Wisconsin's Approach to Variation in Traffic Data

Bruce Aunet

Wisconsin Department of Transportation

4802 Sheboygan Avenue, Rm. 933

Madison, WI 53707-7913

Traffic data exhibits considerable variability, both spatially and temporally. Given limited resources and the large geographic coverage required for data collection efforts, short period (24-hours to 7-day) traffic data collection must often serve to represent average conditions. Yet in order for short period traffic counts to most accurately reflect annual average daily traffic (AADT) estimates, for instance, we must apply factors to take into account temporal variation, particularly seasonal (monthly) and day-of-week variation. To narrow the scope of the study, this paper will focus solely on total volume count data. Sharpening the focus further, the paper will deal with practical approaches to handling one of the largest sources of temporal variation in traffic data – seasonal variation. The techniques presented, however, have broader applicability to other sources of variation in traffic data. The paper will attempt to show how a combination of approaches – using statistical measures such as the coefficient of variation, statistical procedures such as cluster analysis, plots of monthly traffic factors, and geographical mapping of continuous count sites – can produce seasonal factor groups and seasonal adjustment factors to substantially account for seasonal variation and thus produce more accurate AADT estimates for end uses.

INTRODUCTION

While written from the perspective of a knowledge worker or end user of traffic data (one who routinely uses and analyzes traffic data in the course of his work), the traffic data program analyst represents the primary intended audience for this paper is. It can also, however, help provide a better frame of reference for any end user of traffic data.

OVERVIEW

Annual average daily traffic (AADT) represents probably the most widely and frequently used traffic data statistic. Traffic data users, both within and outside of traffic and transportation agencies, need AADT estimates for all major highway segments, and would ideally like them for all road segments – including minor and local roads.

Of course, continuous automatic traffic recording (ATR) monitors provide the most direct and accurate AADT estimates, even given the reality of frequent missing or biased data due to construction, detours, or equipment failures. It does not seem financially practical – or even necessary, given statistical sampling – to continuously monitor traffic at every road segment location where a need exists for accurate traffic data.

In fact, the recommended practice (*1,2*) involves taking a short period count, typically for 48 hours, and then making adjustments to these short period counts for seasonality and day-of-week, based on numerical factors developed from information obtained at the ATRs. In addition, the recommended practice includes the application of an axle adjustment factor – obtained from vehicle classification sites – if the equipment used to collect the short period traffic counts adds axle impulses rather than vehicles.

In actuality, a number of ATRs get installed to collect traffic data for a specific location (prior to the design of a major project, for instance), and thus don't represent a true random spatial sample. Still, it seems both efficient and effective to make more extended use of the expansive and high quality ATR data. One such use involves summarizing the data in order to develop generalized traffic patterns applicable to locations with similar traffic patterns. Even if we don't have a strict random sample of ATR locations, the assumption of a quasi-random spatial sample of ATRs seems reasonable.

OBJECTIVES AND FOCUS

The main objectives of this paper include a discussion of some practical statistical approaches to dealing with variation in short term traffic counts so as to arrive at more reliable and accurate measures of AADT. Since dealing with all sources of traffic variation involves too broad a scope for this paper, we will focus on by far the largest source of this variation (in Wisconsin, at least) – seasonal variation over the different seasons or months of the year.

In addition, we will focus on the practical and usable in our approaches, not the obscure, abstract, overly complicated, or academic. This seems particularly appropriate for a paper aimed at practitioners, by a practitioner, from a conference session entitled, "Workshop - Traffic Data 101: A Statistical Primer for traffic Data Practitioners".

ANALYSIS OF SEASONALITY

Traffic flows exhibit considerable variability, both spatially and temporally. Absent this variability, we could have very simple traffic data collection programs. We would not need nearly as many count locations and a count taken at any time would pretty much reflect typical conditions. Yet we know from simple observation that traffic flows vary substantially across geographic areas – depending on surrounding land uses – and across time – depending on hour of the day, day of the week, and (in many states) month of the year.

This simple fact of life obviously complicates traffic data collection efforts, especially when we often seek to obtain estimates of "average" conditions, such as AADT. Yet, fortunately, we can detect and use underlying patterns in traffic flows to simplify our programs. And again, we will focus here on seasonal variation.

Coefficient of Variation

While we typically seek measures that represent average or mean conditions, such as AADT, we know that we need to pay attention to simple statistics that show the variance or the standard deviation of the resulting measure to get an idea of the reliability and accuracy of our measures. Yet we really want to compare deviations across various volume levels, so we need a standardized measure not affected by the magnitude of the numbers involved. Thus we use the coefficient of variation (CV) – defined as the standard deviation divided (or standardized) by the mean, and multiplied by 100 to get a percentage. Standard statistical software packages provide the CV, and some, such as the Statistical Analysis System (SAS), express it in percentage terms by default.

The calculation of the CV assumes calculating the AADT as the arithmetic mean of the daily traffic at the ATR. For all practical purposes, especially when the ATR contains mostly complete data for the year, this provides the same AADT as the AADT calculated as the arithmetic mean of the seven annual average days of the week for each month, as recommended in the AASHTO Guide (*2*).

Variation in Unfactored Daily Traffic Counts and AADT Error Reduction

While the recommended practice calls for factoring short term traffic counts, and some sort of factoring now seems like generally accepted practice, a brief review of why we need to factor short term counts (if we indeed do want AADT estimates) can help explain the importance of factoring.

Because we want to improve the accuracy of short term traffic counts adjusted to estimates of AADT, we need only concern ourselves with those months of the year when we take short period counts. In northern states – because snow plows would rip up road tube counters – that typically ranges from March or April thru October or November. Removing the lower volume winter months in the northern climes helps reduce the overall variation in seasonal traffic flows.

A recent study involving one urban ATR in Cedar Rapids, Iowa (*3*) found that unfactored daily counts (over the April-October period) had an average deviation of 9% from AADT, and improved an average of

25% (to 6.7%) with the application of statewide group factors. Previous work in examining 49 ATRs in Wisconsin (*4*) found a coefficient of variation in daily traffic, ranging from 13% in urban areas to 23% in rural areas, and as much as 40% in recreational areas. The application of seasonal group adjustment factors to simulated 48-hour counts resulted in a reduction of 65% in the CV.

Because of the stability of commuter traffic in urban areas, urban traffic flows exhibit the least amount of seasonal variability. The more initial seasonal variation in daily traffic (in rural and recreational areas), the more room for improvement resulting from the application of seasonal adjustment factors.

Identifying Seasonal Traffic Flow Patterns

Given the clear need to seasonally adjust short period traffic counts to obtain reliable estimates of AADT, especially outside of urban areas, how do we go about defining seasonal factor groups? In most cases we follow the procedures recommended by the *Traffic Monitoring Guide (TMG)* (*1*), using months as a convenient measure of seasonal variation. Using weeks might provide a more refined measure of seasonal variation, but because of the affects of holidays, we would need to make certain to use data from the ATRs for the current year, and not the previous year, from which to develop the factors.

For the actual measure of monthly variation (for use in the analysis procedures), we prefer to use the inverse of the monthly adjustment factors. While AADT divided by monthly average daily traffic (MADT) results in the calculated (and multiplicatively applied) monthly factors, MADT divided by AADT shows the actual patterns of monthly traffic. This can look more intuitively appealing when viewed in plots, for instance. Yet either approach – using the monthly factors directly or their inverse – can work equally well.

When it comes to analytic procedures, we recommend using a combination of approaches – from examining plots of monthly traffic at each ATR, to examining tables of CVs for each ATR, to examining the results of cluster analysis, to examining geographic mappings of ATRs in preliminary groupings.

Visually Examining Plots of Month of Year Traffic Variation at ATRs

One of the most straightforward, and perhaps ultimately most useful, methods of grouping continuous count stations involves simply plotting the monthly factors (or their inverse) for each site – using the plotting procedures available in your statistical or spreadsheet software – and visually examining the results. Not only do these simple plots allow you to literally see the various seasonal patterns, but they come in particularly handy when interpreting some of the other methods for grouping ATRs that we will discuss next. Figures 1-4 show some of the typical seasonal patterns we encounter.

Figure 1

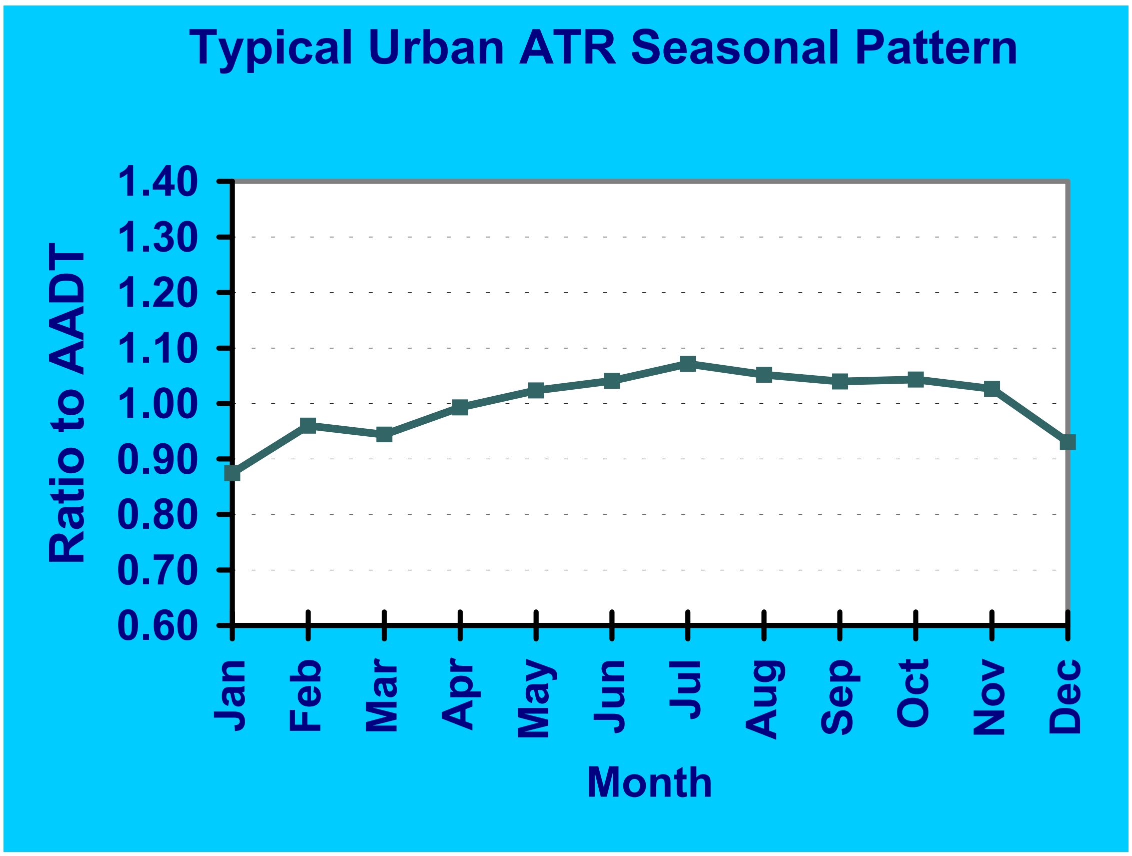


Figure 2

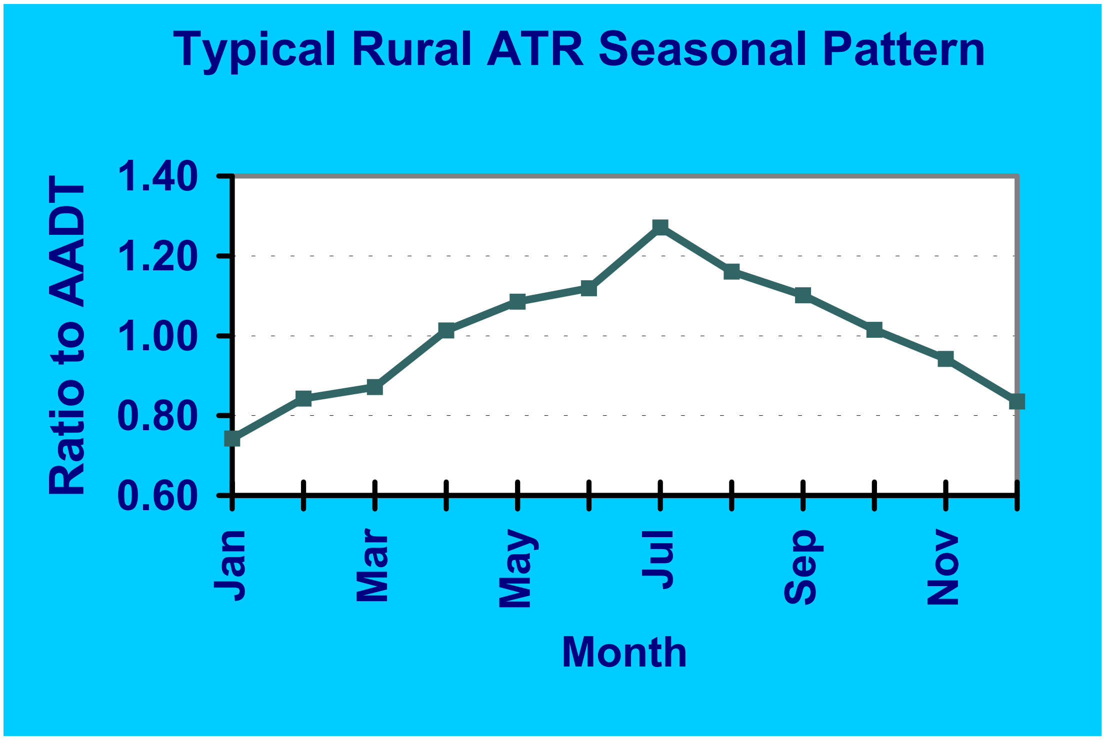
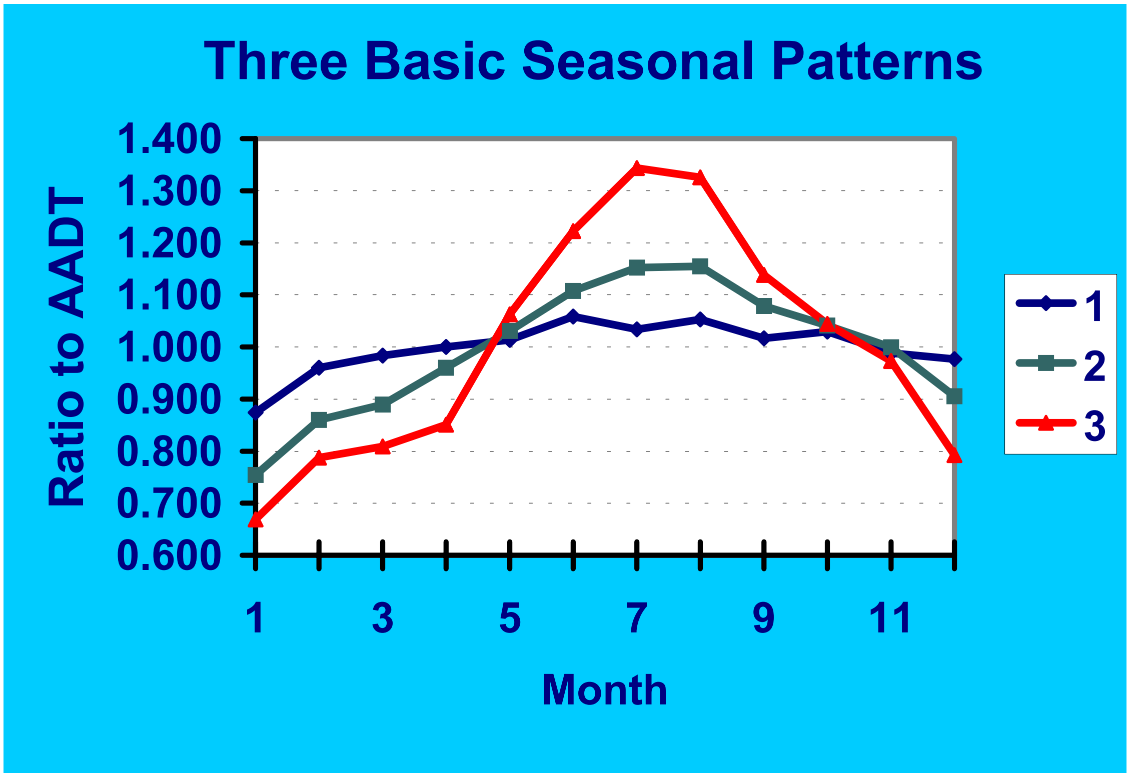


Figure 3

Error! Not a valid embedded object.

Figure 4



A first cut at establishing seasonal factor groups would simply involve manually classifying the plots into groups with similar patterns based on visual observation.

Dealing with Outliers

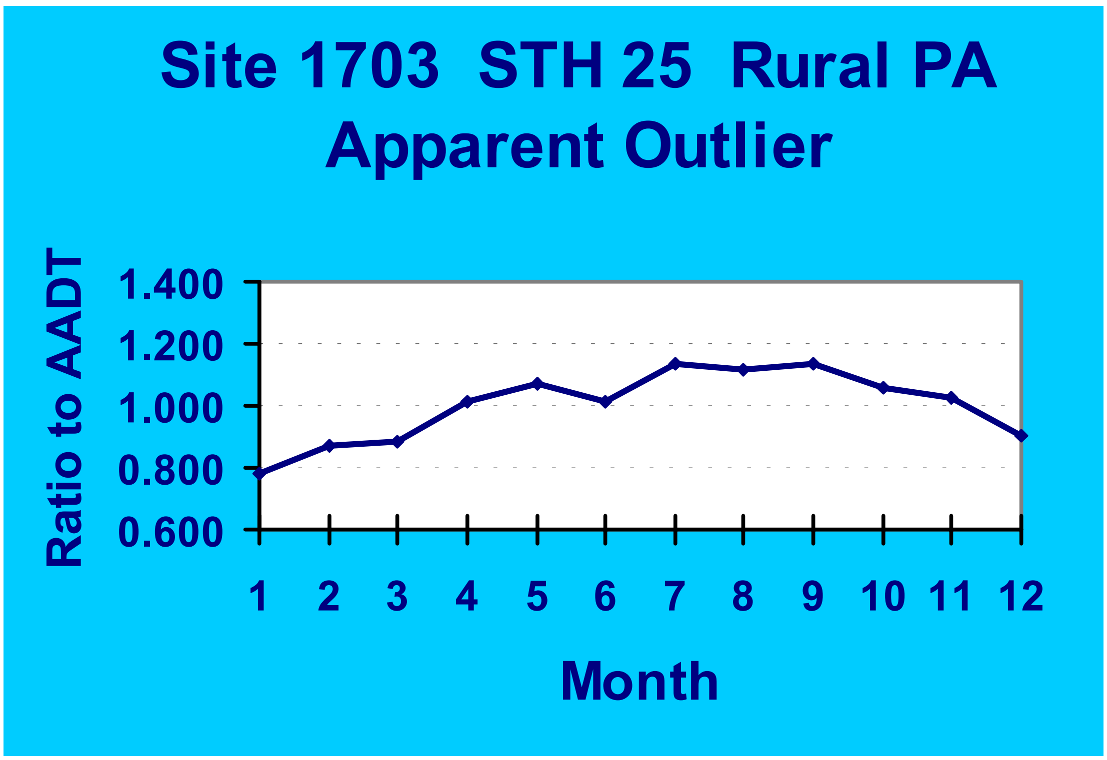
While the data shown for the four ATRs in Figures 1-4 don't appear to have any apparent problems, typically a visual examination of the plotted data will reveal potentially "problem" data. One of the most time consuming tasks in any data analysis project involves dealing with outliers – data that significantly deviates from the vast majority of the rest of the data. These "problem" data can significantly affect the results of many statistical analyses. Cluster analysis, for instance, seems particularly sensitive to outliers. One aberrant month can place a station in its own group.

We don't want to throw out or exclude data from our analyses, however, just because they deviate from the norm or complicate our analyses. Some "outliers" may well reflect actual patterns in the data. In many cases, however, further investigation into the causes of the outliers – the time consuming part of the project – turns out to result from equipment problems, or the temporary effects of highway construction or detours.

In those cases we in fact do not want such atypical situations to bias the results of our analyses when we really want to get at "average" or typical conditions. We then need to exclude such data from our analyses.

Figure 5 shows a continuous count site with an apparent aberration in the month of June. To more systematically check for outliers, we recommend using from 3-5 years of data in your seasonality analysis, given that seasonal trends at a given station appear stable over time. That way our plots will more clearly show outliers, as shown in Figure 6, which if left in the analysis data set, could skew or bias the results.

Figure 5



Examining Tables of CVs for each ATR

A second simple step in analyzing seasonal traffic flow patterns and grouping the ATRs into seasonal factor groups involves calculating some simple statistics, particularly the CV (in this case of the monthly factors), using a statistical or spreadsheet software package, to get some summary statistics to go along with the plots. Including some additional summary statistics, such as the minimum and maximum values, and the range, also can prove helpful. Table 1 shows an example for some ATRs in Wisconsin.

Figure 6

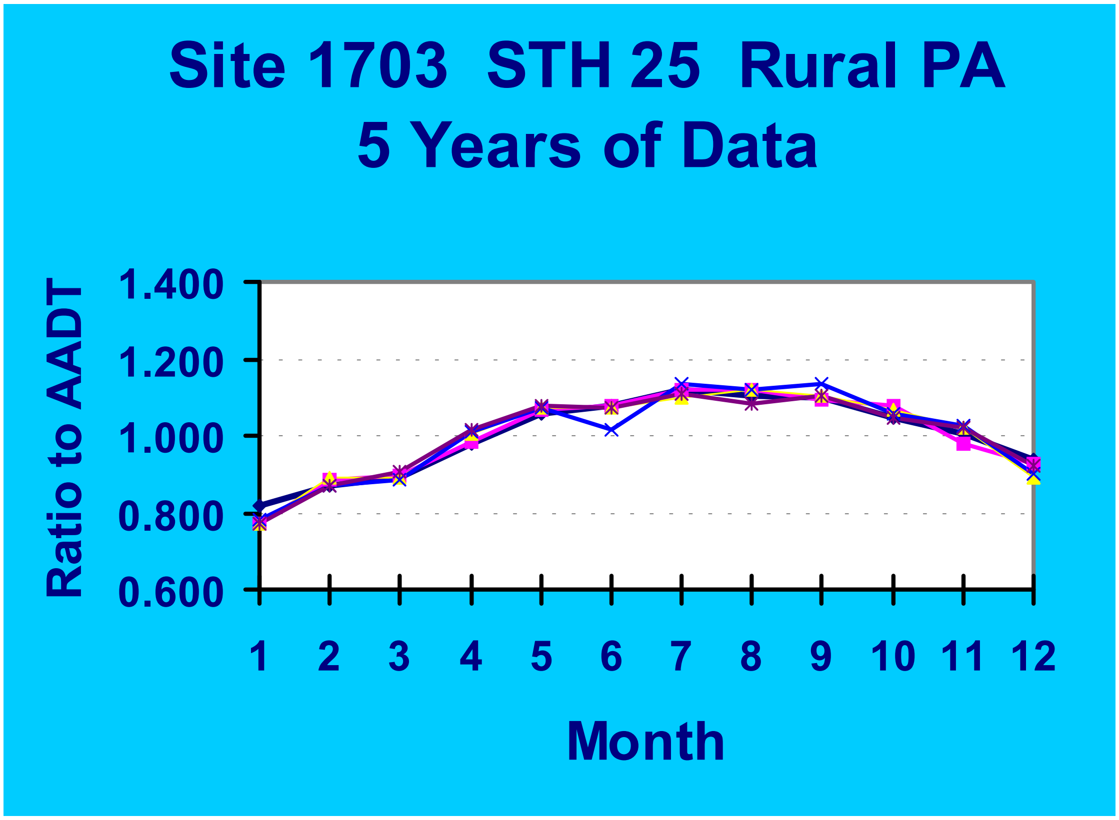


Table 1

Monthly Factor Summary Statistics for Some ATRs in Wisconsin

Site	*Func. Class.*	*Minimum*	*Maximum*	*Range*	*CV*
1102	1	.82	1.35	.53	15.7
2801	1	.86	1.22	.36	10.6
1602	2	.74	1.45	.71	23.9
3101	2	.71	1.54	.83	24.8
0101	6	.87	1.27	.40	11.2
4004	11	.99	1.06	.07	2.4
0502	14	.94	1.03	.09	2.8
1310	16	.94	1.07	.07	3.7

When we compare the CVs with the plots, we find that the plots showing the most variation in the monthly factors also, not surprisingly, have the highest percent CVs. Having both plots and numbers to examine proves quite useful.

Using Cluster Analysis

Cluster analysis, a statistical technique for identifying natural groupings in the data based on variation in the data, provides an objective approach to defining seasonal factor groups among ATR stations. We need not understand the mathematical algorithms used in cluster analysis in order to use it as a tool and interpret the results. Both the *TMG (1)* and the SAS software manuals (*5*) provide good descriptions and examples of cluster analysis so we will not go into those details here.

Suffice it to say that cluster analysis begins with each observation (or ATR station, in our case) in its own group, and then groups observations together, based on similarity, until we end up with just one group. Typically we will want to look at the groupings that result with three to six groups (near the top of the cluster tree). If we find a group or more with just one or a few stations in it, we need to examine the data from those stations for outliers that may have biased the results. In some cases we may indeed have a unique pattern in some portion of the state, but we need to systematically analyze why that might exist and we need to make sure we have not inadvertently included data containing the temporary affects of highway construction or detours or equipment problems

While cluster analysis does <u>not</u> identify the optimal number of groups in the data – that takes some *a priori* analysis – the results do provide some guidance in this direction (in terms of the cubic clustering criteria and the change in r-squared, for instance). We agree with the *TMG* that it makes sense to end up with from three to six seasonal factor groups, certainly with distinctions between urban and rural patterns.. We find it useful to plot out the resulting groupings on a map, using GIS mapping software, and to display the ATRs in each different group with a different color and symbol. Having the data geographically displayed can help interpret the seasonal patterns. Here again, knowledge of the conditions in your own state can prove invaluable

While we have described several approaches to identifying seasonal traffic flow patterns above, in actual practice they can proceed in any order. In fact, we might start with cluster analysis, and then get into plots of monthly factors and tables of CVs to help explain the results and identify outliers.

Ultimately, deciding on the number of seasonal factor groups to use involves a trade off between the two major sources of error in seasonally adjusted short period traffic counts. We refer to those sources of error as “aggregation” error and “assignment” error. Aggregation error results because we use the monthly factors across a number of ATRs to calculate a group mean, and a mean estimate, by definition, has error built into it. Assignment error results because some highway segments (from the vast majority where we don't have ATRs) may get assigned to the wrong seasonal factor group. Obviously, the more groups we have, the greater the likelihood of this assignment error occurring.

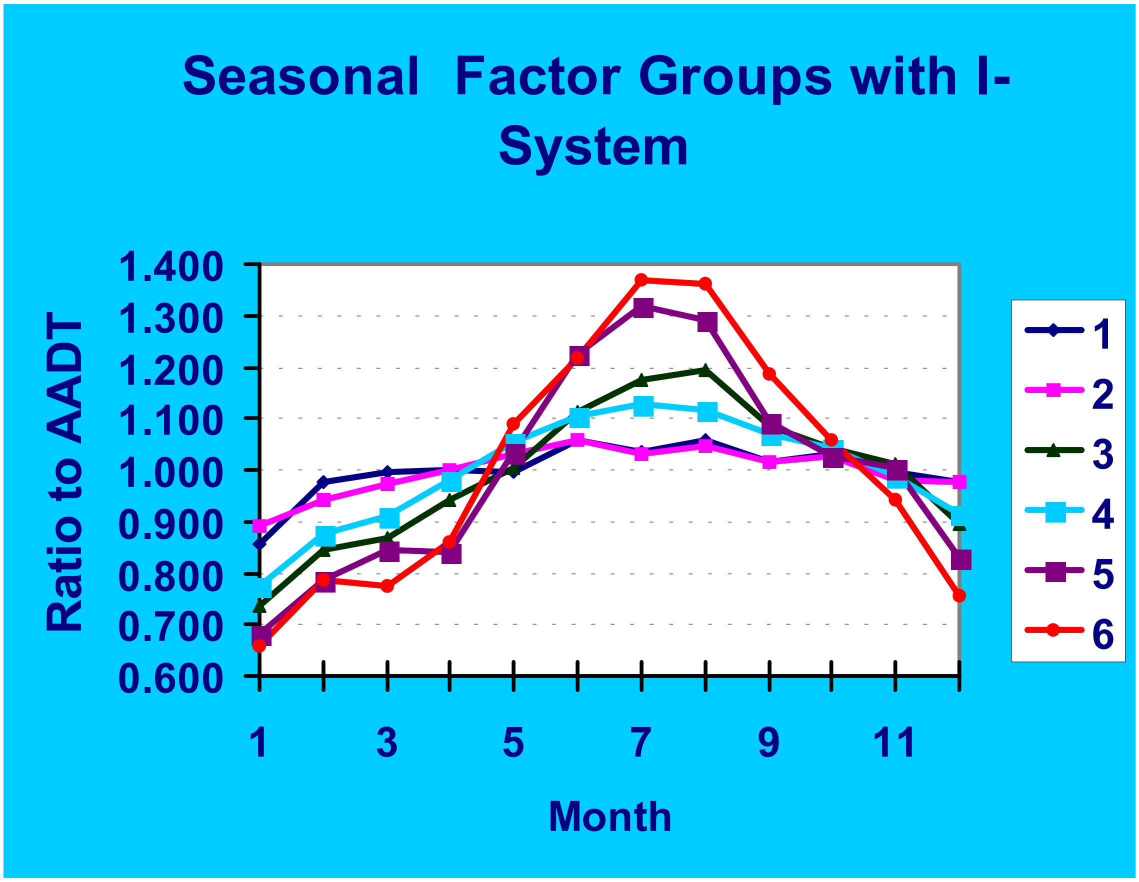
Results

While we have not yet completed our analysis of our seasonal factoring process – identifying seasonal groups, deciding on the final number of groups to use, deciding on how to assign highway segments to groups, etc. – preliminary results indicate we still have the same three basic factor groups we identified almost two decades ago: urban, rural, and recreational. So it certainly appears the basic seasonal patterns remain stable over time.

Given that we now have more ATRs in ex-urban fringe or commuter areas than before – and some of that likely occurred due to urban build-out and ex-urban development – we notice the possibility of a fourth group falling between the urban and rural groups in terms of seasonality. However, the added difficulty of properly assigning highway segments to this potential new group may outweigh the benefits of adding it as a new group. We have yet to reach a decision on this.

When we superimpose the Interstate highway system on top of these three basic groups, we get six groups (with portions of the Interstate system falling into each of the basic three groups), as shown in Figure 7. The urban interstate and other urban highways show very little difference in monthly factors. Rural interstate ATRs show higher summer peaking than the rural other ATRs, and in recreational areas, other highways show more summer peaking than ATRs on interstate highways leading to recreational areas.

Figure 7



DEVELOPING PRECISION ESTIMATES FOR AADT ESTIMATES

Evaluating the different sources of error in AADT estimates, as discussed above, highlights the importance of developing a program to simulate AADT estimates in order to develop actual estimates of errors. This simply involves simulating short period traffic counts with ATR data, applying the appropriate group adjustment factors, and then comparing the resulting simulated AADT estimates with the actual AADT values at the ATRs from which we simulated the short counts. (This presupposes that you can readily get daily traffic estimates from your ATRs through your traffic data software programs.) From these results we can develop precision estimates for our short period factored AADT estimates, such as conclusions that "our urban AADT estimates have a mean absolute percent error (MAPE) of 5%", or "our urban AADT estimates have accuracy of plus or minus 10% at the 95% confidence level". The estimates depend on how the results turn out and on how we present them, given the number of choices available

(MAPE, root-mean-squared percentage error (RMSPE), plus or minus error with associated confidence levels, etc.).

Even though we know we have less precision in our recreational areas (with their highly seasonal nature), we consider it important, both from a data user's perspective and for real "truth-in-data", to develop these precision estimates for AADT estimates. Some recent studies also have used such simulation programs to evaluate the accuracy of AADT estimates (*3,6,7*).

The simulation program, once developed, has many additional purposes, including evaluating the use of: various numbers of factor groups and the use of different data collection durations (from 24 hours to 7 days, for instance) – all by evaluating the magnitudes and changes in the different sources of error in AADT estimates (e.g., aggregation error, assignment error, and sampling error).

FURTHER RESEARCH

The difficulty in assigning some highway segments to factor groups highlights the need for further research in this area. If we could easily identify some independent demographic, economic, or land use related variables related to seasonal traffic patterns – possibly using the statistical technique, discriminant analysis – then we could have more confidence in the assignment of our highway segments to seasonal factor groups. Highway functional classification, while it certainly meets the criteria of "easily identifiable", does not seem to relate well with seasonal patterns in traffic flows, at least in our state. You can see this by the significant variance in the CVs across ATRs on the same highway functional system as shown in Table 2.

Table 2

Statistics for Some Rural Interstate Highway ATR Sites

Site	*Func. Class.*	*Minimum*	*Maximum*	*Range*	***CV***
1102	1	.82	1.35	.53	**15.7**
2801	1	.86	1.22	.36	**10.6**
5101	1	.93	1.14	.20	**6.8**

CONCLUSIONS

Variation in traffic data exists as a fact of life. We recommend spending most of your traffic data analysis efforts on attempting to measure or quantify this variation, and from this, developing approaches (such as factoring short period counts to estimates of AADT) for best dealing with and reducing this variation. We also recommend bringing to bear a combination of descriptive and analytic statistical approaches to analyzing your data. Finally, we recommend marketing your traffic data as well.

Marketing your traffic data has many benefits, especially as we enter into a more data-driven decision-making world. Such marketing has the effects of preventing data collection programs from becoming the first target in times of budget cuts. At the same time we must ensure that our data have relevancy, timeliness, and measurable accuracy. By "relevancy" we mean data that we have summarized and delivered to knowledge workers and decision-makers in a form they can use. "Timelines" defines itself: if knowledge workers and decision-makers don't receive the data they need when they need it, it won't get used, and essentially becomes irrelevant. Finally, by "measurable accuracy" we mean developing precision estimates for AADT estimates, a basic "truth-in-data" principle, and important information for knowledge workers.

REFERENCES

1. *Traffic Monitoring Guide.* Federal Highway Administration, U.S. Department of Transportation, 1995.
2. *AASHTO Guidelines for Traffic data Programs.* AASHTO Standing Committee on Planning, Highway Subcommittee on Traffic Engineering, Joint Task Force on Traffic Monitoring Standards, 1992.
3. Granato, Sam. The Impact of Factoring Traffic Counts for Daily and Monthly Variation in Reducing Sample Counting Error. *Transportation Conference Proceedings*, 1998.
4. Aunet, Bruce. *A Statewide Traffic Counting Program for the Wisconsin Department of Transportation.* Wisconsin Department of Transportation, 1987.
5. *SAS/STAT User's Guide.* SAS Institute, Inc., 1988.
6. Hu, Patricia , T. Wright, and T. Esteve. Traffic Count Estimates for Short-Term Traffic Monitoring Sites: Simulation Study. *Transportation Research Record,* No. 1625, 1998, pp.26-34.
7. Davis, Gary. Accuracy of Estimate of Mean Daily Traffic: A Review. *Transportation Research Record,* No. 1593, 1997, pp. 12-16.