ACCURACY OF ANNUAL TRAFFIC FLOW ESTIMATES FROM AUTOMATIC COUNTS

by

Garwyn Phillips
(Local Government Operational Research Unit)

The work described in this Report was sponsored by
the Transport and Road Research Laboratory

Any views expressed in this Report are not necessarily those of the
Transport and Road Research Laboratory, nor of the
Department of the Environment or the Department of Transport

Traffic Systems Division
Traffic Engineering Department
Transport and Road Research Laboratory
Crowthorne, Berkshire
1979
ISSN 0305–1315
CONTENTS

Abstract 1

1. Introduction 1

2. Definition of terms 2

3. Theoretical basis 3

  3.1 Model based on factoring by days 4
  3.2 Model using G-factors 5
  3.3 Comparison of models 5

4. Estimating annual flow 7

  4.1 Estimation based on factoring by days 7
  4.2 Estimation using G-factors 9
  4.3 Practical application of results 10

5. Conclusions 11

6. Acknowledgements 13

7. Reference 13

8. Glossary 14

9. Appendix 1: Classification of sites of fifty-point census 20

10. Appendix 2: Sample variance in estimated traffic flow 21

11. Appendix 3: Statistical analysis for factoring by days and using G-factors 23

(C) CROWN COPYRIGHT 1979

Extracts from the text may be reproduced, except for commercial purposes, provided the source is acknowledged
Ownership of the Transport Research Laboratory was transferred from the Department of Transport to a subsidiary of the Transport Research Foundation on 1st April 1996.

This report has been reproduced by permission of the Controller of HMSO. Extracts from the text may be reproduced, except for commercial purposes, provided the source is acknowledged.
ACCURACY OF ANNUAL TRAFFIC FLOW ESTIMATES FROM AUTOMATIC COUNTS

ABSTRACT

Automatic traffic counters are used for on-going studies involving regular monitoring of traffic flows. The counters are rotated regularly among a number of sites throughout the year. Local authorities frequently use this sort of procedure to obtain estimates of annual flow on a network of roads. In this report, several methods of deriving these estimates are examined. Recommendations are given for the most appropriate counting schedules and a simple method of assessing the accuracy of the estimates is provided.

1. INTRODUCTION

Highway Authorities often monitor traffic flows on their road network using a procedure in which automatic traffic counters are moved from site to site according to a pre-determined schedule. A number of decisions must be made when such a counting procedure is designed. In particular, it must be decided how many sites to cover with each counter and how long a counter should be kept at a site. Practical considerations suggest that counters should remain at a site for a week or a multiple of a week. When deciding which particular scheduling procedure is most appropriate, there are two main considerations to be borne in mind:

The purpose of the count, particularly what is to be estimated. This study concentrates on estimating total unclassified annual traffic flow. However, the methods developed might be also applied to other estimates, such as classified annual flow.

The accuracy of the estimate and the efficiency of the procedure. In this study, accuracy is expressed in terms of a deviation to either side of the predicted annual flow within which the true annual flow is most likely to lie. A particular sampling procedure may be considered efficient if it results in an accurate estimate of the annual flow from a relatively short duration of counting. Clearly a balance will be sought between the cost of sampling and the level of accuracy required.

In considering estimation procedures based on automatic traffic counts an important factor to be taken into account is the nature of the sites surveyed. In fact, the difference between sites can be so large as to swamp other causes of variability in predictions. In view of this we considered classifying road sites in order to improve annual flow estimates. A recent study by TRRL\(^1\) identified four broad types of sites, whose seasonal characteristics can be inferred from their descriptive titles. These types are:

1. Urban and/or commuter sites
2. Non-recreational low flow (rural) sites
3. Rural long distance sites
4. Recreational sites.

This classification, while subjective, is of great practical use in that most sites can readily be assigned to one of these types.
The basic source of data for the study was the fifty-point census organised by TRRL for 1975. This consisted of records of traffic flow for each hour in the year at each of fifty sites distributed across the country. These sites classified into the four road types are listed in Appendix 1.

2. DEFINITION OF TERMS

The sampling procedures considered in this report are all of the type where one counter is regularly rotated in a set order among a number of sites so that the duration of each count at each site is constant. For example, where four sites must be covered the counter may be installed at each site for one week per month, or for a continuous period of three months per year. These procedures are illustrated in the following diagram.

The terms used in describing these procedures are defined below.

PROCEDURE A procedure is defined as a set number of sites (N) covered by a counter which is kept at each site for a set duration (R) before being moved to the next site.

YEAR The year is taken as consisting of fifty-two weeks, and the minimum counting period at a site is one week. The first day of the week is taken as a Monday and week number one is the first week containing four days or more of the new year in accordance with British Standard 4760: 1971.

DURATION The duration of a count (R) is the length of time a counter is installed at a site before being moved to the next site. For practical reasons the duration is always a multiple of one week in length and is the same for all sites.

CYCLE The cycle length (L) is defined as the number of weeks taken to complete one count at all sites. At the end of a cycle, counting restarts at the first site. The cycle length in weeks is equal to the product of the number of sites and the duration of count: \( L = NR \).
The schedule number (S) gives the order in which the sites are sampled. For example, consider a procedure where a counter is rotated among four sites, with a duration of one week at each site at a time. The first schedule will consist of counting at the first site on the first, fifth, ninth, etc weeks throughout the year. (If there are N sites covered there will be N schedules.)

In considering automatic counts there are two prime questions of interest. The first is: what is the expected accuracy of the annual flow estimate for each site covered by a counter? To answer this question it is necessary to examine the performance of each counting schedule of a procedure separately. The second question is: how do different procedures compare in terms of accuracy and efficiency? To answer this it is necessary to examine the average performance of all the schedules of each procedure.

To take account of these two aspects of automatic counts the models used to estimate annual flows each have two forms: one refers to a particular schedule of a procedure, and the other to all schedules of a procedure. In these particular models, the accuracy of the annual flow estimate is the same as the accuracy of the estimator used to produce this estimate.

In comparing the two methods, use was made of the concept of coefficient of variation to measure accuracy. The coefficient of variation (C) of the estimator is defined as the square root of the variance (Var) expressed as a percentage of its mean. For example, if the mean value of the estimator is G, then its coefficient of variation is given by:

$$C = \frac{\sqrt{\text{Var}}}{G} \times 100\%$$

The accuracy of the estimator can then be described by means of the following statements:

- Only in about one case in three will the true value of the estimator differ from the mean by more than C per cent.
- Only in about one case in twenty will the true value of the estimator differ from the mean by more than 2C per cent.

3. THEORETICAL BASIS

There are several ways of estimating the annual traffic flow at a site using the sample procedures described in Section 2. In this section two of these are examined in detail with regard to their statistical validity and practical usefulness. They are:

- Estimation based on factoring by days.
- Estimation using specially derived factors.

For both of these, there are two forms of model, according to whether a procedure, or an individual schedule of a procedure is being considered, as mentioned in the previous section.
3.1 Model based on factoring by days

A simple method of estimating the annual flow at a site is to factor up the count at each site by the ratio of the number of days in the year to the number of days sampled. For simplicity, this method will be referred to as factoring by days. The merit of this method is that no specially derived factors are needed. However, it is essential that a balanced schedule of counting is chosen for each site to eliminate the seasonal effect on variation; this means that, for example, a winter count should exactly balance a summer count.

The statistical models used in this report are based on a ratio method in which the estimator is given by the ratio of the annual flow to the sample flow measured by the counter. This ratio method assumes that the distribution of the annual flow estimator is normal under particular conditions (eg for a given type of site). Consequently, a reasonable approximation to the actual estimator at a particular site is its mean value over all sites of that type, and a measure of confidence can be applied to the estimate using the standard deviation of the distribution.

The first form of the model for estimating annual flow at a particular site is as follows. For a given set of counting periods (defined by R, N and S) the annual flow estimator is:

\[ G_i = \frac{365}{D_S} + \alpha_i \]  

where \( G_i = \frac{Y_i}{Z_i} \)

- \( G_i \) = multiplicative factor to annual flow for site i
- \( D_S \) = number of days counted for schedule S
- \( Y_i \) = annual flow at site i
- \( Z_i \) = total counter flow at site i
- \( \alpha_i \) = normally distributed random variable with a zero mean and constant variance.

The second form of the model, used for comparing procedures, is that the annual flow estimator for a given procedure (R,N) is given by:

\[ G_{is} = \frac{365}{D_S} + \alpha_{is} \]  

where \( G_{is} = \frac{Y_{i}}{Z_{is}} \)

- \( G_{is} \) = multiplicative factor to annual flow for site i and schedule s
- \( Z_{is} \) = total counter flow at site i for schedule s
- \( \alpha_{is} \) = normally distributed random variable with a zero mean and constant variance.

It may be noted that the estimator \( 365/D_S \) is the reciprocal of the sample fraction.
3.2 Model using G-factors

The appropriateness of the model described in Section 3.1 depends on the assumptions made about the error terms $\alpha_i$ and $\alpha_{is}$. If the means of these errors are not zero then the estimation of the annual flow based on factoring by days will be biased. In such cases an alternative model can be used which requires the derivation of a multiplicative factor (called a G-factor). The G-factors may be regarded as seasonal adjustment factors. As such they can cater for counting schedules in which an exact seasonal balance of traffic flows is not achieved. However, the method has the practical disadvantage that separate G-factors for each schedule have to be derived.

The first form of the model, used for measuring the accuracy of the predicted annual flow at a particular site, requires the derivation of a G-factor for each schedule (S) of a procedure (R,N). This form of the model is:

$$ G_i = G_{S} + \beta_i $$

(3)

where

- $G_i = Y_i/Z_i$ as before
- $G_S$ = mean multiplicative factor over all sites for schedule S
- $\beta_i$ = normally distributed random variable with a zero mean and constant variance.

When comparing procedures a slightly different form of the model is required which considers all schedules and all sites. For a particular procedure (R,N) the annual flow estimator is given by:

$$ G_{is} = G_{S} + \beta_{is} $$

(4)

where

- $G_{is} = Y_i/Z_{is}$ as before
- $\beta_{is}$ = normally distributed random variable with a zero mean and constant variance.

3.3 Comparison of models

In this section the models described in the previous sections are compared in terms of their statistical validity and practical usefulness.

In the model described in Section 3.2, individual G-factors are derived for each schedule for a particular procedure (equations 3 and 4). Statistical tests showed that it is reasonable to assume that the error term is normal and that it is not correlated to the annual flow. This is true both when individual schedules are considered (error term is $\beta_i$) and when an overall procedure is considered (error term is $\beta_{is}$). This observation was found generally true for various combinations of count lengths and number of sites per counter, provided at least two full cycles of counting were obtained in a year.

The alternative model described in Section 3.1 is based on factoring a count by days (equations 1 and 2). The main difference from the G-factor model is that in this model the values of the error term ($\alpha_i$) for a particular schedule do not always average zero. This occurs if the counting periods chosen do not balance out seasonal variation. However, when considering all the schedules of the procedure together, it can be assumed that the error term ($\alpha_{is}$) is normal.
Clearly, when factoring a count by days it is advisable to choose a procedure which provides a seasonal balance for all sites irrespective of the order of scheduling. An example of such a procedure is where a counter is rotated among four sites with a duration of one week at each site. The error terms $\alpha_i$ and $\sigma_{15}$ are then well described by normal distributions with zero mean, and are uncorrelated with the annual flow.

The conclusion from these observations is that the models based on derived G-factors (equations 3 and 4) and on factoring by days (equations 1 and 2) are both appropriate, though care is needed in applying the latter to individual schedules because of the possibility of bias for unbalanced procedures.

The performance of the two methods was compared in terms of the average performance over all schedules of a procedure. That is, the form of the models compared was that given in equations (2) and (4). Figure 1 shows the comparison for the case of a counter rotated among four rural long distance sites for various durations ($R$). The following observations which can be made from this figure are generally true for the other road types and for different numbers of sites per counter:

The coefficient of variation gradually rises as the duration increases. A large increase occurs when the duration is eight weeks; this corresponds to a procedure in which it is not possible to complete two cycles of counting in one year. The rise is a result of seasonal variation. With at least two counts at a site in a year the seasonal effects at each period balance each other to some extent. Such a balancing cannot occur with just one count in the year.

For counts of less than eight weeks duration, a somewhat higher coefficient of variation results from factoring by days than from using a mean G-factor. This may be expected, since the measure of accuracy of the G-factor model refers only to the thirteen sites sampled, whereas the measure of accuracy for the other model refers to all rural long distance sites. If all such sites in the country are considered, the general coefficient of variation of the G-factor may well be higher than that given in Figure 1. Bearing this in mind, the performance of the other model, based on factoring by days, is good.

The following is a summary of the strengths and weaknesses of each model:

The method based on factoring by days is practical and very easy to use but depends on seasonal effects being eliminated by a balanced schedule of counting at each site. Given this, the method performs quite favourably when compared with the G-factor method.

The method which uses specially derived G-factors performs well in taking seasonal variation and bank holiday distortions into account. However, in order to do so separate G-factors must be derived for each schedule of a procedure and for each road type. It is therefore less simple to use than factoring by days. Also, this method is likely to be less accurate than appears from the analysis when it is applied to other years and to road sites not in the fifty-point census.
4. ESTIMATING ANNUAL FLOW

In Section 3 two methods were found suitable for estimating annual flow using traffic flows recorded by automatic counters: factoring by days and using derived G-factors. In this section, the whole range of practical procedures is examined, using both methods. This examination includes a simple way of assessing the accuracy of these procedures. The results for the most satisfactory procedures are given at the end of the section.

The procedures considered were for cases where up to twenty-six sites were covered by one automatic counter. The maximum duration of counting at each site varied from one week when twenty-six sites were covered to thirteen weeks when two sites were covered.

There are two basic considerations in deciding on a counting procedure. The first is, how many sites should be covered (N); the second is, how long the counter should remain at each site (R). An illustration of the effect of various values of N and R on the accuracy of the annual flow estimate is given in Figure 2. The model used is that based on G-factors (Section 3.2) and the figure is for rural long distance sites, though the following observations are applicable to all road types:

There is a discontinuous jump in the coefficient of variation when the length of the cycle exceeds twenty-six weeks.

For cycle lengths of less than twenty-six weeks there is a gradual increase in the coefficient of variation as either the number of sites (N) or the duration of count (R) increases.

These observations suggest that since \( L = RN \), it is the cycle length (L) which is the important parameter determining how accurate the annual flow prediction will be for a given procedure. The rise in variability when the cycle length exceeds twenty-six weeks can be ascribed to seasonal variation. It is not possible to assess the seasonal nature of traffic flows at a site if only one period of counting occurs at that site. The only method of avoiding seasonal bias is to choose a neutral time of the year (i.e. spring or autumn) for counting. However, in rotating a counter among sites it is impossible for every site to have had the counter at a neutral period. The conclusion is that on no account should an automatic counting procedure be used that does not allow at least two complete counting cycles in the year. Therefore, the following discussion is confined to those procedures that satisfy this requirement.

4.1 Estimation based on factoring by days

An accurate estimation of annual flow depends on selecting a counting procedure that eliminates as much as possible all causes of variation. An examination of the accuracy of various procedures showed that a simple relationship exists in which the variance of the estimator (for a given road type) for a particular procedure and schedule depends on three variables:

First, it depends on the value of the estimator for that schedule, \( 365/D_S \). This is because the larger the value of the mean the larger will be the deviation either side of the mean.
Second, it depends on the sample fraction. Clearly, the more days in the year that are sampled, the more accurate should be the estimate. The variance depends on the proportion $P$, of the number of days not sampled to the total number of days in the year.

Third, it depends on the length of the counting cycle $L$, of the procedure. The shorter the cycle length the more account is taken of seasonal variation.

The relationship for the variance of the error term $a_{is}$ of the estimator (given in equation 2) is given by:

$$\text{Var} (a_{is}) \propto \left( \frac{365}{D_S} \right)^2 P^2 L$$

(5)

The theoretical justification for this relationship is set out in Appendix 2.

Instead of using the variance to express accuracy, a more convenient term is the coefficient of variation as defined in Section 2. From this definition and the above relationship (5) the following equation can be derived:

$$C_S = aP \sqrt{L}$$

(6)

where 'a' = a constant whose value depends on the type of site considered

$C_S$ = the coefficient of variation of the estimator at a site for a given schedule $S$.

This equation allows the accuracy of the estimator for any schedule to be assessed. In fact, the proportional error in the estimator is the same as the proportional error in the annual flow prediction. Hence, the accuracy of the predicted annual flow can also be assessed. In other words, for any regular counting procedure the accuracy of the annual flow estimate at a site is given by equation (6).

In order to compare the overall performance of different procedures the average accuracy of all the schedules of a procedure should be considered. In this case, the average proportion of days not sampled ($P$) is given by $(N-1)/N$. For example, if four sites were covered then at each site an average of three quarters of the year will not be sampled. By analogy with equation (6), the overall coefficient of variation $C$ for all schedules of a procedure would then be:

$$C = a \left( \frac{N-1}{N} \right) \sqrt{L}$$

(7)

For simplicity this equation can be expressed as:

$$C = a \times$$

where $X = \left( \frac{N-1}{N} \right) \sqrt{L}$, that is, the product of the proportion of days in the year not counted and the square root of the cycle length.
The linear relationship between C and X is shown for each road type in Figure 3 and the statistical analysis of this relationship is discussed in Appendix 3. It can be seen from Figure 3 that the linear relationship is good, especially considering that the data points refer to all possible counting procedures (from a procedure where one counter covers two sites for thirteen weeks at a time, to a procedure where one covers twenty-six sites for one week at a time).

The estimated values of the constant term 'a' for the method based on factoring by days for each road type is given in Table 1. It can be seen that the value of 'a' is a good indication of the type of site. As such, this parameter may prove useful in giving an objective measure of site type. There are two reasons why a road type may have a large value of 'a'. The first is that the seasonal variation in the traffic flows may be large (as illustrated by the recreational sites). The second is that the average level of flow at a site may be low (as illustrated by the non-recreational low flow sites).

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>The values of 'a' for two methods of estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factoring by days</strong></td>
<td><strong>Using G-factors</strong></td>
</tr>
<tr>
<td></td>
<td><strong>95% Confidence Interval</strong></td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
<tr>
<td>Urban/commuter</td>
<td>0.61</td>
</tr>
<tr>
<td>Non-recreational low flow</td>
<td>1.02</td>
</tr>
<tr>
<td>Rural long distance</td>
<td>1.20</td>
</tr>
<tr>
<td>Recreational</td>
<td>2.72</td>
</tr>
</tbody>
</table>

It may be noted from Figure 3 that there is some tendency for the coefficient of variation for large values of X to be higher than expected from a linear relationship. It would be interesting to examine the reasons for this departure from linearity by examining traffic data for other years. However, in practical terms the effect of non-linearity can be neglected except for recreational sites with a counter covering more than twenty sites in a year.

4.2 Estimation using G-factors

The use of G-factors in estimating annual flow may be regarded as a refinement to the method based on factoring by days. This refinement takes account of bias which may be introduced if the seasonal effects for a schedule are not balanced. Because of this the relationships derived in Section 4.1 will serve also for estimation based on G-factors, though the general level of the coefficients of variation will be lower.

The relationship between C and X (based on equation 7) for the case where the annual flow estimator is a G-factor is shown for each road type in Figure 4. The linear relationship is best for the urban/commuter sites while less good for the recreational sites. The statistical analysis of these figures is discussed in Appendix 3, where the best estimate of the slope 'a' is obtained, together with a confidence interval. These are reproduced in Table 1.
As might be expected, the greater the seasonal variation of the road type, the greater the disparity between
the value of 'a' based on factoring by days, and the value based on the use of G-factors. However, the improvement
by using G-factors is likely to be less good in practice than is indicated by Table 1. This is because the seasonal effect
for a particular schedule may vary from year to year, and this has not been taken into account. Therefore, before a
method of estimation based on G-factors can be recommended the analysis should be repeated over a number of years.

4.3 Practical application of results

Estimating annual flow by factoring by days is a straightforward practical method to apply which allows the
accuracy of the estimate to be simply assessed, using equation 7.

The fact that the value of 'a' depends on the road type means that the accuracy of the estimate for a given
procedure is different for the different road types. It may be desirable to arrange the method of counting in such a
way that the accuracy of the estimate of annual flow is the same for all sites surveyed. This may be achieved, using
equation (7), by varying either the duration of the count or the number of sites covered by one counter. The more
practical method is to vary the number of sites per counter for each type of site, with the duration of count at a
site being fixed at a convenient length. It is then necessary to solve the following equation for N:

\[ \frac{N-1}{\sqrt{N}} = \frac{C}{a\sqrt{R}} \]  \hspace{1cm} (8)

where \( N \) = number of sites of a particular road type covered by the counter

\( a \) = appropriate constant for the particular road type

The number of sites per counter for each road type needed to achieve a constant overall level of accuracy is
given in Table 2 for various durations of count. If a counter is rotated regularly among the number of sites stated
then the coefficient of variation in the estimate of annual flow will be 2 per cent. It must be remembered that the
accuracy achieved is an average for all schedules of the procedure. A well balanced choice of procedure is required
to ensure that the accuracy of each schedule is similar.

<table>
<thead>
<tr>
<th>Duration of count R</th>
<th>Maximum number of sites per counter (N) to achieve a coefficient of variation of 2 per cent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban/Commuter</td>
</tr>
<tr>
<td>1 week</td>
<td>13</td>
</tr>
<tr>
<td>2 weeks</td>
<td>7</td>
</tr>
<tr>
<td>3 weeks</td>
<td>5</td>
</tr>
<tr>
<td>4 weeks</td>
<td>4</td>
</tr>
</tbody>
</table>

TABLE 2
Counting procedures to give a uniform level of accuracy
(Based on factoring by days)
It may be noted that in practice the value of \( N \) must be integer and, therefore, a simple practical approximation to the above equation is given by:

\[
N = \frac{1}{R} \left( \frac{C}{a} \right)^2 + 2
\]

........................... (9)

This equation generally works well but may be misleading for occasions where the solution for \( N \) is 2. In such cases it is advisable to check with the first, more accurate formulation of equation (8).

From Table 2 a programme of counting for all sites under consideration can be devised. Similar tables can be constructed for other levels of accuracy by solving equation (8). It is preferable to use one counter for one type of site as far as possible.

For example, if counting for a duration of one week at each site is practical then a cross section of twenty-six sites can be surveyed using four counters which will result in an accuracy of 2 per cent as measured by the coefficient of variation. The first counter would cover thirteen urban/commuter sites, the second six non-recreational low flow sites, the third five rural long distance sites and the fourth two recreational sites. Such an allocation makes the most efficient use of the counters from a statistical point of view, though practical considerations may require a modification of the ideal.

As another example, consider the problem of how many sites may be covered by a single automatic counter in order to achieve an accuracy (as expressed by the coefficient of variation) of 5 per cent in the annual flow prediction. Suppose that the sites are all of the rural long distance type and the minimum practical duration of counting at a site is a fortnight.

For this problem, equation (9) is solved using the following values: \( R = 2 \) weeks, \( C = 5 \) per cent, \( a = 1.20 \). The solution is that if an automatic counter is regularly rotated every fortnight, among rural long distance sites, the number of sites required to achieve an average coefficient of variation of 5 per cent will be ten. The estimate of the annual flow is obtained by factoring the count at each site according to the number of days sampled.

5. CONCLUSIONS

This report has examined procedures in which an automatic traffic counter is regularly rotated among a given number of sites throughout the year. The duration of count was kept the same and was always a week or multiple of a week. Two aspects were of prime interest: first, the accuracy of the annual flow estimate at a particular site (which depends on the particular schedule chosen); second, the comparative performance of different procedures (averaged over all schedules).

Two methods of estimation were examined. The first factored up the count at each site by the ratio of the number of days in the year to the number of days sampled (called factoring by days). The second used multiplicative factors (called G-factors), derived for each schedule of each procedure considered. The following results were obtained:
The accuracy of the estimate of annual flow depends on the extent to which the sampling procedure eliminates causes of variation, notably those due to seasonal effects. For this reason, reasonable estimates from either method were only obtained when every site was sampled at least twice a year. In other words, it is necessary for the cycle length not to exceed twenty-six weeks.

The accuracy of the annual flow estimate at a site depends on two variables: the cycle length, and the sample fraction. The relationship is a simple one in which the coefficient of variation $C_S$ is given by:

$$C_S = aP\sqrt{L \text{ per cent}}$$ .......................... (6)

$P$ = proportion of the number of days not sampled at a site to the total number of days in the year.

$L$ = cycle length, which is equal to the product of the number of sites $N$ covered by the counter and the duration $R$ in weeks that the counter is placed at any one site ($L = RN$).

The performance of different procedures may be compared by making use of another simple relationship. This is one in which the coefficient of variation $C$ depends on the number of sites per counter and the duration of the count:

$$C = a\left(\frac{N-1}{N}\right)^{1/2} L \text{ per cent}$$ .......................... (7)

An interesting parameter of both equations (6) and (7) is the constant of proportionality ‘$a$’. The value of this term depends on the type of road considered. The more seasonal the type of site and the lower the average traffic flow the higher is the value of ‘$a$’. As a result, this term may be of considerable value as an objective measure of road type.

The value of the constant term ‘$a$’ for each road type when factoring by days is as follows:

- Urban/commuter: 0.61
- Non-recreational low flow: 1.02
- Rural long distance: 1.20
- Recreational: 2.72

For the best procedure to achieve a given level of accuracy, the duration of count should be the smallest integral number of weeks that is practicably convenient. The number of sites that may be covered by one counter to achieve a given level of accuracy can be found by using equation (9).

$$N = \frac{1}{R} \left(\frac{C}{a}\right)^2 + 2$$ .......................... (9)

Factoring by days provides a reliable and accurate method of estimating the annual flow and is also very easy to apply. Using derived G-factors appears to provide even more accurate estimates of the annual flow. However, since the G-factor estimates were derived from data from one year covering a relatively small number of
sites the accuracy is likely to be less good than at first appears, primarily because seasonal effects change from year to year. In addition it is less easy to make use of G-factors, since tables would be required for each road type and each schedule of a procedure.

The above results indicate that further work is desirable to study the performance of G-factors in other years in order to establish whether a real improvement on the method of factoring by days can be achieved.

In the meantime, the annual flow should be estimated by factoring by days based on a schedule of counting that is as seasonally balanced as possible. The recommendations for using an automatic traffic counter to monitor annual traffic flows are therefore as follows:

To estimate the annual flow at a road site using an automatic counter, a regular rotation should be used with at least two counts occurring at each site in a year.

The recommended method of estimation is to factor up the count by the ratio of the number of days in the year to the number of days sampled, i.e.

\[
\text{Estimate of annual flow} = \frac{365}{\text{No. of days sampled}} \times \text{Total measured flow}
\]

To achieve comparable accuracy at each site it is important to obtain a seasonal balance of counting (for example, two counts in winter balancing two counts in summer).

To decide on the most appropriate procedure the following approach is recommended. First, decide on the minimum duration of count that is practically convenient. (This must be a fixed multiple of a week.) Second, decide on the desired level of accuracy, as expressed in terms of the coefficient of variation. The maximum number of sites \( N \) of a particular road type that can be covered to achieve the required level of accuracy is then indicated by equation (9).

6. ACKNOWLEDGEMENTS

This study, undertaken by the Local Government Operational Research Unit, was commissioned by Traffic Systems Division of the Traffic Engineering Department of TRRL. The author is particularly grateful to Mr D H Mathews for the assistance and advice provided in the course of the study.

7. REFERENCE

### 8. GLOSSARY

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>A constant the value of which depends on the type of site.</td>
</tr>
<tr>
<td>(C)</td>
<td>Coefficient of variation of estimator: defined as the standard deviation expressed as a percentage of the mean.</td>
</tr>
<tr>
<td>(D_S)</td>
<td>Number of days for schedule (S).</td>
</tr>
<tr>
<td>(G_S)</td>
<td>Mean (G)-factor for all sites for schedule (S).</td>
</tr>
<tr>
<td>(G_i)</td>
<td>The actual (G)-factor for site (i).</td>
</tr>
<tr>
<td>(G_{is})</td>
<td>The actual (G)-factor for site (i) and schedule (s).</td>
</tr>
<tr>
<td>(i)</td>
<td>Denotes the number of the site for a given site type.</td>
</tr>
<tr>
<td>(L)</td>
<td>The cycle length in weeks for the procedure considered (as given by (N) and (R)).</td>
</tr>
<tr>
<td>(N)</td>
<td>The number of sites covered by a single counter.</td>
</tr>
<tr>
<td>(P)</td>
<td>Proportion of the number of days not sampled at a site to the total number of days in the year.</td>
</tr>
<tr>
<td>(R)</td>
<td>Duration in weeks that a counter is installed at a site before being moved to the next site.</td>
</tr>
<tr>
<td>(S)</td>
<td>The schedule of the procedure which gives the order in which the (N) sites are surveyed.</td>
</tr>
<tr>
<td>(s)</td>
<td>Denotes the number of the schedule.</td>
</tr>
<tr>
<td>(X)</td>
<td>Equals the product of the proportion of days in the year not counted at a site and the square root of the cycle length of the procedure considered.</td>
</tr>
<tr>
<td>(Y_i)</td>
<td>The total annual flow at site (i).</td>
</tr>
<tr>
<td>(Z_i)</td>
<td>Total traffic flow recorded by counter at site (i).</td>
</tr>
<tr>
<td>(Z_{is})</td>
<td>Total traffic flow recorded by counter at site (i) for schedule (s).</td>
</tr>
<tr>
<td>(\alpha_i, \alpha_{is})</td>
<td>Error terms assumed to be normally distributed with zero mean and constant variance.</td>
</tr>
<tr>
<td>(\beta_i, \beta_{is})</td>
<td></td>
</tr>
<tr>
<td>(\text{Var})</td>
<td>Variance: defined as the mean squared deviation of a variable from its mean.</td>
</tr>
</tbody>
</table>
Fig. 1 STABILITY OF ESTIMATORS BASED ON G-FACTORS AND FACTORING BY DAYS

Fig. 2 EFFECT ON ACCURACY OF NUMBER OF SITES AND DURATION
Fig. 3 ACCURACY OF COUNTING PROCEDURES USING G-FACTORS
Fig. 3 (cont) ACCURACY OF COUNTING PROCEDURES USING G-FACTORS

(c) RURAL LONG DISTANCE SITES

(d) RECREATIONAL SITES
Fig. 4 ACCURACY OF COUNTING PROCEDURES - FACTORING BY DAYS

(a) URBAN/COMMUTER SITES

(b) NON RECREATIONAL LOW FLOW SITES
Fig. 4 (cont) ACCURACY OF COUNTING PROCEDURES - FACTORING BY DAYS

(c) RURAL LONG DISTANCE SITES

(d) RECREATIONAL SITES
9. APPENDIX 1
Classification of sites of fifty-point census

This appendix gives the location of the fifty road sites grouped according to the type of site they represent.

<table>
<thead>
<tr>
<th>URBAN/COMMUTER (14 sites)</th>
<th>NON-RECREATIONAL LOW FLOW (8 sites)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Site No.</strong></td>
<td><strong>Road No.</strong></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>A602</td>
</tr>
<tr>
<td>2</td>
<td>C5</td>
</tr>
<tr>
<td>3</td>
<td>A312</td>
</tr>
<tr>
<td>8</td>
<td>B2056</td>
</tr>
<tr>
<td>9</td>
<td>B4012</td>
</tr>
<tr>
<td>11</td>
<td>C156</td>
</tr>
<tr>
<td>14</td>
<td>12a</td>
</tr>
<tr>
<td>15</td>
<td>A1</td>
</tr>
<tr>
<td>19</td>
<td>A58</td>
</tr>
<tr>
<td>22</td>
<td>B5251</td>
</tr>
<tr>
<td>25</td>
<td>A43</td>
</tr>
<tr>
<td>26</td>
<td>A614</td>
</tr>
<tr>
<td>39</td>
<td>A525</td>
</tr>
<tr>
<td>48</td>
<td>A727</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RURAL LONG DISTANCE (13 sites)</th>
<th>RECREATIONAL (11 sites)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Site No.</strong></td>
<td><strong>Road No.</strong></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>4</td>
<td>A3</td>
</tr>
<tr>
<td>6</td>
<td>A272</td>
</tr>
<tr>
<td>7</td>
<td>A35</td>
</tr>
<tr>
<td>10</td>
<td>A45</td>
</tr>
<tr>
<td>18</td>
<td>A59</td>
</tr>
<tr>
<td>21</td>
<td>A595</td>
</tr>
<tr>
<td>23</td>
<td>A621</td>
</tr>
<tr>
<td>27</td>
<td>A49</td>
</tr>
<tr>
<td>28</td>
<td>A454</td>
</tr>
<tr>
<td>30</td>
<td>A44</td>
</tr>
<tr>
<td>42</td>
<td>A96</td>
</tr>
<tr>
<td>43</td>
<td>A1</td>
</tr>
<tr>
<td>45</td>
<td>A92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MISCELLANEOUS (4 sites) — very low traffic flows</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Site No.</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>49</td>
</tr>
<tr>
<td>50</td>
</tr>
</tbody>
</table>
10. APPENDIX 2

Sample variance in estimated traffic flow

The automatic traffic count is mounted at a site for \( n \) full weeks of the year. This sample count is then used to estimate the annual traffic flow \( Y_i \).

Consider first the estimation of the average annual weekly traffic flow (AAWT). For each site in a particular year the weekly traffic flow will have a distribution with mean \( \mu_i \) and variance \( \sigma_i^2 \).

Consider the case where the weeks are randomly sampled without replacement; that is \( n \) different weeks are selected from the total of fifty-two weeks in the year. Then if \( (w_1, \ldots, w_n) \) are the sample flows the mean flow \( w \) is an estimate of AAWT, such that:

\[
E(w) = \mu_i
\]

and

\[
\sigma^2(w) = (1-f) \frac{\sigma_i^2}{n}
\]

where \( f \) is the sample fraction.

The actual method of sampling differs from simple random in two important respects.

First, the actual method of sampling is systematic rather than random, in that for a particular schedule we simply go through the year and count traffic for every \( N \)th week where \( N \) is the total number of schedules in the year. Furthermore, each schedule that may be selected consists of a set of weeks, and these sets are mutually exclusive. As a result the last schedule is dependent on the others and the variance of the estimated AAWT is given by:

\[
\sigma^2(w) = \frac{(N-1)}{N} \cdot (1-f) \cdot \frac{\sigma_i^2}{n}
\]

Second, the actual method of sampling allows for the selection of \( R \) adjacent weeks at a time. The flows in adjacent weeks will be highly correlated. Therefore the important factor reducing the variance \( \sigma_i^2 \) is the number of separate periods of counting \( \rho \) at the site, rather than the total weeks sampled \( n \). In this case the variance of the estimated AAWT is:

\[
\sigma^2(w) = \left( \frac{N-1}{N} \right) \cdot (1-f) \cdot \frac{\sigma_i^2}{\rho}
\]

The value of \( \rho \) is given by:

\[
\rho = \frac{52}{RN}
\]

so that:

\[
\sigma^2(w) = \left( \frac{N-1}{N} \right) \cdot (1-f) \cdot \frac{RN}{52} \cdot \frac{\sigma_i^2}{i}
\]
This equation gives the variance in the estimated AAWT for the particular sampling procedure being used. From this
an expression for the variance of the estimated annual flow at the site $Y_i$ may be obtained, since:

$$\hat{Y}_i = 52 \cdot w$$

so that:

$$\sigma^2(\hat{Y}_i) = 52 \cdot \left(\frac{N-1}{N}\right) \cdot (1-f) \cdot RN \sigma_i^2$$

(14)

Since the cycle length $L$ equals $RN$ and the sample fraction $f$ equals $1/N$ equation (14) can be simplified to:

$$\sigma^2(\hat{Y}_i) = 52 \cdot p^2 \cdot L \cdot \sigma_i^2$$

(15)

where $P$ equals $(1-f)$ and is the proportion of days not sampled. The only variable in this expression that depends
on the particular site chosen is the variance of weekly traffic flows ($\sigma_i^2$).

This variable depends primarily on two factors: the actual level of flow at the site $Y_i$, and the extent of the
seasonal variation at the site. Since the latter would not be known in practice the best that can be done is to use an
average value $A_I$ for the particular type of site I under consideration so that the best estimate of $\sigma_i^2$ is

$$\hat{\sigma}_i^2 = k \cdot A_I$$

where $k$ is a constant.

The estimated variance of the predicted annual flow is related to the level of flow, the proportion of days
not sampled and the cycle length as follows:

$$\hat{\sigma}^2(\hat{Y}_i) \propto Y_i^2 \cdot p^2 \cdot L$$

(16)

Since the proportional error in the predicted annual flow is the same as the proportional error in the estimator
$365/D_S$, the variance in the error term $\alpha_{is}$ given in equation (2) will be given by:

$$\text{Var}(\alpha_{is}) \propto \left(\frac{365}{D_S}\right)^2 \cdot p^2 \cdot L$$

(17)
11. APPENDIX 3

Statistical analysis for factoring by days and using G-factors

The overall coefficient of variation $C$ for all schedules of a procedure is given by equation (7) of the main text as:

$$C = a \left( \frac{N-1}{N} \right)^{\frac{1}{L}}$$

which for simplicity can be expressed as:

$$C = a X$$

For each observation $X_j$ there is a value for the coefficient of variation $C_j$ given by:

$$C_j = aX_j + e_j$$

where $e_j \sim N(0, \sigma_j^2)$

In considering the G-factor model it seems reasonable to suppose that the variance of the error term ($\sigma_j^2$) increases as $X_j$ increases. This is supported by the data (Figures 3 and 4). This relationship is assumed linear, so that:

$$\sigma_j^2 = kX_j$$

where $k$ is a constant. The slope of the regression 'a' can be estimated from equation (2) by a weighted linear combination of the independent random variables $C_j/X_j$, so that:

$$a = \frac{\sum w_j C_j}{X_j}$$

where $w_j$ are the weights which sum to unity to give an unbiased estimate of 'a'. Since 'a' is a weighted linear combination its variance $\sigma_a^2$ is given by:

$$\sigma_a^2 = \frac{\sum w_j^2 \sigma_j^2}{X_j^2}$$

It can be shown by calculus that the variance of 'a' is minimised when the weights are given by

$$w_j = \frac{X_j}{\Sigma X_j}$$

On substituting values of $w_j$ and $\sigma_j$ in equation (4), and simplifying the following equation is obtained:

$$\sigma_a^2 = \frac{k}{\Sigma X_j}$$
This equation can be used to produce confidence intervals for the estimate of 'a'. The value of k can be estimated from the data by the expression:

\[ \hat{k} = \frac{1}{n_j} \sum \frac{e_j^2}{X_j} \]  

(23)

The estimated parameters of the above model are given in the following table for the two models based on factoring by days and on the G-factors.

**Values of parameters for model of equation (19) and (20)**

<table>
<thead>
<tr>
<th>Road Type*</th>
<th>Based on factoring by days</th>
<th>Based on G-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cj</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>119</td>
<td>198</td>
<td>232</td>
</tr>
<tr>
<td>( \hat{a} )</td>
<td>0.61</td>
<td>1.02</td>
</tr>
<tr>
<td>( \hat{k} )</td>
<td>0.05</td>
<td>0.16</td>
</tr>
<tr>
<td>( \sigma_a )</td>
<td>0.016</td>
<td>0.029</td>
</tr>
<tr>
<td>95 per cent confidence interval for 'a'</td>
<td>0.58</td>
<td>0.96</td>
</tr>
<tr>
<td>0.64</td>
<td>1.08</td>
<td>1.28</td>
</tr>
<tr>
<td>( \Sigma X_j )</td>
<td>194</td>
<td></td>
</tr>
</tbody>
</table>

* Road types numbered as given in Section 1.
ABSTRACT

ACCURACY OF ANNUAL TRAFFIC FLOW ESTIMATES FROM AUTOMATIC COUNTS: Garwyn Phillips (Local Government Operational Research Unit): Department of the Environment Department of Transport, TRRL Supplementary Report 515: Crowthorne, 1979 (Transport and Road Research Laboratory). Automatic traffic counters are used for on-going studies involving regular monitoring of traffic flows. The counters are rotated regularly among a number of sites throughout the year. Local authorities frequently use this sort of procedure to obtain estimates of annual flow on a network of roads. In this report, several methods of deriving these estimates are examined. Recommendations are given for the most appropriate counting schedules and a simple method of assessing the accuracy of the estimates is provided.

ISSN 0305-1315