Improving Group Assignment and AADT Estimation Accuracy of Short-term Traffic Counts using Historical Seasonal Patterns & Bayesian Statistics

Ming Zhonga,*, Ehsan Bagherib, Jim Christiec

a Department of Civil Engineering, University of New Brunswick, Fredericton, N.B., Canada E3B 5A3
b Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, O.N., Canada N2L 3G
c Faculty of Science, Applied Science & Engineering, University of New Brunswick, St. John

Abstract

Annual Average Daily Traffic (AADT) is one of the most important pieces of data being used widely in planning, design, operation and management of roads and facilities. A reliable estimate of AADT has always been one of the main interests of transportation and highway agencies. Traditionally, AADTs are estimated for the majority of road segments of a network from short-term traffic counts (STTCs or coverage counts) by applying a set of expansion factors derived from permanent traffic counts (PTCs). Literature indicates that the FHWA (Federal Highway Administration) method may be most widely used. In this method, roads in the same functional class are assumed to have similar traffic patterns, and the factors derived from PTC data collected from the class is used to convert STTCs to AADT estimates. However, it should be noted that, because roads from a functional class do not necessarily have similar seasonal traffic variations, this method may sometimes produce large estimation errors. In this regard, this paper proposes a novel pattern-matching method, which constructs a seasonal traffic variation profile for a short-term counting site using all historical counts available and then use this profile to assign the site to a PTC or a PTC group. In addition, a Bayesian approach is developed to explicitly consider and show the “risk or uncertainty” associated with assigning each short-term counting site to different PTC or PTC groups. Study results based on the simulated STTCs from a permanent counter on a winter recreational road in Alberta, Canada show that the new method proposed can limit the 95th percentile (P95) AADT estimation error to less than 13%, in contrast to 21.7% from the FHWA method. Moreover, it should be noted that the proposed method will not impose any additional monitoring cost, or make any change to existing traffic monitoring programs, and therefore it will be easy for highway agencies to incorporate the proposed method in their practice.

© 2012 Published by Elsevier B.V. Selection and/or peer review under responsibility of Beijing Jiaotong University [BJU], Systems Engineering Society of China (SESC) Open access under CC BY-NC-ND license.

Keyword: short-term traffic counts (STTCs); AADT estimation; group assignment; historical data; bayesian

* Corresponding author. Tel.: +1-506-452-6324; fax: +1-506-453-3568.
E-mail address: ming@unb.ca.
1. Introduction

The importance of reliable estimates of travel demand for effective planning, design, and management of roads and facilities is well known by transportation engineers. Project selection, pavement design, capacity analysis, safety analysis, air quality, and traffic simulation are the six applications of traffic data identified in the AASHTO Guideline for Traffic Data Programs (AASHTO, 1992). Traffic statistics such as Annual Average Daily Traffic (AADT), Vehicle Miles Traveled (VMT), Design Hourly Volume (DHV), and Average Daily Vehicle Distance Traveled (ADVDT) are important parameters used by most transportation agencies in their projects. These agencies commit significant financial and human resources to collecting traffic data and estimating these parameters.

The Traffic Monitoring Guide (FHWA, 2001) describes two basic components recommended for Traffic Monitoring Programs (TMPs): continuous count program and short-term count program. The continuous count programs measure the traffic flow or volume using Permanent Traffic Counters (PTCs). Note, in some cases, TMPs are used to collect vehicle classification and weight information. These data are collected at 15-min intervals and stored at 1 hour intervals, for 365 days of the year. Information provided by these counters is used to study the temporal variations of traffic volume like time-of-day, day-of-week, and seasonal traffic patterns on the roadways, which will be used to convert Short-Term Traffic Counts (STTCs) into AADTs.

Short-term count programs are used for covering a large number of road segments without PTCs as a comprehensive coverage program. And this is why short-term traffic counts are often called “coverage counts”. In practice, one short-term count is usually collected for a road segment every few years and the collection periods usually vary from 1 to 7 days. Due to the variation in traffic volume, the recommended minimum counting period is 48 hours for rural roads and 24 hours for urban roads (Leduc, 2008). Expansion factors are required to convert these STTCs to AADTs. The factors are based on the traffic seasonality so a method is required to find a PTC (or a PTC group) with similar seasonal pattern to the STTC site to derive and apply the appropriate factors.

The literature review indicates that most transportation agencies use FHWA factoring methods to estimate AADTs from STTCs, and in this process they simply use road functional class as the criteria to assign short-term traffic counts to Automated Traffic Counter (ATR) or Permanent Traffic Counter (PTC) factor groups (FHWA, 2001). While this practice is widely used among many highway agencies in North America, it may produce large AADT estimation errors in some cases due to the fact that one cannot ensure every road within a functional class has a similar seasonal traffic pattern. In mitigating the weakness of the FHWA factor method, a regression based approach has been developed by some Canadian Provinces. The method produces AADT estimates with higher accuracy than the factoring approach. However, in order to carry out the analysis, it requires several STTCs for a particular road segment during a counting year. Therefore, the monitoring cost is much higher than the factoring approach.

Improved methods are proposed in this paper, which use all historical counts collected to date for a road segment to create its seasonal traffic pattern. Two methods and corresponding Bayesian approaches are developed and tested using PTC data from a winter recreational road in Alberta and their results are benchmarked with the Federal Highway Administration (FHWA) method. Following this introduction, a brief literature review is provided. Study data and methodology are introduced hereafter. Study results are then presented. Finally, conclusions are drawn based on the study results.

2. Literature review

The Traffic Monitoring Guide (FHWA, 2001) introduces the Federal Highway Administration (FHWA)
factoring method, which may be one of the most widely used approaches among agencies for estimating AADTs from STTCs. In the guide, the factoring approach is recommended as the following: (1) permanent traffic counters (PTCs) or automatic traffic recorders (ATRs) are grouped based on the highway functional classes (e.g. Expressway, Arterial, Collector…); (2) Sets of traffic volume expansion factors are calculated for each group; (3) a short term count conducted on a road segment is assigned to the corresponding PTC or ATR group based on its functional class; and (4) finally AADT is estimated by applying expansion factors from that particular group to the most recent STTC. The problem with this method is that it does not account for the fact that not all the roads within a functional class are necessary to have similar seasonal traffic patterns. In another words, it is possible that some roads from a given functional class (e.g., arterial) show very different seasonal patterns (e.g. regional commuter and winter recreational). Since the main purpose of expansion factors is to take traffic seasonality into consideration, such a method may produce large AADT estimation errors in certain cases.

Another method is the regression-based approach, which has been used by several Canadian highway agencies. For this method, several short-term counts are collected during the sampling or counting year for a road segment, and these short-term counts are assigned to a permanent traffic counter based on regression analysis. The aim of this approach is to find a permanent counter whose traffic variation is similar to that of the short-term counting road segment by comparing the above STTC data to the permanent counters data from the same periods. By using the least square linear regression method, for each pair of permanent and short-term counts, the coefficient of determination, R^2, is calculated, and the pair with the highest values of R^2 is selected and assigned to each other; and the regression model developed in the process is used to estimate AADT in the next step (Robichaud and Gordon, 2003).

It is reported that the regression method used in British Columbia produced AADTs that have a lower accuracy than the factoring approach, since the short-term counts are only collected during one season of the year. However, it produces AADTs with higher accuracy for the province of New Brunswick and Prince Edward Island, where counts from several more days in all four seasons are collected (Robichaud and Gordon, 2003). Concerning the data collection, the regression approach requires several counts over four seasons to match a short-term counting site to a PTC or an ATR site. This may not be suitable for most agencies as most of them use FHWA method and they only count a road segment once (48-hour) every a few years. Obviously, the regression method requires additional resources to implement.

In addition to above practices, Davis and Guan (1996) proposed a method for systematically assigning short-term counts to traffic pattern groups. In their study, an optimal fourteen day data collection plan is introduced within the sampling year. 14 counts need to be collected on specific days designated by their method and a Bayesian approach is used to assign road segments to appropriate factor groups. The results show that the method provides AADT estimates that are, on average, within the ±5 percent error of the actual AADT, and in the worst case within ±20 percent. Since the optimal counting plan is very difficult to implement, the recommended method is to collect two counts of one full-week in different months. Although the method provides a “data-driven” approach for assigning short-term counts to factor groups, it would still be difficult for most jurisdictions to change their current monitoring programs from counting a road segment 48-hour every few years to collecting the required two one-week counts for each STTC site. And similar to the regression method, this method requires more human/financial resources than most existing traffic monitoring programs.

Sharma et al. (2001) compared artificial neural networks and the FHWA factor approach for estimating AADTs from short-term traffic counts on low volume roads. The results showed that, when only one single 48-hour short-term count is used, the factoring method has better results than the neural network approach under the assumption that we have perfect knowledge of assigning short-term counts to their “home” PTC or ATR groups. When two short-term counts are used, the results from the neural network and factoring methods are close to each other. But the advantage of the neural network method is
that it does not require the ATR grouping and STTC assignment. On the other hand, the accuracy of the factoring method depends on the appropriate group assignment and in practice this assignment is subjective in nature. However, the problem with the neural network approach is that it is difficult for practitioners to understand and implement.

**Fig. 1. Study area (Alberta, Canada) and locations of PTCs**

### 3. Study methodology and data

A review of current short-term traffic monitoring practices indicates that most transportation agencies simply use road functional class as the criteria for assigning short-term traffic counts to PTC factor groups (FHWA, 2001). A few others use methods such as regression analysis which require extra data collection efforts. Further, all historical counts collected prior to the most recent year are ignored. It should be noted, however, historical counts are likely to contain important information related to traffic seasonality and growth trend for a particular monitoring site. The proposed method, therefore, will use all counts collected to date (from a given short-term traffic monitoring site) to better understand its seasonal traffic pattern. The seasonal pattern built based on all historical counts will provide useful information for improving assignment of short-term counts (to factor groups) and resulting AADT estimations.

In the proposed method, counts from previous years are used to create a seasonal traffic pattern. These collected counts are stored in a database based on day of the week and month of the year. Growth factors calculated from a nearby PTC are applied to these historical counts to convert them to the present traffic volumes. After each new short-term count is collected, it will be added into the dataset, and gradually after several collection cycles an estimated seasonal traffic pattern for the STTC site can be developed.
To demonstrate this method, data from 357 permanent traffic counters between the years 2002 and 2009 from the province of Alberta, Canada are used in this study. Short-term counts are simulated from the PTC # 50431450 located on a winter recreational road and the other PTCs are used as the candidates of best-match PTC of the simulated short-term counting site. The study area, all PTCs used (marked by black triangles) and the simulated STTC site (the location of PTC # 50431450) (marked by the golden star) are shown in the Fig. 1. Please note that this PTC is located on an expressway close to the Rocky Mountains with an AADT of 8,174 vehicles/day in 2009. The simulated short-term counts are then used to build the seasonal traffic pattern of the study site.

Prior to matching patterns, 48-hour short-term counts are converted to the monthly average daily traffic (MADTs) of the most recent year by using a growth factor derived from a nearby PTC and a “Day-of-the-Month” factor from a nearby PTC located on the same functional class road. Please see Bagheri (2011) for the rationales of using the factors in such ways. And then the seasonal variation at PTC or ATR sites is compared to that of the STTC site using a COV method introduced below. In this process, ratios of the monthly average daily volumes from a STTC site and those from corresponding months of each PTC are calculated, and the coefficients of variation of these ratios are computed using Equation 1 and 2 below. Then the PTC or PTC group with the smallest COV is selected as the best-match and its data is then used to develop expansion factors for estimating AADT from the most recent STTC collected.

\[
\text{Ratio}(i) = \frac{\text{MADT}_{\text{STTC}, i}}{\text{MADT}_{\text{PTC}, i}} \quad \text{for those months where STTCs are collected} \quad (1)
\]

\[
\text{COV} = \frac{\sigma(\text{Ratio}(1), \text{Ratio}(2), \ldots \text{Ratio}(n))}{u(\text{Ratio}(1), \text{Ratio}(2), \ldots \text{Ratio}(n))} \quad (2)
\]

It should be noted that the above calculations are repeated after each time a new STTC is collected and the seasonal pattern at the STTC site is extended. Once the above COV Ratio is calculated for each pair of STTC/PTC site, Bayesian algorithm can then be used to update the posterior probability of “best-match PTC” for the STTC site under consideration. Fig. 2 shows how Bayesian algorithm is used in this study.

Bayesian algorithm used in this study works as the following.

(1) Assign an initial prior probability to each PTC as a starting point using the Equation (3) below. If only one 48-hour count is collected for a STTC site, then the probability of a PTC being the best-match to the STTC site is 1/n, where n is the number of PTCs or PTC groups being considered:

\[
P(\text{PTC}_i) = \frac{1}{n} \quad (3)
\]

(2) Once a new STTC is collected, its seasonal pattern will be extended. The likelihood function of a given PTC being the best-match to the STTC site is then updated as.

\[
\begin{align*}
\frac{P(B \mid \text{PTC}_1)}{P(B \mid \text{PTC}_1)} & = \left( \frac{\text{COV}_1}{\text{COV}_1} \right)^2, \\
\frac{P(B \mid \text{PTC}_2)}{P(B \mid \text{PTC}_2)} & = \left( \frac{\text{COV}_2}{\text{COV}_2} \right)^2, \\
& \vdots \\
\frac{P(B \mid \text{PTC}_n)}{P(B \mid \text{PTC}_n)} & = \left( \frac{\text{COV}_n}{\text{COV}_n} \right)^2
\end{align*}
\]

Please note that \(P(B \mid \text{PTC}_i)\) = Given the data and seasonal pattern of PTCi, the likelihood of the estimated pattern B of the STTC site under study is simulated from PTCi. In this case, the ratio of likelihoods between the two PTCs is inversely proportional to the square of corresponding COV values calculated using Equation (2). Please also note that the sum of all probabilities (of the STTC site belonging to a given PTC) is equal to “1”, as shown on the top of the Equation (4).
Fig. 2. Bayesian Assignment of a STTC site to individual PTC (group) and AADT Estimation

(3) Given an estimated seasonal pattern at the STTC site, calculate the posterior probability of a given PTC is the best-match using:

\[
P(PTC_i | B) = \frac{P(PTC_i)P(B | PTC_i)}{\sum_{j=1}^{n} P(PTC_j)P(B | PTC_j)}
\]

(5) Please note that \( P(PTC,B) = \) Given the estimated pattern B of the STTC site under study, the probability of PTCi is the best-match of the STTC site.

The above COV Ratio method (with and without incorporating Bayesian algorithm) is used to assign
STTCs into individual PTCs or PTC groups and AADTs are then estimated based on the expansion factors from the best-match PTC or PTC group. The corresponding AADT estimation results based on the FHWA method are also provided for comparison purposes. The AADT estimation accuracy from each of the five methods: (1) COV Ratio/Individual PTC/Non-Bayesian (hereafter referred as COV-1); (2) COV Ratio/Individual PTC/Bayesian (hereafter referred as COV-2); (3) COV Ratio/PTC Group/Non-Bayesian (hereafter referred as COV-3); (4) COV Ratio/PTC Group/Bayesian (hereafter referred as COV-4) and (5) FHWA Method, is quantified using the absolute percent error (APE) based on the following equation:

\[
APE = \frac{|\text{Estimated AADT} - \text{Actual AADT}|}{\text{Actual AADT}} \times 100
\]

4. Study results

In order to test the model, 110 sets of historical short-term counts are simulated using PTC # 50431450 between March and October and over different years (2002-2009). It should be noted that the time period from March to October is used because this is the typical “counting season” in Canadian jurisdictions. Using the five methods introduced above and the simulated historical counts as inputs, 110 seasonal patterns are estimated for the short-term count site and the best-match PTC is found for each of them. Among these PTCs, PTC # 50027610, and 50431050 are selected with much higher frequency than others as the best match for the short-term count site. By plotting the seasonal patterns of these PTCs and that for the short-term count site (which is really that of the PTC #50431450), one can observe the similarity of these patterns as shown in Fig. 3.

Table 1 above compares the AADT estimation accuracy of the above five methods. It can be seen that in general, the COV Ratio methods are more accurate than the FHWA method. For example, the 95th percentile error (P95 error) for the FHWA is nearly double those from COV-3 and COV-4. It is also considerably higher than those from COV-1 and COV-2. These results reinforce our previous speculation that FHWA’s functional-class-based assignment method cannot accurately take traffic seasonality into consideration and therefore it may result in large AADT estimation errors in some cases. In addition, it can also be found that the COV Ratio methods based on PTC groups (COV-3 and COV-4) are superior to those based on individual PTCs (COV-1 and COV-2). This in turn supports the notion in the practice that the expansion factors derived from a PTC group tend to be more stable than those from an individual PTC and consequently the estimated AADT tend to be more accurate.

Fig. 3. seasonal pattern of the STTC site versus the seasonal patterns of the top two best-match PTCs
Table 1. Comparison of AADT estimation accuracy (% error) of the five methods

| Method                      | COV-1 (COV Ratio Method/Individual PTC/Non-Bayesian) | COV-2 (COV Ratio Method/Individual PTC/Bayesian) | COV-3 (COV Ratio Method/PTC Group/Non-Bayesian) | COV-4 (COV Ratio Method/PTC Group/Bayesian) | FHWA Method |
|-----------------------------|----------------------------------------------------|-------------------------------------------------|-------------------------------------------------|------------------------------------------|-------------|
| Minimum                     | 2.21                                               | 2.03                                            | 1.56                                            | 1.74                                     | 1.71        |
| p25                         | 5.04                                               | 5.83                                            | 3.80                                            | 3.80                                     | 3.72        |
| Median                      | 7.68                                               | 11.10                                           | 4.83                                            | 5.30                                     | 7.17        |
| p85                         | 9.49                                               | 13.47                                           | 8.55                                            | 9.64                                     | 13.29       |
| p95                         | 12.90                                              | 19.55                                           | 10.31                                           | 11.27                                    | 21.74       |

It is interesting to notice that the Bayesian-based COV methods (COV-2 and COV-4) produce larger errors in general than the corresponding non-Bayesian-based methods (COV-1 and COV-3). This can be found by comparing the corresponding statistics listed in Table 1. For example, the median and P95 errors of the COV-1 method are all smaller than those from the COV-2. Also, similar results can be observed between COV-3 and COV-4. However, it should be noted that the differences between COV-3 and COV-4 are much smaller than those between COV-1 and COV-2. For instance, the difference between the P95 error of COV-3 and COV-4 is only about 1%, but the difference between the P95 error of COV-1 and COV-2 is more than 6%.

The results and discussions above may indicate that non-Bayesian-based COV methods are better because of higher AADT estimation accuracy achieved. In addition, they are also simpler. However, it should be noted that the marginally higher errors from the Bayesian-based COV methods result from the fact that Bayesian algorithms require several more steps, when compared to the non-Bayesian approaches, to recover from a “wrong assignment”. In addition, the Bayesian-based COV methods proposed in this paper can explicitly consider the “risk or uncertainty” related to assigning STTC to different PTC or PTC groups. The Bayesian framework provided can explicitly show the probability of a given STTC site associated with different PTC or PTC groups, and therefore it can inform traffic engineers the “risk” of assigning STTC site into a particular group. The Figure 4 below shows how the probability distribution of the “Home PTC” of the simulated STTC site (based on data from PTC # 50431450) changes, when the STTC seasonal pattern grows by adding newly collected STTCs over years. It is clear from the Fig. 4 how such a probability distribution changes over the period from Year 2003 (where the probability of the simulated STTC belonging to different PTC is equal to 1/n) to 2009 based on growing seasonal traffic pattern of the STTC site. At the beginning years, such as 2004 and 2005, because there are only a few counts available to build the seasonal pattern of the STTC site, there are high risks for assigning the site to a wrong “Home PTC”. However, as the newer counts are collected at the site and the seasonal pattern is extended, the risk for assigning to wrong PTCs are reduced and only a few PTCs are selected as “Home” at the end of study period, e.g., Year 2008 and 2009. It should be noted that these a few PTCs selected at the end of the study period all have very similar seasonal patterns as the STTC site and therefore the risk of assigning the STTC site to a wrong “Home PTC” is nearly zero.

The Bayesian framework proposed here is capable of not only explicitly considering the risk of “wrong assignment” of STTC site during each counting year, but also visually showing how such probability distributions change from one year to next. The visualization of such changes in the probability distributions will in turn help traffic engineers and managers understand (1) whether the risk of assignment a STTC site to a wrong “Home PTC” is high or not; and (2) if a STTC should be collected in recent years to further reduce such risk. The proposed method provides a similar “data-driven” approach as Davis and Guan (1996), however, without imposing any additional monitoring cost. As it can
Fig. 4. Probability distributions of “Home PTC” over the study period from the Bayesian algorithm
be seen from the Fig. 4 that once a confidence probability threshold is reached (e.g., higher than 0.75 by Year 2007), traffic engineer can confidently assign a STTC site to the “Home PTC” identified. After that, the traffic engineer can take advantage of this firm relationship by further reducing data collection activities at the STTC sites or exploring the seasonal pattern of the “Home PTC” to improve the effectiveness of data collection at the STTC site. For example, “Home PTC” has a large amount of data available to explore its temporal traffic variations. Traffic engineer can use the knowledge discovered from the PTC data to design a more effective data collection plan at the associated STTC sites. For instance, if the “Home PTC” is known to have a “Winter Recreational” seasonal pattern, then it would suggest that collecting short-term traffic counts at skiing season is most effective in capturing traffic variation at the site, which will in turn help improve the AADT estimation accuracy. If it is known to have a “Summer Recreational” pattern, then collecting STTC at a summer month would be more effective.

5. Conclusion and recommendation

The factoring method recommended by the FHWA has been widely used among highway agencies to estimate AADTs from short-term traffic counts (STTCs) or coverage counts for a wide range of roads, which are not possible to be covered by permanent traffic counters (PTCs) or automatic traffic recorders (ATRs). It is speculated that the fact that this method uses the functional class as the criterion to assign STTC sites into PTC groups is inconsistent with the primary goal of applying expansion factors for accounting for traffic seasonality, which is solely based on the similarity of seasonal patterns of the STTC site and corresponding PTC or PTC group. The reason is obvious, that is, road segments from the same functional class could have very different seasonal patterns. Using the expansion factors derived from the PTCs with different seasonal patterns in the process of estimating AADTs from STTCs could result in large estimation errors and the study results from this paper clearly show that.

This paper presents a novel pattern matching method – COV Ratio method, which uses all available historical STTCs to date for estimating AADTs. The results show a significant improvement in the AADT estimation with a P95 error of less than 12% for PTC group-based COV Ratio methods, in comparison to 21.74% from the FHWA method. It is found that, in general, the methods based on PTC groups (COV-3 and COV-4) are more accurate than those based on individual PTCs (COV-1 and COV-2) and non-Bayesian methods (COV-1 and COV-3) are better than corresponding Bayesian methods (COV-2 and COV-4). Nevertheless, the Bayesian framework provided in this paper can explicitly show the probability of a given STTC site being assigned into different PTC or PTC groups, and therefore it can inform traffic engineers the “risk” of assigning STTC site into a particular group. Further, it provides a “confidence level” information to traffic engineering about their STTC assignments. This information can in turn be used to guide a more cost-effective STTC collection plan.

It should be noted that better accuracy of the proposed methods is achieved without imposing any additional cost or any change to the current data collection plans, which makes it superior than other approaches proposed in previous research.

Acknowledgements

The authors are thankful to the Alberta Transportation department for the data used in this study and the Natural Science and Engineering Research Council (NSERC), Canada for the financial support.
References

AASHTO. (1992). AASHTO Guidelines for Traffic Data Programs. American Association of State Highway and Transportation Officials.

Davis, G. A., Guan, Y. (1996). Bayesian assignment of coverage counts location to factor groups and estimation of mean daily traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 1542, pp. 30–37.

FHWA. *Traffic Monitoring Guide*. U.S Department of Transportation, Federal Highway Administration, Office of Highway Policy Information, 2001.

Robichaud, K., Gordon, M. (2003). An assessment of data collection techniques for Highway Agencies. *Transportation Research Record: Journal of the Transportation Research Board*, 1855, pp. 129–135.

Sharma, S. C., Lingras, P., Liu, G.X., Xu, F. (2001). Application of neural networks to estimate AADT on low-volume roads. *Journal of Transportation Engineering*, 1719, pp. 426–432.