the region affects the amount of inter-regional (as opposed to intra-regional) traffic. For example, if the point of interest was located on an interstate that connects between two major cities, this parameter will be significant than if it was located on a rural state highway.

However, this chapter focuses only on using land use patterns and demographics to determine hourly volume proportions. Network structure and topology is postponed for later study. Hence, the assumption imposes a limitation on transferability of current results to only sites with similar location characteristics to those of the sample data set used in the analysis.

6.3 METHODOLOGY AND MATHEMATICAL MODEL

For a location on the network, the observed traffic volume could be assumed to be composed of a mix of trips that are made for different purposes. The proportions of this mix vary by location characteristics, land use, and network topography up/down stream of the observation location. Hence, it is acceptable to assume that the observed hourly traffic volumes at a specific location, as depicted in Figure 6-1, are composed of the weighted sum of different hourly trip profiles for different purposes as shown in Equation (6-1)

$$\bar{P}_{H \times 1}^i = [f]_{H \times R}^i * \bar{w}_{R \times 1}^i \quad (6-1)$$

$$\begin{bmatrix} p_0^i \\ p_1^i \\ p_2^i \\ \dots \\ p_h^i \\ \dots \\ p_H^i \end{bmatrix} = \begin{bmatrix} f_{0,1}^i & f_{0,2}^i & \dots & f_{0,r}^i & \dots & f_{0,R}^i \\ f_{1,1}^i & f_{1,2}^i & \dots & f_{1,r}^i & \dots & f_{1,R}^i \\ f_{2,1}^i & f_{2,2}^i & \dots & f_{2,r}^i & \dots & f_{2,R}^i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f_{h,1}^i & f_{h,2}^i & \dots & f_{h,r}^i & \dots & f_{h,R}^i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ f_{H,1}^i & f_{H,2}^i & \dots & f_{H,r}^i & \dots & f_{H,R}^i \end{bmatrix} * \begin{bmatrix} w_1^i \\ w_2^i \\ \dots \\ w_r^i \\ \dots \\ w_R^i \end{bmatrix} \quad (6-2)$$

Where,

$\bar{P}_{H \times 1}^i$ - The predicted hourly volume proportions vector at location (i) over (H) time intervals,

p_h^i - The predicted volume proportions at location (i) and during time interval (h),
where $\sum_{h=0}^H p_h^i = 1$,

$[f]_{H \times R}^i$ - Profile matrix for hourly proportions at location (i) over (H) time intervals and for (R) trip purposes,

f_{hxr}^i - Hourly proportion at location (i) and during time interval (h) and purpose (r), where
 $\sum_{h=0}^H f_{hr}^i = 1$ for all trip purposes

$\bar{w}_{R \times 1}^i$ - The weights vector at (i) for (R) trip purposes,

w_r^i - Weight at location (i) for trip purpose (r), where $\sum_{r=1}^R w_r^i = 1$

To illustrate the concept behind this compositional model, Figure 6-2 depicts schematically three trip purposes, from work, to work, and other. The weights for the profiles of the three trip purposes illustrated are then assumed to predict the hourly volume proportions. In this study the scope of the analysis will be limited only to the following trip purposes:

1. From home to work,
2. From work to home,
3. Other.

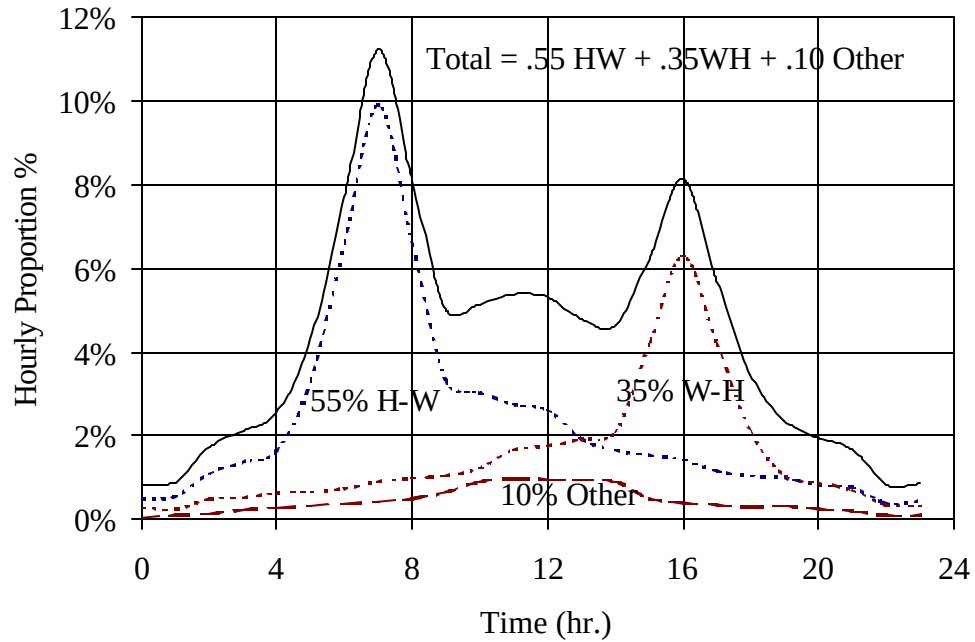


Figure 6-2 Conceptual Example for Estimating Hourly Volume Proportions

As mentioned earlier, hourly link volumes depend on two categories of parameters: demographic / land use activity-based parameters, and network topography based parameters. Under both categories, a set of sub-parameters could be identified. Some of these parameters are dependent on each other and some are not. Then, ideally the model presented in Equation (6-1) could be refined to the following:

$$\bar{P}_{Hx1}^i = [f(y^i)]_{HxR} * \bar{w}(x^i, t^i)_{Rx1} \quad (6-3)$$

Where y^i , x^i and t^i are location, land use, topography characteristics. And since we will be focusing in this chapter only on demographics and land use patterns, Equation (6-3) will be:

$$\bar{P}_{Hx1}^i = [f(y^i)]_{HxR} * \bar{w}(x^i)_{Rx1} \quad (6-3b)$$

Indeed, it would have been optimal to estimate the model in Equation (6-3b) with all the parameters mentioned above simultaneously, if achievable in reasonable computational times. Due to the large number of variables that are expected to be in the model and due to the heteroskedasticity of the problem, it was essential to assume that both sets of variables y^i and x^i are independent. This assumption will be examined later on in the analysis and its significance will be evaluated. Hence, a two-stage estimation scheme will be followed where in the first phase weights will be estimated for all observation

points. In the second phase, a parametric function to predict the weights will be estimated.

6.3.1 Determination of Hourly Proportions Matrix

The process for estimating the hourly proportions matrix $[f(y^i)]_{H \times R}$ for all trip purposes at the location of interest consists of three phases. The first phase is the trafficshed calculations, the second phase is the hourly departure proportions for each trip purpose, and the third phase is the conversion of departure profiles into arrival profiles at the location of interest.

6.3.1.1 *TRAFFICSHED Definition*

For a specific location where there is an interest to estimate the hourly volume proportions, we define up and downstream trafficsheds to represent the major producers and attractors of traffic passing through the location of interest as illustrated in Figure 6-3.

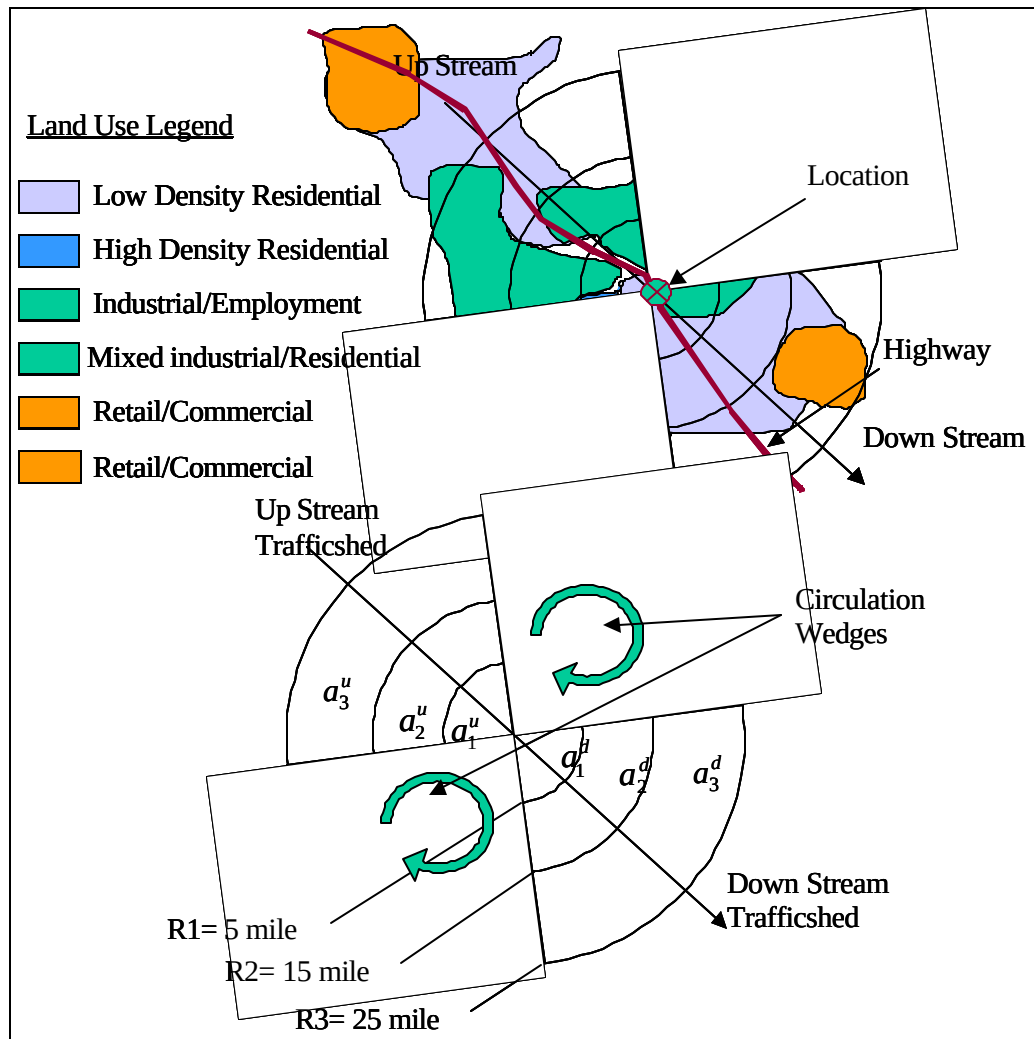


Figure 6-3 Trafficshed Determination Method

Figure 6-4 illustrates the cumulative distribution for trip duration in the state Connecticut based on the results obtained from the National Personal Transportation Survey (NPTS) conducted by the USDOT and the Census Bureau in 1996. Note that 50%, 90%, and 99.9% of home to work or home to shopping trips are less than or equal to 21, 30, and 42 minutes, respectively. That is nearly equivalent to 15, 20, and 30 miles, assuming an average speed of 40 mph. Hence, the trafficsheds are defined by drawing three circles 5, 15, and 25-mile radiuses and centered at the location of interest. Along a line tangential to the highway at the location of interest, two lines at 45 degrees are drawn. Then, the two wedges are defined with 45 degrees on either side of the tangential line, as shown in Figure 6-3. The 45-degree wedges assume that the 45 degree wedges perpendicular to the tangent are circulation wedges, i.e. traffic generated in those wedges will not pass through the location of interest and will use local roads instead. The upstream and

downstream traffic sheds are then defined by six areas $a_1^u, a_2^u, a_3^u, a_1^d, a_2^d$ & a_3^d as shown in Figure 6-3.

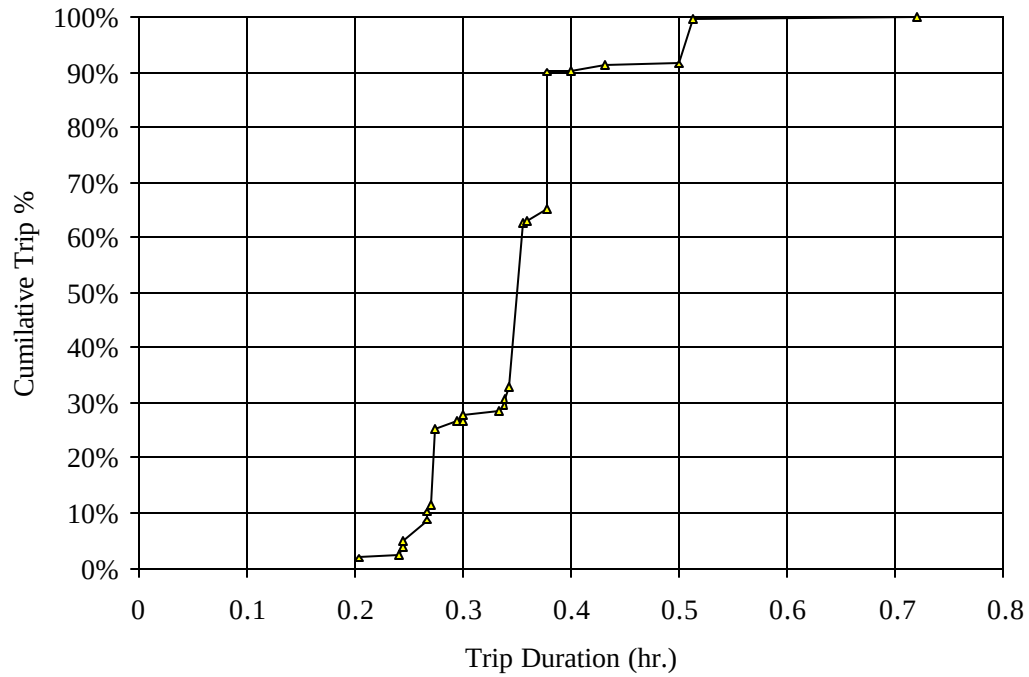


Figure 6-4 Cumulative Distribution for Home-Based Trip Duration in Connecticut

6.3.1.2 Hourly Departure Profiles

In order to determine the hourly proportions profile for each trip purpose, the hourly departure times for the three trip purposes defined earlier are shown in Figures 6-5a, 5b & 6. It is important to remember that these profiles report the time at departure. Hence, the actual observed profile at the location of interest will be a variation of the departure profile. In the following section a methodology for adjusting these profiles is given.

The results of these profiles could be summarized in a matrix $[f]_{H \times R}$ that is not site specific and is generic for the state of Connecticut.

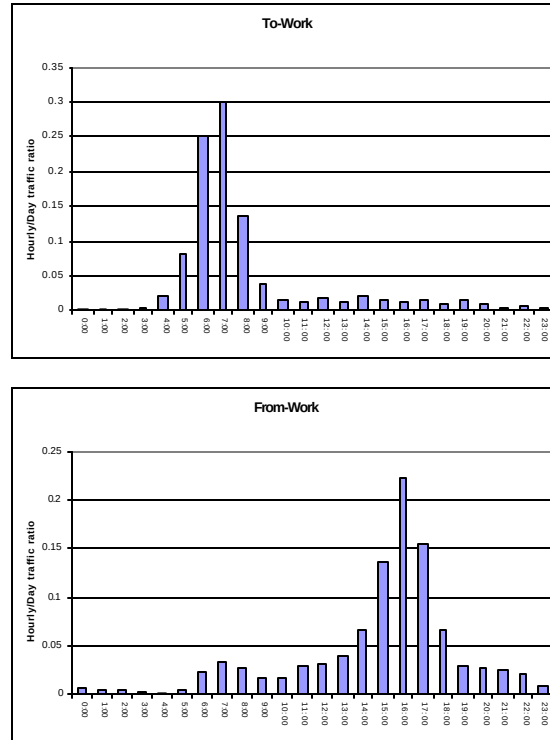


Figure 6-5 a & b Departure Profiles for home/work trips (NPTS 1996)

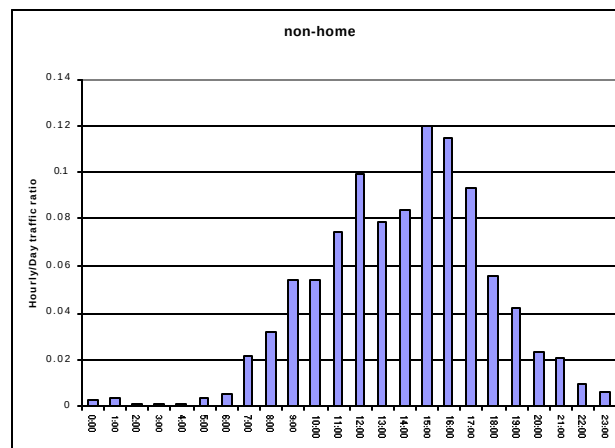


Figure 6-6 Departure profiles for other trip purposes (NPTS 1996)

6.3.1.3 Trip Purpose Hourly Arrival Matrix $[f(y^i)]_{H \times R}$:

The traffic arrival rate at the observation station is basically the departure rate profile for that purpose with an offset to account for the travel time between trip origin and the

observation station. The study team developed a method for determining the offset values for different trip purpose profiles. Not surprising, the results of the offset adjustment was insignificant and with little difference from the original departure profile. To explain this result, the reader could refer to Figure 6-4 that summarizes the average trip durations within the state of Connecticut. It is noted that nearly 90% of trips are less than or equal to 0.4 hour. Since trip departure profiles obtained from the NPTS are aggregated on hourly increments, which more than double the trip duration, it is explainable that such offset adjustment will not be observed. Hence, in this study we will use the trip departure rates as is without offset adjustments. Note that such simplification will have limited impact on the results, but the resulting simplification of the approach is worth it.

6.3.2 Weights Estimation Model

Since the sum of weights for all trip purposes should be equal to 1 and since the sum of all hourly volume proportions in a single day should equal also be equal to 1, an intercept w_o is introduced into the model.

$$\begin{bmatrix} p_{0:00}^i \\ p_{5:00}^i \\ p_{6:00}^i \\ \dots \\ p_h^i \\ \dots \\ p_{23:00}^i \end{bmatrix} = \begin{bmatrix} 1 & f_{0,twork}^i & f_{0,fwork}^i & f_{0,other}^i \\ 1 & f_{5,twork}^i & f_{5,fwork}^i & f_{5,other}^i \\ 1 & f_{6,twork}^i & f_{6,fwork}^i & f_{6,other}^i \\ \dots & \dots & \dots & \dots \\ 1 & f_{h,twork}^i & f_{h,fwork}^i & f_{h,other}^i \\ \dots & \dots & \dots & \dots \\ 1 & f_{23,twork}^i & f_{23,fwork}^i & f_{23,other}^i \end{bmatrix} * \begin{bmatrix} w_0^i \\ w_{twork}^i \\ w_{fwork}^i \\ w_{other}^i \end{bmatrix} \quad (6-4)$$

Also, it was observed that during early morning hours, traffic volume proportions were very low, so the first four hours of the day (from 1:00 to 5:00 AM) were considered to be the complementary proportion in the regression model. The hourly proportion for this period can be estimated as follows:

$$p_{0:00}^i = 1 - \sum_{h=5:00}^{23:00} p_h^i \quad (6-5)$$

Consequently, the time periods considered in the regression model were from 5:00 AM to midnight. The results from this step will be a set of weights for each site in the data set. Following this step, a relationship between trafficshed characteristics and weights will then be developed. This relationship will be in the following form:

$$w_r(x^i) = a_r + b_r * x^i \quad (6-6)$$

where,

w_r - Weigh for trip purpose (r)

a_r & b_r - Coefficients for trip purpose (r)

x^i - Is land use characteristics at site (i) , such that $x^i = \frac{R_{up}}{R_{down}}$, and

R_{up}, R_{down} - the ratios between population and employment in the up and down stream trafficsheds respectively

6.4 RESULTS AND DISCUSSION

6.4.1 Estimation Of Trip Purpose Weights

The data used in this analysis consist of the hourly traffic volumes obtained from 15 automatic traffic recorder (ATR) station locations in the state of Connecticut, with a single station in each of two directions at each location, covering the period from 1996 through 2000. Estimating hourly volume proportions using this approach demonstrated a good fit for the original data. The R^2 values for different sites varied between 0.84 and 0.95, which indicates that the model could explain most of the data variance. Figure 6-7 shows the distribution of the estimated weights. The intercept w_0 demonstrated less variation among the sites considered in this study. This result implies that the intercept represents a common factor that could not be explained by the independent variables. The estimated weight for trip purposes other than work related was the highest, nearly 50%-70% of all trips. This could be due to the topological location of the state of Connecticut in New England, in that it is located between two major metropolitan areas (New York & Boston). Trips from/to work, each on average contributed to nearly 10% of the total daily traffic volume. Also, the variance in the weights on from work trips differed from site to site. This could be due to variation in population and network densities among the sites.

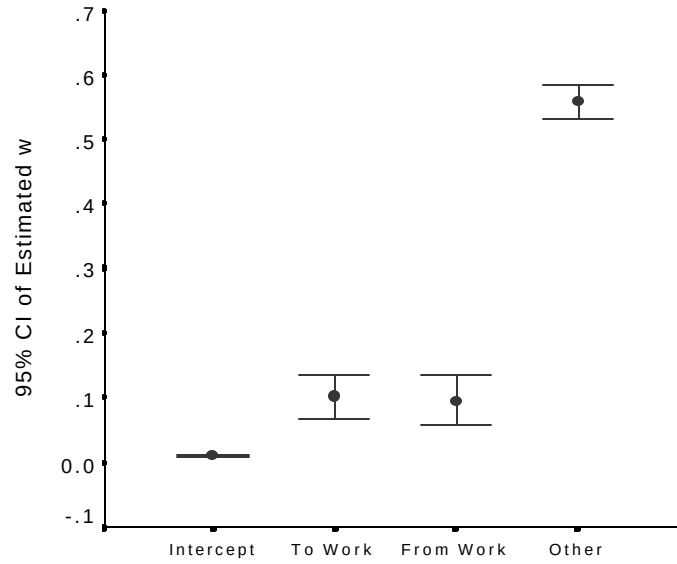


Figure 6-7 Distributions for Estimated Weights by Trip Purpose

6.4.2 Estimation of Weight Functions

The weights for each trip purpose theoretically are functions of land use and network topography as discussed earlier. However, we are only considering land use characteristics in this part. Hence, the weights for each trip purpose w_r could be defined as a function of land use site characteristics, $w_r(x^i)$ as in Equation 6-3. The weights obtained from the previous step were used to fit the weight prediction functions for the three trip purposes considered in the analysis. Figures 6- 8, 9 and 10 illustrate estimated weights and the best-fit function for the three trip purposes. The R^2 for the home to work and work to home trips were 0.689 and 0.665, respectively, indicating a reasonable fit for both. On the other hand, the R^2 for non-work trips was 0.337, which is relatively a poor fit. This poor fit is expected, since we did not consider the network topology in estimating the weights function. The estimated weight prediction functions are:

$$w_{ToWork}(x^i) = 0.28 - 0.158x^i$$

$$w_{FromWork}(x^i) = -0.093 + 0.178x^i$$

$$w_{Other}(x^i) = 0.648 - 0.085x^i.$$

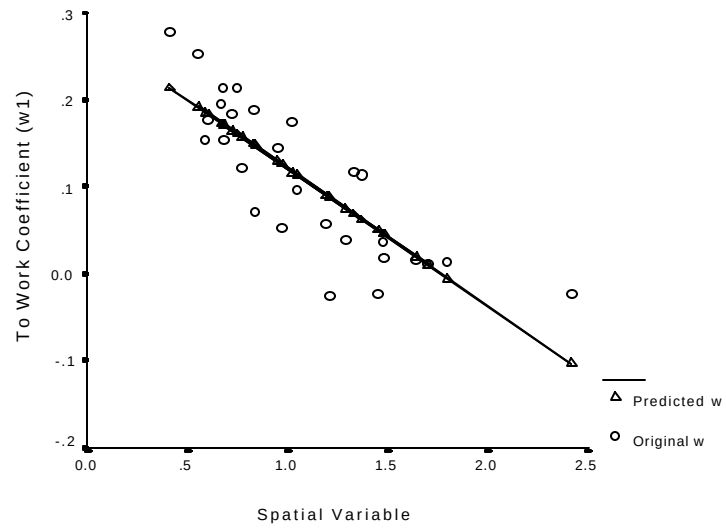


Figure 6-8 Estimated Weights for Work to Home Trips

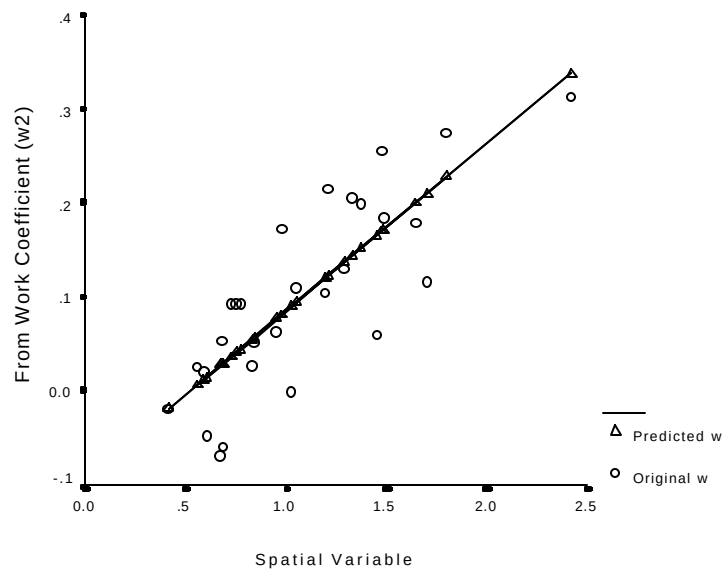


Figure 6-9 Estimated Weights for Home to Work Trips

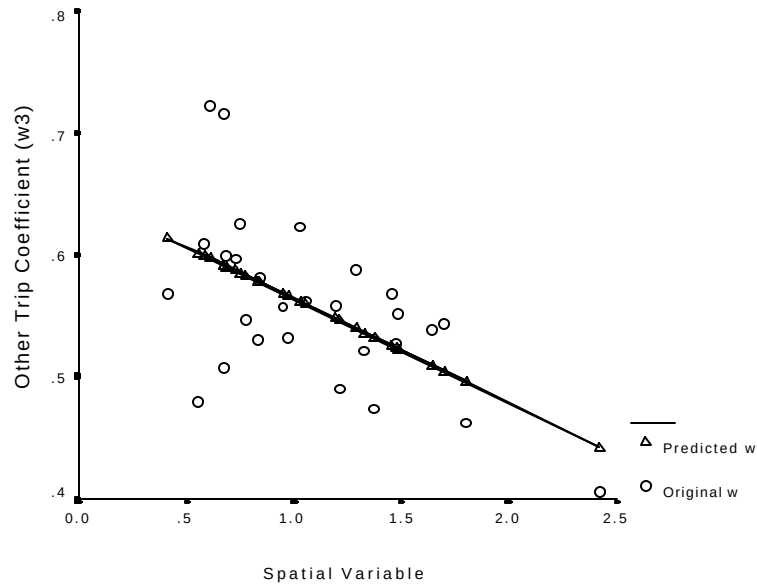


Figure 6-10 Estimated Weights for Other Trips

The predicted weights were then used in the previously defined volume proportion prediction functions to predict proportions for each station in the data set. The resulting predicted proportions were then compared to the actual proportions for all stations at each hour, and 95 percent confidence intervals computed for each hour. Figure 6-11 illustrates these confidence intervals. All of these errors are in the range of 1%, indicating a good estimation of hourly proportions.

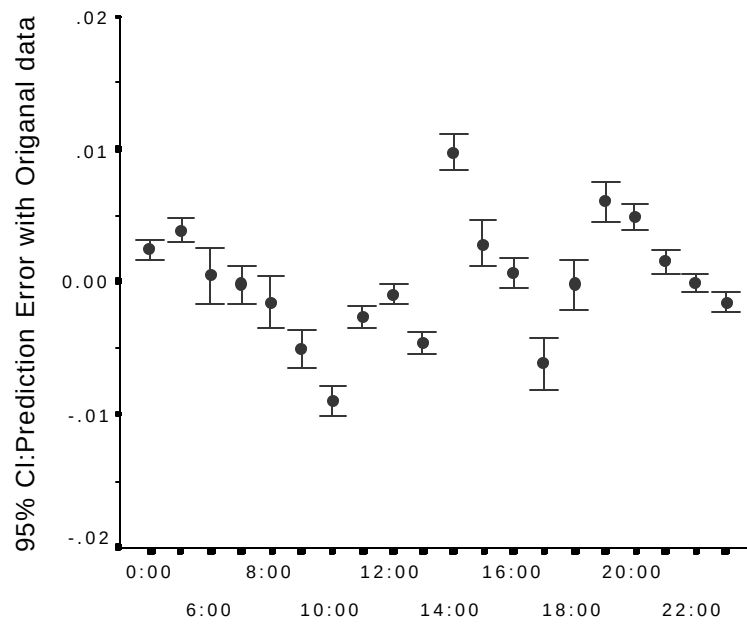


Figure 6-11 Error Distributions for Prediction Using the Estimated Weights

6.5 SUMMARY AND CONCLUSION

In this chapter a simplified method for estimating hourly traffic volumes using upstream and downstream trafficshed land use data and AADT was presented. This method was based on the assumption that traffic volume on a link is a weighted sum of traffic made for different purposes. The weights are a function of land use and network topology. In this part of the research, we considered only land use data.

The results demonstrated the validity of the method. However, there is a transferability limitation on the current results. The current weight functions could be applied only on sites with network topology similar to those in the data set used in the analysis. This barrier will be overcome in the future after incorporation network topology in the weight functions and the trafficshed calculations.

7 SUMMARY AND FURTHER STUDIES

The objective of this work is to learn how to calculate approximately hourly proportion models for estimating highway link hourly volumes, which are important for models in air quality assessment, vehicle crash prediction and transportation planning. Estimating accurate and reliable hourly proportion models is actually very challenging, because there are many factors affecting the hourly proportions. These factors in general include 1) geometric and operational features, socio-economic characteristics and land use patterns associated with the highway network, and 2) temporal factors, such as hour, day of week and month. If all these factors are included, an hourly proportion model most likely becomes too complex to estimate. Consequently, the primary concern here is to find a way to simplify such models.

Following is the specific research work conducted and described in this report:

1. The variation of the hourly proportions by hour, day of week and month are first investigated to gain a better understanding of the effects of these factors.
2. Hourly proportion models considering only these factors are then estimated, based on the findings of the investigation.
3. Procedures are further established to group the factors, so that they can be incorporated in hourly proportion models involving other site-specific factors.
4. A prototype method for estimating models to predict proportions for specific highway locations is described and demonstrated.

The following sections give a summary of the methodology and results of this work. In addition, some application issues and potential further studies are discussed a little later.

7.1 METHODOLOGY AND RESULTS

The effects of hour, day of week and month on the hourly proportions are investigated here using ANOVA statistical procedure, with the hourly traffic counts collected at the ART stations on Connecticut freeways as input. The primary purpose of this exercise is to identify whether these factors interact with each other, since an hourly proportion model can be significantly simplified, if these factors do not interact with each other. The interaction effects of these factors are tested using three sets of models in three different levels:

1. Full model with hour, day of week and month;
2. Hourly model with day of week and month;
3. Hourly model with only weekday and month.

The test results show that the interaction effects (as well as the main effects) are significant at over 95 percent confidence level when all three factors are included. When confining to each hour but including all seven days in the week, the interaction effect (as well as the main effects) is again significant at over 95 percent confidence level. However, when further confining to weekday only, the test results reveal that the interaction effect is in general not significant, while the main effects of weekdays and months are still significant.

These findings lead to our hourly proportion models. Specifically, these models are estimated separately at each hour and for weekday, Friday, Saturday and Sunday, individually. For weekdays, the models include the main effect terms of month and weekday, but not their interactions. (Note that the models are significantly simplified since the interaction term is omitted.) For Friday, Saturday and Sunday, the models include only the main effect term for month. The model parameters are estimated using least squares methods. These models are expected to provide us with more accurate and reliable hourly proportion estimates, since it is specific to hour, day of week and month.

Also based on the ANOVA results, procedures are established to group hour, day of week and month. First, Tukey's Comparison of Means procedure is used in combination with an engineering criterion to compare the means of the hourly proportions corresponding to each weekday and month at each hour. This is to identify the months or weekdays for which the means are not significantly different. It is considered that these months or weekdays can be grouped. Then, proper mutual exclusive groupings of months and weekdays are produced at each hour based on the comparison results.

Considering that it is often not necessary to estimate hourly volumes (in turn, hourly proportions) for each hour in the day (as opposed to certain hour groups), the 24 hours in the day are further grouped into:

1. Early morning (0:00 to 5:00),
2. A.M. peak (5:00 to 9:00),
3. Mid-day (9:00 to 15:00),
4. P.M. peak (15:00 to 19:00),
5. Evening (19:00 to 0:00).

For each of these time periods, overall groupings of weekdays and months are further produced at the end.

An examination of the accuracy of the hourly proportion models estimated based on the groupings indicates that these groupings are appropriate to be used in hourly proportion models. In addition, an examination of the final groupings also reveals that the months and weekdays can be grouped to a considerable level. This means that when these

groupings are used, the hourly proportion models can be significantly simplified, which in turn make the models easier to estimate and more desirable to use in practice.

A prototype model for predicting hourly link volume proportions using location-related variables was estimated using observed hourly volumes available from permanent count stations. The preliminary estimation results presented here show that the prototype model shows promise for predicting hourly traffic volumes accurately as a function of the population and employment patterns upstream and downstream of the highway link. Other characteristics specific to the highway network context in which the link is situated are expected to help improve predictions even more. These would include the availability of parallel routes between the upstream and downstream population and employment areas and routes connecting them to the highway on which the link is situated.

7.2 APPLICATION AND FUTURE WORK

Because the resulting models and groupings are separately produced at each study location in this work, they are only valid or applicable for the study locations. As a result, the models cannot be directly used to estimate hourly proportions for a highway location of interest and the groupings cannot be directly used in an hourly proportion model involving the temporal factors (*i.e.*, hour, day of week and month) and other site-specific variables. This issue needs to be addressed in further studies.

One approach to make the hourly proportion models applicable for a highway link of interest may be to establish a procedure to select a model to use for the highway link. This may involve only a simple comparison of the daily traffic pattern at the location of interest with the daily traffic patterns at each of the study locations. If one of the later traffic patterns is similar to the former traffic pattern, the corresponding model and hourly proportions can be used to estimate hourly volumes for the location of interest. The daily traffic pattern for the location of interest may be obtained by short-term (*e.g.*, 24 hours) traffic counts at the location.

Alternatively, it may be done by first categorizing (or grouping) the study locations. This categorization should be performed based on adequate understanding of the variations of the hourly proportion by locations. Hence, factors that result in the variations in the hourly proportions by locations (*i.e.*, the site specific variables) needs to be identified and their effects need to be investigated. This requires traffic count data and other site-specific data for a reasonable large number of highway locations; thus further data collected are almost definitely necessary. Once the study locations are categorized, the hourly proportion models are re-estimated for each of the categories. The only problem left is to put the location of interest into one of the categories, which should be an easy task.

A better approach is to gather hourly volumes at a larger sample of highway locations and estimate "w-factors" for day of week and month of year for each location. A carefully designed statistical estimation experiment may be able to account for missing month and day of week observations in order to estimate parameters that may be used anywhere on the sampled highway network. The results from Chapters 3, 4 and 5 would be used to

reduce the number of combinations of month and day of week that would need to be sampled at each traffic count observation point. At this point, we know the significant and relevant variation through the year and through the week for each permanent counting station. What is still required is to use these results to decide which grouping patterns to use at locations where the full annual variation is not available.

Hence, all of the hourly proportions observed at each location -- even for periods of less than a full calendar year -- could be classified by month and day of week group, and the w-factors again estimated for each group as a function of the population and employment factors and the daily trip purpose trip start time distributions. The result would be a set of models for predicting the w-factors at any location on the road network for a given time of year and day of the week. This ultimately is what is needed for truly accurate estimation of traffic volume by time of day for all locations on the highway network.

REFERENCES

- Abraham, B. and Ledolter, J. 1983. *Statistical Methods for Forecasting*. John Wiley & Sons, Inc., New York, N. Y.
- Allen, W. G. Jr., and Schultz, G. W. 1996. "Congestion-Based Peak Spreading Model." *Transportation Research Record 1556*, TRB, National Research Council, Washington, D.C., 8-15.
- Cambridge Systematics, Inc. 1994. *Short-Term Travel Model Improvement*.
- ConnDOT. (1994) *State of Connecticut Traffic Monitoring System for Highways*, Connecticut Department of Transportation.
- Crevo, C. C. and Virkud, U. 1994. "Practical Approach to Deriving Peak-Hour Estimates from 24-Hour Travel Demand Model." *Transportation Research Record 1443*, TRB, National Research Council, Washington, D.C., 30-37.
- Daganzo, C. F. and Sheffi Y. 1977. "On Stochastic Models of Traffic Assignment." *Transportation Science*, 11(3)
- Daganzo, C.F. 1994. The cell transmission model: a simple dynamic representation of highway traffic. *Transportation Research B*, 28 (4), 269 –287.
- Daganzo, C.F. 1995a. The cell transmission model part II: network traffic. *Transportation Research B*, 29 (2), 79 –93.
- Daganzo, C.F. 1995b. Properties of link travel time functions under dynamic loads. *Transportation Research B*, 29 (2), 93 –98.
- Daly, A. J., Gunn, H. F., Hungerink, G. J., Kroes, E. P., and Mijjer, P. D. 1990. "Peak-Period Proportions in Large-Scale Modeling." *Proceedings of Seminar H held at the PTRC Transport and Planning Summer Annual Meeting*, PTRC Education and Research Services Ltd., 215-226.
- Deakin, Harvey, Skabardonis, Inc. 1993. *Manual of Regional Transportation Modeling Practice for Air Quality Analysis*.
- Evans, S. P. 1973. "A Relationship between the Gravity Model for Trip Distribution and the Transportation Problem in Linear Programming" *Transportation Research*, 7, 39-61.
- Evans, S. P. 1976. "Derivation and Analysis of Some Models for Combined Trip Distribution and Assignment," *Transportation Research*, 10, 35-57.
- FHWA. 1994. *Workshop on Transportation Air Quality Analysis*. U.S. Department of Transportation, Federal Highway Administration, FHWA-141-94-011.
- FHWA. 1995. *Traffic Monitoring Guide*. U.S. Department of Transportation, Federal Highway Administration.
- Fisk, C. 1980. "Some Developments in Equilibrium Traffic Assignment Methodology", *Transportation Research B*, 14, 243-256.
- Florian, M., Hearn, D. 1995. "Network equilibrium models and algorithms." In: Ball, M.O., et al. (Eds.), *Handbooks in Operations Research and Management Science*, 8. Network Routing. Elsevier Science, The Netherlands.
- Gunawardena, N. R., Sinha, K. C., and Fricker, J. D. 1996. "Development of Peak-Hour and Peak Directional Factors for Congestion Management Systems." *Transportation Research Record 1552*, TRB, National Research Council, Washington, D.C., 8-18.

- Gwynn, D. 1967. "Relationship of Accident Rates and Accident Involvements with Hourly Volumes." *Traffic Quarterly*, 21 (3), 407-418.
- Horowitz, J. L. 1982. *Air Quality Analysis for Urban Transportation Planning*. MIT Press, Cambridge, Mass.
- Ivan, J. N., Pasupathy, R. K., and Ossenbruggen, P. J. 1999. "Differences in Causality Factors for Single and Multi-Vehicle Crashes on Two-Lane Roads." *Accident Analysis and Prevention* 31, 695-704.
- Janson, B., 1991. "Dynamic Traffic Assignment With Arrival Time Costs." *Transportation Research B* 25, 143 –161.
- Jayakrishnan, R., Tsai, W.K., Chen, J., Chen, A., 1995. "A Dynamic Traffic Assignment Model With Traffic Flow Relationship." *Transportation Research C* 3, 51 –82.
- Kumar, A. and Levinson, D. 1995. "Temporal Variation on Allocation of Time." *Transportation Research Record* 1493, TRB, National Research Council, Washington, D.C., 118-127.
- Loudon, W. R., Ruiter, E. R., and Schlappi, M. L. 1988. "Predicting Peak-Spreading Under Congested Conditions." *Transportation Research Record* 1203, TRB, National Research Council, Washington, D.C., 1-9.
- Magnanti, T. L., and Wong, R.T. 1984. "Network Design and Transportation Planning Models: Models and Algorithms." *Transportation Science*, 18, 1-55.
- Montgomery, D. C. 1991. *Design and Analysis of Experiments*. 3rd ed, John Wiley & Sons, Inc., New York, N. Y.
- Peat, Marwick, Mitchell & Co. 1972. *An Analysis of Urban Area Travel by Time of Day*. Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.
- Powell, W and Sheffi, Y. 1982. "The Convergence of Equilibrium Algorithms with Predeter-mined Step Sizes", *Transportation Science*, 16 (1), 45-55.
- Ran, B, Lo, H., and Boyce, D. 1996. "A Formulation And Solution Algorithm For A Multi-Class Dynamic Traffic Assignment Problem." In Lesort (Ed.), *Transportation and Traffic Theory*, 195 –216.
- Ran, B., and Boyce, D. 1996. *Modeling Dynamic Transportation Networks. An Intelligent Transportation System Oriented Approach*, second revised ed. Springer, Heidelberg.
- Robertson, H. D., Hummer, J. E., and Nelson, D. C. 1994. *Manual of Transportation Engineering Studies*. Prentice-Hall, Inc., Englewood Cliffs, N. J.
- SAS Institute Inc. 1990. *SAS/STAT User's Guide, Volume 2*. Version 6, 4th ed, Gary, NC.
- Sheffi, Y., 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, NJ.
- Sheffi, Y. and Powell, W. "A Comparison of Stochastic and Deterministic Traffic Assignment Over Congested Networks." *Transportation Research*, 15B, 53-64.
- Smith, M.J. 1993. "A New Dynamic Traffic And The Existence And Calculation Of Dynamic User Equilibria On Congestion Capacity-Constrained Road Networks." *Transportation Research B* 27, 49 –63.
- SPSS Inc. 1998. *SPSS Base 8.0, Applications Guide*. Chicago, IL.
- The Mathworks, Inc. 1998. *MATLAB The Language of Technical Computing, Using MATLAB Version 5.2*, Natick, MA.

- Van Every, B. E. and George, A. T. 1981. "Hourly Traffic Volume Patterns Throughout the Day in Melbourne." *Australian Road Research*, 11 (1), 60-69.
- Wardrop, J. 1952. "Some Theoretical Aspects of Road Traffic Research" *Proceedings of the Institute of Civil Engineers, Part 2*, 325-378.
- Zhao, M., Garrick, N. W., and Achenie, L. K. 1998. "Data Reconciliation-Based Traffic Count Analysis System." *Transportation Research Record 1625*, TRB, National Research Council, Washington, D.C., 12-17.
- Zhou, M., and Sisiopiku, V. 1997. "Relationship Between volume to Capacity Ratios and Accident Rates." *Transportation Research Record 1581*, TRB, National Research Council, Washington, D.C., 47-52.

APPENDIX A: MODEL ADEQUACY CHECKING

In the ANOVA, it is assumed that the errors (or residuals) of the model are normally and independently distributed with zero mean and constant but unknown variance. The ANOVA results cannot be considered to be totally valid if these assumptions are significantly violated. Hence, the validity of these assumptions needs to be checked as an integrated part of ANOVA.

The normal distribution assumption can be checked using a normal probability plot of the residuals. This plot should resemble a straight line, if the residuals are normally distributed. The equal-variance assumption can be checked by constructing a residual vs. fitted value plot. This plot should reveal no obvious patterns, or in other words, the residuals should be randomly scattered around zero. Since the ANOVA is generally robust to these two assumptions for fixed effect models, which is the case for our models, moderate departure from these assumptions is not of great concern here. The independence assumption can be checked using plots of the residuals in time order of data collection. A tendency of runs of positive and negative values implies that this assumption is violated. This assumption is critical to ANOVA and its validity should be carefully examined.

Figures A.1 through A.3 show the normal probability, residual vs. fitted value and time series plots, respectively, for the Hourly Model I (Equation 3.12) for Station 9027-3. As can be seen, the normal probability plot shows slight deviation of a straight line, and the residual vs. fitted values plot reveals that the residuals slightly decrease when the fitted values increases. This indicates a moderate violation of the normality and equal variance assumptions, which is acceptable here because our models are generally robust to these two assumptions as discussed earlier. The time series plot of the residuals shows no strong tendency of runs of positive and negative values, meaning that the residuals are generally independent. Therefore, it is concluded that the model assumptions are generally valid in this case.

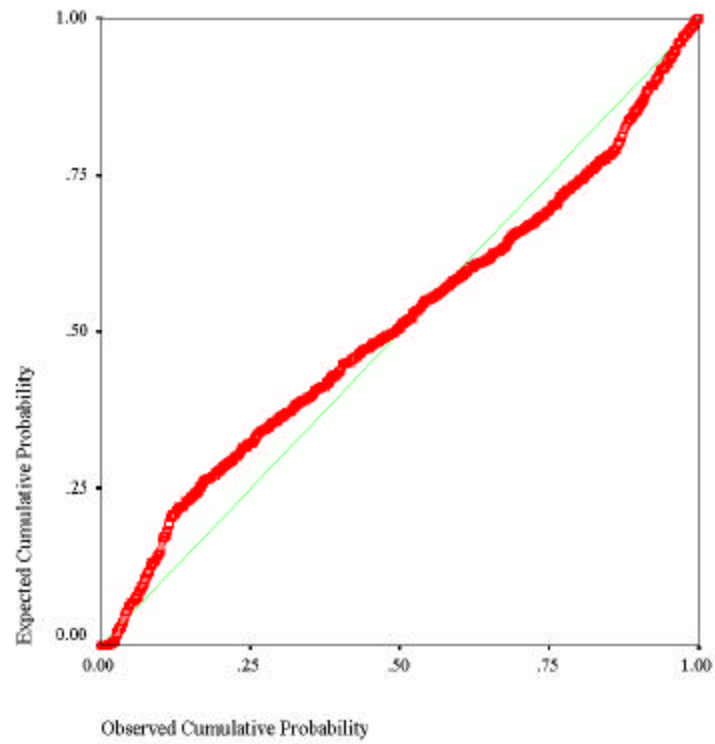


Figure A 0-1 Normal Probability Plot of Residuals (Hourly Model I for Station 9027-3)

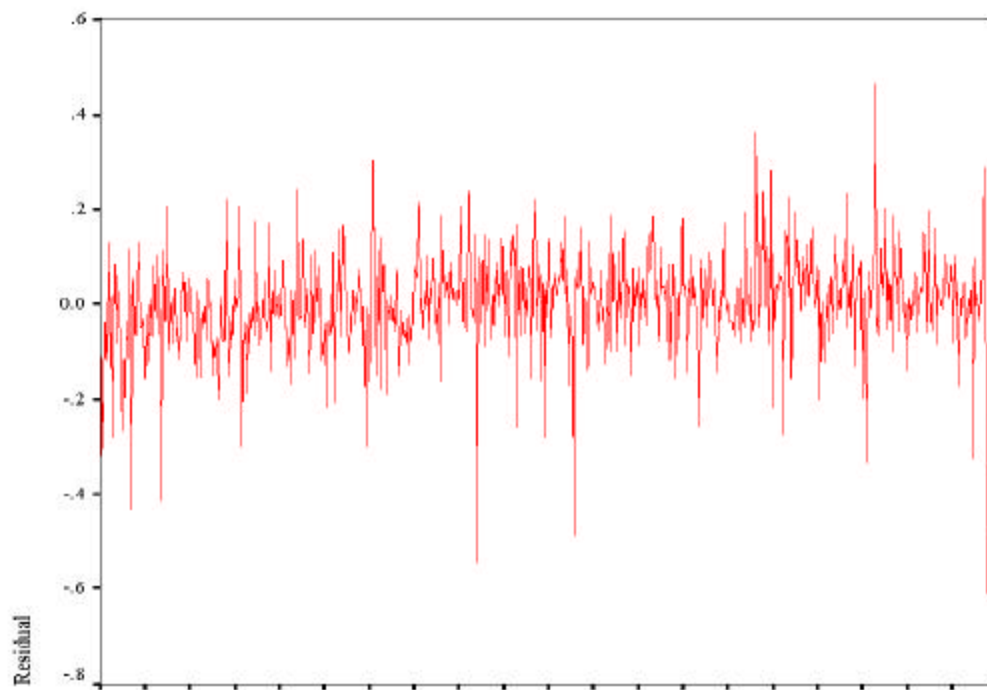


Figure A 0-2 Time Series Plot of Residuals (Hourly Model I for Station 9027-3)

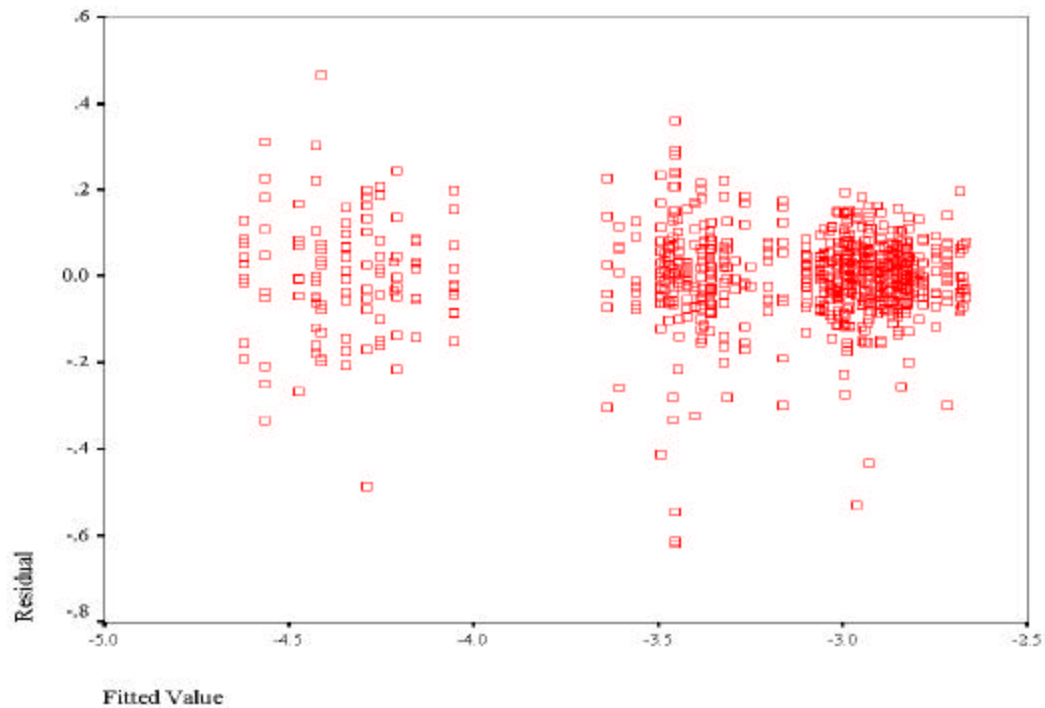


Figure A 0-3 Residual vs. Fitted Value Plot (Hourly Model I for 9027-3)

In addition, Figures A.4 through A.6 show the normal probability, residual vs. fitted value and time series plots, respectively, for Hourly Model II in Chapter 3. Here the normal probability plot shows very slightly deviation of a straight line, and the residual vs. fitted values plot reveals no obvious pattern, indicating again that the normality and equal variance assumptions are generally valid. Also, the time series plot of the residuals shows no strong tendency of runs of positive and negative values. Therefore, we conclude that the model assumptions are again generally valid in this case.

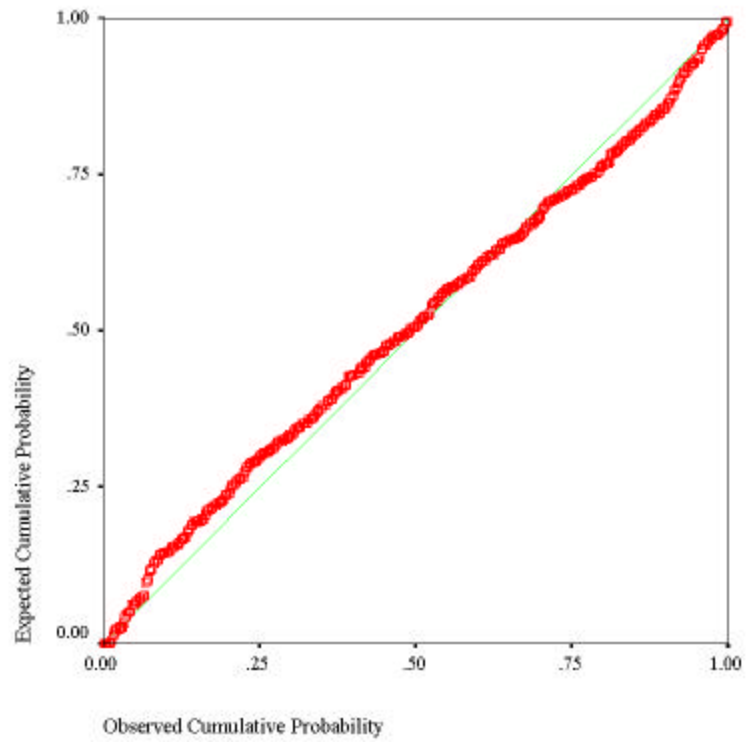


Figure A 0-4 Normal Probability Plot of Residuals (Hourly Model II for Station 9027-3)

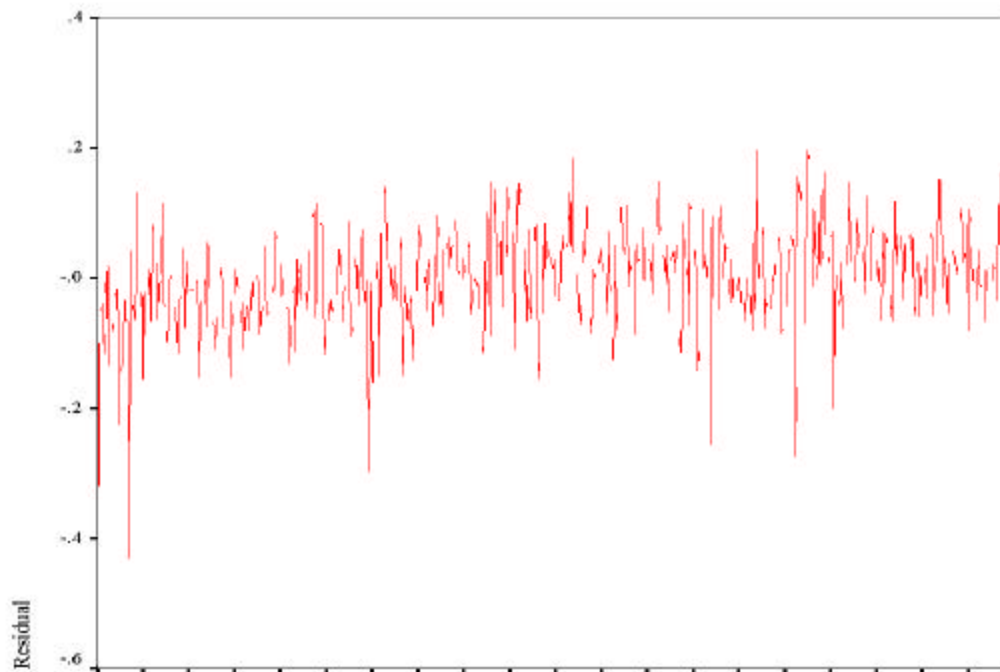


Figure A 0-5 Time Series Plot of Residuals (Hourly Model II for Station 9027-3)

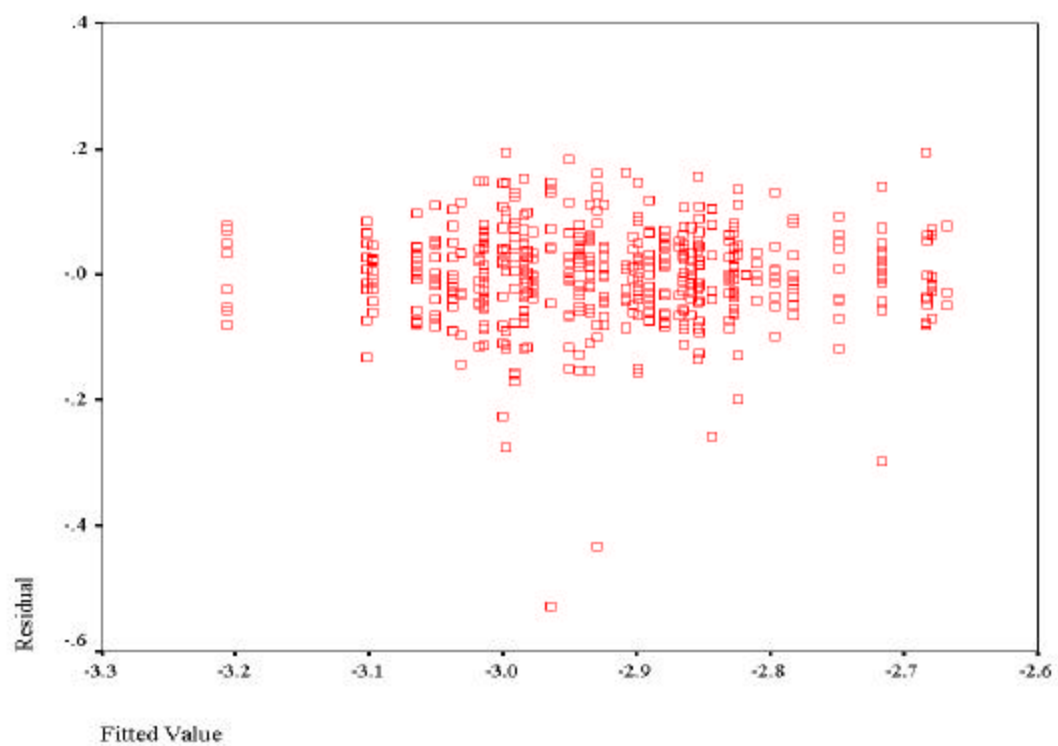


Figure A 0-6 Residual vs. Fitted Value Plot (Hourly Model II for 9027-3)