Title
Factors Affecting Vehicle Occupancy Measurement

Authors
Levine, Ned
Wachs, Martin

Publication Date
1996-12-01

Factors Affecting Vehicle Occupancy Measurement

Ned Levine

Ned Levine & Associates
4400 East West Highway, #333
Bethesda, MD 20814

Martin Wachs

University of California Transportation Center
University of California, Berkeley
Berkeley, CA 94720-1720

Working Paper
December 1996

UCTC No. 350

The University of California Transportation Center
University of California at Berkeley

The University of California Transportation Center

The University of California Transportation Center (UCTC) is one of ten regional units mandated by Congress and established in Fall 1988 to support research, education, and training in surface transportation. The UC Center serves federal Region IX and is supported by matching grants from the U.S. Department of Transportation, the California Department of Transportation (Caltrans), and the University.

Based on the Berkeley Campus, UCTC draws upon existing capabilities and resources of the Institutes of Transportation Studies at Berkeley, Davis, Irvine, and Los Angeles; the Institute of Urban and Regional Development at Berkeley; and several academic departments at the Berkeley, Davis, Irvine, and Los Angeles campuses. Faculty and students on other University of California campuses may participate in Center activities. Researchers at other universities within the region also have opportunities to collaborate with UC faculty on selected studies.

UCTC's educational and research programs are focused on strategic planning for improving metropolitan accessibility, with emphasis on the special conditions in Region IX. Particular attention is directed to strategies for using transportation as an instrument of economic development, while also accommodating to the region's persistent expansion and while maintaining and enhancing the quality of life there.

The Center distributes reports on its research in working papers, monographs, and in reprints of published articles. It also publishes *Access*, a magazine presenting summaries of selected studies. For a list of publications in print, write to the address below.

University of California
Transportation Center

108 Naval Architecture Building
Berkeley, California 94720
Tel: 510/643-7378
FAX: 510/643-5456

FACTORS AFFECTING VEHICLE OCCUPANCY MEASUREMENT

ABSTRACT

Factors affecting vehicle occupancy measurement were examined with the aim of improving state vehicle occupancy monitoring programs. A comparison was conducted of five data sets looking at the effects on average vehicle occupancy (AVO) of time of day, day of week, road types, HOV lanes, locational differences, and traffic volume. It was found that AVO was higher in the afternoons, on Saturdays, and on HOV lanes. Inconsistent differences were found for the other variables, though there were considerable locational variations. Based on these factors, suggestions are made for drawing samples to represent regional and corridor-level vehicle occupancy levels.

FACTORS AFFECTING VEHICLE OCCUPANCY MEASUREMENT

INTRODUCTION

Vehicle occupancy measurement is an increasingly important part of transportation congestion management and can be used for evaluating the efficiency of the road system, HOV lanes, or particular congestion reduction programs. Employer trip reduction ordinances, subsidies to ridesharing programs, and parking cashout legislation are examples of many legislative and programmatic attempts to increase vehicle occupancies (Higgins, 1995). In order to evaluate the effectiveness of those policies, there is greater interest than there has ever been in measuring vehicle occupancy in a statistically responsible way (Turnbull, Henk, and Christiansen, 1991).

This article will identify factors that affect the measurement of vehicle occupancy in order to develop a sampling frame for adequately measuring vehicle occupancy. Our aim is to advance the methodology for vehicle occupancy measurement so that state and local transportation agencies can integrate this measurement into larger transportation policies for reducing congestion and improving the efficiency of road systems.

State Vehicle Occupancy Monitoring Practices

To date, the literature on vehicle occupancy monitoring includes four general reviews of practice and theory. Ulberg and McCormack (1988) reviewed factors affecting the accuracy of vehicle occupancy counts, such as weather, time-of-day, day-of-week, and season. Rutherford, Kinchen and Jacobsen (1990) reviewed state practices in monitoring high occupancy vehicle (HOV) lane violations, while Turnbull, Stokes and Henk (1991) have reviewed current practices in the evaluation of HOV facilities. Finally, Levine and Wachs (1994) updated these reviews.

In spite of considerable diversity in administrative context and geographical settings, there appears to be a common method which is used throughout the country to count the number of occupants in vehicles. Typically, counts are conducted by observers who record vehicle occupancy on a manual count board. With a couple of exceptions, observers generally count for short periods, varying from five to thirty minutes. Measurements are usually taken during the morning peak period, though some states also conduct evening peak period counts. Observations are typically taken in the middle of the week (Tuesdays, Wednesdays, or Thursdays). Separate counts are usually taken of the number of vehicles and the number of occupants. The most common measure used, average vehicle occupancy (AVO), is calculated by dividing the total number of occupants by the total number of vehicles.

Monitoring programs vary considerably in the locations chosen. In Texas, California, and Washington, D.C., special attention is paid to the central business district. In Oregon, and also to a certain extent in Texas, sites with higher traffic volume are chosen. In a recently established program in Washington, emphasis was on creating a statistically valid random sample. All programs are somewhat constrained locationally by safety and visibility factors.

The literature also suggests that little work has been done to statistically analyze the accuracy of average vehicle occupancy estimates. There have been only four federally funded studies applying a sound sampling methodology to measure vehicle occupancy (Ferlis, 1979; Ferlis, 1981; Fisher, Williams and Boyd, 1980; Hupp and Palombo, 1980).

Changing Transportation Patterns

Many of these practices emerged during the 1970s in response to the emphasis on reducing home-to-work congestion that occurred after the 1973 oil crisis. The morning period was chosen primarily because a high proportion of morning trips are work related, and freeways were chosen because of their heavy usage for work trips. Over time, however, travel patterns have become more diffuse and irregular. Home-to-work trips have declined as a proportion of all trips (Richardson and Gordon, 1989); nationally, commute trips constitute less than thirty percent of trips. Similarly, the spatial redistribution of employment centers towards the suburbs has altered the highly concentrated nature of home-to-work trips (Cervero, 1996; Pisarksi, 1996; Levine and Glickfeld, 1996). In addition, increasing numbers of workers have altered their travel patterns to avoid peak hour congestion through such means as variable working hours, staggered work days, telecommuting, or seeking work further on the periphery of the metropolitan area (see Wachs, Taylor, Levine, and Ong, 1993).

The notion of a highly congested travel network due to geographic centralization and concentrated hours is becoming less realistic every year as congestion in major metropolitan areas spreads and its duration increases. This is not to suggest that traffic is lessening in the downtown areas during peak periods. Rather, the peak periods are lengthening and the differences in congestion between peak periods and non-peak periods are decreasing over time. Also, in some metropolitan areas, suburban congestion is growing in importance in comparison with inner-city congestion.

There is a need, therefore, to reshape the sampling methodology by which vehicle occupancy is monitored. In particular, the sampling frame needs to be reformed in order to better estimate vehicle occupancy for an entire region or for particular corridors.

Factors Affecting Average Vehicle Occupancy

To develop a methodology for improving vehicle occupancy measurement, a study was constructed to examine factors which affect AVO. Intuitively, a number of variables could be related:

1. Time-of-day
2. Day-of-week
3. Road Types
4. Locational Variations
5. Traffic Volume

Several studies have looked at these. A study in Detroit by the Southeast Michigan Council of Governments showed that vehicle occupancy differed by geographic area, highway class, and time-of-day (Hupp and Palombo, 1980). Different routes yielded different AVO. Vehicle occupancy also varied by time of day; the morning peak period had lower average vehicle occupancy than the mid-day period and the evening peak period.

Another study in Atlanta by the Georgia Department of Transportation collected vehicle occupancy data for eight stratified geographic areas and highway classes for a twelve-hour time period and for both directions at each location (Fisher, Williams and Boyd, 1980). The results showed that AVO varied between different locations and typically increased throughout the day. A study in the Minneapolis area surveyed vehicle occupancy inbound into the Minneapolis and St. Paul central business districts and concluded that there were significant variations in average vehicle occupancy for different days of the week, but that these variations depended on location (Benke and Sjoberg, 1977). It also found that there were significant differences in vehicle occupancy for different areas of a corridor.

In the Seattle area, vehicle occupancy data were collected at eighteen regional sites over a fifteen-month period, but no predictable trends were found in vehicle occupancy by type of facility, season, day of week, time of day, traffic volume, level of transit service, or distance to CBD (Lester, Dare and Roach, 1979). A research project in Phoenix examined factors that have the greatest influence on vehicle occupancy (Arizona Department of Transportation, 1989). Trip purpose was found to be an important factor. Other determinants, such as household income and trip distance, were uncorrelated with AVO. The report showed that vehicle occupancy varied greatly by time of day and that the lowest AVO was associated with home-based work trips.

In summary, there is little consistency in the conclusions of studies on the importance of spatial and temporal differences in vehicle occupancy. Several studies have shown significant differences between routes and for different time periods. But at least one study did not find such differences.

Bias in Vehicle Occupancy Counting

It should also be mentioned that vehicle occupancy counts are prone to underestimation in that passengers who are not visible to the observers are not counted (e.g., small children in the back seat). Unfortunately, there is not much one can do about this other than to train observers to be attentive to this problem.

The use of video systems may increase accuracy, but even here there are problems. In an experiment conducted by the California Highway Patrol and the California Department of Transportation (Caltrans), photographic interpreters disagreed with visual observers on approximately 20 to 25 percent of photographed vehicles (McFadden and Innes, 1986). Another experiment using videotaping for enforcement purposes demonstrated that human observers cannot accurately determine occupancy unless they make roadside contact with automobiles (Billheimer, 1990). In this experiment, highway patrol officers pursued suspected carpool lane violators when cued by persons viewing the road directly and by persons viewing the road through video cameras. The study found that both the direct observers and the video observers overestimated the number of vehicles violating occupancy requirements. Video interpretation and human observers were further compared. The counts made through video observation were consistently more accurate than manual counts by experienced Caltrans observers, but the cost was about double. An occupancy study conducted by the New Jersey Department of Transportation provided a similar assessment of about two hours of viewing time per hour of video footage (Kuziw, 1993).

The use of technology to improve observations may help in increasing accuracy. But until the procedures are well established and the cost reduced, the use of human observers will probably remain the major means for vehicle occupancy counts.

METHODOLOGY

Empirical Analysis of Five Data Sets

There have been no studies which have adequately sampled all the above mentioned factors affecting vehicle occupancy. Therefore, in order to examine factors that affect AVO in California, five different data sets were compiled, each of which optimized particular variables affecting AVO. The data sets and their names are:

1. Los Angeles-Ventura 1988-1992. Caltrans District 7 conducted observations at fifty freeway locations in Los Angeles and Ventura counties between 1988 and 1992 (Kluza, 1991). The data set had 706 measurements for the fifty locations including several HOV lanes. Counts were taken in one direction only, for traffic headed into downtown Los Angeles. About eighty percent of the observations were taken during the morning rush hour (6:30 to 8:30 a.m.) with the remaining twenty percent taken during the late afternoon hours (3:00 to 5:00 p.m.). The observations were taken on Tuesdays, Wednesdays and Thursdays. Separate counts were made of motorcycles and buses along with estimates of how full they were. However, these have been combined with passenger vehicle counts into a single AVO estimate.

2. SANDAG 1985; SANDAG 1990. The San Diego Association of Governments (SANDAG) conducted observations at 100 locations in San Diego County during the springs of 1985 and 1990, based on an initial sample from 1981 (San Diego Association of Governments, 1982; 1985; 1990). Counts were taken in one direction during the morning rush hours (6:30 to 8:30 a.m.). There was no information on the day of the week that the count was taken. The 100 locations represented six different road types. AVO was calculated only for automobiles, vans, pickup trucks and light-duty trucks. Separate counts were taken of buses and heavy-duty trucks.

3. Los Angeles 1994. We conducted observations at twelve locations along the Santa Monica Freeway corridor in April and May 1994 (Levine and Wachs, 1994). The twelve locations were sampled from the California Highway Performance Monitoring Systems (HPMS) that fell within one mile of both sides of the Santa Monica Freeway (California Department of Transportation, 1990; 1993). The twelve locations represented five different types of roads. Observations were paired, starting with a random start, and were then sequenced (e.g., Wednesday and Friday, Thursday and Saturday, Friday and Monday). Thus, each location was measured both earlier and later in the week. For each observation day at a location, observations were conducted at three different time periods: 6:30 to 8:30 a.m., 11:00 a.m. to 1:00 p.m., and 4:00 to 6:00 p.m. Several of the observation periods had to be dropped for technical reasons, so that there were 59 complete observation periods in the sample. For a measure of AVO, all occupants and all vehicles have been counted, including motorcycles, trucks and buses. Because of the January 17, 1994 earthquake, the Santa Monica freeway was being reconstructed during this period; most probably, AVO was higher during this period than at normal times.

4. Sacramento 1994. We also conducted observations at six locations in the Sacramento area in May 1994 (Levine and Wachs, 1994). The particular corridor chosen was that paralleling Interstate 99 south of downtown within one mile of the segment in each direction. Two locations along Highway 50 and Interstate 80 were also included in the Sacramento survey. One observation location on State Highway 99 had an HOV lane. Again, observations were conducted in paired comparisons on two separate days, one early in the week and one later in the week. - Mondays through Saturdays, at three different time periods: 6:30 to 8:30 a.m., 11:00 a.m. to 1:00 p.m., and 4:00 to 6:00 p.m. There were, therefore, 36 complete observations for the six locations.

Each data set defined occupants and vehicles slightly differently. The Los Angeles-Ventura 1988-92 data base did not include buses and motorcycles, but did include trucks. The SANDAG data set did not include buses, motorcycle and heavy trucks, but did include light trucks. The 1994 Sacramento and Los Angeles data sets included all vehicles, though buses were measured separately. However, each data set was analyzed by itself so that these differences should not appreciably affect the results. We analyzed the results by both unweighted and weighted (by traffic volume) estimates of AVO. Because the results were similar, we are only presenting the unweighted estimates.

RESULTS

Data Set Differences in AVO

There are differences in AVO among the five data sets. Table 1 presents the average AVO across all measurement locations, along with the standard deviation and the range of the different locational AVOs.

The AVOs vary from 1.204 to 1.450. There is also considerable variation within each data set. The standard deviations vary from 0.068 for the 1988-92 Los Angeles-Ventura data set (observations that were taken only on freeways) to 0.16 for the SANDAG 1985 and Los Angeles 1994 data sets.

Time of Day

The Los Angeles-Ventura, Sacramento 1994 and Los Angeles 1994 data bases examined changes in AVO by time of day. All three data sets took the morning period as 6:30 to 8:30 a.m. The Sacramento and Los Angeles data sets took the noon period as 11 a.m. to 1 p.m. There was a slight difference in the afternoon time periods; Los Angeles- Ventura measured this as 3:00 to 5:00 p.m. whereas the Sacramento 1994 and Los Angeles 1994 data sets took this as 4:00 to 6:00 p.m.

Table 2 presents the results. There is a consistent increase in AVO from the morning rush hours till the afternoon rush hours. The difference between the morning and afternoon periods is significant for the Los Angeles-Ventura data. The incremental changes for the Sacramento and Los Angeles data sets are not significant in that the change from morning to noon and from noon to afternoon are gradual. However, the difference between the morning and afternoon periods is highly significant for the Sacramento and Los Angeles data sets, as with the Los Angeles-Ventura data set ($F=4.92$, $p<=.05$ for Sacramento and $F=5.66$, $p<=.05$ for Los Angeles).

In other words, AVO is higher in the afternoon than in the morning, with the AVO for the noon period falling in between. Further, the mean AVO for the afternoon and morning periods in both data sets are not significantly different from the mean AVO for all three periods. This suggests that by measuring only the morning and afternoon periods of the day, a reasonably close estimate for the entire day can be obtained by combining these estimates (weighting each by traffic volume).

Day of Week

The Los Angeles-Ventura, Sacramento 1994 and Los Angeles 1994 data bases include information on day of week (Table 3). In the Los Angeles-Ventura data base, there is not a significant difference in AVO between Tuesdays, Wednesdays and Thursdays. Similarly, the 1994 Sacramento and Los Angeles data bases do not show a significant change between Mondays and Fridays. The State of California and Caltrans had designated Thursdays as 'Ride Sharing Day' during the time of this study. For the Los Angeles-Ventura and Los Angeles data bases, there is no evidence that Thursdays had significantly higher AVO. However, for Sacramento, Thursdays had the second highest AVO; it is possible that with the large number of government workers in Sacramento, the government commitment to Thursday ridesharing had an effect.

However, on Saturdays there is a considerably higher AVO in the two data sets that measure it; this difference is statistically significant ($F=11.80$, $p<=.01$ for

Sacramento and F=5.70, p<.05 for Los Angeles). Saturdays should be treated differently from weekdays in taking vehicle occupancy counts.

Road Types

There are considerable variations in AVO by type of road. Each of the data sets defined road types in unique ways. The two SANDAG data sets distinguished six classes: freeways, expressways, six-lane arterials, four-lane arterials, collector roads, and local roads. The HPMS systems, from which the 1994 Sacramento and Los Angeles samples were drawn, distinguished five classes: interstate highways, other freeway segments, principal arterials, minor arterials, and collector roads. To compare these, each of these road types have been grouped into one of four categories:

1. Freeways/expressways
2. Major arterials
3. Minor arterials
4. Local roads.

Table 4 presents the analysis of AVO by road types. There are some consistencies and a lot of inconsistencies. With the two SANDAG data sets, AVO consistently decreases as the level of road classification increases; freeways have the lowest AVO and local roads the highest. In Sacramento, however, there is not a significant difference between the two road types examined. In Los Angeles, the Santa Monica freeway had a lower AVO than arterials, but the major arterials had a higher AVO than the minor arterials. The Los Angeles case, however, was measured at a time shortly after the January 1994 earthquake and it is possible that travel patterns were substantially different during this period than at other times.

It is possible that there are some local differences in the use of road segments. In San Diego, work trips tend to be focused on freeways and major arterials. In Los Angeles, the freeways are a major means for crosstown travel, for work as well non-work purposes. In Sacramento, the freeways measured do not seem to be different than non-freeway segments. More data will be necessary to draw any conclusions.

HOV Lanes and AVO

HOV lanes are added to the highway system for the explicit purpose of increasing vehicle occupancy, and this study gave us an opportunity to partially judge their effectiveness. AVO was examined by whether a freeway link had an HOV lane or not. Only two of the data sets had HOV links -- Los Angeles-Ventura and Sacramento 1994. The Santa Monica freeway had a temporary HOV link open during the counting period, but none of the sampled HPMS points were on it. Los Angeles-Ventura had two types of HOV facilities -- a normal HOV lane along with mixed-flow lanes, and a bus/HOV facility that allows both buses and vehicles with three or more passengers; we have distinguished between these two facilities.

For observation points where there was an HOV lane, we combined HOV counts with the regular lanes. Thus, it is possible to see whether the total AVO increases for those links which include an HOV lane. However, we can't easily compare the HOV lanes with the non-HOV lanes on these facilities. Table 5 presents the results. Even though the number of observation periods for those locations with HOV lanes is small, it is clear that higher vehicle occupancies are being carried on links with HOV lanes. Both the Los Angeles-Ventura and Sacramento 1994 data sets show increased AVO for HOV segments; the difference is not significant in the Sacramento case, but it is in the expected direction. Further, with the Los Angeles-Ventura data, the mixed bus/HOV facility has an even higher AVO. HOV lanes are definitely associated with higher AVO.

Locational Variations

Implicit in many of these comparisons are differences between locations. AVO is not consistent throughout the transportation network, but varies by district, area and, even, neighborhood. This can be seen by looking at the standard deviations and at the range of AVO estimates by data source (Table 1). There are considerable differences in AVO between locations. Some of these are due to different time periods and road classes being compared. But even if only freeway segments on non-HOV links for the morning hours in the latest year are taken (data not presented), there are meaningful differences between travel routes in AVO. We further examined the relationship between AVO and two geographical variables - population density and whether the location was near a central business district, but found no relationship. It is clear that there are significant differences in AVO between locations, but we have not found a simple variable which explain the variation. Consequently, it is essential in documenting vehicle occupancy for a region to sample many locations to capture some of the variability.

Traffic Volume

The final variable examined was traffic volume. The sampled links for each data set were divided into two groups - 'high' volume links (those having measured volumes greater than the median volume for the data set) and 'low' volume links (those having measured volumes less than the median volume for the data set). Table 6 presents the results.

For the 1988-92 Los Angeles-Ventura data, there is a significant difference between the two types of links; those with higher traffic volumes have significantly higher AVO than links with lower traffic volumes. On the other hand, the 1994 Sacramento and Los Angeles data sets do not show significant differences and, in fact, show trends in the opposite direction. For the 1994 Sacramento and Los Angeles data, the effect of different road types may obscure the relationship. More research on this point is necessary.

Consistencies Among the Models

The analysis of the five data sets reveals some consistencies. First, AVO increased throughout the day; this was found in the three data sets that included afternoon observations. Second, Saturdays had significantly higher AVO than weekdays. This result was found in both the Sacramento and Los Angeles data sets, the only two in which measurements were taken on Saturdays.

Third, within weekdays, there were not consistent results for changes in AVO. This implies that measurements can be comfortably taken on Mondays through Fridays without the need for additional stratification. Fourth, it appears that freeway facilities with HOV lanes had higher AVO than those without such facilities. This outcome was found in the Los Angeles-Ventura data set, but only partially in the Sacramento data set. The higher AVO on HOV segments could indicate that the HOV policies have been somewhat effective. If HOV lanes are lightly utilized, then the AVO found might not be appreciably higher than the non-HOV facilities.

In terms of the other variables examined, the results are less clear. It is apparent that there are substantial locational differences in AVO. In the Los Angeles-Ventura data set (the only one where spatial variation was examined), a simple relationship to location was not found. Most likely, there are particular routes and locations with higher or lower AVO than the average. Also, no consistency was found for the effect of traffic volumes.

Need for a Representative Sampling Frame

Most highway agencies that collect AVO data on a continuing periodic basis use rules-of-thumb for determining count locations and frequencies. These rules-of-thumb are not based upon sampling theory, but rather seem based on precedent. However, the data presented here show that there are important differences in transportation conditions that effect vehicle occupancy levels. Further, these are more complex than the simple assumption that vehicle occupancy measurements conducted on freeways in a direction oriented toward a CBD during the morning period and in the middle of the week are somehow 'typical'. On the contrary, our data suggest that measurements must occur in the afternoons as well as in mornings, that Saturdays have significantly higher AVO than weekdays, and that there are substantial differences in types of roads and locations.

In addition, each urban area may have unique characteristics that require additional stratification (e.g., road types, CBD). When it comes to developing a sample frame for selecting a random sample of locations to measure regional or corridor vehicle occupancy, state and local departments of transportation need to incorporate these differences. Without taking into consideration these important variables, a biased estimate of AVO will be obtained.

VEHICLE OCCUPANCY SAMPLING STRATEGIES

To illustrate how these factors can be incorporated into a sampling frame, we will show how samples can be constructed for estimating regional AVO and corridor AVO. It should be noted that there is no single best method for drawing a random sample. Each sampling method optimizes certain variables, while ignoring others. The following procedure is reasonable for measuring daytime AVO at a regional level at a sufficient level of precision and with a limited potential for bias. We apply it at the regional level, though it can be applied to sub-regional, city, or, even, smaller areas. However, it should be clear that other designs could be used to increase precision for other variables.

We constructed the sample using the Los Angeles data. Our analysis showed that AVO varied by time of day, between weekdays and Saturdays, and by type of road with freeway segments having the lowest AVO and major arterials having the highest AVO. Thus, there is a need to incorporate these variables in any sampling design. For other geographical areas, additional variables may be important.

The formula for calculating the number of link-days that need to be sampled is computed from Ferlis (1981), and is similar to formulas used to define sample sizes for household surveys (Kish, 1965, 49-52):

$$N = \frac{Z^2 * \sigma^2_l}{T^2} \qquad (1)$$

where N is the number of observations to be sampled, Z is a plus or minus value on the standard normal sampling distribution of AVO estimates within which a particular percentage of the AVO measurements fall (e.g., 95 percent), &l is the standard deviation of AVO across locations, and T is an acceptable tolerance between the estimated AVO and the true AVO. Since the true value of &l is not usually known, an estimate based on a sample is used.

In the design, we want to represent the daytime regional AVO for a given year within a tolerance of +/- 0.03. Further, four additional categories will have their daytime AVOs for that year represented within a tolerance of +/- 0.04:

1. All freeway segments,
2. Morning periods,
3. Afternoon periods, and
4. Saturdays.

The design uses the HPMS data base (California Department of Transportation, 1990; 1993). It stratifies by traffic volume and by time of day, but assumes that a high proportion of vehicle trips are made on freeways (the HPMS estimated that 80% of all annual average daily traffic was on freeway segments in Los Angeles County). Since the HPMS sample is a random sample of all public road segments in California, it can be used as a basis for a sample of segments for vehicle occupancy measurement.

Sampling Procedure

The appendix illustrates this with an example of drawing a sample from the HPMS data base for Los Angeles County. Because of the length of the actual data base, only a few pages are shown. The procedure for drawing the sample was as follows:

1. All segments in the HPMS sample were listed by segment names and by annual average daily traffic volume (AADT);

2. All road segments were sorted in descending order of AADT;

3. It was assumed that the daytime standard deviation of AVO for the region was 0.16; this estimate was consistent with Table 1. Therefore, using formula (1), in order to represent the entire region within +/- 0.03 of the true daytime AVO for the region, there needs to be at least 109 observations in the entire sample;

4. However, since AVO for morning periods and for afternoon periods are being estimated within +/- 0.04, there must be a minimum of 61 observations during the morning periods and 61 observations during the afternoon. To obtain these error limits, 61 different links are to be selected and are measured twice in one day. This gives a total sample size of 122 observation periods for the year, which would yield a daytime AVO with a tolerance well within +/- 0.03 of the correct value;

5. In order to avoid potential bias from missing segments with large traffic volumes, the 40 highest volume segments were automatically selected and the remaining 21 segments were sampled based on a systematic interval. The top 40 were selected based on a graph of AADT on the Y-axis and rank order on the X-axis; there was a natural break at rank 40. The 40 largest segments in the HPMS data base were sampled and removed from the data base. These will be referred to as the 100-percent sampled segments. In the example in the appendix, these 40 links accounted for 43.6 percent of the total AADT for the entire sample;

6. The remaining non-100-percent segments were sorted by traffic volume in descending order (i.e., with the link having the largest traffic volume of the remaining segments listed first, with the link having the next largest traffic volume of the remaining segments listed next, and so forth);

7. Among the non-100-percent segments, the average annual daily traffic volumes were summed to yield the total estimated daily traffic volume for these remaining segments;

8. A cumulative count of the AADT for the non-100-percent sample segments was taken. That is, starting with the first segment in the list (the link with the largest traffic volume after excluding the 100-percent sampled segments), the traffic volume for that link was listed. For the next link on the list (the link with the next largest traffic volume of the remaining segments), the traffic volume for that segment was combined with the traffic volume from the first segment on the list. This process was continued until all segments were accumulated. The sum of the traffic volumes for the last segment on the list was equal to the total estimated traffic volume for the non-100-percent segments, calculated in step 7 above;[1]

9. A link sampling interval was calculated by dividing the total annual average daily traffic volume for the remaining segments by 21 (i.e., 61 - 40); call this interval, i. A random number between 1 and i was generated; call this number, s. The road segment containing traffic volume, s, within its range was selected. The interval, i, was added to s and the segment with a range which contains that number was selected next. This process was continued, adding i to the previous sum and selecting that segment in which the number falls within the range, until no more segments could be selected. From this list, 21 segments were selected (in addition to the 40 segments that were automatically selected). The appendix illustrates the process.

Observation Procedure

To implement vehicle occupancy measurements based on this sample, the following procedure can be used.

1. It is necessary to take observations on Monday through Saturday, with an approximately equal number of observations for each day of the week. Starting from the first day in January, each day is given a sequential number starting with 1 and continuing to 365. That is, each day of the year will be identified by a unique number. All Sundays are removed from the list. This should leave about 313 or 314 numbers corresponding to Mondays-Saturdays (depending on whether it is a Leap Year). Holidays can also be removed from the list to yield the remaining eligible days;

2. Each of the selected road segments is randomly assigned one of the numbers corresponding to the eligible days. A simple way to do this is to randomly generate a three digit number. If the number falls between 001 and the maximum number of eligible days (313 or 314 at a maximum), then the day corresponding to that number is assigned to the road segment. If the random number is greater than the eligible number, then it is dropped and another random number is generated;

3. This procedure will assign a particular day to each of the 61 road segments. Since each day of the week has an equal chance of being included (except for Sundays and any holidays that have been excluded), there should be an approximately equal number of Mondays, Tuesdays, Wednesdays, etc.; and

4. Finally, each selected road segment is observed on the selected day. Two measurements are taken, in the morning between 6:30 and 8:30 a.m. and in the afternoon between 4:00 and 6:00 p.m.

This design will yield a sample size of 122 different link-observation periods (61 links with two observation periods for each). Further, freeway segments are sampled with a high probability since they account for a sizeable proportion of the region's trips. Morning and afternoon periods will be equally represented, and Saturdays will represent approximately one-sixth of all observations.

Once the data are collected, an AVO and a standard deviation can be calculated for each observation session (selected location, l, and time period, p). Then, for all observation periods for each of the two strata separately (100-percent and non-100-percent), a stratum AVO can be produced by weighting each observation AVO by its proportion of the stratum traffic volume for the *selected* links. Finally, the two strata AVOs are combined into a regional AVO. The 100-percent stratum AVO is multiplied by the proportion of the total HPMS traffic volume accounted for by the 40 segments. The non-100-percent stratum AVO is multiplied by the proportion of total HPMS traffic volume accounted for by the non-100-percent strata before selecting the 21 segments. The two estimates are then added together to form a regional AVO (see formulas in Levine and Wachs, 1994). A standard error of estimate and 95% confidence intervals can also be calculated.

Adapting the Procedure to Measure AVO Along Parallel Corridors

The sampling method described above is aimed at measuring AVO for an entire region. Since the method uses the HPMS sample frame, it will not necessarily give accurate estimates along a particular freeway corridor. For example, in both demonstration tests in Los Angeles and Sacramento, a particular freeway corridor was chosen from which the HPMS locations within one mile of the freeway in each direction were used. In these cases, however, most of the HPMS links measured routes bringing traffic to and from the freeway rather than paralleling it.

Defining Corridors

The methodology needs to be modified for application to corridors. Conceptually, the logic of sampling for a corridor is not different from sampling for observations in an entire region. One constructs a 'universe' of road links and samples from them with a probability proportional to traffic volume. The major problem, however, is a conceptual one. What is a freeway corridor? If it is assumed that a major freeway segment forms the core of the corridor, then a corridor can be considered as a linear area extending several miles in either direction along the length of the freeway. However, this depends on the unique geography of the particular freeway segment. In some cases, there are meaningful corridors, while in other cases there are not.

In Los Angeles, for example, most freeways have parallel alternative routes. Los Angeles has a grid street pattern and the freeways tend to follow this. There are usually many alternative routes, even within a short distance. Corridors are easily defined. To illustrate this, the Santa Monica Freeway (I-10) segment that was measured in the field test will be used. Figure 1 below shows a map of the central Los Angeles freeways and the major arterials. The particular test segment was on the Santa Monica Freeway (I-10) between the San Diego Freeway (I-405) and the Harbor Freeway (I-110). The segment runs for 9.4 miles in an east-west direction and connects the west side of Los Angeles with downtown. It is a heavily traversed freeway and carries traffic in both directions during most hours of the day.

In Figure 2, all north-south streets and east-west streets that traverse short distances have been eliminated. The result is all the major arterials that parallel the Santa Monica Freeway, within the range of the sampled segments. As can be seen, there are numerous roads that run parallel to the freeway. Some run the entire span between the San Diego and Harbor Freeways, while others extend over only part of the distance. Several streets are parallel for part of the segment, but curve away for other parts of the segment. These east-west streets represent meaningful alternatives to the Santa Monica Freeway, on different days and at different times of the day.

On the other hand, along other freeway segments, there may be no parallel roads. In Sacramento, for example, along Interstate 99 south of downtown, there were not any major parallel roads. The 'corridor', therefore, is only the freeway. Sacramento has a pattern of freeways radiating from the central business district. This creates a unique problem in defining corridors since surface streets generally follow a grid pattern.

Thus, the first task in taking a sample to measure corridor AVO is to identify which roads would truly represent parallel routes around a stretch of freeway. A second task is to determine how far away from the freeway segments in both directions a parallel corridor extends. Again, this depends on geography and on local traffic patterns. Unique, local knowledge is necessary to define the extent of the corridor.

Illustration of Selecting Alternative Routes around a Freeway Segment

To illustrate the corridor concept, using the Santa Monica Freeway segment, a two mile linear zone along the freeway segment was arbitrarily defined (Figure 3). This reduces the number of parallel streets that have to be considered. Some of these appear to be reasonable choices, particularly at the western edge of the segment. Others, however, appear to be closer to the Hollywood Freeway (I-101) which travels in a northwest-southeast direction. Consequently, the number of streets was further reduced to those that meaningfully constitute alternatives to the Santa Monica Freeway, based on our understanding of traffic conditions in the area (Figure 4). In this map, parallel arterials that are closer to the Hollywood Freeway and arterials that run only a short distance parallel to the Santa Monica Freeway have been removed. The result is six streets that parallel either the entire east-west length that the Santa Monica Freeway segment traverses or the vast majority of that length.

Selecting Measurement Locations within the Selected Corridor

The logic of selecting measurement locations to represent a corridor is similar to that for selecting locations to represent a region, except the sampling 'universe' is now the street and freeway segments that form the corridor.

Assuming that the corridor standard deviation is around 0.16 and that we wish to represent the corridor within +/- 0.04 of the true corridor AVO, then according to formula (1), a sample size of 61 observations will be necessary. This could be allocated to 30 or 31 different locations, measured twice a day. Even if the tolerance is reduced to +/- 0.05, then there would still need to be 39 observations to achieve this level of precision. In the case of the Santa Monica Freeway example, with seven roads (the freeway and six parallel arterials), each road would need to be measured at 5 or 6 separate locations. These can be selected randomly using a procedure such as that presented above. Each separate location is randomly assigned to a single day in the year and is observed on that day during a morning and an afternoon period. The corridor AVO would then be determined as the mean of the AVOs at these locations, weighted by their traffic volume.

CONCLUSION

Monitoring the vehicle occupancy in a region or along a particular corridor can be important to detect changes in the efficiency of road use over time or in the distribution of different types of vehicle occupancies over time (e.g., whether bus occupancy increases or not). To do this, a precise and unbiased sample must be selected so that the observations truly represent regional vehicle occupancy levels. It is not acceptable for a state or local department of transportation to haphazardly select locations and time periods and then expect the AVO estimates from such observations to accurately represent regional or, even, corridor vehicle occupancy levels. There is simply too much bias in such a method. One must use a systematic sampling theory approach.

ACKNOWLEDGMENTS

This study was funded by a grant from the Office of Traffic Improvement, California Department of Transportation to the Lewis Center for Regional Policy Studies and the Institute of Transportation Studies at UCLA. The authors wish to acknowledge the support of Les Jones, project manager, Don Ochoa, technical consultant, and Norm Roy, director of the office. In addition, we want to thank our research assistants: Janice Murabayashi, Sparky Harris, Patricia Larsen, Samuel Lau and Mary Jane Breinholt.

REFERENCES

Arizona Department of Transportation (1989). *Vehicle Occupancy Determinants. Final Report*. Federal Highway Administration, Washington, DC.

Benke, R. J. and Sjoberg, R. A. (1977). Auto occupancy parameter variations. *Traffic Engineering*, 7.

Billheimer, John W. *Use of Videotape in HOV Lane Surveillance and Enforcement: Final Report*. Washington D.C., March 1990.

California Department of Transportation (1990). *Highway Performance Monitoring System: California Report for 1989 Calendar Year*. August. Sacramento.

California Department of Transportation (1993). *Highway Performance Monitoring System: Los Angeles Count*. November. Digital Format. Sacramento.

Cervero, Robert (1996). Jobs-housing balance revisited: Trends and impacts in the San Francisco Bay Area. *Journal of the American Planning Association*, 62, 492-511.

Ferlis, R. A. (1979). Field Data Collection and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy. *Transportation Research Record*, 701.

Ferlis, R. A. (1981). *Guide for Estimating Urban Vehicle Classification and Occupancy*. March. Federal Highway Administration, Washington, D.C.

Fisher, Fred R., Williams, G. Jack, and Boyd, J. Phillip (1980). Atlanta Vehicle Occupancy Monitoring. *Transportation Research Record*, 779.

Higgins, Thomas J. (1995). Congestion management systems: evaluation issues and methods. *Transportation Quarterly*, 49, 25-42.

Hupp, R. and Carmine Palombo (1980). Evaluation of the FHWA Vehicles Classification and Automobile Occupancy Sampling Manual. *Transportation Research Record*, 779.

Kish, Leslie (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Klusza, Ron (1991). *Vehicle Occupancy Monitoring*. California Department of Transportation, Los Angeles.

Kuziw, George. Manager, Bureau of Data Resources, New Jersey Department of Transportation. Personal Communication, 9 July 1993.

Lester, M., Dare, J. W., and Roach, William T. (1979). Techniques for Monitoring Automobile Occupancy: Research in the Seattle Area. *Transportation Research Record*, 701.

Levine, Ned and Wachs, Martin (1994). *Methodology for Vehicle Occupancy Measurement*. Report Submitted to The California Air Resources Board, and The Office of

Traffic Improvement, California Department of Transportation. October 27. Lewis Center for Regional Policy Studies and the Institute for Transportation Studies School of Public Policy, University of California, Los Angeles.

Levine, Ned and Glickfeld, Madelyn (1996). Land use conflicts among California jurisdictions. In preparation.

McFadden, Bill and Jim Innes. *Photographic Enforcement of HOV Lanes*. 1986.

Maricopa Association of Governments (1993). *1992 Study of Occupancy and Vehicle Classification in the Metropolitan Phoenix Area*. Prepared by Lee Engineering. Phoenix, AZ.Pisarski, Alan E (1996). *Commuting in America, II: The Second National Report on Commuting Patterns and Trends*. Lansdowne, VA: Eno Transportation Foundation, Inc.

Richardson, Harry and Peter Gordon (1989). Counting Nonwork Trips: The Missing Link in Transportation, Land Use, and Urban Policy." *Urban Land*. September, 6-18.

Rutherford, Scott G., Kinchen, Ruth K., and Jacobson, Leslie N. (1990). Agency Practice for Monitoring Violations of HOV Facilities. *Transportation Research Record*, 1280.

San Diego Association of Governments (1982). *Vehicle Occupancies and Classifications at 100 Sites*. May. San Diego.

San Diego Association of Governments (1985). *San Diego Regional Vehicle Occupancy and Classification Study - 1985*. August. San Diego.

San Diego Association of Governments (1990). *San Diego Regional Vehicle Occupancy and Classification Study - 1990*. September. San Diego.

Turnbull, Katherine F., Henk, Russell H., and Christiansen, Dennis L. (1991). *Suggested Procedures for Evaluating the Effectiveness of Freeway HOV Facilities*. February, Technology Sharing Program, U.S. Department of Transportation, Washington, DC.

Turnbull, Katherine F., Stokes, Robert W., and Henk, Russell H. (1991). Current Practices in Evaluating Freeway HOV Facilities. *Transportation Research Record*, 1299, 63-73.

Ulberg, Cy and McCormack, Edward (1988). Accuracy and Other Factors Affecting a Continuous Vehicle Occupancy Monitoring Program. *Transportation Research Record*, 1206.

Wachs, Martin, Taylor, Brian, Levine, Ned, and Ong, Paul (1993). The Changing Commute: A Case Study of the Jobs/Housing Relationship Over Time. *Urban Studies*, 30, 1711- 1729.

Table 1

VARIATIONS IN AVO AMONG DATA SETS
Mean, Standard Deviation and Range of Link Estimates

	Data Set				
	Los Angeles-Ventura 1988-92	SANDAG 1985	SANDAG 1990	Sacramento 1994	Los Angeles 1994
Number of locations	162	100	100	36	59
AVO across all observations[1]	1.204	1.285	1.245	1.364	1.450
Standard deviation (σ_T)	0.068	0.157	0.131	0.152	0.155
Minimum	1.066	1.103	1.087	1.136	1.174
Maximum	2.256	2.100	1.946	1.754	1.857

[1] Differences in AVO between data sets are significant at $p \leq .001$ ($F=82.48$).

Table 2

CALCULATED AVO BY TIME PERIOD

	Data Set					
	Los Angeles-Ventura 1988-92		Sacramento 1994		Los Angeles 1994	
Time Period	AVO	N	AVO	N	AVO	N
Morning	1.187	(433)	1.289	(12)	1.398	(20)
Noon Period	-		1.374	(12)	1.440	(19)
Afternoon	1.265	(97)	1.428	(12)	1.513	(20)
Significance Test	F = 80.95 p<=.001		F = 2.84 n.s.		F = 3.09 n.s.	

Table 3

CALCULATED AVO BY DAY OF WEEK

	Data Set					
	Los Angeles-Ventura 1988-92		Sacramento 1994		Los Angeles 1994	
Day of Week	AVO	N	AVO	N	AVO	N
Monday	-		1.363	(6)	1.473	(9)
Tuesday	1.199	(199)	1.242	(6)	1.496	(6)
Wednesday	1.197	(202)	1.345	(6)	1.376	(12)
Thursday	1.210	(171)	1.425	(6)	1.374	(9)
Friday	-		1.274	(6)	1.451	(11)
Saturday	-		1.533	(6)	1.542	(12)
Significance Test	F = 1.25 n.s.		F = 4.26 $p<=.01$		F = 2.19 n.s.	

Table 4

CALCULATED AVO BY ROAD TYPES

	Data Set							
	SANDAG 1985		SANDAG 1990		Sacramento 1994		Los Angeles 1994	
Road Type	AVO	N	AVO	N	AVO	N	AVO	N
Freeway/ Expressway	1.188	(17)	1.167	(17)	1.379	(18)	1.346	(13)
Major Arterial	1.241	(49)	1.214	(49)	1.348	(18)	1.525	(3)
Minor Arterial	1.352	(24)	1.294	(24)	-		1.477	(43)
Local	1.506	(10)	1.419	(10)	-		-	
Significance Test	F = 17.05 p<=.001		F = 13.61 p<=.001		F = 0.37 n.s.		F = 4.413 p<=.05	

Table 5

CALCULATED AVO BY HOV AND NON-HOV SEGMENTS

	Data Set			
	Los Angeles-Ventura 1988-92		Sacramento 1994	
Link Type	AVO	N	AVO	N
HOV (bus)	1.402	(24)	-	
HOV (normal)	1.360	(14)	1.451	(6)
Non-HOV Freeway	1.186	(693)	1.343	(12)
Significance Test	F = 214.82 p<=.001		F = 2.03 n.s.	

Table 6

CALCULATED AVO BY HIGH AND LOW TRAFFIC VOLUMES

	Data Set					
	Los Angeles-Ventura 1988-92		Sacramento 1994		Los Angeles 1994	
Link Type	AVO	N	AVO	N	AVO	N
'High' Volume	1.216	(286)	1.316	(18)	1.413	(30)
'Low' Volume	1.187	(286)	1.411	(18)	1.489	(29)
Significance Test	F = 18.62 p<=.001		F = 3.81 n.s.		F = 3.79 n.s.	

ENDNOTES

[1] These cumulated sums actually represent ranges. That is, the first segment on the list represents a range from 1 to some number, X. The second link on the list represents a range from X+1 to another number, Y, and so forth.

Figure 1

CENTRAL LOS ANGELES FREEWAYS

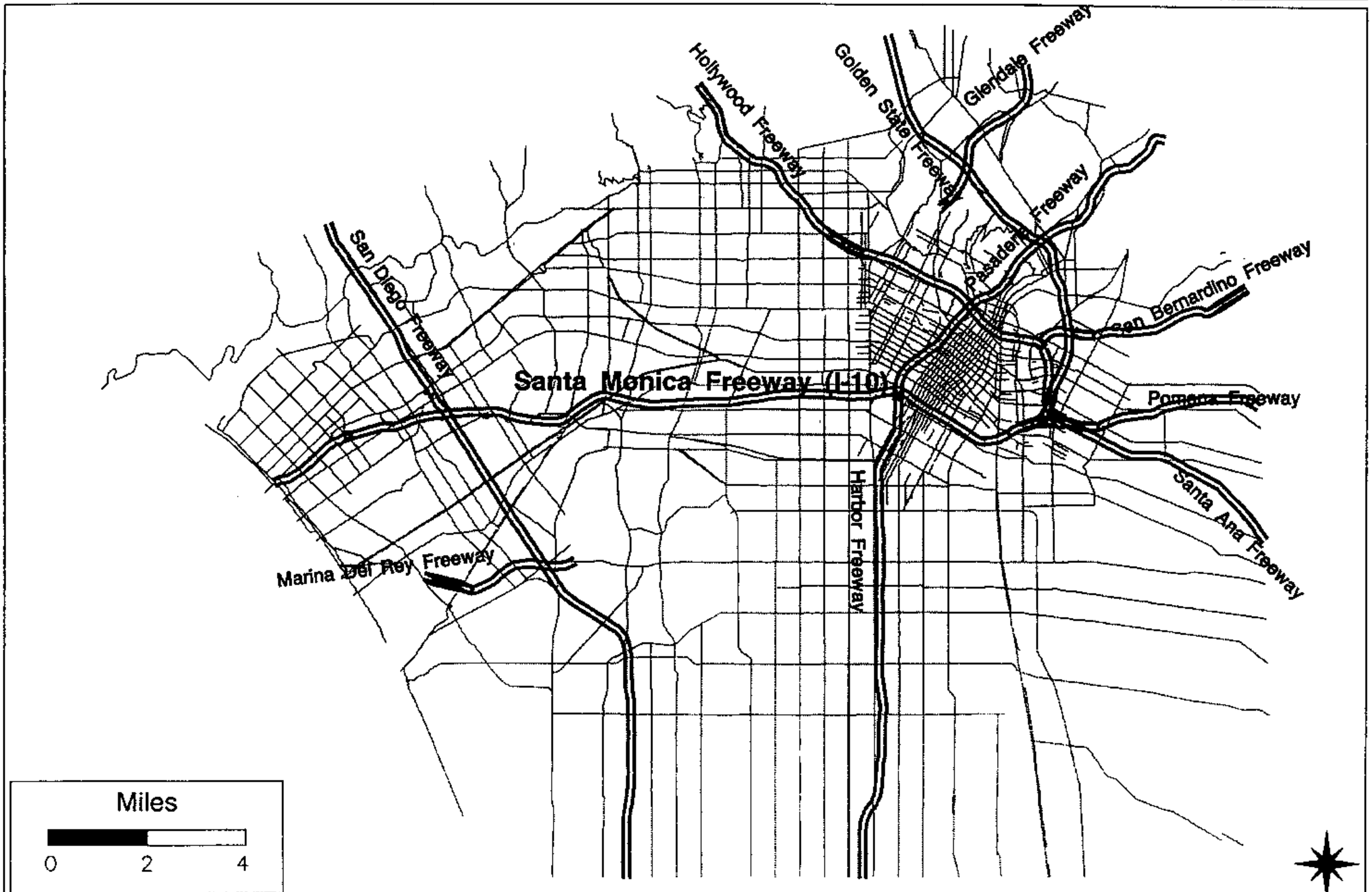


Figure 2

EAST-WEST ARTERIALS 'PARALLEL' TO SANTA MONICA FREEWAY

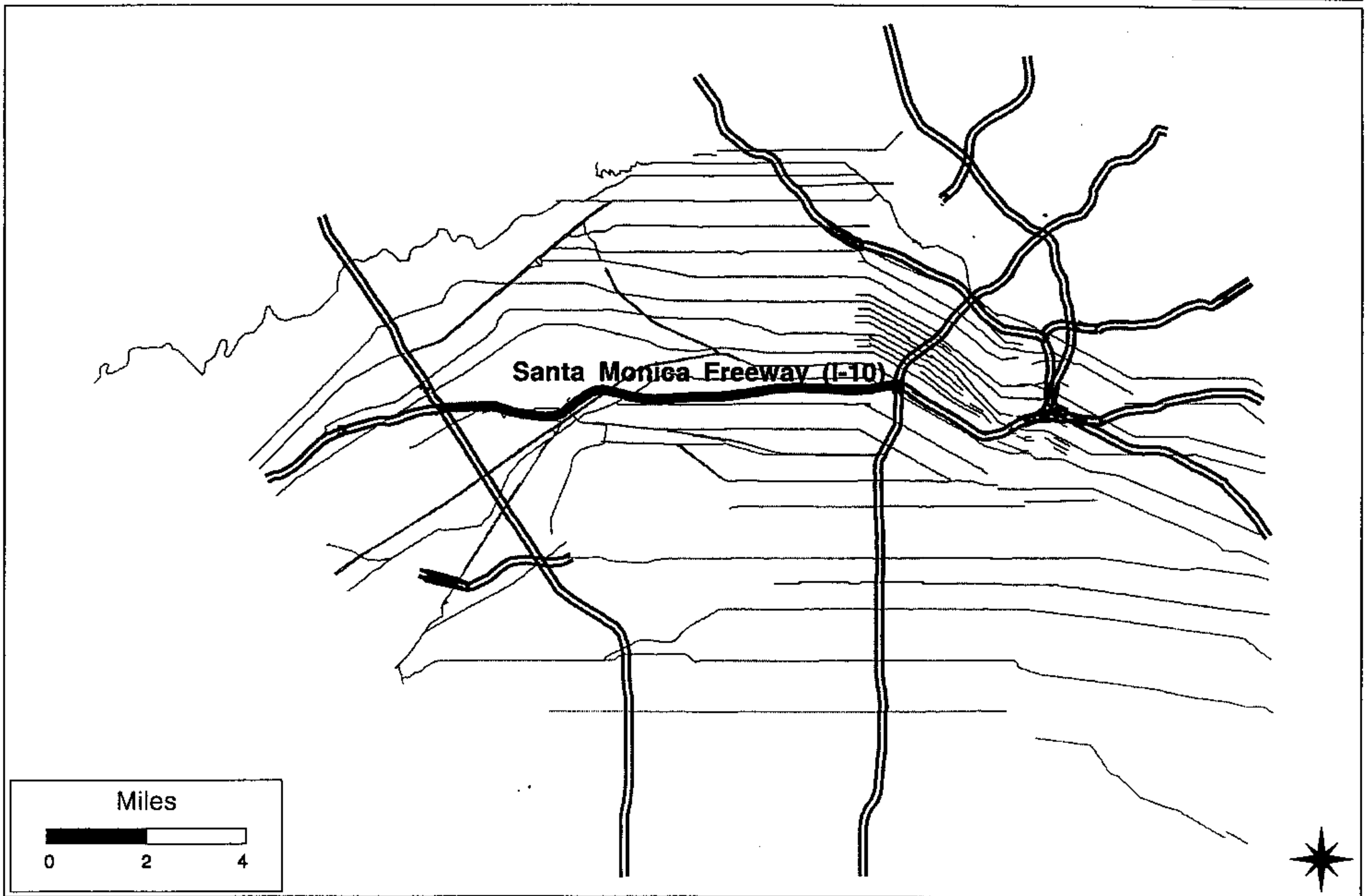


Figure 3

TWO MILE ZONE AROUND SANTA MONICA FREEWAY SECTION

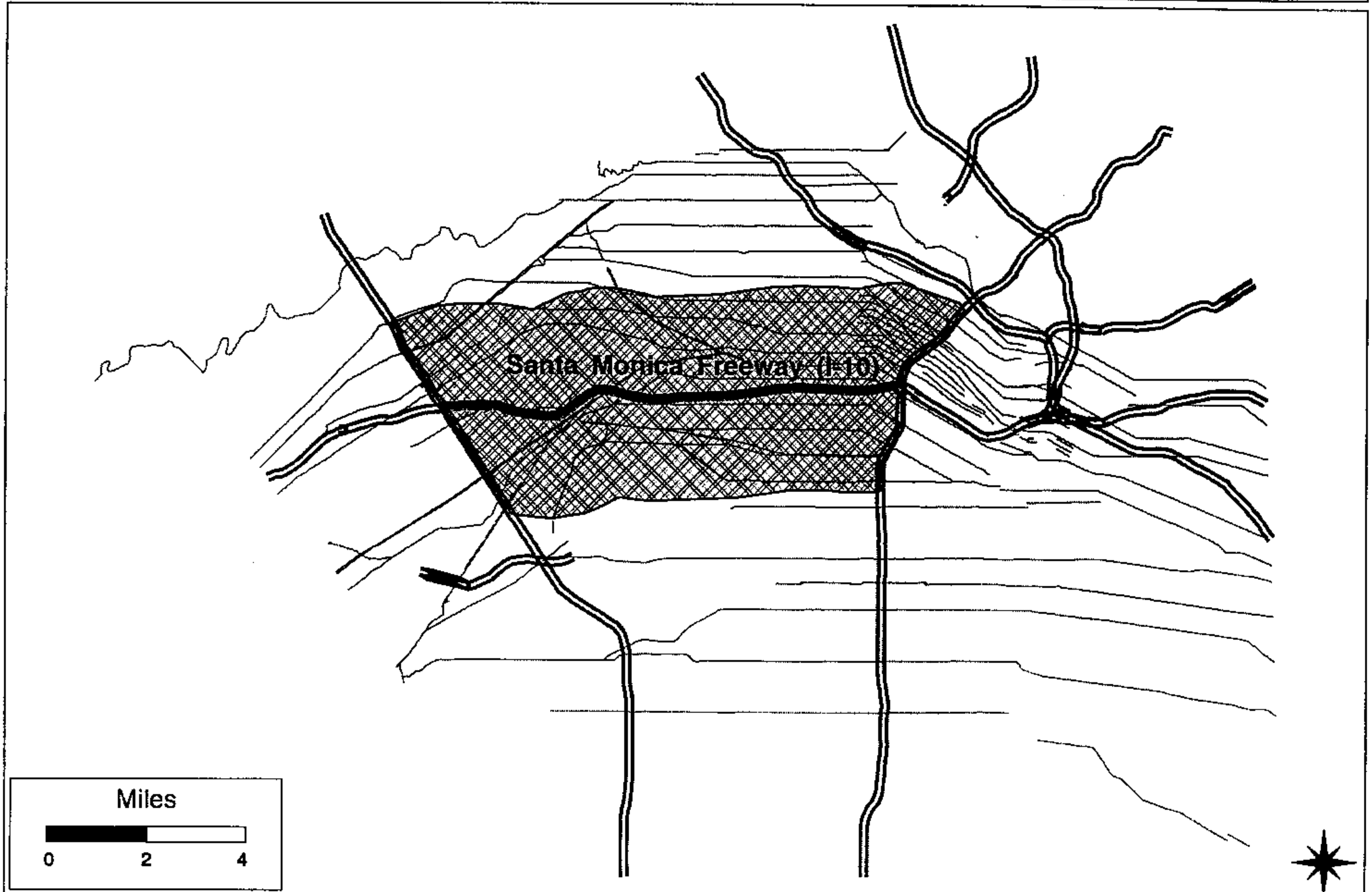
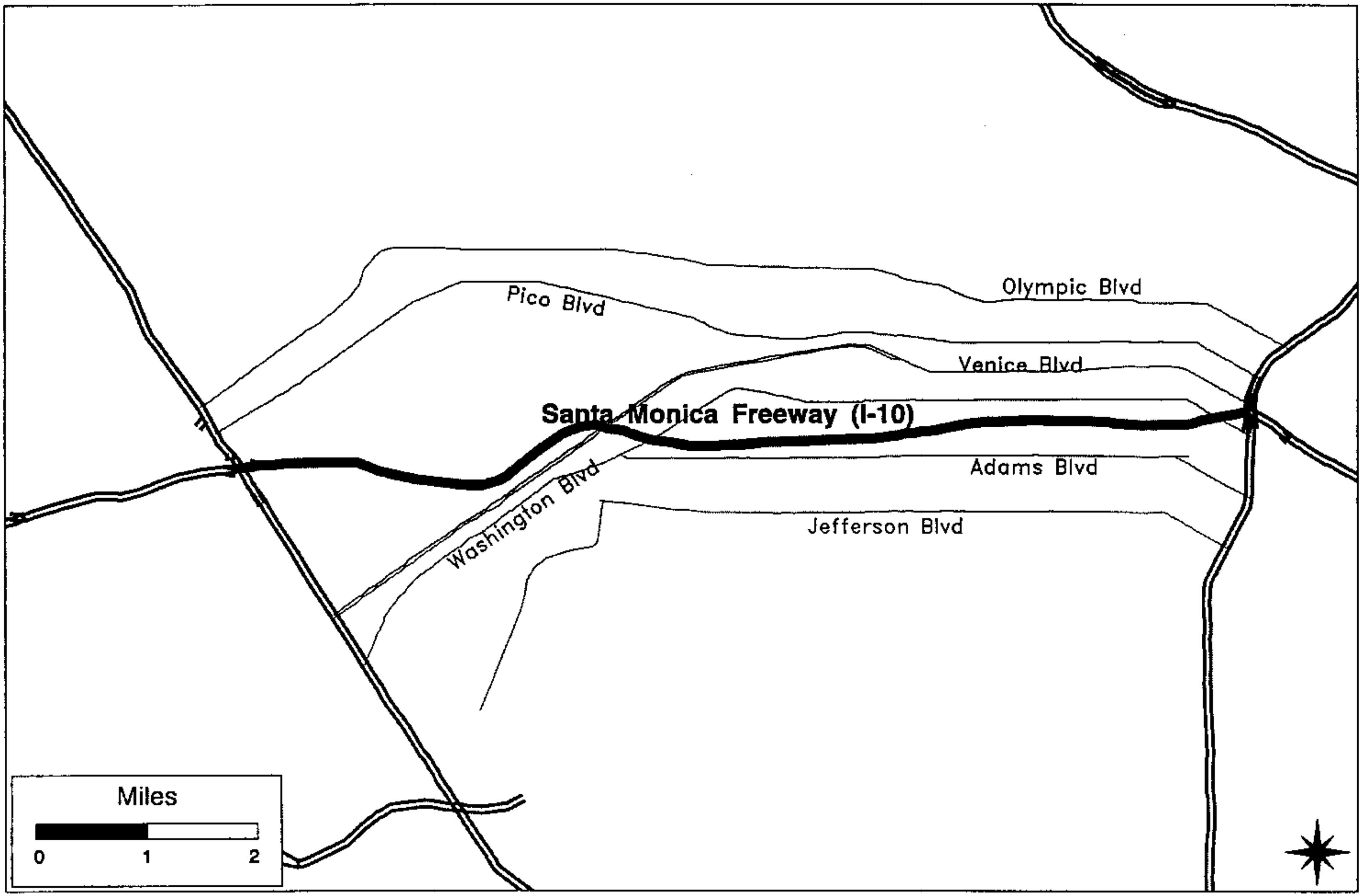


Figure 4

SELECTED STREETS FOR VEHICLE OCCUPANCY COUNTS



APPENDIX

SELECTING A SAMPLE TO ESTIMATE REGIONAL AVERAGE VEHICLE OCCUPANCY

A. 100% PROBABILITY SEGMENTS

Route	Section	Length	AADT	Cumulative AADT	Sampled Segments	AADT of Sampled Segments	Strata Proportional Weight
0010	11390	3.45	333100	n.a.	X	333100	0.0327
0405	26000	3.5	312700	n.a.	X	312700	0.0307
0405	29500	3.5	291200	n.a.	X	291200	0.0286
0010	14840	3.55	286000	n.a.	X	286000	0.0281
0110	21400	2.3	283900	n.a.	X	283900	0.0279
0101	17150	3.12	279100	n.a.	X	279100	0.0274
0405	33000	5.6	275800	n.a.	X	275800	0.0271
0405	38600	0.8	271000	n.a.	X	271000	0.0266
0405	13000	3.6	270800	n.a.	X	270800	0.0266
0405	16600	1	267900	n.a.	X	267900	0.0263
0010	5430	0.97	265800	n.a.	X	265800	0.0261
0405	24300	1.7	265400	n.a.	X	265400	0.0261
0091	11700	2.9	263300	n.a.	X	263300	0.0259
0101	13900	3.2	262900	n.a.	X	262900	0.0258
0405	17600	3.6	262700	n.a.	X	262700	0.0258
0010	6400	0.81	261200	n.a.	X	261200	0.0257
0405	21200	1	261000	n.a.	X	261000	0.0256
0101	1570	2.83	259800	n.a.	X	259800	0.0255
0405	23400	0.9	258200	n.a.	X	258200	0.0254
0405	22200	1.2	258000	n.a.	X	258000	0.0254
0405	4900	1.02	256800	n.a.	X	256800	0.0252
0101	9220	1.12	256000	n.a.	X	256000	0.0252
0091	15100	1.8	251600	n.a.	X	251600	0.0247
0405	6100	1.5	249700	n.a.	X	249700	0.0245
0405	4300	0.6	246600	n.a.	X	246600	0.0242
0405	3300	1	246000	n.a.	X	246000	0.0242
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	table continues	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
			10176700			10176700	1.0000

B. SAMPLED SEGMENTS

Random Start (S) = 442676.8
Sampling Interval (i) = 560226.1

Route	Section	Length	AADT	Cumulative AADT	Random Start + Sampling Interval	Sampled Segments	AADT of Sampled Segments	Strata Proportional Weight (Selected Segments Only)
0405	42400	3.9	192400	192400				
0710	13900	1.1	190000	382400				
0010	38510	0.46	189000	571400	442676.80	X	189000	0.1084
0710	17000	1.4	188000	759400				
0710	15000	0.7	185000	944400				
0710	15700	1.3	184400	1128800	1002902.9	X	184400	0.1058
0005	27080	4.15	182000	1310800				
0057	0	1.29	179000	1489800				
0134	9000	4.34	175400	1665200	1563129.0	X	175400	0.1006
0405	46300	0.6	173600	1838800				
0110	24500	1.28	167100	2005900				
0060	26530	2.86	161800	2167700	2123355.1	X	161800	0.0928
0005	45580	0.3	156000	2323700				
0710	12000	1	155500	2479200				
0210	43200	1.2	155000	2634200				
0710	9400	2.6	148200	2782400	2683581.2	X	148200	0.0850
0101	31050	1.73	144100	2926500				
0710	6880	2.52	144000	3070500				
0101	36180	2.01	144000	3214500				
0005	45880	3.15	135000	3349500	3243807.3	X	135000	0.0774
0005	49030	2.41	130800	3480300				
0605	21100	2.44	128800	3609100				
0710	24600	1.9	126500	3735600				
0014	25300	1.73	126000	3861600	3804033.4	X	126000	0.0723
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	table continues	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
			11764748				1743630	1.0000