Estimation of Hourly Allocation Factors for the Central California Ozone Study Using State-wide Model Data and Real Time Traffic Data

Final Report

Prepared for

California Air Resources Board
2020 L Street
Sacramento, CA 95812-2815

Prepared by

Terence Lam
Post-Doctoral Researcher

Debbie Niemeier
Principal Investigator

Kiettipong Jierranaitanakit
Research Assistant

Department of Civil and Environmental Engineering
University of California
One Shields Ave.
Davis, CA 95616

July 2002

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

As part of ongoing efforts to improve the precision of emission estimates in the Central California Ozone Study (CCOS), observed traffic data are being used in conjunction with travel demand model information to obtain more accurate hourly estimates of traffic volumes. Improved hourly estimates of traffic are required because travel demand models typically produce volumes in blocks of time, some as long as the full 24 hours. For photochemical modeling, hourly estimates are required.

The purpose of this report is to estimate hourly allocation factors for CCOS applying a disaggregation method that makes use of observed traffic data to predict the hourly allocation of traffic volumes for links in travel demand network. The report consists of seven sections. The study purpose and study domain are presented in Section 2. This is followed by a discussion of travel demand model network and traffic counter data. In Section 4, the analytical method and data summaries statistics are described. The analysis results are presented in Section 5. Section 6[1] presents a comparison of DTIM emission results using the allocation factors derived with the new disaggregation method vis-à-vis the conventional trip diary method. The report ends with conclusion in Section 7.

2.0 PURPOSE AND STUDY DOMAIN

2.1 Purpose

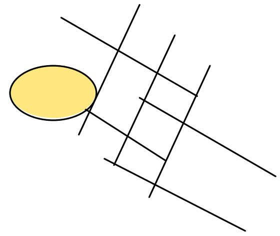
In travel demand forecasting, roadway (or link) traffic volumes are estimated by modeling period (e.g., AM peak or PM peak), where periods can include any number of hours. Alternatively, photochemical air quality models, such as EPA's Urban Airshed Model (UAM), require hourly volumes as an input for photochemical modeling. The hourly running stabilized emissions needed as input for air quality modeling have traditionally been estimated using a post-processor, such as the Direct Travel Impact Model (DTIM) [Caltrans, 1998 #35]. As the region-wide emissions modeling tool for California, DTIM is designed to produce estimates of total on-road vehicle emissions inventories, while also serving as the interface to prepare gridded data for input to photochemical air quality models, such as the UAM [Caltrans, 1994 #43; Caltrans, 1998 #35].

As Figure 2.1 illustrates, in DTIM travel diary information is used to estimate the proportion of trips beginning or ending in each hour. The total period volume (e.g., AM peak, 7a-9a) produced by the travel demand models is then disaggregated to each hour within that period. Travel diaries are surveys in which respondents are asked to record travel activities over a 2-5 day period and from these travel diaries, we know how many trips each household makes and the time at which each trip begins and ends.

[1] Section 6 was not completed by the project end data. Input data, which were required for DTIM runs, were not received from Alpine Geophysics, Inc.

Figure 2.1 Current Method for Establishing Hourly Volumes

$Link_j$ (AM:3hrs) = 2000



AQ Models require hourly estimates

From travel diaries:

hhid	subid	pers	diary	dayofwk	md	tottrip	tripnum	purpose	type	mode	begtime	endtime
4	0	1	1	3	204	4	1	1	1	1	715	730
4	0	1	1	3	204	4	2	8	1	1	1600	1615
4	0	1	1	3	204	4	3	6	2	1	1845	1900
4	0	1	1	3	204	4	4	8	2	1	2030	2045
4	0	1	2	4	205	3	1	1	1	1	710	725
4	0	1	2	4	205	3	2	7	3	1	1500	1525

There are several problems with using travel diary trip ends to estimate the hourly stabilized volumes. First, the surveyed hourly breakdowns of trip ends represent the proportions of trips beginning or ending in any given hour, rather than the actual stabilized on-road travel occurring in each hour. The other problem with using the trip end data is that the same proportions are applied across all modeling links (roadway segments) in the transportation network. That is, under the traditional method there is no way to spatially distinguish different hourly profiles (e.g., the suburbs may peak earlier than the CBD). Thus, the same temporal allocation is used spatially.

A method was developed to disaggregate the period-based demand model link volumes into hourly volumes [Lin, 1998 #8] and to incorporate spatial variability in the disaggregation [Hicks, 2001 #10]. This method uses observed hourly traffic count data to stochastically estimate hourly allocation factors, which in turn represent the expected value of traffic occurring during each hour within a modeling period. Thus, these allocation factors represent statistical estimations of the proportions of observed hourly on-road traffic and can be easily used to disaggregate the period-based travel demand model link volumes into hourly profiles, which can then be directly input into models such as DTIM. More important, observations can also be statistically clustered into groups of similar temporal traffic pattern, allowing spatial variability to be introduced.

Note also that since the total vehicle miles of travel (VMT) is not changing (and thus, total estimated running stabilized emissions are constant with and without the application of the method), the new method primarily results in a shifting of emissions from hour to hour by taking into account what is actually observed on the roadway. In other words, the estimated total daily emissions will be roughly the same using the default method (i.e., travel surveys) and the new method (i.e., using observed volumes), as long as the estimated travel demand model roadway VMT remains the same. It is possible that a minor change in the estimated <u>total</u> daily emissions will occur because hourly changes in the diurnal temperature profile (i.e., some hours will have higher (or lower) temperatures associated with them).

Understanding shifts in predicted diurnal emissions is important for better predicting ambient conditions. This can be best illustrated by examining the relationship between ozone, NO_x, and VOC's. In general, when the ratio of VOC's to NO_x is greater than about 8-10, NO_x tends to promote ozone formation; when the ratio is less than about 8-10, NO_x tends to restrict ozone formation (Finlayson-Pitts and Pitts, 2000). The ratio can vary not only by time of day but also by location (Seinfield and Pandis, 1998). In the South Coast air basin, mobile sources are responsible for about 85% of the estimated year 2000 NO_x annual average emissions and about 54% of the estimated year 2000 ROG annual average emissions. So understanding how mobile emissions can vary both temporally and spatially provides insight on how the ratio might vary as well, which, in turn, can improve our understanding of how to best implement control strategies.

2.2 Study Domain

The study domain extends from Redding in the north to the Mojave Desert in the south, and from the Pacific Ocean in the west to the Sierra Nevada Mountains in the east (Refer to the web-site of California Air Resource Board on **http://www.arb.ca.gov/airways/ccos/ccos.htm**.) For illustrative purpose, the approximate study domain is traced out on a California Transportation Department's (Caltrans) traffic management district map in Figure 2.2. As shown in Figure 2.2, the study domain approximately encompasses the San Francisco Bay Area, the Sacramento Metropolitan Area, the San Joaquin Valley, Ventura County and northern region of the Los Angeles Metropolitan Area.

Figure 2.2 Study Domain

3.0 DATA USED IN STUDY

3.1 Travel Demand Network

A number of travel demand networks were provided to us during the course of the study. The first travel demand network coverage was provided to us in a standard format (i.e., Arc-Info file format) for commonly used geographic software on June 16, 2001. The Arc-Info file contained the network information with a geographical projection of the network links as Lambert Conformal Conic, which generally preserves the actual geographical contour with a slight distortion. The original file also contained important attribute information (or known as the attribute table) of the model network such as internal link ID, the co-ordinates of a link's beginning and ending point (or A-node and B-node, in technical parlance), link length, daily volume, and average speed of each link. This information is important to understanding the identity, the direction, and speed of traffic for each link in the roadway network. For example, because each link had a unique ID and A-B nodes, we could easily locate any link in the graphical representation of the model network. With the help of a local map, we were then able to match a given link in the model network to an actual traffic counter location.

In general, each link in the network has a pair of unique A-B nodes co-ordinates. When we conducted the QA/QC on the first network we received, we found many links, which were not geographically proximate, sharing identical A-B nodes coordinates. This created problems for both our subsequent matching of links with counter locations and for aggregation of links' traffic volumes. Since the links' traffic volumes in both directions need to be aggregated for analysis purpose, duplication of A-B nodes would thus produce inaccurate computation of traffic volumes for some links. In addition, there was a large disparity between some links' volumes in two directions. These disparities seemed implausible because the links at issue represented major freeways, which generally had similar daily traffic volumes in both directions. For instance, a certain link representing a segment of the Interstate 5 freeway had a daily traffic volume of 44,338 vehicles per day for one direction denoted by its A-B nodes (A8372 to B8263) but had a daily volume of 5,687 vehicles per day for the traffic in the opposite direction with corresponding B-A nodes (A3444 to B3428).

We received a revised Arc-Info file on August 4, 2001 that included a more resolute network and more detailed attribute table including traffic volumes in different modeling periods: AM, PM, Midday, and Off Peak. Many of the aforementioned problems were fixed in this coverage.

Information concerning the network coverage is presented in Table 3.1. The study network coverage was merged by Alpine Geophysics, Inc. from the statewide model network and a number of Metropolitan Planning Organizations (MPOs) travel demand model networks. For example, the network coverage of the San Francisco Bay Area was provided by the Metropolitan Transportation Commission (MTC) and the Sacramento Area of Council of Governments provided SACMET, the network coverage of the Sacramento Metropolitan Area. The network coverage of the San Joaquin Valley was provided by a number of its constituent counties transportation agencies such as Kern, Tulare, Kings, and Fresno Counties and likewise for the Los Angeles Metropolitan Area, which includes Los Angeles, Ventura, Santa Barbara, and San Bernardino.

Since the study domain does not include Orange, Riverside, and San Diego Counties, these travel model network areas were excluded, leaving a total of 122,803 CCOS links. Amongst the 122,803 links, there are 56,566 pairs of links, which represent the same segment of roadways but record the roadway traffic in opposite directions. To facilitate subsequent analysis associated with the links' model volumes, these pairs of links' volumes were aggregated without differentiation in traffic direction. These 56,566 links were merged with other 9,671 links, which represent a unique segment of roadway and whose model volumes have been aggregated from the volumes of two traffic directions. As a result, there are a total of 66,237 model links for use in subsequent analysis.

Also shown in Table 3.1, among the 66,237 CCOS links provided by Alpine Geophysics, Inc., there are 62,923 links with non-zero daily volumes. These include 28,139 links, which have volumes available for four modeling periods (i.e., AM, MD, PM, and Off-peak), 13,910 links with volumes for three modeling periods only (i.e., AM, PM, and Off-peak), and 20,874 links with daily volumes only. Of the 20,874 daily volume only links, 5,360 links of the CCOS network come from the statewide travel demand model. Daily volumes but not period-based volumes are produced for the 20,874 links; the lack of period volumes is usually associated with high computational and data costs.

3.2 Traffic Count Data Collection

Two types of traffic count data were collected. A sub-contractor was employed to collect data on roadways that were not part of Caltrans' sampling program as provided to us in late spring. We also requested and received the Caltrans' count data that were collected as part of the sampling program for the summer.

<u>Sub-Contractor Count Data</u>

A sub-contractor was employed to collect traffic count data on 120 locations in various parts of the study domain. These 120 traffic counter sites were selected in an earlier Scoping Study conducted with CARB. The locations of these sites are listed in Table 3.2. These 120 traffic counter sites are scattered among four Caltrans' traffic management districts, namely, 3, 4, 6, and 10 (See Figure 2.2). These four districts approximately cover the San Francisco Bay Area, the Sacramento Metropolitan Area, and the San Joaquin Valley. As shown in Table 3.2, there are 19, 36, 31, and 34 counter sites in each respective district (i.e., Districts 3, 4, 6, and 10.).

The primary data collection efforts were conducted by the subcontractor between July and September of 2000. The actual dates that traffic counts data were collected vary with each location (See Tables 3.2 and 3.3 for more details.). Data were collected from the starting date to the ending date for most of the locations but data were not available on some days for a few locations. Missing days were often due to malfunction of the traffic counter machines. In addition to the summer months of 2000, the subcontractor continued collecting traffic data at many locations in the San Joaquin Valley (Districts 6 and 10) through early February in 2001.

Caltrans Count Data

Traffic data were also provided by the California Transportation Department (Caltrans), which collected hourly traffic volumes at 44 locations in the study domain as part of their summer sampling program. These 44 counter sites were mainly concentrated in the San Francisco Bay Area in northern California. For example, 32 counter sites were located in District 4 while 8 were in the Sacramento Metropolitan Area and the San Joaquin Valley (See Table 3.3). Of the 44 locations, 23 locations were also collected by the sub-contractor. Accounting for the 23 duplicate sites yields a total of 141 locations for which observed hourly traffic volumes are available. The 141 counter sites are plotted using the traffic network in Figure 3.1 and the network resolution of three main study areas are magnified for easier viewing in Figures 3.1(a)-3.1(c).

3.3 Quality Assurance/ Quality Control Analysis

3.3.1 Issues with the Data

The raw traffic data for the 141 counter locations were stored in a combination of excel or text files. Each file generally contained a week's hourly traffic volumes at each location for one direction of travel. Given that there were 8 weeks' observed traffic volumes and two directions of traffic per location, there were a large number of files to process.

The data processing task was further hampered by the irregular formats and other data problems. For example, the hourly volumes and site ID's were stored in different formats in excel and text files. In fact, while the text files could be sorted and processed using a code written in C++ programming language, data stored in excel files had to be manually sorted and extracted.

Other data problems also included a considerable number of files that contained site ID's that did not match their purported locations. In some files, traffic volumes were assigned to four different traffic directions for the same location. For these reasons, most processed files had to be carefully checked manually.

Other Quality Assurance (QA) and Quality Control (QC) measures were also developed to handle minor input errors. For example, hourly traffic volumes for some locations were double-counted on certain dates and averaged afterwards. It also happened that the dates recorded for data for certain locations were implausible. These dates were corrected by a reasonable approximation after the complete data archives for affected locations were checked. Hourly traffic volumes were also reviewed by day of week, date, and by location to identify any anomalies that might have resulted from tube or counter malfunctions.

However, we did not replace the zero hourly volumes that were recorded on a number of locations for a varied length of time (See Table 3.4) for two reasons: 1) these zero hourly volumes recordings were likely meaningful due to the overall low traffic volumes on some locations (e.g., Sites 670 and 671), 2) discarding zero volumes data together with the locations' good hourly volumes data for other periods of time deplete the sample size for reliable estimation.

3.3.2 Missing Data

Before the models/allocation factors can be estimated, the 141 counter sites need to be matched to the corresponding links of the CCOS travel demand model network. Based on the site location information shown in Tables 3.2 and 3.3, we were able to match network links to 135 actual traffic counter sites using local street maps. Six sites could not be matched to the model network links with reasonable confidence because of insufficient location information or the low resolution of the model network in some areas (These six sites' ID's are 400, 443, 444, 524, 1164, and 1611, respectively).

The proportion of observed volumes relative to total network links is less than 1% (0.2%) or 135 out of 66,237 links (See the lower panel of Table 3.1). This low number of observed count locations implies that hourly volumes for 99.8% (or 66,102) of the links in the model network need to be predicted. It is also important to note that the 141 sample locations were chosen specifically to maximize generalizability and representativeness.

For reasons that will be discussed in detail in subsequent sections, the model volume resolution allows easier identification with observed traffic data in the case of matched links and likewise with the matched links' estimated allocation factors in the case of unmatched links. In the original application of the method using the network covering the Los Angeles and San Diego areas with more than 1800 observed volume locations, hourly volumes were predicted for the unmatched links when traffic volumes were available for at least three periods of a day (e.g., AM, PM, and off-peak).

However, one of the problems we did not anticipate with CCOS was a lack of period-based volumes for links in the CCOS network provided by Alpine Geophysics, Inc. As will be seen later, only 8 of 135 matched links have volumes for 4 periods of a day (i.e., AM, Mid-day, PM, Off-peak) and 23 links have volumes for 3 periods of a day (e.g., AM, PM, and Off-peak). The remaining 104 matched links have only aggregate daily volume (i.e., 24-hour). The lack of period-based volumes reduces the representativeness of the estimated allocation factors for the matched links. Since the unmatched links' predicted hourly volumes are based on the matched links' allocation factors, these are also less accurate. Fortunately, as well be seen, this prediction problem is somewhat mitigated by the fact that 63.6% (or 42,049) of the 66,102 unmatched links have volumes for at least three periods of day (i.e., AM, PM, and Off-peak).

Table 3.1 Descriptive Summary of CCOS Travel Demand Network Links

Network	Link Counts	Links with zero daily volumes	Links with non-zero daily volumes
Statewide	5,360	249	5,111
Kern	10,111	484	9,627
Tulare	9,663	344	9,319
Kings	4,367	289	4,078
Fresno	17,009	755	16,254
Los Angeles	18,879	124	18,755
Ventura	2,140	48	2,092
Santa Barbara	5,596	1,401	4,195
San Bernardino	6,627	198	6,429
Orange	6,189	51	6,138
Riverside	6,991	133	6,858
San Diego	48,263	8,019	40,244
MTC	27,432	3,641	23,791
SACMET	15,627	148	15,479
Total	184,254	15,884	168,370
Total excluding Orange, San Diego, Riverside[1]	122,803		
Network Information Pertinent to Analysis			
1. No. of links showing one direction[2]	9,671		
2. No. of links showing two directions[3]	56,566 (x 2)		
3. No. of links considered in analysis	66,237[4]		
4. No. of links with volumes for 4 periods (AM, MD, PM, Off-Peak)	28,139		
5. No. of links with volumes for 3 periods (AM, PM, Off-Peak)	13,910		
6. No. of links with daily volumes only	20,874		
7. No. of links with non-zero daily volumes	62,923[5]		
8. No. of links with matching count data	135		
9. No. of links without matched count data	66,102[6]		

[1] Orange, San Diego, and Riverside are not included in the study domain.
[2] These links' A-B nodes do not have corresponding links with A-B nodes in opposite direction.
[3] These links' A-B nodes have corresponding links with A-B nodes in opposite direction. The daily volumes on both corresponding links are combined for subsequent analyses.
[4] It is obtained by adding the number of links showing one direction and the number of links showing two directions multiplied by two, i.e., 9,671+56,566x2=66,237. [5] It is the sum of (4) No. of links with volume for 4 periods, (5) No. of links with volumes for 3 periods, and (6) No. of links with daily volumes only.
[6] It is the difference between (3) No. of links considered in analysis and (8) No. of links with matching count data.

Table 3.2 Traffic Count Data Collected by the Sub-contractor

(The 23 locations that both subcontractor and Caltrans sampled are shaded in Tables 3.2 and 3.3.)

Site ID	District	County	Location	Date	Missing Data
254	3	YOL	MOSSY CREEK BRIDGE	7/20/00-9/1/00	
259	3	YOL	WEST MAIN ST/COUNTY ROAD 98	7/19/00-9/1/00	
306	3	BUT	JCT RTE 191 NORTH	7/19/00-8/31/00	
307	3	BUT	PULGA-MILL CREEK MAINT STATIO	7/20/00-8/31/00	
308	3	SUT	JCT RTE 99	7/19/00-8/31/00	
335	3	YOL	WEST SACRAMENTO JCT RTE 50	7/19/00-9/1/00	
358	3	BUT	BIGGS HIGHWAY	7/19/00-8/31/00	
360	3	BUT	JCT RTE 162 WEST	7/12/00-8/31/00	
361	3	BUT	JCT RTE 162 EAST	7/18/00-8/31/00	
393	3	SUT	JCT RTE 113	7/18/00-8/31/00	
395	3	SUT	BARRY ROAD	7/18/00-8/31/00	
397	3	SUT	YUBA CITY LINCOLN ROAD	7/20/00-9/1/00	
399	3	SUT	FRANKLIN ROAD	7/18/00-8/31/00	
400	3	SUT	YUBA CITY BRIDGE STREET	7/18/00-8/31/00	
409	3	SAC	RANCHO SECO ROAD	7/19/00-9/3/00	
417	3	SUT	SUTTER CAUSEWAY BRIDGE	7/19/00-8/31/00	
422	3	YOL	COUNTY ROAD 18C	7/19/00-9/1/00	
443	3	SAC	RYDE JCT RTE 220 WEST	7/19/00-9/3/00	
444	3	SAC	PAINTERSVILLE BR RIVER ROAD	7/19/00-9/3/00	
492	4	CC	BRENTWOOD CHESTNUT STREET	7/19/00-9/1/00	
493	4	CC	JUNCTION BRYON HIGHWAY	7/19/00-8/31/00	
513	4	NAP	.2-MI N/O NAPA/SOL CNTY LINE	7/19/00-9/1/00	
515	4	ALA	JCT RTE 123 SAN PABLO AVENUE	7/18/00-9/1/00	
531	4	SOL	JCT RTE 37 MARINE WORLD PKWY	7/16/00-9/1/00	
533	4	SOL	JCT RTE 29 SONOMA BOULEVARD	7/19/00-9/1/00	
535	4	ALA	ALAMEDA ISLAND DRIVE	7/19/00-9/1/00	
571	4	SCL	LAWRENCE EXPRESSWAY	7/15/00-8/31/00	
573	4	SCL	PALO ALTO EMBARCADERO ROAD	7/17/00-8/31/00	
605	4	SM	JCT RTE 84 WOODSIDE ROAD	7/15/00-8/31/00	

Table 3.2 (continued) Traffic Count Data Collected by the Sub-contractor

607	4	SM	SAN MATEO JCT RTE 92	7/15/00-8/31/00	
608	4	SM	BURLINGAME PENINSULA AVENUE	7/15/00-9/1/00	
609	4	SM	MILLBRAE CENTER STREET	7/15/00-9/1/00	
610	4	SM	S SAN FRANCISCO HICKEY BLVD	7/17/00-8/31/00	
611	4	SM	DALY CITY HILLSIDE BOULEVARD	7/15/00-8/31/00	
630	4	ALA	FREMONT SOUTH JCT RTE 238	7/21/00-8/31/00	
631	4	ALA	FREMONT NORTH JCT RTE 238	7/19/00-8/31/00	
632	4	ALA	NORTH JCT RTE 680	7/14/00-9/3/00	9/2/00-9/3/00
633	4	ALA	SCOTTS CORNER S JCT RTE 680	7/14/00-8/31/00	
634	4	ALA	LIVERMORE STANLEY BOULEVARD	7/14/00-8/31/00	
635	4	ALA	LIVERMORE WEST JCT RTE 580	7/21/00-9/5/00	9/2/00-9/4/00
649	4	SM	SAN GREGORIO JCT RTE 1	7/19/00-9/3/00	
651	4	SM	WOODSIDE JCT RTE 35	7/19/00-9/18/00	
652	4	SM	WOODSIDE JCT RTE 280	7/19/00-9/6/00	9/2-9/5/00
654	4	SM	REDWOOD CITY S JCT RTE 101	7/16/00-9/14/00	7/30-8/5
669	4	SOL	JCT RTE 12 RIO VISTA	7/17/00-9/18/00	9/2-9/7
670	4	SOL	CACHE SLOUGH FERRY	7/17/00-9/18/00	9/2-9/7
671	4	SOL	JCT RTE 220	7/17/00-9/18/00	9/2-9/7
679	4	SM	HALF MOON BAY JCT RTE 1	7/19/00-9/13/00	8/11-8/14,8/17-8/18
681	4	SM	JCT RTE 35 SOUTH	7/19/00-8/26/00	8/11-8/14
682	4	SM	JCT RTE 35 NORTH	7/15/00-8/31/00	
696	4	ALA	JCT RTE 880 DAVIS STREET EAS	7/16/00-9/1/00	
733	4	ALA	FREMONT JCT RTE 680 MISSION R	7/20/00-8/31/00	
734	4	ALA	N JCT RTE 84 NILES CANYON RD	7/22/00-8/31/00	8/10-8/45
735	4	ALA	UNION CITY DECOTO ROAD	7/19/00-7/29/00	
736	4	ALA	HAYWARD JCT RTES 92 AND 185	7/17/00-8/31/00	
994	6	FRE	MERCED AVENUE	7/10/00-2/4/01	1/26/01-2/1/01
997	6	FRE	NORTH JCT RTE 198	7/10/00-2/3/01	
1006	6	KIN	JCT RTE 198	7/15/00-2/2/01	8/17-8/24,9/23-9/28
1007	6	KIN	10TH AVENUE	7/15/00-2/2/01	8/18-8/25,1/12-1/17/01

Table 3.2 (continued) Traffic Count Data Collected by the Sub-contractor

1026	6	TUL	TULARE JCT RTE 137	7/15/00-1/31/01	8/23-8/25
1030	6	TUL	AVENUE 352 OIL WELL ROAD	7/14/00-9/4/00	
1031	6	TUL	AVENUE 416 EL MONTE WAY	7/14/00-2/1/01	
1032	6	TUL	AVENUE 432	7/14/00-2/1/01	
1072	6	TUL	TULARE SOUTH WEST STREET	7/15/00-1/31/01	
1098	6	TUL	BALCH PARK DRIVE	7/15/00-1/31/01	
1110	6	KIN	MAIN GATE LEMOORE NAV AIR ST	7/15/00-2/2/01	
1115	6	TUL	JCT RTE 65 SOUTH	7/15/00-2/2/01	
1133	6	TUL	SOUTH JCT RTE 63	8/1/00-2/1/01	
1134	6	TUL	JCT RTE 245 STAFFORDS CORNER	8/1/00-2/1/01	
1141	6	TUL	WOODLAKE JCT RTE 245	7/14/00-2/1/01	10/9-10/16
1142	6	TUL	JCT RTE 198	7/14/00-2/1/01	1/25-2/1/01
1153	6	TUL	JCT RTE 216 NARANJO BOULEVAR	7/14/00-2/1/01	1/10-1/11/01
1154	6	TUL	JCT RTE 201 W STAFFORDS CORNE	7/14/00-2/1/01	
1158	6	FRE	AVENAL CUTOFF ROAD	7/13/00-2/2/01	
1159	6	FRE	JCT RTE 145	7/13/00-2/2/01	
1164	6	KIN	JCT RTE 33	7/13/00-2/2/01	
1613	10	SJ	STOCKTON SOUTH JCT RTE 5	7/20/00-1/26/01	9/1-11/30
1614	10	SJ	STOCKTON AIRPORT WAY	7/20,7/29,8/13/,8/20, 8/31,12/1, 12/9, 12/16,12/23,12/30, 1/12/01,1/20, 1/26	All other dates
1646	10	CAL	VALLEY SPRINGS JCT RTE 26	7/20/00-1/31/01	9/1-11/27
1647	10	CAL	TOYON JCT RTE 26	7/20/00-1/31/01	9/1-11/28
1656	10	SJ	LOWER SACRAMENTO ROAD	7/20/00, 8/4, 8/18, 8/25, 8/31,12/1, 12/9, 12/16, 12/26, 12/29, 1/5/01, 1/12, 1/26	All other dates
1661	10	SJ	LODI NORTH JCT RTE 99	7/29/00, 8/4, 8/18, 8/31, 12/1, 12/9, 12/16, 12/22, 12/29, 1/5, 1/12/01, 1/19,-1/26	All other dates

Table 3.2 (continued) Traffic Count Data Collected by the Subcontractor

1662	10	SJ	CLEMENTS EAST JCT RTE 88	7/29/00, 8/4, 8/11, 8/25, 8/31, 12/1, 12/9, 12/15, 12/22, 12/29, 1/5/01, 1/12, 1/19, 1/26	All other dates
1663	10	SJ	LOCKEFORD WEST JCT RTE 88	7/20/00-1/31/01	
1664	10	SJ	SACRAMENTO-SAN JOAQUIN COUNTY	7/20/00-1/31/01	9/1-11/30
1676	10	AMA	JCT RTE 88	7/20/00-1/28/01	1/24-1/28
1678	10	SJ	HWY26 A JENNY LIND RD	11/27/00-2/4/01	11/27
1679	10	CAL	VALLEY SPRINGS W JCT RTE 12	7/20/00-1/31/01	9/1-11/28, 1/10-1/11/01
1694	10	SJ	STOCKTON JCT RTE 4	7/20/00-1/31/01	9/1-11/30
1698	10	SJ	LINDEN FLOOD RD/FRONT ST	7/21/00-1/31/01	9/1-11/27
1711	10	AMA	DEPOT ROAD	7/20/00-2/1/01	9/1-11/30
1731	10	CAL	MOKELUMNE HILL JCT RTE 26	7/20/00-1/31/01	9/1-11/30
1748	10	AMA	JCT RTE 124	7/20/00-1/31/01	9/1-11/30
1756	10	AMA	JCT RTE 26 RED CORRAL ROAD	7/20/00-1/28/01	9/1-11/30, 1/25-1/28
1773	10	SJ	EIGHT MILE ROAD	7/20/00-1/31/01	9/1-11/30
1777	10	SJ	CLEMENTS JCT RTE 12 EAST	7/20/00-1/31/01	9/1-11/30
1814	10	AMA	MICHIGAN BAR ROAD	7/20/00-2/1/01	9/1-11/28
1816	10	AMA	IONE JCT RTE 124 NORTH	7/20/00-1/31/01	9/1-11/30
1818	10	AMA	IONE JCT RTE 124 SOUTH	7/20/00-1/31/01	9/8-11/28
1838	10	AMA	JCT RTE 88	7/20/00-2/1/01	7/30-11/28
3001	6	TUL	RD 80 /AVE 328	8/7/00-2/1/01	8/25-8/29, 1/21-1/22/01
3002	6	TUL	RD 80 /AVE 384	8/7/00-2/1/01	10/27, 11/13-11/14
3003	6	TUL	RD 204 /SHWY198	8/7/00-10/10/00	8/26-8/28
3004	6	TUL	RD 108 /AVE 264	8/9/00-1/31/01	
3005	6	KIN	AVENAL CUTOFF E /NEVADA AVE	8/8/00-9/24/00	
3006	6	KIN	AVENAL CUTOFF RD /PLOYMOUTH AVE	8/8/00-2/2/01	
3007	6	TUL	LOVERS LN N /AVE 256	8/9/00-1/30/01	1/12
3008	6	TUL	AVE 416E /KINGS RIVER	8/9/00-2/1/01	
3009	6	TUL	GRANGEVILLE BOULEVARD E/22ND AVE	8/8/00-2/2/01	9/12-9/13

Table 3.2 (continued) Traffic Count Data Collected by the Sub-contractor

3010	6	KIN	CALDWELL FROM MOONEY BLVD TO COURT STREET	8/9/00-1/31/01	
4001	10	STA	CLARIBEL RD BETWEEN LANGWORTH RD & WELLSFORD RD	7/27/00-1/31/01	9/1-11/30
4002	10	STA	LADD ROAD W/O AMERICAN	7/30/00-2/2/01	9/1-11/30
4003	10	STA	YOSEMITE AVENUE BETWEEN H AND I STREET	7/27/00-1/31/01	9/1-11/30
4004	10	STA	OAKDALE RD. N/O MORRIL RD	7/27/00-1/31/01	9/1-11/30
4005	10	STA	CLAVS ROAD S/OCLARIBEL ROAD	7/27/00-1/31/01	9/1-11/30
4006	10	STA	COFFEE RD. N/0 MORILL RD.	8/6/00-2/2/01	9/1-11/30
4007	10	STA	SISK RD N/O WALLASEY RD	7/29/00-1/31/01	9/1-11/30
4008	10	STA	CARVER RD S/O LADD RD	7/30/00-2/2/01	9/1-11/30
4009	10	STA	ROSELLE AVE. N/O CLARIBEL RD	7/27/00-1/31/01	9/1-11/30
4010	10	STA	TERMINAL AVE. S/O CLARIBEL RD	8/3/00-1/31/01	9/1-11/30
Total Observations			120		

Note: The following locations with site ID's cannot be matched to links in the travel demand model network: 400, 443, 444, 524, 1164, and1611.

Table 3.3 Traffic Count Data Collected by Caltrans

Site ID	District	County	Location	Dates	Missing Dates
242	3	SAC	045SACRAMENTO, DEL PASO ROAD	7/01//00-9/30/00	7/3-7/12, 8/8-9/8
360	3	BUT	573JCT. RTE. 162 WEST	7/01/00-7/31/00	
364	3	BUT	306JCT. RTE. 149 SOUTHEAST	6/01//00-9/30/00	
427	3	YOL	719WOODLAND, MAIN STREET	7/01//00-9/30/00	
492	4	CC	540BRENTWOOD, CHESTNUT STREET	7/08/00-7/15/00	
493	4	CC	123BYRON HIGHWAY	7/08/00-7/16/00	
495	4	CC	541MC EWEN ROAD	7/07/00-7/14/00	
496	4	CC	523MARTINEZ, ALHAMBRA BOULEVARD	7/07/00-7/14/00	
500	4	CC	121WILLOW PASS ROAD	7/25/00-08/01/00	
502	4	CC	122ANTIOCH, A STREET/LONETREE	7/18/00-7/25/00	
515	4	ALA	125JCT. RTE. 123	8/04/00-8/11/00	
517	4	ALA	027OAKLAND, JCT. RTE. 580	8/04/00-8/11/00	
518	4	ALA	124OAKLAND, MORAGA AVENUE	8/04/00-8/11/00	
524	4	ALA	130OAKLAND, JCT RTE 580	7/11/00-7/18/00	
535	4	ALA	131ALAMEDA, ISLAND DRIVE	8/03/00-8/09/00	
558	4	SOL	330VALLEJO, JCT. RTE. 37 WEST	9/06/00-9/13/00	
566	4	SOL	342JCT. RTE. 113 SOUTH	9/06/00-9/13/00	
571	4	SCL	181LAWRENCE EXPRESSWAY	7/07/00-7/13/00	
605	4	SM	246JCT. RTE. 84	9/07/00-9/14/00	
607	4	SM	054SAN MATEO, JCT. RTE. 92	8/05/00-8/13/00	
608	4	SM	185BURLINGAME, PENINSULA AVENUE	9/08/00-9/18/00	
610	4	SM	187SOUTH SAN FRANCISCO, HICKEY	9/08/00-9/18/00	
630	4	ALA	136FREMONT, JCT. RTE. 238	9/06/00-9/13/00	
637	4	ALA	135NEWARK, SOUTH JCT. RTE. 880	9/06/00-9/13/00	
649	4	SM	055SAN GREGORIO, JCT. RTE. 1	9/09/00-9/18/00	
651	4	SM	190WOODSIDE, JCT. RTE. 35	9/09/00-9/18/00	
652	4	SM	202WOODSIDE, JCT. RTE. 280	9/07/00-9/14/00	
669	4	SOL	344WEST JCT. RTE. 12	9/09/00-9/18/00	
670	4	SOL	345CACHE SLOUGH FERRY	9/09/00-9/18/00	

Table 3.3 (Continued)   Traffic Count Data Collected by Caltrans

671	4	SOL	346JCT. RTE. 220 EAST	9/09/00-9/18/00	
675	4	ALA	032HAYWARD, JCT. RTES. 185/238	8/03/00-8/11/00	
679	4	SM	195HALF MOON BAY, JCT. RTE. 1	8/03/00-8/10/00	
681	4	SM	057JCT. RTE. 35 SOUTH	8/03/00-8/10/00	
682	4	SM	196JCT RTE 35 NORTH	8/03/00-8/10/00	
683	4	SM	197JCT. RTE. 280	8/03/00-8/09/00	
685	4	SM	199SAN MATEO, JCT. RTE. 101	8/03/00-8/10/00	
686	4	SM	910FOSTER CITY BOULEVARD	7/09/00-8/19/00	
696	4	ALA	141JCT. RTE. 880	8/08/00-8/15/00	
701	4	SM	191MENLO PARK, JCT. RTE. 101	8/07/00-8/15/00	
704	4	SM	191MENLO PARK, JCT. RTE. 101	8/08/00-8/15/00	
1611	10	SJ	054ROBERTS ISLAND ROAD	7/26/00-12/27/00	11/27/00
1663	10	SJ	123LOCKEFORD, JCT. RTE. 88	7/26/00-12/31/00	10/30/00
1674	10	AMA	084JCT. RTE. 124 SOUTH	7/01/00-9/21/00	
1676	10	AMA	316JCT. RTE. 88	7/01/00-12/27/00	10/30/00
Total Observations			44		

Note: The following locations with site ID's cannot be matched to links in the travel demand model network: 400, 443, 444, 524, 1164, and1611.

Table 3.4 List of Zero Hourly Volumes by Site ID

Site ID	Zero Hourly Volumes		
	Date	Day of Week	Time of Day
242	---	---	--
254	---	---	---
259	7/19/00, 8/2/00	Wed, Wed	12AM-5PM, 12AM-3PM
306	7/19/00	Wed	12AM-12PM
307	---	---	---
308	7/19/00	Wed	12AM-1PM
335	---	---	---
358	7/19/00	Wed	12AM-1PM
360	6/2/00	Fri	12AM-7AM
361	---	---	---
364	---	---	---
393	7/18/00	Tue	12AM-2PM
395	7/18/00	Tue	12AM-11PM
397	---	---	---
399	---	---	---
400	---	---	---
409	7/19	Wed	12AM-2PM
417	---	---	---
422	8/3/00	Thur	12AM-1PM
427	---	---	---
443	8/3/00	Thur	12AM-10AM
444	---	---	---
492	7/19/00	Wed	12AM-2PM
493	---	---	---
495	7/6/00	Thur	12AM-2PM
496	---	---	---
500	---	---	---
502	---	---	---
513	7/19/00	Wed	12AM-2PM
515	---	---	---
517	---	---	---
518	---	---	---
524	---	---	---
531	---	---	---
533	7/19/00,7/24	Wed, Mon	12AM-2PM, 12AM-12PM
535	7/19/00	Wed	12AM-2PM
558	---	---	---
566	---	---	---
571	---	---	---
573	---	---	---

Table 3.4(continued) List of Zero Hourly Volumes by Site ID

605	---	---	---
607	---	---	---
608	---	---	---
609	---	---	---
610	7/26/00	Wed	12AM-2PM
611	8/6/00	Sun	12AM-1PM
630	---	---	---
631	7/19/00	Wed	12AM-2PM
632	---	---	---
633	---	---	---
634	---	---	---
635	9/5/00	Tue	12AM-11AM
637	---	---	---
649	7/25,8/30,9/8	Tue, Wed, Fri	2AM-3AM;1AM, 3AM;12AM-10AM
651	---	---	---
652	7/19, 9/6	Wed, Wed	12AM-9AM, 12AM-11AM
654	---	---	---
669	8/16-8/17	Wed-Thur	12AM-5AM, 12AM-5AM
670	7/18,7/21, 7/24-7/28,7/31-8/9, 8/11,8/12, 8/14-8/15, 8/19-8/20, 8/22-8/25, 8/28, 8/31-9/1, 9/9, 9/12, 9/14-9/15, 9/17-9/18	All days of week	Mostly between 12AM-6AM
671	7/18-7/21, 7/23-7/29,7/31-8/10, 8/11-8/12,8/14-8/25,8/29, 8/31, 9/1, 9/9-9/10, 9/12-9/17	All days of week	Mostly between 12AM-6AM
675	---	---	---
679	8/19	Sat	12AM-10AM
681	---	---	---
682	---	---	---
683	---	---	---
685	---	---	---
686	---	---	---
696	---	---	---
701	---	---	---
704	---	---	---
733	---	---	---
734	---	---	---
735	7/28-7/29	Fri-Sat	12AM-11AM

Table 3.4(continued) List of Zero Hourly Volumes by Site ID

736	7/17,7/27	Mon,Thur	12AM-1PM, 12AM-2PM
994	10/12/00	Thur	12AM-2PM
997	---	---	---
1006	8/25, 9/28	Fri, Thur	12AM-3PM,12AM-1PM
1007	1/12/01,1/18/01	Fri, Thur	12AM-12PM
1026	8/25	Fri	12AM-4PM
1030	---	---	---
1031	---	---	---
1032	12/11/-12/15/00,12/24, 12/26-12/27	Mon-Fri, Sun, Tue-Wed	Mostly between 12AM-6AM
1072	---	---	---
1098	8/1,8/13, 8/15,8/25,8/31-9/1, 9/6,9/7,9/19, 10/2, 10/10-10/11,10/13, 10/24, 10/26, 10/30, 11/3, 11/7-11/9, 11/13-, 11/16, 11/27, 12/1, 12/4, 12/11-12/15, 12/24, 12/26-12/27, 1/2/01, 1/4/01, 1/5/01, 1/8-1/10, 1/14-1/17, 1/21-1/22, 1/24-1/25, 1/27-1/28, 1/31	All days of week	Mostly between 12AM-6AM and 12AM-12PM
1110	---	---	---
1115	2/2/01	Fri	12AM-2PM
1133	---	---	---
1134	12/22	Fri	2AM-3AM, 11AM
1141	1/11-1/16	Thur-Tue	12AM-8AM, 12AM-2PM
1142	7/14,9/5,9/8, 9/11,9/21-9/23, 10/2, 10/11-10/12, 10/21, 10/23, 10/27-10/28, 11/1, 11/5-11/6, 11/8-11/10, 11/13, 11/15, 11/17, 11/20-11/22, 11/24, 11/28, 12/3-12/5, ½, 1/11-1/15, 1/17, 1/23/01	All days of week	12AM-5AM
1153	---	---	---

Table 3.4(continued) List of Zero Hourly Volumes by Site ID

1154	7/14, 7/17, 8/13, 8/16, 8/24-8/25, 8/28, 9/4, 9/6, 9/8-9/9, 9/12-9/14, 9/20, 9/29, 10/4,-10/6, 10/11, 10/13-10/14, 10/16-10/18, 10/20, 10/24-10/25, 10/27, 10/30-11/10, 11/13-11/16, 11/27-1130, 12/5-12/6, 12/8, 12/11-12/14, 12/16, 12/21-12/22, 12/31, 1/7/01, 1/9-1/10/01, 1/14-1/15, 1/18, 1/22-1/26, 1/29-2/1/01	All days of week	Mostly between 12AM-3AM
1158	---	---	---
1159	7/29, 1/1/01	Sat, Tue	12AM-12PM, 1AM-4AM
1164	---	---	---
1611	7/25	Tue	12AM-12PM
1613	7/20, 1/6/01	Thur, Thur	12AM-1PM, 12AM-4PM
1614	7/29, 8/20, 12/9, 12/16,12/23, 12/30, 1/26/01	Sat, Sun, Sat, Sat, Sat, Sat, Fri,	Mostly between 12AM-3PM
1646	---	---	---
1647	1/24/01, 1/25/01	Wed, Thur	7AM-12AM, 12AM-4PM
1656	12/9, 12/16, 12/29, 1/5/01, 1/12/01	Sat, Sat, Fri, Fri, Fri	12AM-1PM, 12AM-11AM, 12AM-4PM, 12AM-3Pm, 12AM-12PM
1661	12/9, 12/16, 12/22, 12/29, 1/5/01, 1/12/01, 1/19/01, 1/26/01	Sat, Sat, Fri (x 6)	12AM-12PM, 12AM-3PM, 12AM-3PM, 12AM-11PM, 12AM-2PM, 12AM-12PM
1662	12/9, 12/15, 12/22, 12/29, 1/5/01, 1/12/01, 1/19, 1/26,	Sat, Fri (x 7)	12AM-6PM, 12AM-2PM, 12AM-3PM, 12AM-11AM, 12AM-11AM, 12AM-1PM, 12AM-11PM
1663	---	---	---
1664	---	---	---
1674	---	---	---
1676	---	---	---
1678	1/28-1/30/01, 2/1/01	Sun-Tue, Thur	12AM-6AM, 12AM-9AM, 6PM-5AM
1679	1/16-1/18/01	Tue-Thur	1AM-6AM, 12AM-4AM, 12AM-5AM
1694	1/6/01, 1/12	Sat, Fri	12AM-2PM, 12AM-2PM

Table 3.4(continued) List of Zero Hourly Volumes by Site ID

1698	---	---	---
1711	---	---	---
1731	---	---	---
1748	8/25	Fri	12AM-1PM
1756	7/20	Thur	12AM-10AM
1773	--	---	---
1777	---	---	---
1814	---	---	---
1816	7/20	Thur	12AM-4PM
1818	---	---	---
1838	---	---	---
3001	8/7, 8/29, 1/23/01, 1/28/01	Mo, Tue, Tue, Sun	1AM-5AM, 10AM; 12AM-12PM, 12am-12PM, 1AM-4AM, 10AM, 12PM
3002	10.4, 11/15, 1/9	Wed, Wed, Tue	12AM-12PM; 12AM-1PM; 12AM-4AM, 1PM
3003	8/29,	Tue	12AM-2PM
3004	12/29	Fri	12AM-6AM; 12PM
3005	---	---	---
3006	---	---	---
3007	10/13, 10/30	Fri, Mon	12AM-4PM, 12AM-11AM
3008	---	---	---
3009	8/31, 9/14, 10/11	Thur, Thur, Wed	12AM-12PM, 12AM-11AM, 12AM-5AM
3010	---	---	---
4001	11/12, 7/27, 1/20	Wed, Thur, Sat	12AM-5PM, 12AM-7PM, 12AM-3PM
4002	7/30	Sun	12AM-7PM
4003	7/27	Thur	12AM-6AM
4004	7/27, 12/11, 1/1/01	Thur, Mon, Mon	12AM-8AM, 12AM-4PM, 12AM-3PM
4005	7/27	Thur	12AM-6PM
4006	8/6, 12/30	Sun, Sat	12AM-8PM, 12AM-5PM
4007	7/29, 8/6	Sat, Sun	12AM-6PM, 12AM-6PM
4008	7/30, 1/11	Sun, Thur	12AM-7PM, 1AM
4009	7/27	Thur	12AM-8PM
4010	8/3	Thur	12AM-8PM

Figure 3.1 Distribution Map of Traffic Counter Sites

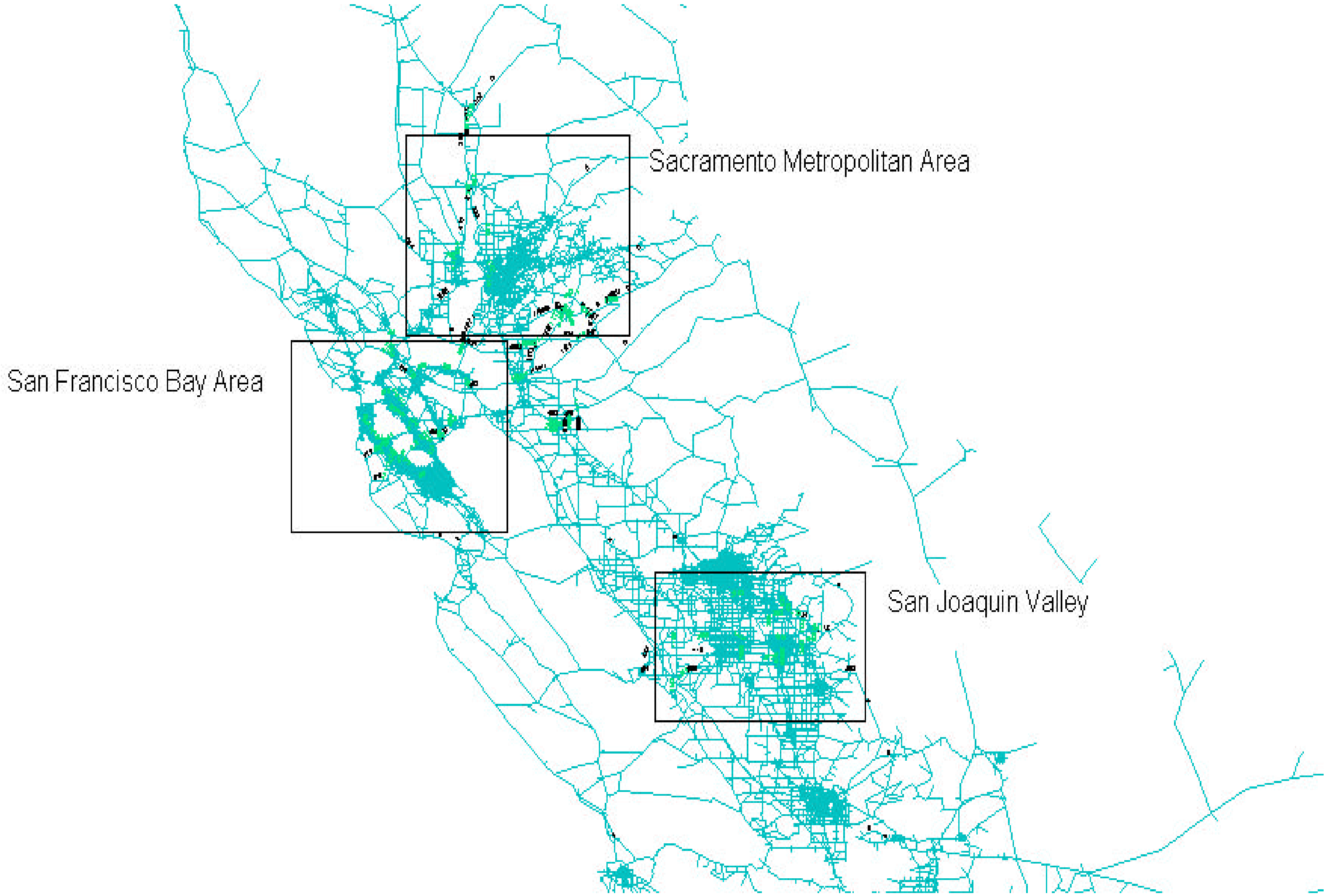


Figure 3.1(a) Location of Traffic Counter Sites (by Site ID) in the Sacramento Metropolitan Area
(Counter locations with higher average daily volumes on weekend than on weekdays are marked with a circle)

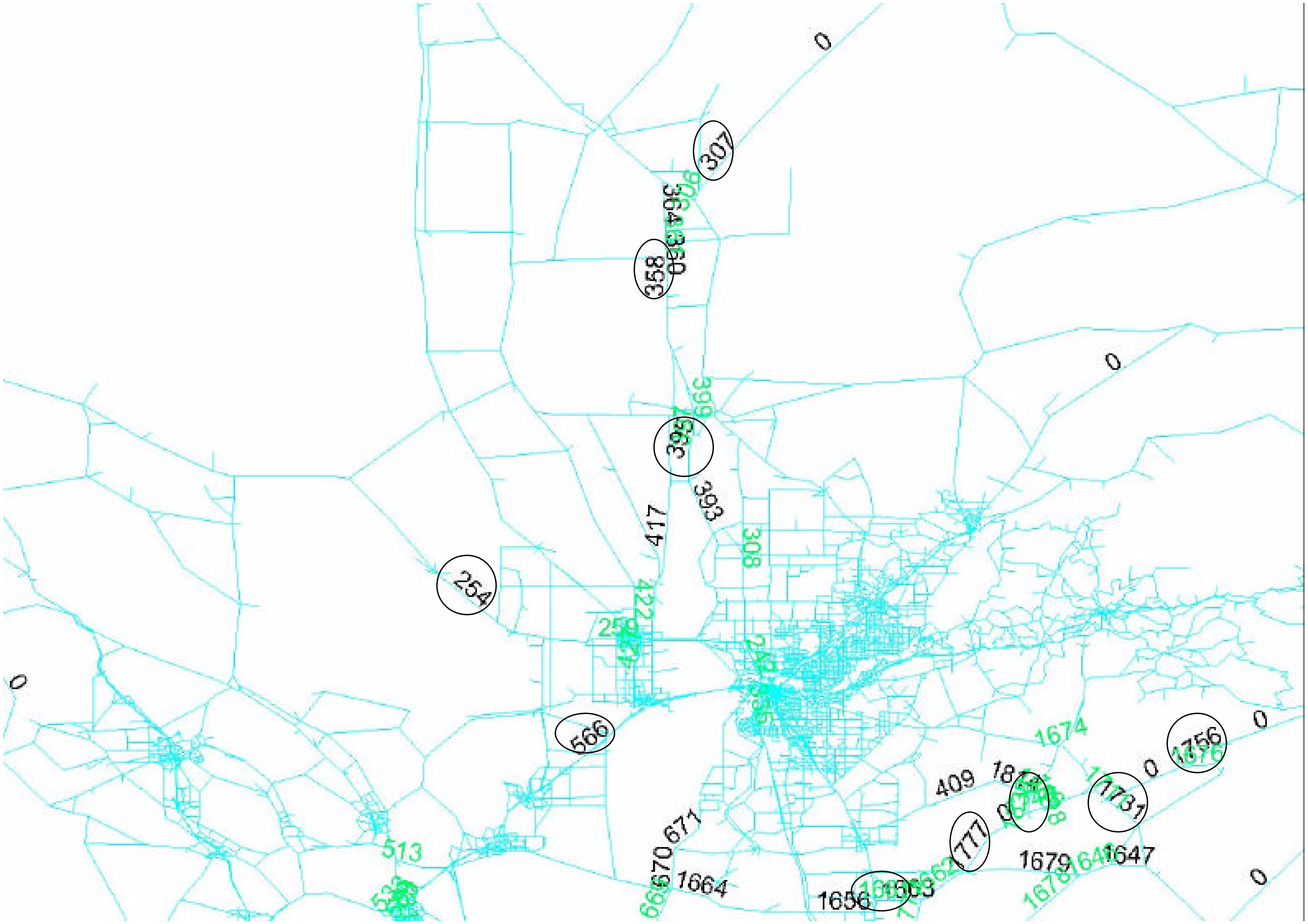


Figure 3.1(b) Location of Traffic Counter Sites (by Site ID) in the San Francisco Bay Area

(Counter locations with higher average daily volumes on weekend than on weekdays are marked with a circle)

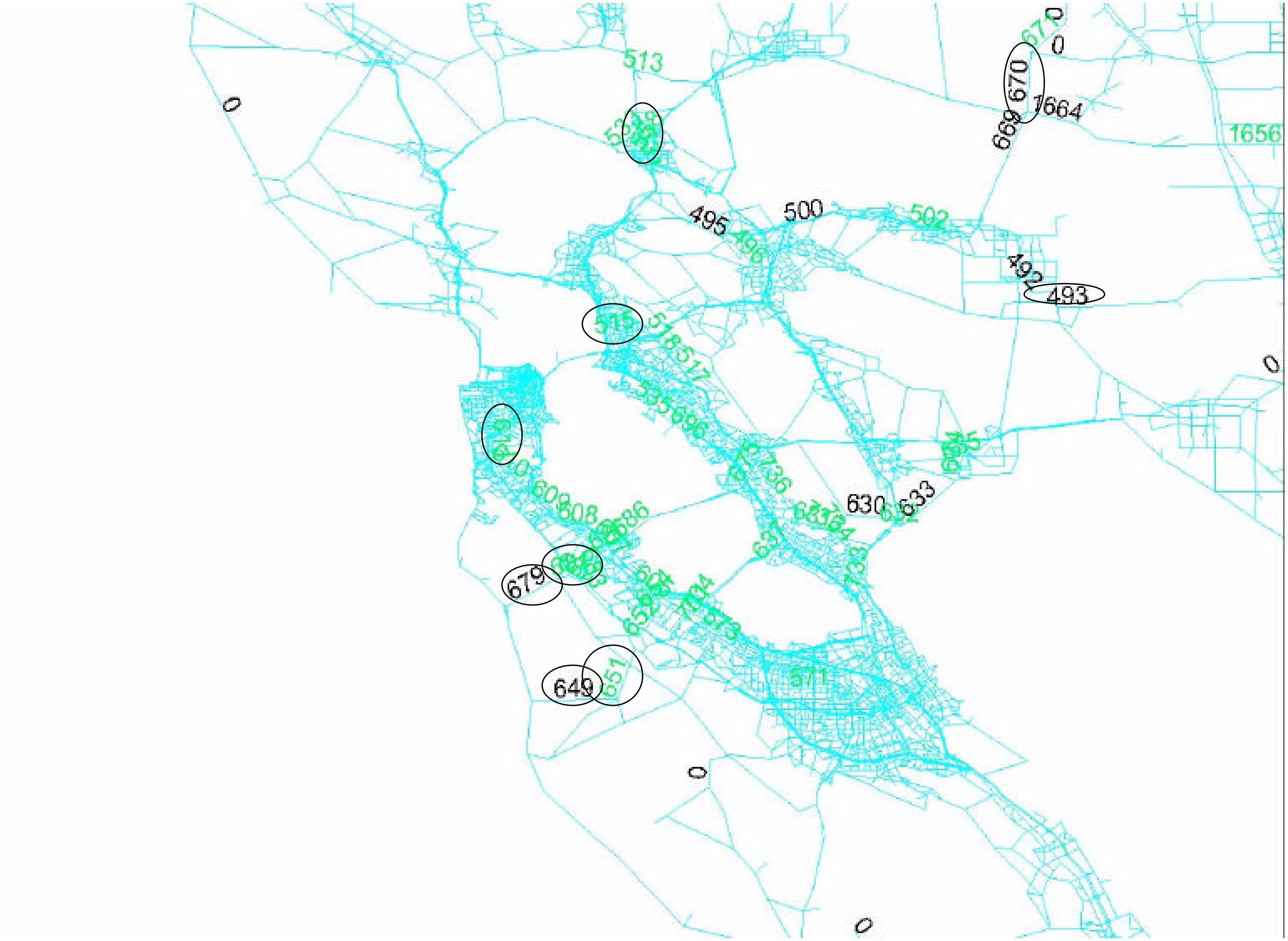
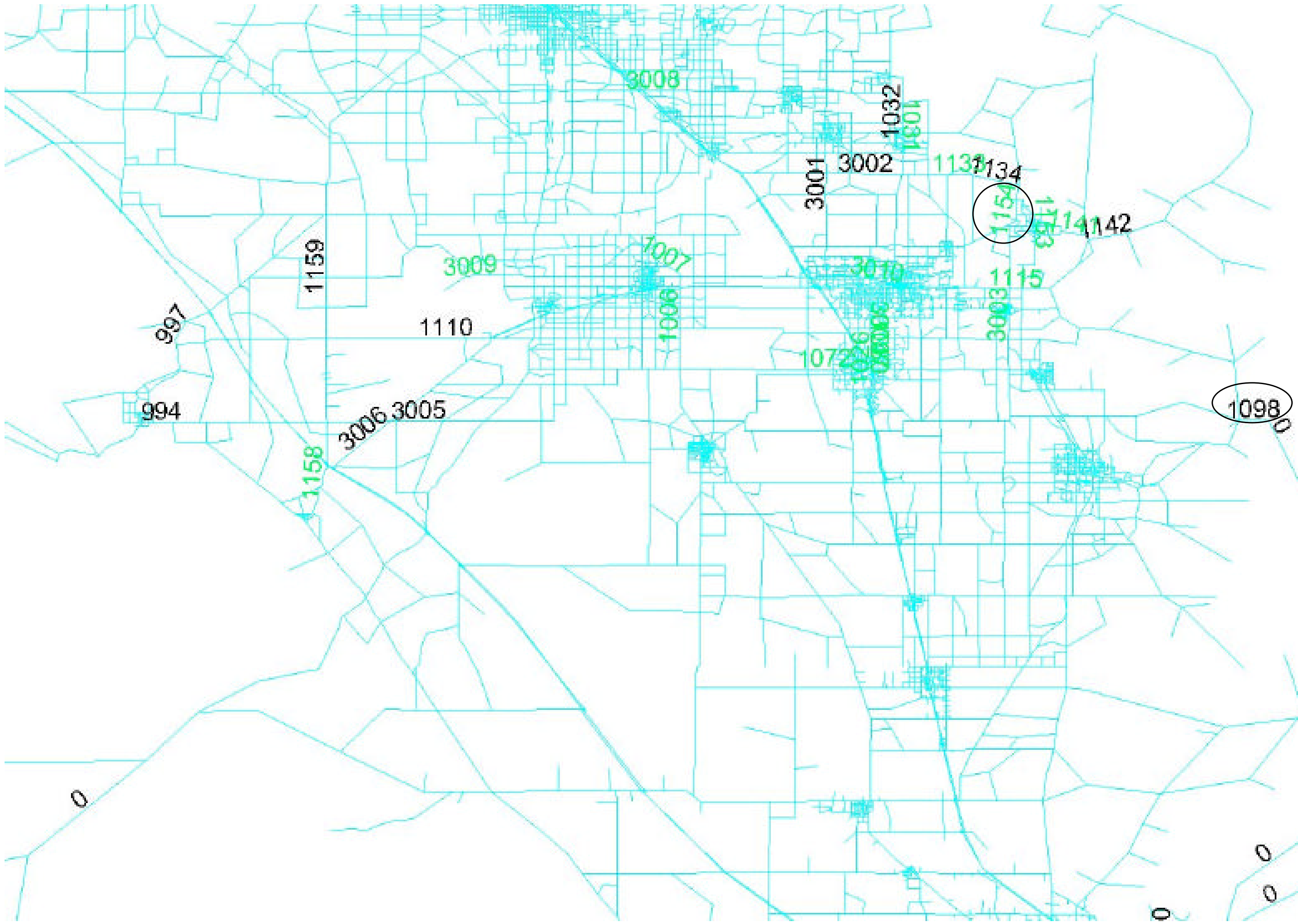


Figure 3.1(c) Location of Traffic Counter Sites (by Site ID) in the San Joaquin Valley
(Counter locations with higher average daily volumes on weekend than on weekdays are marked with a circle)



4.0 ANALYSIS METHODS

The major steps in the allocation method can be summarized as:

1. The matched links' observed hourly volumes are reduced to mathematical combinations known as Principal Components (PC's), which have two desirable properties: high explanatory power for the 24 hourly volumes, for which they serve proxies, and the principal components are uncorrelated.
2. The matched links are assigned into different clusters based on statistics computed using the obtained principal components. Simply speaking, the statistics transmit important information such as spatial and temporal characteristics of each link into a quantifiable measure of distance. In this way, links that are closest to each other, that is, they bear a certain degree of spatial and temporal similarities, are grouped into one cluster. This clustering method used was the average agglomerative hierarchical clustering method.
3. Allocation factors are estimated using the matched links' model volumes and the 24 hourly volumes in a multivariate multiple regression (MMR) model for each cluster.
4. The links without matching observed traffic data are used to derive a statistic that measures their statistical distance from all the available clusters. Similar to step 2, the unmatched links are assigned to the closest cluster.
5. The unmatched links' model volumes are disaggregated into hourly volumes using the allocation factors specific to the cluster in which they were assigned in step 4.

4.1 Step 1:Links with Matching Observed Traffic Data

Before the matched links are clustered, P principal components are constructed as linear combinations of the 24-hourly traffic counts for each location as follows,

$$Y_{kp} = \sum_{i=1}^{24} q_i \, x_{ip}^{k} , \qquad (1)$$

where Y_{kp} denotes the pth principal component for the kth location, x_i is the ith hour volume at kth location and $\boldsymbol{q}_i$ are the eigenvectors.

Although as many principal components as variables can be constructed, i.e., P=24 in this case, only a few are actually used to simplify computation involving the large number of variables. Since the principal components are used as proxies for the 24 hourly volumes, they are constructed such that the first principal component has decreasing explanatory power in terms of total variance of the 24 hourly volumes. For example, the first principal component should account for most of the total variance. Principal components are also designed to be uncorrelated with each other.

4.2 Step 2: Clustering of Matched Links

By using the principal components instead of the 24-hourly traffic counts, the computation of the Euclidean distance, *d(j,k),* between the *j*th and *k*th location is simplified and can be written as

$$d(j,k) = \left[\sum (Y_{jp} - Y_{kp})^2 \right]^{1/2} \qquad (2)$$

Closest locations are grouped into clusters using an average agglomerative hierarchical clustering method, which begins by assigning each of *K* count locations to a unique cluster so that each of the *K* clusters contains a single entity; this is followed by computation of the Euclidean distance between each pair of clusters, *d(j,k)*. The pair of clusters having the minimum distance between them is merged so that there are *K-1* clusters remaining after merging. This process is repeated until an appropriate number of clusters are determined based on a suitable criterion. One commonly used criterion to determine the number of clusters is the pseudo-F ratio value, which is the ratio of the mean square error between clusters to the mean square error within clusters,

$$F = \frac{\left(\sum_{i}^{n} (Y_i - \bar{Y})^2 - \sum_{k}^{G} \sum_{i}^{n} (Y_i - \bar{Y}_k)^2 \right) / (G-1)}{\left(\sum_{k}^{G} \sum_{i}^{n} (Y_i - \bar{Y}_k)^2 \right) / (n-G)} \qquad (3)$$

where

G=number of clusters,
K=index for cluster k,
P=number of principal components,
n=number of observations, i.e., $G \times P$,
Y=principal component.

A large F-ratio value implies that more of the variation can be attributed to the clusters than to randomness, a preferred condition.

4.3 Step 3: Estimation of Hourly Allocation Factors

Hourly allocation factors can be estimated with a multivariate multiple regression (MMR) model that assumes correlation across the observed hourly counts within each modeling period. If there are *J* modeling periods, T_j is the number of hours within modeling period *j*, and *i* represents any hour within a 24-hour period; so *i=1,…,24*. Note also that t_j represents the subset of hours *i* in modeling period *j*. Let *c = 1,2,…,C* denote the cluster ID's identifying spatial location and k_c = *1,2,…,*K_c the index for locations in Cluster *c*, with K_c representing the total number of locations in Cluster *c*. The counter volume observed on the *kth* location in Cluster *c* at hour *i* can be formulated as

$$y_{i,k_c} = b_i^c \, x_{k_c}^j + e_{i,k_c}, \qquad i \in t_j, \; j=1,\ldots,J. \tag{4}$$

where y_{i,k_c} is the observed volume for cluster *c* for hour *i* contained in the subset of t_j, x_{k_c} is the estimated travel demand volume for location contained in cluster *c* during period *j*, b_i is the proportion of the demand volume occurring during hour *i*, and e_{i,k_c} represents the model error term. There is no intercept (b_0) included in the equation since by definition the proportion of period volumes assigned to each hour sum to one. The error terms are distributed with a mean of zero and a variance of s_i^2, $e_{i,k_c} \sim (o, s_i^2)$; that is, the variance in error terms occurs across hours but not across locations within clusters. Also, the error terms are independent and identically distributed (i.i.d.). If inferences about the value of *b* were necessary, then an assumption of normality would also be required.

4.4 Steps 4 and 5: Links without Matching Count Data

It is common that observed traffic counts are far fewer in scope than the roadways represented by the travel demand network. For example, approximately 99% of the 66,237 links in our demand network do not have matching count data. This is a very common condition for large area spatial estimation problems. For these unmatched links, Hicks and Niemeier (2001) suggested classifying them to the predetermined clusters according to the squared statistical distance between each specific link and the cluster's mean volume vectors. The squared statistical distance is calculated as follows:

$$D_c = (x_0 - x_c)' S^{-1} (x_0 - x_c) \tag{5}$$

where $\boldsymbol{x_0}$ denotes the vector of model volumes for a certain unmatched link 0, $\boldsymbol{x_c}$ the vector of average model volumes for Cluster *c*, and $\boldsymbol{S^{-1}}$ denotes the inverse of the covariance matrix of the pooled cluster model volumes.

$\boldsymbol{S^{-1}}$, the inverse of the covariance matrix of the pooled cluster model volumes, is used to standardize the statistical distance. In doing this, the differences in variation among modeling periods can be accounted in the statistical distance, assuming that the volumes within each modeling period and cluster are normally distributed with equal covariance among clusters.

After determining the distance between the new link and each predetermined cluster, a new link is classified into the closest cluster (i.e., the cluster with smallest value of $\boldsymbol{D_c}$) and the allocation factors obtained previously for each cluster are used to disaggregate the model volumes into hourly volumes.

4.5 Possible Classification Problems Using Daily Volumes

In theory, the squared statistical distance, $\boldsymbol{D_c}$, can be computed using the model links', both matched and unmatched, total daily volumes instead of period-based volumes as in Hicks and Niemeier's original application. For example, if $\boldsymbol{x_c}$, the vector of average model volumes for

Cluster c , and $\boldsymbol{S^{-1}}$,the inverse of the covariance matrix of the pooled cluster model volumes, are computed using the matched links' period-based volumes, $\boldsymbol{D_c}$ can then be calculated using unmatched links' period-based volumes; vice versa for calculation of $\boldsymbol{D_c}$ using model links' total daily volumes.

However, using links total daily volumes weakens the power of squared statistical distance to classify unmatched links and hence increases the probability of assigning unmatched links into wrong cluster. This can occur because the covariance matrix of the pooled cluster model volume, S, reduces to the pooled clusters' variance of daily volume. In addition, $\mathbf{(x_0 - x_c)' (x_0 - x_c)}$ now measures only the squared difference between an unmatched link's average model volume and each cluster's average daily volume. As a result, the squared statistical distance depends solely on the average daily volumes of each predetermined cluster.

Possible problems arising from using model links' daily volumes for classification of unmatched links is illustrated by a simple hypothetical example. Suppose that the MMR procedure yielded two sets of estimated allocation factors for clusters, 1 and 2, respectively. Cluster 1 is assumed to consist of predominantly model links representing urban roadways while Cluster 2 consists of model links representing rural roadways. For this geographical reason, let us further assume that Cluster 1's average daily volumes (x_1) be larger than Cluster 2's (x_2). Furthermore, the two clusters' allocation factors are assumed to exhibit distinctly different temporal traffic patterns, for instance, traffic profile in Cluster 1 displays a single peak in the afternoon (PM-peak) whereas Cluster 2's displays a single peak in the morning (AM-peak).

A certain unmatched model link 0, which represents a segment of an urban roadway has an average daily volume of x_0 ($x_2 > x_0$, for sake of exposition). The link's average daily volume is low because the roadway segment is not an important arterial, which also explains why count data are unavailable. It is logical to assume that the urban roadway link 0 represents exhibits a temporal traffic pattern similar to that of Cluster 1 (i.e., PM-peak). However, following Hicks and Niemeier's method to classify unmatched links to predetermined clusters by using the statistical distance of Equation 5, link 0 is assigned to cluster 2 because Equation 5 is simply a comparison of average daily volumes of clusters 1 and 2, i.e., $w(x_0 - x_1)^2 > w(x_0 - x_2)^2$,where w is the reciprocal of the pooled variance of daily link volumes in clusters 1 and 2. Consequently, link 0's hourly volumes, which are disaggregated using Cluster 2's allocation factors are mis-estimated.

The illustrative example presents a conundrum in the direct application of Hicks and Niemeier's method using our data. Recall that only 8 of 135 matched links have volumes for 4 periods of a day (i.e., AM, Mid-day, PM, Off-peak) while 23 links have volumes for three periods of a day (i.e., AM, PM, and Off-peak). The remaining 104 matched links have only aggregate daily volume (i.e., 24-hour). The large number of unmatched links with missing period-based volumes precludes the use of period-based volumes in actual analysis. On the other hand, direct application of Hicks and Niemeier's method using model links' aggregate daily volumes will be very likely subject to the misclassification problems we just discussed. This is because the model network consists of a large number of unmatched links representing both roadways in urban (e.g., the San Francisco Bay Area) as well as rural (e.g., the San Joaquin Valley) areas. Consequently, it is reasonable to expect that rural roadways exhibit a temporal traffic pattern different from roadways in urban area. In fact, as shown in Hick and Niemeier (2001), traffic on roadways in different localities were found to exhibit different temporal pattern (AM or PM peak).

4.6 Modified Classification Procedure

To attenuate the previously discussed misclassification problems, we modified the method used to classify unmatched links. Since 63.6% (or 42,049) of the 66,102 unmatched links have volumes for at least three periods of day, it is apparent that the main difficulty in calculating the statistical distance (Equation 5) using period-based volumes lies with the matched links' missing period-based volumes.

After the matched links have been clustered as described earlier, allocation factors are estimated for each cluster using the model links' aggregate daily volumes in the MMR procedure. Given the duration of each time period, which were defined in the travel demand model (See Table 4.10 for definition of each period), we imputed volumes for the four time periods from the matched links' aggregate daily volumes using the previously estimated allocation factors in each predetermined cluster. For example, $\hat{x}_{ic}^{am}$, matched link *i*'s volume for AM peak (6am-9am) in Cluster *c* can be imputed as

$$\hat{x}_{ic}^{am} = b_{c}^{am6} x_{ic}^{wd} + b_{ic}^{am7} x_{ic}^{wd} + b_{c}^{am8} x_{ic}^{wd} \quad (6)$$

where b_{c}^{am6} is the allocation factor at 6am for Cluster *c*; x_{ic}^{wd} is the aggregate daily volume for matched link *i* in Cluster *c*.

These imputed period-based volumes are then used to calculate a hybrid covariance matrix, $\hat{S}$; vectors of average volumes for Cluster c , $\hat{x}_c$, and hence $\hat{D}_c$, the imputed statistical distance using Equation 5. After the statistical distance has been calculated using imputed period-based volumes for every unmatched links with period-based volumes, the links are assigned to the closest cluster. For the unmatched links without period-based volumes, the links are classified based on the statistical distance computed using the model links' aggregate daily volumes as before.

5.0 CLUSTERING AND ALLOCATION FACTOR RESULTS

In this section, we present a general description of the observed count data followed by the clustering and estimation results.

5.1 Exploratory Data Analysis

An overall summary of the traffic counts data is presented in Table 5.1. The daily volumes, which are obtained by averaging daily volumes across direction of traffic, days and counter locations, averages 5,575 vehicles per day for weekday (Monday through Friday) and averages lower with 4,634 vehicles per day for weekend (Saturday and Sunday). Table 5.2 shows a more detailed summary of the daily average volumes by counter site. It is generally consistent with Table 5.1. Note that 23 counter locations (shaded in Table 5.2.) record higher average daily volumes on weekends than on weekdays. We marked 21 of the 23 locations (two locations with ID Nos.: 443 and 44, could not be matched to any link in the CCOS network) with circle for easy identification using previously discussed maps in Figures 3.1(a-c).

Table 5.1 Summary Statistics of Traffic Count Data

Days of Week	Frequency	Average Daily Volume
Monday-Friday (Weekday)	17,886	5,575
Saturday-Sunday (Weekend)	7,061	4,634
Monday-Sunday (Whole Week)	24,947	5,308

The observed traffic volumes are tabulated in more detail as hourly averages by districts for weekdays and weekdays in Table 5.3. The 24-hour traffic profiles for weekdays suggest two peaks, one in the morning and one in the afternoon. The weekend traffic profile is like an inverted bell: traffic is light early in the morning and late in the evening but fairly heavy in between these periods. The weekday traffic profile is probably a result of combining two directions traffic volumes on roadways. The weekend traffic tends to be more clustered in the afternoon. Moreover, as shown in column 7 of Table 5.3 (columns 6 and 7 in the second part of Table 5.3), the hourly traffic volumes in District 6, which covers the San Joaquin Valley are significantly lower compared to those in other more urbanized districts.

The hourly volumes in Table 5.3 are also plotted in Figures 5.1-5.6. As shown in Figure 5.1, the 24-hour traffic profiles on weekdays are more like a saddle whereas the weekend traffic has the shape of an inverted bell. The average hourly volumes for the weekday and the 7-day traffic profiles increase in the early morning becoming slightly flat after peaking at 8AM. The traffic picks up again around 12PM until it peaks at 5PM and begins to decrease steeply afterwards. In contrast, traffic on the weekend increases and decreases more gradually after peaking for a shorter period of time around 1PM. Note that the dip at 12PM on weekend is caused by significant number of zero volumes around 12PM (See Table 3.4). These zero volumes were not discarded because doing so would deplete the sample size, especially for weekend traffic and zero volumes were likely meaningful for some locations with low total traffic volumes. Table 5.2(a) shows the average daily volumes re-computed without the zero hourly volumes for both weekdays and weekends. In comparison with Table 5.2, the weekday average daily volumes change little for most of the locations except for a number of locations, for which a significant number of traffic volumes were discarded (e.g., sites 670 and 671). In contrast, the weekend average daily volumes change more significantly and were missing for 27 locations because of the relatively small number of observations. As will be seen later, the significant number of zero hourly volumes causes abruptions in weekend's 24-hour traffic profiles.

The 24-hour traffic profiles for 7 days for the four study districts 3, 4, 6, and 10 (i.e., the last four columns of Table 5.3) are shown in Figure 5.2. All the 24-hour traffic profiles for the four study districts are fairly similar. Traffic is generally heavier in Districts 3 and 4, which include more urbanized areas such as the San Francisco Bay Area and the Sacramento Metropolitan Area in northern California. In contrast, Districts 6 and 10, which cover the San Joaquin Valley in central California have appreciably lower hourly volumes.

After comparing the 24-hour weekday and weekend traffic profiles in the four study districts as shown in Figures 5.3-5.6, it appears that roadways in more urbanized districts (i.e., Districts 3 and 4) display a more distinctly saddle-shaped weekday traffic profiles with traffic peaks in both the morning and in the afternoon (See Figures 5.3 and 5.4). In contrast, the average weekday traffic on roadways in the less urbanized Districts 6 and 10, display a peak more distinctly in the afternoon than in the morning (Figures 5.5 and 5.6). On weekends, the 24-hour traffic profile in all of the four study districts bear close resemblance to each other: an inverted bell-shaped curve with heavier volumes from noon till 5PM.

5.2 Clustering Results of Links with Matching Count Data

As discussed in preceding sections, principal components were constructed as the linear combinations of the 24 hourly volumes of the model links with matched count data. Theoretically, as many principal components as the explained variables (24 in this case) can be constructed but only a few are used to simplify the mathematical operation involving the explained variables. Table 5.4 shows the cumulative proportion of variance within the 24 hourly volumes for the weekday and weekend that are explained by each of the 24 possible principal components. For both the weekday and weekend traffic, 18 principal components are sufficient to describe most of the variation exhibited within the 24 hourly volumes. As also shown in Table 5.4, the first two principal components combine to explain more than half of the total variance within the 24 hourly volumes---51% and 61% for the weekday and weekend, respectively. Although additional principal components help explain more of the total variance, it is more difficult to interpret these additional principal components in terms of temporal traffic patterns As shown in Table 5.5, magnitude of the remaining principal components' (i.e., PC3-PC8) coefficients fluctuate around the time of day.

For this reason as well as for the sake of simplifying the subsequent analyses, these first two principal components are used as in place of the 24 hourly volumes, weekday and weekend alike.

The linear combinations of these 2 principal components for both the weekday and weekend are presented in Table 5.5. For example, the following principal components are written in terms of the 24 hourly volumes for weekday:

$$PC_1^{weekday} = -0.07x_i^{am1} + 0.004x_i^{am2} + 0.025x_i^{am3} + 0.067x_i^{am4} + 0.166x_i^{am5} + 0.439x_i^{am6} + 0.539x_i^{am7} + 0.375x_i^{am8} + 0.148x_i^{am9} + 0.028x_i^{am10} - 0.033x_i^{am11} - 0.022x_i^{am12} - 0.193x_i^{pm1} - 0.183x_i^{pm2} - 0.136x_i^{pm3} - 0.136x_i^{pm4} - 0.081x_i^{pm5} - 0.075x_i^{pm6} - 0.244x_i^{pm7} - 0.287x_i^{pm8} - 0.210x_i^{pm9} - 0.093x_i^{pm10} - 0.062x_i^{pm11} - 0.030x_i^{pm12}$$

$$PC_2^{weekday} = 0.078x_i^{am1} + 0.038x_i^{am2} + 0.017x_i^{am3} - 0.002x_i^{am4} + 0.058x_i^{am5} + 0.066x_i^{am6} + 0.111x_i^{am7} - 0.051x_i^{am8} + 0.023x_i^{am9} - 0.076x_i^{am10} - 0.218x_i^{am11} - 0.336x_i^{am12} - 0.257x_i^{pm1} - 0.312x_i^{pm2} - 0.251x_i^{pm3} - 0.199x_i^{pm4} - 0.1371x_i^{pm5} - 0.198x_i^{pm6} + 0.282x_i^{pm7} + 0.381x_i^{pm8} + 0.317x_i^{pm9} + 0.279x_i^{pm10} + 0.219x_i^{pm11} + 0.166x_i^{pm12}$$

Likewise for weekend,

$$PC_1^{weekend} = -0.159x_i^{am1} - 0.104x_i^{am2} - 0.073x_i^{am3} - 0.046x_i^{am4} - 0.035x_i^{am5} - 0.012x_i^{am6} - 0.035x_i^{am7} + 0.048x_i^{am8} + 0.087x_i^{am9} + 0.088x_i^{am10} + 0.130x_i^{am11} + 0.547x_i^{am12} + 0.185x_i^{pm1} + 0.270x_i^{pm2} + 0.214x_i^{pm3} + 0.191x_i^{pm4} + 0.1721x_i^{pm5} + 0.055x_i^{pm6} - 0.185x_i^{pm7} - 0.297x_i^{pm8} - 0.294x_i^{pm9} - 0.276x_i^{pm10} - 0.266x_i^{pm11} - 0.206x_i^{pm12}$$

$$PC_2^{weekend} = 0.031x_i^{am1} + 0.032x_i^{am2} + 0.041x_i^{am3} - 0.062x_i^{am4} + 0.104x_i^{am5} + 0.203x_i^{am6} + 0.237x_i^{am7} + 0.254x_i^{am8} + 0.209x_i^{am9} + 0.100x_i^{am10} - 0.040x_i^{am11} + 0.472x_i^{am12} - 0.345x_i^{pm1} - 0.292x_i^{pm2} - 0.353x_i^{pm3} - 0.260x_i^{pm4} - 0.193x_i^{pm5} - 0.080x_i^{pm6} - 0.219x_i^{pm7} - 0.169x_i^{pm8} - 0.049x_i^{pm9} + 0.064x_i^{pm10} + 0.106x_i^{pm11} + 0.085x_i^{pm12}$$

where the 24-hourly volume variables are denoted by x_i. The coefficients represent the weights allotted to each hourly traffic volume. For example, the first principal component, $PC_1^{weekday}$, for weekday traffic has large positive coefficients for hourly volumes at 6AM-8AM (β_6 -β_9) and large negative coefficients at 7PM-9PM (β_{19} -β_{21}). These large coefficients, positive or negative regardless, imply that their corresponding hourly volume variables are more important to $PC_1^{weekday}$ and possibly suggest strong morning but weak evening peak characteristics for the weekday traffic. Large coefficients of the principal components for weekday and weekend traffic are highlighted in Table 5.5.

The four principal components (2 for weekday, 2 for weekend) are used to calculate the statistical (Euclidean) distance between any pair of counter locations according to Equation 2. Afterwards, the count locations are continuously grouped into clusters until an appropriate number of clusters is determined by comparison of the pseudo-F ratio values in Equation 3.

Since the pseudo-F value is the ratio of the mean square error between clusters to the mean square error within clusters, initially increasing the number of clusters usually allows better distinction of locations with significantly different characteristics. A larger proportion of overall variation is thus attributed to the clusters and hence a large F-ratio. Too many clusters, however, reduces the number of elements in each cluster and poses problems for subsequent cluster-specific estimation.

As shown in Table 5.6, the pseudo F-ratio generally increases as the number of clusters increases for both the weekday and weekend hourly volumes. For the weekday hourly volumes, the pseudo-F ratio value increases dramatically from 6.7 to 43 before decreasing slightly to 35.4 when the number of clusters increases from 3 to 4. In the case of weekend hourly volumes, the F-ratio increases significantly from 11.8 to 21.8 when the number of clusters increases from 2 to 3. These large jumps in F-ratio values suggest that the additional cluster accounts for a significant proportion of the variations within clusters. Although a larger number of clusters yields a larger F-ratio, the number of locations in each cluster is too few to allow reliable estimations within each cluster. In fact, the small sample size constrains even the suggested 4 and 3 clusters for weekday and weekend traffic. For example, in one cluster for weekday traffic, there were only 2 locations for weekday traffic (site ID's: 1711 and 3005). As a result, locations were grouped into two clusters for both weekday and weekend hourly volumes.

5.3 Results of Allocation Factors Estimation

After the number of clusters was determined, we proceeded to estimate the allocation factors for each cluster. Since period-based volumes are available for only 31 matched links, allocation factors are estimated using each matched link's aggregate daily volume. By the construct of the estimation model, the 24 hourly allocation factors obtained using the aggregate volume sum to one (i.e., $\sum_{i=1}^{24} b_i^k = 1$).

As shown in Table 5.7, allocation factors are estimated for a 7-day week, weekdays, and weekend. These estimated allocation factors are also plotted against time of day by cluster in Figures 5.7, 5.8, and 5.9. Figure 5.7 presents the plot of allocation factors using a 7-day week of traffic data. For the roadways represented by the matched links in Cluster 1, the allocation factor increases steeply from 5AM to 7AM, which is then followed by a gradual increase until the allocation factor achieves its highest value at 6PM and decreases steeply afterwards. In contrast, Cluster 2's profile of allocation factors has a more distinct peak at 5PM in the afternoon. However, the results should be interpreted with a caveat that there are only four matched links in Cluster 2 (See column 2 of Table 5.7).

Profiles of the weekday allocation factors in Figure 5.8 suggest a temporal pattern of traffic with peaks in the morning (8AM) and in the evening (8PM) for locations in Cluster 1. Profiles of allocation factors for locations in Cluster 2, however, exhibit a temporal pattern of a distinctly heavier peak of traffic in the evening. As shown in Figure 5.9, the profiles of weekend allocation factors are inverted bell-shaped. These inverted bell-shaped allocation factors profiles suggest that the weekend traffic on locations in Cluster 1 exhibits a hump of heavy traffic between 2PM and 6 PM and tapers off in other periods of day. Cluster 2's allocation factor takes an abrupt decrease at 12AM. As discussed previously, this abrupt drop is caused by a significant number of zero hourly volumes recorded at a number of links during a two-hour period (i.e., 12AM-1PM). The effect of these zero volume recordings becomes more obvious in a small sample size such as Cluster 2's (i.e., 6 locations).

5.4 Classification of Links without Matching Count Data

After allocation factors have been estimated for each predetermined cluster, links without matched count data are classified in two ways: 1) the method used in the SCOS study, that is, the method proposed by Hick and Niemeier (2001) and 2) the method using a modified procedure discussed in Section 4.6.

5.4.1 Method 1: Using Model Links' Aggregate Daily Volumes

As discussed previously, since 110 of 141 links do not have period-based volumes (AM, Mid-day, PM, Off-peak), classifying the unmatched links using period-based volumes may yield unreliable results because of the small sample size. For this reason, we applied Hick and Niemeier's classification procedure using aggregate daily volumes as inputs in Equation 5. In order to classify an unmatched link, the link's statistical distances, D as denoted in Equation 5, from each predetermined cluster need to be computed. For reasons that have already been discussed, calculating the statistical distance using models' aggregate daily volumes requires only each cluster's mean daily volume and the pooled variance of daily volume.

As shown in Table 5.8, Cluster 1 has a larger mean daily volume than Cluster 2 for both weekday and weekend traffic, namely, 20,113 vehicles per day (Cluster 1) versus 19,005 vehicles per day (Cluster 2) on weekdays and 20,198 vehicles per day (Cluster 1) versus 5,664 vehicles per day (Cluster 2) on weekend. The pooled variances (denoted as S in Equation 5), 826,227,450 for weekday and 830,132,477 for weekend, are used to standardize the variations in each cluster in the calculation of the statistical distance. Since a link's statistical distance to each predetermined cluster depends on only three parameters, namely, the mean daily volumes of clusters 1 and 2 and the pooled variance, the cluster an unmatched link is assigned to is only a function of the mean daily volumes of clusters 1 and 2. For example, all the unmatched links, whose weekday daily volumes are greater than Cluster 1's mean daily volume (i.e., 20,113 vehicles per day) will be classified into Cluster 1 and likewise for those unmatched links, whose weekday volumes are less than cluster 2's mean daily volume (i.e., 19,005 vehicles per day). It is less apparent only for those unmatched links, whose model volumes fall between the mean daily volumes of clusters 1 and 2. This reasoning applies similarly to classification of unmatched links based on weekend traffic. The classification based on weekend traffic should be taken with a caveat that there are only 6 locations in the Cluster 2.

The classification results of the 66,102 unmatched links using Method 1 are shown in the upper panel of Table 5.9. When weekday traffic is considered, 12,719 links are assigned to Cluster 1 and 53,383 to Cluster 2. This is similar for weekend traffic when 19,266 unmatched links are assigned to Cluster 1 and 46,836 links to Cluster 2. As expected, the resulting mean daily volume of those unmatched links is greater in Cluster 1 than in Cluster 2, though Cluster 1 contains less than one third of links in Cluster 2. This is because the unmatched links with larger daily volumes are assigned to cluster 1 while those with smaller daily volumes are assigned to Cluster 2.

Distributions of all the links, matched or unmatched, using daily volumes for both weekend and weekday are drawn in Figures 5.10(a-c) and 5.11(a-c). In particular, the resolution of network coverage in three areas, namely, the Sacramento Metropolitan Area, the San Francisco Bay Area and the San Joaquin Valley are magnified in Figures 5.10(a)-5.10(c) and 5.11(a)-5.11(c) for a closer examination. As shown in Figure 5.10, it appears that most of the links in Cluster 1 (represented by black line) are concentrated in urban areas and most notably the northern part of the Los Angeles Metropolitan Area (See Table 3.1 for the CCOS study network coverage.) based on weekday traffic. It is slightly different for traffic on weekend. Figure 5.11 shows that more links in less urbanized areas belong to Cluster 1 when classified using weekend daily volumes. As shown in more resolute networks in Figures 5.10(a)-5.10(c) and 5.11(a)-5.11(c), most of the links classified in Cluster 1 appear to represent important roadways in each highlighted area.

5.4.2 Method 2: Using Imputed Period-Based Volumes

As the foregoing classification results show, Method 1, which uses model links' daily volumes relies heavily on the mean daily volume of each predetermined cluster. As a result, links with high daily volumes are assigned to the cluster with the higher mean daily volume with little consideration of temporal traffic variation. To compensate for Method 1's deficiency, we imputed all the matched links' missing period-based volumes from their daily volumes using the previously determined allocation factors in their assigned clusters. These imputed period-based volumes were then used to compute the statistical distance (Equation 5) instead of daily volumes for each unmatched link.

Equation 5 thus has more parameters, that is, four mean period-based volumes for each predetermined cluster and a pooled covariance matrix of eight elements, to determine an unmatched link's statistical distance from each predetermined cluster. However, this modified classification procedure does not apply to unmatched links, for which period-based volumes are not available (A total of 20,874 links according to Table 3.1.). These links are classified using Method 1, which uses daily volume as inputs. A more detailed discussion of the imputation and calculation of statistical distance can be found in Appendix A.

As an illustration, the pooled mean imputed period-based volumes for both weekday and weekend are presented in Table 5.10. As can be seen, the four time periods are not evenly divided: AM peak consists of only a period of three hours from 6AM through 9AM while off-peak has a span of 11 hours from 7PM through 6AM. The peak of traffic shifts later on weekend as evident in the smaller mean volumes in AM peak but larger mean volumes in subsequent periods compared to weekdays.

The classification results using imputed period-based volumes are summarized in the lower panel of Table 5.9. Compared to classification results using Method 1, links are more evenly assigned to clusters 1 and 2, for example, 35,525 links are assigned to Cluster 1 using Method 2 versus 12,719 links using Method 1. The mean daily volume of Cluster 1 also decreases more than half from 46,601 vehicles per an average weekday using Method 1 to 17,710 vehicles per an average weekday using Method 2. Similar findings are also observed in classification results based on weekend traffic. For example, 25,785 unmatched links are assigned to Cluster 1 compared with 19,266 using Method 1. Cluster 1's mean daily volume also decreases to 18,005 vehicles per average day in weekend from 36,167 vehicles per day using Method 1. These findings indicate that the temporal traffic variations displayed within the period-based volumes are important in classification of the 66,102 unmatched links.

The distribution of all the links by cluster is plotted using the network coverage in Figures 5.12(a-c) and 5.13(a-c) for both the weekday and weekend traffic. Network coverage of specific geographic areas such as the Sacramento Metropolitan Area, the San Francisco Bay Area, and the San Joaquin Valley are drawn in a larger scale in Figures 5.12(a)-5.12(c) and 5.13(a)-5.13(c).

Consistent with the findings in Table 5.9, there are more unmatched links assigned to Cluster 1 compared to using Method 1 for both weekday and weekend traffic in Figures 5.12 and 5.13. The links in Cluster 1 appear to be more concentrated in areas, where population densities are generally higher. In addition to the three aforementioned urbanized areas (i.e., the Sacramento Metropolitan Area, the San Francisco Bay Area, and the Los Angeles Metropolitan Area), these areas include Fresno in Fresno County and Bakersfield in Kern County, both of which are in the San Joaquin Valley. As discussed earlier, roadways in these urbanized areas have temporal patterns different from those in the rural areas but were previously assigned to Cluster 2 by Method 2 because of their relatively small daily volumes.

5.5 Misclassification of Matched Links

In order to evaluate the reliability of Methods 1 and 2, we computed the rate of correctly classified links using both Methods 1 and 2. Following each respective method's described procedures, clusters were predicted for the 135 matched links using the predetermined allocation factors. The matched links' predicted clusters were then compared to their original clusters obtained using observed traffic counts based on weekday or weekend traffic. The detailed comparison results by weekday and weekend are presented in Tables A3.1 and A3.2 in Appendix A but a brief summary is presented in Table 5.11.

The first column of Table 5.11 lists the clusters, which matched links were assigned to using observed traffic counts. The next four columns list the clusters predicted using daily volumes (i.e., Method 1) and using imputed period-based volumes (i.e., Method 2). For example, based on weekday traffic, 23 of 92 links (25%) and 35 of 43 links (81.4%) were correctly predicted using Method 1 into clusters 1 and 2, respectively (The number of correctly predicted clusters are bolded in Table 4.10. Method 2 correctly predicted 33 of 92 links (35.9%) in Cluster 1 and 25 of 43 links (58.1%) in Cluster 2. Based on weekend traffic (See lower panel of Table 5.11), Method 1 correctly predicted 48 of 129 links in Cluster 1 and 6 of 6 links in Cluster 2. In contrast, 51 of 129 links (39.5%)in Cluster 1 and 5 of 6 links (83.3%) in Cluster 2 were correctly predicted using Method 2.

From comparison results, there appears to be no significant difference between Method 1 and Method 2 in terms of their classification success rate. The classification results of the 135 matched links using Method 2 are probably due to three main reasons: first, the sample size of

135 links is not large enough to support the clustering analysis. Second, the sample size of Cluster 2, especially for weekend, is very small for reliable estimate of success rate of classification. Third, only 8 of 135 matched links (6%) have 4 period-based volumes (i.e., AM, Mid-day, PM, Off-peak) and 23 (17.3%) links have volumes for three periods (i.e., AM, PM, and Off-peak). This implies that 110 or 82.7% of the matched links have only aggregate daily volumes. According to Method 2, these links' clusters were predicted using their daily volumes (i.e., Method 1) instead of period-based volumes.

Table 5.2 Summary Statistics of Traffic Count Data by Site ID
(Sites with higher daily volume on weekend compared to weekday are shaded)

Site ID	Average Daily Volume	Average Daily Volume (Mon-Fri)	Average Daily Volume (Sat-Sun)
242	51790	55270	44692
254	1021	926	1274
259	3571	3717	3153
306	3024	3226	2377
307	1043	1038	1056
308	6363	6611	5681
335	9686	10078	8562
358	6153	6042	6454
360	5330	5443	5050
361	6288	6353	6097
364	11954	12658	10226
393	6237	6300	5985
395	8786	8785	8790
397	13031	13046	12990
399	14068	14613	12567
400	15083	15717	13341
409	832	846	797
417	2921	2987	2745
422	2211	2267	2057
427	3063	3325	2430
443	1416	1334	1613
444	1573	1502	1745
492	6884	7264	5954
493	8986	8841	9303
495	21030	22804	15263
496	22169	24349	15630
500	64189	67046	55617
502	36377	37959	33476
513	13723	13967	13054
515	12864	12812	12998
517	24037	25626	19270
518	34603	36261	29627
524	71680	76663	61712
531	11843	12176	10947
533	13360	13531	12928
535	12561	13625	9636
558	65842	64584	69614
566	55531	54165	59631
571	22869	22890	22815

Table 5.2 (continued) Summary Statistics of Traffic Count Data by Site ID

573	21668	22847	17421
605	20694	21443	18868
607	26930	29081	22270
608	13065	13772	11357
609	19826	20866	16705
610	8291	8683	7376
611	14187	14175	14221
630	10120	10651	8608
631	7251	7935	5429
632	10710	12191	7378
633	4576	5007	3497
634	12483	13143	10833
635	18420	19786	14848
637	56909	60549	45988
649	990	775	1470
651	2603	2447	2956
652	14480	16533	9008
654	21945	23313	15789
669	979	1075	733
670	96	96	99
671	113	114	113
675	20459	22135	15429
679	8692	8402	9484
681	12457	12630	11972
682	14040	13932	14306
683	41164	45236	29966
685	72219	76068	60671
686	45464	47400	40623
696	15503	16392	13130
701	30351	33754	20142
704	24841	27865	15770
733	13612	14576	11121
734	14529	15602	11846
735	5606	6377	4167
736	23150	23788	21185
994	2549	2726	2122
997	1827	1956	1519
1006	2823	3108	2124
1007	3316	3563	2696
1026	5029	5246	4496
1030	2626	2689	2483
1031	4560	4594	4476

Table 5.2 (continued) Summary Statistics of Traffic Count Data by Site ID

1032	696	720	638
1072	1961	2023	1809
1098	370	345	431
1110	4842	5283	3739
1115	5697	6208	4424
1133	1935	2014	1736
1134	448	458	420
1141	1942	2032	1715
1142	574	618	464
1153	2662	2762	2407
1154	242	240	245
1158	1705	1747	1600
1159	1094	1184	865
1164	634	675	529
1611	5671	5963	4886
1613	8884	10585	5738
1614	5872	7848	4884
1646	3018	3048	2937
1647	2525	2667	2160
1656	5009	5126	4250
1661	3534	3484	3738
1662	2535	2609	2214
1663	3363	3499	3018
1664	7703	8161	6606
1674	4178	4210	4098
1676	1078	1107	1003
1678	1801	1902	1543
1679	3888	3899	3857
1694	8801	10048	6459
1698	2042	2189	1562
1711	6907	7010	6615
1731	2482	2404	2711
1748	3799	3710	4032
1756	3784	3673	4094
1773	4261	4311	4131
1777	4226	4117	4519
1814	724	755	633
1816	4107	4138	4019
1818	2095	2241	1720
1838	1008	1059	878
3001	3246	3524	2533
3002	1457	1553	1203

Table 5.2 (continued)   Summary Statistics of Traffic Count Data by Site ID

3003	2567	2881	1764
3004	4201	4411	3682
3005	1613	1782	1245
3006	920	968	798
3007	2965	3158	2474
3008	4017	4183	3596
3009	1286	1461	834
3010	9635	10087	8498
4001	2442	2643	1953
4002	2787	3133	2055
4003	7554	8250	5949
4004	4066	4290	3533
4005	4185	4523	3372
4006	1539	1739	971
4007	2361	2647	1788
4008	982	1120	711
4009	2510	2708	2047
4010	1212	1325	953
Average Volume	**5308**	**5575**	**4634**

Table 5.2(a) Summary Statistics of Traffic Count Data by Site ID without Zero Hourly Volumes

(Sites with higher daily volume on weekend compared to weekday are shaded)

Site ID	Average Daily Volume	Average Daily Volume (Mon-Fri)	Average Daily Volume (Sat-Sun)
242	51790	55303	44692
254	1049	955	1288
259	4108	4366	3440
306	3735	4057	2971
307	1118	1133	1085
308	6610	6901	5826
335	10020	10579	8562
358	6538	6581	6454
360	5344	5467	5050
361	6555	6751	6097
364	11954	12658	10226
393	6501	6646	5985
395	8948	8990	8790
397	13778	14141	12990
399	14288	14977	12567
400	14811	15399	13341
409	872	883	849
417	2935	3008	2745
422	2271	2358	2057
427	3063	3325	2430
443	1461	1387	1613
444	1585	1504	1745
492	6763	7015	6066
493	9873	10100	9509
495	21851	24046	15263
496	22169	24349	15630
500	64189	67046	55617
502	36377	37959	33476
513	15847	16049	15042
515	13058	13082	12998
517	24037	25626	19270
518	34603	36261	29627
524	71680	76663	61712
531	13434	13434	
533	16647	16932	15945
535	13586	14268	10432
558	65842	64584	69614

Table 5.2(a)(continuted) Summary Statistics of Traffic Count Data by Site ID without Zero Hourly Volumes

566	55531	54165	59631
571	22869	22890	22815
573	21668	22847	17421
605	22608	23561	20442
607	26930	29081	22270
608	13311	13991	11582
609	20026	21110	16558
610	10688	11074	9670
611	14527	14498	14613
630	10165	10706	8271
631	7186	8124	4490
632	11658	13413	7857
633	5175	5693	3362
634	13562	14261	11933
635	17413	18287	13915
637	56909	60549	45988
649	1027	822	1560
651	2732	2707	2844
652	17238	19220	9637
654	21945	23313	15789
669	1244	1399	913
670	---	---	---
671	---	---	---
675	20459	22135	15429
679	10447	10331	11174
681	13378	13386	13330
682	14040	13932	14306
683	41164	45236	29966
685	72219	76068	60671
686	45464	47400	40623
696	18891	19398	15951
701	30351	33754	20142
704	24841	27865	15770
733	14806	15641	12300
734	16658	17809	13699
735	15386	16738	12005
736	24629	26006	21185
994	3103	3103	---
997	2182	2182	---
1006	3751	3822	2871
1007	4246	4289	3241

Table 5.2(a)(continuted) Summary Statistics of Traffic Count Data by Site ID without Zero Hourly Volumes

1026	5912	5953	4952
1030	2890	2891	2880
1031	5168	5168	---
1032	803	803	---
1072	2186	2190	2137
1098	479	461	534
1110	5702	5748	4172
1115	6930	6930	---
1133	2201	2201	---
1134	512	512	---
1141	2227	2231	2142
1142	710	710	---
1153	3053	3050	3235
1154	286	249	323
1158	2029	2029	---
1159	1441	1441	---
1164	763	766	697
1611	5675	5970	4886
1613	10719	11495	8063
1614	10082	10082	---
1646	3757	3657	4099
1647	3142	3219	2903
1656	7183	7183	---
1661	6511	6511	---
1662	3911	3911	---
1663	3552	3713	3155
1664	8683	9110	7616
1674	4178	4210	4098
1676	1139	1197	1031
1678	2145	2145	---
1679	4809	4806	4818
1694	10077	10399	8918
1698	2390	2442	1951
1711	7624	8044	6712
1731	2760	2784	2711
1748	4263	4217	4371
1756	4390	4392	4386
1773	4745	4802	4609
1777	4755	4619	5082
1814	908	917	865
1816	4590	4784	4166

Table 5.2(a)(continuted) Summary Statistics of Traffic Count Data by Site ID without Zero Hourly Volumes

1818	2416	2523	2042
1838	1269	1252	1406
3001	3866	3866	---
3002	1759	1759	---
3003	3275	3275	---
3004	4843	4856	4513
3005	1956	1956	---
3006	1101	1101	---
3007	3464	3464	---
3008	4514	4514	---
3009	1789	1789	---
3010	10920	10920	---
4001	2832	3011	2350
4002	3206	3297	2854
4003	8571	9046	7314
4004	4846	4984	4446
4005	4815	5104	4035
4006	1768	1854	1317
4007	2747	2812	2520
4008	1151	1201	989
4009	2885	3022	2522
4010	1395	1481	1157
Average Volume	8084	8164	7832

Table 5.3 Average Hourly Volume by Weekday and Weekend for Locations in Different Districts

Time of Day	Pooled Hourly Volume (7 days)	Pooled Hourly Volume (Mon-Fri)	Pooled Hourly Volume (Sat-Sun)	Hourly Volume for District 3 (7 days)	Hourly Volume for District 4 (7 days)	Hourly Volume for District 6 (7days)	Hourly Volume for District10 (7 days)
1	53	45	73	105	139	30	26
2	34	28	47	68	83	19	18
3	27	24	34	60	61	14	16
4	28	28	26	69	56	14	19
5	51	58	32	124	119	22	37
6	117	139	59	267	275	52	88
7	195	236	91	387	492	92	144
8	249	298	124	518	639	110	182
9	270	308	174	493	746	123	190
10	286	305	238	492	762	144	204
11	298	303	284	518	768	151	222
12	286	297	257	538	729	130	231
13	349	347	353	574	914	182	251
14	346	349	339	579	884	182	258
15	366	375	343	608	931	195	268
16	384	402	337	662	982	197	283
17	390	416	326	705	1014	190	284
18	365	387	309	681	950	171	270
19	337	354	293	538	963	172	201
20	270	276	253	422	786	142	145
21	214	214	213	354	626	108	116
22	175	174	177	306	502	89	92
23	131	130	136	234	369	69	66
24	89	86	96	172	239	47	42
Frequency	24947	17886	7061	2808	3634	13025	5480

Table 5.3 (continued) Average Hourly Volume by Weekday and Weekend for Locations in Different Districts

Time of Day	Volume for District 3 (Mon-Fri)	Volume for District 3 (Sat-Sun)	Volume for District 4 (Mon-Fri)	Volume for District 4 (Sat-Sun)	Volume for District 6 (Mon-Fri)	Volume for District 6 (Sat-Sun)	Volume for District10 (Mon-Fri)	Volume For District 10 (Sat-Sun)
1	90	143	118	193	25	40	22	37
2	60	89	68	121	16	27	15	24
3	57	69	52	84	13	19	15	19
4	72	59	56	56	14	14	21	16
5	139	88	140	63	24	16	43	22
6	310	155	338	111	61	30	104	46
7	457	206	609	191	110	48	172	73
8	610	279	776	287	132	56	211	105
9	544	361	873	421	140	79	206	149
10	506	454	827	595	156	114	207	196
11	516	523	779	740	158	134	220	227
12	529	562	752	672	146	89	227	242
13	560	610	904	940	185	175	244	270
14	569	603	886	877	188	167	253	269
15	611	601	944	898	204	171	270	263
16	684	604	1014	898	212	161	292	260
17	745	601	1066	880	207	148	297	251
18	721	575	994	833	184	140	281	240
19	545	520	1027	796	182	148	202	198
20	415	441	822	692	146	133	140	158
21	341	388	640	588	108	107	110	131
22	292	342	509	484	89	88	87	105
23	223	264	368	371	69	70	61	76
24	162	199	234	253	46	49	39	50
Frequency	2031	777	2628	1006	9277	3748	3950	1530

Figure 5.1 Average Hourly Volumes by Weekday & Weekend

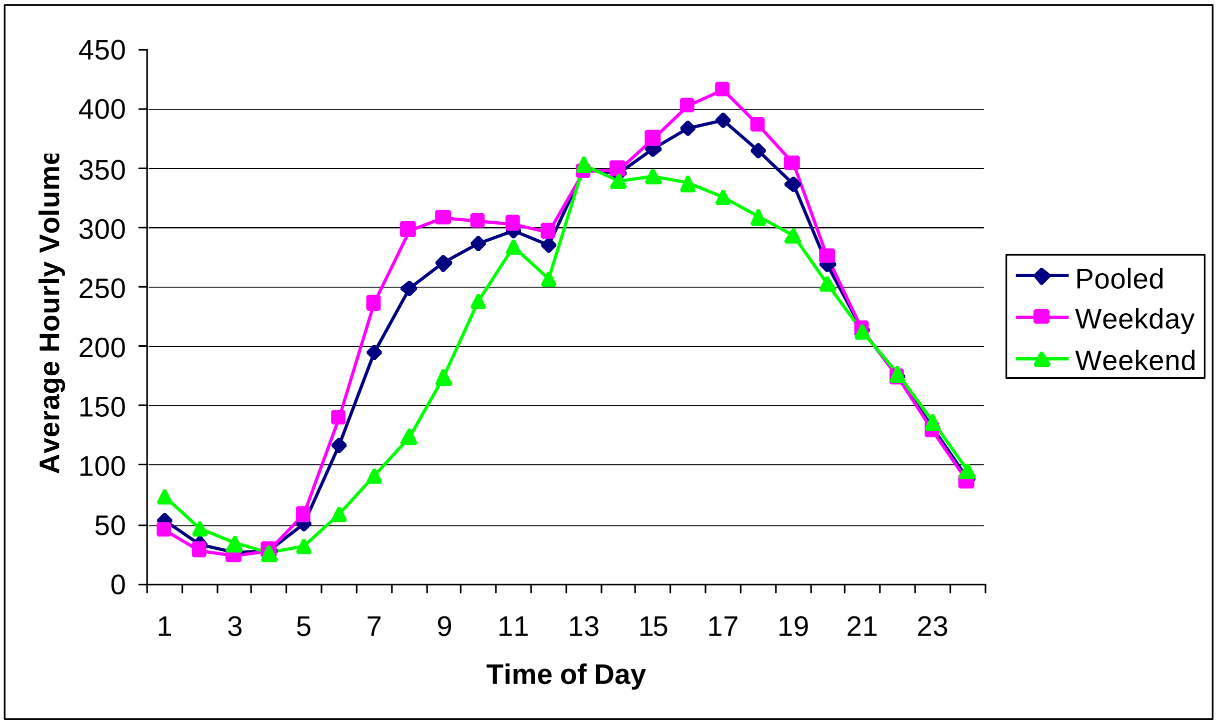


Figure 5.2 Average Hourly Volumes for 7 Days by District

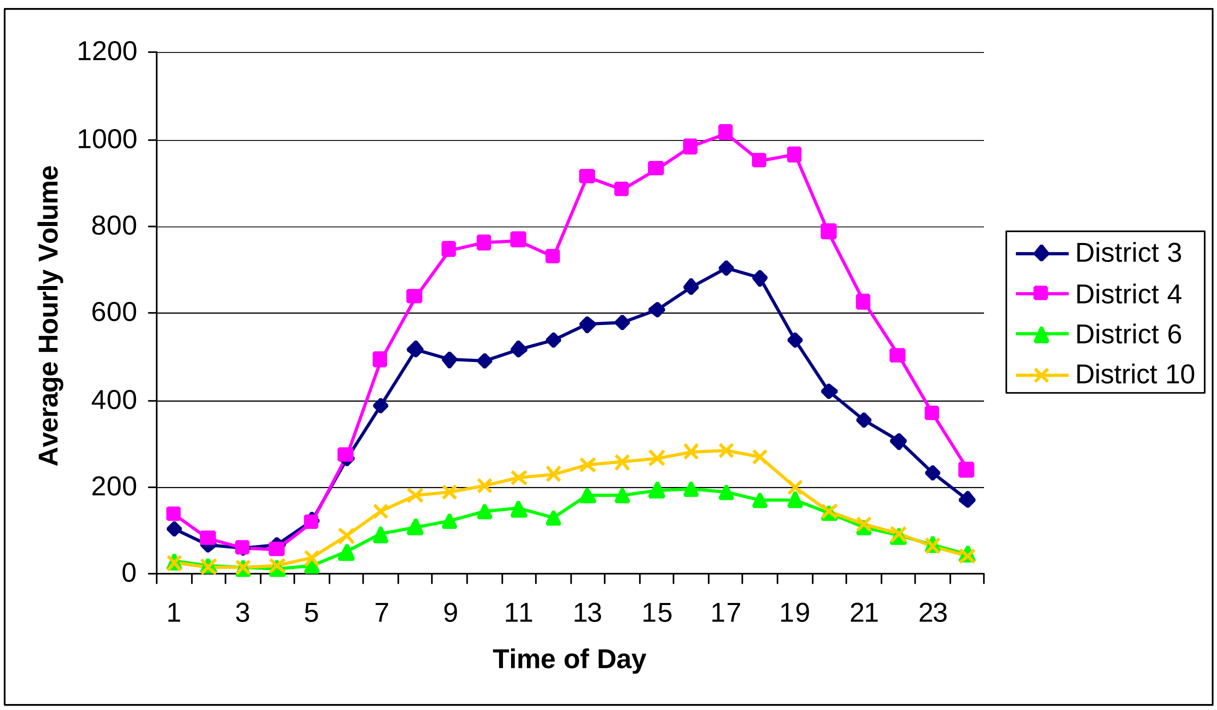


Figure 5.3 Average Hourly Volumes for District 3

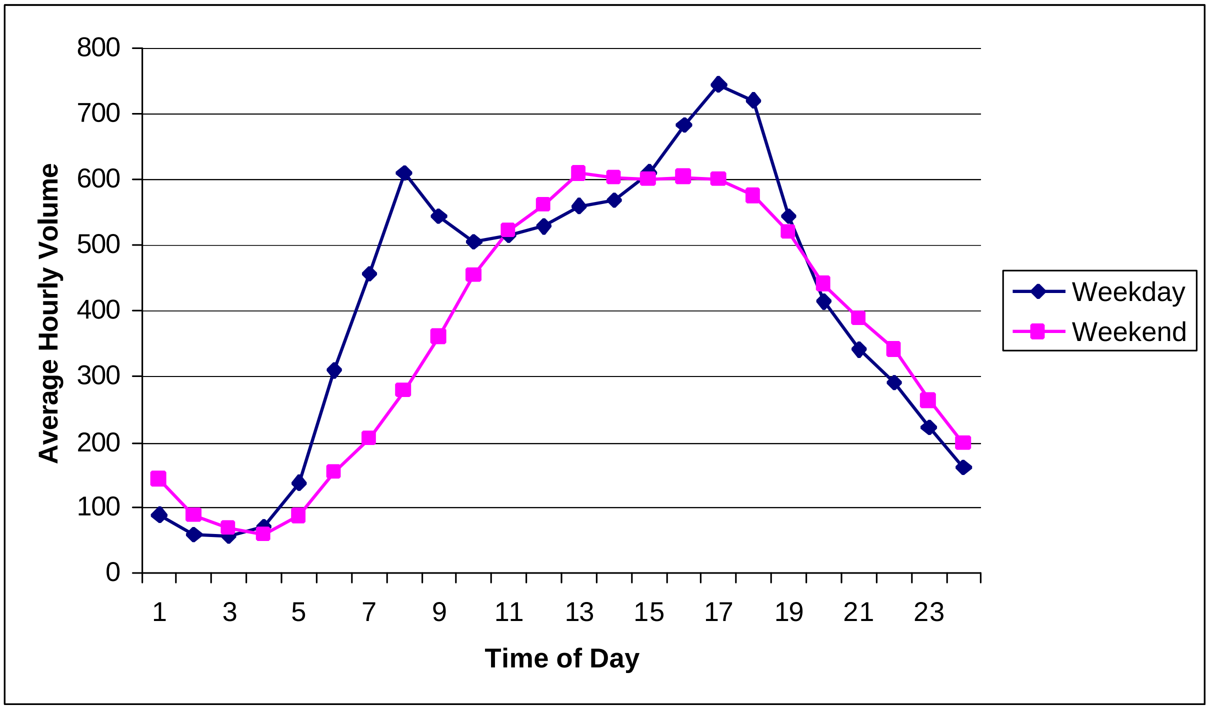


Figure 5.4 Average Hourly Volumes for District 4

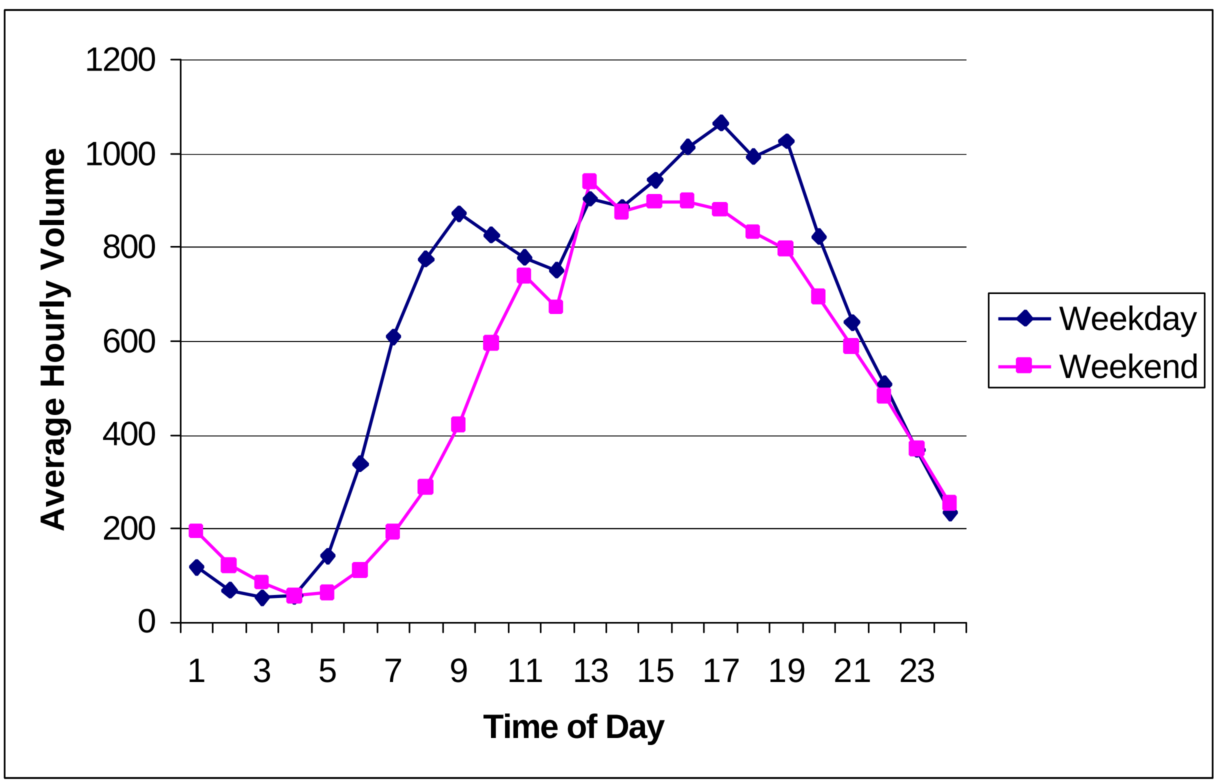


Figure 5.5 Average Hourly Volumes for District 6

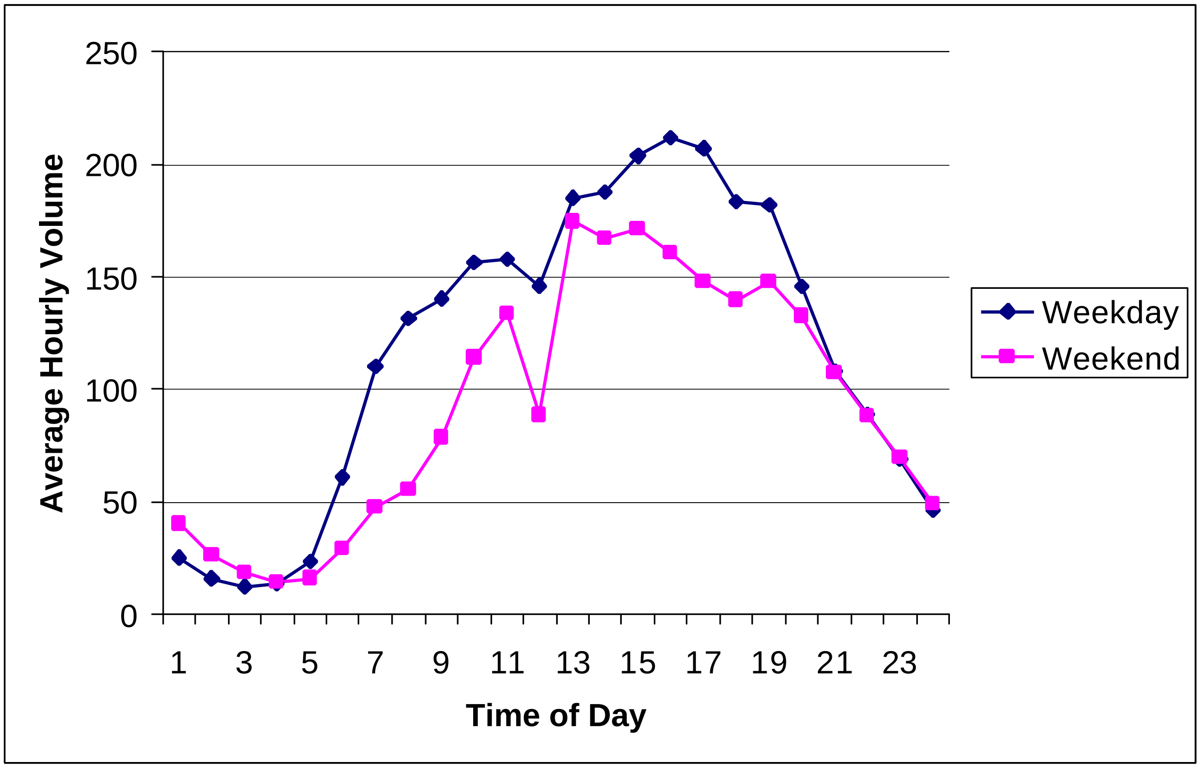


Figure 5.6 Average Hourly Volumes for District 10

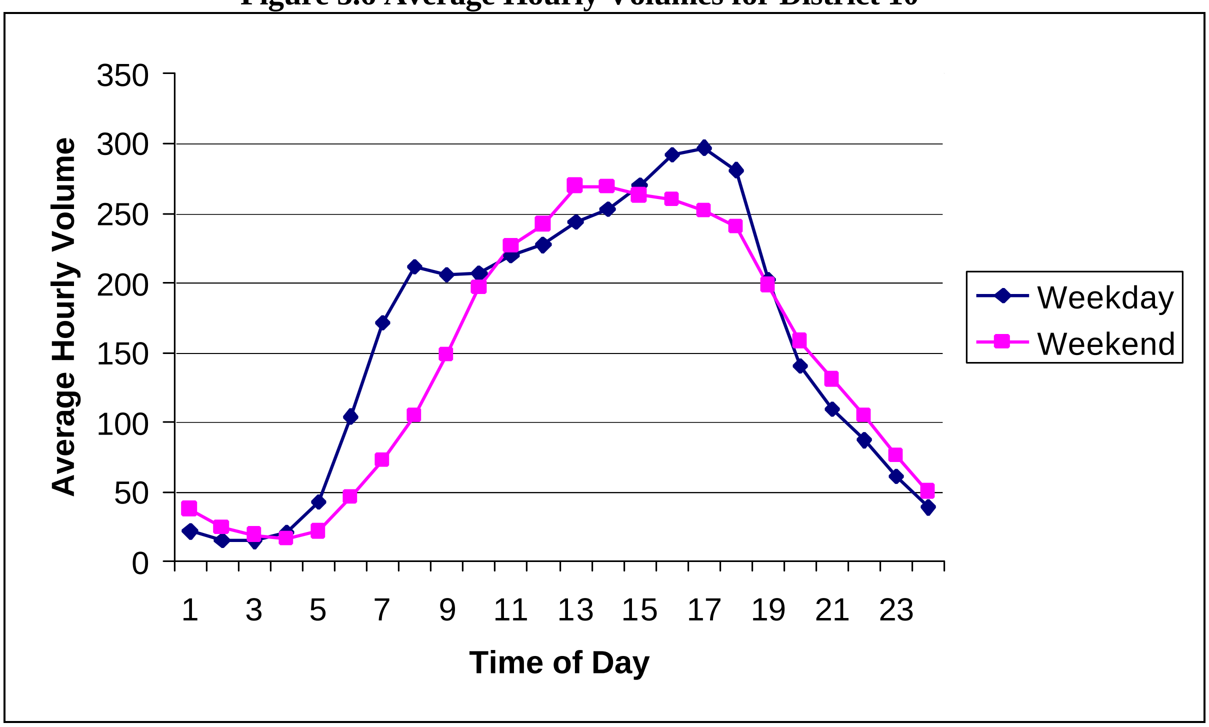


Table 5.4 Proportion of Variance Explained by Principal Components

Principal Components	% of Variance Explained	
	Weekday	Weekend
1	0.34	0.35
2	0.51	0.61
3	0.67	0.76
4	0.76	0.81
5	0.82	0.86
6	0.86	0.89
7	0.89	0.91
8	0.91	0.93
9	0.93	0.94
10	0.94	0.95
11	0.96	0.96
12	0.97	0.97
13	0.98	0.98
14	0.98	0.98
15	0.99	0.99
16	0.99	0.99
17	0.99	0.99
18	1.00	1.00
19	1.00	1.00
20	1.00	1.00
21	1.00	1.00
22	1.00	1.00
23	1.00	1.00
24	1.00	1.00

Table 5.5 Linear Combinations of Principal Components

Eigenvectors[1] (coefficients)	Principal Components (PC) for Weekday Traffic							
	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8
β_1	-0.007	0.078	-0.007	-0.023	-0.130	0.070	-0.139	-0.064
β_2	0.004	0.038	-0.020	-0.029	-0.097	0.057	-0.081	-0.064
β_3	0.025	0.017	-0.035	-0.045	-0.083	0.059	-0.058	-0.073
β_4	0.067	-0.002	-0.052	-0.084	-0.120	0.039	-0.029	-0.080
β_5	0.166	0.058	-0.152	-0.200	-0.072	0.153	0.068	-0.209
β_6	0.439	0.066	-0.269	-0.257	0.085	0.303	0.487	-0.127
β_7	0.539	0.111	-0.128	-0.125	0.242	-0.236	-0.260	0.490
β_8	0.375	-0.051	0.518	0.335	-0.297	-0.432	0.328	-0.111
β_9	0.148	0.023	0.098	0.463	0.337	0.238	-0.276	-0.040
β_{10}	0.028	-0.076	-0.204	0.431	0.181	-0.014	-0.121	-0.108
β_{11}	-0.033	-0.218	-0.240	0.214	-0.013	0.014	-0.039	-0.019
β_{12}	-0.022	-0.336	-0.037	0.093	-0.255	0.219	-0.104	-0.330
β_{13}	-0.193	-0.257	-0.209	0.090	-0.119	-0.149	0.156	0.218
β_{14}	-0.183	-0.312	-0.201	0.014	-0.026	-0.041	0.113	0.256
β_{15}	-0.136	-0.251	-0.095	-0.125	-0.002	-0.218	0.189	0.229
β_{16}	-0.136	-0.199	0.113	-0.327	0.214	-0.296	-0.259	-0.129
β_{17}	-0.081	-0.137	0.257	-0.311	0.438	-0.181	0.028	-0.406
β_{18}	-0.075	-0.198	0.571	-0.121	0.025	0.554	0.052	0.425
β_{19}	-0.244	0.282	0.051	0.156	0.379	-0.020	0.242	0.051
β_{20}	-0.287	0.381	-0.035	0.097	0.102	-0.017	0.209	0.000
β_{21}	-0.210	0.317	-0.013	0.018	-0.089	0.027	0.141	-0.003
β_{22}	-0.093	0.279	0.043	-0.076	-0.275	-0.143	-0.046	0.101
β_{23}	-0.062	0.219	0.022	-0.113	-0.239	-0.036	-0.299	0.047
β_{24}	-0.030	0.166	0.025	-0.078	-0.186	0.050	-0.305	-0.054

[1] Large magnitude of the coefficients, positive or negative alike, indicate that relatively larger weight are given to the corresponding inputs in the composition of the principal components. They are shaded for easier viewing.

Table 5.5 (continued) Linear Combinations of Principal Components

Eigenvectors[1] (coefficients)	Principal Components (PC) for Weekend Traffic							
	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8
β_1	-0.159	0.031	0.072	0.008	-0.123	-0.096	-0.238	-0.243
β_2	-0.104	0.032	0.029	0.034	-0.087	-0.038	-0.145	-0.177
β_3	-0.073	0.041	0.008	0.042	-0.055	-0.020	-0.115	-0.157
β_4	-0.046	0.062	-0.019	0.085	-0.023	0.006	-0.054	-0.087
β_5	-0.035	0.104	-0.065	0.123	0.018	0.051	0.029	-0.089
β_6	-0.012	0.203	-0.198	0.288	-0.076	0.170	0.136	0.107
β_7	-0.035	0.237	-0.214	0.323	-0.058	0.269	0.289	-0.032
β_8	0.048	0.254	-0.149	0.228	0.111	0.066	0.106	0.074
β_9	0.087	0.209	-0.298	-0.038	0.221	-0.005	-0.207	0.064
β_{10}	0.088	0.100	-0.448	-0.328	0.153	-0.149	-0.156	0.170
β_{11}	0.130	-0.040	-0.267	-0.318	0.102	-0.366	0.004	-0.141
β_{12}	0.547	0.472	0.488	-0.335	-0.186	0.075	0.082	0.115
β_{13}	0.185	-0.345	-0.176	-0.109	-0.151	-0.252	0.543	-0.055
β_{14}	0.270	-0.292	-0.047	-0.036	-0.250	0.414	0.015	-0.221
β_{15}	0.214	-0.353	-0.075	0.051	-0.319	0.121	-0.396	-0.069
β_{16}	0.191	-0.260	0.125	0.390	-0.032	-0.236	-0.231	0.646
β_{17}	0.172	-0.193	0.293	0.278	0.409	-0.269	0.299	-0.225
β_{18}	0.055	-0.080	0.213	-0.048	0.617	0.223	-0.220	-0.090
β_{19}	-0.185	-0.219	0.091	-0.160	0.183	0.328	0.014	0.061
β_{20}	-0.297	-0.169	0.064	-0.236	0.007	0.243	0.195	0.256
β_{21}	-0.294	-0.049	0.096	-0.267	-0.021	0.065	0.163	0.204
β_{22}	-0.276	0.064	0.148	-0.086	-0.134	-0.174	-0.002	0.190
β_{23}	-0.266	0.106	0.171	0.057	-0.171	-0.229	-0.008	0.023
β_{24}	-0.206	0.085	0.157	0.055	-0.132	-0.196	-0.102	-0.325

[1] Large magnitude of the coefficients, positive or negative alike, indicate that relatively larger weight are given to the corresponding inputs in the composition of the principal components. They are shaded for easier viewing.

Table 5.6 Pseudo F-Ratio Values by Number of Clusters

No. of clusters	Pseudo F-ratio[1]	
	Weekday	Weekend
1	---	---
2	8.9	11.8
3	6.7	21.9
4	43	15.5
5	35.4	13.6
6	46.8	35.7
7	45.4	47.8
8	40.8	44.2
9	45	39.5
10	68.8	40.3
11	68	50.3
12	67.8	57.9
13	69.2	78.7
14	82.5	77.3
15	86.7	77.6
16	88.1	87.1
17	84.3	85.7
18	91	100
19	90.6	102
20	101	104

[1] F-ratio measures the ratio of the mean square error between clusters to the mean square error within clusters and a large F-ratio value is preferred.

Table 5.7 Estimated Allocation Factors Using Aggregate Daily Volumes

Allocation Factors	7-Day Week		Weekday (Mon-Fri)		Weekend (Sat-Sun)	
	Cluster 1	Cluster 2	Cluster1	Cluster2	Cluster 1	Cluster 2
β_1	0.011	0.003	0.008	0.009	0.017	0.017
β_2	0.006	0.002	0.005	0.005	0.011	0.011
β_3	0.005	0.002	0.004	0.003	0.008	0.007
β_4	0.004	0.003	0.005	0.003	0.005	0.005
β_5	0.008	0.005	0.013	0.004	0.006	0.004
β_6	0.022	0.016	0.035	0.011	0.011	0.007
β_7	0.035	0.027	0.053	0.023	0.017	0.012
β_8	0.049	0.038	0.066	0.042	0.025	0.016
β_9	0.051	0.036	0.060	0.051	0.035	0.029
β_{10}	0.051	0.036	0.055	0.049	0.048	0.045
β_{11}	0.053	0.044	0.052	0.049	0.060	0.060
β_{12}	0.056	0.058	0.052	0.054	0.066	0.021
β_{13}	0.062	0.075	0.055	0.064	0.073	0.090
β_{14}	0.063	0.072	0.056	0.065	0.074	0.084
β_{15}	0.066	0.084	0.060	0.070	0.072	0.090
β_{16}	0.071	0.108	0.066	0.077	0.073	0.090
β_{17}	0.073	0.114	0.069	0.080	0.072	0.078
β_{18}	0.074	0.084	0.071	0.082	0.068	0.065
β_{19}	0.064	0.055	0.060	0.071	0.062	0.073
β_{20}	0.052	0.039	0.046	0.059	0.053	0.063
β_{21}	0.043	0.035	0.037	0.047	0.047	0.047
β_{22}	0.036	0.028	0.031	0.038	0.041	0.038
β_{23}	0.026	0.021	0.023	0.027	0.032	0.029
β_{24}	0.018	0.015	0.016	0.018	0.022	0.020
Observation	131	4	92	43	129	6

Figure 5.7 Allocation Factors for 7-Day Week

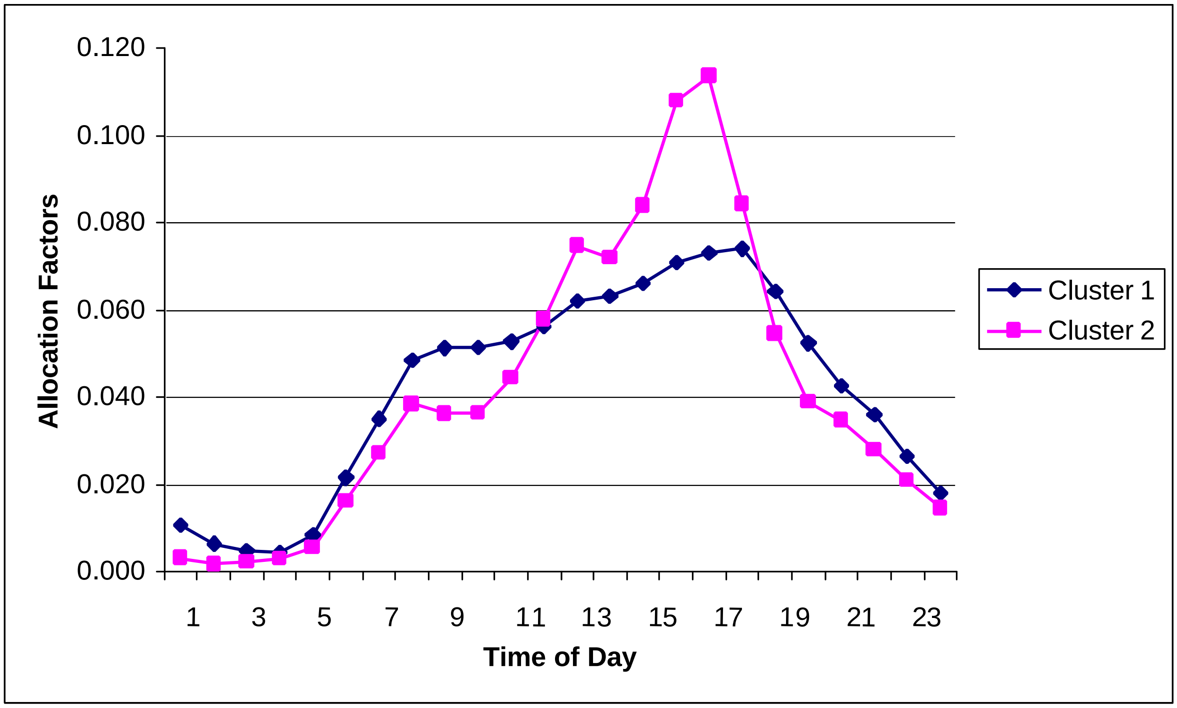


Figure 5.8 Allocation Factors for Weekday

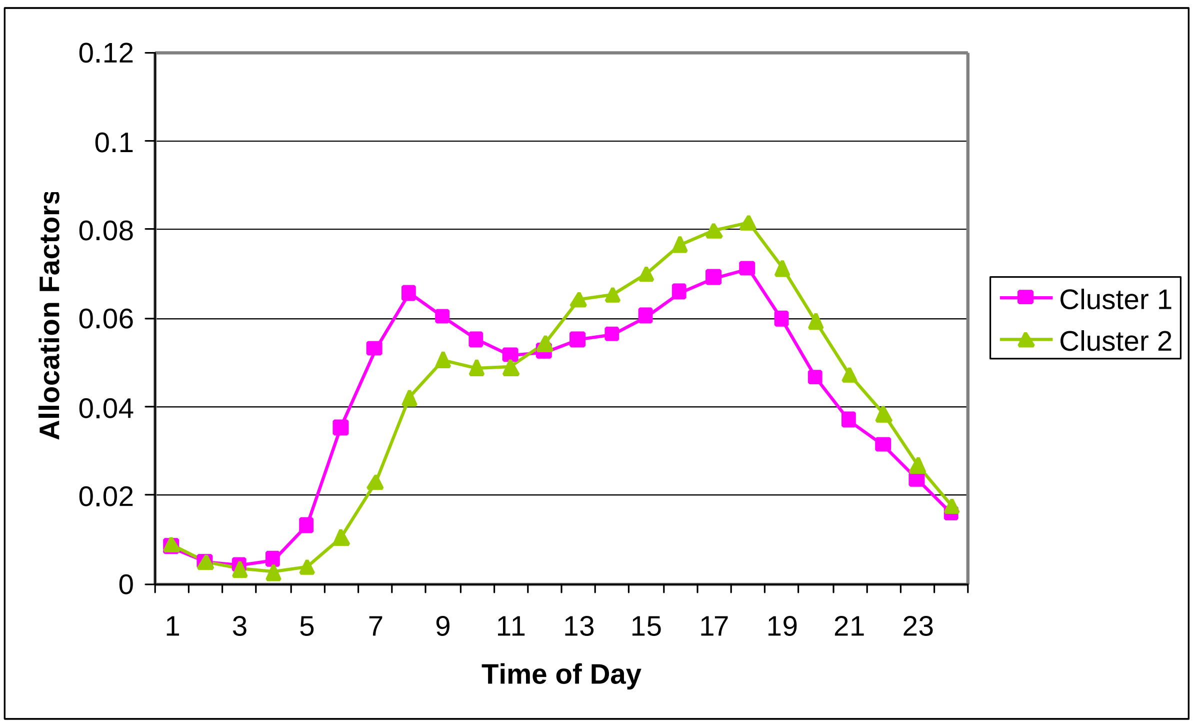


Figure 5.9 Allocation Factors for Weekend

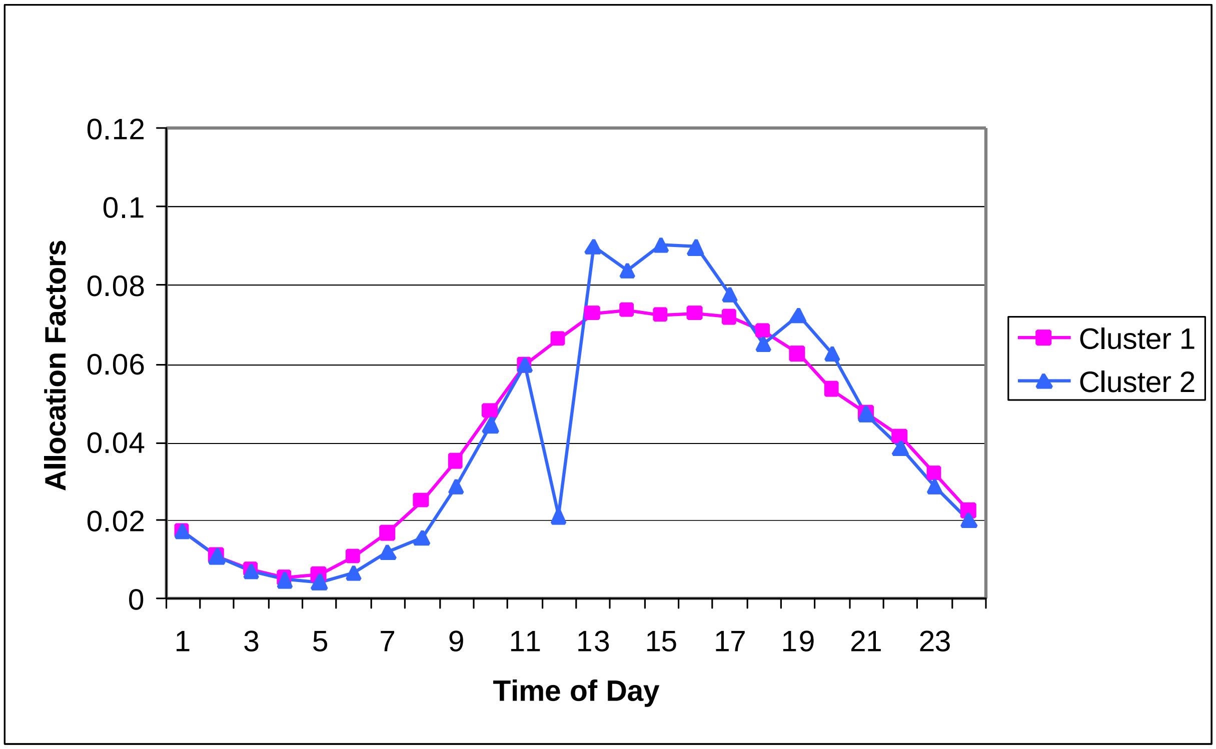


Table 5.8 Descriptive Statistics of Traffic Count Data by Cluster for 135 Matched with Model Links

CLUSTER	No. of Locations	Mean Daily Volume	Pooled Variance (Daily Volume)
7-Day Week			
1	131	19,820	835,021,782
2	4	9,342	211,689,638
Total	135	19,509	818,014,961
Weekday (Monday-Friday)			
1	92	20,113	656,918,517
2	43	19,005	1203,822,285
Total	135	19,754	826,227,450
Weekend (Saturday-Sunday)			
1	129	20,198	859,143,560
2	6	5,664	23,046,253
Total	135	19,542	830,132,477

Table 5.9 Comparison of Classification Results Using Methods 1 and 2

Cluster	No. of links	Mean Daily Volume	Variance of Daily Volume
Weekday by Method 1 (Using Daily Volumes)			
1	12,719	46,601	1,859,993,526
2	53,383	5,351	27,781,521
Weekend by Method 1			
1	19,266	36,167	1,440,611,524
2	46,836	3,877	13,441,347
Weekday by Method 2 (Using Imputed Period-based Volumes)			
1	35,525	17,710	1,069,001,973
2	30,577	8,151	102,676,302
Weekday by Method 2			
1	35,785	18,005	1,040,721,055
2	30,317	7,721	120,052,681

Table 5.10 Imputed Period-Based Volumes for 135 Links with Matched Traffic Count Data

	AM Peak (6am-9am)	Mid-Day (9am-3pm)	PM Peak (3pm-7pm)	Off-Peak (7pm-6am)
Monday-Friday				
Mean	2,667	6,251	5,408	5,556
Std. Dev.	3,709	9,379	8,183	7,777
Saturday-Sunday				
Mean	1,315	6,649	5,546	6,395
Std. Dev.	2,006	10,073	8,139	8,830

Figure 5.10 Clustering Results Using Method 1 Based on Weekday Traffic

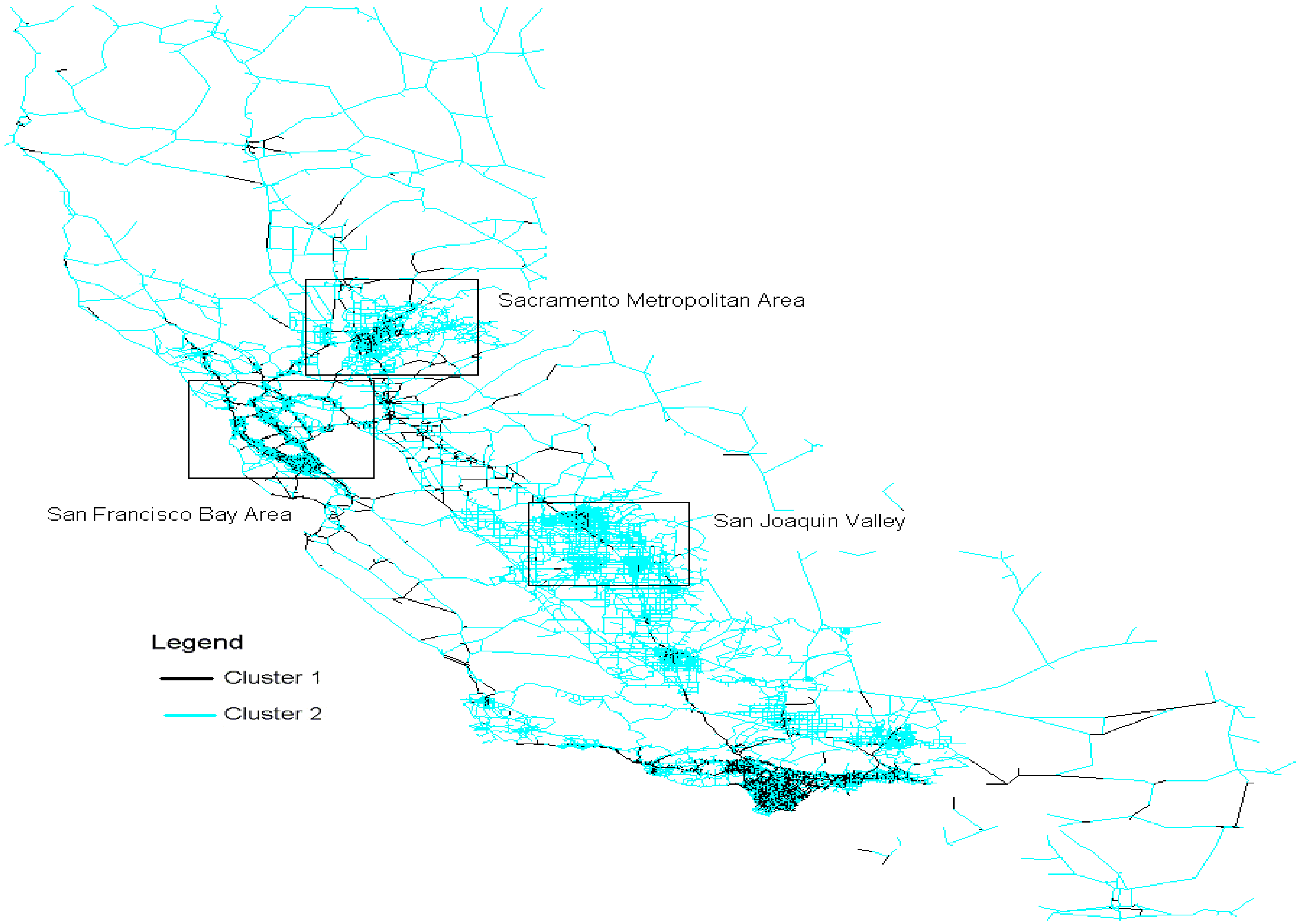


Figure 5.10(a) **Clustering Results for the Sacramento Metropolitan Area Using Method 1 Based on Weekday Traffic**

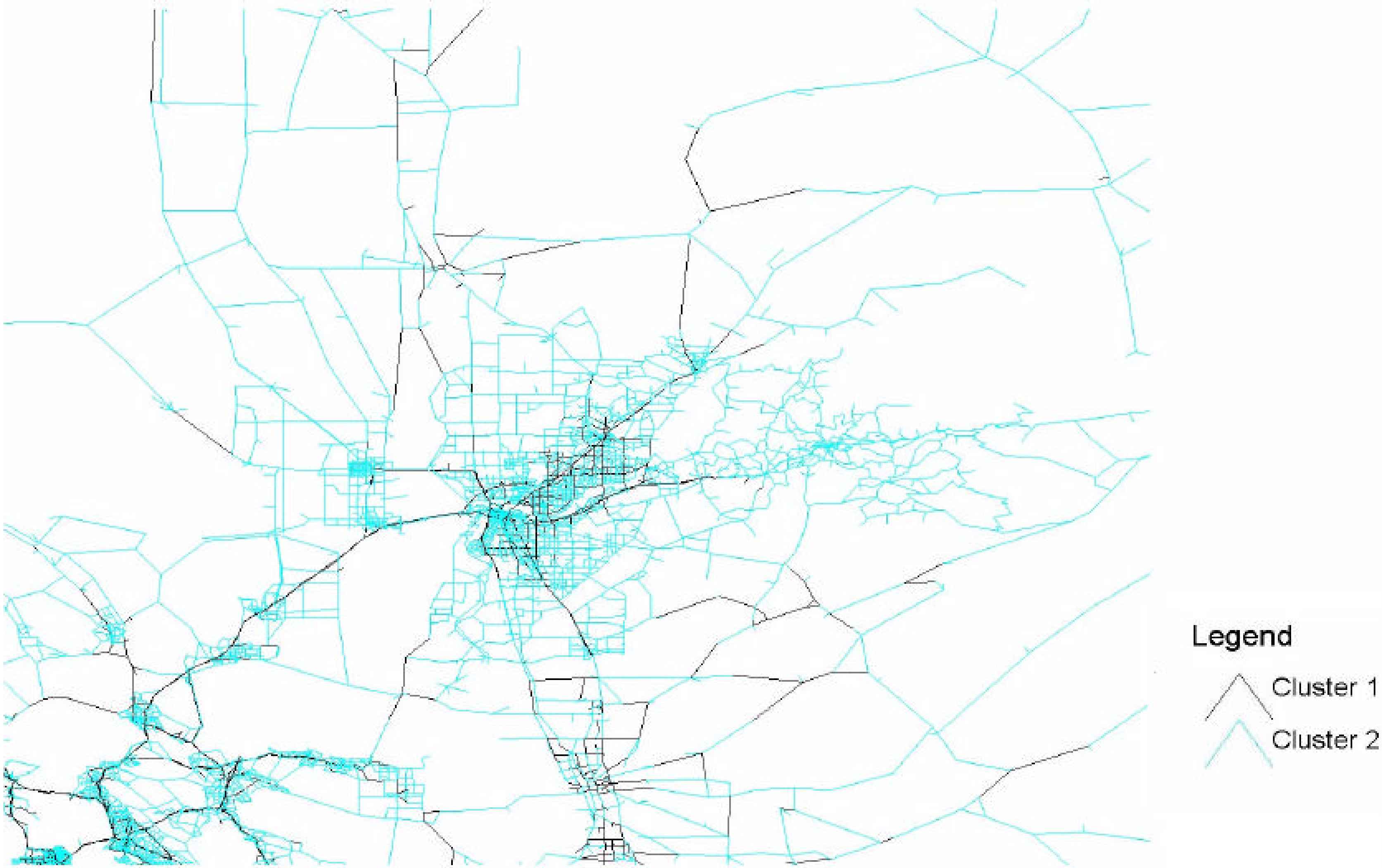


Figure 5.10(b) Clustering Results for the San Francisco Bay Area Using Method 1 Based on Weekday Traffic

Figure 5.10(c) Clustering Results for the San Joaquin Valley Using Method 1 Based on Weekday Traffic

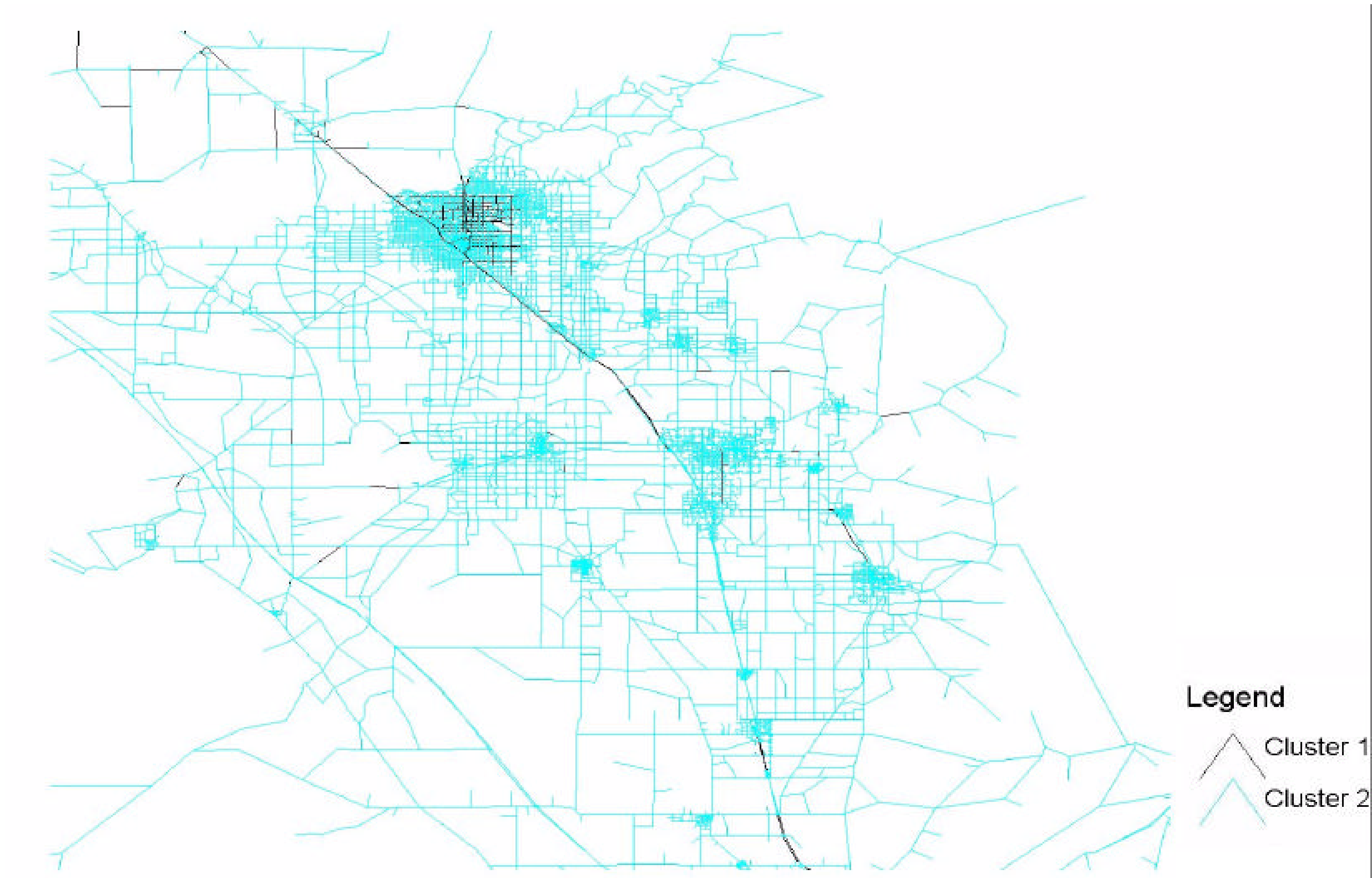


Figure 5.11 Clustering Results Using Method 1 Based on Weekend Traffic

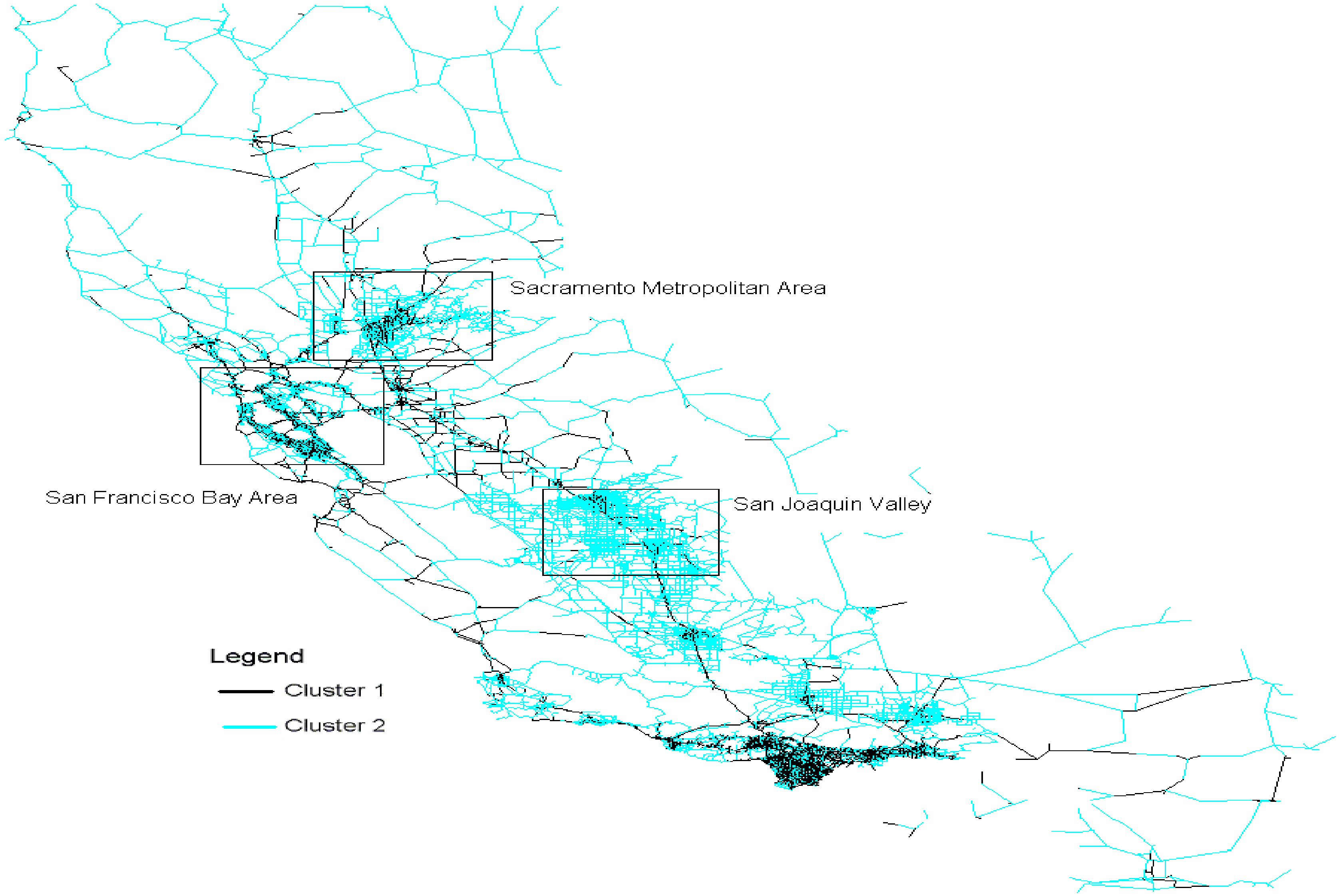


Figure 5.11(a) Clustering Results for the Sacramento Metropolitan Area Using Method 1 Based on Weekend Traffic

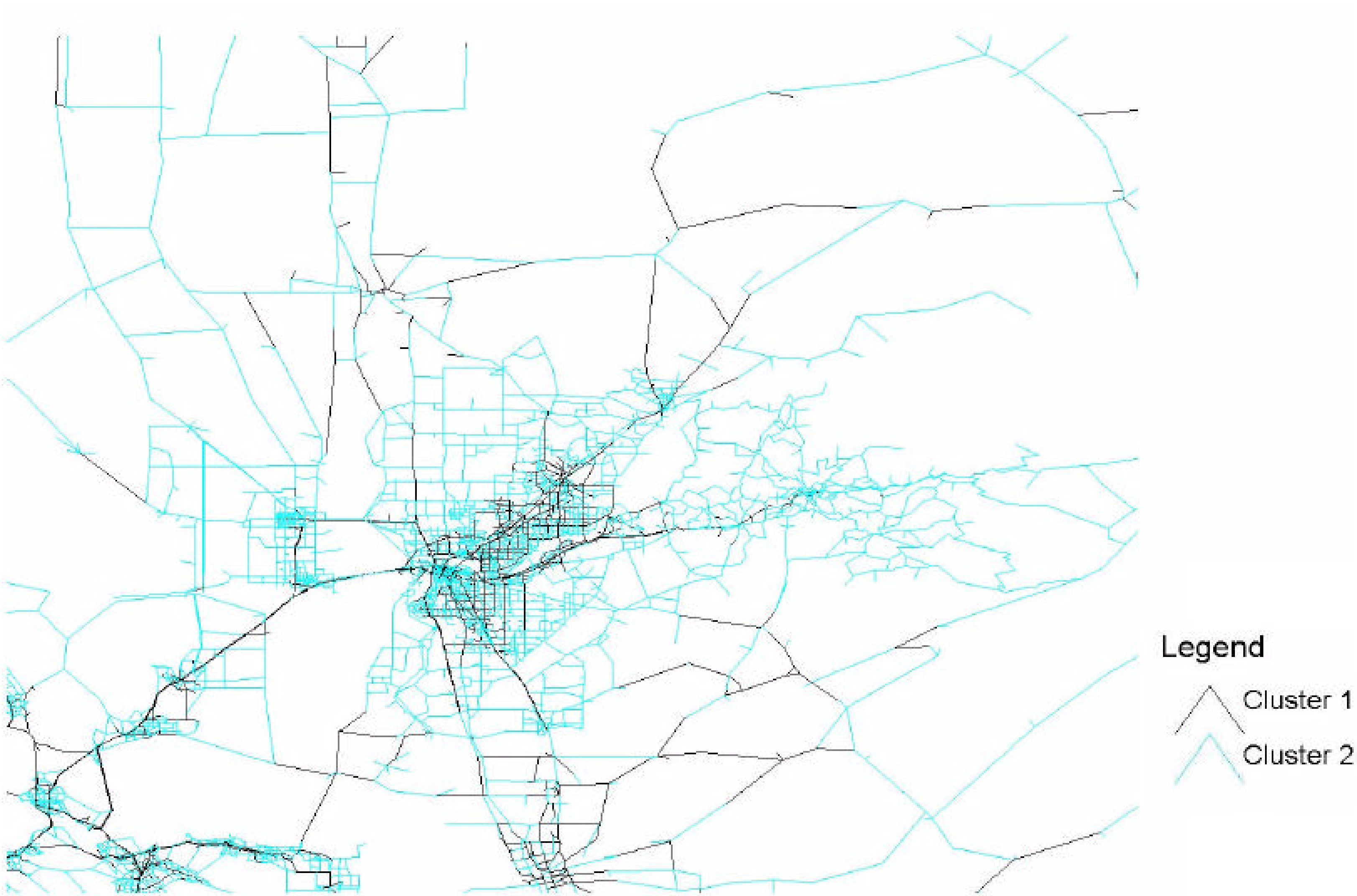


Figure 5.11(b) Clustering Results for the San Francisco Bay Area Using Method 1 Based on Weekend Traffic

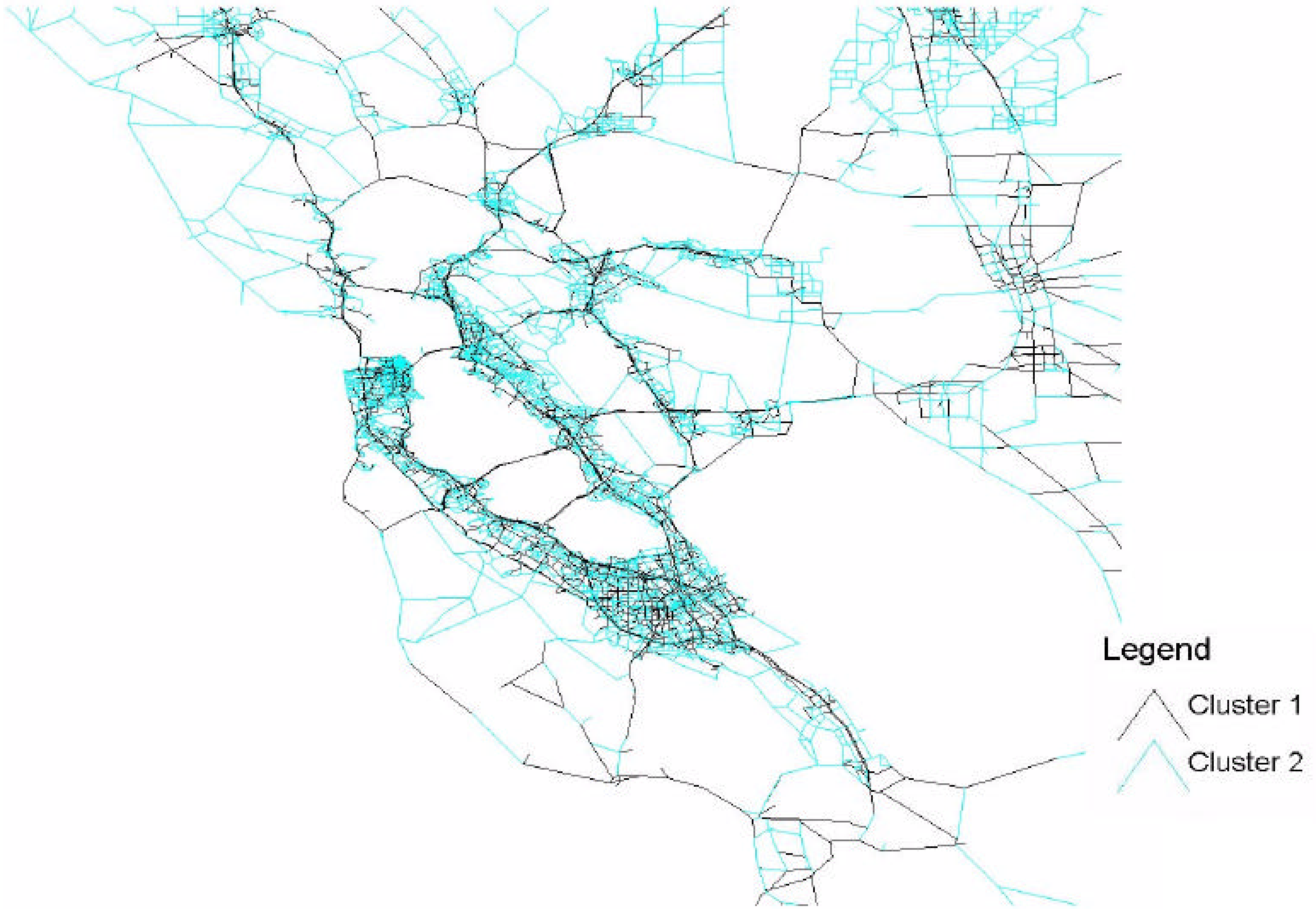


Figure 5.11(c) Clustering Results for the San Joaquin Valley Using Method 1 Based on Weekend Traffic

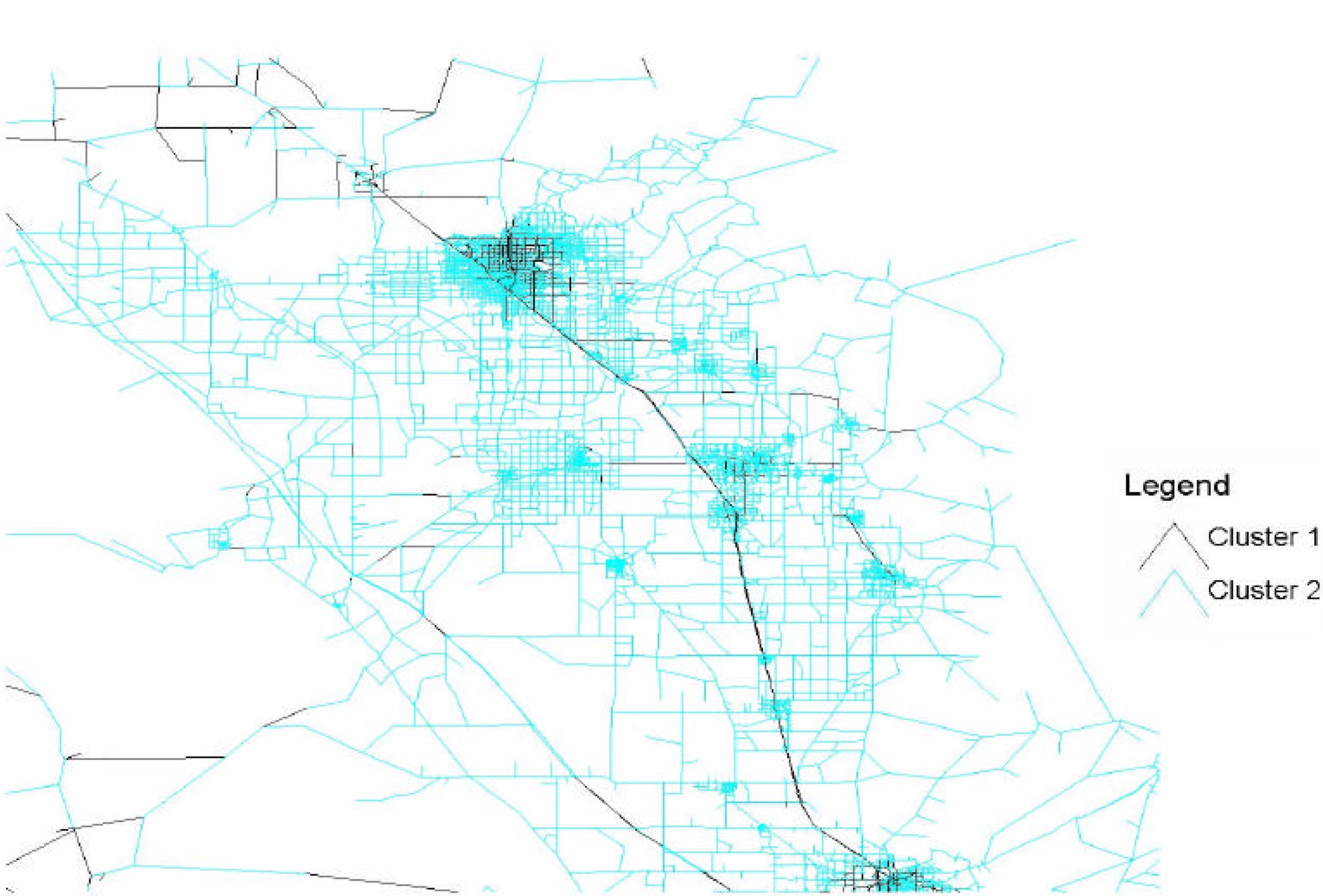


Figure 5.12 Clustering Results Using Method 2 Based on Weekday Traffic

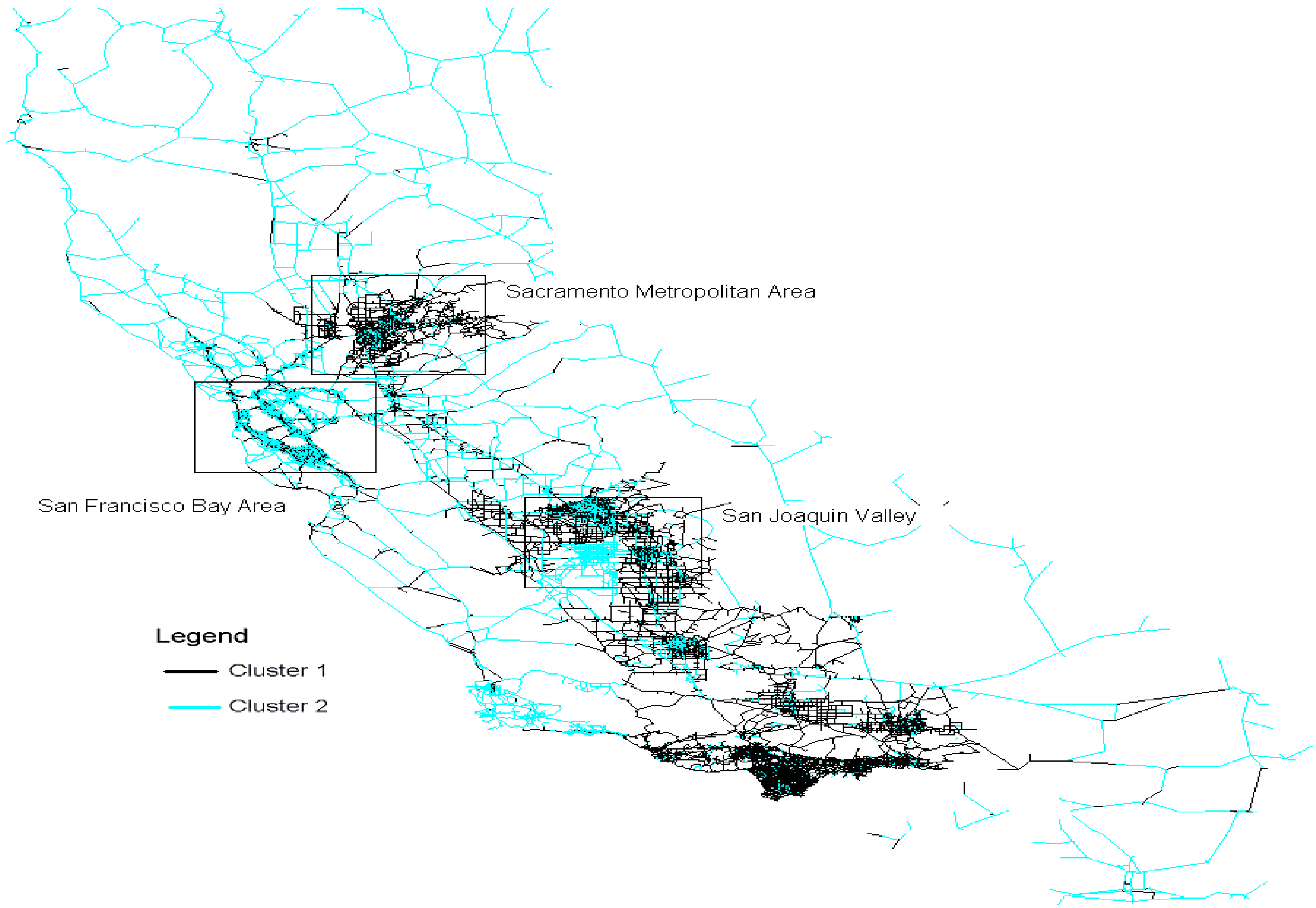


Figure 5.12(a) Clustering Results for the Sacramento Metropolitan Area Using Method 2 Based on Weekday Traffic

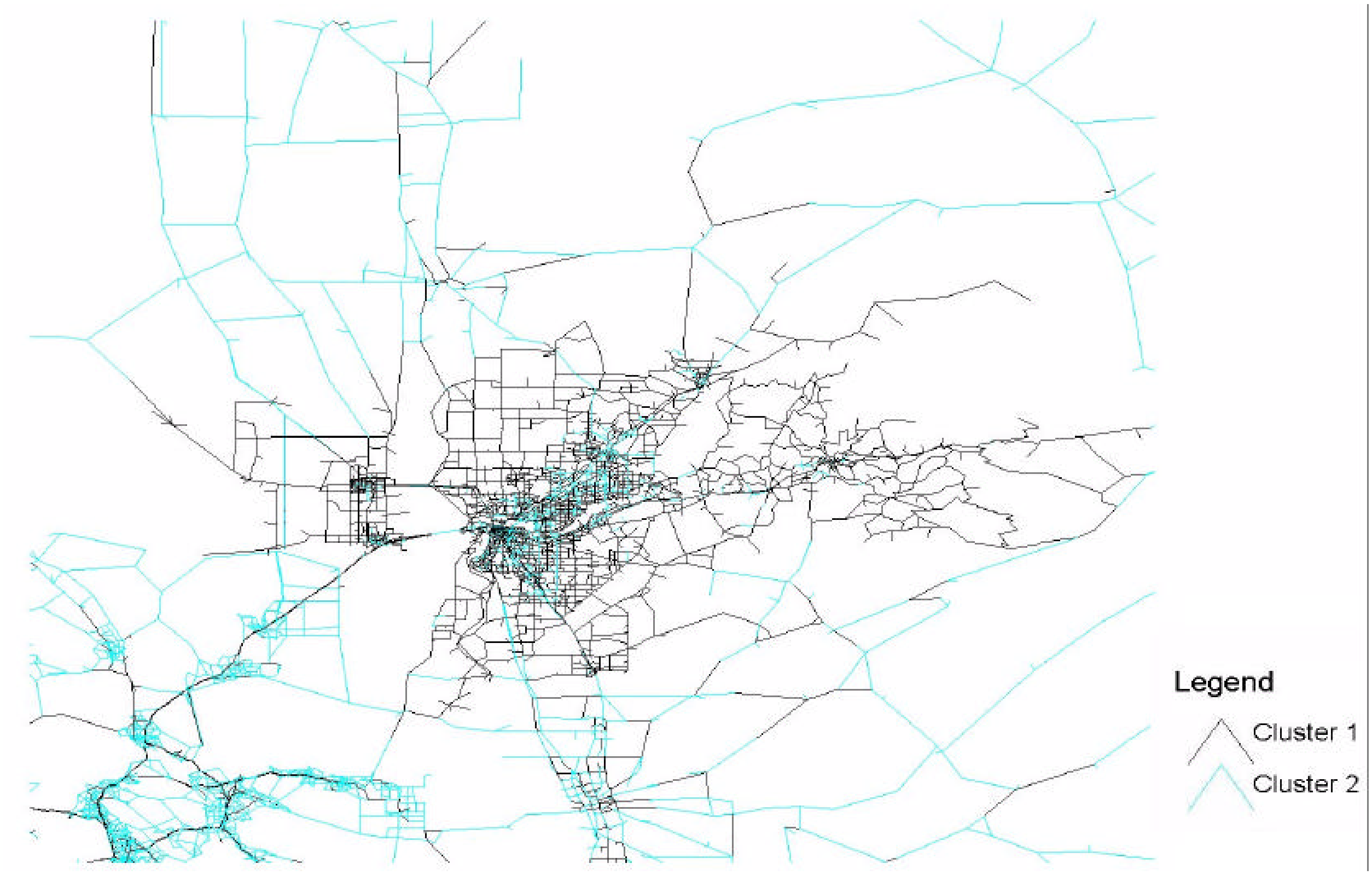


Figure 5.12(b) Clustering Results for the San Francisco Bay Area Using Method 2 Based on Weekday Traffic

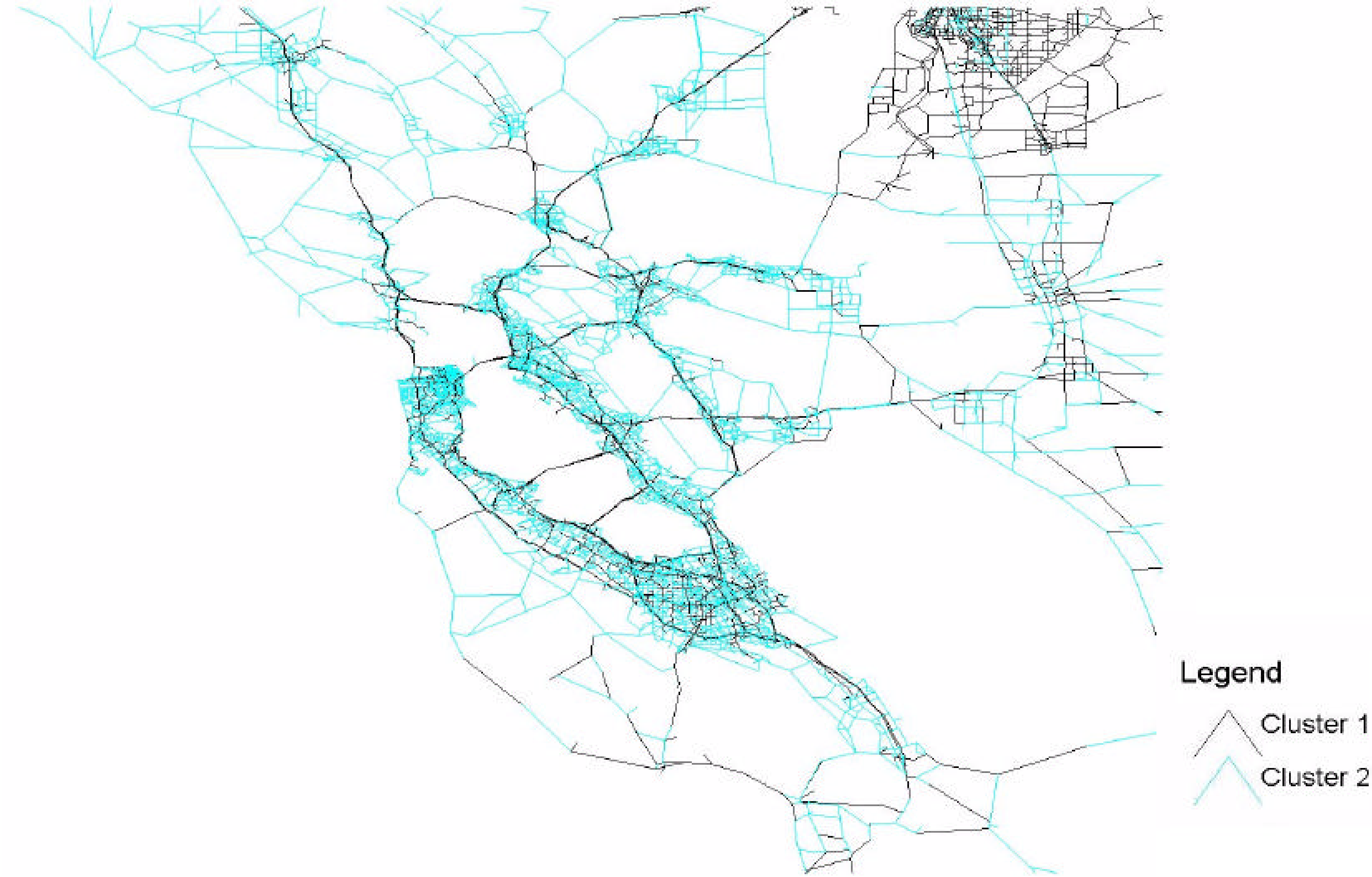


Figure 5.12(c) Clustering Results for the San Joaquin Valley Using Method 2 Based on Weekday Traffic

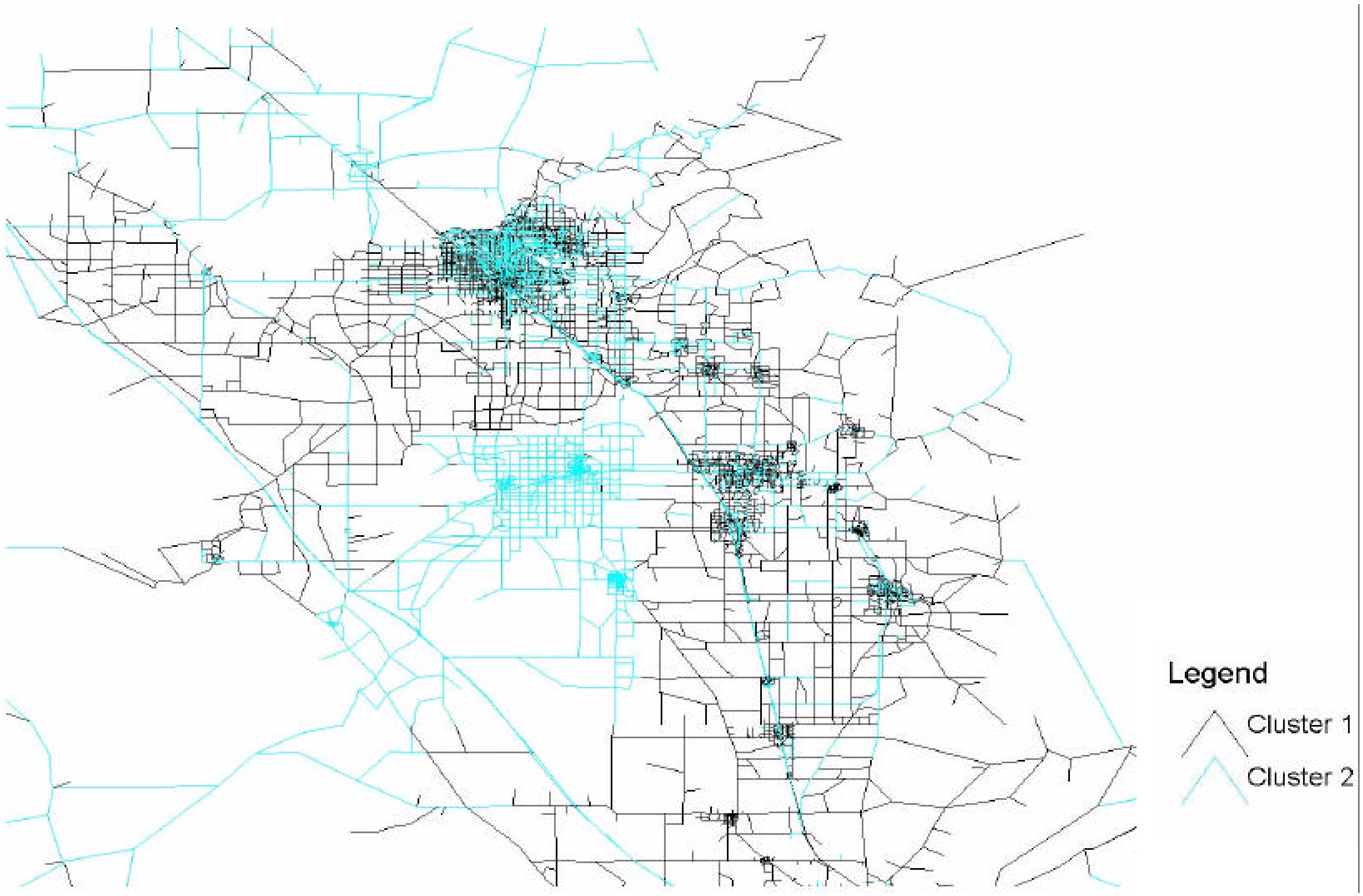


Figure 5.13 Clustering Results Using Method 2 Based on Weekend Traffic

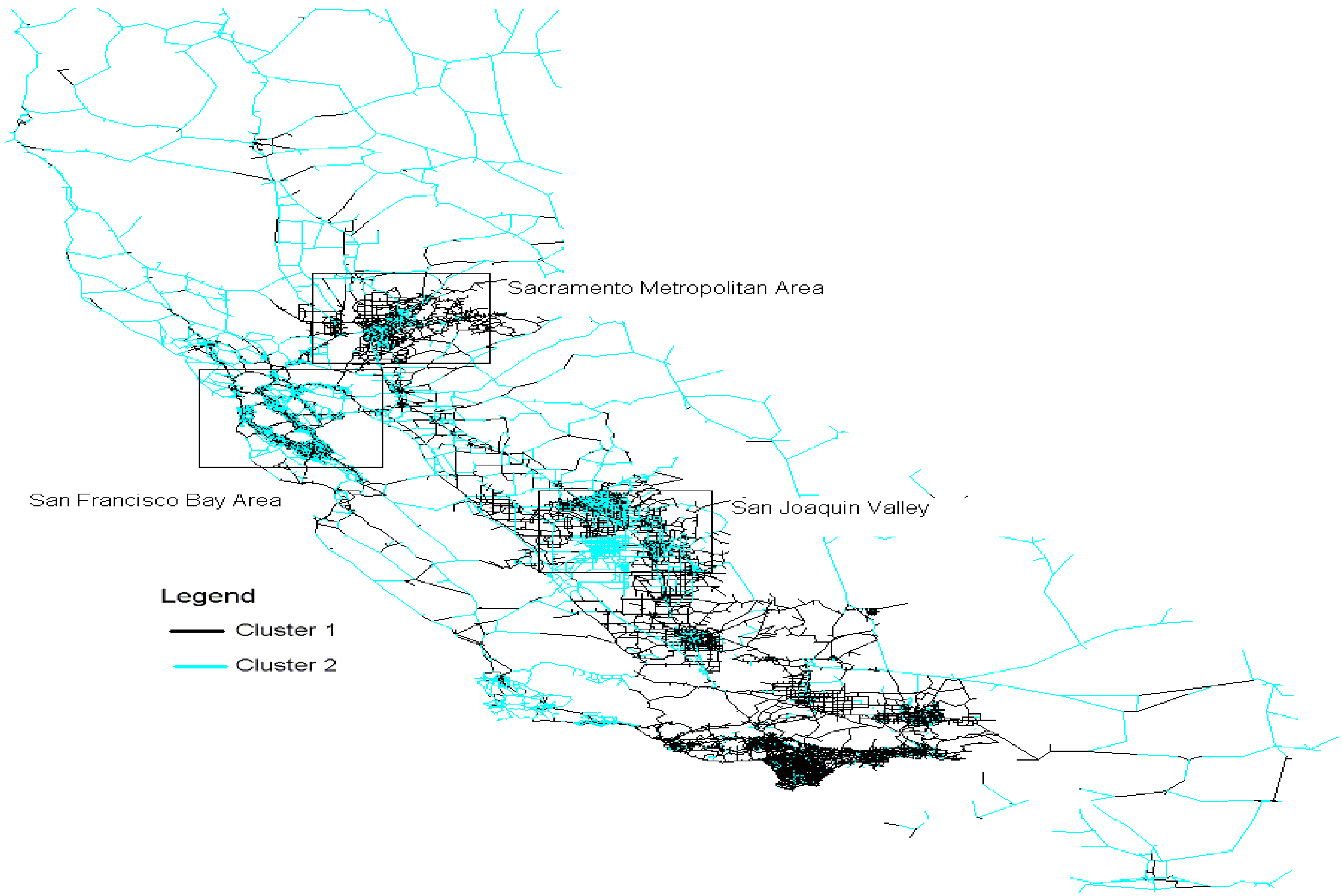


Figure 5.13(a) **Clustering Results for the Sacramento Metropolitan Area Using Method 2 Based on Weekend Traffic**

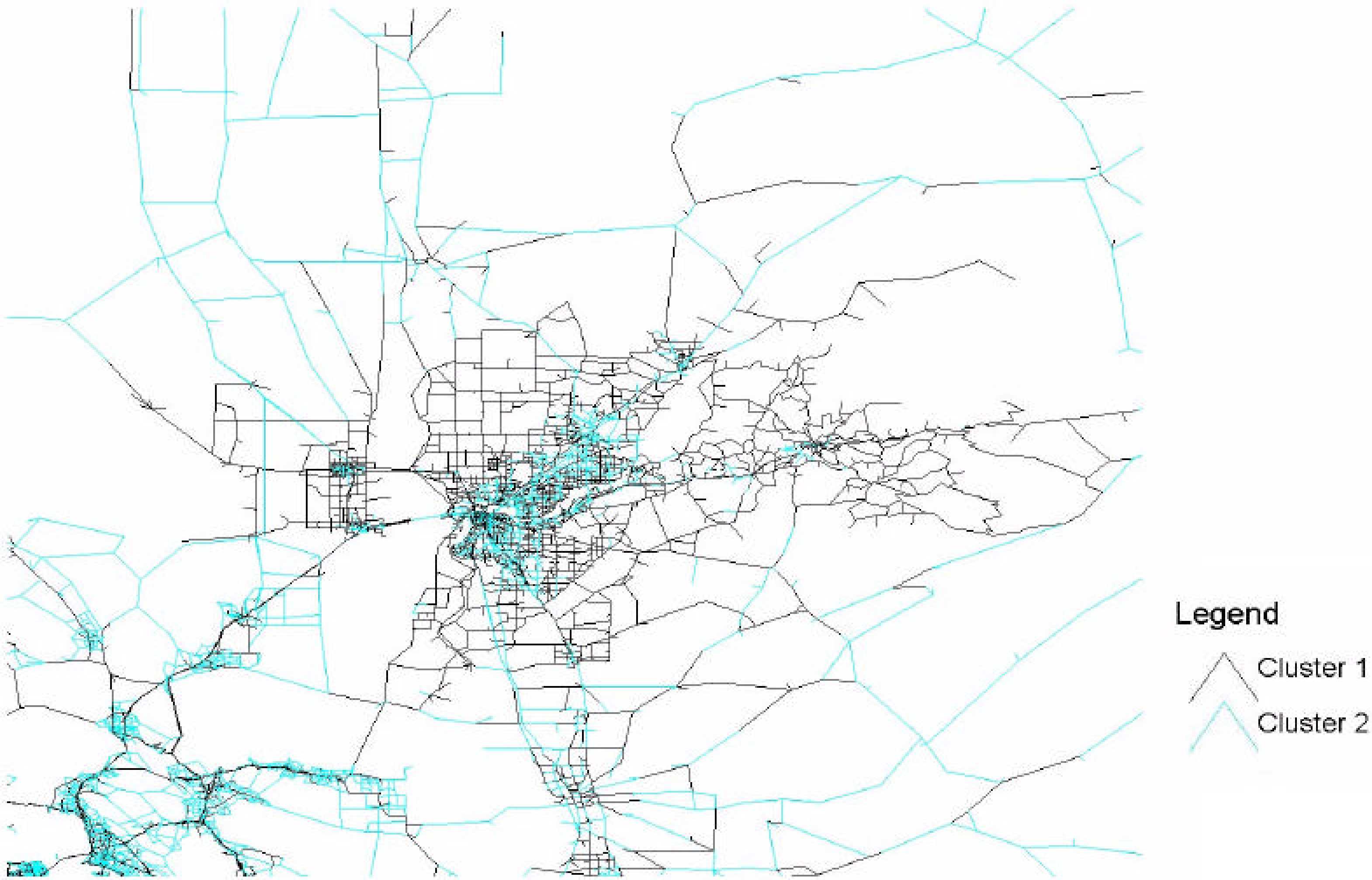


Figure 5.13(b) Clustering Results for the San Francisco Bay Area Using Method 2 Based on Weekend Traffic

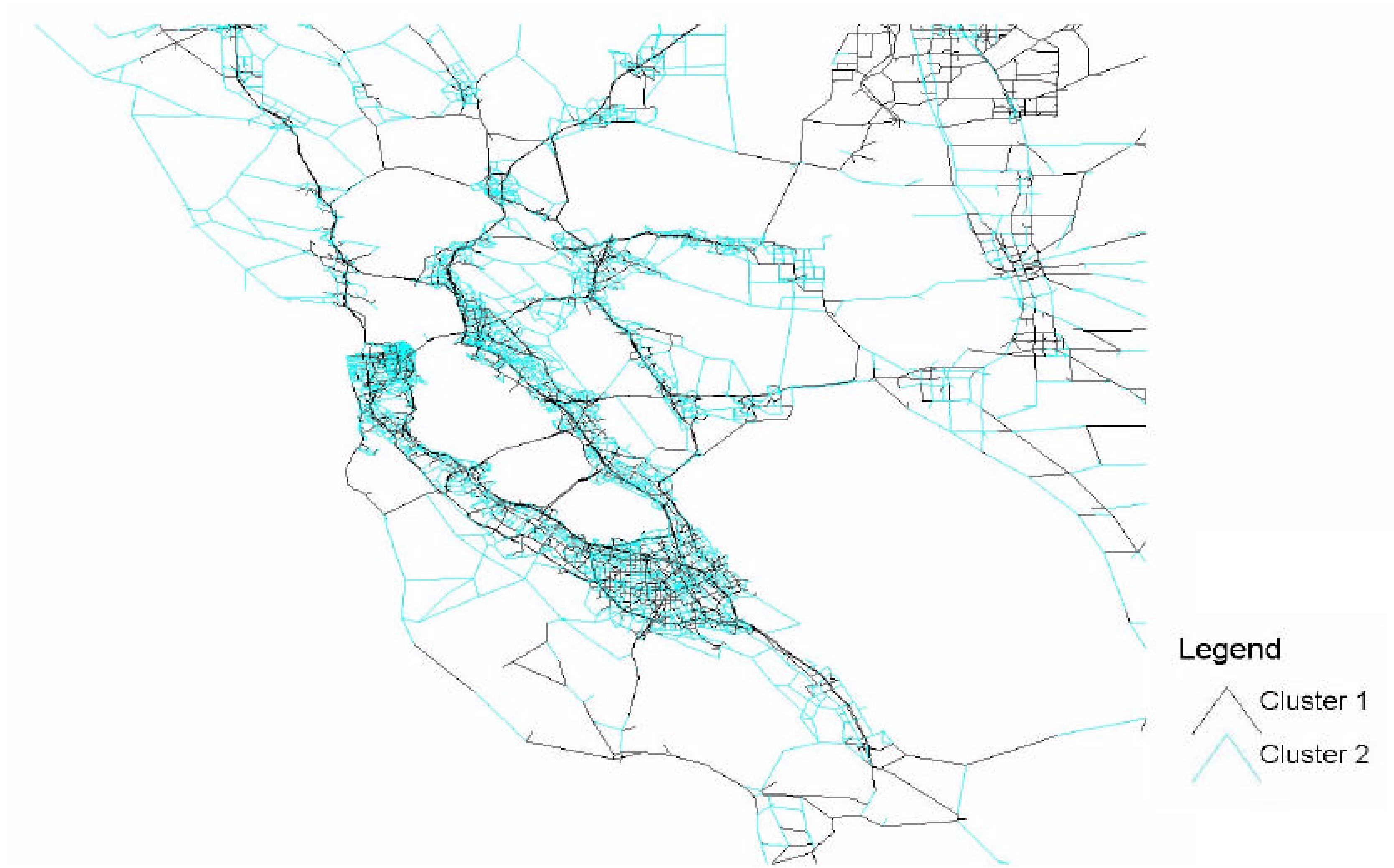


Figure 5.13(c) Clustering Results for the San Joaquin Valley Using Method 2 Based on Weekend Traffic

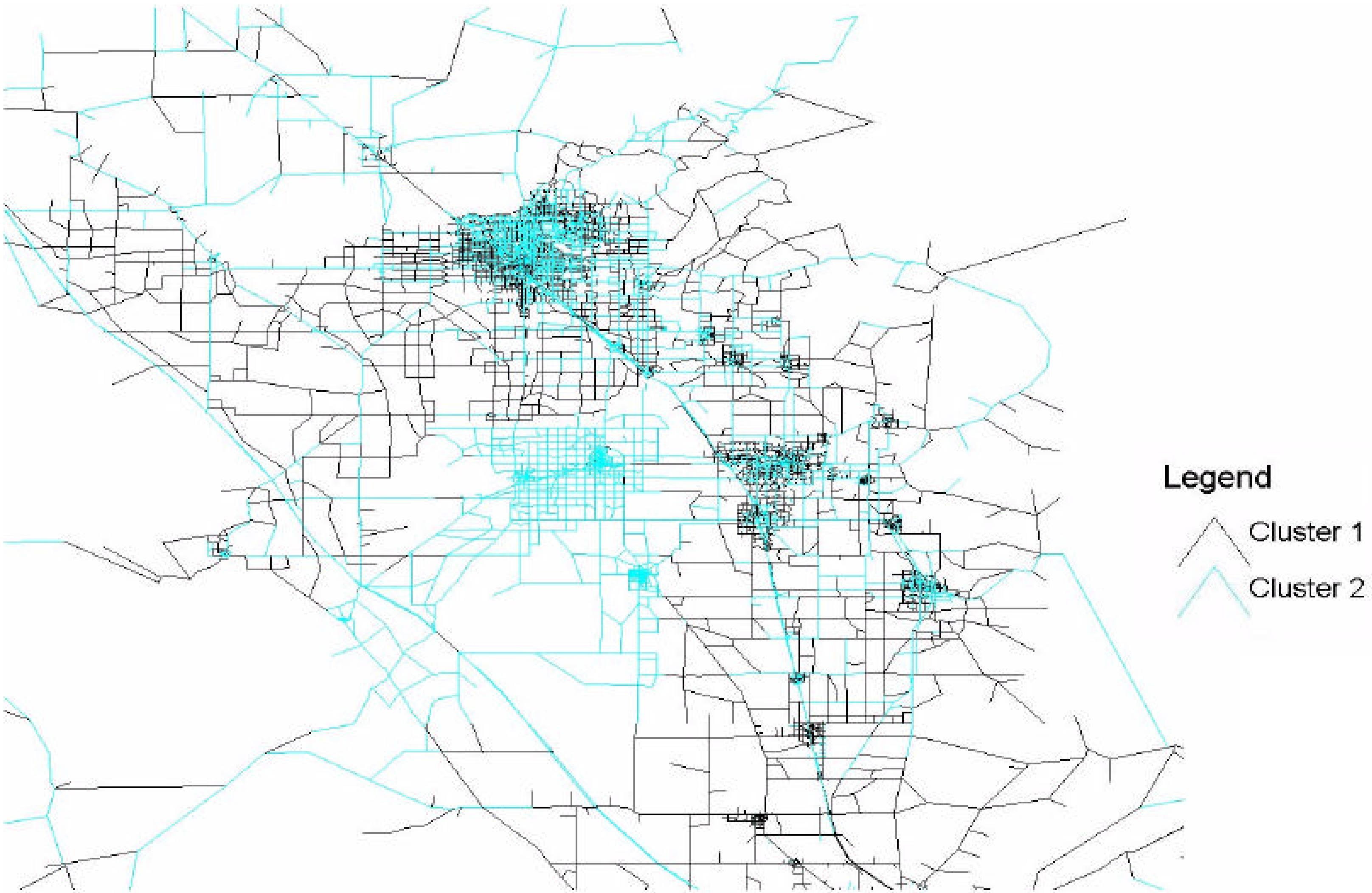


Table 5.11 Cluster Prediction Results for 135 Links with Matched Count Data

Correct cluster	**Cluster predicted using daily volumes**		**Cluster predicted using imputed volumes**		**Total***
Weekday (Monday-Friday)					
	1	2	1	2	
1	**23** **25%**	69 75%	**33** **35.9%**	59 64.1%	92 100%
2	8 18.6%	**35** **81.4%**	18 41.9%	**25** **58.1%**	43 100%
Weekend (Saturday-Sunday)					
1	**48** **37.2%**	81 62.8%	**51** **39.5%**	78 60.5%	129 100%
2	0 0%	**6** **100%**	1 16.7%	**5** **83.3%**	6 100%

* 6 counter locations cannot be located to match model links.

6.0 DTIM RUNS

Section 6 was not completed by study end date. Input data were not received from Alpine Geophysics, Inc and DTIM runs could not be completed without this information.

7.0 CONCLUSION

In travel demand modeling, roadway (or link) traffic volumes are typically produced in periods of time, which can be as long as the full 24 hours. A post-processor, such as the California Direct Travel Impact Model (DTIM), is frequently employed to disaggregate these period-based volumes from travel demand models to hourly volumes required by photochemical models. A common method is to use trip information from a travel diary and travel model volumes to produce a set of allocation factors. These allocation factors are used to disaggregate the total volumes in each modeling period into hourly estimates. The main problems in using travel diary trip information to produce allocation factors are twofold. First, the proportions of trips beginning and ending in any given period in a travel diary do not represent actual on-road traffic. The other problem with using the trip data is that the same allocation factors are applied across all modeling links in the transportation network without consideration of the spatial characteristics of the model links.

In this report, we applied a method to estimate allocation factors using observed traffic counts. The traffic counts were collected by the California Transportation Department (Caltrans) and by a sub-contractor from 141 locations in 4 Caltrans' traffic management districts. These four districts comprise the San Francisco Bay Area, the Sacramento Metropolitan Area, and the San Joaquin Valley. Plots of each district's 24 average hourly volumes suggest that districts covering more urbanized areas (e.g., the San Francisco Bay Area) do not only have higher traffic volumes, they also have different temporal traffic pattern compared to districts covering mainly rural area (e.g., the San Joaquin Valley). For example, Districts 3 and 4, where the San Francisco Bay Area and the Sacramento Metropolitan Area are encompassed, have 24-hour traffic profiles resembling a saddle with peaks of traffic both in the morning and in the afternoon. Districts 6 and 10, which cover a larger stretch of rural area, have inverted bell-shaped traffic profiles with a distinct peak of traffic in the afternoon.

Following our method, 135 of the 141 counter locations were successfully matched to links in the model network using local maps. A standard clustering method was applied to cluster the 135 successfully matched counter locations based on weekday and weekend traffic. This clustering method requires simplification of each location's 24-hourly volumes into linear functions (or principal components). Based on the linear functions, locations with similar spatial and temporal characteristics of traffic were clustered. Using the weekday traffic, the 135 locations were grouped into 2 clusters; 92 locations in one and 43 in another. There were also 2 clusters, 129 and 6 locations, respectively when weekend traffic was considered.

After locations had been clustered, allocation factors were estimated for each determined cluster. For weekdays, the 24-hourly allocation factors profiles show peaks in the morning and in the afternoon for one cluster (Cluster 1) and a distinct afternoon peak for another cluster (Cluster 2). For weekend, Cluster 1, which has 129 locations, has a 24-hourly allocation factors profile like an inverted bell with an afternoon peak. In contrast, the allocation profile of another cluster (Cluster 2) also resembles an inverted bell but is less smooth because of the relatively small sample size (6 locations). Since there were far fewer observed traffic counts than links in travel demand model (i.e., 66,237 links), links that are represented in the model network but do not have matching observed traffic counts need to be classified to predetermined clusters. These predetermined clusters' allocation factors were applied to these unmatched links so as to disaggregate model links' volumes into hourly volumes.

One of the problems with classifying the unmatched links using the current method is that 110 of 135 matched links had only daily volumes but not period-based volumes. As a result, 66,102 unmatched links were classified using daily volumes. Based on weekday traffic, 12,719 (Cluster 1) and 53,383 (Cluster 2) links were assigned to the two predetermined clusters while 19,266 and 46,836 links were assigned to clusters 1 and 2, respectively for weekend traffic. There is also a large disparity in the mean daily volumes of these clustered unmatched links. For example, the mean daily volume of the 12,719 links (i.e., 46,601 vehicles per day) is almost 9 times larger than that of 53,383 links (i.e., 5,351 vehicles per day) in case of weekday traffic. The disparities in the classification results of the unmatched links are probably because model links' daily volumes alone are generally not good indicators of the temporal traffic patterns.

To resolve this problem, we imputed the missing period-based volumes for each link with matching data using predetermined allocation factors and the matched links' daily volumes. All the unmatched links with period-based volumes were then classified based on the imputed period-based volumes instead of daily volumes. The classification results of unmatched links using matched links' imputed period-based volumes showed a more even number of links assigned to each cluster. The difference in each cluster's mean average volume also became smaller. For example, 35,525 and 30,577 links were assigned to clusters 1 and 2, respectively, which have respective mean daily volumes of 17,710 vehicles per day and 8,151 vehicles per day. These significant differences are likely due to facts that imputed period-based volumes are able to capture the temporal traffic variation.

To evaluate the reliability of classification using daily volume vis-à-vis using imputed period-based volumes, we predicted clusters for the 135 matched links using daily volumes as well as imputed period-based volumes. The prediction results using these two respective methods are not significantly different. Both methods predicted clusters correctly for 58 out of 135 matched links (43%) based on weekday traffic; method using daily volumes correctly predicted clusters for 54 links (40%) while method using imputed period-based volumes 56 links (41.5%). These results are mainly because only 31 matched links have period-based volumes, and whose clusters could be predicted using imputed period-based volumes. The remaining 110 links could only be predicted using daily volumes.

8.0 REFERENCES

J. Hicks and D.A. Niemeier (2001), “Improving the resolution of gridded-hourly mobile emissions: incorporating spatial variability and handling missing data,” *Transportation Research Part D*. No. 6 , pp. 153-177

9.0 APPENDIX A

A1 Calculation of Statistical Distance Using Daily Volumes

As discussed in the text, an unmatched link's the statistical distance (defined in Equation 5) from a predetermined cluster reduces to depend on two parameters: mean daily volume and pooled variance. For example, using information from Table 5.8, the statistical distance of an unmatched link 0 to clusters 1 and 2 on weekdays or on weekend were computed as follows:

$$D_1^{weekday}(x_0) = {(x_0 - 20113)^2}/{826227450}$$

$$D_2^{weekday}(x_0) = {(x_0 - 19005)^2}/{826227450}$$

$$D_1^{weeekend}(x_0) = {(x_0 - 20198)^2}/{830132477}$$

$$D_2^{weeekend}(x_0) = {(x_0 - 5664)^2}/{830132477}$$

Link 0 was assigned to the cluster with smallest value of $\hat{D}_i$ **.**

A2 Calculation of Statistical Distance Using Imputed Period-based Volumes

The statistical distance of Equation 5 was calculated in the following steps:

1. $\hat{x}_{ik}^{t_j}$, matched link *i*'s volume for t_j time period (i.e., AM, Mid-Day, PM, and Off-Peak) in cluster *k* was imputed. For example, matched link *i*'s AM-based volumes (6AM-9AM) was imputed as follows:

$$\hat{x}_{ik}^{am} = b_k^{am6} x_{ik}^{wd} + b_{ik}^{am7} x_{ik}^{wd} + b_k^{am8} x_{ik}^{wd} \qquad (6)$$

   where b_k^{am6} is the allocation factor at 6am for cluster *k*; x_{ik}^{wd} is the aggregate daily volume for matched link *i* in cluster *k*.

2. Vectors of mean period-based volumes by weekday and weekend for each predetermined cluster ***i*****,** $\hat{x}_i$ (Table A2.1) and pooled covariance matrix by weekday (Table A2.2) and weekend (Table A2.3), $\hat{S}$ were computed.

3. The new statistical distance, $\hat{D}_i$, was computed using $\hat{S}$ and $\hat{x}_i$ for each unmatched link to each predetermined cluster. For example, the statistical distance of an unmatched link 0 to clusters 1 and 2 on weekdays or on weekend were computed as follows:

$$\hat{D}_1^{weekday}(x_0) = \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 3201 \\ 6350 \\ 5332 \\ 5455 \end{pmatrix}\right]'$$

$$\times \begin{bmatrix} 1.37E8 & 3.06E8 & 2.67E8 & 2.60E8 \\ 3.06E8 & 8.8E8 & 7.59E8 & 6.93E8 \\ 2.67E8 & 7.59E8 & 6.7E8 & 6.23E8 \\ 2.60E8 & 6.93E8 & 6.23E8 & 6.05E8 \end{bmatrix}^{-1} \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 3201 \\ 6350 \\ 5332 \\ 5455 \end{pmatrix}\right],$$

$$\hat{D}_2^{weekday}(x_0) = \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 1663 \\ 6271 \\ 5768 \\ 6002 \end{pmatrix}\right]'$$

$$\times \begin{bmatrix} 1.37E8 & 3.06E8 & 2.67E8 & 2.60E8 \\ 3.06E8 & 8.8E8 & 7.59E8 & 6.93E8 \\ 2.67E8 & 7.59E8 & 6.7E8 & 6.23E8 \\ 2.60E8 & 6.93E8 & 6.23E8 & 6.05E8 \end{bmatrix}^{-1} \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 1663 \\ 6271 \\ 5768 \\ 6002 \end{pmatrix}\right],$$

$$\hat{D}_1^{weekend}(x_0) = \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 1355 \\ 6927 \\ 5713 \\ 6551 \end{pmatrix}\right]'$$

$$\times \begin{bmatrix} 4.02E7 & 1.43E8 & 1.28E8 & 1.52E8 \\ 1.43E8 & 1.01E9 & 8.12E8 & 8.59E8 \\ 1.28E8 & 8.12E8 & 6.62E8 & 7.10E8 \\ 1.52E8 & 8.59E8 & 7.10E8 & 7.8E8 \end{bmatrix}^{-1} \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 1355 \\ 6927 \\ 5713 \\ 6551 \end{pmatrix}\right],$$

$$\hat{D}_2^{weekend}(x_0) = \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 793 \\ 1859 \\ 1598 \\ 2873 \end{pmatrix} \right]'$$

$$\times \begin{bmatrix} 4.02E7 & 1.43E8 & 1.28E8 & 1.52E8 \\ 1.43E8 & 1.01E9 & 8.12E8 & 8.59E8 \\ 1.28E8 & 8.12E8 & 6.62E8 & 7.10E8 \\ 1.52E8 & 8.59E8 & 7.10E8 & 7.8E8 \end{bmatrix}^{-1} \left[\begin{pmatrix} x_0^{am} \\ x_0^{md} \\ x_0^{pm} \\ x_0^{op} \end{pmatrix} - \begin{pmatrix} 793 \\ 1859 \\ 1598 \\ 2873 \end{pmatrix} \right]$$

4. Link 0 was assigned to the closest cluster, that is, the smallest $\hat{D}_i$.

A3 Predicting Clusters for 135 Matched Links

The clusters of 135 matched links were predicted using aforementioned procedures in Sections A1 and A2. For the 110 links, which had only daily volumes but no period-based volumes, their clusters could only be predicted using daily volumes. The prediction results based on weekday traffic and on weekend traffic were presented in Tables A3.1 and A3.2, respectively.

As shown in Tables A3.1 and A3.2, there are six locations, which cannot be matched to any links in the model network and thus do not have period-based volumes nor daily volumes. All the imputed period-based volumes are highlighted and all the misclassified predicted clusters are marked with asterisks.

Table A2.1 4 Mean Period-based Volumes by Cluster for 135 Matched Links

	AM Peak (6am-9am)	Mid-Day (9am-3pm)	PM Peak (3pm-7pm)	Off-Peak (7pm-6am)
Monday-Friday				
Cluster 1	3,201	6,350	5,332	5,455
Cluster 2	1,603	6,271	5,768	6,002
Saturday-Sunday				
Cluster 1	1,355	6,927	5,713	6,551
Cluster 2	793	1,859	1,598	2,873

Table A2.2 Covariance Matrix (*S*) for 135 Matched Links for weekday traffic (Monday-Friday)

	AM Peak (6am-9am)	Mid-Day (9am-3pm)	PM Peak (3pm-7pm)	Off-Peak (7pm-6am)
AM Peak	13,758,852	30,623,544	26,721,215	26,041,453
Mid-Day	30,623,544	87,961,272	75,928,017	69,338,134
PM Peak	26,721,215	75,928,017	66,959,645	62,310,976
Off Peak	26,041,453	69,338,134	62,310,976	60,474,179

Table A2.3 Covariance Matrix (S) for 135 Matched Links for weekend traffic (Saturday-Sunday)

	AM Peak (6am-9am)	Mid-Day (9am-3pm)	PM Peak (3pm-7pm)	Off-Peak (7pm-6am)
AM Peak	4,024,464	14,326,710	12,787,823	15,243,249
Mid-Day	14,326,710	101,455,747	81,183,830	85,934,148
PM Peak	12,787,823	81,183,830	66,237,258	71,046,317
Off Peak	15,243,249	85,934,148	71,046,317	77,974,149

Table A3.1 Clustering Results for Matched Links Using Weekday Volumes
(Imputed volumes are shaded and misclassified clusters are marked with asterisks.)

ID	AM Volume	Mid-day Volume	PM Volume	Off-Peak Volume	Daily Volume	Cluster obtained using counter data	Cluster predicted using daily volume	Cluster predicted using imputed volumes
242	16571	8987	17879	31132	74569	1	1	1
254	689	314	774	1272	3049	1	2*	1
259	2277	1021	2758	5103	11159	1	2*	1
306	1645	3536	2851	2667	10700	1	2*	2*
307	2370	5093	4106	3841	15410	1	2*	2*
308	1864	1052	2060	3727	8703	1	2*	1
335	1393	337	1720	3210	6660	2	2	1*
358	1080	2320	1870	1750	7020	1	2*	2*
360	963	2070	1669	1561	6264	1	2*	2*
361	1714	3683	2970	2778	11145	1	2*	2*
364	1899	4081	3290	3078	12347	1	2*	2*
393	490	1053	849	794	3185	1	2*	2*
395	663	1425	1149	1075	4312	1	2*	2*
397	1719	3695	2979	2787	11179	1	2*	2*
399	1173	2520	2032	1901	7626	1	2*	2*
400						1		
409	74	36	65	146	321	1	2*	1
417	2023	4348	3506	3280	13157	1	2*	2*
422	792	311	905	1372	3380	1	2*	1
427	4891	2546	6428	9005	22870	1	1	1
443						1		
444						1		
492	1507	3239	2612	2443	9802	1	2*	2*

Table A3.1 (continued) Clustering Results for Matched Links Using Weekday Volumes

493	996	2142	1727	1615	6480	1	2*	2*
495	5800	12466	10050	9403	37719	1	1	1
496	562	1209	974	912	3657	1	2*	2*
500	10139	44636	41388	38166	134329	2	1*	1*
502	13040	28026	22595	21139	84800	1	1	1
513	2579	5542	4468	4180	16768	1	2*	2*
515	1022	4500	4173	3848	13543	2	2	2
517	4770	10253	8266	7733	31022	1	1	1
518	435	1915	1775	1637	5762	2	2	2
524						1		
531	29	127	117	108	381	2	2	2
533	2755	5921	4773	4466	17915	1	2*	2*
535	1937	8526	7906	7290	25658	2	1*	1*
558	14876	31971	25776	24115	96737	1	1	1
566	13351	28693	23133	21643	86820	1	1	1
571	5207	22925	21257	19602	68992	2	1*	1*
573	877	3859	3578	3300	11613	2	2	2
605	1833	8069	7482	6900	24284	2	1*	1*
607	13653	60107	55733	51394	180886	2	1*	1*
608	416	1834	1700	1568	5518	2	2	2
609	0	0	0	0	0	2	2	2
610	270	1189	1102	1017	3578	2	2	2
611	787	3466	3214	2964	10431	2	2	2
630	520	2289	2123	1958	6890	2	2	2
631	1802	3873	3123	2922	11720	1	2*	2*
632	2198	4724	3808	3563	14293	1	2*	2*
633	2200	4728	3812	3566	14305	1	2*	2*

Table A3.1 (continued) Clustering Results for Matched Links Using Weekday Volumes

634	735	3234	2999	2765	9732	2	2	2
635	1258	5538	5135	4735	16667	2	2	2
637	3713	16347	15158	13977	49195	2	1*	1*
649	0	0	0	0	0	2	2	2
651	549	2417	2242	2067	7275	2	2	2
652	6200	13324	10742	10050	40317	1	1	1
654	9229	19834	15990	14960	60013	1	1	1
669	2020	4340	3499	3274	13133	1	2*	2*
670	3794	8153	6573	6150	24669	1	1	1
671	85	183	148	138	555	1	2*	2*
675	7106	15272	12312	11519	46209	1	1	1
679	3970	8532	6878	6435	25815	1	1	1
681	3391	7287	5875	5496	22049	1	1	1
682	2521	5417	4367	4086	16391	1	2*	2*
683	6171	13263	10693	10004	40131	1	1	1
685	21842	46943	37846	35407	142038	1	1	1
686	13554	29129	23484	21971	88138	1	1	1
696	1400	6164	5715	5270	18549	2	2	2
701	4401	19375	17965	16566	58306	2	1*	1*
704	9238	19854	16007	14975	60074	1	1	1
733	87	381	353	326	1147	2	2	2
734	3331	7160	5772	5400	21664	1	1	1
735	1006	2163	1744	1631	6544	1	2*	2*
736	4446	9556	7704	7208	28914	1	1	1
994	944	1883	1231	3491	5667	2	2	2
997	782	1270	1002	2057	3842	1	2*	1
1006	1517	3259	2628	2458	9862	1	2*	2*

Table A3.1 (continued) Clustering Results for Matched Links Using Weekday Volumes

1007	620	1332	1074	1005	4030	1	2*	2*
1026	675	1487	997	2802	4474	2	2	2
1030	1582	3524	3000	6081	10662	1	2*	2*
1031	693	1290	1094	2096	3882	2	2	1*
1032	742	1359	1135	2215	4090	2	2	1*
1072	389	721	599	1183	2170	2	2	1*
1098	144	272	231	444	819	2	2	1*
1110	782	1681	1355	1268	5086	1	2*	2*
1115	2385	4646	3709	7886	13981	2	2	2
1133	791	1214	995	1888	3673	1	2*	1
1134	791	1221	995	1888	3673	2	2	1*
1141	671	1026	842	1575	3088	2	2	1*
1142	1990	3665	3109	5992	11091	1	2*	2*
1153	422	799	679	1302	2404	2	2	1*
1154	937	1488	1220	2321	4478	2	2	1*
1158	1220	2621	2113	1977	7931	1	2*	2*
1159	811	1568	1095	2835	4743	1	2*	2*
1164						1		
1611						1		
1613	1021	2194	1769	1655	6638	1	2*	2*
1614	3230	6941	5596	5236	21003	1	1	1
1646	145	312	251	235	943	1	2*	2*
1647	1097	2358	1901	1779	7135	1	2*	2*
1656	1031	4541	4211	3883	13666	2	2	2
1661	2327	10247	9501	8761	30836	2	1*	1*
1662	219	965	895	825	2905	2	2	2
1663	1019	2190	1766	1652	6627	1	2*	2*

Table A3.1 (continued) Clustering Results for Matched Links Using Weekday Volumes

1664	242	520	419	392	1574	1	2*	2*
1674	2201	4730	3813	3567	14311	1	2*	2*
1676	1903	4090	3297	3085	12374	1	2*	2*
1678	117	251	202	189	759	1	2*	2*
1679	2443	5250	4233	3960	15885	1	2*	2*
1694	1364	2931	2363	2211	8870	1	2*	2*
1698	1972	4238	3417	3197	12823	1	2*	2*
1711	470	2434	1863	946	5714	1	2*	2*
1731	1864	4006	3230	3022	12121	1	2*	2*
1748	796	1710	1379	1290	5175	1	2*	2*
1756	1765	3793	3058	2861	11477	1	2*	2*
1773	1793	3854	3107	2907	11662	1	2*	2*
1777	1270	2729	2200	2059	8258	1	2*	2*
1814	425	914	737	690	2766	1	2*	2*
1816	1678	3606	2907	2720	10912	1	2*	2*
1818	238	512	413	386	1550	1	2*	2*
1838	119	256	206	193	774	1	2*	2*
3001	1682	3211	2716	5265	9663	2	2	2
3002	1001	1622	1354	2554	4909	1	2*	1
3003	1386	2575	2265	4139	7791	1	2*	2*
3004	2120	4072	3117	7015	12253	2	2	2
3005	58	299	229	116	703	1	2*	2*
3006	925	1988	1603	1499	6015	1	2*	2*
3007	1602	3578	3025	6141	10767	2	2	2
3008	2738	4876	3336	8680	14755	1	2*	2*
3009	778	1673	1349	1262	5062	1	2*	2*
3010	1574	3045	2898	4691	9164	2	2	1*

Table A3.1 (continued) Clustering Results for Matched Links Using Weekday Volumes

4001	6703	14406	11614	10866	43589	1	1	1
4002	6979	14998	12092	11313	45382	1	1	1
4003	2490	5352	4315	4037	16193	1	2*	2*
4004	943	4153	3851	3551	12498	2	2	2
4005	13648	29331	23647	22124	88750	1	1	1
4006	1662	3572	2880	2694	10807	1	2*	2*
4007	552	2432	2255	2080	7320	2	2	2
4008	2209	4748	3828	3581	14366	1	2*	2*
4009	1999	4296	3463	3240	12998	1	2*	2*
4010	1521	3268	2635	2465	9888	1	2*	2*
Total no. of links with correctly predicted clusters (out of 135 (%))							**58(43%)**	**58(43%)**

Table A3.2 Clustering Results for Matched Links Using Weekend Volumes (Imputed volumes are shaded and misclassified clusters are marked with asterisks.)

ID	AM Volume	Mid-day Volume	PM Volume	Off-Peak Volume	Daily Volume	Cluster obtained using counter data	Cluster predicted using daily volume	Cluster predicted using imputed volumes
242	16571	8987	17879	31132	74569	1	1	1
254	689	314	774	1272	3049	1	2*	1
259	2277	1021	2758	5103	11159	1	2*	2*
306	562	3803	3057	3279	10700	1	2*	2*
307	809	5477	4402	4722	15410	1	1	1
308	1864	1052	2060	3727	8703	1	2*	1
335	1393	337	1720	3210	6660	1	2*	2*
358	369	2495	2005	2151	7020	1	2*	2*
360	329	2226	1789	1919	6264	1	2*	2*
361	585	3961	3184	3415	11145	1	2*	2*
364	648	4388	3527	3783	12347	1	2*	2*
393	167	1132	910	976	3185	1	2*	2*
395	226	1533	1232	1321	4312	1	2*	2*
397	587	3973	3193	3425	11179	1	2*	2*
399	400	2710	2178	2337	7626	1	2*	2*
400						1		
409	74	36	65	146	321	1	2*	1
417	691	4676	3758	4031	13157	1	1	1
422	792	311	905	1372	3380	1	2*	1
427	4891	2546	6428	9005	22870	1	1	1
443						1		
444						1		
492	515	3484	2800	3003	9802	1	2*	2*

Table A3.2 (continued) Clustering Results for Matched Links Using Weekend Volumes

493	340	2303	1851	1986	6480	1	2*	2*
495	1980	13406	10775	11557	37719	1	1	1
496	192	1300	1045	1121	3657	1	2*	2*
500	7052	47744	38373	41160	134329	1	1	1
502	4452	30140	24224	25984	84800	1	1	1
513	880	5960	4790	5138	16768	1	1	1
515	711	4814	3869	4150	13543	1	1	1
517	1629	11026	8862	9505	31022	1	1	1
518	302	2048	1646	1766	5762	1	2*	2*
524						1		
531	20	135	109	117	381	1	2*	2*
533	941	6367	5118	5489	17915	1	1	1
535	1347	9120	7330	7862	25658	1	1	1
558	5079	34383	27634	29641	96737	1	1	1
566	4558	30858	24801	26603	86820	1	1	1
571	3622	24522	19709	21140	68992	1	1	1
573	610	4128	3317	3558	11613	1	2*	2*
605	1275	8631	6937	7441	24284	1	1	1
607	9496	64291	51673	55425	180886	1	1	1
608	290	1961	1576	1691	5518	1	2*	2*
609	0	0	0	0	0	1	2*	2*
610	188	1272	1022	1096	3578	1	2*	2*
611	548	3707	2980	3196	10431	1	2*	2*
630	362	2449	1968	2111	6890	1	2*	2*
631	615	4166	3348	3591	11720	1	2*	2*
632	750	5080	4083	4380	14293	1	1	1
633	751	5084	4086	4383	14305	1	1	1

Table A3.2 (continued) Clustering Results for Matched Links Using Weekend Volumes

634	511	3459	2780	2982	9732	1	2*	2*
635	875	5924	4761	5107	16667	1	1	1
637	2583	17485	14053	15074	49195	1	1	1
649	0	0	0	0	0	2	2	2
651	250	2388	2349	2288	7275	2	2	2
652	2117	14330	11517	12354	40317	1	1	1
654	3151	21330	17144	18389	60013	1	1	1
669	689	4668	3752	4024	13133	1	1	1
670	1295	8768	7047	7559	24669	1	1	1
671	29	197	159	170	555	1	2*	2*
675	2426	16424	13200	14159	46209	1	1	1
679	1355	9175	7374	7910	25815	1	1	1
681	1158	7837	6299	6756	22049	1	1	1
682	861	5826	4682	5022	16391	1	1	1
683	2107	14264	11464	12297	40131	1	1	1
685	7457	50484	40575	43522	142038	1	1	1
686	4627	31326	25178	27006	88138	1	1	1
696	974	6593	5299	5684	18549	1	1	1
701	3061	20723	16656	17866	58306	1	1	1
704	3154	21352	17161	18407	60074	1	1	1
733	60	408	328	351	1147	1	2*	2*
734	1137	7700	6189	6638	21664	1	1	1
735	344	2326	1869	2005	6544	1	2*	2*
736	1518	10277	8260	8860	28914	1	1	1
994	944	2014	1231	3491	5667	1	2*	2
997	782	1366	1002	2057	3842	1	2*	2
1006	518	3505	2817	3022	9862	1	2*	2

Table A3.2 (continued) Clustering Results for Matched Links Using Weekend Volumes

1007	212	1432	1151	1235	4030	1	2*	2
1026	675	1468	997	2802	4474	2	2	2
1030	1582	3790	3000	6081	10662	1	2*	2*
1031	693	1380	1094	2096	3882	1	2*	2*
1032	742	1454	1135	2215	4090	1	2*	2*
1072	389	771	599	1183	2170	1	2*	2*
1098	144	269	231	444	819	2	2	1*
1110	267	1808	1453	1558	5086	1	2*	2*
1115	2385	4969	3709	7886	13981	1	1	2*
1133	791	1305	995	1888	3673	1	2*	2*
1134	791	1305	995	1888	3673	1	2*	2*
1141	671	1098	842	1575	3088	1	2*	1
1142	1990	3942	3109	5992	11091	1	2*	2*
1153	422	854	679	1302	2404	1	2*	2*
1154	937	1592	1220	2321	4478	1	2*	2*
1158	416	2819	2266	2430	7931	1	2*	2*
1159	811	1686	1095	2835	4743	1	2*	2*
1164								
1611								
1613	348	2359	1896	2034	6638	1	2*	2*
1614	359	4045	8146	8453	21003			
1646	50	335	269	289	943	1	2*	2*
1647	375	2536	2038	2186	7135	1	2*	2*
1656	233	2632	5300	5500	13666			
1661	1619	10960	8809	9448	30836	1	1	1
1662	153	1033	830	890	2905	1	2*	2*
1663	348	2355	1893	2031	6627	1	2*	2*

Table A3.2 (continued) Clustering Results for Matched Links Using Weekend Volumes

1664	83	559	450	482	1574	1	2*	2*
1674	751	5086	4088	4385	14311	1	1	1
1676	650	4398	3535	3792	12374	1	2*	2*
1678	40	270	217	233	759	1	2*	2*
1679	834	5646	4538	4867	15885	1	1	1
1694	466	3153	2534	2718	8870	1	2*	2*
1698	673	4558	3663	3929	12823	1	2*	2*
1711	300	2031	1632	1751	5714	1	2*	2*
1731	636	4308	3463	3714	12121	1	2*	2*
1748	272	1839	1478	1586	5175	1	2*	2*
1756	603	4079	3279	3517	11477	1	2*	2*
1773	612	4145	3331	3573	11662	1	2*	2*
1777	434	2935	2359	2530	8258	1	2*	2*
1814	145	983	790	848	2766	1	2*	2*
1816	573	3878	3117	3344	10912	1	2*	2*
1818	81	551	443	475	1550	1	2*	2*
1838	41	275	221	237	774	1	2*	2*
3001	1682	3434	2716	5265	9663	1	2*	2*
3002	1001	1745	1354	2554	4909	1	2*	2*
3003	1386	2769	2265	4139	7791	1	2*	2*
3004	2120	4022	3117	7015	12253	2	2	2
3005	37	250	201	215	703	1	2*	2*
3006	316	2138	1718	1843	6015	1	2*	2*
3007	1602	3827	3025	6141	10767	1	2*	2*
3008	2738	5244	3336	8680	14755	1	1	2*
3009	266	1799	1446	1551	5062	1	2*	2*
3010	1574	3008	2898	4691	9164	2	2	2

Table A3.2 (continued) Clustering Results for Matched Links Using Weekend Volumes

4001	2288	15493	12452	13356	43589	1	1	1
4002	2383	16130	12964	13906	45382	1	1	1
4003	850	5755	4626	4962	16193	1	1	1
4004	656	4442	3570	3830	12498	1	2*	2*
4005	4659	31544	25353	27194	88750	1	1	1
4006	567	3841	3087	3311	10807	1	2*	2*
4007	384	2602	2091	2243	7320	1	2*	2*
4008	754	5106	4104	4402	14366	1	1	1
4009	682	4620	3713	3983	12998	1	1	1
4010	519	3514	2825	3030	9888	1	2*	2*
Total no. of links with correctly predicted cluster (out of 135 (%))							**54(40%)**	**56(41.5%)**