1991

Traffic Data Selection: An Evaluation of Siting Criteria for Permanent Traffic Recorders

Richard Ledbetter
TriMet

Kenneth Dueker
Portland State University

Milan Krukar
Oregon Department of Transportation

Vern Tabery
Oregon Department of Transportation

TRAFFIC DATA COLLECTION:
AN EVALUATION OF SITING CRITERIA
FOR PERMANENT TRAFFIC RECORDERS

by

Richard Ledbetter, Operations Planner
Tri-County Metropolitan Transportation District (Tri-Met)
4012 SE 17th Avenue
Portland, Oregon

Kenneth J. Dueker, Ph.D., Director
Center For Urban Studies
Portland State University
PO Box 751
Portland, OR 97207

Milan Krukar
Transportation Economist
Oregon Department of Transportation
Transportation Building
Salem, OR 97310

Vern Tabery
Traffic Analysis Supervisor
Oregon Department of Transportation
Transportation Building
Salem, OR 97310

1991

ABSTRACT

Traffic volume data are needed for design and construction zone traffic management. In addition, continuous traffic volume data (collected by automatic traffic recorders) is needed to factor short counts (24 hours or more) collected at key sites on the states highway network.

This study evaluates the procedures used by the Oregon Department of Transportation for collecting continuous traffic volume data to determine: (1) if the current location and number of automatic traffic recorders (ATRs) on Oregon's highway network is adequate for estimating monthly seasonal factors, and (2) if the current procedure of using group means, based on geographical regions, to calculate seasonal adjustment factors is the best method. The use of simple regression versus (1) cluster specific means from computer generated clusters, and (2) weighted means based on the distance from one ATR to the three closest ATRs that triangulate the point, are evaluated as techniques for calculating monthly seasonal adjustment factors from ATR data. These adjustment factors are used to factor volume data where only short counts are available. The results show the current geographical location of ATRs used by ODOT compare favorably, statistically, with other siting criteria. However, it is shown that the task of factoring short count data for making inferences about traffic volume count data from known locations to unknown locations on the highway network can be improved by siting additional ATRs in high volume, small urban areas. In addition, two new areas of research are addressed which have application to real time integrated traffic data collection.

INTRODUCTION

Improved technology for counting, classifying, and weighing vehicles in the highway traffic stream is becoming increasingly available. Automatic vehicle classifiers (AVC), automatic traffic recorders (ATR), weigh-in-motion (WIM) and automatic vehicle identification (AVI) equipment are some of the equipment currently available for use by state transportation agencies to collect traffic data. The traffic volume and classification data is used for design, planning, maintenance, enforcement, and vehicle taxation. Automatic traffic recorders (ATRs) placed at specific sites to continuously monitor traffic volumes for each 24 hour period throughout the year, provide the most accurate means to collect traffic volume counts on the highway network. However, they are also the most costly. Data collection programs aimed at improving efficiency and reducing costs must, therefore, look to alternative methods to monitor traffic volumes and make inferences to other locations on the highway system where continuous counts are not taken.

The Oregon State Highway Division currently operates a 115 station network of automatic traffic recorders (ATRs) throughout the state. In addition, portable counters are used for obtaining traffic information for special projects and situations. Manual counts are also taken periodically and for special projects to obtain information on traffic volumes and vehicle types and classes.

The Federal classification of Oregon's ATRs, based on the Traffic Monitoring Guide (TMG) recommended road groups, are shown in Table 1. The TMG recommends five groupings. The Oregon State Highway Division has expanded these to six groups, the difference being recreational which is split between Interstate and other.

Table 1: Location of ATRs by TMG Grouping		
Road Group	**No. ATRs**	**Percent**
Interstate Rural	10	9%
Other Rural	46	40%
Interstate Urban	10	9%
Other Urban	9	6%
Interstate Rec.	4	3%
Other Rec.	36	31%
Total	115	100%

Source: Oregon Department of Transportation

The purpose of this study is to examine the existing ATR sites on the Oregon Highway Network to determine how well they satisfy objective statistical and design criterion for site selection, and if the current geographical clusters used by ODOT are still spatially valid. Alternative methods for calculating seasonal adjustment factors are identified and assessed. Recommendations for improving site selection for a continuous data collection system are provided. Two areas of future research are identified. The first focuses on the statistical validity of non-random continuous count data and the second, the use of a shrink wrap point set (a-hulls) for spatial clustering of ATR locations.

LITERATURE REVIEW SUMMARY

The most accurate method to monitor changes in traffic volumes at specific locations on the highway network is to continuously count traffic for each 24 hour period throughout the year by installing Automatic Traffic Recorders (ATRs). This method, although efficient for collecting large amounts of volume count data, is very costly. For state transportation agencies, it is not feasible to install such technology at every location on the highway network where traffic count data is needed.

A less costly, but statistically valid procedure is to collect data on a selected sample of highway sections. Generally, counts are taken at pre-determined frequencies over a period of days. A permanent counter is used to adjust the short-term count to an estimated Average Daily Traffic (ADT). The accuracy of the estimated ADT is affected both by the short-term count and by the adjustment factor. The basic assumption embodied in this procedure is that if road sections can be put into groups such that each group contains sections of highways with similar traffic characteristics, data collected on a selected sub-sample in any one group will provide traffic data representative of all sections within that group.

The primary factors used by states for grouping highway segments are the functional class of the highway as given in the Highway Performance Monitoring System (HPMS) and its Annual Average Daily Traffic (AADT). Estimates of the coefficients of variation (ratio of the standard deviation to the mean) of the AADT of groups formed by the standard FHWA procedures tend to be high, which means that large sample sizes are needed to obtain accurate results *(1)*.

A study of traffic monitoring practices in relation to the TMG recommended procedures was undertaken by the New Mexico State Highway and Transportation Department *(2)*. Their analysis of ATR volume counts and seasonal factors identified a high correlation between functional classification and seasonal variation. As a result, the annual and monthly adjustment factors used are based on summary statistics from ATRs on the same highway functional classification. Individual permanent counters are used to factor coverage counts throughout the state, and to make annual ADT adjustments when counts are not taken. Individual counters must be used to control extensive areas of highway. The study determined that beyond a two-mile distance, the mean statistic from ATRs on the same functional classification provides a more adequate count adjustment factor than individual ATR locations.

A study conducted by Ritchie for the Washington Department of Transportation *(3)* evaluated several alternative methods for performing seasonal factoring of traffic volume data including cluster analysis of ATRs; procedures contained in the TMG; and a revised Federal Highway Administration procedure using linear regression. The study indicates that the use of a computer-based cluster analysis routine was somewhat problematic. The clusters computed were not consistent across years and assignment of road sections to a specific cluster was difficult. The adopted approach stratifies the state highway system by geographic region and functional classification. The various strata are examined to determine which have similar seasonal patterns for grouping. Seven factor groups were defined by functional class of road and county boundaries. Seasonal factors for each month of the year were derived for each of the seven factor groups using regression analysis. The dependent variable was the *AADT* for each factor group for each month and the independent variable was the 24-hour short-count volumes (*VOL*) that could be formed for each ATR from 72-hour Tuesday-Thursday counts for the same month. The regression coefficient of the short-count volume is then the derived seasonal factor for that factor group and month.

Ritchie reports that the regression results revealed the presence of heteroscedasticity for some factor groups. This problem can lead to biased estimates of the true variance which in turn effect tests of statistical hypothesis concerning the regression coefficients. It was necessary to employ a standard transformation of the data (dividing each term of the regression equation by (*VOL*) to reduce the regression equation to a homoscedastic form).

The results found that the precision levels for estimated AADT vary as a function of the number of ATRs in each factor group. Ritchie found little improvement in relative precision was obtained beyond six to eight ATRs per group. However, he also indicates that there may be other reasons for maintaining large numbers of ATRs in any group, such as the automatic collection of vehicle classification data. Ritchie did not test different grouping schemes for the purpose of classifying ATRs. Instead, factor groups were defined by functional class of road and county boundaries. In addition, all state highways classified as urban comprise one group regardless of the county location.

An examination of the traffic counting procedures used in Wisconsin *(4)* found that the current seasonal factoring procedure required excessive subjective judgement and time consuming manual manipulation in both the grouping of ATRs and the process of assigning highway sections to defined factor groups. Based on their analysis of the program, Wisconsin adopted the use of a computerized clustering routine to group ATRs. Derived seasonal factors from existing ATR volume count data are used as the grouping variable. Results of the procedures indicate that the use of weekly rather than monthly factors provide no significant increase in accuracy in estimating AADT. Further, monthly factors capture the overwhelming proportion of seasonal variation and are much more stable than weekly factors. In addition, Wisconsin found that the six (6) factor groups they currently used resulted in instability, with some ATRs changing groups each year after the computer cluster procedure. This finding is consistent with Ritchie's finding in his Washington study. To help alleviate this problem, Wisconsin further analyzed traffic flow patterns across the states and concluded that three factor groups was optimal. The Wisconsin study predicts that AADT estimates based on three factor groups should be accurate at the 95% confidence interval within plus or minus 9% for urban highways (Group 1), plus or minus 16% for rural highways (Group 2), and within plus or minus 27% for tourism/recreation area highways (Group 3). Neither Wisconsin nor Ritchie experimented with different geographical groupings. The groups used were taken as given.

A recent paper by Gur and Hocherman *(5)* examined different methods to estimate AADT on the urban arterial road network based on traffic counts on the rural road network. This is the same problem of inferring traffic volumes from a known section of roadway to other sections where counts are not taken. The methods evaluated using actual count data included: (1) estimation of traffic volumes based on road characteristics; (2) estimation of traffic volumes based on past adjustment factors; and (3) use of network connectivity information to update volume estimates.

The study found a significant association between link volume data and type of road, and an apparent association between volume and region. However, no systematic time trend between years, seasons, or months was found. The authors concluded that the use of a system-wide uniform growth factor (method 2) was superior to individual growth factors by route type and geographical region. A more complex procedure focused on updating link volume estimates based on volume changes in connected links. Upon calibration of the model with 1985 data, the results showed large and inconsistent variations in the extent and even direction of volume changes. These findings tend to support the use of clustering techniques for the purpose of determining seasonal adjustment and growth factors, rather than network analysis.

In a draft report "Seasonality Analysis of Colorado's 1986 ATR Data *(6)*," 1986 traffic volume data collected from 54 ATR sites was used to examine seasonality patterns and to recommend an appropriate grouping scheme. Colorado uses a computer generated clustering technique to group its 54 ATRs based on AADT. These groups were compared with the Traffic Monitoring Guide proposed groups developed by the FHWA. The mean seasonal factors and coefficient of variation for each group were calculated to determine statistical reliability. The findings showed both procedures to be reliable, and neither could be proclaimed as "superior." The report did conclude that the use of group factors, as opposed to point factors for each ATR as recommended by the TMG, are appropriate.

Florida's traffic counting program *(7)* operates on the assumption that continuous counters replicate the traffic patterns within contiguous areas. Based on this assumption, ATRs are assigned to individual counties for the purpose of developing weekly factors for use in factoring a single 24-hour count (ADT) to a 24-hour AADT. This was the only program reviewed that is attempting to make point estimates from known ATR locations to other point locations where continuous counts are not taken. Actual statistical results were not reported.

A new classification system developed for the state of Virginia's rural highway system *(8)* breaks down each highway in the rural area into homogeneous highway links with similar traffic characteristics such as AADT and seasonal variations in traffic volumes. Variables that have a significant effect on AADT were identified through a literature search and used as surrogates to cluster the highway links. The candidate variables selected included the following:

- FHWA functional class,
- Functional use,
- Land use of the county in which the link is located,
- Population of the county in which the link is located, and
- Type of terrain.

The conclusion from the authors was that the computerized highway link classification procedure for traffic volume counts was statistically accurate. The important considerations for Oregon are that the procedure does not require knowing the ADT of a specific link before it is assigned to a factor group. In addition, sole reliance on the FHWA functional class stratification is not necessary. Application of this method to point locations has not been tested.

A standard reference text is the Guide for Traffic Volume Counting. The Guide documents two important characteristics that paved the way for clustering of traffic volume data:

- The pattern of monthly variation of traffic volume persists over long stretches of highway, and
- The pattern of monthly variation of traffic volumes persists over long periods of time.

The group mean factor from cluster analysis may be applied to short-term counts located on road sections that fit within the group. The Guide recommended a minimum number of four ATRs were needed for valid results.

A NCHRP study of heavy vehicle monitoring dealt with deployment of the technology on the highway system for both enforcement and planning *(9)*. For planning purposes, they conclude: "A relatively small number of sites would be used as permanent recording stations to obtain seasonal and weekly factors, and a much larger number of sites would be used for short-term classification and weight studies." They also conclude that the TMG recommended approach of randomly selecting sites for the data collection sections used in the HPMS is inefficient, because there is a high correlation in traffic characteristics from section to section over long segments of highway, particularly for the Interstate System. This suggested to them that states might easily develop a more efficient data collection program by selecting weigh-in-motion (WIM) sites that are representative of longer segments.

STUDY METHODOLOGY

The focus of the study is an analysis of siting methods for ATRs for traffic volume data collection, and for estimating seasonal adjustment factors. These continuously collected data can be used to factor counts, classification, and weights collected to meet the requirements of HPMS, and to meet the needs of Oregon's Department of Transportation (ODOT) for factoring data collected at non-continuous and manual short-count locations.

The lack of consistent findings from the literature as to the best method to use led to the creation of a database of annual and monthly volume counts for the years 1984 through 1988 from Oregon's 115 permanent traffic recorders. Cluster analysis was used to segment the ATRs into groups so that the within group sum of squares was minimized and the between group sum of squares was maximized. The grouping variables included the Natural Log of AADT (LogAADT) of each ATR location, the Urban/Rural Class (URC) of the highway where the ATR is located, the Federal Functional Classification (FED) as contained in the TMG, the state plane X, Y coordinates of the ATRs, the percent of heavy vehicles for each recorder, the design hour volume (highest 30th hour), and a factor index controlling for the range in volumes between January and August compared to the AADT for each of the 5 years.

The optimal number of groups is determined by comparing the coefficient of variation (ratio of the standard deviation to the mean) for each group and for different combinations of grouping variables. Clusters were completed for 3, 4, 5, 7, and 9 groups. Based on these groups, the coefficient of variation in AADT as a function of the number of clusters of ATRs was calculated. Once an optimal number of groups and appropriate grouping variable(s) were determined, seasonal adjustment factors for each group were calculated using (1) cluster means; (2) cluster specific regression coefficients; (3) triangulation (mean based on the average means of three ATRs which triangulate a given ATR within a specific cluster group, and weighted by the distance from the target ATR to the three ATRs forming the triangle); and (4) ATR specific regional adjustment factors currently used by the Oregon Department of Transportation.

In order to compare the statistical accuracy for each method, the ADT for January 1988 and August 1988 for each ATR was multiplied by the seasonal factors derived from each of the four methods. Estimated AADT for each cluster and for each ATR location within a cluster was then compared to the actual volume data for each ATR site. The absolute error between the "ground truth data" and the point estimate was used to evaluate the accuracy of each method.

The final step is to compare the volume distribution by highway group for the 115 ATR sites with the 5,195 coverage count locations use by ODOT to determine whether additional or fewer ATRs are needed to accurately represent the universe of traffic volumes and to maintain a specified level of statistical precision.

The coordinates of individual ATRs were provided by ODOT for the purpose of conducting spatial analysis of existing sites. Because the x, y coordinates for five of the ATRs were not available, the data set was reduced to 110 ATR sites for analysis.

McQueen's K-means method using the Euclidean metric (defined by equation 1) was used as the clustering technique *(10)*. The analysis was conducted using the "SYSTAT" statistical software package *(11)*.

$$d^2_{ij} = \sum_{h=1}^{n} (x_{ih} - s_{jh})^2 \qquad (1)$$

where

d^2_{ij} = squared Euclidean distance,

x_{ih} = value of variable h for case i,

x_{jh} = value of variable h for case j,

n = number of variables

Estimation Procedures

It was the intent of the study that the grouping variables allow for easy assignment of coverage count locations to a specific cluster. The coefficient of variation was frequently used in the literature to measure the accuracy of assigning ATRs to factor groups. Oregon's departure from previous studies of traffic counting techniques that use grouping methods focuses on the use of volume data in addition to the type and classification of roadway. The assignment variables included the urban/rural classification, and federal volume class. The ultimate purpose is to use point data that is readily available to make inferences about traffic volumes at other point locations.

Once the optimal number of clusters and cluster elements were determined, the research focused on the best method to calculate cluster specific seasonal factors to estimate AADT. Cluster means, cluster specific regression factors, geographical cluster mean factors currently used by ODOT, and factors based on triangulation were compared for statistical accuracy.

A basic functional relationship of AADT to monthly volume counts is

$$AADT = ADT (Fs) (Fa) (Fg) \tag{2}$$

where

ADT = average daily volume for specific month;

Fs = seasonal factor for the count month;

Fa = weekday axle correction factor if ADT is in axles, equal to 1 if ADT is in vehicles; and

Fg = growth factor if ADT is not a current year count, equal to 1 if ADT is a current year count.

Because the data used in this study was vehicle volume counts for the years 1984 through 1988, the axle adjustment factor was not required. In addition, a growth factor was not applied because the estimates for AADT were based on composite factors for the five year period and applied to January 1988 and August 1988 monthly volumes. These months were chosen because they represent the highest variance in volumes for estimating seasonal adjustment factors. If the procedure was valid for these months, it would be applicable for the remaining months that experience less extremes in volume fluctuations. The following sections provide a brief description of the estimation procedures for each technique.

Current ODOT Seasonal Factors

The Oregon Department of Transportation currently divides the state into four geographical areas representing the coastal areas (Area 12), the Columbia River

Gorge (Area 13), the Willamette Valley region (Area 11), and the Central and Eastern regions of Oregon (Area 14). This division is based on economic development regions established during the 1950's. Each geographical area provides the spatial factor for estimating seasonal adjustment factors based on the following HPMS highway classifications:

UIS/11 - Urban Interstates	STURB/11 - State Urban
RIS/11 - Rural Interstates	STRUR/11 - State Rural
RIS/13 - Rural Interstates	STRUR/12 - State Rural
RIS/14 - Rural Interstates	STRUR/13 - State Rural
	STRUR/14 - State Rural

The adjustment factors are calculated as the mean factor (*AADT/ADT*) from volume counts at continuous counter locations. The mean factor is based on composite year calculations normally covering a 3 to 5 year period. For the purpose of this study, the composite years covered the period 1984 - 1988. The factors were applied to ATR volume counts for January and August 1988 for each spatial group and ODOT highway classification to estimate AADT.

Cluster Specific Means

The January and August mean seasonal adjustment factor (*AADT/ADT*) was calculated for each of the 10 sub-groups of the east-west spatial clusters for the same five year period (1984-1988) used to calculate the ODOT factors. Continuous counter data for January 1988 and August 1988 from each ATR within a cluster was then multiplied by this factor to determine an estimate of AADT for 1988. The computerized clustering of ATRs into the east-west groupings was modified slightly to retain the geographical integrity of the U. S. 97 segmentation. The following ATRs were arbitrarily changed from their original grouping:

Table 2: ATRs reassigned from their original group

Recorder	Route	County	East	West
3014	ORE211	Clackamas	0	X
7001	US26	Crook	X	0
18020	ORE39	Klamath	X	0
20017	ORE58	Lane	0	X
26012	Crown Point	Multnomah	0	X

0 = Original Cluster

X = Final Cluster

Regression Factors

Based on Ritchie's approach, seasonal adjustment factors for each cluster were estimated from the following regression equation:

$$AADT = bADT + u \quad (3)$$

where AADT and ADT are as defined previously. The regression coefficient (*b*) becomes the seasonal adjustment factor and *u* the error term. The presence of heteroscedasticity in some of the clusters resulted in the application of the "Jackknife" technique to estimate the regression coefficients. This procedure has been successful in producing unbiased estimates of parameters that vary in a systematic way over space. The jackknifed parameters are estimated from the difference between the parameter for all the data and that for all places except the one in question. Once calculated, they may be treated as ordinary spatial data and analyzed by ordinary statistical means (Stetzer, 1982).

Triangulation

The seasonal adjustment factor for a specific ATR was calculated as a weighted average of the three ATRs from the same cluster that triangulate the point forming a convex hull. The weighting variable was the distance in feet from the given ATR to each of the three ATRs forming the triangle.

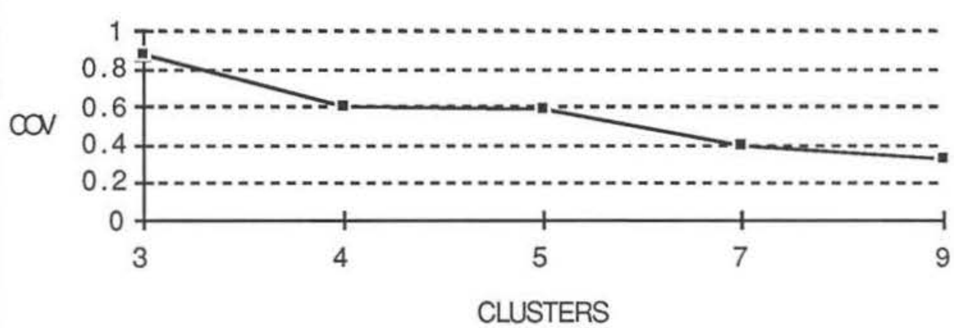
STUDY FINDINGS

Optimal Number of Clusters

The lack of consensus from the literature review led to earlier than anticipated empirical analysis to determine whether spatial clustering of seasonal, design hour and heavy truck factors exist in Oregon. A clustering of ATRs by seasonal factor, design hour volume, or percent heavy trucks did not yield logical spatial groups nor logical types of highway groups.

ATRs clustered by AADT alone indicates that five clusters are needed to significantly reduce the coefficient of variation (cv) and that additional clusters do not significantly reduce this measure. The results of this analysis are shown in Figure 1.

Figure 1. Coefficient of Variation in AADT as Function of the Number of Clusters of ATR



An examination of the clusters formed by grouping on log of AADT, Urban/Rural Class (URC), and Federal Functional Class (FED), shows that when 5 clusters are used, rural low class, low volume roads and rural high class, low volume roads define their own clusters. However, a definite spatial pattern was not apparent in any of the cluster groups. A final clustering scheme using the log of AADT, the percent of heavy trucks, the urban/rural designation, the federal functional classification, a five year growth factor, and a dispersion factor for the range in traffic volumes between January and August compared to the AADT for each of the five years, resulted in an identifiable spatial pattern dividing ATRs east of U.S. 97 from ATRs west of U.S. 97. This east-west segmentation was not apparent before inclusion of the growth factor and monthly variations in volume counts.

The 35 ATRs in the east region and the 75 ATRs in the west region were re-clustered using log of AADT, URC, and FED. This procedure resulted in three optimal groups east of U.S. 97, and four optimal groups west of U.S. 97. The west region was further divided into 7 sub-groups based on federal functional class and volume characteristics.

Adjustment Factors

The triangulation approach was used for a random sample of 12 ATRs from each of the ODOT geographical areas. The estimate of the seasonal adjustment factor and the absolute error for each sample ATR were compared to the actual AADT estimated for 1988. The results showed this approach to be inferior to the use of mean factors or regression factors for estimating AADT. A modification to this procedure using "a-hulls" is discussed later in the paper.

The seasonal adjustment factors based on ODOT's geographical clusters are compared to the factors estimated from the regression technique using Jackknife parameters in Table 3. The regression technique results in a smaller interval at the

90% confidence level for all factor groups for January and for all groups except rural interstates in area 14 (Eastern Oregon) for August. The R^2 for each cluster exceeded .90.

Table 3: ODOT Factors compared to Regression Factors								
ODOT Geographical Factors					**Regression Factors (Jackknife)**			
AREA	JanSF	CI90%	AugSF	CI90%	JanSF	CI90%	AugSF	CI90%
UIS/11	1.10	0.06	0.96	0.06	1.09	0.01	0.95	0.04
RIS/11	1.26	0.03	0.79	0.03	1.24	0.01	0.81	0.02
RIS/13	1.40	0.21	0.77	0.04	1.34	0.12	0.77	0.01
RIS/14	1.49	0.08	0.75	0.03	1.47	0.03	0.77	0.07
URB/11	1.08	0.03	0.93	0.04	1.07	0.01	0.92	0.01
RUR/11	1.27	0.06	0.80	0.03	1.25	0.02	0.80	0.02
RUR/12	1.35	0.06	0.70	0.06	1.35	0.02	0.70	0.02
RUR/13	1.31	0.05	0.81	0.04	1.31	0.03	0.85	0.01
RUR/14	1.48	0.07	0.75	0.02	1.44	0.02	0.76	0.02

Table 4 provides a comparison of the actual AADT for 1988 (AADT88) and estimates of AADT for 1988 based on monthly volume data for January (JAN88) and August (AUG88) using the ODOT cluster mean factors and the Jackknife regression factors. A 90% confidence interval is calculated for each factor group. The slight improvement realized by using the Jackknife parameters does not warrant ODOT changing its current procedures.

Table 4: Estimates of AADT for 1988

Cluster	AADT88	JAN88	CI 90%	AUG88	CI 90%
ODOT					
UIS/11	77319	76284	17549	79268	16292
RIS/11	24077	23632	8743	23665	7458
RIS/13	10279	9223	2775	10003	2683
RIS/14	7093	7178	2734	6979	2097
STURB/11	43718	42655	23938	43249	23866
STRUR/11	3787	3753	778	3689	719
STRUR/12	5850	5770	1019	5720	1042
STRUR/13	2701	2517	1740	2620	1680
STRUR/14	2510	2414	886	2448	892
JACKKNIFE					
UIS/11	77319	75591	17389	78442	16123
RIS/11	24077	23257	8604	24264	7647
RIS/13	10279	8827	2656	10003	2683
RIS/14	7093	7082	2697	7165	2152
STURB/11	43718	42260	23717	42784	23609
STRUR/11	3787	3694	766	3689	719
STRUR/12	5850	5770	1676	5720	1714
STRUR/13	2701	2517	1740	2750	1763
STRUR/14	2510	2349	862	2480	904

The next step was to calculate the group mean factors for the east/west spatial groups and 10 sub-clusters. As with the ODOT estimates, the size of the 90% confidence interval is a direct function of the number of ATRs in each cluster. Only Urban Interstates in area 11 (Willamette Valley) form the same group for both clustering schemes. The results are presented in Table 5.

Table 5: Seasonal Adjustment Factors Based on East-West Groups

CLUSTER	JanSF	CI90%	AugSF	CI90%
E1/Rural IS	1.39	0.09	0.80	0.03
E2/Rural MOD[3]	1.45	0.06	0.77	0.02
E3/Rural LOW	1.37	0.12	0.81	0.04
W1/Rural IS	1.27	0.05	0.80	0.02
W1/Urban IS	1.10	0.06	0.97	0.06
W2/Rural HI[1]	1.23	0.05	0.89	0.02
W2/Rural MED[2]	1.36	0.08	0.73	0.06
W3/Rural LOW	1.43	0.42	0.73	0.18
W3/Rural MOD	1.26	0.12	0.81	0.05
W4/Urban HI	1.10	0.04	0.93	0.04

1. HI = High Volume 2. MED = Medium Volume 3. MOD = Moderate Volume

Because of the differences in clustering criteria, a cluster by cluster comparison of absolute error is not possible. Instead, the percent of ATRs within each cluster that is within 10% of the actual AADT is used to show the accuracy of the two procedures. Figure 2 (ODOT) and Figure 3 (east/west groupings) below show the percent of estimates for January and August that fall within 10% of the actual mean volume for each cluster. The ODOT clusters on average contain a higher percentage of ATRs within the 10% criteria.

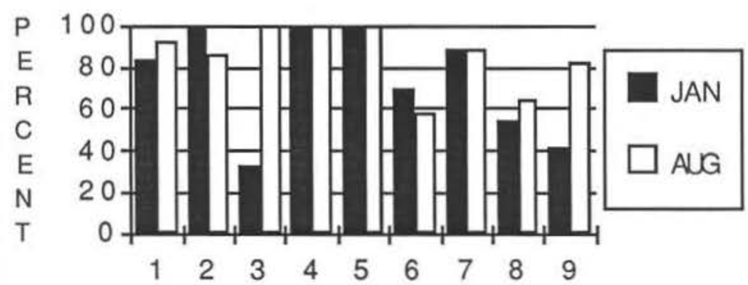


Figure 2: ODOT AADT ESTIMATES WITHIN 10% ABSOLUTE ERROR OF ACTUAL MEAN VOLUME

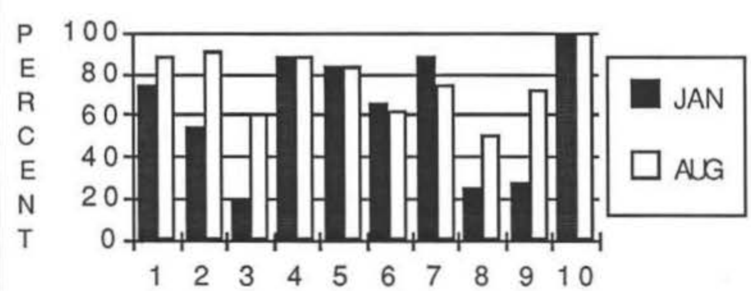


Figure 3: EAST/WEST AADT ESTIMATES WITHIN 10% ABSOLUTE ERROR OF ACTUAL MEAN VOLUME

The findings from the above analysis indicate that the spatial clusters used by ODOT still provide a representative spatial segmentation of the traffic volumes by highway class. In addition, the use of cluster mean factors for factoring coverage count volumes is still an acceptable methodology compared to other statistical techniques.

The question that remains is whether the number of ATRs and their location is providing traffic volume data that accurately represents the universe, and that allows for reliable statements to be made about the Oregon highway system. This is accomplished in two steps.

The first step is to compare the volume counts at manual count locations with counts at permanent locations where "ground truth" data exists for each highway class. Areas where the absolute error is too large may indicate locations where additional ATRs are needed. Table 6 provides a comparison of the mean volumes, standard errors and coefficient of variation (cv)) for the 110 ATR locations and the 4700 coverage count locations used to estimate AADT. The data is broken down by spatial cluster and highway class.

Table 6: Mean Volumes at ATR Locations

Cluster	#Cases	Mean Volume	Standard Deviation	Coefficient of Variation
ATR DATA				
UIS/11	12	70118	32259	0.46
RIS/11	7	21654	11538	0.53
RIS/13	3*	9237	2587	0.28
RIS/14	3*	6424	2449	0.38
STURB/11	6*	39355	31711	0.80
STRUR/11	33	3508	2413	0.69
STRUR/12	8*	5243	2564	0.49
STRUR/13	11	2542	3342	1.30
STRUR/14	27	2314	2619	1.10
COVERAGE COUNT DATA				
UIS/11	95	71080	32176	0.45
RIS/11	85	24945	13039	0.52
RIS/13	43	7876	3283	0.41
RIS/14	26	5703	1022	0.18
STURB/11	784	25908	23329	0.90
STRUR/11	2274	4253	2620	0.62
STRUR/12	395	7211	3981	0.55
STRUR/13	325	2339	2754	1.20
STRUR/14	674	3093	2784	0.90

Source: ODOT * Additional ATRs needed

The data indicates that the sampling rate for Rural Interstates in the Columbia River Gorge Area (RIS/13), State Rural Highways in the Coastal Region (STRUR/12), Rural Interstates in Eastern Oregon (RIS/14), and State Urban Roads in the Willamette Valley (STURB/11) are not representative of the volume changes occurring in these areas. For example, ATR data for rural interstates in eastern Oregon (RIS/14) show a higher mean volume (6,424) than that recorded at coverage count sites (5,703). However, there is more variation in the data at ATR sites as evidenced by the standard deviation (SD=2,449)) and coefficient of variation (cv=.38). The 3 ATRs used to represent the universe tend to capture higher volumes. An additional ATR placed at a lower volume site may improve the variability in the count data and more accurately represent the coverage count mean volume. Similarly, ATRs on state rural roads on the Oregon coast (STRUR/12) tend to capture lower volumes when compared to coverage count sites. Additional ATRs placed at higher volume sites would help reduce the difference in recorded mean volumes, and ultimately provide a better estimate of cluster specific adjustment factors.

Second, AADT estimates with a standard of precision of ± 10 percent at the 90% level of confidence was the preferred target for factored volume counts. Given this standard and the total variation in AADT for the Oregon highway system, a minimum sample size for recording point estimates of traffic volumes can be derived as follows:

$$n = [(Z_{\alpha/2} * S)^2/e/2] \quad (4)$$

where

n = the number of sites to be randomly sampled;

$Z_{\alpha/2}$ = the critical z value associated with a level of confidence of $(1-\alpha)$;

S = the standard deviation of traffic volumes system wide; and

e = the level of precision required from the sample estimate

In examining equation (4), three important points must be considered. The quantities e, $1-\alpha$, and S must all be specified before n can be found. In addition, equation (4) can be used even when the underlying distribution is not normal, provided that the resulting sample size is large enough to invoke the Central Limit Theorem. What is needed, however, is a random sample of traffic volume counts. Thus, the method of data collection becomes very important to the question of "how many ATRs are needed?"

With coverage count locations, counts can be taken at randomly selected sites. Alternatively, ATR counts are taken at the same sites on a continuous basis. These volume counts are not independent as is required for simple random sampling. Rather, the ATR data constitutes a cluster sample where coverage count data represents the primary sampling unit. A sampling plan for ATRs must account for the magnifying effects of the systematic correlation of volume data on the overall sampling error. What this means is that the presence of a high level of correlation in the ATR data requires a larger sample size to achieve the same level of precision as a simple random sample would.

The proportionate increase in sample size resulting from the non-random sampling of ATR volume data is similar to the phenomenon called the "design effect" (11). If the correlation between the data is known or can be estimated, the "design effect" can be removed. This is a necessary first step before statements can be made about the number of ATR needed to replicate the universe at a specified level of precision for a specific highway class.

Finally, it was noted during the study that coverage counts taken at sites within the city limits of high volume areas such as Astoria and Lincoln City on the Oregon coast, and Bend in Central Oregon, are not used in the calculation of mean volumes for their respective spatial group and highway class due to large variances from ATR data from the same areas. This same situation was not observed however, in the coastal city of Newport. What was discovered was that coverage count data in this area recorded volumes within the city were significantly higher than those recorded by the ATR located 3 miles south of town. As a result, there is a significant difference in the mean volume calculated for state rural roads in area 12 (STRUR/12) when ATR and coverage count data are compared. One recommendation for decreasing the variance between ATR data and factored coverage count data is to place additional recorders within the city limits of Newport to better replicate the average daily traffic occurring on the highway within the urban area. A second approach is to form a new class of highway called "small urban" and group similar locations into their own cluster. Monthly adjustment factors could then be calculated from these groups and used to factor coverage count data to AADT with more precision and less variation.

Conclusions

The use of ODOT's four geographical clusters and geographical cluster means for calculating seasonal adjustment factors fulfills the majority of the department's data collection needs and federal reporting requirements. However, the data collected at permanent recorder sites should replicate the average traffic volumes being recorded at coverage count locations in the same area and functional class of highway, to provide a statistically valid estimate of AADT at non-count sits. To the degree the permanent traffic recorders do not replicate coverage count average volumes, the greater will be the error in estimates of AADT.

ATR sites are generally located in rural areas, except for urban freeways. Consquently, the volumes observed in the coverage count program for state highways in small urban areas are often not reflected at ATR sites. If this is the case, it may not be appropriate to apply factors derived from current geographic cluster means to ATR sites in small urban areas that exhibit volumes that are greater than the volumes used to derive the factors.

In the case of Oregon, it is recommended that a new class of highway called "small urban" be formed to account for the higher volumes being recorded in urban areas on the coast and central Oregon. These segments of highways would form their own group for the purpose of calculating adjustment factors. Currently, traffic volumes in these small urban centers are not used in calculating adjustment factors. This same approach may be beneficial to other states experiencing large variances between permanent counter data and coverage count volumes in specific areas. If this procedure is used, volumes from the designated areas should be compared statistically to determine if there are significant differences geographically. If there is, seasonal factors based on a geographical clustering of ATR sites in these areas should be developed.

A similar problem may exist in large urban areas, because the ATR sites are usually located on freeways, not on urban arterials. However, this problem can be dealt with by an urban counting program.

Future Research Areas

It is recommended that research be undertaken to determine the systematic bias present in continuous count data ("design effect") and its effect on sample size. Removal of the effect may eliminate the need for ATRs at some locations on the highway network.

It is believed by the authors that one of the shortcomings of the current clustering methodology has been the inability to reproduce or improve upon with greater resolution the state's designated spatial areas (i.e., the coast, valley, interstate corridors, etc.). The problem is twofold. First, unless the clustering is performed on the spatial data without respect to the statistical data (x - y coordinates only), the statistical data renders the technique impracticable. Secondly, when convex hulls are drawn about the clustered point sets, outliers within the clusters distort the proper spatial representation. The current technique does not allow for distance constraints on the spatial clustering.

In order to achieve the most representative spatial clustering, statistical clustering should be performed first. Each of the resulting clusters will then serve as input point sets for the spatial clustering. Delaunay triangulation is then performed on each point set. A distance constraint can now be easily imposed by dropping from the triangulation all edges which exceed the threshold distance. This action will result in the current point set (statistical clusters) being broken up into two or more subsets.

At this point, instead of using convex hulls to define the subset areas, a-hulls are used. The difference between convex hulls and a-hulls is that a-hulls are not by definition convex, although they can be. A-hulls are best described as a shrink-wrapped point set. A is a parameter which can be set to control the shape of the hull around the point. The advantage to using the shrink wrapping technique is that the overlap between point sets is minimized, thereby reducing ambiguities associated with assigning a point to a spatial zone.

The above procedure can be performed to each statistical cluster. It is then possible to combine various a-hulls resulting from the statistical clustering to achieve the desired level of aggregation for further analysis. This methodology may provide clusters that are superior to the existing spatial groups or the clusters formed by statistical criteria in this study, for the purpose of calculating adjustment factors to factor coverage counts to AADT.

REFERENCES

1. Garber, Nicholas J. and Faramarz Bayat-Mokhtari, "A Computerized Highway Link Classification System for Traffic Volume," Transportation Research Record 1090, TRB, Washington, D.C.

2. Albright, David, Chris Blewett, and Rachel Goldstine, "Development of State Traffic Monitoring Standards: Implementing the Traffic Monitoring Guide in New Mexico," Transportation Research Board 68th Annual Meeting, January 22-26, 1989, Washington D.C., Paper No. 880241.

3. Ritchie, Stephen G. and Mark E. Hallenbeck, "Evaluation of a Statewide Highway Data Collection Program," Transportation Research Record 1090, TRB, Washington, D.C.

4. State of Wisconsin, Wisconsin Department of Transportation, A Statewide Traffic Counting Program For The Wisconsin Department Of Transportation. Madison, Wisconsin, January 1987.

5. Gur, Yehuda J., "Optimal Design of Traffic Counts," Transportation Research Board 68th Annual Meeting, January 22-26, 1989, Washington, D.C., Paper No. 880552.

6. Esteve, M., Draft Report, "Seasonality Analysis of Colorado's 1986 ATR Data," Federal Highway Administration, May 1987.

7. Florida Department of Transportation, "Traffic Data Collection in Florida," unpublished report submitted to FHWA, circa 1988.

8. Ritchie, Stephen, G., "A Statistical Approach to Statewide Traffic Counting," Transportation Research Record 1090, TRB, Washington D.C.

9. Grenzeback, Stowers, and Boghani, NCHRP Heavy Vehicle Monitoring Project Report #303, 1988.

10. Everitt, Brian, Cluster Analysis, Halsted Press, 1974.

11. Wilkinson, Leland, Systat: The System for Statistics, Evanston, IL 1988.