# Exploratory Statistical and Geographical Freight Traffic Data Analysis

**Eric Moreno,** Senior Researcher, Instituto Mexicano del Transporte, Eric.Moreno@imt.mx

**Roberto Aguerrebere**, Transport Integration Coordinator, Instituto Mexicano del Transporte,

Roberto.Aguerrebere@imt.mx

**Agustin Bustos**, Associated Researcher, Instituto Mexicano del Transporte,

Agustin.Bustos@imt.mx

**Alfonso Balbuena**, Associated Researcher, Instituto Mexicano del Transporte,

Alfonso.Balbuena@imt.mx

**Abstract.** Data from freight traffic roadside surveys in Mexican highways are analyzed in order to find consistent patterns or systematic relationships between variables characterizing this traffic. Patterns traced are validated by contrasting against new data sets, allowing so a pattern's description refinement.

In a first stage truck traffic is characterized with respect to vehicle counting size, month of the year, geographical area and hourly behavior, yielding results similar to those of common analysis practices of traffic engineering. In a second stage the analysis of freight traffic variables like: traffic composition by type of freight vehicle or service class detected is followed.

The prime objective in this analysis is to reveal variable relationships or tendencies of freight data sets providing helpful predictions for decision making under uncertainty. As a second objective, the analysis is extended for the most reliable patterns to statistical inference in order to support hypothesis characterizing freight variables with the typical confidence levels used in decision making for managing and controlling of road freight traffic.

Techniques employed are those usual in the fields of Data Mining and Exploratory Data Analysis: frequency tables, histograms, box plots or correlation matrices. Besides, geographical representations of variable distributions with assistance of Geographic Information System (GIS) software is used.

Preliminary conclusions obtained from the analysis of the hourly behavior of truck traffic from several field surveys have shown the range of vehicle counting size variations, the identification of daily peak periods or the distribution of the type of vehicles present in freight traffic, amongst other results.

## 1. INTRODUCTION

In 1991, the Communications and Transport Secretariat of Mexico and the Mexican Transportation Institute started a field survey to gather basic information on freight traffic in Mexican highways. Since then, this field survey is achieved annually, collecting important data concerning vehicles and transport moving on the major Mexican highways. The vehicle types observed, their actual weights and sizes, the kind of commodities carried and the principal origins and destinations are among the data obtained in this field survey. As a product of this annual task, data bases have been integrated, and some statistical reports have been published (*1, 2*).

This paper describes the seasonal appearance and the geographical distribution showed by some variables related to data gathered in the field study referred to. The main objective is to get a first glance at the temporary and regional behavior of these variables as a step towards a better knowledge of the freight transport characteristics in the principal Mexican highways. Such a knowledge contributes to an improved understanding of the road freight flow phenomenon and can suggest action lines for a more accurate management of the necessary survey requirements or for a more effective application of control and inspection procedures to the hauling activity on the highway network.

## 2. THE EXPLORATORY ANALYSIS APPROACH

When data are analyzed for the purpose of studying a certain phenomenon, it is usual to begin with some assumptions expecting to be confirmed by the same data; this is the traditional approach in the statistical hypothesis testing. On the contrary, the exploratory approach tries to find consistent patterns or regular relationships between variables, and in a subsequent stage, confirm them on new data sets. This exploratory approach advances through three basic steps: a) data exploration, b) pattern or relationships determination and c) verification of findings on new data sets.

Following this approach, freight traffic data for the 1991-1996 period was analyzed, obtaining some findings that were confirmed with the 1997 data. The geographical presentation of the data was supported by the use of a GIS environment, allowing an easier and direct view of the regional characteristics of the information.

## 3. SEASONAL FEATURES OF THE FREIGHT VEHICLE FLOWS

Data for the field survey are gathered 24 hours a day on temporary sites located every year on different sites of the main road network during several days of a week. The selection of sites is designed to progressively cover most of the intercity freight flows in Mexico. Five years were analyzed: 1991, 1993, 1994, 1995 and 1996 giving a total of 522,804 freight vehicles registered in 74 survey stations, that is an average of 10,456 freight vehicles by year and 7,065 freight vehicles by station. The 1997 data, used to compare results recorded a total of 126,687 freight vehicles in 20 survey stations. For 1992 there were no hourly data, so this year was cut off from the sample.

### 3.1 Basic Description of Data

The first vehicle traffic characteristics examined were the daily average and the minimum and maximum of the hourly average.

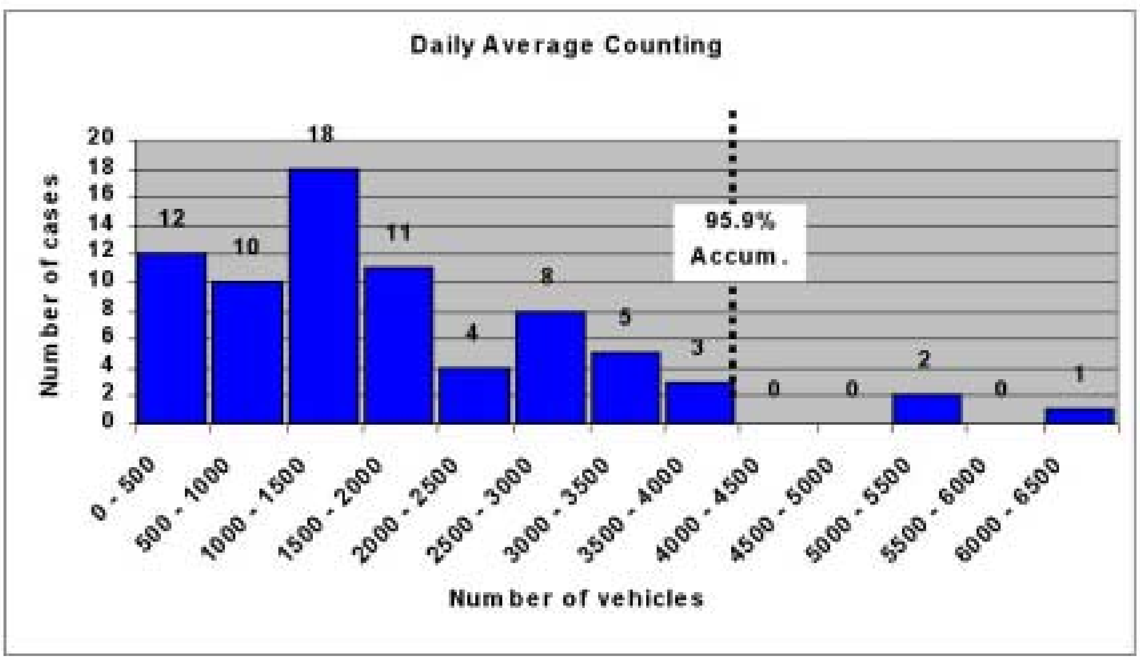


Figure 1

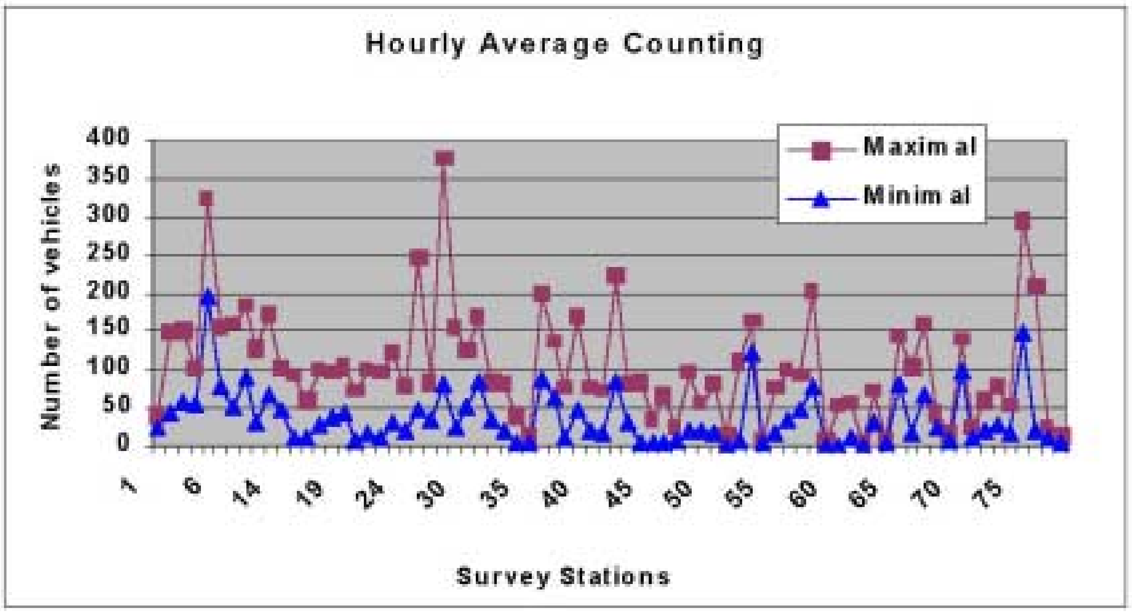


Figure 2

Fig. 1 shows that in 95.9% of the cases the daily average vehicle counting is below 4000. In particular, the global maximum took place at station “Tepotzotlán” gauged in 1991 with an Average Annual Daily Traffic of 6,376 freight vehicles. Fig. 2 displays maximum and minimum values of the hourly averages counting at survey stations, suggesting that most maximum values are under 250 vehicles, and most minimum values are no more than 100 vehicles. These values give a first idea of the vehicular flows magnitudes to evaluate the material requirements for survey conducting.

The maximum value for the hourly average counting is a basic figure used to estimate the vehicular flow that could be expected in the peak hours. This maximum value changes with Average Annual Daily Traffic in the stations. Upon relating the values of observed Average Annual Daily Traffic against the corresponding maximum values of the hourly average counting, a lineal relationship was found. Figure 3 shows this relationship with the regression equation and the $R^2$ coefficient, indicating a good fit for the model. This result suggests that the values of maximum hourly average counting are approximately 5.77% of the corresponding values of

Average Annual Daily Traffic, giving a first approach of the maximum values of hourly average counting that can be expected in the stations.

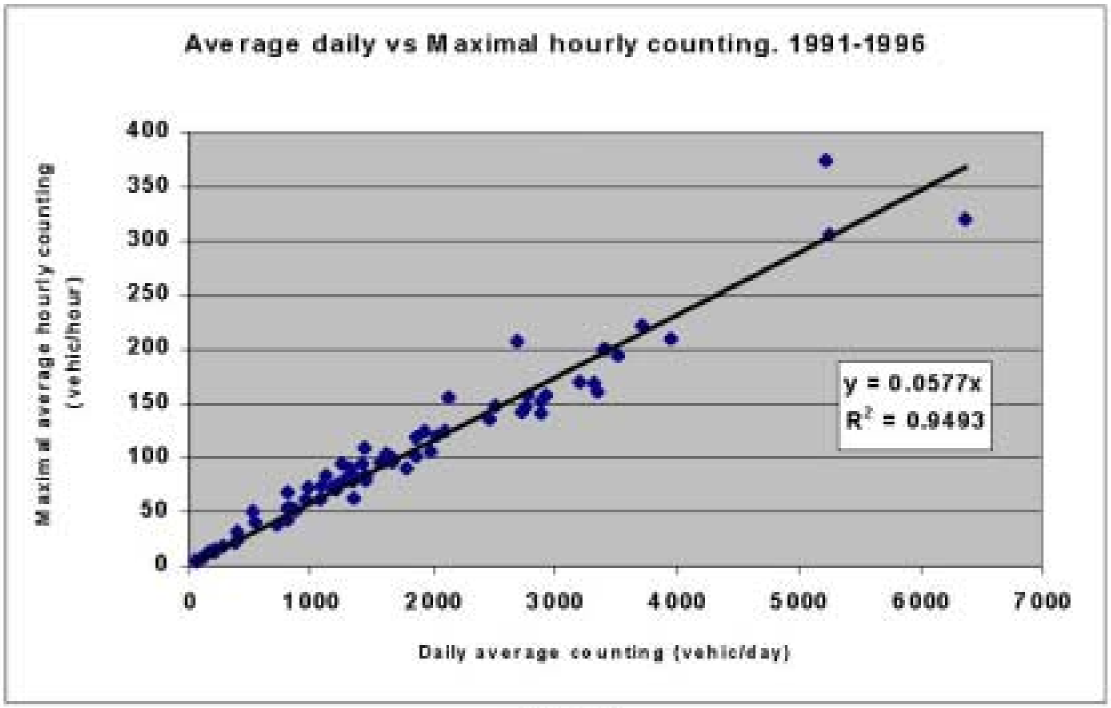


Figure 3

### 3.2 Variations in the Hourly Percentages of Vehicle Flows

The variation of the vehicular flow along the day was another aspect studied. Since the total vehicular counts registered varied from the order of hundreds until the dozens of thousands, the percentage of the total flow was examined in every hour of the day for each station. Figure 4 shows the average hourly percentages of the total flow, indicating a similar behavior for every year of the analysis.

The main flow period occurs from 08: 00 to 18: 00 hours approximately, presenting values of 4% to 6% of the total flow. Likewise, the relaxed periods run from 23: 00 to 05: 00 hours with values below the 4% of the total flow (notice that the mean percentage is 4.17%= 100%/ 24hrs).

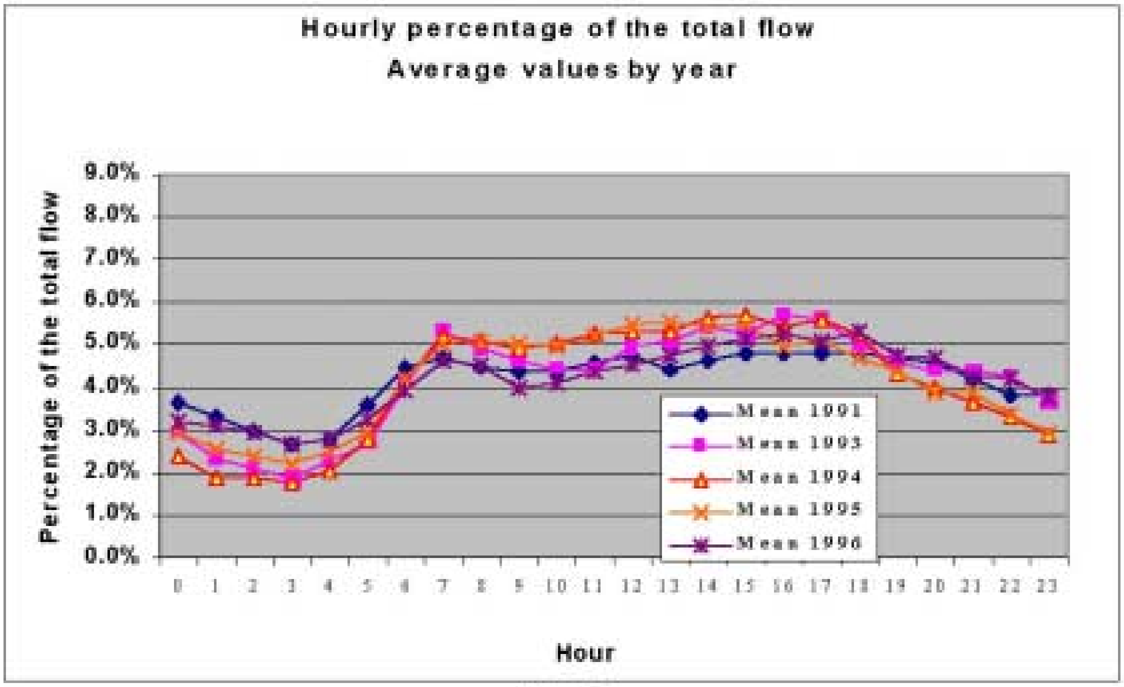


Figure 4

The number of stations classified by month of survey is as follows:

- January 1
- February 4
- March 3
- April 1
- May 5
- June 6
- July 10
- August 7
- September 10
- October 18
- November 7
- December 2

The average percentages of the flow for each month are shown in Fig. 5, taking only months with 2 or more survey stations. Curve shapes are in general similar to those of the Fig. 4 (per year curves), in what to peak and valley periods refer, however, per month curves show greater dispersion.

Concerning the day of the week in which data were gathered Figure 6 exhibits the average percentage of the flow for Tuesday, Wednesday, Thursday and Friday. This graph shows a likeness of the curves with the yearly and monthly ones, although with two much lower dispersion.

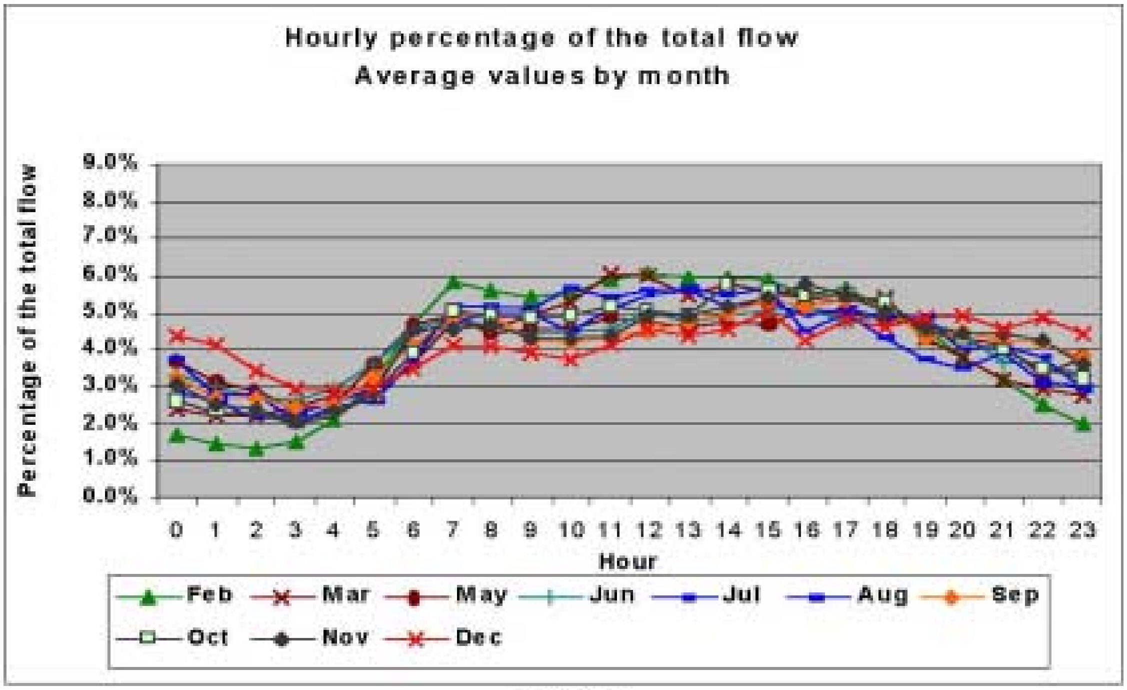


Figure 5

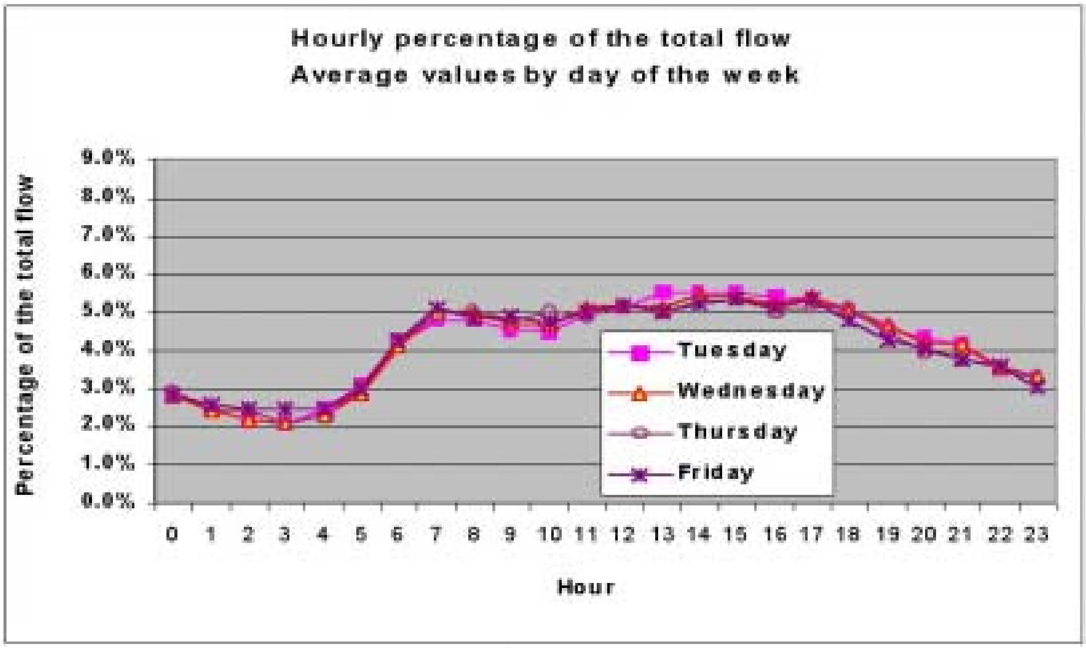


Figure 6

The different graphs shown give a visual idea of the behavior of the hourly percentage of the vehicular flow in the survey stations. A more impartial judgment of this likeness is achieved through the correlation coefficients between the average values generating these curves. Table 1 contains the correlations between these averages hourly percentages of flow in the five years studied, confirming so the resemblance observed in Figure 4.

**TABLE 1. Correlations**
**Mean percentages of total flow**
**1991-1996**

| | MEAN 1991 | MEAN 1993 | MEAN 1994 | MEAN 1995 | MEAN 1996 |
|---|---|---|---|---|---|
| MEAN 1991 | ---- | | | | |
| MEAN 1993 | 0.943 | ---- | | | |
| MEAN 1994 | 0.929 | 0.954 | ---- | | |
| MEAN 1995 | 0.901 | 0.922 | 0.985 | ---- | |
| MEAN 1996 | 0.936 | 0.963 | 0.912 | 0.859 | ---- |

For the case of the monthly classification, all the correlation coefficients between average percentages of flow of the months with 2 or more stations were calculated, resulting good values in all cases except December. For most of the cases, the correlation coefficients ranged from 0.789 (minimum for August-November) to 0.975 (maximum for September-November). For December, the correlations ranged from 0.387 (minimum for February-December) to 0.699 (maximum for September-December), indicating a different pattern for this month.

For the averages referred to the day of the week, Table 2 shows the correlation coefficients obtained.

**TABLE 2. Correlations**
**Mean percentages of total flow by day of the week**

| | Tuesday | Wednesday | Thursday | Friday |
|---|---|---|---|---|
| Tuesday | ---- | | | |
| Wednesday | 0.991 | ---- | | |
| Thursday | 0.979 | 0.989 | ---- | |
| Friday | 0.978 | 0.987 | 0.989 | ---- |

The high values of correlation in this last table confirm the great likeness shown previously in Figure 6.

### 3.3 Comparisons with 1997 Data

The analysis results of the 1991 to 1996 data were compared with 1997 data, comprising 20 stations gauged during the months of September, October and November.

Table 3 displays the frequency distribution of Average Annual Daily Traffic, for both groups of data (1991-1996 & 1997). From the table, the accumulated 95% for 1997, occurs at the 3001 to 3500 vehicles interval, whereas the 95.9% for the period 1991-1996 happens at the 3501 to 4000 vehicles interval. For both groups of data, the interval with the greatest number of cases (survey stations) was that of 1001 to 1500 vehicles, representing a 24.3% for 1991-1996 data and a 30% for the 1997 data. The table shows two extreme values greater than 6000 vehicles, both corresponding to surveys at "Tepotzotlán" station, the first in 1991 and the last in 1997.

| TABLE 3 | 1991-1996 Data | | | 1997 Data | | |
|---|---|---|---|---|---|---|
| **Counting interval (vehicles)** | **Number of stations** | **Percentage** | **Accum. Percentage** | **Number of stations** | **Percentage** | **Accum. Percentage** |
| 0 to 500 | 12 | 16.2 | 16.2 | 5 | 25.0 | 25.0 |
| 501 to 1000 | 10 | 13.5 | 29.7 | 5 | 25.0 | 50.0 |
| **1001 to 1500** | 18 | **24.3** | 54.1 | 6 | **30.0** | 80.0 |
| 1501 to 2000 | 11 | 14.9 | 68.9 | 1 | 5.0 | 85.0 |
| 2001 to 2500 | 4 | 5.4 | 74.3 | 1 | 5.0 | 90.0 |
| 2501 to 3000 | 8 | 10.8 | 85.1 | 0 | 0.0 | 90.0 |
| 3001 to 3500 | 5 | 6.8 | 91.9 | 1 | 5.0 | **95.0** |
| 3501 to 4000 | 3 | 4.1 | **95.9** | 0 | 0.0 | 95.0 |
| 4001 to 4500 | 0 | 0.0 | 95.9 | 0 | 0.0 | 95.0 |
| 4501 to 5000 | 0 | 0.0 | 95.9 | 0 | 0.0 | 95.0 |
| 5001 to 5500 | 2 | 2.7 | 98.6 | 0 | 0.0 | 95.0 |
| 5501 to 6000 | 0 | 0.0 | 98.6 | 0 | 0.0 | 95.0 |
| 6001 to 6500 | 1 | 1.4 | 100.0 | 0 | 0.0 | 95.0 |
| and greater... | 0 | 0.0 | 100.0 | 1 | 5.0 | 100.0 |
| **TOTALS:** | 74 | | | 20 | | |

The maximum values of the hourly average counting for the period 1991-1996 was compared with those of 1997. Figure 7 shows these values for the two groups of data indicating that for the great majority of stations of 1997, the maxima of the hourly average counting look alike those of the period 1991-1996, not surpassing the 200 vehicles. The only exception is again the station "Tepotzotlán" with more than 600 vehicles.

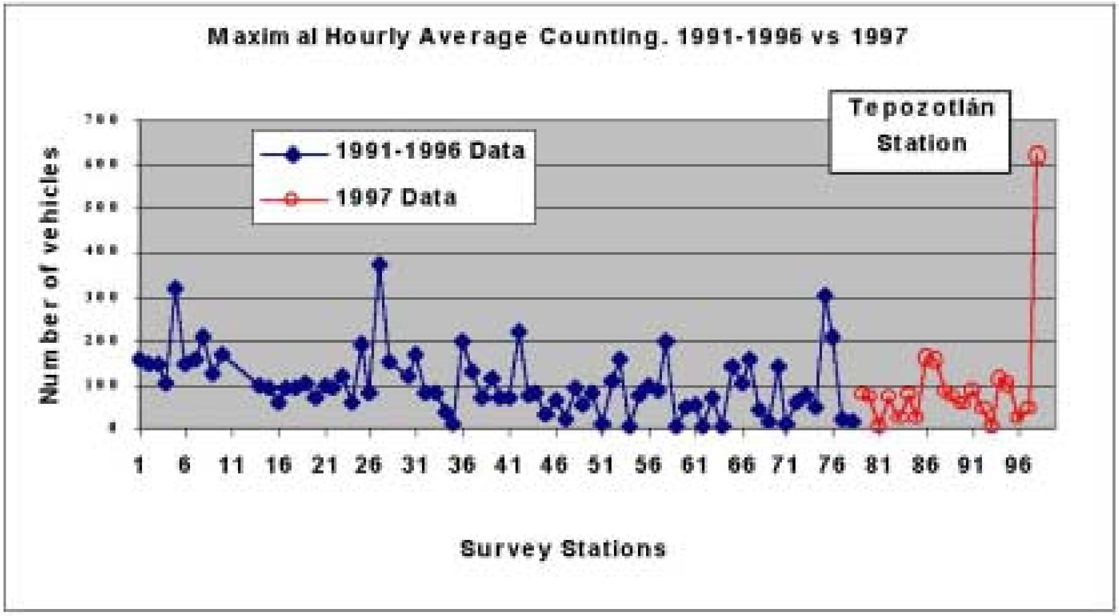


Figure 7

The lineal relationship found between the Average Annual Daily Traffic and the maximum of the hourly average counting in the analysis of data 1991-1996 was confirmed with the 1997 data. For this purpose, the equation Y= 0.0577* X, referred in Figure 3 was used. After comparing the values predicted by the previous equation with the observed data of maximum hourly average counts in 1997, the straight line of the Figure 8 was obtained. In this figure a regression equation also appears, with a slope of 0.9225 (very near to 1) and a coefficient $R^2$ = 0.9742 denoting a good fit. The fact that the slope of this line is close to one, suggests a good likeness between the values predicted by the equation and the values observed in 1997.

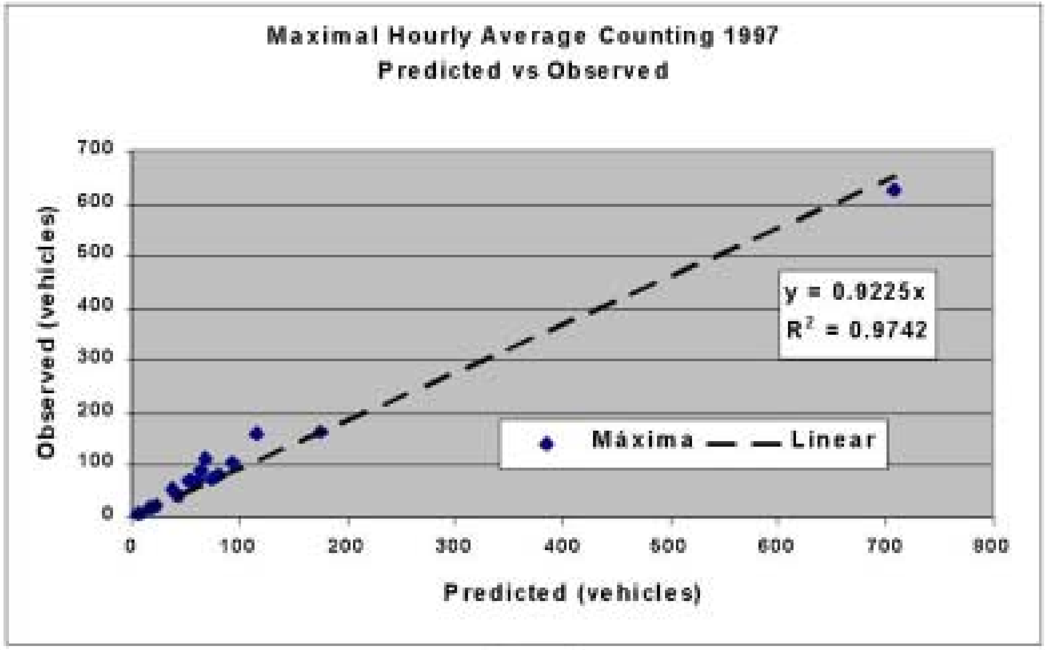


Figure 8

Finally, the variations of the hourly average percentage of flow observed in 1991-1996 were compared with the corresponding to 1997.

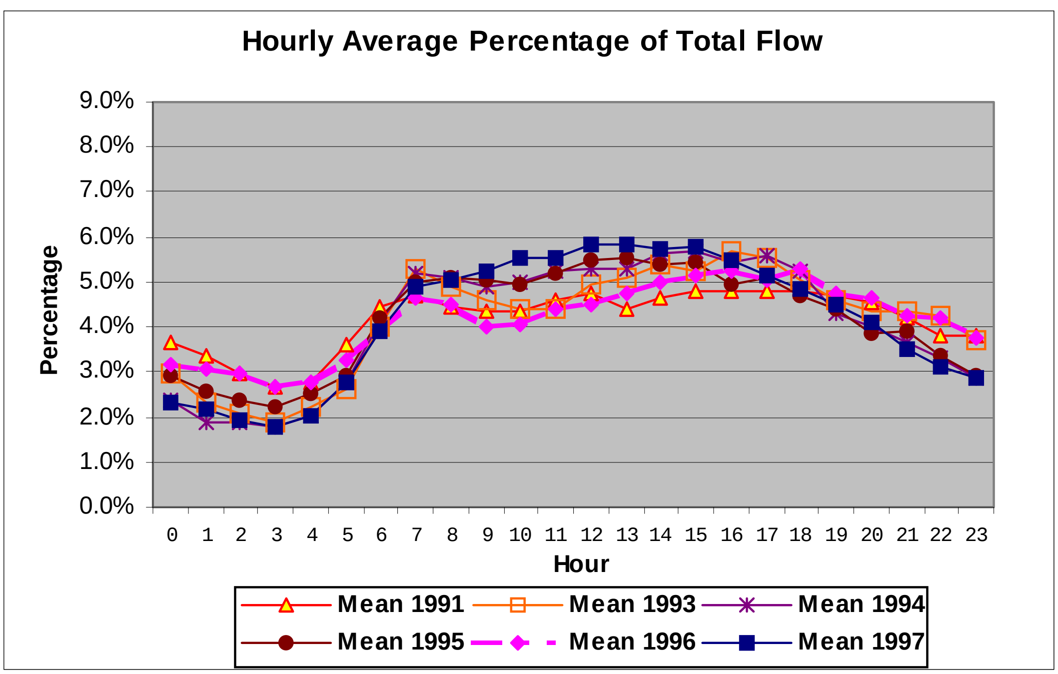


Figure 9

Figure 9 shows the set of curves for the six years: 1991, 1993, 1994, 1995, 1996 and 1997, showing a similar behavior of the variations of 1997 to the others years, particularly in the location of the peak period at 8: 00 to 18: 00 and the more relaxed one from 02: 00 to 04: 00 hours.

## 4. EXPLORATORY GEOGRAPHICAL ISSUES.

Data of the 94 survey stations corresponding to 1991-1997 period were used to create an appropriated database that could be analyzed in a Geographic Information System.

Traffic variables included for each geo-referenced station were: the Average Annual Daily Traffic (AADT), the proportions of the five most common types of vehicles within the AADT, the type of license (for hire or private freight transport) and whether the survey station was on a toll or on a free road, among other variables.

For each of these variables different map views were produced in order to identify any geographical pattern.

### 4. 1 Volumes of Vehicular Traffic.

Observed volumes of the average annual daily traffic (AADT) ranged from 58 to 12261 freight vehicles; this range was equally divided in 9 class intervals of width 1356 vehicles.

Figure 10 shows 92 of the 94 stations (97.8%) having AADT values under 5481 freight vehicles. The two remaining survey stations (orange and gray dots) were installed near the “Tepotzotlán” toll, booth in 1991 (6376 AADT) and 1997 (12261 AADT).

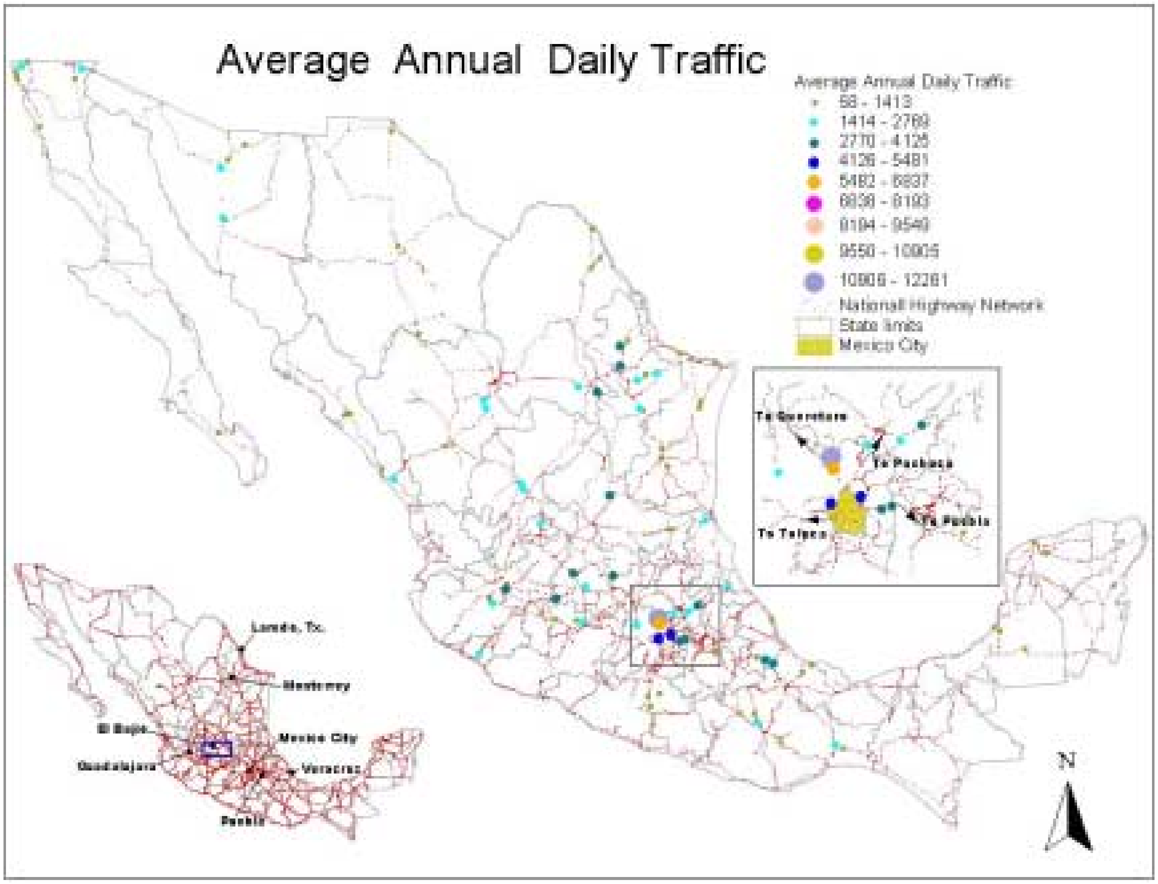


**Figure 10**

Both sites were located in the common exit from Mexico City Metropolitan Zone towards the main road gateway entering the United States (Laredo, Texas) as well as towards important economic regions inside the country: "El Bajío", Guadalajara and Monterrey. [1]

Figure 10 also shows other major traffic volumes correspond to survey stations also located in the surroundings of Mexico City Metropolitan Zone.

Another interesting feature is the trace of the blue-gray dots, corresponding to corridors in the same AADT range (with around 3500 freight vehicles). In fact, these stations are located in corridors communicating the Mexico City Metropolitan Zone [2] with major destinations above mentioned (Laredo Tx., Guadalajara, Monterrey, Puebla and Veracruz).

[1] The difference between the traffic volume of the two survey stations located in Tepotzotlán can be explained by the evolution of freight traffic suffered between these two years, considering the effect of NAFTA, which started in 1994. Although it could also be influenced by seasonal factors because the first station was installed in the month of June while the second station was installed in November.

[2] Mexico City Metropolitan Zone has a population of more than 18 million of people.

A fourth group of survey stations can be identified with light-blue dots (with around 2000 freight vehicles). These stations are located near cities which have a second level of economic and demographic importance, such as interior cities like Oaxaca and Torreon, population centers located on the north border like Tijuana or located near maritime ports like Tampico, Mazatlan and Manzanillo.

Survey stations presenting the minor AADT appear to be close and/or related to a fifth hierarchical level of cities, located in less densely inhabited regions like Baja California Peninsula, north region (Chihuahua), south and southeast of Mexico (Guerrero and Yucatan).

The freight AADT patterns at survey stations can be explained by the economic and demographic importance of the cities nearby stations. Therefore the largest traffic volumes observed arouse close to the Mexico City Metropolitan Zone, decreasing to small values at low population communities.

### 4.2 Percentage of "for hire" and "private" freight transport included in freight AADT.

The freight vehicles classification was based on their license plates according to Mexican Official Norm NOM-001-SCT-2-1994 that establishes the alpha-numeric characters for every motor vehicle and trailers according to its permitted use.

Figure 11 shows that in the Southeast, "El Bajío", Michoacan and Jalisco truck traffics show a tendency towards the use of a relatively higher proportion of private carriers. On the contrary, the case on the corridors from Mexico City Metropolitan Zone to the north of Mexico, exhibit a prevalence of traffic of "for hire " services.

Other patterns or relationships with actual regional economic activities can be disclosed with further geographical analysis.

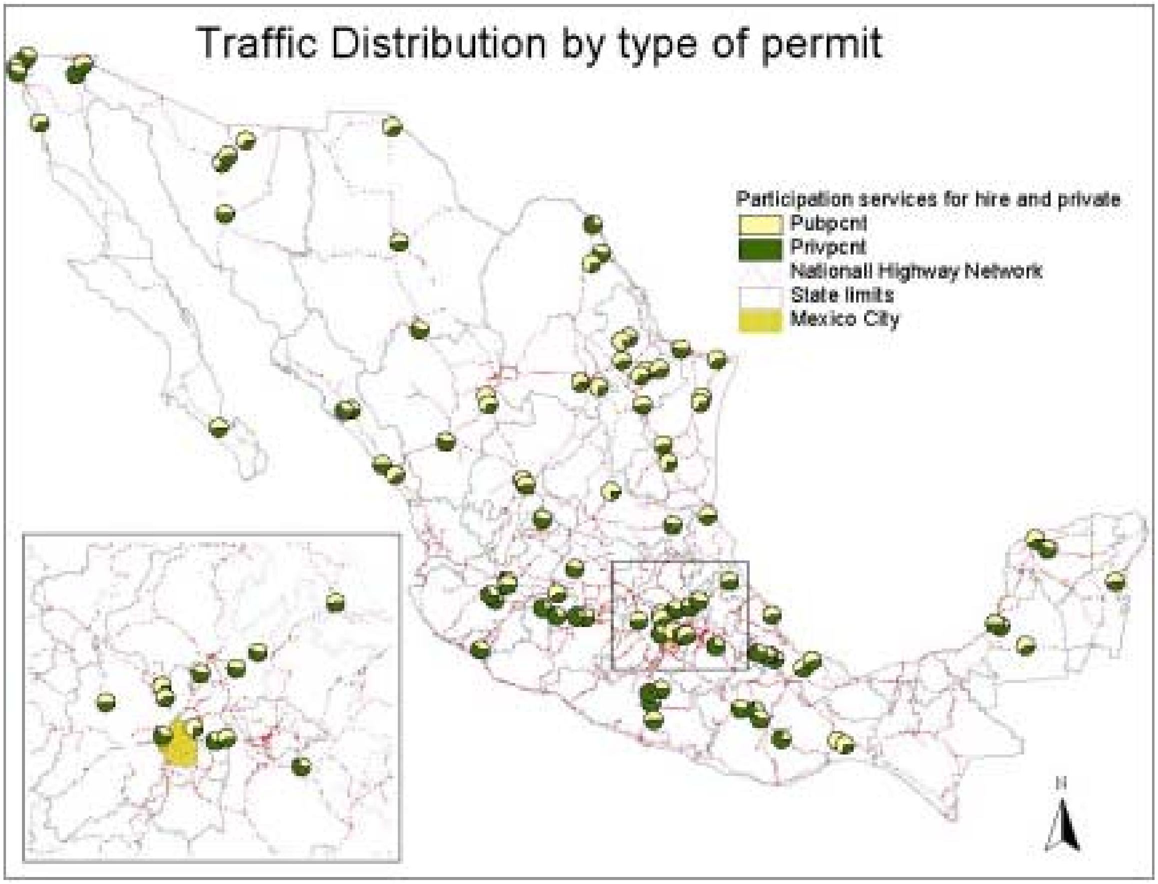


**Figure 11**

### 4.3 Traffic volumes not classified by type of permit.

A certain amount of vehicles could not be classified in every survey station. This situation arises because of the lack of license plates, the incorrect registration of their characters or the fact that vehicles had foreign, mainly U.S., license plates.

The percentage of vehicles that could not be classified for each station varies between 1% and 57%. 60% of survey stations presented a proportion of vehicles that could not be classified inferior to the 10%, while the 97% of the stations presented a proportion of less than 21% of vehicles without classifying.


**Figure 12**

In the map of traffic volumes not classified by type of permit, shown in figure 12, it is possible to observe that the remaining 3% (more than 21% of nonclassified vehicles) corresponds to three stations that could be taken as isolated cases. The first case was in Durango, with a range of vehicles without classifying between the 21% and the 30% while the two remaining cases were near Tijuana, which had 41% and 57% vehicles without classifying regarding to the total of the freight vehicles counting.

Undoubtedly, among the reasons for so high figures of unclassified vehicles; there is the significant presence of U.S. plate vehicles in the Mexican north border zone, where free circulation of U.S. vehicles is allowed on a temporary basis. Border cities population and companies take advantage of this allowance to buy light trucks at lower prices, because of they are available in the U.S.

In addition to this most probable reason, Tijuana is one of the Mexican border cities with the largest number of "maquiladora " plants, which carry into and out of their facilities a vast number of daily shipments from and to U.S. This activity is probably carried out mainly by U.S. licensed vehicles.

Other reasons that stand out are the common use of U.S. trailers for international freight transportation between our two countries and special federal government periodically has granted to farmers to temporally import used light trucks for farming activities.

**4.4. Geographical distribution of types of vehicle.**

Five type of freight vehicles are the most frequently used in the Mexican main road network, according to the surveys discussed in this paper, by number of axles.

- Two axles (C2). This broad configuration, which includes light pick-ups as well as 17.5 ton Gross Vehicle Weight (GVW) vehicles, had a participation that varied between 11.5 and 80% with regard to the total number of vehicles surveyed in each site. Figure 13 shows that C2 participation is bigger in short routes that in long ones. The highest levels of participation happened in the surroundings of Mexico City except the stations that are in the highway to Laredo, which is the common access to the corridors toward the north and west of the Country. Following Mexico City surroundings, the major participation of this configuration happened along the Gulf of Mexico and in the Peninsula of Yucatán. On the other hand, in the north and northeast the percentage of this type of vehicle is smaller.
- Three axles (C3). This configuration had a participation between 5 and 40%, however, in most of the survey stations its participation varies between 11 and 25%. The stations with longer participation of this type of vehicle were located in the "Bajío" and west of Mexico, where a lot of farming activities takes place, while the stations with minor participation of this type of vehicle were located in the north border, mainly in the northwest.

- Three axles tractor with a two axles semitrailer (T3S2). This configuration varies in its participation between 3 and 74%, although in almost the whole country their values of participation are between the 17 and the 32%. In the north of the country these values reached up to 45% while the stations in which they had a major participation were in the northeast, particularly between Saltillo and Nuevo Laredo (neighbor city of Laredo, Tx).
- Three axles tractor with a three axles semitrailer (T3S3). Their participation varies between 0 and 42%. A minor participation of 17% is had in the “Bajío” as well as in the states of Oaxaca and Guerrero and on the roads between Durango and Cd. Juarez and between Mexico City and Tampico as shown in figure 14. In the Pacific, the Peninsula of Yucatán and the Isthmus of Tehuantepec their participation reaches 25% while the stations with major participation are in the more industrious northeast of Mexico.
- Full articulated truck (tractor with a semitrailer and a trailer T3S2R4). This type of vehicle has a participation between 0 and 11%. The smallest participation is in the Occident of the country and Oaxaca. On other hand, larger than 5% values happened in isolated cases like Monterrey - Reynosa and in the Cuernavaca - Acapulco toll highways. However, in both cases this could be due to a very low absolute AADT in these survey stations, varying 96 and 109 freight vehicles.

Finally, the maps of percentages of vehicle type C2 and T3S3 (figures 13 and 14) show that the longer the routes get, the bigger the vehicles used on these routes. Besides, the routes towards the United States border use bigger vehicles than routes towards the Southeast of Mexico.

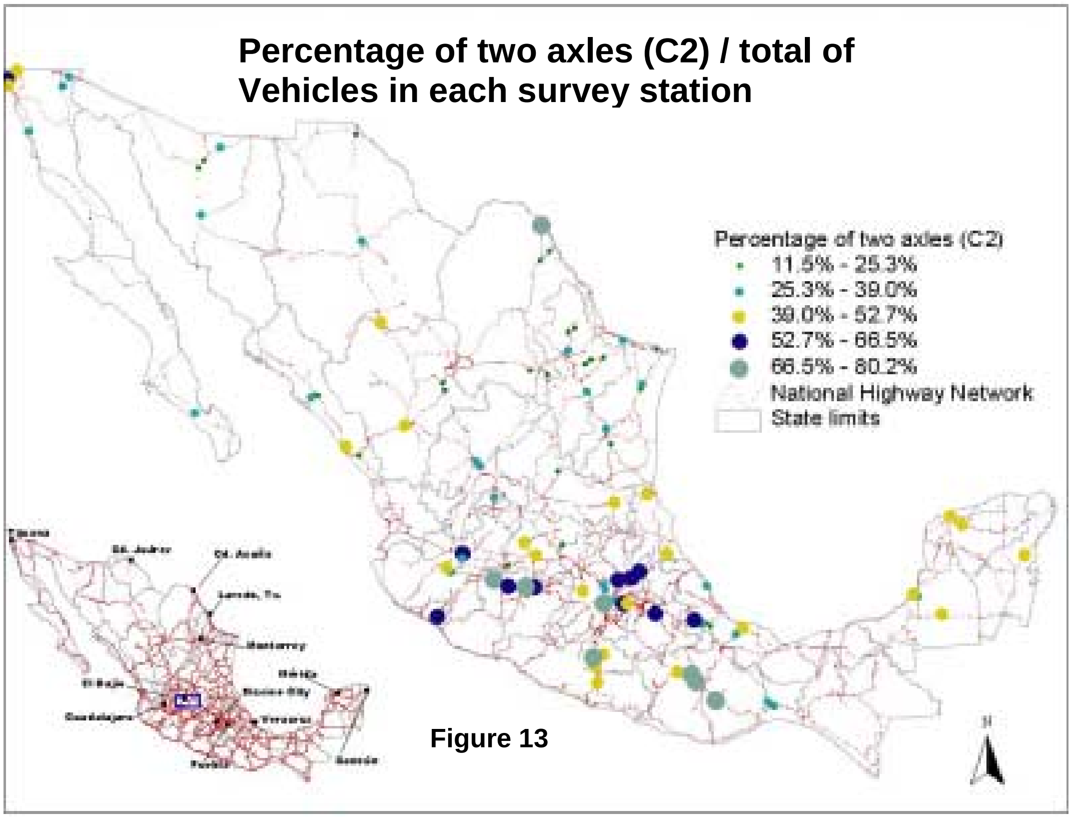


Figure 13

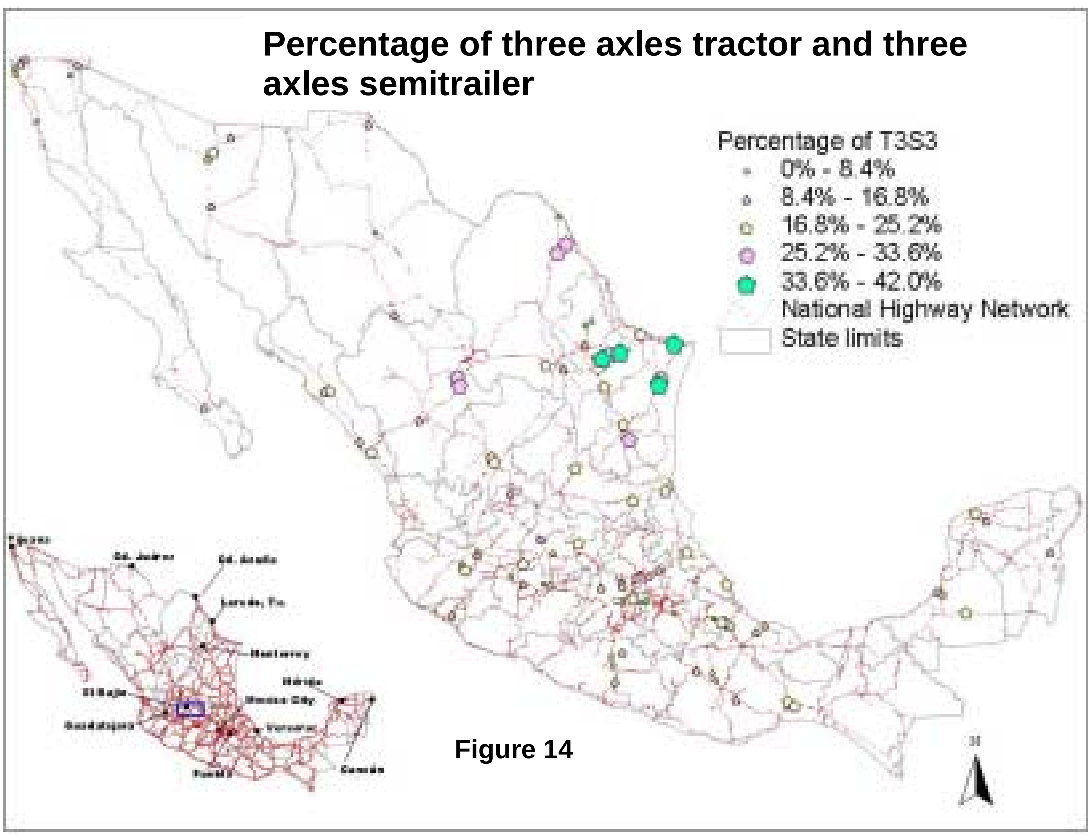


Figure 14

**4.5. Volumes of traffic according the existence of toll on the road use.**

Some of the survey stations were located simultaneously in highways that communicated the same O-D's , with the difference that in a case was on a free highway and the other was on a toll highway.

Figure 15 identifies groups of highways that communicate an origin with the same destination. As it can be observed in this table, a great difference in traffic freight volumes is presented in both types of highways.

The stations that were located on a free highway presented a much larger AADT than those located in the alternative toll highways. The traffic volume in survey stations located in free highways came to be in some cases 20 times larger than the volume in stations located on toll highways.

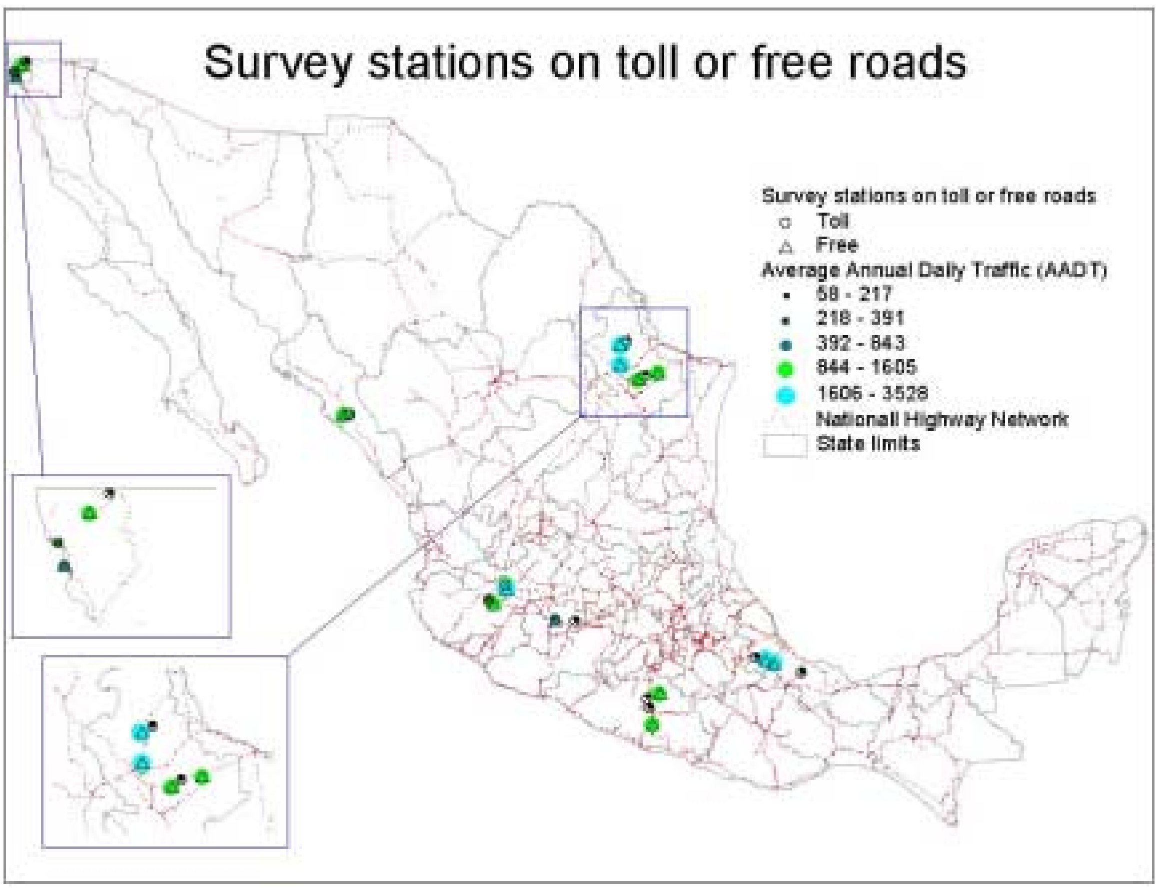


**Figure 15**

## 5. CONCLUSIONS

The exploratory approach used in the statistical analysis of the data suggests an iterative process that could be repeated annually with the information gathered by the field surveys. In this way, having the complete data of 1991 to 1998, the exploratory analysis could be made with the 1991-1997 series, leaving the 1998 data for confirming the tendencies and detected patterns.

In the geographical analysis, the addition of new series of data to the GIS databases will permit a progressively wide covering of Mexico's territory, allowing to confirm or reform the patterns detected in the first analysis.

Based on the previous analysis the following conclusions are presented:

1. In near the 95% of the stations of survey, the Average Annual Daily Traffic is up to 4000 vehicles, representing a mean intensity of flow up to 2.78 vehicles per minute. As for the maxima of the hourly average counting, approximately 95% of the stations, present values up to 250 vehicles, implying so mean hourly intensities of up to 4.17 vehicles per minute. The lineal relationship found between average counting and maximum of the hourly average counting suggests that the latter is approximately a 5.77% of the former. This information gives an idea of the necessary resources for the rising of the surveys.

2. The hourly average curves for the different years, months and days, do not show notable differences in behavior. However, these curves suggest that the period of high intensities in the day are of 8: 00 to 18: 00 hours, with values of the 4% to the 6% of the total flow, and the period more relaxed is of 23: 00 to 05: 00 of the following day, with flows below the 4% of the total.

3. As a particular case stands out the station "Tepotzotlán," with data very far from the rest of the sample. Being an "atypical" station, it should be treated separately of the rest of the sample. This situation is explained by Tepotzotlán position in a common exit from Mexico City Metropolitan Zone to important corridors

4. The classification between free and toll roads with same O-Ds shows a larger freight volume in free roads than in toll roads.

5. The vehicular classification by type of permit shows a geographic specialization such that private freight vehicles are more utilized in southeast and central zones while “for hire” vehicles are more utilized in the most important corridors from Mexico City Metropolitan Zone to the north of the country. The classification also shows an important volume or unclassified vehicles. The largest volumes of these vehicles are in cities near the northwest border. Despite this, it is necessary to improve the capture and classification of license vehicular plates, after analyzing the type of vehicle and model, among other characteristics related to freight vehicles.

## REFERENCES

1. Rascon, O., Barousse, M. and Ventura G., *Standards and Statistical Analysis of Live-Loads at Bridges in Mexico (in Spanish)*. Technical Publication No. 97, Mexican Transportation Institute, Queretaro, Mexico, 1997.
2. Mendoza, A., Gutierrez, J.L. and Dontchev, P. *Statistical Field Study of Domestic Road Transportation. Survey Stations installed during 1995 and 1996 (in Spanish)*. Technical Document No. 20, Mexican Transportation Institute, Queretaro, Mexico, 1999.