Traffic volume estimation
from short period traffic counts

Magne Aldrin, Norwegian Computing Center

Abstract

This paper considers the problem of estimating the yearly traffic volume at a count site, when traffic counts are available for only a limited part of the year, perhaps only a few hours or days. A new method for estimating annual average daily traffic (AADT) based on regression is presented. In addition to being more precise than the traditional factor approach, the new method supplies the precision of the AADT estimate as a function of the sample design. This precision function may be used to optimize the sampling design before the actual counting is performed. Separate AADT estimates may be combined in various ways, for instance to an estimate of annual vehicle distance travelled (AVDT) within a specific region.

The new method is applied to traffic data from Oslo, Norway. For each count site, hourly counts of number of vehicles within five length classes are available in both directions of the road. The method provides AADT estimates and their precision for each length class within each direction, as well as for weighted sums of separate AADT estimates.

1 Introduction

Exact knowledge of yearly traffic volume parameters such as annual average daily traffic (AADT) can only be provided by permanent automatic traffic recorders under perfect conditions. In Oslo, Norway there are slightly more than 30 permanent count sites, each registering the traffic in one direction of a road. These are usually located in pairs, one for each direction of the road, but a pair is here regarded as two separate sites. At those permanent count sites the traffic are registered almost continuously, but with some periods of missing data due to, for instance, failure of the counting equipment. The number of vehicles are counted separately for 5 length classes, the smallest of which consists of vehicles between 0 and 5.5 meters. For the present study, we have data of 32 count stations for the

years 1994, 1995 and 1996. We will consider hourly count data, but the raw data may have even finer time resolution.

In addition to the permanent count sites, the traffic is counted periodically or sporadically on several hundred other count sites. For these count sites, the AADT estimates must be based on short period traffic counts, of lengths usually varying from a few hours to two weeks, say. With such a limited amount of data, it is certainly important to use estimation methods that utilize the information of the available data in an efficient way. Traffic data typically have very strong systematically variations within a day, a week and a year, with similar patterns at various roads. Therefore, it should be possible to use data from the permanent count sites to get information about the traffic on the periodic count sites. This paper presents a new method for estimating AADT and similar traffic parameters, with the intention of being more precise than the traditional factor approach (see for instance Sharma, Gulati and Rizak 1996).

The factor approach for estimating AADT may be divided into two steps. Step F1 involves calculations on data from permanent count sites, and step F2 involves calculations on short period data for the actual count site for which the AADT estimate is needed:
- F1: The permanent count sites with (almost) complete data are divided into a small number of groups with similar traffic patterns. In Oslo there are 3 such groups. For each group, a factor curve is calculated that describes typical variation within a year and within a week. In Oslo, each factor curve is a product of curves describing yearly, weekly and daily variation.
- F2: When estimating AADT for a site with sample counts, the site is first assigned to one of these groups. Then the appropriate factor curve is applied to the sample counts to produce an estimate of AADT.

Certainly there is no need to do step F1 each time step F2 is performed. Step F1 may for instance be performed once a year for the purpose of updating the factor curves.

The assignment of a site to one group will often be based on the available data for this site. If these are few (short counting period), the site may erroneously be assigned to the wrong group. On the other hand, when there are many data, the best factor curve will still not fit the data perfectly. From a statistical point of view the factor approach yields overfitting when there are few data, and underfitting when there are many data.

The new estimation method presented here is designed to overcome this problem. It is based on a model, the complexity of which is adaptive with respect to the

amount of data available. When the counting period is very short (e.g. 6 hours), the model is very simple. When the counting period is longer (e.g. 2 weeks), the model is more complex and is able to fit the data more precisely. Due to this adaptivity, the method is able to produce more precise estimates than the factor approach for the same amount of data, or equally precise estimates for less data. The method is called the basis curve method. It consists of four steps, where the first step B1 corresponds to step F1 of the factor approach, and the three next steps B2, B3 and B4 corresponds to step F2:

- B1: A set of so called basis curves is calculated from data from permanent count sites. The basis curves are functions of time (see Section 2 and Figure 1) that account for the main structures in the data.
- B2: A model is chosen with a complexity (i.e. number of basis curves) appropriate for the amount of available count data.
- B3: The number of vehicles per hour is estimated for all unobserved hours in a year, using the observed count data to fit the model with a certain combination of the basis curves.
- B4: The AADT estimate is given as the daily average estimated or, whenever available, observed number of vehicles for all hours in the year.

The basis curves method differs from the factor curves in two ways: Firstly, instead of using one out of a set of curves, the basis curve method combines several curves. Secondly, the model complexity, represented by the number of curves to be combined, is chosen adaptively with respect to the amount of data available.

The precision of the AADT estimates are provided as a function of the actual sample design, i.e. as a function of what time and for how long the traffic has been counted. This precision function (see Section 3) is estimated using the data from the permanent count sites. This is not limited to the specific estimation procedure used in steps B1 to B4, so in principle the factor approach could be extended with this feature as well.

The basis curve method as well as the precision functions are calibrated by a simulation experiment based on real data from permanent count sites. Sharma, Gulati and Rizak (1996) performed a similar type of simulation experiment on Canadian traffic data for assessing the uncertainty of the factor approach. However, they reported the uncertainty for selected sample designs only, where in the present paper the uncertainty is provided as a function of an arbitrary sample design.

The basis curve method is an extension of Aldrin (1995). The steps B2, B3 and B4 and the precision evaluation involve relatively simple calculations, what have been implemented in Excel. In step B1 more complicated calculations are needed, and

a prototype C++ code is implemented. Another alternative to the factor approach is presented in Lingras and Adamo (1996), who use both linear and non-linear regression (neural nets) methods.

In Section 2 the new estimation method is presented. Section 3 treats uncertainty evaluation, and Section 4 describes the simulation experiment and a comparison between the factor approach and our new method. Section 5 contains some conclusions.

2 The new estimation method

The traffic level may vary considerably from one count site to another. However, the variations over a year, a week and a day are usually very similar from site to site, especially within the same car length classes. Here we consider the first length class only. The basic element of our construction is a function of time that accounts for the average variation patterns of the traffic at the permanent count sites. This function is called the first basis curve, and denoted $b_1(t)$, where t is a specific hour of the year. A precise definition of the basis curves is given in Appendix A. Let y_{it} denote the number of vehicles in hour t at site i. As a first approximation, the hourly number of vehicles at the i-th count site is modelled as

$$y_{it} = c_i \cdot \exp(b_1(t)) \quad . \qquad (1)$$

Here, the constant c_i varies between the sites due to their different traffic levels, but the variation over time is common for all sites, i.e. the basis curve $b_1(t)$ does not depend on i. The basis curve may be decomposed into four components

- trend, a long term increase or decrease,
- yearly seasonal variation repeated each year,
- special days (Easter, Christmas and other holidays),
- weekly seasonal variation repeated each week.

These four components of the first basis curve are shown in Figure 1. The upper panel shows the increasing trend, the second panel shows the yearly seasonal variation with typical low traffic in winter and mid-summer, and the third panel shows that the traffic is typically low on special days. The special days may change from year to year, and here we show1994.The weekly seasonal pattern in the fourth panel is repeated throughout the year. The traffic is typically lower at night time and weekends. We also see the morning and afternoon rush as two peaks each weekday.

Figure 1 First decomposed basis curve for length class 0-5.5 meters.

In equation (1), all sites have their own specific level, but exactly the same variation over time. A better approximation is attained if each site is allowed to have separate amplitudes α_{1i} of the variations taken care of by $b_1(t)$. This may be expressed as

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t)) \quad . \qquad (2)$$

Certainly, the model (2) will not fit the observed data exactly. A second basic curve $b_2(t)$, again common for all sites, is introduced to improve the fit. This yields a more detailed model

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t)) \quad , \qquad (3)$$

where the α_{1i}'s and α_{2i}'s are specific for each site. The second basis curve $b_2(t)$ represents an adjustment to the first basis curve. However, the second basis curve is less interpretable than the first one, and, therefore, it is not shown here.

We continue to build more and more accurate models by increasing the number of basis curves, and the general model with K basis curves is

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t) + \ldots + \alpha_{Ki} b_K(t)) \quad . \qquad (4)$$

In our application on data from Oslo, the maximum value of K has been set to 8. The first step of the basis curve method is now:

- B1: The 8 basis curves are found by estimating the model (4) from the huge amount of data from the permanent count sites (see Appendix A for details). The site specific coefficients (α's and c's) and the common basis curves are both estimated, but only the latter will be used in later steps. In the next steps these basis curves are considered as known functions.

Now assume that we have only a short period of traffic counts for a site. We want to estimate the yearly traffic volume at this site, based on a model like (1)- (4). The next steps of the basis curve method are:

- B2: Choose an appropriate model. To avoid overfitting as well as underfitting, the model complexity should be adapted to the amount of data available. If the

count period is very short (6 hours, say), one should use a very simple model with few coefficients that need to be estimated, maybe model (1) or (2). If the count period is relatively long (2 weeks, say), one should use a more complex model, such as model (4) with K equal to 7 or 8. In order to help in the choice of the appropriate model, we have devised a function that for a given sample period, suggest a K in (4). This is similar to the function for assessment of estimation uncertainty, which is treated in section 3.

- B3: The model eventually chosen is estimated from the available count data. This means that the site-specific coefficients c_i, α_{1i}, α_{2i}, ..., α_{Ki} are estimated, whereas the basis curves are now treated as known functions. The estimation is done by ordinary linear regression with the logarithm of y_{it} as the response variable and the basis curves as explanatory variables. The estimated model further gives estimates for the number of vehicles for every hour of the year, denoted as $\hat{y}_{it}$.

- B4: The single hour estimates are then combined to an estimate of AADT as

$$\text{estimate of AADT at site i} = \sum_{t \in \text{obs}} y_{it} + \sum_{t \in \text{unobs}} \hat{y}_{it} , \quad (5)$$

where the first sum is taken over all hours with counts, and the second sum over all hours without counts.

We have so far considered estimation for vehicles in the first length class. Exactly the same methodology is used for the other four length classes, but with different basis curves and coefficients (α's and c's).

3 Uncertainty evaluation and optimization of sampling designs

The uncertainty of an AADT estimate depends on the sampling design, that is, when and how long the traffic is counted. In general the uncertainty will decrease when the counting period becomes longer, but it also depends on what time of the day or week the counts are done. The uncertainty will also typically as a function of the traffic volume. We have divided all hours in a week into 9 categories. One category is hours between 7.00 and 9.00, another is hours between 9.00 and 15.00. The error of the AADT estimate is modelled as a function of the number of counted hours between 7.00 and 9.00, the number of counted hours between 9.00 and 15.00, the number of counted hours within the 7 remaining categories, and an estimate of AADT.

$$z_j = 0.1 + \text{number of counted hours in the j-th category} \quad . \qquad (6)$$

The constant 0.1 is added to avoid problems when there is categories without counts. Then

$$\text{standard error of AADT estimate} = \sqrt{\gamma_0 \cdot z_1^{\gamma_1} \cdot \ldots \cdot z_9^{\gamma_9} \cdot (\text{AADT-estimate})^{\gamma_{10}}} \quad , \qquad (7)$$

where the γ-s are coefficients which are estimated by a simulation experiment based on real data from the permanent count sites.

The function (7)is specific for each length class, and in general the uncertainty increases with increasing length class. We do not present these functions in detail, but rather give an example on how the function for the first length class varies with the sampling design. First, assume that counts are available for full 24-hours periods on weekdays, and that between 1 and 5 such periods has been counted. Figure 2 shows how the relative standard error (i.e. the standard error of the estimate divided on the true AADT) decreases from 9% for 24 hours of counts to 6.6% for $5 \cdot 24$ hours of counts. If the count data were independent, we would expect that the standard error would decrease as $1/(\sqrt{\text{number of observations}})$, but since the count data are highly correlated over time, the decrease is much slower.

Figure 2 Relative standard error of AADT estimate in % as a function of counted weekdays.

Figure 3 shows a contour plot of the relative standard error as a function of the number of counted hours between 7.00 and 9.00, and between 9.00 and 15.00, with no counts in other periods. An 8-hours period of counts from 7.00 to 15.00 would give a relative standard error of about 13.5% (the point x=2, y=6 on the figure). The value of an extra counted hour depends on how many hours are already counted in this as well as in the other cathegories.To get maximal precision with a minimum number of counted hours, one should spread the counts over different parts of the day.

Figure 3 Contour plot of relative standard error of AADT estimate in %as a function of counted hours between 7.00 and 9.00, and between 9.00 and 15.00.

The precision functions can be used to construct confidence intervals for the AADT estimates. However, the precision functions may also be used before any counting is performed. One can use the functions to compare several alternative designs with the same cost, and choose the design that gives the best precision. This may be done for a specific count site.

In addition to AADT, we may be interested in annual vehicle distance travelled (AVDT) within a specific region. If each count site is related to a road with a given length, one gets an estimate of AVDT found as a weighted sum of AADT estimates for each site, where the weights are the road lengths. The precision of the AVDT estimate depends on the sampling design at each count site, but also on the true, but unknown, AADT values. In practice, the unknown AADTs must be replaced by their estimates. If one need an estimate of the precision function before the counting is actually performed, one may roughly guess the magnitudes of the separate AADTs.The resulting approximate precision function may be used to find an optimal sampling plan for the whole region. This allows for balancing between counting longer periods on heavy traffic roads, and counting shorter periods on roads with little traffic.

4 Simulation from real data. Calibration and comparison with the factor approach.

The new basis curve method has been compared to the traditional factor approach by a simulation experiment based on real data. The rules for choosing the number of basis curves (Section 2) and the precision functions (Section 3) have been found from a parallel experiment. The real data consist of (almost) continuous count data from 32 count sites for the years 1994, 1995 and 1996. Thus for these years, we know the true AADTs for each count site.

The simulation procedure is like this: A count site is temporarily removed from the permanent data, and plays the role as a short periodic count site. The basis curves are calculated from the remaining permanent data. Then, for a specific year, a count period with length between two hours and two weeks is generated randomly from the excluded site. The real counts in this period act as the short period counts. Based on these counts, the AADT are estimated, and the error relative to the true AADT are calculated. For each of the three years, several sampling designs are generated for the removed count site. The procedure is repeated such that each count site has played the role as a short periodic count site once.

For the first length class, the average absolute error for the factor approach was 9%. This was reduced to 7.2% by using the basis curve method. This may be

regarded as a moderate improvement, but if one look at Figure 2, we notice that in order to reduce the error from 9% to 7.2%, one has to increase the amount of data by a factor of 3. This means that compared to the factor approach, the basis curve method is able to produce equally good AADT estimates with only 1/3 of the amount of data.

For the other length classes, the results were much more in favour of the basis curve method. However, this was not a fair comparison, because the factor curves are not calibrated for these length classes.

5 Conclusions

The new basis curve method is able to produce more precise AADT estimates than the factor approach. The main reason is that the basis curve method is based on a model, the complexity of which is adapted to the available amount of count data.

The uncertainty of the AADT estimates can be calculated for any sampling design. As far as we know, in other studies such uncertainty has only been calculated for a few, specific designs.

Estimates of annual vehicle distance travelled (AVDT) in a specific area may be calculated as a straight forward weighted sum of separate AADT estimates. The precision of an AVDT estimate for a given sampling plan may roughly be calculated before any countings are done. This may be used to optimize the total sampling plan for the whole area.

Acknowledgements

The development of the basis curve method has been funded by “Vegdirektoratet” and “SFT”. I thank Arnoldo Frigessi and Ola Haug for valuable comment to drafts of this paper.

References

Aldrin, M. A statistical approach to the modelling of car traffic. *Traffic Engineering + Control*, **36**, 1995, 489-493.

Aldrin, M. A new method for traffic volume estimation based on count - Calibrated to short period counts from Oslo. Technical report STAT/07/97. Norwegian Computing Center, Oslo, 1997 (in Norwegian).

Davies, P.T. and Tso, M.K.-S. Procedures for reduced rank regression. *Applied Statistics*, **31**,1995, 244-255.

Lingras, P. and Adamo, M. Average and peak traffic volumes: neural nets, regression, factor approaches. *Journal of Computing Civil Engineering*, **10**, 1996, 300-306.

Sharma, S. C, Gulati, B. M. and Rizak, S. N. Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering*, **122**, 1996, 430-439.

Appendix A Definition of basis curves

Taking the logarithm of model (4) yields the additive model

$$\log y_{it} = \log c_i + \sum_{k=1}^{K} \alpha_{ki} b_k(t) \quad . \qquad (8)$$

In the following we show how the basis curves $b_k(t)$ are defined through a multivariate regression model. For each length class, the logarithm of hourly count data for the 32 count sites are put together into a matrix **Y**, with 32 columns and 26304 rows (one row for each hour of the years 1994 to 1996). The variation in **Y** is modelled as a function of 203 explanatory variables related to time:
 - a linear trend, time is the only variable
 - yearly seasonal effect, with17 variables, sine- and cosine functions with period 1year, 1/2 year, 1/3 year etc.
 - special days, 17 indicator variables
 - weekly seasonal variation, 168 indicator variables, one for each hour in a week

The explanatory variables are put together into a 26304x203 matrix **X**. **Y** and **X** is related through

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{XB} + \mathbf{E} \quad , \tag{9}$$

where $\mathbf{B}_0$ is a matrix of intercepts for each count site (all rows are identical), **B** is a matrix of regression coefficients, and **E** a matrix of random errors. By the method of reduced rank regression (Davies and Tso 1982), the regression matrix **B** may be decomposed into min(32,203)=32 terms as

$$\mathbf{B} = \sum_{k=1}^{32} \beta_k \alpha_k^T \quad , \tag{10}$$

where the β_k's are 203-dimensional vectors and the α_k's are 32 dimensional vectors. The decomposition is unique, and has the desirable property that the first term explains as much as possible of the variation in **Y**, the second term explains as much as possible of the variation not explained by the first term etc.

The basis curves are then defined through the corresponding decomposition of **XB**:

$$\mathbf{XB} = \mathbf{X}\sum_{k=1}^{32} \beta_k \alpha_k^T = \sum_{k=1}^{32} (\mathbf{X}\beta_k)\alpha_k^T = \sum_{k=1}^{32} b_k \alpha_k^T \quad . \tag{11}$$

Here, the k-th basis curve $b_k = \mathbf{X}\beta_k$ is a vector with length equal to the number of hours (26304), and $b_k(t)$ is it's value for the t-th hour. For the i-th response, the model on logarithmic scale can be written as (8), and transforming back yields models like (1)- (4).

Traffic volume estimation
from short period traffic counts

Magne Aldrin, Norwegian Computing Center

Abstract

This paper considers the problem of estimating the yearly traffic volume at a count site, when traffic counts are available for only a limited part of the year, perhaps only a few hours or days. A new method for estimating annual average daily traffic (AADT) based on regression is presented. In addition to being more precise than the traditional factor approach, the new method supplies the precision of the AADT estimate as a function of the sample design. This precision function may be used to optimize the sampling design before the actual counting is performed. Separate AADT estimates may be combined in various ways, for instance to an estimate of annual vehicle distance travelled (AVDT) within a specific region.

The new method is applied to traffic data from Oslo, Norway. For each count site, hourly counts of number of vehicles within five length classes are available in both directions of the road. The method provides AADT estimates and their precision for each length class within each direction, as well as for weighted sums of separate AADT estimates.

1 Introduction

Exact knowledge of yearly traffic volume parameters such as annual average daily traffic (AADT) can only be provided by permanent automatic traffic recorders under perfect conditions. In Oslo, Norway there are slightly more than 30 permanent count sites, each registering the traffic in one direction of a road. These are usually located in pairs, one for each direction of the road, but a pair is here regarded as two separate sites. At those permanent count sites the traffic are registered almost continuously, but with some periods of missing data due to, for instance, failure of the counting equipment. The number of vehicles are counted separately for 5 length classes, the smallest of which consists of vehicles between 0 and 5.5 meters. For the present study, we have data of 32 count stations for the

years 1994, 1995 and 1996. We will consider hourly count data, but the raw data may have even finer time resolution.

In addition to the permanent count sites, the traffic is counted periodically or sporadically on several hundred other count sites. For these count sites, the AADT estimates must be based on short period traffic counts, of lengths usually varying from a few hours to two weeks, say. With such a limited amount of data, it is certainly important to use estimation methods that utilize the information of the available data in an efficient way. Traffic data typically have very strong systematically variations within a day, a week and a year, with similar patterns at various roads. Therefore, it should be possible to use data from the permanent count sites to get information about the traffic on the periodic count sites. This paper presents a new method for estimating AADT and similar traffic parameters, with the intention of being more precise than the traditional factor approach (see for instance Sharma, Gulati and Rizak 1996).

The factor approach for estimating AADT may be divided into two steps. Step F1 involves calculations on data from permanent count sites, and step F2 involves calculations on short period data for the actual count site for which the AADT estimate is needed:

- F1: The permanent count sites with (almost) complete data are divided into a small number of groups with similar traffic patterns. In Oslo there are 3 such groups. For each group, a factor curve is calculated that describes typical variation within a year and within a week. In Oslo, each factor curve is a product of curves describing yearly, weekly and daily variation.
- F2: When estimating AADT for a site with sample counts, the site is first assigned to one of these groups. Then the appropriate factor curve is applied to the sample counts to produce an estimate of AADT.

Certainly there is no need to do step F1 each time step F2 is performed. Step F1 may for instance be performed once a year for the purpose of updating the factor curves.

The assignment of a site to one group will often be based on the available data for this site. If these are few (short counting period), the site may erroneously be assigned to the wrong group. On the other hand, when there are many data, the best factor curve will still not fit the data perfectly. From a statistical point of view the factor approach yields overfitting when there are few data, and underfitting when there are many data.

The new estimation method presented here is designed to overcome this problem. It is based on a model, the complexity of which is adaptive with respect to the

amount of data available. When the counting period is very short (e.g. 6 hours), the model is very simple. When the counting period is longer (e.g. 2 weeks), the model is more complex and is able to fit the data more precisely. Due to this adaptivity, the method is able to produce more precise estimates than the factor approach for the same amount of data, or equally precise estimates for less data. The method is called the basis curve method. It consists of four steps, where the first step B1 corresponds to step F1 of the factor approach, and the three next steps B2, B3 and B4 corresponds to step F2:

- B1: A set of so called basis curves is calculated from data from permanent count sites. The basis curves are functions of time (see Section 2 and Figure 1) that account for the main structures in the data.
- B2: A model is chosen with a complexity (i.e. number of basis curves) appropriate for the amount of available count data.
- B3: The number of vehicles per hour is estimated for all unobserved hours in a year, using the observed count data to fit the model with a certain combination of the basis curves.
- B4: The AADT estimate is given as the daily average estimated or, whenever available, observed number of vehicles for all hours in the year.

The basis curves method differs from the factor curves in two ways: Firstly, instead of using one out of a set of curves, the basis curve method combines several curves. Secondly, the model complexity, represented by the number of curves to be combined, is chosen adaptively with respect to the amount of data available.

The precision of the AADT estimates are provided as a function of the actual sample design, i.e. as a function of what time and for how long the traffic has been counted. This precision function (see Section 3) is estimated using the data from the permanent count sites. This is not limited to the specific estimation procedure used in steps B1 to B4, so in principle the factor approach could be extended with this feature as well.

The basis curve method as well as the precision functions are calibrated by a simulation experiment based on real data from permanent count sites. Sharma, Gulati and Rizak (1996) performed a similar type of simulation experiment on Canadian traffic data for assessing the uncertainty of the factor approach. However, they reported the uncertainty for selected sample designs only, where in the present paper the uncertainty is provided as a function of an arbitrary sample design.

The basis curve method is an extension of Aldrin (1995). The steps B2, B3 and B4 and the precision evaluation involve relatively simple calculations, what have been implemented in Excel. In step B1 more complicated calculations are needed, and

a prototype C++ code is implemented. Another alternative to the factor approach is presented in Lingras and Adamo (1996), who use both linear and non-linear regression (neural nets) methods.

In Section 2 the new estimation method is presented. Section 3 treats uncertainty evaluation, and Section 4 describes the simulation experiment and a comparison between the factor approach and our new method. Section 5 contains some conclusions.

2 The new estimation method

The traffic level may vary considerably from one count site to another. However, the variations over a year, a week and a day are usually very similar from site to site, especially within the same car length classes. Here we consider the first length class only. The basic element of our construction is a function of time that accounts for the average variation patterns of the traffic at the permanent count sites. This function is called the first basis curve, and denoted $b_1(t)$, where t is a specific hour of the year. A precise definition of the basis curves is given in Appendix A. Let y_{it} denote the number of vehicles in hour t at site i. As a first approximation, the hourly number of vehicles at the i-th count site is modelled as

$$y_{it} = c_i \cdot \exp(b_1(t)) \quad . \qquad (1)$$

Here, the constant c_i varies between the sites due to their different traffic levels, but the variation over time is common for all sites, i.e. the basis curve $b_1(t)$ does not depend on i. The basis curve may be decomposed into four components
- trend, a long term increase or decrease,
- yearly seasonal variation repeated each year,
- special days (Easter, Christmas and other holidays),
- weekly seasonal variation repeated each week.

These four components of the first basis curve are shown in Figure 1. The upper panel shows the increasing trend, the second panel shows the yearly seasonal variation with typical low traffic in winter and mid-summer, and the third panel shows that the traffic is typically low on special days. The special days may change from year to year, and here we show1994.The weekly seasonal pattern in the fourth panel is repeated throughout the year. The traffic is typically lower at night time and weekends. We also see the morning and afternoon rush as two peaks each weekday.

Figure 1 First decomposed basis curve for length class 0-5.5 meters.

In equation (1), all sites have their own specific level, but exactly the same variation over time. A better approximation is attained if each site is allowed to have separate amplitudes α_{1i} of the variations taken care of by $b_1(t)$. This may be expressed as

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t)) \quad . \tag{2}$$

Certainly, the model (2) will not fit the observed data exactly. A second basic curve $b_2(t)$, again common for all sites, is introduced to improve the fit. This yields a more detailed model

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t)) \quad , \tag{3}$$

where the α_{1i}'s and α_{2i}'s are specific for each site. The second basis curve $b_2(t)$ represents an adjustment to the first basis curve. However, the second basis curve is less interpretable than the first one, and, therefore, it is not shown here.

We continue to build more and more accurate models by increasing the number of basis curves, and the general model with K basis curves is

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t) + \ldots + \alpha_{Ki} b_K(t)) \quad . \tag{4}$$

In our application on data from Oslo, the maximum value of K has been set to 8. The first step of the basis curve method is now:

- B1: The 8 basis curves are found by estimating the model (4) from the huge amount of data from the permanent count sites (see Appendix A for details). The site specific coefficients (α's and c's) and the common basis curves are both estimated, but only the latter will be used in later steps. In the next steps these basis curves are considered as known functions.

Now assume that we have only a short period of traffic counts for a site. We want to estimate the yearly traffic volume at this site, based on a model like (1)- (4). The next steps of the basis curve method are:

- B2: Choose an appropriate model. To avoid overfitting as well as underfitting, the model complexity should be adapted to the amount of data available. If the

count period is very short (6 hours, say), one should use a very simple model with few coefficients that need to be estimated, maybe model (1) or (2). If the count period is relatively long (2 weeks, say), one should use a more complex model, such as model (4) with K equal to 7 or 8. In order to help in the choice of the appropriate model, we have devised a function that for a given sample period, suggest a K in (4). This is similar to the function for assessment of estimation uncertainty, which is treated in section 3.

- B3: The model eventually chosen is estimated from the available count data. This means that the site-specific coefficients c_i, α_{1i}, α_{2i}, ..., α_{Ki} are estimated, whereas the basis curves are now treated as known functions. The estimation is done by ordinary linear regression with the logarithm of y_{it} as the response variable and the basis curves as explanatory variables. The estimated model further gives estimates for the number of vehicles for every hour of the year, denoted as $\hat{y}_{it}$.

- B4: The single hour estimates are then combined to an estimate of AADT as

$$\text{estimate of AADT at site i} = \sum_{t \in \text{obs}} y_{it} + \sum_{t \in \text{unobs}} \hat{y}_{it} \quad , \qquad (5)$$

where the first sum is taken over all hours with counts, and the second sum over all hours without counts.

We have so far considered estimation for vehicles in the first length class. Exactly the same methodology is used for the other four length classes, but with different basis curves and coefficients (α's and c's).

3 Uncertainty evaluation and optimization of sampling designs

The uncertainty of an AADT estimate depends on the sampling design, that is, when and how long the traffic is counted. In general the uncertainty will decrease when the counting period becomes longer, but it also depends on what time of the day or week the counts are done. The uncertainty will also typically as a function of the traffic volume. We have divided all hours in a week into 9 categories. One category is hours between 7.00 and 9.00, another is hours between 9.00 and 15.00. The error of the AADT estimate is modelled as a function of the number of counted hours between 7.00 and 9.00, the number of counted hours between 9.00 and 15.00, the number of counted hours within the 7 remaining categories, and an estimate of AADT.

$$z_j = 0.1 + \text{number of counted hours in the j-th category} \quad . \qquad (6)$$

The constant 0.1 is added to avoid problems when there is categories without counts. Then

$$\text{standard error of AADT estimate} = \sqrt{\gamma_0 \cdot z_1^{\gamma_1} \cdot \ldots \cdot z_9^{\gamma_9} \cdot (\text{AADT-estimate})^{\gamma_{10}}} \quad , \qquad (7)$$

where the γ-s are coefficients which are estimated by a simulation experiment based on real data from the permanent count sites.

The function (7)is specific for each length class, and in general the uncertainty increases with increasing length class. We do not present these functions in detail, but rather give an example on how the function for the first length class varies with the sampling design. First, assume that counts are available for full 24-hours periods on weekdays, and that between 1 and 5 such periods has been counted. Figure 2 shows how the relative standard error (i.e. the standard error of the estimate divided on the true AADT) decreases from 9% for 24 hours of counts to 6.6% for $5 \cdot 24$ hours of counts. If the count data were independent, we would expect that the standard error would decrease as $1/(\sqrt{\text{number of observations}})$, but since the count data are highly correlated over time, the decrease is much slower.

Figure 2 Relative standard error of AADT estimate in % as a function of counted weekdays.

Figure 3 shows a contour plot of the relative standard error as a function of the number of counted hours between 7.00 and 9.00, and between 9.00 and 15.00, with no counts in other periods. An 8-hours period of counts from 7.00 to 15.00 would give a relative standard error of about 13.5% (the point x=2, y=6 on the figure). The value of an extra counted hour depends on how many hours are already counted in this as well as in the other cathegories.To get maximal precision with a minimum number of counted hours, one should spread the counts over different parts of the day.

Figure 3 Contour plot of relative standard error of AADT estimate in %as a function of counted hours between 7.00 and 9.00, and between 9.00 and 15.00.

The precision functions can be used to construct confidence intervals for the AADT estimates. However, the precision functions may also be used before any counting is performed. One can use the functions to compare several alternative designs with the same cost, and choose the design that gives the best precision. This may be done for a specific count site.

In addition to AADT, we may be interested in annual vehicle distance travelled (AVDT) within a specific region. If each count site is related to a road with a given length, one gets an estimate of AVDT found as a weighted sum of AADT estimates for each site, where the weights are the road lengths. The precision of the AVDT estimate depends on the sampling design at each count site, but also on the true, but unknown, AADT values. In practice, the unknown AADTs must be replaced by their estimates. If one need an estimate of the precision function before the counting is actually performed, one may roughly guess the magnitudes of the separate AADTs.The resulting approximate precision function may be used to find an optimal sampling plan for the whole region. This allows for balancing between counting longer periods on heavy traffic roads, and counting shorter periods on roads with little traffic.

4 Simulation from real data. Calibration and comparison with the factor approach.

The new basis curve method has been compared to the traditional factor approach by a simulation experiment based on real data. The rules for choosing the number of basis curves (Section 2) and the precision functions (Section 3) have been found from a parallel experiment. The real data consist of (almost) continuous count data from 32 count sites for the years 1994, 1995 and 1996. Thus for these years, we know the true AADTs for each count site.

The simulation procedure is like this: A count site is temporarily removed from the permanent data, and plays the role as a short periodic count site. The basis curves are calculated from the remaining permanent data. Then, for a specific year, a count period with length between two hours and two weeks is generated randomly from the excluded site. The real counts in this period act as the short period counts. Based on these counts, the AADT are estimated, and the error relative to the true AADT are calculated. For each of the three years, several sampling designs are generated for the removed count site. The procedure is repeated such that each count site has played the role as a short periodic count site once.

For the first length class, the average absolute error for the factor approach was 9%. This was reduced to 7.2% by using the basis curve method. This may be

regarded as a moderate improvement, but if one look at Figure 2, we notice that in order to reduce the error from 9% to 7.2%, one has to increase the amount of data by a factor of 3. This means that compared to the factor approach, the basis curve method is able to produce equally good AADT estimates with only 1/3 of the amount of data.

For the other length classes, the results were much more in favour of the basis curve method. However, this was not a fair comparison, because the factor curves are not calibrated for these length classes.

5 Conclusions

The new basis curve method is able to produce more precise AADT estimates than the factor approach. The main reason is that the basis curve method is based on a model, the complexity of which is adapted to the available amount of count data.

The uncertainty of the AADT estimates can be calculated for any sampling design. As far as we know, in other studies such uncertainty has only been calculated for a few, specific designs.

Estimates of annual vehicle distance travelled (AVDT) in a specific area may be calculated as a straight forward weighted sum of separate AADT estimates. The precision of an AVDT estimate for a given sampling plan may roughly be calculated before any countings are done. This may be used to optimize the total sampling plan for the whole area.

Acknowledgements

The development of the basis curve method has been funded by “Vegdirektoratet” and “SFT”. I thank Arnoldo Frigessi and Ola Haug for valuable comment to drafts of this paper.

References

Aldrin, M. A statistical approach to the modelling of car traffic. *Traffic Engineering + Control*, **36**, 1995, 489-493.

Aldrin, M. A new method for traffic volume estimation based on count - Calibrated to short period counts from Oslo. Technical report STAT/07/97. Norwegian Computing Center, Oslo, 1997 (in Norwegian).

Davies, P.T. and Tso, M.K.-S. Procedures for reduced rank regression. *Applied Statistics*, **31**,1995, 244-255.

Lingras, P. and Adamo, M. Average and peak traffic volumes: neural nets, regression, factor approaches. *Journal of Computing Civil Engineering*, **10**, 1996, 300-306.

Sharma, S. C, Gulati, B. M. and Rizak, S. N. Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering*, **122**, 1996, 430-439.

Appendix A Definition of basis curves

Taking the logarithm of model (4) yields the additive model

$$\log y_{it} = \log c_i + \sum_{k=1}^{K} \alpha_{ki} b_k(t) \quad . \qquad (8)$$

In the following we show how the basis curves $b_k(t)$ are defined through a multivariate regression model. For each length class, the logarithm of hourly count data for the 32 count sites are put together into a matrix **Y**, with 32 columns and 26304 rows (one row for each hour of the years 1994 to 1996). The variation in **Y** is modelled as a function of 203 explanatory variables related to time:

- a linear trend, time is the only variable
- yearly seasonal effect, with17 variables, sine- and cosine functions with period 1year, 1/2 year, 1/3 year etc.
- special days, 17 indicator variables
- weekly seasonal variation, 168 indicator variables, one for each hour in a week

The explanatory variables are put together into a 26304x203 matrix **X**. **Y** and **X** is related through

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{XB} + \mathbf{E} \quad , \tag{9}$$

where $\mathbf{B}_0$ is a matrix of intercepts for each count site (all rows are identical), **B** is a matrix of regression coefficients, and **E** a matrix of random errors. By the method of reduced rank regression (Davies and Tso 1982), the regression matrix **B** may be decomposed into min(32,203)=32 terms as

$$\mathbf{B} = \sum_{k=1}^{32} \beta_k \alpha_k^T \quad , \tag{10}$$

where the β_k's are 203-dimensional vectors and the α_k's are 32 dimensional vectors. The decomposition is unique, and has the desirable property that the first term explains as much as possible of the variation in **Y**, the second term explains as much as possible of the variation not explained by the first term etc.

The basis curves are then defined through the corresponding decomposition of **XB**:

$$\mathbf{XB} = \mathbf{X}\sum_{k=1}^{32} \beta_k \alpha_k^T = \sum_{k=1}^{32} (\mathbf{X}\beta_k)\alpha_k^T = \sum_{k=1}^{32} b_k \alpha_k^T \quad . \tag{11}$$

Here, the k-th basis curve $b_k = \mathbf{X}\beta_k$ is a vector with length equal to the number of hours (26304), and $b_k(t)$ is it's value for the t-th hour. For the i-th response, the model on logarithmic scale can be written as (8), and transforming back yields models like (1)- (4).

Traffic volume estimation from short period traffic counts

Magne Aldrin, Norwegian Computing Center

Abstract

This paper considers the problem of estimating the yearly traffic volume at a count site, when traffic counts are available for only a limited part of the year, perhaps only a few hours or days. A new method for estimating annual average daily traffic (AADT) based on regression is presented. In addition to being more precise than the traditional factor approach, the new method supplies the precision of the AADT estimate as a function of the sample design. This precision function may be used to optimize the sampling design before the actual counting is performed. Separate AADT estimates may be combined in various ways, for instance to an estimate of annual vehicle distance travelled (AVDT) within a specific region.

The new method is applied to traffic data from Oslo, Norway. For each count site, hourly counts of number of vehicles within five length classes are available in both directions of the road. The method provides AADT estimates and their precision for each length class within each direction, as well as for weighted sums of separate AADT estimates.

1 Introduction

Exact knowledge of yearly traffic volume parameters such as annual average daily traffic (AADT) can only be provided by permanent automatic traffic recorders under perfect conditions. In Oslo, Norway there are slightly more than 30 permanent count sites, each registering the traffic in one direction of a road. These are usually located in pairs, one for each direction of the road, but a pair is here regarded as two separate sites. At those permanent count sites the traffic are registered almost continuously, but with some periods of missing data due to, for instance, failure of the counting equipment. The number of vehicles are counted separately for 5 length classes, the smallest of which consists of vehicles between 0 and 5.5 meters. For the present study, we have data of 32 count stations for the

years 1994, 1995 and 1996. We will consider hourly count data, but the raw data may have even finer time resolution.

In addition to the permanent count sites, the traffic is counted periodically or sporadically on several hundred other count sites. For these count sites, the AADT estimates must be based on short period traffic counts, of lengths usually varying from a few hours to two weeks, say. With such a limited amount of data, it is certainly important to use estimation methods that utilize the information of the available data in an efficient way. Traffic data typically have very strong systematically variations within a day, a week and a year, with similar patterns at various roads. Therefore, it should be possible to use data from the permanent count sites to get information about the traffic on the periodic count sites. This paper presents a new method for estimating AADT and similar traffic parameters, with the intention of being more precise than the traditional factor approach (see for instance Sharma, Gulati and Rizak 1996).

The factor approach for estimating AADT may be divided into two steps. Step F1 involves calculations on data from permanent count sites, and step F2 involves calculations on short period data for the actual count site for which the AADT estimate is needed:

- F1: The permanent count sites with (almost) complete data are divided into a small number of groups with similar traffic patterns. In Oslo there are 3 such groups. For each group, a factor curve is calculated that describes typical variation within a year and within a week. In Oslo, each factor curve is a product of curves describing yearly, weekly and daily variation.
- F2: When estimating AADT for a site with sample counts, the site is first assigned to one of these groups. Then the appropriate factor curve is applied to the sample counts to produce an estimate of AADT.

Certainly there is no need to do step F1 each time step F2 is performed. Step F1 may for instance be performed once a year for the purpose of updating the factor curves.

The assignment of a site to one group will often be based on the available data for this site. If these are few (short counting period), the site may erroneously be assigned to the wrong group. On the other hand, when there are many data, the best factor curve will still not fit the data perfectly. From a statistical point of view the factor approach yields overfitting when there are few data, and underfitting when there are many data.

The new estimation method presented here is designed to overcome this problem. It is based on a model, the complexity of which is adaptive with respect to the

amount of data available. When the counting period is very short (e.g. 6 hours), the model is very simple. When the counting period is longer (e.g. 2 weeks), the model is more complex and is able to fit the data more precisely. Due to this adaptivity, the method is able to produce more precise estimates than the factor approach for the same amount of data, or equally precise estimates for less data. The method is called the basis curve method. It consists of four steps, where the first step B1 corresponds to step F1 of the factor approach, and the three next steps B2, B3 and B4 corresponds to step F2:

- B1: A set of so called basis curves is calculated from data from permanent count sites. The basis curves are functions of time (see Section 2 and Figure 1) that account for the main structures in the data.
- B2: A model is chosen with a complexity (i.e. number of basis curves) appropriate for the amount of available count data.
- B3: The number of vehicles per hour is estimated for all unobserved hours in a year, using the observed count data to fit the model with a certain combination of the basis curves.
- B4: The AADT estimate is given as the daily average estimated or, whenever available, observed number of vehicles for all hours in the year.

The basis curves method differs from the factor curves in two ways: Firstly, instead of using one out of a set of curves, the basis curve method combines several curves. Secondly, the model complexity, represented by the number of curves to be combined, is chosen adaptively with respect to the amount of data available.

The precision of the AADT estimates are provided as a function of the actual sample design, i.e. as a function of what time and for how long the traffic has been counted. This precision function (see Section 3) is estimated using the data from the permanent count sites. This is not limited to the specific estimation procedure used in steps B1 to B4, so in principle the factor approach could be extended with this feature as well.

The basis curve method as well as the precision functions are calibrated by a simulation experiment based on real data from permanent count sites. Sharma, Gulati and Rizak (1996) performed a similar type of simulation experiment on Canadian traffic data for assessing the uncertainty of the factor approach. However, they reported the uncertainty for selected sample designs only, where in the present paper the uncertainty is provided as a function of an arbitrary sample design.

The basis curve method is an extension of Aldrin (1995). The steps B2, B3 and B4 and the precision evaluation involve relatively simple calculations, what have been implemented in Excel. In step B1 more complicated calculations are needed, and

a prototype C++ code is implemented. Another alternative to the factor approach is presented in Lingras and Adamo (1996), who use both linear and non-linear regression (neural nets) methods.

In Section 2 the new estimation method is presented. Section 3 treats uncertainty evaluation, and Section 4 describes the simulation experiment and a comparison between the factor approach and our new method. Section 5 contains some conclusions.

2 The new estimation method

The traffic level may vary considerably from one count site to another. However, the variations over a year, a week and a day are usually very similar from site to site, especially within the same car length classes. Here we consider the first length class only. The basic element of our construction is a function of time that accounts for the average variation patterns of the traffic at the permanent count sites. This function is called the first basis curve, and denoted $b_1(t)$, where t is a specific hour of the year. A precise definition of the basis curves is given in Appendix A. Let y_{it} denote the number of vehicles in hour t at site i. As a first approximation, the hourly number of vehicles at the i-th count site is modelled as

$$y_{it} = c_i \cdot \exp(b_1(t)) \quad . \qquad (1)$$

Here, the constant c_i varies between the sites due to their different traffic levels, but the variation over time is common for all sites, i.e. the basis curve $b_1(t)$ does not depend on i. The basis curve may be decomposed into four components
- trend, a long term increase or decrease,
- yearly seasonal variation repeated each year,
- special days (Easter, Christmas and other holidays),
- weekly seasonal variation repeated each week.

These four components of the first basis curve are shown in Figure 1. The upper panel shows the increasing trend, the second panel shows the yearly seasonal variation with typical low traffic in winter and mid-summer, and the third panel shows that the traffic is typically low on special days. The special days may change from year to year, and here we show1994.The weekly seasonal pattern in the fourth panel is repeated throughout the year. The traffic is typically lower at night time and weekends. We also see the morning and afternoon rush as two peaks each weekday.

Figure 1 First decomposed basis curve for length class 0-5.5 meters.

In equation (1), all sites have their own specific level, but exactly the same variation over time. A better approximation is attained if each site is allowed to have separate amplitudes α_{1i} of the variations taken care of by $b_1(t)$. This may be expressed as

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t)) \quad . \qquad (2)$$

Certainly, the model (2) will not fit the observed data exactly. A second basic curve $b_2(t)$, again common for all sites, is introduced to improve the fit. This yields a more detailed model

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t)) \quad , \qquad (3)$$

where the α_{1i}'s and α_{2i}'s are specific for each site. The second basis curve $b_2(t)$ represents an adjustment to the first basis curve. However, the second basis curve is less interpretable than the first one, and, therefore, it is not shown here.

We continue to build more and more accurate models by increasing the number of basis curves, and the general model with K basis curves is

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t) + \ldots + \alpha_{Ki} b_K(t)) \quad . \qquad (4)$$

In our application on data from Oslo, the maximum value of K has been set to 8. The first step of the basis curve method is now:

- B1: The 8 basis curves are found by estimating the model (4) from the huge amount of data from the permanent count sites (see Appendix A for details). The site specific coefficients (α's and c's) and the common basis curves are both estimated, but only the latter will be used in later steps. In the next steps these basis curves are considered as known functions.

Now assume that we have only a short period of traffic counts for a site. We want to estimate the yearly traffic volume at this site, based on a model like (1)- (4). The next steps of the basis curve method are:

- B2: Choose an appropriate model. To avoid overfitting as well as underfitting, the model complexity should be adapted to the amount of data available. If the

count period is very short (6 hours, say), one should use a very simple model with few coefficients that need to be estimated, maybe model (1) or (2). If the count period is relatively long (2 weeks, say), one should use a more complex model, such as model (4) with K equal to 7 or 8. In order to help in the choice of the appropriate model, we have devised a function that for a given sample period, suggest a K in (4). This is similar to the function for assessment of estimation uncertainty, which is treated in section 3.

- B3: The model eventually chosen is estimated from the available count data. This means that the site-specific coefficients c_i, α_{1i}, α_{2i}, ..., α_{Ki} are estimated, whereas the basis curves are now treated as known functions. The estimation is done by ordinary linear regression with the logarithm of y_{it} as the response variable and the basis curves as explanatory variables. The estimated model further gives estimates for the number of vehicles for every hour of the year, denoted as $\hat{y}_{it}$.

- B4: The single hour estimates are then combined to an estimate of AADT as

$$\text{estimate of AADT at site i} = \sum_{t \in \text{obs}} y_{it} + \sum_{t \in \text{unobs}} \hat{y}_{it} \quad , \tag{5}$$

where the first sum is taken over all hours with counts, and the second sum over all hours without counts.

We have so far considered estimation for vehicles in the first length class. Exactly the same methodology is used for the other four length classes, but with different basis curves and coefficients (α's and c's).

3 Uncertainty evaluation and optimization of sampling designs

The uncertainty of an AADT estimate depends on the sampling design, that is, when and how long the traffic is counted. In general the uncertainty will decrease when the counting period becomes longer, but it also depends on what time of the day or week the counts are done. The uncertainty will also typically as a function of the traffic volume. We have divided all hours in a week into 9 categories. One category is hours between 7.00 and 9.00, another is hours between 9.00 and 15.00. The error of the AADT estimate is modelled as a function of the number of counted hours between 7.00 and 9.00, the number of counted hours between 9.00 and 15.00, the number of counted hours within the 7 remaining categories, and an estimate of AADT.

$$z_j = 0.1 + \text{number of counted hours in the j-th category} \quad . \qquad (6)$$

The constant 0.1 is added to avoid problems when there is categories without counts. Then

$$\text{standard error of AADT estimate} = \sqrt{\gamma_0 \cdot z_1^{\gamma_1} \cdot \ldots \cdot z_9^{\gamma_9} \cdot (\text{AADT-estimate})^{\gamma_{10}}} \quad , \qquad (7)$$

where the γ-s are coefficients which are estimated by a simulation experiment based on real data from the permanent count sites.

The function (7)is specific for each length class, and in general the uncertainty increases with increasing length class. We do not present these functions in detail, but rather give an example on how the function for the first length class varies with the sampling design. First, assume that counts are available for full 24-hours periods on weekdays, and that between 1 and 5 such periods has been counted. Figure 2 shows how the relative standard error (i.e. the standard error of the estimate divided on the true AADT) decreases from 9% for 24 hours of counts to 6.6% for $5 \cdot 24$ hours of counts. If the count data were independent, we would expect that the standard error would decrease as $1/(\sqrt{\text{number of observations}})$, but since the count data are highly correlated over time, the decrease is much slower.

Figure 2 Relative standard error of AADT estimate in % as a function of counted weekdays.

Figure 3 shows a contour plot of the relative standard error as a function of the number of counted hours between 7.00 and 9.00, and between 9.00 and 15.00, with no counts in other periods. An 8-hours period of counts from 7.00 to 15.00 would give a relative standard error of about 13.5% (the point x=2, y=6 on the figure). The value of an extra counted hour depends on how many hours are already counted in this as well as in the other cathegories.To get maximal precision with a minimum number of counted hours, one should spread the counts over different parts of the day.

Figure 3 Contour plot of relative standard error of AADT estimate in %as a function of counted hours between 7.00 and 9.00, and between 9.00 and 15.00.

The precision functions can be used to construct confidence intervals for the AADT estimates. However, the precision functions may also be used before any counting is performed. One can use the functions to compare several alternative designs with the same cost, and choose the design that gives the best precision. This may be done for a specific count site.

In addition to AADT, we may be interested in annual vehicle distance travelled (AVDT) within a specific region. If each count site is related to a road with a given length, one gets an estimate of AVDT found as a weighted sum of AADT estimates for each site, where the weights are the road lengths. The precision of the AVDT estimate depends on the sampling design at each count site, but also on the true, but unknown, AADT values. In practice, the unknown AADTs must be replaced by their estimates. If one need an estimate of the precision function before the counting is actually performed, one may roughly guess the magnitudes of the separate AADTs.The resulting approximate precision function may be used to find an optimal sampling plan for the whole region. This allows for balancing between counting longer periods on heavy traffic roads, and counting shorter periods on roads with little traffic.

4 Simulation from real data. Calibration and comparison with the factor approach.

The new basis curve method has been compared to the traditional factor approach by a simulation experiment based on real data. The rules for choosing the number of basis curves (Section 2) and the precision functions (Section 3) have been found from a parallel experiment. The real data consist of (almost) continuous count data from 32 count sites for the years 1994, 1995 and 1996. Thus for these years, we know the true AADTs for each count site.

The simulation procedure is like this: A count site is temporarily removed from the permanent data, and plays the role as a short periodic count site. The basis curves are calculated from the remaining permanent data. Then, for a specific year, a count period with length between two hours and two weeks is generated randomly from the excluded site. The real counts in this period act as the short period counts. Based on these counts, the AADT are estimated, and the error relative to the true AADT are calculated. For each of the three years, several sampling designs are generated for the removed count site. The procedure is repeated such that each count site has played the role as a short periodic count site once.

For the first length class, the average absolute error for the factor approach was 9%. This was reduced to 7.2% by using the basis curve method. This may be

regarded as a moderate improvement, but if one look at Figure 2, we notice that in order to reduce the error from 9% to 7.2%, one has to increase the amount of data by a factor of 3. This means that compared to the factor approach, the basis curve method is able to produce equally good AADT estimates with only 1/3 of the amount of data.

For the other length classes, the results were much more in favour of the basis curve method. However, this was not a fair comparison, because the factor curves are not calibrated for these length classes.

5 Conclusions

The new basis curve method is able to produce more precise AADT estimates than the factor approach. The main reason is that the basis curve method is based on a model, the complexity of which is adapted to the available amount of count data.

The uncertainty of the AADT estimates can be calculated for any sampling design. As far as we know, in other studies such uncertainty has only been calculated for a few, specific designs.

Estimates of annual vehicle distance travelled (AVDT) in a specific area may be calculated as a straight forward weighted sum of separate AADT estimates. The precision of an AVDT estimate for a given sampling plan may roughly be calculated before any countings are done. This may be used to optimize the total sampling plan for the whole area.

Acknowledgements

The development of the basis curve method has been funded by “Vegdirektoratet” and “SFT”. I thank Arnoldo Frigessi and Ola Haug for valuable comment to drafts of this paper.

References

Aldrin, M. A statistical approach to the modelling of car traffic. *Traffic Engineering + Control*, **36**, 1995, 489-493.

Aldrin, M. A new method for traffic volume estimation based on count - Calibrated to short period counts from Oslo. Technical report STAT/07/97. Norwegian Computing Center, Oslo, 1997 (in Norwegian).

Davies, P.T. and Tso, M.K.-S. Procedures for reduced rank regression. *Applied Statistics*, **31**,1995, 244-255.

Lingras, P. and Adamo, M. Average and peak traffic volumes: neural nets, regression, factor approaches. *Journal of Computing Civil Engineering*, **10**, 1996, 300-306.

Sharma, S. C, Gulati, B. M. and Rizak, S. N. Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering*, **122**, 1996, 430-439.

Appendix A Definition of basis curves

Taking the logarithm of model (4) yields the additive model

$$\log y_{it} = \log c_i + \sum_{k=1}^{K} \alpha_{ki} b_k(t) \quad . \qquad (8)$$

In the following we show how the basis curves $b_k(t)$ are defined through a multivariate regression model. For each length class, the logarithm of hourly count data for the 32 count sites are put together into a matrix **Y**, with 32 columns and 26304 rows (one row for each hour of the years 1994 to 1996). The variation in **Y** is modelled as a function of 203 explanatory variables related to time:
- a linear trend, time is the only variable
- yearly seasonal effect, with17 variables, sine- and cosine functions with period 1year, 1/2 year, 1/3 year etc.
- special days, 17 indicator variables
- weekly seasonal variation, 168 indicator variables, one for each hour in a week

The explanatory variables are put together into a 26304x203 matrix **X**. **Y** and **X** is related through

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{XB} + \mathbf{E} \quad , \tag{9}$$

where $\mathbf{B}_0$ is a matrix of intercepts for each count site (all rows are identical), **B** is a matrix of regression coefficients, and **E** a matrix of random errors. By the method of reduced rank regression (Davies and Tso 1982), the regression matrix **B** may be decomposed into min(32,203)=32 terms as

$$\mathbf{B} = \sum_{k=1}^{32} \beta_k \alpha_k^T \quad , \tag{10}$$

where the β_k's are 203-dimensional vectors and the α_k's are 32 dimensional vectors. The decomposition is unique, and has the desirable property that the first term explains as much as possible of the variation in **Y**, the second term explains as much as possible of the variation not explained by the first term etc.

The basis curves are then defined through the corresponding decomposition of **XB**:

$$\mathbf{XB} = \mathbf{X}\sum_{k=1}^{32} \beta_k \alpha_k^T = \sum_{k=1}^{32} (\mathbf{X}\beta_k)\alpha_k^T = \sum_{k=1}^{32} b_k \alpha_k^T \quad . \tag{11}$$

Here, the k-th basis curve $b_k = \mathbf{X}\beta_k$ is a vector with length equal to the number of hours (26304), and $b_k(t)$ is it's value for the t-th hour. For the i-th response, the model on logarithmic scale can be written as (8), and transforming back yields models like (1)- (4).

Traffic volume estimation from short period traffic counts

Magne Aldrin, Norwegian Computing Center

Abstract

This paper considers the problem of estimating the yearly traffic volume at a count site, when traffic counts are available for only a limited part of the year, perhaps only a few hours or days. A new method for estimating annual average daily traffic (AADT) based on regression is presented. In addition to being more precise than the traditional factor approach, the new method supplies the precision of the AADT estimate as a function of the sample design. This precision function may be used to optimize the sampling design before the actual counting is performed. Separate AADT estimates may be combined in various ways, for instance to an estimate of annual vehicle distance travelled (AVDT) within a specific region.

The new method is applied to traffic data from Oslo, Norway. For each count site, hourly counts of number of vehicles within five length classes are available in both directions of the road. The method provides AADT estimates and their precision for each length class within each direction, as well as for weighted sums of separate AADT estimates.

1 Introduction

Exact knowledge of yearly traffic volume parameters such as annual average daily traffic (AADT) can only be provided by permanent automatic traffic recorders under perfect conditions. In Oslo, Norway there are slightly more than 30 permanent count sites, each registering the traffic in one direction of a road. These are usually located in pairs, one for each direction of the road, but a pair is here regarded as two separate sites. At those permanent count sites the traffic are registered almost continuously, but with some periods of missing data due to, for instance, failure of the counting equipment. The number of vehicles are counted separately for 5 length classes, the smallest of which consists of vehicles between 0 and 5.5 meters. For the present study, we have data of 32 count stations for the

years 1994, 1995 and 1996. We will consider hourly count data, but the raw data may have even finer time resolution.

In addition to the permanent count sites, the traffic is counted periodically or sporadically on several hundred other count sites. For these count sites, the AADT estimates must be based on short period traffic counts, of lengths usually varying from a few hours to two weeks, say. With such a limited amount of data, it is certainly important to use estimation methods that utilize the information of the available data in an efficient way. Traffic data typically have very strong systematically variations within a day, a week and a year, with similar patterns at various roads. Therefore, it should be possible to use data from the permanent count sites to get information about the traffic on the periodic count sites. This paper presents a new method for estimating AADT and similar traffic parameters, with the intention of being more precise than the traditional factor approach (see for instance Sharma, Gulati and Rizak 1996).

The factor approach for estimating AADT may be divided into two steps. Step F1 involves calculations on data from permanent count sites, and step F2 involves calculations on short period data for the actual count site for which the AADT estimate is needed:
- F1: The permanent count sites with (almost) complete data are divided into a small number of groups with similar traffic patterns. In Oslo there are 3 such groups. For each group, a factor curve is calculated that describes typical variation within a year and within a week. In Oslo, each factor curve is a product of curves describing yearly, weekly and daily variation.
- F2: When estimating AADT for a site with sample counts, the site is first assigned to one of these groups. Then the appropriate factor curve is applied to the sample counts to produce an estimate of AADT.

Certainly there is no need to do step F1 each time step F2 is performed. Step F1 may for instance be performed once a year for the purpose of updating the factor curves.

The assignment of a site to one group will often be based on the available data for this site. If these are few (short counting period), the site may erroneously be assigned to the wrong group. On the other hand, when there are many data, the best factor curve will still not fit the data perfectly. From a statistical point of view the factor approach yields overfitting when there are few data, and underfitting when there are many data.

The new estimation method presented here is designed to overcome this problem. It is based on a model, the complexity of which is adaptive with respect to the

amount of data available. When the counting period is very short (e.g. 6 hours), the model is very simple. When the counting period is longer (e.g. 2 weeks), the model is more complex and is able to fit the data more precisely. Due to this adaptivity, the method is able to produce more precise estimates than the factor approach for the same amount of data, or equally precise estimates for less data. The method is called the basis curve method. It consists of four steps, where the first step B1 corresponds to step F1 of the factor approach, and the three next steps B2, B3 and B4 corresponds to step F2:

- B1: A set of so called basis curves is calculated from data from permanent count sites. The basis curves are functions of time (see Section 2 and Figure 1) that account for the main structures in the data.
- B2: A model is chosen with a complexity (i.e. number of basis curves) appropriate for the amount of available count data.
- B3: The number of vehicles per hour is estimated for all unobserved hours in a year, using the observed count data to fit the model with a certain combination of the basis curves.
- B4: The AADT estimate is given as the daily average estimated or, whenever available, observed number of vehicles for all hours in the year.

The basis curves method differs from the factor curves in two ways: Firstly, instead of using one out of a set of curves, the basis curve method combines several curves. Secondly, the model complexity, represented by the number of curves to be combined, is chosen adaptively with respect to the amount of data available.

The precision of the AADT estimates are provided as a function of the actual sample design, i.e. as a function of what time and for how long the traffic has been counted. This precision function (see Section 3) is estimated using the data from the permanent count sites. This is not limited to the specific estimation procedure used in steps B1 to B4, so in principle the factor approach could be extended with this feature as well.

The basis curve method as well as the precision functions are calibrated by a simulation experiment based on real data from permanent count sites. Sharma, Gulati and Rizak (1996) performed a similar type of simulation experiment on Canadian traffic data for assessing the uncertainty of the factor approach. However, they reported the uncertainty for selected sample designs only, where in the present paper the uncertainty is provided as a function of an arbitrary sample design.

The basis curve method is an extension of Aldrin (1995). The steps B2, B3 and B4 and the precision evaluation involve relatively simple calculations, what have been implemented in Excel. In step B1 more complicated calculations are needed, and

a prototype C++ code is implemented. Another alternative to the factor approach is presented in Lingras and Adamo (1996), who use both linear and non-linear regression (neural nets) methods.

In Section 2 the new estimation method is presented. Section 3 treats uncertainty evaluation, and Section 4 describes the simulation experiment and a comparison between the factor approach and our new method. Section 5 contains some conclusions.

2 The new estimation method

The traffic level may vary considerably from one count site to another. However, the variations over a year, a week and a day are usually very similar from site to site, especially within the same car length classes. Here we consider the first length class only. The basic element of our construction is a function of time that accounts for the average variation patterns of the traffic at the permanent count sites. This function is called the first basis curve, and denoted $b_1(t)$, where t is a specific hour of the year. A precise definition of the basis curves is given in Appendix A. Let y_{it} denote the number of vehicles in hour t at site i. As a first approximation, the hourly number of vehicles at the i-th count site is modelled as

$$y_{it} = c_i \cdot \exp(b_1(t)) \quad . \qquad (1)$$

Here, the constant c_i varies between the sites due to their different traffic levels, but the variation over time is common for all sites, i.e. the basis curve $b_1(t)$ does not depend on i. The basis curve may be decomposed into four components

- trend, a long term increase or decrease,
- yearly seasonal variation repeated each year,
- special days (Easter, Christmas and other holidays),
- weekly seasonal variation repeated each week.

These four components of the first basis curve are shown in Figure 1. The upper panel shows the increasing trend, the second panel shows the yearly seasonal variation with typical low traffic in winter and mid-summer, and the third panel shows that the traffic is typically low on special days. The special days may change from year to year, and here we show1994.The weekly seasonal pattern in the fourth panel is repeated throughout the year. The traffic is typically lower at night time and weekends. We also see the morning and afternoon rush as two peaks each weekday.

Figure 1 First decomposed basis curve for length class 0-5.5 meters.

In equation (1), all sites have their own specific level, but exactly the same variation over time. A better approximation is attained if each site is allowed to have separate amplitudes α_{1i} of the variations taken care of by $b_1(t)$. This may be expressed as

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t)) \quad . \qquad (2)$$

Certainly, the model (2) will not fit the observed data exactly. A second basic curve $b_2(t)$, again common for all sites, is introduced to improve the fit. This yields a more detailed model

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t)) \quad , \qquad (3)$$

where the α_{1i}'s and α_{2i}'s are specific for each site. The second basis curve $b_2(t)$ represents an adjustment to the first basis curve. However, the second basis curve is less interpretable than the first one, and, therefore, it is not shown here.

We continue to build more and more accurate models by increasing the number of basis curves, and the general model with K basis curves is

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t) + \ldots + \alpha_{Ki} b_K(t)) \quad . \qquad (4)$$

In our application on data from Oslo, the maximum value of K has been set to 8. The first step of the basis curve method is now:

- B1: The 8 basis curves are found by estimating the model (4) from the huge amount of data from the permanent count sites (see Appendix A for details). The site specific coefficients (α's and c's) and the common basis curves are both estimated, but only the latter will be used in later steps. In the next steps these basis curves are considered as known functions.

Now assume that we have only a short period of traffic counts for a site. We want to estimate the yearly traffic volume at this site, based on a model like (1)- (4). The next steps of the basis curve method are:

- B2: Choose an appropriate model. To avoid overfitting as well as underfitting, the model complexity should be adapted to the amount of data available. If the

count period is very short (6 hours, say), one should use a very simple model with few coefficients that need to be estimated, maybe model (1) or (2). If the count period is relatively long (2 weeks, say), one should use a more complex model, such as model (4) with K equal to 7 or 8. In order to help in the choice of the appropriate model, we have devised a function that for a given sample period, suggest a K in (4). This is similar to the function for assessment of estimation uncertainty, which is treated in section 3.

- B3: The model eventually chosen is estimated from the available count data. This means that the site-specific coefficients c_i, α_{1i}, α_{2i}, ..., α_{Ki} are estimated, whereas the basis curves are now treated as known functions. The estimation is done by ordinary linear regression with the logarithm of y_{it} as the response variable and the basis curves as explanatory variables. The estimated model further gives estimates for the number of vehicles for every hour of the year, denoted as $\hat{y}_{it}$.

- B4: The single hour estimates are then combined to an estimate of AADT as

$$\text{estimate of AADT at site i} = \sum_{t \in \text{obs}} y_{it} + \sum_{t \in \text{unobs}} \hat{y}_{it} \quad , \qquad (5)$$

where the first sum is taken over all hours with counts, and the second sum over all hours without counts.

We have so far considered estimation for vehicles in the first length class. Exactly the same methodology is used for the other four length classes, but with different basis curves and coefficients (α's and c's).

3 Uncertainty evaluation and optimization of sampling designs

The uncertainty of an AADT estimate depends on the sampling design, that is, when and how long the traffic is counted. In general the uncertainty will decrease when the counting period becomes longer, but it also depends on what time of the day or week the counts are done. The uncertainty will also typically as a function of the traffic volume. We have divided all hours in a week into 9 categories. One category is hours between 7.00 and 9.00, another is hours between 9.00 and 15.00. The error of the AADT estimate is modelled as a function of the number of counted hours between 7.00 and 9.00, the number of counted hours between 9.00 and 15.00, the number of counted hours within the 7 remaining categories, and an estimate of AADT.

$$z_j = 0.1 + \text{number of counted hours in the j-th category} \quad . \qquad (6)$$

The constant 0.1 is added to avoid problems when there is categories without counts. Then

$$\text{standard error of AADT estimate} = \sqrt{\gamma_0 \cdot z_1^{\gamma_1} \cdot \ldots \cdot z_9^{\gamma_9} \cdot (\text{AADT-estimate})^{\gamma_{10}}} \quad , \qquad (7)$$

where the γ-s are coefficients which are estimated by a simulation experiment based on real data from the permanent count sites.

The function (7)is specific for each length class, and in general the uncertainty increases with increasing length class. We do not present these functions in detail, but rather give an example on how the function for the first length class varies with the sampling design. First, assume that counts are available for full 24-hours periods on weekdays, and that between 1 and 5 such periods has been counted. Figure 2 shows how the relative standard error (i.e. the standard error of the estimate divided on the true AADT) decreases from 9% for 24 hours of counts to 6.6% for $5 \cdot 24$ hours of counts. If the count data were independent, we would expect that the standard error would decrease as $1/(\sqrt{\text{number of observations}})$, but since the count data are highly correlated over time, the decrease is much slower.

Figure 2 Relative standard error of AADT estimate in % as a function of counted weekdays.

Figure 3 shows a contour plot of the relative standard error as a function of the number of counted hours between 7.00 and 9.00, and between 9.00 and 15.00, with no counts in other periods. An 8-hours period of counts from 7.00 to 15.00 would give a relative standard error of about 13.5% (the point x=2, y=6 on the figure). The value of an extra counted hour depends on how many hours are already counted in this as well as in the other cathegories.To get maximal precision with a minimum number of counted hours, one should spread the counts over different parts of the day.

Figure 3 Contour plot of relative standard error of AADT estimate in %as a function of counted hours between 7.00 and 9.00, and between 9.00 and 15.00.

The precision functions can be used to construct confidence intervals for the AADT estimates. However, the precision functions may also be used before any counting is performed. One can use the functions to compare several alternative designs with the same cost, and choose the design that gives the best precision. This may be done for a specific count site.

In addition to AADT, we may be interested in annual vehicle distance travelled (AVDT) within a specific region. If each count site is related to a road with a given length, one gets an estimate of AVDT found as a weighted sum of AADT estimates for each site, where the weights are the road lengths. The precision of the AVDT estimate depends on the sampling design at each count site, but also on the true, but unknown, AADT values. In practice, the unknown AADTs must be replaced by their estimates. If one need an estimate of the precision function before the counting is actually performed, one may roughly guess the magnitudes of the separate AADTs.The resulting approximate precision function may be used to find an optimal sampling plan for the whole region. This allows for balancing between counting longer periods on heavy traffic roads, and counting shorter periods on roads with little traffic.

4 Simulation from real data. Calibration and comparison with the factor approach.

The new basis curve method has been compared to the traditional factor approach by a simulation experiment based on real data. The rules for choosing the number of basis curves (Section 2) and the precision functions (Section 3) have been found from a parallel experiment. The real data consist of (almost) continuous count data from 32 count sites for the years 1994, 1995 and 1996. Thus for these years, we know the true AADTs for each count site.

The simulation procedure is like this: A count site is temporarily removed from the permanent data, and plays the role as a short periodic count site. The basis curves are calculated from the remaining permanent data. Then, for a specific year, a count period with length between two hours and two weeks is generated randomly from the excluded site. The real counts in this period act as the short period counts. Based on these counts, the AADT are estimated, and the error relative to the true AADT are calculated. For each of the three years, several sampling designs are generated for the removed count site. The procedure is repeated such that each count site has played the role as a short periodic count site once.

For the first length class, the average absolute error for the factor approach was 9%. This was reduced to 7.2% by using the basis curve method. This may be

regarded as a moderate improvement, but if one look at Figure 2, we notice that in order to reduce the error from 9% to 7.2%, one has to increase the amount of data by a factor of 3. This means that compared to the factor approach, the basis curve method is able to produce equally good AADT estimates with only 1/3 of the amount of data.

For the other length classes, the results were much more in favour of the basis curve method. However, this was not a fair comparison, because the factor curves are not calibrated for these length classes.

5 Conclusions

The new basis curve method is able to produce more precise AADT estimates than the factor approach. The main reason is that the basis curve method is based on a model, the complexity of which is adapted to the available amount of count data.

The uncertainty of the AADT estimates can be calculated for any sampling design. As far as we know, in other studies such uncertainty has only been calculated for a few, specific designs.

Estimates of annual vehicle distance travelled (AVDT) in a specific area may be calculated as a straight forward weighted sum of separate AADT estimates. The precision of an AVDT estimate for a given sampling plan may roughly be calculated before any countings are done. This may be used to optimize the total sampling plan for the whole area.

Acknowledgements

The development of the basis curve method has been funded by “Vegdirektoratet” and “SFT”. I thank Arnoldo Frigessi and Ola Haug for valuable comment to drafts of this paper.

References

Aldrin, M. A statistical approach to the modelling of car traffic. *Traffic Engineering + Control*, **36**, 1995, 489-493.

Aldrin, M. A new method for traffic volume estimation based on count - Calibrated to short period counts from Oslo. Technical report STAT/07/97. Norwegian Computing Center, Oslo, 1997 (in Norwegian).

Davies, P.T. and Tso, M.K.-S. Procedures for reduced rank regression. *Applied Statistics*, **31**,1995, 244-255.

Lingras, P. and Adamo, M. Average and peak traffic volumes: neural nets, regression, factor approaches. *Journal of Computing Civil Engineering*, **10**, 1996, 300-306.

Sharma, S. C, Gulati, B. M. and Rizak, S. N. Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering*, **122**, 1996, 430-439.

Appendix A Definition of basis curves

Taking the logarithm of model (4) yields the additive model

$$\log y_{it} = \log c_i + \sum_{k=1}^{K} \alpha_{ki} b_k(t) \quad . \tag{8}$$

In the following we show how the basis curves $b_k(t)$ are defined through a multivariate regression model. For each length class, the logarithm of hourly count data for the 32 count sites are put together into a matrix **Y**, with 32 columns and 26304 rows (one row for each hour of the years 1994 to 1996). The variation in **Y** is modelled as a function of 203 explanatory variables related to time:

- a linear trend, time is the only variable
- yearly seasonal effect, with17 variables, sine- and cosine functions with period 1year, 1/2 year, 1/3 year etc.
- special days, 17 indicator variables
- weekly seasonal variation, 168 indicator variables, one for each hour in a week

The explanatory variables are put together into a 26304x203 matrix **X**. **Y** and **X** is related through

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{XB} + \mathbf{E} \qquad , \tag{9}$$

where $\mathbf{B}_0$ is a matrix of intercepts for each count site (all rows are identical), **B** is a matrix of regression coefficients, and **E** a matrix of random errors. By the method of reduced rank regression (Davies and Tso 1982), the regression matrix **B** may be decomposed into min(32,203)=32 terms as

$$\mathbf{B} = \sum_{k=1}^{32} \beta_k \alpha_k^T \qquad , \tag{10}$$

where the β_k's are 203-dimensional vectors and the α_k's are 32 dimensional vectors. The decomposition is unique, and has the desirable property that the first term explains as much as possible of the variation in **Y**, the second term explains as much as possible of the variation not explained by the first term etc.

The basis curves are then defined through the corresponding decomposition of **XB**:

$$\mathbf{XB} = \mathbf{X}\sum_{k=1}^{32} \beta_k \alpha_k^T = \sum_{k=1}^{32} (\mathbf{X}\beta_k)\alpha_k^T = \sum_{k=1}^{32} b_k \alpha_k^T \qquad . \tag{11}$$

Here, the k-th basis curve $b_k = \mathbf{X}\beta_k$ is a vector with length equal to the number of hours (26304), and $b_k(t)$ is it's value for the t-th hour. For the i-th response, the model on logarithmic scale can be written as (8), and transforming back yields models like (1)- (4).

Traffic volume estimation
from short period traffic counts

Magne Aldrin, Norwegian Computing Center

Abstract

This paper considers the problem of estimating the yearly traffic volume at a count site, when traffic counts are available for only a limited part of the year, perhaps only a few hours or days. A new method for estimating annual average daily traffic (AADT) based on regression is presented. In addition to being more precise than the traditional factor approach, the new method supplies the precision of the AADT estimate as a function of the sample design. This precision function may be used to optimize the sampling design before the actual counting is performed. Separate AADT estimates may be combined in various ways, for instance to an estimate of annual vehicle distance travelled (AVDT) within a specific region.

The new method is applied to traffic data from Oslo, Norway. For each count site, hourly counts of number of vehicles within five length classes are available in both directions of the road. The method provides AADT estimates and their precision for each length class within each direction, as well as for weighted sums of separate AADT estimates.

1 Introduction

Exact knowledge of yearly traffic volume parameters such as annual average daily traffic (AADT) can only be provided by permanent automatic traffic recorders under perfect conditions. In Oslo, Norway there are slightly more than 30 permanent count sites, each registering the traffic in one direction of a road. These are usually located in pairs, one for each direction of the road, but a pair is here regarded as two separate sites. At those permanent count sites the traffic are registered almost continuously, but with some periods of missing data due to, for instance, failure of the counting equipment. The number of vehicles are counted separately for 5 length classes, the smallest of which consists of vehicles between 0 and 5.5 meters. For the present study, we have data of 32 count stations for the

years 1994, 1995 and 1996. We will consider hourly count data, but the raw data may have even finer time resolution.

In addition to the permanent count sites, the traffic is counted periodically or sporadically on several hundred other count sites. For these count sites, the AADT estimates must be based on short period traffic counts, of lengths usually varying from a few hours to two weeks, say. With such a limited amount of data, it is certainly important to use estimation methods that utilize the information of the available data in an efficient way. Traffic data typically have very strong systematically variations within a day, a week and a year, with similar patterns at various roads. Therefore, it should be possible to use data from the permanent count sites to get information about the traffic on the periodic count sites. This paper presents a new method for estimating AADT and similar traffic parameters, with the intention of being more precise than the traditional factor approach (see for instance Sharma, Gulati and Rizak 1996).

The factor approach for estimating AADT may be divided into two steps. Step F1 involves calculations on data from permanent count sites, and step F2 involves calculations on short period data for the actual count site for which the AADT estimate is needed:

- F1: The permanent count sites with (almost) complete data are divided into a small number of groups with similar traffic patterns. In Oslo there are 3 such groups. For each group, a factor curve is calculated that describes typical variation within a year and within a week. In Oslo, each factor curve is a product of curves describing yearly, weekly and daily variation.
- F2: When estimating AADT for a site with sample counts, the site is first assigned to one of these groups. Then the appropriate factor curve is applied to the sample counts to produce an estimate of AADT.

Certainly there is no need to do step F1 each time step F2 is performed. Step F1 may for instance be performed once a year for the purpose of updating the factor curves.

The assignment of a site to one group will often be based on the available data for this site. If these are few (short counting period), the site may erroneously be assigned to the wrong group. On the other hand, when there are many data, the best factor curve will still not fit the data perfectly. From a statistical point of view the factor approach yields overfitting when there are few data, and underfitting when there are many data.

The new estimation method presented here is designed to overcome this problem. It is based on a model, the complexity of which is adaptive with respect to the

amount of data available. When the counting period is very short (e.g. 6 hours), the model is very simple. When the counting period is longer (e.g. 2 weeks), the model is more complex and is able to fit the data more precisely. Due to this adaptivity, the method is able to produce more precise estimates than the factor approach for the same amount of data, or equally precise estimates for less data. The method is called the basis curve method. It consists of four steps, where the first step B1 corresponds to step F1 of the factor approach, and the three next steps B2, B3 and B4 corresponds to step F2:

- B1: A set of so called basis curves is calculated from data from permanent count sites. The basis curves are functions of time (see Section 2 and Figure 1) that account for the main structures in the data.
- B2: A model is chosen with a complexity (i.e. number of basis curves) appropriate for the amount of available count data.
- B3: The number of vehicles per hour is estimated for all unobserved hours in a year, using the observed count data to fit the model with a certain combination of the basis curves.
- B4: The AADT estimate is given as the daily average estimated or, whenever available, observed number of vehicles for all hours in the year.

The basis curves method differs from the factor curves in two ways: Firstly, instead of using one out of a set of curves, the basis curve method combines several curves. Secondly, the model complexity, represented by the number of curves to be combined, is chosen adaptively with respect to the amount of data available.

The precision of the AADT estimates are provided as a function of the actual sample design, i.e. as a function of what time and for how long the traffic has been counted. This precision function (see Section 3) is estimated using the data from the permanent count sites. This is not limited to the specific estimation procedure used in steps B1 to B4, so in principle the factor approach could be extended with this feature as well.

The basis curve method as well as the precision functions are calibrated by a simulation experiment based on real data from permanent count sites. Sharma, Gulati and Rizak (1996) performed a similar type of simulation experiment on Canadian traffic data for assessing the uncertainty of the factor approach. However, they reported the uncertainty for selected sample designs only, where in the present paper the uncertainty is provided as a function of an arbitrary sample design.

The basis curve method is an extension of Aldrin (1995). The steps B2, B3 and B4 and the precision evaluation involve relatively simple calculations, what have been implemented in Excel. In step B1 more complicated calculations are needed, and

a prototype C++ code is implemented. Another alternative to the factor approach is presented in Lingras and Adamo (1996), who use both linear and non-linear regression (neural nets) methods.

In Section 2 the new estimation method is presented. Section 3 treats uncertainty evaluation, and Section 4 describes the simulation experiment and a comparison between the factor approach and our new method. Section 5 contains some conclusions.

2 The new estimation method

The traffic level may vary considerably from one count site to another. However, the variations over a year, a week and a day are usually very similar from site to site, especially within the same car length classes. Here we consider the first length class only. The basic element of our construction is a function of time that accounts for the average variation patterns of the traffic at the permanent count sites. This function is called the first basis curve, and denoted $b_1(t)$, where t is a specific hour of the year. A precise definition of the basis curves is given in Appendix A. Let y_{it} denote the number of vehicles in hour t at site i. As a first approximation, the hourly number of vehicles at the i-th count site is modelled as

$$y_{it} = c_i \cdot \exp(b_1(t)) \quad . \qquad (1)$$

Here, the constant c_i varies between the sites due to their different traffic levels, but the variation over time is common for all sites, i.e. the basis curve $b_1(t)$ does not depend on i. The basis curve may be decomposed into four components

- trend, a long term increase or decrease,
- yearly seasonal variation repeated each year,
- special days (Easter, Christmas and other holidays),
- weekly seasonal variation repeated each week.

These four components of the first basis curve are shown in Figure 1. The upper panel shows the increasing trend, the second panel shows the yearly seasonal variation with typical low traffic in winter and mid-summer, and the third panel shows that the traffic is typically low on special days. The special days may change from year to year, and here we show1994.The weekly seasonal pattern in the fourth panel is repeated throughout the year. The traffic is typically lower at night time and weekends. We also see the morning and afternoon rush as two peaks each weekday.

Figure 1 First decomposed basis curve for length class 0-5.5 meters.

In equation (1), all sites have their own specific level, but exactly the same variation over time. A better approximation is attained if each site is allowed to have separate amplitudes α_{1i} of the variations taken care of by $b_1(t)$. This may be expressed as

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t)) \quad . \tag{2}$$

Certainly, the model (2) will not fit the observed data exactly. A second basic curve $b_2(t)$, again common for all sites, is introduced to improve the fit. This yields a more detailed model

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t)) \quad , \tag{3}$$

where the α_{1i}'s and α_{2i}'s are specific for each site. The second basis curve $b_2(t)$ represents an adjustment to the first basis curve. However, the second basis curve is less interpretable than the first one, and, therefore, it is not shown here.

We continue to build more and more accurate models by increasing the number of basis curves, and the general model with K basis curves is

$$y_{it} = c_i \cdot \exp(\alpha_{1i} b_1(t) + \alpha_{2i} b_2(t) + \ldots + \alpha_{Ki} b_K(t)) \quad . \tag{4}$$

In our application on data from Oslo, the maximum value of K has been set to 8. The first step of the basis curve method is now:

- B1: The 8 basis curves are found by estimating the model (4) from the huge amount of data from the permanent count sites (see Appendix A for details). The site specific coefficients (α's and c's) and the common basis curves are both estimated, but only the latter will be used in later steps. In the next steps these basis curves are considered as known functions.

Now assume that we have only a short period of traffic counts for a site. We want to estimate the yearly traffic volume at this site, based on a model like (1)- (4). The next steps of the basis curve method are:

- B2: Choose an appropriate model. To avoid overfitting as well as underfitting, the model complexity should be adapted to the amount of data available. If the

count period is very short (6 hours, say), one should use a very simple model with few coefficients that need to be estimated, maybe model (1) or (2). If the count period is relatively long (2 weeks, say), one should use a more complex model, such as model (4) with K equal to 7 or 8. In order to help in the choice of the appropriate model, we have devised a function that for a given sample period, suggest a K in (4). This is similar to the function for assessment of estimation uncertainty, which is treated in section 3.

- B3: The model eventually chosen is estimated from the available count data. This means that the site-specific coefficients c_i, α_{1i}, α_{2i}, ..., α_{Ki} are estimated, whereas the basis curves are now treated as known functions. The estimation is done by ordinary linear regression with the logarithm of y_{it} as the response variable and the basis curves as explanatory variables. The estimated model further gives estimates for the number of vehicles for every hour of the year, denoted as $\hat{y}_{it}$.

- B4: The single hour estimates are then combined to an estimate of AADT as

$$\text{estimate of AADT at site i} = \sum_{t \in \text{obs}} y_{it} + \sum_{t \in \text{unobs}} \hat{y}_{it} \quad , \qquad (5)$$

where the first sum is taken over all hours with counts, and the second sum over all hours without counts.

We have so far considered estimation for vehicles in the first length class. Exactly the same methodology is used for the other four length classes, but with different basis curves and coefficients (α's and c's).

3 Uncertainty evaluation and optimization of sampling designs

The uncertainty of an AADT estimate depends on the sampling design, that is, when and how long the traffic is counted. In general the uncertainty will decrease when the counting period becomes longer, but it also depends on what time of the day or week the counts are done. The uncertainty will also typically as a function of the traffic volume. We have divided all hours in a week into 9 categories. One category is hours between 7.00 and 9.00, another is hours between 9.00 and 15.00. The error of the AADT estimate is modelled as a function of the number of counted hours between 7.00 and 9.00, the number of counted hours between 9.00 and 15.00, the number of counted hours within the 7 remaining categories, and an estimate of AADT.

$$z_j = 0.1 + \text{number of counted hours in the j-th category} \quad . \tag{6}$$

The constant 0.1 is added to avoid problems when there is categories without counts. Then

$$\text{standard error of AADT estimate} = \sqrt{\gamma_0 \cdot z_1^{\gamma_1} \cdot \ldots \cdot z_9^{\gamma_9} \cdot (\text{AADT-estimate})^{\gamma_{10}}} \quad , \tag{7}$$

where the γ-s are coefficients which are estimated by a simulation experiment based on real data from the permanent count sites.

The function (7)is specific for each length class, and in general the uncertainty increases with increasing length class. We do not present these functions in detail, but rather give an example on how the function for the first length class varies with the sampling design. First, assume that counts are available for full 24-hours periods on weekdays, and that between 1 and 5 such periods has been counted. Figure 2 shows how the relative standard error (i.e. the standard error of the estimate divided on the true AADT) decreases from 9% for 24 hours of counts to 6.6% for $5 \cdot 24$ hours of counts. If the count data were independent, we would expect that the standard error would decrease as $1/(\sqrt{\text{number of observations}})$, but since the count data are highly correlated over time, the decrease is much slower.

Figure 2 Relative standard error of AADT estimate in % as a function of counted weekdays.

Figure 3 shows a contour plot of the relative standard error as a function of the number of counted hours between 7.00 and 9.00, and between 9.00 and 15.00, with no counts in other periods. An 8-hours period of counts from 7.00 to 15.00 would give a relative standard error of about 13.5% (the point x=2, y=6 on the figure). The value of an extra counted hour depends on how many hours are already counted in this as well as in the other cathegories.To get maximal precision with a minimum number of counted hours, one should spread the counts over different parts of the day.

Figure 3 Contour plot of relative standard error of AADT estimate in %as a function of counted hours between 7.00 and 9.00, and between 9.00 and 15.00.

The precision functions can be used to construct confidence intervals for the AADT estimates. However, the precision functions may also be used before any counting is performed. One can use the functions to compare several alternative designs with the same cost, and choose the design that gives the best precision. This may be done for a specific count site.

In addition to AADT, we may be interested in annual vehicle distance travelled (AVDT) within a specific region. If each count site is related to a road with a given length, one gets an estimate of AVDT found as a weighted sum of AADT estimates for each site, where the weights are the road lengths. The precision of the AVDT estimate depends on the sampling design at each count site, but also on the true, but unknown, AADT values. In practice, the unknown AADTs must be replaced by their estimates. If one need an estimate of the precision function before the counting is actually performed, one may roughly guess the magnitudes of the separate AADTs.The resulting approximate precision function may be used to find an optimal sampling plan for the whole region. This allows for balancing between counting longer periods on heavy traffic roads, and counting shorter periods on roads with little traffic.

4 Simulation from real data. Calibration and comparison with the factor approach.

The new basis curve method has been compared to the traditional factor approach by a simulation experiment based on real data. The rules for choosing the number of basis curves (Section 2) and the precision functions (Section 3) have been found from a parallel experiment. The real data consist of (almost) continuous count data from 32 count sites for the years 1994, 1995 and 1996. Thus for these years, we know the true AADTs for each count site.

The simulation procedure is like this: A count site is temporarily removed from the permanent data, and plays the role as a short periodic count site. The basis curves are calculated from the remaining permanent data. Then, for a specific year, a count period with length between two hours and two weeks is generated randomly from the excluded site. The real counts in this period act as the short period counts. Based on these counts, the AADT are estimated, and the error relative to the true AADT are calculated. For each of the three years, several sampling designs are generated for the removed count site. The procedure is repeated such that each count site has played the role as a short periodic count site once.

For the first length class, the average absolute error for the factor approach was 9%. This was reduced to 7.2% by using the basis curve method. This may be

regarded as a moderate improvement, but if one look at Figure 2, we notice that in order to reduce the error from 9% to 7.2%, one has to increase the amount of data by a factor of 3. This means that compared to the factor approach, the basis curve method is able to produce equally good AADT estimates with only 1/3 of the amount of data.

For the other length classes, the results were much more in favour of the basis curve method. However, this was not a fair comparison, because the factor curves are not calibrated for these length classes.

5 Conclusions

The new basis curve method is able to produce more precise AADT estimates than the factor approach. The main reason is that the basis curve method is based on a model, the complexity of which is adapted to the available amount of count data.

The uncertainty of the AADT estimates can be calculated for any sampling design. As far as we know, in other studies such uncertainty has only been calculated for a few, specific designs.

Estimates of annual vehicle distance travelled (AVDT) in a specific area may be calculated as a straight forward weighted sum of separate AADT estimates. The precision of an AVDT estimate for a given sampling plan may roughly be calculated before any countings are done. This may be used to optimize the total sampling plan for the whole area.

Acknowledgements

The development of the basis curve method has been funded by “Vegdirektoratet” and “SFT”. I thank Arnoldo Frigessi and Ola Haug for valuable comment to drafts of this paper.

References

Aldrin, M. A statistical approach to the modelling of car traffic. *Traffic Engineering + Control*, **36**, 1995, 489-493.

Aldrin, M. A new method for traffic volume estimation based on count - Calibrated to short period counts from Oslo. Technical report STAT/07/97. Norwegian Computing Center, Oslo, 1997 (in Norwegian).

Davies, P.T. and Tso, M.K.-S. Procedures for reduced rank regression. *Applied Statistics*, **31**,1995, 244-255.

Lingras, P. and Adamo, M. Average and peak traffic volumes: neural nets, regression, factor approaches. *Journal of Computing Civil Engineering*, **10**, 1996, 300-306.

Sharma, S. C, Gulati, B. M. and Rizak, S. N. Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering*, **122**, 1996, 430-439.

Appendix A Definition of basis curves

Taking the logarithm of model (4) yields the additive model

$$\log y_{it} = \log c_i + \sum_{k=1}^{K} \alpha_{ki} b_k(t) \qquad . \qquad (8)$$

In the following we show how the basis curves $b_k(t)$ are defined through a multivariate regression model. For each length class, the logarithm of hourly count data for the 32 count sites are put together into a matrix **Y**, with 32 columns and 26304 rows (one row for each hour of the years 1994 to 1996). The variation in **Y** is modelled as a function of 203 explanatory variables related to time:
- a linear trend, time is the only variable
- yearly seasonal effect, with17 variables, sine- and cosine functions with period 1year, 1/2 year, 1/3 year etc.
- special days, 17 indicator variables
- weekly seasonal variation, 168 indicator variables, one for each hour in a week

The explanatory variables are put together into a 26304x203 matrix **X**. **Y** and **X** is related through

$$\mathbf{Y} = \mathbf{B}_0 + \mathbf{XB} + \mathbf{E} \quad , \tag{9}$$

where $\mathbf{B}_0$ is a matrix of intercepts for each count site (all rows are identical), **B** is a matrix of regression coefficients, and **E** a matrix of random errors. By the method of reduced rank regression (Davies and Tso 1982), the regression matrix **B** may be decomposed into min(32,203)=32 terms as

$$\mathbf{B} = \sum_{k=1}^{32} \beta_k \alpha_k^T \quad , \tag{10}$$

where the β_k's are 203-dimensional vectors and the α_k's are 32 dimensional vectors. The decomposition is unique, and has the desirable property that the first term explains as much as possible of the variation in **Y**, the second term explains as much as possible of the variation not explained by the first term etc.

The basis curves are then defined through the corresponding decomposition of **XB**:

$$\mathbf{XB} = \mathbf{X}\sum_{k=1}^{32} \beta_k \alpha_k^T = \sum_{k=1}^{32} (\mathbf{X}\beta_k)\alpha_k^T = \sum_{k=1}^{32} b_k \alpha_k^T \quad . \tag{11}$$

Here, the k-th basis curve $b_k = \mathbf{X}\beta_k$ is a vector with length equal to the number of hours (26304), and $b_k(t)$ is it's value for the t-th hour. For the i-th response, the model on logarithmic scale can be written as (8), and transforming back yields models like (1)- (4).