Optimal Design of Traffic Counts

YEHUDA J. GUR AND IRIT HOCHERMAN

The central traffic counting program in Israel covers the major rural highway network. Week-long mechanical counts are performed on each link every 1 to 2 years. This paper describes a study of possible ways to increase the program's efficiency by counting only a sample of links every year. Two possible methods for updating link volume estimates based on a sample of counts are described: (a) using network connectivity information and (b) using older counts for estimating the rate of change of volumes over time. The methods have been evaluated using actual count data.

In almost every country there is an agency that operates a system for the collection and processing of traffic-count data on a permanent basis. The main purpose of such systems is to provide an overall, reliable picture of the patterns of motor-vehicle travel and traffic load on the road network, and their changes in time.

There are two main uses to traffic-count data:

- On the demand side, they provide information for the description of the amount of vehicular travel, including growth trends and distribution by type of road and location. These data are used for transportation policy decisions and strategic planning. Traffic counts may also assist in the estimation of origin-destination tables.
- From the supply side, estimates of volumes on specific road sections serve as a basis for maintenance planning, for measuring exposure (in accident analysis), and sometimes as input for detailed planning.

The permanent count program also provides adjustment factors that are used for the conversion of traffic counts into volume estimates. Examples include such factors as corrections for seasonal effects, and the conversion of average annual daily volume (AADT) into design hourly volume (DHV).

In Israel, the Central Bureau of Statistics (CBS) performs the systematic traffic counts on the rural road network. About 800 road sections are counted for 1 week every 1 or 2 years by mechanical counters. Some manual classified counts are also performed to determine the percent of trucks in traffic. In addition, three fixed counting points are located on major arterial roads. The counting program is run quite efficiently by two full-time field crews and a small office staff.

Urban road sections are not included in the program. Various local counting programs are maintained by individual municipalities, but they do not provide a satisfactory basis for a comprehensive picture of traffic trends.

The use of the CBS count data is limited to the supply aspects. The demand side uses are hampered by the absence of counts on urban roads. There is no information on the split of travel between urban and rural roads in the different regions and the way it changes with time. Thus it is not possible to estimate the total amount of travel in the system or the geographic distribution of traffic loads. Moreover, the definition of a road as urban or interurban is sometimes arbitrary and certainly subject to change over time.

In a separate study, Gur et al. (*1*) investigated the feasibility of performing traffic counts on the urban network, and concluded that it was technically feasible. The study recommended that urban arterial roads be incorporated into the CBS counting program. It also recommended that the additional road sections would be added to the system without an increase in the permanent traffic counting staff, that is, without a significant change in the number of road sections that are counted every year.

The purpose of the study that is described here is to examine ways to incorporate the arterial urban road network in the count scheme, assuming that the total number of sections counted is fixed and given. The basic strategy used is to reduce the frequency of counting by sampling the sections to be counted every year and by devising a method for estimating the traffic volumes on those sections that were not counted in a specific year.

On the basis of the findings reported by Gur et al. (*1*), it can be assumed that in Israel there are no systematic seasonality effects. That is, the weekly traffic counts are random drawings from a distribution with expectation $VOL(l,n)$, where VOL is the average annual weekly volume of section l during year n.

The major thrust of this work is to develop a method for estimating AADT based on sample counts and an optimal sampling scheme. In other words, we search for an efficient method for estimating the current traffic volume in a specific road section, given its location and physical characteristics, its own traffic counts in previous years, and current traffic counts of other sections.

THE DATA

Three data files from three different sources were used for the study:

- Traffic count data base constructed by the CBS, which included all the counts for the years 1983 to 1986.
- The physical characteristics file, which was constructed as part of the study on the basis of information from the Department of Public Works.

Transportation Research Institute, Technion–Israel Institute of Technology, Haifa, Israel.

- The interurban network file, which described the statewide arterial roads as a connected network, compiled by the Israel Institute for Transportation Planning.

The three files were edited for errors and inconsistencies and concatenated. The resulting file included the following information for each road section: The identification numbers of the section and its two nodes, average weekday traffic volume in each of the years 1982 to 1985 where the section had been counted, and various attributes such as class of road, region, number of lanes, length, and a dichotomous variable denoting proximity to a metropolitan area. This data file was used to develop and test the various methods that were examined in the project.

ESTIMATION OF TRAFFIC VOLUMES BASED ON ROAD CHARACTERISTICS

We have examined the link volume data for significant trends. Most striking is the association between a link's volume and its type of road. There is also an apparent association between the volume and the region. We found no apparent systematic time trend between years, seasons, or months.

Our first attempt at estimating the traffic volume of a section is based on the mentioned associations by a regression model of the annual weekday traffic volumes (AWDT) as a function of the various sections' attributes. The independent attributes are

- Class of road (major arterial, minor arterial, collector, or local*)
- Region (north*, Haifa, center, Jerusalem, south, or Golan)
- Number of lanes (1 or 2)
- Area (1, rural; 2, metropolitan).

Each nonstarred value of the first two attributes is represented by a dummy variable. For example, the dummy variable for the value "major arterial" is 1 if the section belongs to a major arterial road, and 0 otherwise, and it measures the additional volume that is ascribed to the section being a major arterial, relative to a local road. This regression model is basically identical to an analysis of variance (ANOVA) model.

A separate model was calibrated for each of the years 1982 to 1985; the model for 1985 is displayed in Table 1. The coefficients of the models are quite similar for all years; the proportion of variance explained, R^2, ranges between 70 and 74 percent, which may be considered a relatively high explanatory power. At the same time, the models' estimation errors are rather high, with a root mean square error (RMSE) of around 6500 vehicles per day. Such large errors make the model estimates unacceptable, particularly for estimating traffic volumes of individual road sections.

ESTIMATION OF TRAFFIC VOLUMES BASED ON PAST VALUES

Our next attempt at estimating the traffic volume of a section uses the count information on the same section from the prior year. The regression equation for 1985 is as follows:

$$U_n = -0.174 + 1.075 * C_{n-1}$$

where U is the traffic volume for year n and C is the number of cars counted ($R^2 = 0.97$; RMSE = 2.08).

TABLE 1 REGRESSION MODEL FOR AVERAGE TRAFFIC VOLUME IN A ROAD SECTION BY ITS PHYSICAL CHARACTERISTICS, FOR 1985 (1,000 VEH.)

Variable	Beta	S.D.	T-Value	P-Value
Constant	-24.03	1.67	-14.4	0.0001
Area	10.52	1.01	10.4	0.0001
No of Lanes	13.57	1.38	9.8	0.0001
Type of Road:				
Major Arterial	11.95	1.32	9.0	0.0001
Minor Arterial	4.99	0.79	6.4	0.0001
Collector	2.42	0.84	2.9	0.004
Region:				
Haifa	-0.30	0.98	-0.3	0.76
Center	4.16	0.89	4.7	0.001
Jerusalem	-2.52	1.33	-1.9	0.059
South	-1.35	0.94	-1.4	0.155
Golan	-2.68	1.30	-2.1	0.040

$R2 = 0.70$; $RMSE = 6.5$;

This model is much better than the one based on the physical characteristics; it has a much higher explanatory power and the RMSE is much smaller. Ignoring the small constant, the model suggests that 1985 link volumes can be estimated by assuming a uniform 7.5 percent growth in 1984 volumes.

The good fit of the model (or the high correlation coefficient) may result from the huge dispersion of counts in different types of road sections. To test this we have calculated separate correlation coefficients for each combination of region and type of road. All of the correlation coefficients were greater than 0.9, showing that the correlation between counts of the same link in two consecutive years is indeed high.

Correlation coefficients of similar magnitudes are observed also between the counts of 1983 and 1984, and even with a 2-year gap, between 1983 and 1985.

This finding is meaningful. It suggests that a promising method for estimating current link volumes should be based on updating older counts. However, the model's coefficient (i.e., the annual growth rate) might differ between years and must be estimated anew for every year. This issue is discussed in the following sections.

CONNECTIVITY INFORMATION AS AN AID IN UPDATING VOLUME ESTIMATES

The Problem

Standard statistical methods treat links in the system as independent members of a population. In reality, links are connected into a network, and many trips use not one, but a chain of connected links. Hence, it is likely that volumes are

correlated—a volume change in one link is associated with similar changes in related links. We have tried to formulate a procedure that uses network connectivity information to improve the accuracy of link volume estimates. Assume that outdated (say, 1-year-old) volume estimates on all links in the network are given, as well as updated volume estimates (say, based on counts) for some of the links. Our problem is to provide estimates of the updated volumes on all links.

At the outset, it is tempting to try a comprehensive approach—the network can be described as a connected graph, and volumes can be estimated using assignment-type network analysis models. For example, procedures for estimating trip tables based on link volumes (*2*) can use volume estimates on some links to produce consistent flow estimates on other links. However, in closer examination it becomes clear that this approach is not compatible with the problem at hand; such an approach is dependent on information about land use and travel patterns. It might be suitable for planning, but not for providing traffic volume data, which is considered basic data and must be mostly self contained.

Principles of the Approach

A method for updating link volume estimates that is aided by connectivity information has been developed. Basic definitions are presented in Figure 1. Each link is treated separately; consider the analysis of link O, which we will term the link of interest (LOI). Each one of its nodes (1 and 2 in Figure 1) connect to a number of link chains; for example, node 1 connects to chains I and II.

Each chain consists of links that belong to the same route (as defined by the highway authority). Advancing along a chain for a set number of links, we might hit a link with an updated count; for it, we know the relative and absolute volume changes. This information may be used to estimate the volume's change in link O. The process is repeated independently for each of the chains connecting to the two nodes.

We have examined a wide range of procedures for weighting and summing the data on volume changes on each of the chains, in order to get an estimate of the expected changes on the LOI. Factors that we have accounted for include:

- Whether a chain is along the same route as the LOI;
- The distance (measured by the number of intersections) between the counted link and the LOI;
- The volumes on the links along each chain and their stability; and
- The number of chains with and without counts.

It is beyond the scope of this paper to describe in full the various methods.

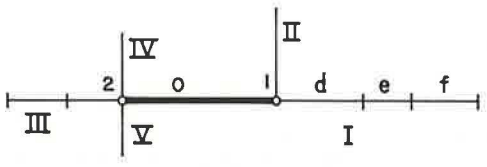

FIGURE 1 Elements of the model.

Testing

The most promising versions of the model are evaluated using the CBS data. The model is used to estimate 1985 link volumes based on 1984 counts for most links and on 1985 counts for a random sample of links. A range of sampling rates is tested. The model's estimates are compared to the available 1985 counts.

The model's estimates of 1985 volumes are also compared to estimates that are obtained by factoring the 1984 volumes by estimated aggregate growth rates in the amount of vehicle miles traveled (VMT). The following growth factors are used:

- FT: A uniform factor for the whole system,
- FC: Different factors by class of road,
- FD: Different factors by district, and
- FCD: Factors by class and district.

Typical results are presented in Table 2. Trends, which have appeared repeatedly in the different experiments, include:

- The model's performance is not superior, and in most cases mostly inferior, to simple factoring.
- Factoring using one universal factor for the whole system has performed best in most cases.
- The worst results have been obtained by use of detailed factoring by both route type and district.

In detailed examination, we have found that the model behaves as expected; that is, the estimated volume change is a reasonable resultant of the changes in the neighboring links. However, the actual counts do not behave this way—large and inconsistent variations in the extent and even direction of the volume changes are apparent.

Conclusions

The attempt to formulate a method for improving the estimates of link volume changes, by using connectivity information has not succeeded—simple estimates of the change, using a gross, system-wide growth factor, prove superior to the estimates of a rather involved model.

The findings in this section reinforce our previous findings, namely that a system-wide uniform growth factor is a powerful and promising means to update link volume estimates.

The methods for the inclusion of the connectivity information are somewhat arbitrary; possibly, there can be found more successful methods. However, considering the fact that we have tried a rather wide range of versions, revolutionary improvements seem to us unlikely.

TABLE 2 NETWORK-BASED ESTIMATION—TEST RESULTS

Nominal Rate (%)	Counts	Estimate's RMS Error (K veh.) Model	FT	FC	FD	FCD
20	49	3.12	2.98	3.52	3.02	3.45
50	126	3.24	2.76	2.85	2.75	2.96
80	201	2.34	1.76	1.87	1.79	2.15

It is possible to speculate on possible reasons for these results:

- In a rather small country like Israel, factors that cause system-wide changes in travel are dominant, compared to local variations. Such factors might include vehicle fleet growth, the economic situation, or amount of tourism.
- Because of the large data errors caused by technical problems, as well as the natural variability in weekly volumes during a year, the counts provide only a rough estimate of the AWDT, with a large random error. It is possible that the benefits incurred from using a large body of data to obtain a reliable estimate of one system-wide growth factor more than compensate for ignoring rather minor detailed trends within the network.

It will be more instructional to refer to the experiment described here as another example that shows that elaboration does not necessarily lead to added accuracy. In some cases, the tendency to complicate models by adding detail (and stratifications) does not improve, and might harm, the model's performance. Even if there is a "behavioral" justification for adding details, their ultimate inclusion must be decided according to the model's performance. The elaboration and precision of a model should be tempered by our limited ability to define and measure relevant relationships.

SAMPLING

The Problem

On the basis of previous findings, we know that the use of one universal systemwide change (growth) factor provides a promising way for updating link volumes over time. We also know that there is a high correlation between counts of the same link in different years. Considering these findings, we examine possible ways to design a traffic counting program that is based on an annual count of a sample of links and uses a system-wide expansion factor.

There are two interrelated design problems: How to estimate the annual volume change and how to select the sample of links each year. In selecting the sample, it is desirable to count each link as frequently as possible; more recent counts provide a better base for volume updates. Frequent counts also correct for irregular, large volume changes that may occur in individual links (e.g., because of increased local activity). In order to achieve the highest possible count frequency for all links, it is necessary to change the links in the sample every year.

The procedure should also provide as reliable an estimate as possible of the annual growth rate. The growth rate can best be estimated by counting the same links in consecutive years. In such a procedure, the estimation error of the growth factor will not be affected by the random variations among the links in the different samples; that is, for efficient estimation of the annual growth rate, it is best to count the same links in consecutive years. (We have not examined other promising methods for estimating aggregate growth, such as using fuel sales or similar external data.)

In this section we examine different ways for estimating the growth factor. If we find that it is indeed necessary to rely on counts of the same links in consecutive years, we will have to find a way to select an optimal rate of replacement of links in the sample. For simplicity of the argument, and considering the counting techniques in Israel, we (a) ignore the existence and role of permanent counting stations, (b) assume a fixed budget and a uniform cost per count, and (c) examine annual sampling rates in the range of 30 to 70 percent.

Link Volume Updating

We consider estimation methods where the estimates of individual link volumes are consistent with the estimates of the amount of travel. The population of links in the system is well defined and known. Volume estimates are to be provided annually and simultaneously for all the links in the system. The proposed estimation procedure is applied as follows.

At the start of any year n, there exists, for each link l in the system, a volume estimate for the last year, $U_{n-1,l}$. During the year a subset of links, C_n is counted; the counts provide current volume estimates, $C_{n,l}$ for these links. On the basis of available information, U and C, a growth factor for the whole system F_n, is calculated. We discuss later how F_n is estimated. Volume estimates for each link are calculated as follows:

$$U_{n,l} = \begin{cases} C_{n,l} & \text{if } l \in C_n \\ U_{n-1,l} * F_n & \text{if } l \notin C_n \end{cases} \tag{1}$$

This method provides volume estimates $U_{n,*}$ for all the links in the system. These estimates are used as a base for estimating the volumes for year $n + 1$, and so on. They also provide the information for calculating the amount of travel in any subset of links by:

$$VMT_{n,S} = \sum_{l \in S} U_{n,l} * L_l \tag{2}$$

where L_l is the length of link l, and S is any selected subset of links.

Estimating the Growth Factor, F_n

The growth factor, F_n, is defined as the annual rate of change in the amount of travel:

$$F_n = VMT_n/VMT_{n-1}$$

or

$$F_n = \sum_l U_{n,l} \Big/ \sum_l U_{n-1,l} \tag{3}$$

Three different methods for estimating F_n have been examined:

1. On the basis of independent VMT estimations. With the counts at year n, the total system's VMT can be estimated by

$$VMT_n = \frac{1}{W_n} * \sum_{l \in C_n} (C_{n,l} * L_l) \tag{4}$$

where

$$W_n = \sum_{l \in C} L_l \Big/ \sum_l L_l \tag{5}$$

W_n is the fraction (by length) of the counted links in year n out of the total link population. From here the estimate of the growth rate by the first method, F_n^a is obtained by Equation 3. In this method, different links are used to estimate VMT_n and VMT_{n-1}. Also, the method is consistent; the VMT estimates using Equations 2 and 4 are identical.

2. On the basis of all $l \in C$. Here, the growth rate is estimated by comparing the current year counts to the previous year volume estimates, for all the counted links:

$$F_n^b = \frac{W_{n-1}}{W_n} * \frac{\sum_{l \in C_n} C_{n,l} * L_l}{\sum_{l \in C_n} U_{n-1,l} * L_l} \tag{6}$$

3. On the basis of only links that are counted in both years. This method is very similar to the previous one, in that F_n^c is estimated by

$$F_n^c = \frac{\sum_{l \in C_{n,n-1}} C_{n,l} * L_l}{\sum_{l \in C_{n,n-1}} C_{n-1,l} * L_l} \tag{7}$$

where $C_{n,n-1}$ contains only links that are counted in both years.

Comparative Evaluation of the Three Estimation Methods

Qualitatively, the three methods differ in the amount and type of the information they use. The first method uses all counts made in the last 2 years. VMT_n and VMT_{n-1} are estimated independently, based on two different sets of links.

The second and third methods provide a direct estimate of F_n. Both use the volumes in consecutive years of one set of links for estimating the rate of volume change, thus exploiting the correlation between consecutive counts of the same link. The second method uses all the counts of year n. It uses the information from counts in earlier years embedded in the estimated volumes U_{n-1}. However, it uses only indirectly the counts in year $n - 1$.

The third method is similar to the second, but it uses only a subset of the links, those that were counted on both years; it also does not take advantage of counts of earlier years. Compared to the second method, its major advantage is that it relies only on actual counts, rather than on volume estimates that are subject to estimation errors.

The three methods differ also in their use of the information. The third method provides direct estimate of the volume growth rate, by examining the change in individual links. In the first method, growth is estimated indirectly, without taking advantage of the correlation in volumes of the same links.

Cochran (*3*) discusses these two estimation methods for the case where the sample sizes in both cases are equal; he shows that the third method is preferable when:

$$R_{n,n-1} > 0.5 * F_n \tag{8}$$

where $R_{n,n-1}$ is the correlation in volumes in the 2 years. In our case, the correlation coefficients are much higher. At the same time, the sample size in the third method is likely to be much smaller. It is equal to the size in the first method only if the same links are counted in both years—an unfeasible case.

The major deficiency in the second method is that it does not consider the quality of the estimate, $U_{l,n-1}$; in particular, the age of the last count. However, if links are counted in a reasonable frequency, and care is taken to count a link whenever a major change in its volume is suspected, then it is likely that the added information will make this method superior in many instances.

The three methods were compared quantitatively, using the count data for the years 1983 to 1985. The 1985 volumes were estimated based on the 1983 volumes for all links, and random samples of counts for 1984 and 1985. Sampling rates of 0.3, 0.5, and 0.7 were tested. Estimated growth rates by the three methods were compared to the 1985 counts. The procedure was repeated a few times, in order to control random effects. Typical results are summarized in Table 3.

The test results show

- The three methods perform well in the high sampling rates.
- The first method is inferior in the medium and low sampling rates.
- The second method, which uses all the link counts, performs best in all sampling rates.
- The third method performs surprisingly well, considering the small number of links in the sample with counts in consecutive years with only about 20 in the 30 percent samples.

Similar results have been obtained in comparing the standard estimation errors of individual links.

Conclusion

Both the qualitative and quantitative analysis indicate that the most effective method for estimating the annual rate of change is by estimating the average VMT change in all the counted links (Equation 6). The method provides satisfactory estimates even at a 30 percent sampling rate. The method was tested using data for two consecutive annual count samples. In actual applications, the expected gap between counts is somewhat longer. However, this difference is not likely to have a critical effect.

For the problem at hand there is no particular need to count the same links in consecutive years. This result is valid only

TABLE 3 ESTIMATED VMT CHANGE RATE (1983–1985)

sampling rate / method	30%	50%	70%	true
1	.99- 1.09	1.18-1.28	1.08-1.12	1.13
2	1.09-1.10	1.06-1.11	1.09-1.12	
3	1.06	1.08-1.12	1.08-1.09	

for relatively high sampling rates where each link is counted every 2 to 4 years.

The qualitative analysis has indicated that links should be counted whenever an irregular change in their volumes is suspected.

SUMMARY AND CONCLUSIONS

We have examined a number of issues that are related to the use of samples of links in an on-going traffic counting program. We have found trends in link volumes, and in particular a high correlation of volumes on a link between different years. By exploiting these trends, it is possible to get reliable estimates of the link volumes for years when they are not counted.

On the basis of the findings of the study, a gross annual sampling rate of about 40 percent is recommended, of which about a quarter of the count is devoted to special cases, such as new roads and areas where significant changes are suspected.

REFERENCES

1. Y. J. Gur, A. S. Hakkert, and I. Hocherman. *Traffic Counting in Israel—Analysis and Evaluation*. Publication no. 185-85. Transportation Research Institute, Technion, Israel, 1985.
2. M. Turnquist and Y. J. Gur. Estimating Link Volumes Based on Link Volume Counts. *Transportation Research Record 750*, TRB, National Research Council, Washington, D.C., 1979.
3. W. G. Cochran. *Sampling Techniques*, John Wiley and Sons, Inc., New York, 1977, Chapter 8.

Publication of this paper sponsored by Committee on Transportation Data and Information Systems.