Transportation Planning and Technology

ISSN: 0308-1060 (Print) 1029-0354 (Online) Journal homepage: http://www.tandfonline.com/loi/gtpt20

Empirical analysis of short period traffic counts and their efficiency: the case of Indian traffic

Pranamesh Chakraborty & Partha Chakroborty

To cite this article: Pranamesh Chakraborty & Partha Chakroborty (2017): Empirical analysis of short period traffic counts and their efficiency: the case of Indian traffic, Transportation Planning and Technology, DOI: 10.1080/03081060.2017.1340021

To link to this article: http://dx.doi.org/10.1080/03081060.2017.1340021

Published online: 19 Jun 2017.

TRANSPORTATION PLANNING AND TECHNOLOGY, 2017
https://doi.org/10.1080/03081060.2017.1340021

Empirical analysis of short period traffic counts and their efficiency: the case of Indian traffic

Pranamesh Chakraborty and Partha Chakroborty

Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur, India

ABSTRACT
Short period traffic counts (SPTCs) are conducted routinely to estimate the annual average daily traffic (AADT) at a particular site. This paper uses Indian traffic volume data to methodically and extensively study the effect of four aspects related to the design of SPTCs. These four aspects are: (i) for how long, (ii) on which days should SPTCs be carried out, (iii) how many times, and (iv) on which months should SPTCs be carried out? The analyses indicate that the best durations for conducting SPTCs are 3 days (starting with a Thursday) and 7 days, for total traffic and truck traffic, respectively. Further, these counts should be repeated twice a year keeping a separation of two months between the counts to obtain good estimates of AADT at minimal cost. An additional outcome of this study has been the determination of seasonal factor values for roads in developing economies, like India.

ARTICLE HISTORY
Received 31 December 2015
Accepted 15 May 2017

KEYWORDS
Short period traffic count; annual average daily traffic; traffic volume data; seasonal factors; developing economies; case study

Introduction

Traffic volume counts are an essential part of highway planning programs. Traffic volume data are used for a variety of purposes, including estimation of road revenues, estimation of loads for pavement design and maintenance planning, and forecasting of vehicle emissions. Each of these types of studies requires annual traffic data.

The locations where traffic volume data are recorded continuously throughout the year are known as permanent traffic counters (PTCs). Installation of a PTC in each and every road section is neither economically feasible nor is it required. The general practice is to collect short period traffic counts (SPTCs; say, for 7 days, 3 days, etc.) on road segments and to adjust them using Seasonal Factors (*SFs*) to predict the annual average daily traffic (*AADT*). The *SF* values are obtained using PTC data from similar road segments.

The primary purpose of this study is to see how SPTCs can be made more efficient as estimators of *AADT*. Ideally *AADT* should be obtained by summing the flows on all the days of the year. SPTC, in that sense, represents a sample and like all samples is prone to errors. The idea here is to see how this sampling process can be improved so as to obtain estimates of *AADT* with greater accuracy with minimum deployment of resources to conduct the SPTCs.

CONTACT Pranamesh Chakraborty pranameshbesu@gmail.com Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur-208016, Uttar Pradesh, India

The important questions that need to be answered before conducting any SPTC are: (i) for how long should the data be collected, that is, duration; (ii) when should the data be collected (which days of a week, which months, etc.), that is, time and (iii) how many times a year should the data be collected, that is, frequency, so as to obtain accurate estimates with minimal resources? In other words, how can the duration, time and frequency of SPTCs be decided so that the resources for data collection are most effectively utilized?

This paper uses Indian data to analyse: (i) the impact as well as marginal impact of duration on accuracy, (ii) the effect of days of the week in which data is collected (for a given duration) on accuracy, (iii) the impact as well as marginal impact of frequency on accuracy, and (iv) the effect of months of the year in which data is collected (for a given frequency) on accuracy. The analysis shows that duration, time and frequency can be chosen in a way so as to gain in accuracy at no extra cost.

An additional outcome of this study has been the determination of *SFs* for roads in developing economies like India. It may be noted that significant differences in travel patterns exist between developed and developing economies.

Literature review and motivation

This section provides a brief overview of past research on the determination of *SFs* (note, seasonal factors are required to convert SPTC data to *AADT*) and the best frequency, time and duration of SPTCs. It also presents the motivation for the present work.

The *Traffic Monitoring Guide* (TMG) of the US Federal Highway Administration (FHWA) provides the guidelines for estimation of *AADT* using PTC and SPTC data. The FHWA procedure (FHWA 2013) consists of four basic steps:

(1) Identification of groups of PTC sites having similar temporal traffic volume variations;
(2) Determination of average seasonal adjustment factor for each road group;
(3) Assignment of the SPTC site to one of the groups defined in step 1 and
(4) Estimation of *AADT* of the required road section using appropriate seasonal adjustment factor.

The TMG recommends three methods for classification of road groups using PTC data: geographical/functional classification, cluster analysis and 'same road' application of adjustment factors. A number of methods have been reported in the literature for classifying PTCs into different homogeneous groups and using their data to determine *SFs* for that group. The four most common methods used for this purpose are: (a) cluster analysis (Sharma 1983; Ritchie 1986; Flaherty 1993); (b) multiple regression analysis (Hallenbeck and Kim 1993; Faghri and Chakroborty 1994; Li, Zhao, and Wu 2004; Zhao, Li, and Chow 2004; Yang et al. 2009); (c) artificial neural networks (ANN) (Faghri and Hua 1995; Lingras 1995) and (d) genetic algorithms (Lingras 2001). Zhao and Park (2004) used geographic weighted regression technique (GWR), a development of multiple regression analysis, to estimate *AADT*. More recently, Duddu and Pulugurtha (2013) used spatial variations in land-use characteristics for estimation of *AADT* and Gecchele et al. (2011) presented a comparison of the various clustering techniques used for grouping PTCs.

In order to estimate *AADT* from SPTCs, it is essential to first classify the study sites into different groups and find out the adjustment factors (*SFs*) of each group. Also, travel patterns in developing economies are different from those in developed economies. So it is worthwhile to study *SF* variation on roads of developing economies rather than using the classification of roads obtained in previous literature. Thus one of the aims here is to study flow variations on Indian roads with a view to predict *SFs* for Indian highways and use them for estimation of *AADT* from SPTCs.

Now, moving on to the primary purpose of this study, as previously mentioned, SPTCs are used to estimate *AADT* for sites that do not have a PTC. Hence, the duration and frequency of traffic counts that can accurately as well as economically estimate *AADT* also need to be determined.

TMG recommends conducting SPTCs of 2-day duration on weekdays. However, it also states that longer duration counts (3 or 7-day) are encouraged based on the availability of resources. Sharma et al. (1996a) compared 1-, 2- and 3-day traffic counts on weekdays to determine the best duration. They concluded that the error for 2-day durations is quite similar to that for a 3-day duration, except for recreational roads. This work, however, does not specify which 2 days in a week are best for traffic counts. Recently, Gecchele et al. (2012) used ANN to assign the 1-, 2- and 3-day SPTCs to the correct PTC group. They concluded that SPTCs should be undertaken on weekdays rather than at weekends and the improvement in taking 3-day counts rather than 2-day counts is not statistically significant. The issue of which days are better for SPTCs was also analysed by Hallenbeck and Kim (1993) who found that truck traffic counts on Thursdays are better compared to Tuesdays and Wednesdays. This conclusion was drawn at an aggregate level by combining the data from all study sites, and hence does not indicate whether the conclusion is valid for every site in their database. Combining data from all sites may create a situation where the chosen strategy may be very good for one site, but not so good for others; while another strategy which is reasonable for many sites may not be chosen. Since *AADT* estimates are site-specific, a strategy that is good for individual sites should be chosen rather than a strategy that may be very good for some sites, while not good for many others.

Traffic counts can also be undertaken multiple times a year. TMG suggests that in order to capture the seasonal variability of road sections, traffic counts should be undertaken in different periods of the year. Lingras (1998) used artificial neural networks to show that two 2-day counts (in July and December) were better than a single 7-day count (in July or December). Other researchers have also tried to determine the number of times in a year traffic counts need to be undertaken for accurate estimation of *AADT* (Sharma and Allipuram 1993; Sharma et al. 1996b). In these studies, traffic counts that are equally spaced over a year are used. For example, in their studies for counting traffic twice a year, the two counts are always separated by 182 days (or approximately six months). Recently, Gastaldi, Gecchele, and Rossi (2014) used one-week seasonal counts taken on consecutive months (i.e. Jan–Feb, Mar–Apr, etc.) for estimation of *AADT* using a Fuzzy C-means algorithm to represent the fuzzy boundaries of road groups.

The literature shows that over the last two decades, there have been, at best, sporadic efforts to analyse the impact of duration, time and frequency of SPTCs on the accuracy of *AADT* estimates. Further, none of these studies carried out systematic and detailed analyses of the data with a view to identifying the best strategy for SPTCs, based on site-specific performance levels.

The motivation here is to systematically study various aspects of SPTCs and their impact on developing a cost-effective and reliable strategy to predict *AADT*. More specifically, in order to obtain sound *AADT* estimates, this study attempts to determine: (i) the number of days for which SPTCs should be undertaken; (ii) if applicable, the days of the week on which SPTCs should be undertaken; (iii) the number of times in a year SPTCs should be carried out; and (iv) if applicable, the months in which SPTCs should be carried out.

It may be pointed out that Hallenbeck and Kim (1993) had noted that *SFs* for truck traffic and automobile traffic are not necessarily the same. In this paper, however, instead of considering truck traffic and car traffic separately, analyses have been undertaken separately for truck traffic and total traffic. The reason for doing this is that predictions on truck traffic data are often required for pavement design and maintenance studies, while information on total traffic is required for traffic engineering and planning purposes.

Analysis and results

In this section, the analyses of best duration (and if applicable, which days of the week) and frequency (and if applicable, which months) for SPTCs for total and truck traffic are presented. Since *AADT* is predicted from SPTC data using *SFs*, *SFs* for Indian roads are estimated first.

In the analyses, daily traffic volume data from toll plazas located in different parts of India have been used. Since these toll plazas collect volume data with vehicle classes throughout the year, they are considered as PTCs. In this study, data from 21 such PTCs on multi-lane National Highways (NHs) and State Highways (SHs) of India, in rural as well as urban settings, are used. The traffic data were collected in the period April 2010–March 2014. However, the duration of traffic data available for different PTCs are different (minimum 1 year and maximum 4 years). The details of the duration of traffic data are given in Table 1. The first column gives the duration (in years) and the second gives the number of sites having the corresponding duration. Table 2 gives the descriptive statistics (mean, standard deviation, minimum and maximum) of the data. It includes *AADT*, annual average daily truck traffic (*AADTT*) and percentage of truck volume, which is defined as the ratio of *AADTT* and *AADT*.

Before using the data from the PTCs, each data set has been individually studied to identify possible outliers or abnormalities either due to data recording errors or due to occurrence of out-of-the ordinary events. For example, since a PTC collects traffic volume data throughout the year, it also includes special days like national holidays, days when public transport was off the roads, days within a period of consecutive holidays, festival days, election days, etc. In this study, these special days have been identified for

Table 1. Duration of traffic data available.

Duration (years)	Number of sites
1	8
2	3
3	7
4	3

Table 2. Descriptive statistics.

	AADT	*AADTT*	% of truck volume
Max	48840	17751	86.5
Min	3553	1184	10.3
Mean	17696	8285	52.6
Std. Dev.	10839	4444	18.5

each PTC. Statistical tests (*t*-test) concluded that there is no evidence of statistical difference between the *AADT* computed over the entire year and the same computed excluding the aforementioned special days. So traffic data for the entire year have been used to determine the *AADT* of that particular year and for further analyses. However, while analysing relationships between SPTCs and *AADT* at a site, the SPTCs avoid these special days.

The following three subsections provide results for *SF* estimation, analysis for the duration of SPTCs and analysis for the frequency of SPTCs, respectively.

Estimation of seasonal factors of traffic

Seasonal factors are required to estimate *AADT* of a site from its SPTCs. Seasonal factor ($SF_j^{m,k}$) of month m for site j in year k is defined as the ratio of monthly average daily traffic ($MADT_j^{m,k}$) to annual average daily traffic ($AADT_j^k$) and is given by:

$$SF_j^{m,k} = \frac{MADT_j^{m,k}}{AADT_j^k}. \tag{1}$$

Three methods are investigated to predict seasonal factors from the PTC data. Three methods are used for estimation of *SFs* of traffic: the average seasonal factor method; cluster analysis and multiple regression analysis.

In the average seasonal factor method, the most simplistic of the three, it is hypothesized that seasonal variations on all types of Indian highways are same. In this method, it is proposed that the *SFs* on any Indian road for month m can be estimated with reasonable accuracy as the average of the seasonal factors for that month obtained from all the PTCs. The average seasonal factor (ASF^m) for month m can be obtained as:

$$ASF^m = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} SF_j^{m,k}}{\sum_{j=1}^{J} K_j}, \tag{2}$$

where $SF_j^{m,k}$ is obtained from Equation (1), J is the total number of PTCs (21 in this study) and K_j is the number of years for which traffic data are available at the jth PTC.

The seasonal factors obtained from this method for total and truck traffic are given in Table 3.

Next, the assumption that all Indian highways have the same seasonal variations is done away with and cluster analysis and multiple regression analysis are used to

Table 3. Seasonal factors obtained using the average seasonal factor method.

	Month											
Traffic type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Total traffic	1.03	1.07	1.04	0.98	1.02	1.00	0.94	0.94	0.94	0.98	1.02	1.05
Truck traffic	1.01	1.07	1.06	0.99	0.99	0.99	0.96	0.94	0.97	0.98	1.00	1.04

estimate *SFs*. However, for the current data set the results from these two methods do not give any significant improvement over the prediction accuracies obtained using the average seasonal factor method. Further, in cluster analysis, the clusters were difficult to define as no clear pattern emerged. In regression analysis, no parameter came out to be statistically significant for all months. [Full details of the results can be found in Chakraborty (2014).] Given that the estimation errors obtained using the average seasonal factor method are of the same order as those obtained in previous studies (Faghri and Chakroborty 1994; Faghri and Hua 1995; Li, Zhao, and Wu 2004; Yang et al. 2009) and that the average seasonal factor method is comparatively simpler than cluster analysis and regression analysis, it is proposed that the average seasonal factor method be used to determine the *SFs* for different months at each of the sites.

Before leaving this section, it needs to be noted that the data set on which the analysis is based consisted of national and state highways (with none leading to recreational sites). It is quite possible that all the sites actually belong to a single cluster and show similar seasonal variations. There could have been distinct patterns in *SFs* (and hence different groups) if the data set included arterials, recreational roads, etc. Unfortunately, such data do not exist for India. In any case, since the primary purpose of this paper is to analyse the effect of duration and frequency of SPTCs on *AADT* estimation, it is considered that the available data suffice. As will be seen later, at least in a limited sense, the analysis shows that it is possible to choose duration and frequency of SPTCs judiciously so as to gain in accuracy (of *AADT* estimates) at little or no extra effort towards conducting SPTCs.

Duration and frequency of SPTCs

As explained earlier, SPTCs are used for estimation of *AADT* for sites that do not have a PTC. The two primary decisions that need to be taken while designing an SPTC exercise are the duration of the count (i.e. for how long traffic count has to be done) and the frequency of the count (i.e. the number of times such counts need to be done in a year) to predict *AADT* accurately. Two associated decisions that also need to be taken are on which days and on which months should these counts be taken. This section presents analyses to determine the most effective duration, days, frequency and months for SPTCs for total and truck traffic.

In any sampling exercise, accuracy of predictions increases as sample size increases. Similarly, as the duration and frequency of SPTCs increase, the error in *AADT* estimation decreases. However, the cost of conducting the survey also grows with increases in duration and frequency of SPTCs. The main objective, therefore, is to discover a duration and frequency (with information on days and months) that is not prohibitively costly while providing estimates of *AADT* with reasonable accuracy.

Before presenting the analysis on determining effective duration and frequency of SPTCs, the yardsticks used to determine the cost of the survey and accuracy of the *AADT* estimates are presented. The cost of conducting a survey is assumed to increase with the number of days over which the survey is conducted. Hence, reducing the duration lowers the cost. Cost rises with increase in frequency of SPTCs too, so reducing the frequency lowers the cost.

In order to determine accuracy, the mean squared error of the deviation in estimated *AADT* from the actual *AADT* is used. In the following, the procedure to obtain the mean squared error is explained.

SPTCs of frequency F, duration of d days and starting on the nth day of the week for a particular site j can be done in various week-sets (w) of the year. Thus, many estimates of *AADT* can be obtained for a given site j and for a particular combination of F, d and n. For example, for $F = 2$, $d = 3$ and $n =$ Tuesday, SPTCs can be conducted in the first week of January and first week of July to get an estimate of *AADT*, or SPTCs can be conducted in the third week of February and second week of May to get an estimate, and so on.

The estimated *AADT* value, $EAADT_j^F(d, n, w)$, for site j when SPTCs have a frequency of F, a duration of d days, starts on the nth day of the week and undertaken in the wth week-set of the year, is given by:

$$EAADT_j^F(d, n, w) = \frac{1}{F}\sum_{f=1}^{F}\frac{ADT_j^f(d, n, w)}{PSF_A^{m_f}}, \tag{3}$$

where $ADT_j^f(d, n, w)$ is the average daily traffic of site j obtained from the fth traffic count of duration d days starting on day n and done on the wth week-set of the year; $PSF_A^{m_f}$ is the predicted *SFs* of month m^f (in which the fth traffic count is undertaken) as predicted by the average seasonal factor method. In this study, the $PSF_A^{m_f}$ are obtained from Table 3.

The deviation in the estimated *AADT* is defined as:

$$D_j^F(d, n, w) = \left(\frac{EAADT_j^F(d, n, w) - AADT_j}{AADT_j}\right) \times 100, \tag{4}$$

where $AADT_j$ is the actual *AADT* obtained from the entire year's traffic count at the same site.

Now, for a given site j and a given combination of F, d and n, there are various values of $D_j^F(d, n, w)$ that can be obtained depending on which weeks the SPTCs are undertaken. If it is assumed that a given F, d and n should in theory give a single true value, then all the values that are obtained can be thought of as the true value with some stochastic error arising out of sampling. It can further be assumed that the true value of this deviation $D_j^F(d, n, w)$ is zero (that is, the process is unbiased). The mean squared error of the estimate of $AADT_j$ (given by Equation (3)) can therefore be obtained as:

$$MSE_j^F(d, n) = (\overline{D_j^F(d, n)})^2 + \frac{1}{W_j^F(d, n) - 1}\sum_{w=1}^{W_j^F(d,n)}(D_j^F(d, n, w) - \overline{D_j^F(d, n)})^2 \tag{5}$$

where $MSE_j^F(d, n)$ is the mean squared error of the estimated *AADT* for site j for a given combination of F, d and n; $W_j^F(d, n)$ is the number of estimates of $AADT_j$ obtained for the same site j and the same combination of F, d and n; and $\overline{D_j^F(d, n)}$ is the mean of all the instances of $D_j^F(d, n, w)$ found using Equation 4. This mean value is obtained as:

$$\overline{D_j^F(d, n)} = \frac{\sum_{w=1}^{W_j^F(d,n)} D_j^F(d, n, w)}{W_j^F(d, n)}. \tag{6}$$

Thus, in this study, the cost of doing the SPTCs is assumed to be proportional to the duration and frequency of SPTCs and the accuracy of *AADT* estimates is taken as $MSE_j^F(d, n)$.

Analysis to determine the most effective duration of SPTCs

This section describes the methodology for determining the best duration of SPTCs. The following values of d are investigated in this study: 14, 7, 5, 3 and 2 days. The values of n (recall n stands for starting day of the SPTCs) used in this study for 5-, 3-, and 2-day SPTCs are Monday, Tuesday, Wednesday, Thursday, Friday, Saturday and Sunday. For 7- and 14-day SPTCs, the value of n is always assumed to be Monday. Thus, an SPTC for $d = 3$ and $n =$ Tuesday means that flow data are collected on Tuesday, Wednesday and Thursday while the SPTC for $d = 7$ and $n =$ Monday means data are collected on Monday through Sunday. Note for $d = 14$ and $d = 7$, n is not really a decision variable, since no matter which day of the week the count starts, each day is represented twice (for $d = 14$) or once (for $d = 7$) in the SPTC data set.

The process to determine the best d proceeds in two steps. First, for every d the best n is determined. Next, the best value of d is determined assuming that for every d the corresponding best value of n will be used.

The methodology to determine the best value of n for total traffic and truck traffic for $d =$ 2, 3 and 5 days is presented next. Note that for 7 and 14-day counts, n is not a meaningful decision variable and is always taken as Monday. Also, during this determination, F is assumed to be 1. In order to determine the best n-value, two measures have been used – the average *MSE* and average ranks of *MSE*. These are explained next.

Average MSE, $AMSE^1(d, n)$: This is defined as the arithmetic mean (average) of the $MSE_j^F(d, n)$ over all the sites, j:

$$AMSE^1(d, n) = \frac{1}{J}\sum_{j=1}^{J} MSE_j^1(d, n), \tag{7}$$

where J is the total number of sites. The measure indicates the average performance of a particular d and n. However, this measure is an aggregate measure and can be low for a particular d and n because that d and n (which will be chosen) is very good for a few sites and not so good for many others. Thus, use of this d and n for a site for which it is not very good will yield large errors. Ideally, a measure that reflects the performance of d and n for individual sites is better. The next measure is one such.

Average rank of MSE, $ARMSE^1(d, n)$: For a given site j and a given value of d, the $MSE_j^1(d, n)$ are ranked in ascending order over the various values of n. Let these ranks, for a given value of d and n, be $RMSE_j^1(d, n)$. Then the average of these ranks over all js for a particular d, $ARMSE^1(d, n)$, given by Equation (8), gives a measure of how well a given value of n fares. The lower the value of $ARMSE^1(d, n)$, the better the performance of that n for the given d. If for a particular value of n, (for a given d), the $MSE_j^1(d, n)$ is the smallest (best) for each site, then $ARMSE^1(d, n)$ takes a value of unity. If on the other hand, it is worse for each of the J sites, then it takes a value of seven because n has seven values:

$$ARMSE^1(d, n) = \frac{1}{J}\sum_{j=1}^{J} RMSE_j^1(d, n). \tag{8}$$

These two measures are used to determine n_b, the best n for a given duration d. Table 4 shows n_b obtained using the two measures for all durations (except for $d = 7$ and 14 days) of SPTCs. A discussion on the results follows.

For total traffic, Table 4 shows that for $d = 2$, the best value of n (n_b) is Thursday since it offers the lowest $AMSE^1(2, n)$and lowest $ARMSE^1(2, n)$. Similarly, for $d = 3$ and 5 days, the best values of n are Thursday and Wednesday, respectively. The analysis indicates that Thursday and Friday are important days for collecting total traffic data irrespective of whether $d = 2$, 3 or 5. Interestingly, when the duration is increased by one day, (i.e. d is increased to 3 from 2 days), the analysis suggests adding Saturday to Thursday and Friday for conducting SPTCs. When two more days are added (i.e. d increases from 3 to 5 days), the analysis suggests adding a day on either side of the Thursday to Saturday span (found to be best for $d = 3$). Figure 1 describes this observation schematically.

For truck traffic, a figure similar to Figure 1 is presented in Figure 2. Unlike for total traffic, the best durations for truck traffic obtained from the two measures do not match exactly. However, as for total traffic, two points emerge: (i) Monday and Tuesday are important days and (ii) as d is increased, the spans for data collection become supersets of the smaller spans (obtained for the smaller durations). Note, the observations on the nested nature of the best days for data collection is valid for $d = 7$ and 14 days too. This is so because for these durations all the days of the week are in any case included for data collection.

Table 4 also shows that for each d and n_b, the $ARMSE^1(d, n_b)$ values are close to unity for total traffic, but are around 2.5 for truck traffic. This indicates that for total traffic, the d and n_b combination is good for each of the sites. While for truck traffic, the $ARMSE^1(d, n_b)$ indicates that, even for the best n for a given d, the estimates are not uniformly good for all the sites. This, to some extent, reduces the confidence in the results on n_b obtained here for truck traffic. Note that using $AMSE^1(d, n_b)$ only would not have brought out this limitation so clearly.

So far the discussion has concentrated on the best n for a given d. Next, the best value of d for total and truck traffic is determined. As discussed before, it is expected that as d increases the accuracy of $AADT$ estimates will also increase. Therefore, in order to determine the best value of d, the percentage improvement in $AMSE^1(d, n)$ per extra day of data collection is used. For a given value of d, the value of n used in this analysis is that of n_b under the $ARMSE^1(d, n)$ column of Table 4.

Note that in the analysis $ARMSE^1(d, n)$ is not used since the rate of change in $ARMSE^1(d, n)$ with d may not be well behaved; this is so because, even when accuracy improves with increased duration, the rank may not change and when it changes it may change due to only a small improvement in accuracy. For example, if a particular

Table 4. Best n (n_b) of total and truck traffic for all durations.

		$AMSE^1(d, n)$		$ARMSE^1(d, n)$	
Traffic type	d (days)	n_b	$AMSE^1(d, n_b)$	n_b	$ARMSE^1(d, n_b)$
Total traffic	2	Thu	66.4	Thu	1.4
	3	Thu	61.1	Thu	1.2
	5	Wed	58.3	Wed	1.6
Truck traffic	2	Mon	91.8	Mon	2.8
	3	Mon	89.6	Sun	2.6
	5	Sat	84.8	Sun	2.4

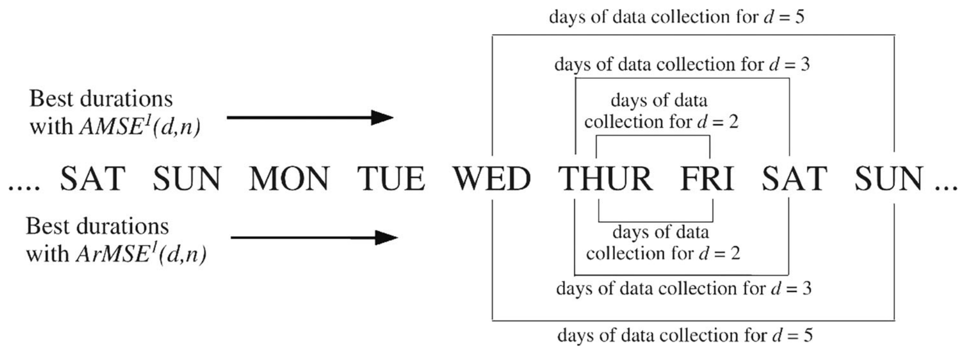

Figure 1. Days for which SPTCs should be conducted for total traffic obtained using $AMSE^1(d, n)$ and $ARMSE^1(d, n)$.

prediction is best and continues to remain the best while improving in accuracy, then it will appear as if d has no influence on the accuracy if $ARMSE^1(d, n)$ is used. Hence, while deciding on the best d, the measure that better reflects the change in accuracy with d is used.

Table 5 presents the $AMSE^1(d, n_b)$ values for all ds. The first column in the table gives the different values of d. The total and truck traffic columns are each divided into two sub-columns. The first gives the $AMSE^1(d, n_b)$ values and the second column gives the percentage improvement in $AMSE^1(d, n_b)$ per extra day of data collection.

As can be seen from Table 5, for total traffic, substantial improvement in accuracy (8%) is obtained by moving from a duration of 2–3 days. Subsequent addition of days does not yield commensurate improvements in $AMSE^1(d, n_b)$. Hence, it is suggested that for total traffic, data should be collected for three days starting with Thursday (note, from Table 4, n_b = Thursday for $d = 3$).

For truck traffic, however, the data indicate that reasonable improvement is obtained by adding days till d is equal to 7 days. Beyond 7 days, collecting data on additional days does not yield significant improvements. Hence, it is suggested that, for truck traffic, 7 days data be collected starting with Monday (or any other day).

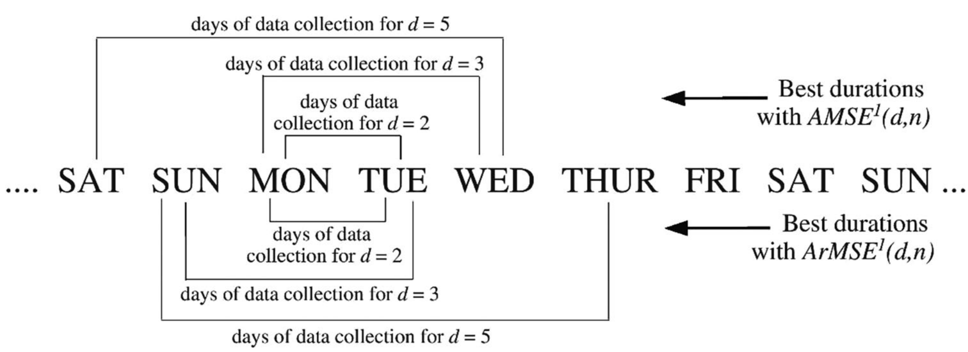

Figure 2. Days for which SPTCs should be conducted for truck traffic obtained using $AMSE^1(d, n)$ and $ARMSE^1(d, n)$.

Table 5. Percentage improvement in $AMSE^1(d, n_b)$ per extra day of data collection.

Duration (d)	Total traffic		Truck traffic	
	$AMSE^1(d, n_b)$	% Improvement per extra day	$AMSE^1(d, n_b)$	% Improvement per extra day
2	66.4	–	91.8	–
3	61.1	$\frac{(66.4-61.1)\times 100}{(3-2)\times 66.4}=8.0$	89.6	$\frac{(91.8-89.6)\times 100}{(3-2)\times 91.8}=2.4$
5	58.3	$\frac{(61.1-58.3)\times 100}{(5-3)\times 61.1}=2.3$	84.8	$\frac{(89.6-84.8)\times 100}{(5-3)\times 89.6}=2.7$
7	56.5	$\frac{(58.3-56.5)\times 100}{(7-5)\times 58.3}=1.5$	76.5	$\frac{(84.8-76.5)\times 100}{(7-5)\times 84.8}=4.9$
14	45.7	$\frac{(56.5-45.7)\times 100}{(14-7)\times 56.5}=2.7$	61.7	$\frac{(76.5-61.7)\times 100}{(14-7)\times 76.5}=2.8$

Notes: $n = n_b$; n_b as per $ARMSE^1(d, n)$ column of Table 4.

In order to compute improvements per extra day of data collection, a fixed benchmark could be used instead of the floating benchmark used in the table. If $d = 2$ is used as the fixed benchmark and calculates the improvement per extra day, then for total traffic the values will be 8.0%, 4.1%, 3.0% and 2.6% for $d = 3$, 5, 7 and 14, respectively; and for truck traffic the values will be 2.4%, 2.5%, 3.3% and 2.7% for $d = 3$, 5, 7 and 14, respectively. From this, it is also clear that the maximum per day improvement is obtained for $d = 3$ in the case of total traffic and $d = 7$ in the case of truck traffic.

It may be noted that, as expected, with increases in the value of d, $AMSE^1(d, n_b)$ improves. Hence, if there are no resource constraints, the duration of SPTCs may be made as long as is feasible to obtain progressively better $AADT$ estimates.

Analysis to determine the most effective frequency of SPTCs

As discussed earlier, the accuracy of the estimated $AADT$ is also impacted by how many times the SPTCs are undertaken in a year. In this section, this dependence of $EAADT_j^F(d, n)$ on F for every value of d is analysed in detail. $F = 1$ means the SPTC is carried out during only one month of a year, $F = 2$ means SPTCs are carried out in two different months and so on. The value of n used (for a given d) in the analysis is the n_b given under the $ARMSE^1(d, n)$ column of Table 4. The results from this analysis are presented here.

The values of $AMSE^F(d, n_b)$ for different values of F and d are calculated. In order to see which value of F is most desirable, the rate of improvement in $AMSE^F(d, n_b)$ for every additional repetition of SPTC is calculated. These rates for every value of d are given in Table 6. The first column gives different values of F (two to six) and the second and

Table 6. Rate of improvement in $AMSE^F(d, n_b)$ with increase in frequency of SPTCs for total and truck traffic.

	Total traffic					Truck traffic				
	d (days)					d (days)				
Freq. (F)	2	3	5	7	14	2	3	5	7	14
2	36.7	33.7	32.1	30.2	26.8	44.1	39.8	38	32.7	25.9
3	11.6	10.6	10.1	9.5	8	14.2	12.1	12.2	10	7.5
4	5.8	5.3	5	4.7	4	7.1	6	6.1	5	3.7
5	3.5	3.2	3	2.8	2.4	4.2	3.6	3.7	3	2.2
6	2.3	2.1	2	1.9	1.6	2.8	2.4	2.4	2	1.5

Notes: $n = n_b$; n_b as per $ARMSE^1(d, n)$ column of Table 4.

third columns give the rate of reduction of $AMSE^F(d, n_b)$ for $d = 2, 3, 5, 7$ and 14 for total and truck traffic, respectively.

The values given in Table 6 suggest that a large improvement (reduction) in $AMSE^F(d, n_b)$ is obtained (for every value of d) by using $F = 2$ instead of $F = 1$. Any further increase in F does not yield similarly large improvements. Hence, it is suggested that, irrespective of the value of d, SPTCs should be repeated twice in a year (i.e. $F = 2$). It should be noted that, as in the case of determining the best value of d, so here, better estimates of $AADT$ can be obtained by continuing to increase F. However, such an increase will cause increased strain on resources. Also, note for reasons same as those enunciated in the previous section, $AMSE^F(d, n)$ is not used while deciding the best F.

Now that the analysis indicates that choosing $F = 2$ is the best, the obvious question is which two-month combination (M) is the best? In order to answer this, $AMSE^2(d, M)$values for different two-month combinations are compared. (Note, the $AMSE^2(d, M)$ for a given two-month combination is simply the average of $MSE^2(d, M)$ values obtained for that two-month combination over all sites.)

The question of which two-month combination is the best (like the question of which n is the best) is also evaluated by taking the average of the ranks over all sites for $MSE_j^2(d, M)$ obtained for different two-month combinations ($ARMSE^2(d, M)$). The $ARMSE^2(d, M)$ measure was explained in detail during the determination of the best n. Note that in this case the best $ARMSE^2(d, M)$ value can be unity and the worst $ARMSE^2(d, M)$ can be 66 (since there are 66 different two-month combinations possible).

Table 7 gives the best two-month combination (denoted as M_b) with respect to $AMSE^2(d, M)$as well as $ARMSE^2(d, M)$. The $AMSE^2(d, M)$ and $ARMSE^2(d, M)$ values of the corresponding M_b are denoted by $AMSE^2(d, M_b)$ and $ARMSE^2(d, M_b)$, respectively. The first column gives the type of traffic (total traffic and truck traffic), the second provides the different values of d (2, 3, 5, 7 and 14) and the third gives the best two-month combination obtained using the $AMSE^2(d, M)$ measure as well as the best value of $AMSE^2(d, M)$. Similarly, the fourth column provides the best two-month combination obtained using the $ARMSE^2(d, M)$ measure as well as the best value of $ARMSE^2(d, M)$.

The table shows that the best two-month combinations from the two measures are in agreement for total traffic for every value of d; however, for truck traffic there is disagreement between $d = 5$ and $d = 14$. The analysis points towards May-September or May-October (depending on the values of d) to be 'good' two-month combinations for total

Table 7. $AMSE^2(d, M_b)$ and $ARMSE^2(d, M_b)$ of total and truck traffic for all durations.

		$AMSE^2(d, M)$		$ARMSE^2(d, M)$	
Traffic Type	d (days)	M_b	$AMSE^2(d, M_b)$	M_b	$ARMSE^2(d, M_b)$
Total Traffic	2	May Oct	9.9	May Oct	4.8
	3	May Sep	9.3	May Sep	8.8
	5	May Sep	7.5	May Sep	7.0
	7	May Oct	10.3	May Oct	9.0
	14	May Oct	2.7	May Oct	2.8
Truck Traffic	2	Jan Oct	11.2	Jan Oct	6.8
	3	Apr May	7.9	Apr May	7.5
	5	Apr May	8.6	Apr Jun	9.5
	7	Mar Jun	12.2	Mar Jun	12.3
	14	Apr Aug	2	Apr Jul	3.5

traffic; however, no such conclusion can be drawn for truck traffic. Further, the $ARMSE^2(d, M_b)$ values are reasonably high; as before, this indicates that the best combination may perform well for some sites but may not be good for others. Hence, it is considered that the analysis of the best two-month combination, though promising (at least for total traffic), remains by-and-large inconclusive.

The goal is therefore modified from trying to determine the best two-month combination to determining the best separation (in terms of months) that should be maintained while conducting SPTCs twice a year. That is, even though the analysis could not determine a particular two-month combination as the best, an attempt is made to see whether analysing the data can indicate how far apart the two SPTCs should be for the best result. This analysis is presented next.

In the previous analysis, both $AMSE^2(d, M)$ and $ARMSE^2(d, M)$ values are evaluated for a particular two-month combination (M). In this case, however, *AMSE* and *ARMSE* values are obtained for cases where SPTCs are conducted in two months separated by the same number of months. These measures are denoted as $AMSE^2(d, S)$ and $ARMSE^2(d, S)$; here S denotes the *separation* between the two months when SPTCs are undertaken. For example, unlike in the previous case, where (January, April) and (August, November) represented two different two-month combinations, in the present analysis they represent the same case (in terms of the value of S) because both (January, April) and (August, November) represent combinations of months separated by a two-month period, that is, $S = 2$. Here analysis is undertaken for the cases where *separation*, S, between the SPTCs is zero month (i.e. SPTCs undertaken in consecutive months), one month (i.e. SPTCs undertaken in (January, March), (February, April), etc.) and so on, until $S = 5$ months (i.e. (January, July), (February, August), etc.). Note, in this case, the smallest value for $ARMSE^2(d, S)$ can be one and the largest value can be six.

Table 8 gives the best separation (denoted by S_b) to be kept between two SPTCs obtained by using $AMSE^2(d, S)$and $ARMSE^2(d, S)$ as measures of accuracy. Table 8 gives the results in a tabular format similar to Table 7 except that instead of M_b it includes S_b.

Table 8 shows that for total traffic, the S_b obtained from $AMSE^2(d, S)$ and $ARMSE^2(d, S)$ do not match for 7- and 5-day durations. However, for truck traffic, the S_b from both approaches are the same for all durations. It is noteworthy that for most of the durations, the best separation turns out to be two months from both the approaches. The results indicate that although it is not possible to determine a particular two-month combination that is best for all sites used in this study, it can, with reasonable confidence,

Table 8. $AMSE^2(d, S_b)$ and $ARMSE^2(d, S_b)$ of total and truck traffic for all durations.

		$AMSE^2(d, S)$		$ARMSE^2(d, S)$	
Traffic type	d (days)	S_b	$AMSE^2(d, S_b)$	S_b	$ARMSE^2(d, S_b)$
Total traffic	2	2	24.3	2	1.5
	3	2	22.8	2	1.5
	5	3	24.3	2	1.5
	7	3	23.1	2	2.3
	14	2	15.7	2	2.0
Truck traffic	2	1	35.7	1	2.0
	3	2	28	2	2.0
	5	2	21.8	2	1.8
	7	2	11.7	2	2.0
	14	2	11.5	2	2.0

be stated that SPTCs are to be undertaken twice a year keeping a *separation* of two months between the counts.

Conclusions

SPTCs are used to estimate *AADT* for sites that do not have a PTC. The important decisions that need to be taken before conducting any SPTC are: (i) the number of days the SPTCs should be undertaken; (ii) if applicable, the days of the week on which SPTCs should be undertaken; (iii) the number of times SPTCs should be carried out in a year; and (iv) if applicable, the months in which SPTCs should be carried out. Attempts have been made in the past to analyse the impact of duration and frequency of SPTCs on the accuracy of *AADT* estimates. However, none of these studies carried out a systematic and detailed study of these effects with a view to identifying the best approach based on site-specific performance levels.

In this study, two measures – average *MSE* (*AMSE*) and average ranks of *MSE* (*ARMSE*) – have been used to determine the best duration, days, frequency and months or months' separations of SPTCs. As is expected, the accuracy of the estimates increased monotonically with increasing duration and frequency. Looking at the accuracy values themselves is therefore not very meaningful when determining ideal duration and frequency. What is more important to analyse here is to see when the marginal benefits from increasing duration and frequency are highest. Hence, in order to answer questions on best duration and frequency, the rates of improvement in *AMSE* with duration and frequency were used as guiding principles. While answering questions on which days or which month combinations are best, the *ARMSE* values were used since they indicate how well a particular set of days or month combination fare at each of the sites. Analysis has been undertaken separately for total traffic and truck traffic.

Although, as expected, the longer the duration of SPTCs, the greater the accuracy of the estimated *AADT*, the best balance between the accuracy of *AADT* estimates and the resource requirement for conducting SPTCs was achieved when the duration is 3 days (starting with Thursday) for total traffic and 7 days for truck traffic. One way of interpreting the result is that SPTCs at a site need to be conducted for 7 days, although from the total traffic standpoint, 3 days are sufficient.

For the determination of the best frequency of SPTCs, previous studies have mostly considered individual traffic counts to be distributed equally over the year. However, this might not be the best possible choice because two traffic counts separated by five months might not depict the most distinct traffic patterns of the year. Hence, in this study, an attempt has been made to determine the months in which traffic counts need to be done to obtain the best estimates of *AADT*. Analysis indicates that SPTCs should be conducted twice a year irrespective of the duration of SPTCs. However, no particular two-month combination that is good for each of the sites could be determined. Hence, the goal was modified from trying to determine the best two-month combination to determining the best separation (in terms of months) that should be maintained while conducting SPTCs twice a year. The results indicate that SPTCs need to be undertaken twice a year keeping a separation of two months between the counts. This also shows that, maintaining a separation of five months between two SPTCs, as was undertaken in previous studies, is not necessarily the best option.

An additional outcome of this study has been determination of seasonal factors for roads in developing economies like India. Since travel patterns in developing economies are different from those in developed economies, it is worthwhile to study seasonal variations on roads of developing economies. The analysis indicated that on Indian roads, for the data used here, there are no discernible differences in annual seasonal factor variation between sites.

In summary, this study provides an effective way to choose duration, time and frequency of SPTCs so as to gain in accuracy of *AADT* estimates with no extra survey cost. In future, the proposed methodology (the *ARMSE* approach) can be applied to other datasets to find out an effective strategy of conducting SPTCs. This approach is expected to produce results which will work well for each site, rather than working exceptionally well for only a few sites.

Acknowledgements

The data used in this study were provided gratis by V. R. Techniche Consultants Pvt. Ltd. The authors gratefully acknowledge this support. The authors also gratefully acknowledge the opportunity to have presented some of the results reported here at the 94th Annual Meeting of the Transportation Research Board held in Washington, D.C. in January 2015.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Chakraborty, P. 2014. "Estimation of AADT for Indian Highways." Unpublished M.Tech thesis, Indian Institute of Technology Kanpur, Kanpur.

Duddu, V. R., and S. S. Pulugurtha. 2013. "Principle of Demographic Gravitation to Estimate Annual Average Daily Traffic: Comparison of Statistical and Neural Network Models." *Journal of Transportation Engineering* 139 (6): 585–595.

Faghri, A., and P. Chakroborty. 1994. "Development and Evaluation of a Statistically Reliable Traffic Counting Program." *Transportation Planning and Technology* 18 (3): 223–237.

Faghri, A., and J. Hua. 1995. "Roadway Seasonal Classification Using Neural Networks." *Journal of Computing in Civil Engineering* 9 (3): 209–215.

Federal Highway Administration (FHWA). 2013. *Traffic Monitoring Guide*. Washington, DC: U.S. Dept. of Transportation.

Flaherty, J. 1993. "Cluster Analysis of Arizona Automatic Traffic Recorder Data." *Transportation Research Record: Journal of the Transportation Research Board* 1410: 93–99.

Gastaldi, M., G. Gecchele, and R. Rossi. 2014. "Estimation of Annual Average Daily Traffic from One-week Traffic Counts. A combined ANN-Fuzzy Approach." *Transportation Research Part C: Emerging Technologies* 47: 86–99.

Gecchele, G., R. Rossi, M. Gastaldi, and A. Caprini. 2011. "Data Mining Methods for Traffic Monitoring Data Analysis: A Case Study." *Procedia-Social and Behavioral Sciences* 20: 455–464.

Gecchele, G., R. Rossi, M. Gastaldi, and S. Kikuchi. 2012. "Advances in Uncertainty Treatment in FHWA Procedure for Estimating Annual Average Daily Traffic Volume." *Transportation Research Record: Journal of the Transportation Research Board* 2308: 148–156.

Hallenbeck, M. E., and S. G. Kim. 1993. "Final Technical Report for Task A: Truck Loads and Flows." Technical Report 9233, Washington State Department of Transportation, Olympia WA.

Li, M. T., F. Zhao, and Y. Wu. 2004. "Application of Regression Analysis for Identifying Factors that Affect Seasonal Traffic Fluctuations in Southeast Florida." *Transportation Research Record: Journal of the Transportation Research Board* 1870 (1): 153–161.

Lingras, P. 1995. "Classifying Highways: Hierarchical Grouping versus Kohonen Neural Networks." *Journal of Transportation Engineering* 121 (4): 364–368.

Lingras, P. 1998. "Traffic Parameter Estimation and Highway Classification: Rough Patterns Using a Neural Networks Approach." *Transportation Planning and Technology* 21 (3): 155–179.

Lingras, P. 2001. "Statistical and Genetic Algorithms Classification of Highways." *Journal of Transportation Engineering* 127 (3): 237–243.

Ritchie, S. G. 1986. "A Statistical Approach to Statewide Traffic Counting." *Transportation Research Record: Journal of the Transportation Research Board* 1090: 14–21.

Sharma, S. C. 1983. "Improved Classification of Canadian Primary Highways According to Type of Road Use." *Canadian Journal of Civil Engineering* 10 (3): 497–509.

Sharma, S. C., and R. R. Allipuram. 1993. "Duration and Frequency of Seasonal Traffic Counts." *Journal of Transportation Engineering* 119 (3): 344–359.

Sharma, S. C., B. M. Gulati, and S. N. Rizak. 1996a. "Statewide Traffic Volume Studies and Precision of AADT Estimates." *Journal of Transportation Engineering* 122 (6): 430–439.

Sharma, S. C., Y. Wu, and P. Kilburn. 1996b. "The Precision of Average Annual Daily Traffic Volume Estimates from Seasonal Traffic Counts: Alberta Example." *Canadian Journal of Civil Engineering* 23 (1): 302–304.

Yang, S., C. Lu, F. Zhao, R. Reel, and J. D. O'Hara. 2009. "Estimation for Seasonal Factors of Similarity-based Traffic for Urban Roads in Florida." *Transportation Research Record: Journal of the Transportation Research Board* 2121 (1): 74–80.

Zhao, F., M. T. Li, and L. F. Chow. 2004. "Alternatives for Estimating Seasonal Factors on Rural and Urban Roads in Florida." Technical Report BD015–03, Florida Department of Transportation, Tallahassee FL.

Zhao, F., and N. Park. 2004. "Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic." *Transportation Research Record: Journal of the Transportation Research Board* 1879: 99–107.