THE EFFECTS OF ERRORS IN ANNUAL AVERAGE DAILY TRAFFIC FORECASTING: STUDY OF HIGHWAYS IN RURAL IDAHO

FINAL REPORT

SEPTEMBER 2004

Report Budget Number KLK253
N04-12

Prepared for

IDAHO TRANSPORTATION DEPARTMENT

Prepared by

NATIONAL INSTITUTE FOR ADVANCED TRANSPORTATION TECHNOLOGY

UNIVERSITY OF IDAHO

Michael Dixon

ABSTRACT

Accurate forecasting of annual average daily traffic data (AADT) minimizes errors in design decisions. Several methods produce unfavorable results in rural Idaho where traffic data are available and growth trends are not identifiable. The classification and regression tree (CART) method can reduce the variability in the AADT annual growth rate. The maximum errors for different data subgroups were calculated and the effects of the prediction errors were evaluated. Following an asphalt overlay, using both the actual and forecasted AADT values, differences in the thickness required for each were evaluated. Second, a level of service analysis studying the differences between the values using both actual and forecasted AADTs showed that significant differences did not occur unless the ESALs were high enough to warrant more than the minimum thickness. In those cases, only ESALs with errors of greater than 20 percent exhibited large differences between the forecasted and actual AADT values. Only eight percent of the cases would have resulted in incorrect design decisions. Because incorrect design decisions rarely occurred in either case, using forecasting methods as those depicted in this study is recommended. The CART method should also be implemented to improve the classification of AADT data points.

Table of Contents

List of Figures

List of Tables

Chapter 1 ––Introduction

This report will cover the basics of AADT estimation, the background and a review of current practices throughout the nation. Next, the data available for this research will be described and trends identified. Then, a few methods that have been used to forecast AADT will be examined. The method chosen to estimate the AADT growth factor for this research is then described, justified, and tested. To take the function of AADT forecasting to the next level, two design applications where AADT was required were examined and the influence of AADT forecasting errors on design was determined.

1.1 Background

Because funding is always an issue in transportation planning, design, and improvement projects, making critical decisions in an informed manner is important. Traffic data are an important source of information for these decisions; as a result the accuracies of these data are imperative. The AASHTO Guidelines for Traffic Data Programs identifies six applications in which traffic data play a significant role. These six include: project selection, pavement design, capacity analysis, safety analysis, air quality, and traffic simulation [1]. The annual traffic volume, annual average daily traffic (AADT), is one traffic record used in these applications. Often, forecasted AADT volumes are required for use in project selection, capacity analysis, and design. Inaccuracies in traffic volume forecasts are responsible for the additional costs associated with over and under design. The costs associated with an under designed project arise when an additional project must satisfy the original inadequacies [1]. Extra materials, labor, and additional right-of-way attainment add to the cost of an over designed project [1].

In pavement design, forecasted values of AADT directly affect the estimation of future pavement deterioration [1]. This affects which roadways are candidates for overlay projects. High errors in AADT forecasts could wrongly influence which roadways planners decide to improve. Also, the required overlay thickness can be influenced greatly by the AADT estimate. This could result in over or under design if the errors in the AADT estimate are large.

Capacity analysis is used in design, planning, and operational analysis, where AADT is used in level of service analysis [1]. Under design could cause a highway project to be at or near capacity upon completion. Over design could waste precious funding that can be used in areas where the need for improvement is more crucial. There must be a range of estimated AADT values that, though not entirely accurate, would allow the correct design decisions to be deduced. The purpose of the research presented here is to develop an improved methodology for estimating future AADT values.

1.2 Review of Literature and Current Methods

In this section, the current practices are examined for both the Idaho Transportation Department and other state agencies, with a focus on the western rural areas of the United States. Previous AADT forecasting research is also explained and examined. Previous research for predicting AADT volumes includes time series models, regression models, additional models, and neural networks. Although many of the studies compare forecasting methods, the effects of the errors are not explained from a design perspective and this is one of the objects of this research.

1.2.1 Overview of Current Practice for Forecasting in Idaho

The Idaho Transportation Department currently uses annual growth rates, calculated from 20 years of past data, to forecast current annual average daily traffic (AADT) volumes to a design year in the future. In Idaho, many of the automatic traffic recorder stations (ATR) have recorded data from 1980 to the present. AADT volumes are calculated using the volumes collected from these ATR stations. Annual growth rates that represent the average percent increase in AADT volume per year are calculated at these ATR stations using Equation 1.1:

$$g = \sqrt[n]{\frac{AADT_t}{AADT_{t-n}}} - 1 \tag{1.1}$$

where

$AADT_t$ = AADT volume recorded during the most recent year t;
$AADT_{t-n}$ = AADT volume recorded n years prior to the year t; and
n = number of years between the most recent (AADT) and past ($AADT_n$) volumes.

The Guidebook to Statewide Travel Forecasting identifies the equation to forecast the AADT volumes as equation 1.2 [2]:

$$AADT_{t+n} = AADT_t(1+g)^n \tag{1.2}$$

where:

$AADT_{t+n}$ = AADT value forecasted n years in the future;
$AADT_t$ = base year AADT value observed during year t;
g = annual growth rate; and
n = number of years into the future for which a forecast is being made.

In Idaho, the accuracy of the forecasts is questionable because the annual growth rates have not been updated on a regular basis. Professionals in Idaho that use the forecasts expressed their concerns with creating a new forecasting method. These professionals want a simple model that is easy to explain, to update, and to understand [3], [4]. It is also important to realize how the forecasting errors actually affect design and planning applications when deciding the required accuracy for such forecasts.

In the urban areas of Idaho, metropolitan planning organizations use calibrated four-step models that represent the operations within the network. Currently, in Idaho there are five MPOs: COMPASS, Bannock Planning Organization, the Bonneville Metropolitan Planning Organization, Kootenai Metropolitan Planning Organization, and Lewis-Clark Valley Metropolitan Planning Organization. Land use, economic and demographic statistics, and the geometry of the network are just some of the parameters that are incorporated into these metropolitan planning models. The first three MPOs mentioned are well established and because this study deals with rural areas not within the metropolitan area, locations in these areas were not included in the scope of this study. The last two planning organizations are recent additions and may not currently have calibrated models. Therefore, rural locations within these areas were included in this project.

1.2.2 Summary of Current Practice in Idaho and Elsewhere

Because there are many techniques for forecasting AADT volumes on rural highways, other departments of transportation were contacted and the different methods were compiled. Like ITD, many departments use the growth factor method. There are other methods represented, as well, such as regression and trend analysis. The results are documented in Table 1.1. Time series and regression seem to be the most common AADT forecasting methods among the various state departments and the applicability of these methods to the Idaho data was evaluated as described in subsequent sections of this report.

TABLE 1.1 Summary of Current Practices for Forecasting in Rural Areas

Department	Method	Uniqueness of the Technique
Idaho Transportation Department	Growth factor using 20 years of past data	
Washington State Department of Transportation	Time series analysis [5]	Works well in rural areas, but not as accurate in urban fringe areas that are not included in an MPO model
Oregon Department of Transportation	Time series analysis [6]	
Montana Department of Transportation	Growth factor [7]	
Utah Department of Transportation	Time series analysis using 20 years of past data [8]	Modified by economic/demographic variables such as population, number of households, and employment.
Colorado Department of Transportation	Time series analysis using 20 years of past data [9]	Currently re-evaluating the forecasting methods
Nevada Department of Transportation	Linear regression [10]	
New Mexico Highway and Transportation Department	Growth factor using 20 years of past data [11]	Updated and evaluated regularly using statistical and conceptual methods
Florida Department of Transportation	Uses planning models whenever possible, but in very rural areas linear regression is used with 10 years of past data [12]	
Wisconsin Department of Transportation	Box-Cox regression using 21 years of historical data -- when a regression line is not significant, annual and flat growth rates are assigned[13]	

1.2.3 Literature Review

1.2.3.1 Time Series Forecasting Methods

Time series forecasting methods assume that past trends will continue into the future. With this assumption, the past data can be used to forecast AADT volumes to a specified year in the future. As cautioned in the Guidebook on Statewide Travel Forecasting circulated by the Federal Highway Administration, time series models must be used with care [2]. Because time series models use past data, this method cannot anticipate unpredictable or random events that could substantially affect the traffic volumes. Research completed by Horowitz and Farmer in 1999 suggested that many departments of transportation use some sort of a time series model for forecasting and implied that most could be providing more accurate forecasts by using a statewide model [14]. The review explained that many state departments of transportation are trying to use urban planning models in rural areas. There are changes to capacity and traffic analysis zone size that are required before the urban model can successfully work as a statewide model [14]. Horowitz suggested that some tasks could be better handled with time series analysis when a practical statewide model could not be created with an efficient use of resources. He went on to recommend that:

[I]t is important that the objectives of the model be described well ahead of any decisions on data collection, model structure, computer software, and budget. The objectives should clearly relate to ongoing policy issues and needs of state transportation plans [14].

1.2.3.1.1 Growth Factors

Many states use growth factors to forecast AADT volumes because of the simplicity of this technique. This method assumes that the past trends in percent increase in traffic volume each year will continue into the future. Any number of years of past data can be used to find a growth factor and using plenty of historical data usually minimizes the effects of spikes in the data. Many methods exist for developing a growth factor and not all are as simply calculated as the Idaho Transportation Department's technique. Memmott explored different methods for determining these growth factors. The growth factors were obtained by finding the curve that best fit the historical data [15]. Memmott showed the importance in examining the trends in past data to insure that the future trend has consistent results. Failing to do so could produce large errors in the volume estimates. He also explained that growth could take many forms between the base year and the projection year and have identical beginning and ending points (Table 1.2). Figure 1.1 shows the phenomenon mentioned [15]. Memmott notes that when finding the trend that fits the historical data most accurately that "overall, the ADT projections are good, with an average error of 28.7 percent." The statement that the projections are "good" does not clarify what the errors mean in the context of design or planning. In other words, this research lacks the explanation of how the errors would affect planning or design decisions. Research that demonstrated how the errors of forecasting methods can affect the design decisions would be beneficial.

TABLE 1.2 Types of Growth

Functional Form			Growth Rate	
$ln(ADT_t)$	=	$a + bt$	b	(1.3)
ADT_t	=	$a + bt$	$\frac{b}{(a+bt)}$	(1.4)
ADT_t^2	=	$a + bt$	$\frac{b}{2(a+bt)}$	(1.5)
$ln(ADT_t)$	=	$a + be^{\frac{-t}{10}}$	$\left(\frac{-b}{10}\right)e^{\left(\frac{-t}{10}\right)}$	(1.6)
$ln(ADT_t)$	=	$a + b[ln(t+1)]$	$\frac{b}{t+1}$	(1.7)
ADT_t	=	$a + b[ln(t+1)]$	$\frac{b}{t[a+b\ln(t+1)]}$	(1.8)

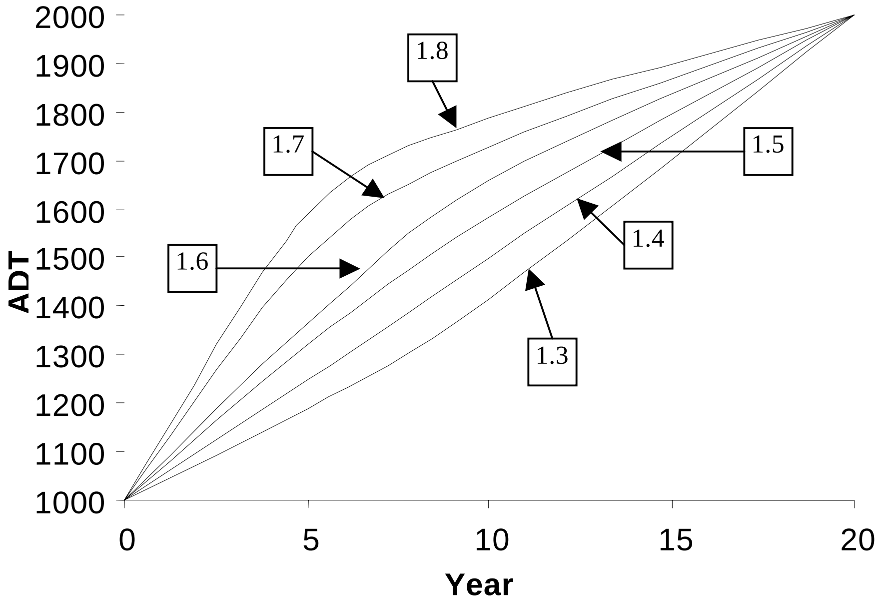

FIGURE 1.1 Different Types of Growth Between the Same Two Points [15]

1.2.3.2 Regression

Linear regression can extrapolate trends in average annual daily traffic into the future. It also uses past trends in data, but it can also incorporate the relationship between economic and demographic variables and the traffic growth pattern. A general example of a regression equation is:

$$AADT_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_{ni} + \varepsilon \tag{1.9}$$

where:

$AADT$	=	value of the dependent AADT value for the ith year;
B_0	=	constant intercept term;
β_j	=	regression coefficient for the jth independent variable;
X_{ni}	=	value of the nth independent variable for the ith year; and
ε	=	error term.

Another type of regression is lagged regression, which is actually a form of time series analysis. In lagged regression, previous values of the dependent variable are one or more of the independent variables. An example of the general form of the lagged regression equation is:

$$AADT_t = \beta_0 + \beta_1 AADT_{t-1} + \beta_2 AADT_{t-2} + \cdots + \beta_n AADT_{t-n} \tag{1.10}$$

where:

β_0	=	constant intercept term;
$\beta_1, \beta_2, \ldots, \beta_n$	=	regression coefficients; and
$AADT_{t-1}, AADT_{t-2}, \ldots, AADT_{t-n}$	=	values of AADT at prior time steps.

A lagged regression equation is also useful when the present AADT value is known with reasonable accuracy. In the forecasting equation, the dependent variable is the future year AADT and the present AADT is one of the independent variables. Variables are chosen on the basis of causal relationship to the traffic volume and high R-square values. The Guidebook on Statewide Travel Forecasting advises the analyst to look at the basis for each variable chosen to be sure that every variable has a causal relationship with the AADT forecast [2]. This means that the relationship between each variable and the traffic volume forecast should be logical.

In the same study cited above, Memmot also considers using multiple-regression for the forecast. These equations include a dummy-term for whether the capacity of the roadway is increased or not throughout the forecast period. The equations he examines all have a logarithmic transformation and are [15]:

$$\ln(ADT_t) = a_1 + a_2 t + a_3 C \tag{1.11}$$

$$\ln(ADT_t) = a_1 + a_2 \ln(t) + a_3 C \tag{1.12}$$

$$\ln(ADT_t) = a_1 + a_2 e^{\left[\frac{-t}{10}\right]} + a_3 C \tag{1.13}$$

where:

C = 1 if the capacity has increased during the forecast period and zero otherwise.

Memmott suggests that this method opens doors to more accurate regression forecasting models. A recommendation is made to study this topic further to find other variables that would significantly predict the traffic demand [15]. These equations are used to forecast the AADT values and the accuracy of the projections is influenced by several factors. One of these factors is the time period of the forecast, where as the amount of time between the base year and projection year increases, the accuracy of the forecast reduces. Also, the stage of development of the surrounding area affects the prediction ability of the model. Developing areas seem to have higher prediction errors (29.2 percent) than developed areas (24.7 percent) [15]. Memmott suggested that the amount of economic activity could have an effect on the accuracies of forecasting models and indicated that forecasting models should take this into account [15]. Similar to the part of the study that used growth factors and trend analysis to forecast volumes, the errors for the regression analysis are not explained in a design or planning sense. Also, this study does not include a multiple regression equation with economic and demographic variables such as population, land use, or employment. Research that added these explanatory variables to regression models and explained how the errors would affect design and planning decisions could be the next natural step.

1.2.3.2.1 Clustering

Often the first step in performing regression analysis is sorting the data into groups with similar characteristics. There are several ways to group transportation data. The Federal Highway Administration has set up functional classes for different types of roadways, which are documented in the Highway Performance Monitoring System Field Book and are shown in Table 1.3 [16]:

TABLE 1.3 The Highway Performance Monitoring System Functional Classes

Rural Functional Systems		Urban Functional Systems	
Interstate	Major Collector	Interstate	Minor Arterial
Other Principle Arterials	Minor Collector	Other Freeway and Expressway	Collector
Minor Arterial	Local	Other Principle Arterial	Local

Usually, the AADT of each section of roadway is also required for grouping. As a result, assigning certain roadways to different functional classes can be a difficult and subjective process [17]. On roadways without ATR stations, the AADT volume is often not measured. Garber and Bayat-Mokhtari researched a method for predicting the current year AADT value that did not require a known, or measured, AADT value and developed an alternative method for grouping the data. The method of Garger and Bayat-Mokhtari has three main steps:

1. Dividing roadways into sections that have a homogeneous traffic volume,
2. Identifying variables that significantly predict AADT, and
3. Grouping the data by similar characteristics.

In the first step, the roadway is broken into sections that have constant traffic characteristics and roadway geometry. A roadway section is defined to begin and end at either major intersections or where the geometry changes significantly. In the second step, the variables that have significant influence in predicting the AADT value were identified using an analysis of variance procedure. The significant predictor variables were [17]:

- FHWA functional class,
- Primary functional use such as recreational, local travel, or commercial,
- Land use of the county in which the roadway section lies,
- Population of the county in which the roadway section lies, and
- Type of terrain.

The third step uses the significant variables to group sections by similar characteristics using clustering capabilities in statistical software. A regression analysis was then run on the clusters of data that were formed in step three. The coefficient of variation of the AADT values for each cluster were lower than the values recommended by FHWA and did not require the initial step of estimating the AADT value on each roadway [17, 18]. FHWA recommends, in the Traffic Monitoring Guide, that the absolute precision of estimates be within 10 percent [18]. Equation 1.14 shows the relationship between the coefficient of variation and the absolute precision. Therefore, the recommended coefficient of variation depends on the number of locations.

$$D = t_{\frac{\alpha}{2},n-1} \frac{C}{\sqrt{n}} \tag{1.14}$$

where

D	=	absolute precision;
t	=	value of Student's t-distribution with 1-a/2 confidence and n-1 degrees of freedom;
C	=	coefficient of variation (equal to the ratio of the standard deviation to the mean); and
n	=	number of ATR locations in sample.

This method did not forecast AADT, but clustered roadways into groups with similar AADT values using variables that significantly predict AADT. This method may be used to forecast AADT if the variables used to predict AADT were forecasted for the forecast year. Faghri and Chakroborty tried to cluster their traffic data into the ideal number of groups. The researchers noticed a problem associated with their clustering technique. Several of the ATR stations would change clusters from year to year [19], which is impractical. The AADT volumes were not forecasted in this study either.

1.2.3.2.2 Forecasting Using Present AADT as Independent Variable

Saha and Fricker developed models to forecast AADT values using disaggregate and aggregate analysis, utilizing data from the years 1970 through 1980 from 154 ATR stations. In disaggregate analysis, a model was developed for each ATR station separately. In aggregate analysis, the ATR stations were grouped by highway type, similar to the convention shown in Table 1.3, and one model was developed for the entire roadway classification. Descriptor variables that were used in the analysis to help model the traffic demand volumes were [20]:

- Annual average daily traffic,
- County vehicle registrations,
- US gasoline price,
- Year,
- County population,
- County households,
- County employment,
- State vehicle registrations,
- State population,
- State households,
- State employment,
- Consumer price index,
- Gross national product, and
- Nationwide per capita disposable income.

An elasticity model was used to estimate the future value of the AADT. One problem the researchers noticed was that the future values of the variables were required. The model used Equation 1.15 [20]:

$$AADT_f = AADT_p\left[1.0 + \sum_{j=1}^{n} \frac{e_j^{(x_{j,f} - x_{j,p})}}{x_{j,p}}\right] \tag{1.15}$$

where

$AADT_f$ = AADT in the future year *f*;
$AADT_p$ = AADT in the present year *p*;
$x_{j,f}$ = value of variable x_j in the future year *f*;
$x_{j,p}$ = value of variable x_j in the present year *p*;
e_j = elasticity of AADT with respect to x_j ; and
n = number of associated variables.

Saha and Fricker explained that, typically, as more causal variables are included in the regression equation the accuracy is improved [20]. They included, however, that the linear regression relationship should be easy to understand and implement and; therefore, should not be extremely complex. There were two necessary requirements that were followed when choosing the variables. First, the variables had to adequately represent the trends in Indiana. Second, the data for these variables had to be easy to obtain and compile in a useful format [20]. Saha and Fricker asserted that each variable included in the models have an understandable and practical relationship with the traffic volume trends. A correlation matrix was evaluated to determine the strength of the relationship between each variable and the future AADT value. Step-wise regression was used to select the variable for inclusion in the model. In the aggregate analysis, county and state population and number of households were the best predictor variables. The forecasts were made up to twelve years into the future. The range of absolute errors for the aggregate analysis was between 0.3 and 30.4 percent with the average absolute error being 15.8 percent. When disaggregate analysis was used, the absolute errors were much lower, the range being between 1.1 to 7.0 and the average absolute error was 4.0 percent [20]. Of course, disaggregate analysis is not efficient because a model must be developed for every ATR station and many highway sections do not have an ATR station. The main problem with the linear regression model is the amount of data points required. Many states have insufficient counts to produce a statistically significant model.

There are some missing points in the research done by Saha and Fricker. One is the drawback that the input variable values must be predicted, leading to an AADT forecast that is based on predictions. Also, the accuracy of this method was not compared to other forecasting methods and there was no mention of how the errors, either aggregate or disaggregate, would affect design.

1.2.3.2.3 Estimation of Present AADT Using a Variety of Independent Variables

A 1998 study completed by Mohamad, Sinha, Kuczek and Scholer used linear regression to predict AADT volumes on county roads in Indiana using economic, demographic, and two-category quantitative variables [21]. The predictor variables that were explored were:

- County population,
- County households,
- County vehicles registration,
- County employment,
- Presence of interstate highways nearby,
- County state roadway mileage,
- Location (urban or rural),
- Access to state highway system, and
- County per capita income.

The models were developed to predict the current AADT on county roads given the significant predictor variables. Forecasts could be performed with these models if future year predictions of the variables were used, but forecasting was not the intention of the research. The researchers used SAS as their primary tool in selecting the variables, creating the models, determining the significance of each predictor variable, and determining the accuracy of the models [21]. This study tested the regression equations on other county roads that were not used in the development of the equations. Because the variables were not normally distributed, the independent variables were standardized and a transformation of the AADT was required. Each variable included in the models was a significant predictor of the future AADT. The full model, that contained all of the available statistically significant variables, accounted for 76.6 percent of the AADT variation. The number of predictor variables was kept as low as possible while still generating a reasonable level of accuracy, meaning that a minor reduction in R-square was accepted to simplify the model. The final model that was selected included four independent variables: location, access, county population, and total arterial mileage of the county and had an R-square of 75.1 percent [21]. The results depicted an absolute average error of 16.78 percent and a range of 1.6 to 34.2 percent [21]. The researchers found this method to be effective because of the models efficiency, cost-effectiveness, and simplicity. This method does not forecast AADT, but can be helpful in this research because it shows that adding complexity to the model does not always drastically increase the accuracy. Again, the study did not compare this method to other prediction techniques or state what these errors, 1.6 to 38.2 percent would mean from a planning or design perspective.

1.2.3.3 Neural Networks

1.2.3.3.1 Estimation of Current AADT

Lam and Jianmin Xu completed a study in Hong Kong in 2000 that compares regression techniques to neural networks [22]. In this study, AADT volumes are predicted from short period counts, but are not forecasted into the future. The neural networks did approximate the AADT more accurately than the regression equation, but only by one to two percent [22]. The primary problems associated with this research are: 1) the neural network method requires many neurons and weighted interconnections between the input and output at each neuron, making it difficult to explain the relationship between input variables and AADT; 2) the study did not forecast the AADT volumes; and 3) because the neural network AADT estimations were only slightly more accurate than the regression equations, the impact of the errors probably would not be crucial in design.

1.2.3.3.2 Forecasting One Hour Ahead

A report by Clark, Chen, and Grant-Muller in 1999 compared neural networks to more traditional techniques. These traditional techniques included time series methods and statistical methods such as regression, smoothing, decomposition, and Box-Jenkins techniques. Traditional methods are more simply explained and rationalized than neural networks. On the other hand, neural networks can predict more complex relationships within the system than the traditional methods can [23]. Several time series models were examined. The first is the least complex assuming that the future observation is the same as the current observation. The fifth is the most complex time series model explored. In this model, the future observation is a function of the current observation and three previous observations.

$$v_{t+1} = v_t \tag{1.16}$$

$$v_{t+1} = 2v_t - v_{t-1} \tag{1.17}$$

$$v_{t+1} = 0.5v_t + 0.5v_{t-1} \tag{1.18}$$

$$v_{t+1} = 0.5v_t + 0.25v_{t-1} + 0.25v_{t-2} \tag{1.19}$$

$$v_{t+1} = 0.25v_t + 0.25v_{t-1} + 0.25v_{t-2} + 0.25v_{t-3} \tag{1.20}$$

An ARIMA model was used for the statistical method because of the smoothing action of the technique. A general form of an ARIMA equation is:

$$Y_t = \delta + \varphi_1 T_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_n Y_{t-n} + \varepsilon_t \tag{1.21}$$

where δ and φ are estimated parameters and ε_t is an error term that is not used in forecasting. An infinite number of terms can be added to the ARIMA model, but each successive term has less and less predictive contribution. The model was created using 24 hours of data and the model was used to forecast the 25th and 26th hourly observation.

The neural network method was compared to the time series and ARIMA methods at four different sites. The neural network had an absolute average error of 11.4 percent with a range of 11.1 to 11.8 percent [23]. The time series methods had an overall average absolute error of 14.2 and a range that varied from 10.7 to 17.7 percent [23]. The ARIMA methods had an overall average absolute error of 14.6 and the range varied from 11.0 to 26.2 percent [23]. The neural network's forecasts had lower errors than the time series and ARIMA methods in three out of the four sites. The average percent errors for the neural network are 2.8 percent lower than for the other methods, but the neural network model is more difficult to create and to explain than the time series and the ARIMA models. Because there is no explanation of how this slight decrease in percent error would affect design and planning decisions, which method actually produces the correct design decision most often is unknown.

1.3 Problem Statement

To use funding in the most effective manner, over and under design must be minimized. There are many methods for forecasting traffic volumes such as regression models and time series trend analysis. Each of these methods can be customized for a particular project.

Of the studies that were examined, not one explained how differences in error associated with each method would affect decisions in design or planning applications. In Memmott's time series study, the average forecasting errors were 28.7 percent and he states that this is an acceptable error, but recommends future research in this area to improve accuracy. [15]. In the study by Saha and Fricker, the disaggregate analysis outperformed the aggregate analysis with average percent errors of 4.0 percent and 15.8 percent, respectively [20]. The study by Mohamad, Sinha, Kuczek, and Scholer declared an average error of 16.8 percent for the present AADT prediction [21]. The study states that the models were developed as "accurately as possible within the limitations of the traffic data" [21]. None of the studies addressed how the forecasting or estimation errors would affect design. Forecasts with errors such as these may or may not give the correct design decisions. The accuracy of these forecasts is important to allow accurate judgment of professionals responsible for design and planning projects that would use these values. A study that found a method appropriate for forecasting in rural Idaho and explained how the AADT forecast errors would affect decisions in design or planning applications would be valuable.

1.4 Research Approach

First, the AADT, demographic, and econometric data were compiled and organized. Only rural locations were used in this study; therefore, urban locations were removed from the data set. Some data such as existing pavement and overlay characteristics and highway characteristics such as the directional distribution factor and the peak hour factor were required, but were not economically feasible to collect. In these cases, valid assumptions were made. These assumptions did not affect the validity of the research because the differences in values were analyzed, not the individual values.

The literature review showed that there are many existing methods that could forecast AADT values with reasonable accuracy and the applicability of these methods for use in this project was assessed using the Idaho data. The methods that were investigated were 1) the Elasticity-based model and Regression 2) Time Series Analysis and 3) Clustering . These methods were found to have unfavorable results for this particular project, but helped to realize the type of method that would be promising.

The classification and regression tree method has transportation applications to date, but has yet to be applied to estimate the growth factors used to forecast AADT values. This method was verified, calibrated, and validated to evaluate its adequacy as a tool for forecasting AADT values.

The AADTs were forecasted from 1990 to 2000 and the forecasted values were compared to the actual 2000 AADT values and the errors were calculated and assessed. The level of service in the design year and the overlay thickness required for the pavement to remain adequate until the design year at each location were calculated using the actual design year AADT and the forecasted values. The errors expected when forecasting AADT and the design implications of these errors were reviewed. Recommendations for practice were made based on these findings.

Chapter 2 ––Idaho Data

Traffic data are important for making informed design decisions. Traffic volumes are one of the primary components of the traffic data. There are different forms of traffic volumes such as annual average daily traffic (AADT), commercial annual average daily traffic (CAADT), and design hourly volume (DHV). Traffic volumes used in highway planning and design are typically provided in the form of AADT. Several types of data that were required for this project were:

- ATR Station Data – There are automatic traffic recorder stations throughout Idaho that constantly record traffic and are considered the most accurate type of AADT data. The Idaho Transportation Department compiles this data annually and provided their ATR records for use in this project. For this project, only ATRs that were located in rural areas were included for the calibration of AADT models, because Metropolitan Planning Organizations have more complex models for forecasting within urban regions.
- Portable Counts – ITD provided records of locations where short or portable counts were taken in both 1990 and 2000, the current and design years for this project. These counts are not as accurate as the ATR station data, but they are considered more accurate than the AADT data where no count was taken. These portable count data were used as the validation sample for this project, because it was important to create a validation sample that was completely independent of the calibration sample.
- Entire AADT Database – ITD provided the entire AADT database, which includes the AADT, CAADT, and functional class among other useful data for most locations on highways in Idaho. The accuracy of this database is difficult to assess primarily because it is not feasible to have ATR stations or to take portable counts at all locations in Idaho. AADT values in this database are estimated based on ATR station and portable count data.
- Economic and Demographic Data – Data such as county population, county employment, the current AADT, annual population growth rate, and the functional classification of the roadway have been used in past research and were included in the data set.
- Other Data – Data required for calculating the design aspects of this report. These were usually estimated and included information such as the traffic directional and peak hour flow rates and the properties of the existing and overlay pavement.

2.1 Trends in Data

There are four notable trends in the ATR data set, two that were expected, one that was not and one that is contradictory. These trends are:

1. There is usually higher traffic growth in high-growth counties.
2. There can be dissimilar traffic growth within a single county.

3. There can be dissimilar traffic growth within a single functional class group.
4. There are instances where traffic volumes increase, but the population of the county decreases over the same period.

The first trend was expected because usually counties with high population growths have larger AADT growth on the highways than in counties with lower population growth. Figure 2.1 compares the trends of the ATR stations in Kootenai, a high growth county, with Bear Lake and Lemhi, both low growth counties. As, expected the traffic growth in the high growth county have steeper slopes than in the low growth counties. This trend occurs throughout the state, but not universally. There are instances in the high growth counties where some ATR stations exhibit low traffic growth.

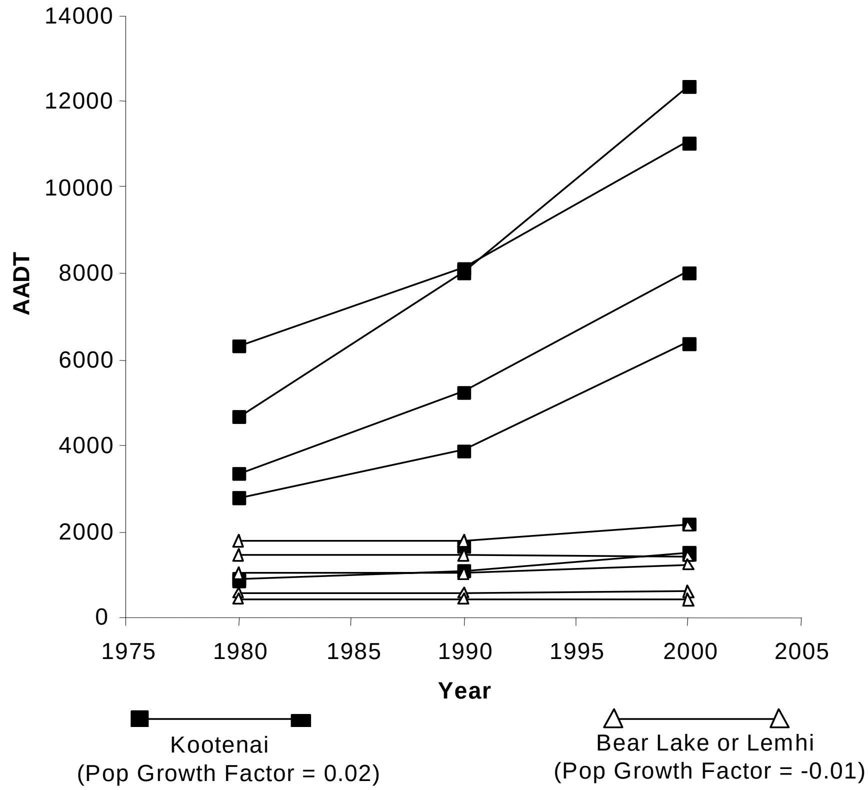

FIGURE 2.1 Trends in High Growth vs. Low Growth Counties

The second trend was also expected because different types of traffic may travel different routes through the county. For instance, commercial vehicles may be more apt to travel on the interstates and local traffic travels on arterials. In any case, it should not be surprising to find variation in traffic growth within a county. Figure 2.2 shows that there can be differences in the AADT growth within a county. In Kootenai County, the AADT growth varies substantially. ATR 27 has fairly low growth but ATR 21 has high growth. This trend occurs in counties throughout the state. Therefore, an annual growth rate assigned by county alone would not be accurate.

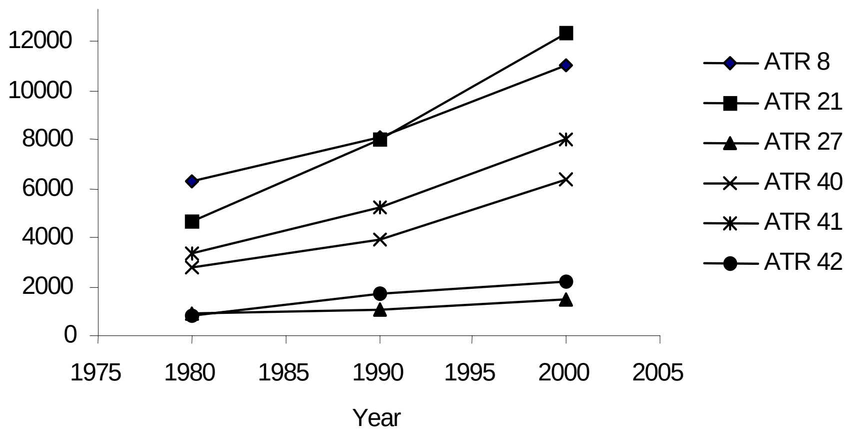

FIGURE 2.2 Differences in Growth within a County

The third trend was unexpected. In some past research, it was found that the trends in traffic growth were similar within the functional class and it was valuable to classify the traffic data based on this characteristic. This was not the case in the Idaho data, however, where there is a large range of traffic growth throughout a functional class. For instance, Figure 2.3 shows functional class 1, rural interstate. Although, most of these stations exhibit fairly high growth, there is still substantial variation within the functional class. Consequently, an annual growth rate assigned only by functional class would not match the actual trends. None of the rural interstate AADT growth is negative even though some of these stations are located in counties with negative growth. Some have a growth rate that is not constant, which could make assigning an annual growth rate to these stations difficult. However, there are stations with similar trends within the rural interstate functional classification group, however. For example, ATR 25 and ATR 7 seem to exhibit very similar AADT trends. These ATR stations are not located in the same county or even in counties that are exhibiting similar growth patterns. Therefore, it is difficult to explain why these ATR stations have similar growth when it seems that the only characteristic they have in common is their functional classification.

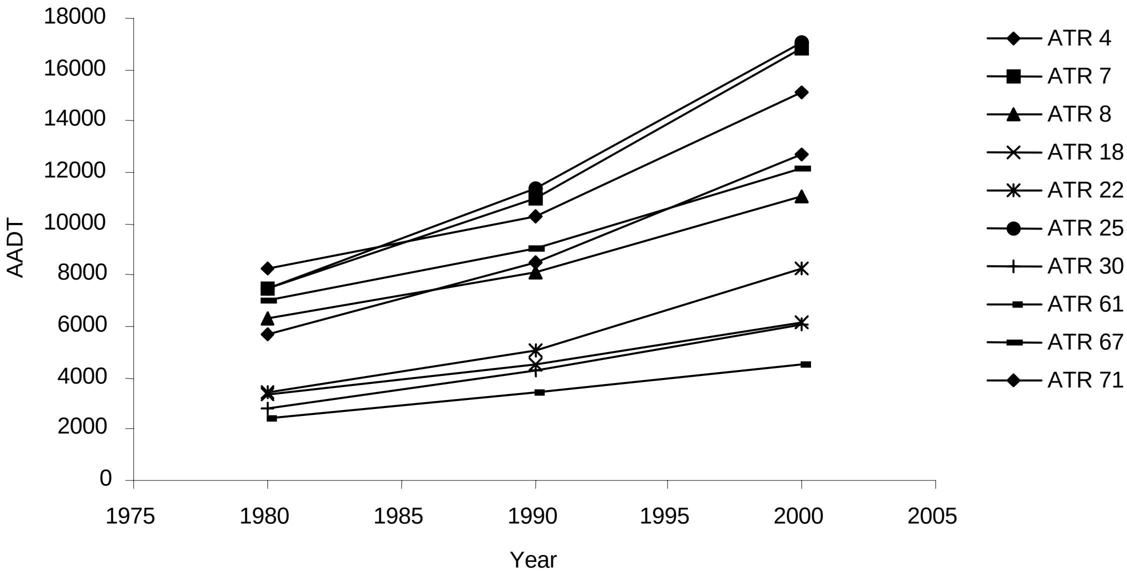

FIGURE 2.3 Function Class 1 (Rural Interstate) AADT Trends

The fourth trend contradicts the first trend in that there are instances in very low growth counties where the traffic continues to grow despite a dramatic decrease in population. One explanation could be that other routes may have become unavailable for traffic traveling through the county. This could increase the traffic on through routes. For instance, Butte County had a population growth rate of negative 1.4 percent between 1980 and 1990. Surprisingly, the ATR stations in that county all had positive growth of around two percent. This trend is shown in several other counties where there is a negative population growth over time, but the AADTs continue to increase. These locations are very difficult to forecast because their trends are atypical.

Because trends can vary greatly within a county and a functional class group, a method that takes several characteristics of the different locations into account is required. There must be a straightforward way to classify the ATR stations into groups with similar attributes and similar annual growth rates such that these characteristics could be used to classify highway stations not located at ATR stations and assign an accurate annual growth rate.

Chapter 3 –Evaluate Existing Methods

There are several mandatory characteristics for a method focusing on the estimation of annual growth rate and forecasting AADT. First, the method must be able to be used on highway sections where an ATR station is not present, and/or where past measured AADT data will not be available. A current AADT value could be obtained by completing a portable count at the site. This one data point may be the only AADT data available. Second, past and current economic and demographic data for the county in which the site is located should be easily obtainable. Third, professionals in Idaho also demand that the method be simple and easy to explain. Therefore, exceptionally complicated and data intensive methods do not fit within the requirements of this project.

In the early stages of this project, several methods for forecasting AADT were identified. Of these, some seemed especially promising and their applicability to the Idaho data was investigated. Those investigated are as follows:

- Regression;
- Time Series;
- Neural Networks; and
- Clustering.

The first three methods are models to forecast AADT. The research for this report attempted to evaluate each of these methods using calibration and validation data sets. These models had unfavorable results when applied to the Idaho data and, therefore, were not chosen as the method to forecast AADT. The subsequent sections of this chapter examine each method, explain the problems found when trying to apply the methods to the Idaho data, and the steps that lead to abandoning these models in search for a more applicable technique.

3.1 Elasticity-Based Method and Regression

One of the methods described earlier was the elasticity-based method which uses regression as a basis; therefore, given the heterogeneity of the ATR data, it is unlikely that this approach would succeed if the underlying methods for regression fail. The first step in this method is to estimate the elasticity of the dependent variable to each independent variable. This is achieved by fitting a linear regression with AADT as the dependent variable with a variety of independent variables that would seem to have an effect on the AADT growth. To adequately fit a linear regression, there has to be a linear correlation between the dependent and independent variables. Several variables that were thought to have some influence over traffic growth were chosen and their correlations with AADT found. County population had the highest correlation, although still low, equal to 0.36. Figure 3 presents a scatter plot of this data, which does not exhibit a trend. Other variables had even lower correlations with AADT and do not reveal recognizable trends, either. Therefore, fitting a linear regression to these data would not provide significant results.

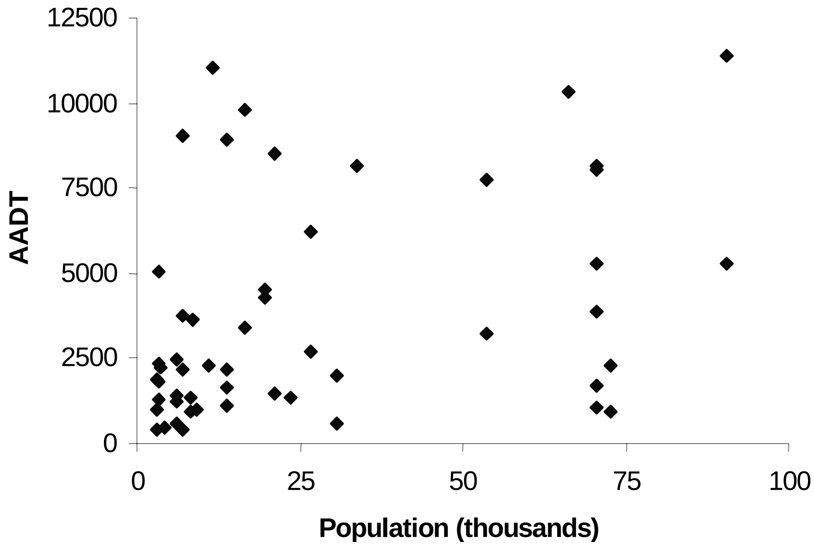

FIGURE 3.1 Population vs. AADT for All Rural ATR Stations

Saha and Fricker's Indiana data were first separated into functional class groups and regressions were run on each group. It appears that, their data were more linearly correlated once this classification was made [20]. In the case of the Idaho data set, this did not occur. Figure 3.2 shows a plot of county population versus AADT for only rural primary arterial highways. This scatter plot, which is representative of the other variables and other functional classes, does not confirm that trends occur; therefore, there would no value in continuing with this method. Because the elasticity-based method failed in the calibration phase, the regression methods had the same fate. Transformations were also attempted to investigate non-linear trends, but no strong identifiable relationships were identified. Consistent with the elasticity-based research done by Saha and Fricker, different types of classification were attempted to help identify some trends, but these efforts were unsuccessful because it seems that the data needed to be classified by not one, but many characteristics. These relationships were manually unidentifiable; therefore, a method to help identify the best way to group the ATR stations was required.

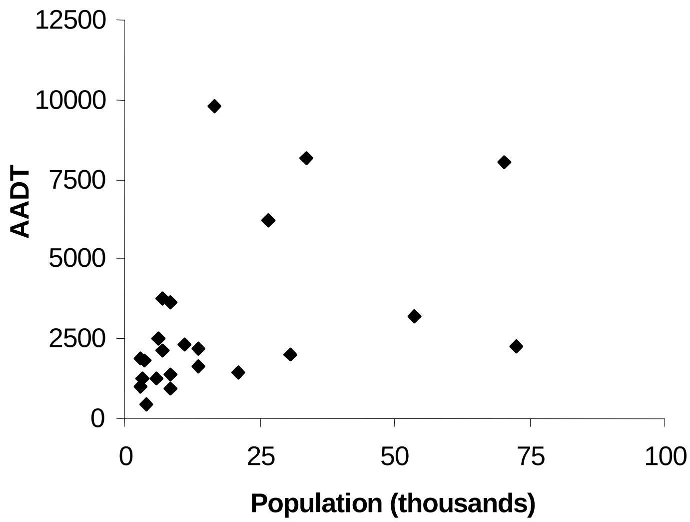

FIGURE 3.2 Population vs. AADT For Only Rural Primary Arterials

The Idaho data are not homogenous across any obvious characteristic, meaning that there are many differences in the trends ATR stations exhibited throughout the state and within counties. Therefore, classifications on combinations of these characteristics may help to identify sets of ATR stations with homogeneous trends. There must be other methods that can address this issue when recognizable trends cannot be revealed. More specifically, there is a need for an efficient method for exploring and classifying the ATR data so that trends can be established between the dependent and independent variables, thereby making regression or elasticity-based methods applicable.

3.2 Time Series

There are many types of time series methods including growth factors and ARIMA models. All the time series methods require measured historical AADT data, at least in the calibration phase. The reason that a method requiring measured historical AADT data would not work is that, in practice, not all highway sections have measured historical data. Some sort of classification method would need to be applied to highway locations based on the current characteristics. As described in the previous sections of this chapter, the Idaho data does not exhibit trends that are easily recognizable; therefore, a method that can classify the ATR stations based on the characteristics as well as the growth trends of the location would be necessary before time series methods could be evaluated.

Assigning annual growth rates is one of the most basic techniques used to forecast AADTs. An annual growth rate cannot be directly calculated for the actual highway locations in the field using any traditional methods where measured historical data are not available. Therefore, the method used in this research, to be practical, could not require measured historical AADT data. The annual growth rates for the ATR stations in the calibration data set were calculated using Equation 1.1. It should be noted that due to heterogeneity with the ATR data, the growth factor method would be more useful if there were a classification method available that created groups with similar characteristics and annual growth rates. If this were practical, then highway locations would be assigned the average annual growth rates based on which group the site fell into.

ARIMA, autoregressive integrated moving average, has not been used to forecast AADT specifically, but has been used in other forecasting applications. ARIMA methods require at least 30 observations without gaps or the parameter approximations may be meaningless [25]. Monthly Idaho ATR traffic count data were available; therefore, 30 observations could have been obtained. Unfortunately, one or more months were frequently missing from these data. To adequately use the ARIMA methodology, the missing months would have to be estimated. Also, past data will not be available in practice. Therefore, as described earlier, some way of classifying this data would be required. As also found in the regression and clustering sections of this chapter, a method is required that could classify the ATR stations based on characteristics that would reduce the variability of the AADT annual growth rate. If this initial process could be completed successfully, then it is possible that ARIMA time series forecasting techniques could be implemented using ATR stations to create the models and assigning these models to highway locations in practice based on the characteristics these highway location exhibit.

3.3 Clustering

Clustering is a method that combines groups of ATR stations based on similarities in the data. This seemed promising given the need for classification established by the review of previous existing methods. There are several clustering methods all having the hierarchical clustering procedure as a basis. The SAS Online Document gives details and descriptions of the primary clustering methods [25]. Each observation begins in a single cluster with similar observations being combined into new clusters, two at a time. Clusters are joined until all observations are included in one cluster. Before performing the cluster analysis, some assumptions are required [25].

- The variables used must be independent of one another.
- The variables must have equal variance.
- The variables must have a small variance.

Variables that were used in the clustering analysis were functional class, county population, annual county population growth rate, and AADT, which were also chosen in the research by Garber and Bayat-Mokhtari [17]. The variables used in this project are thought to be independent of one another because they have low correlations with each other. All the variables do not have low variances, however, and the magnitude of those variances vary. Because variables with large variances have a tendency to have larger effects on the results than those with small variances, it was necessary to consider scaling the variables [25]. The average linkage method for clustering was utilized and the resulting hierarchical tree is shown in Figure 3.3. ATR station numbers are used as identification on the X-axis and the average Euclidean distance between clusters is on the Y-axis. Theoretically, clusters with lower distances will have more similar characteristics. For instance in cluster 1, ATR station 8 and 21 were the first to be merged. These locations have very similar characteristics. For example, they are located in the same county and their AADTs vary by only 89 vehicles per day. ATR station 25 was then merged to the first two. This station, although located in a different county, has a similar county population, AADT, and population growth rate. Therefore, each new addition to the cluster has fewer common characteristics than the stations previously merged, but they should still have similar characteristics that make such a merge logical. The ranges of the variables in the final clusters are shown in Table 3.1. Figure 3.3 shows that ATR station 68 does not fit well within a cluster because the AADT, population, and population growth rate are equal to 8931, 13.8, and 0.0330, respectively. These characteristics are not similar to any of the existing clusters. This problem could also make it difficult for all highway locations in practice to always fit into one and only one cluster.

TABLE 3.1 Cluster Characteristics

Cluster	FC	AADT		Pop		Pop GR		AADT GR		AADT GR Var
		min	max	min	max	min	max	min	max	
1	1,2,7	7733	11362	53.8	90.6	0.0008	0.0161	-0.0026	0.0256	0.000145
2	1,2	6235	11022	7.1	33.8	-0.0024	0.0096	0.0161	0.0396	0.000085
3	2,6,7	912	5311	53.8	90.6	0.0013	0.0161	0.0171	0.0575	0.000284
4	2,6,7	432	3772	2.9	13.8	-0.0224	-0.0069	0.0153	0.0420	0.000066
5	1,2,6	568	5037	3.3	30.7	-0.0034	0.0096	-0.0027	0.0460	0.000279
6	2,6,7	445	2238	3.6	23.8	0.0133	0.0330	0.0151	0.0701	0.000504

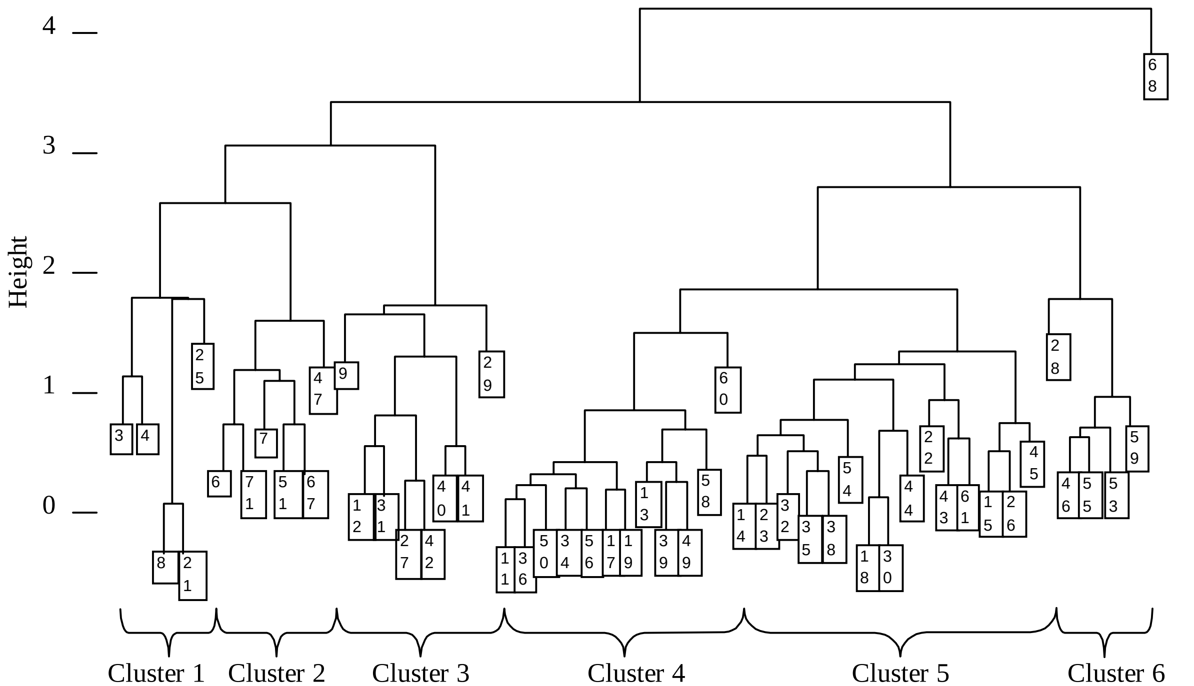

FIGURE 3.3 ATR Hierarchical Cluster Tree

The AADT annual growth rate, AADT GR, in the tables, was not included in the cluster analysis because of the assumption that past measured AADT data will usually be unknown in practice; therefore, the analysis cannot require this variable. There were two main faults of the clustering method for assigning annual growth rates: 1) annual growth rates do not have a narrow range within each cluster and 2) there are gaps between the characteristics in the clusters. The clustering process did create subsets of data with similar characteristics, but the ranges of AADT annual growth rate are still broad. For instance, cluster 6 has an AADT annual growth rate range of 0.0150 to 0.0701. In this case, it would not be reasonable to assign an average annual growth rate from this cluster because the growth rates within the cluster vary substantially. Also, there are gaps in the characteristics. For instance, data points in counties with populations greater than 33,800 and less than 53,800 or with AADTs of greater than 5311 and less than 6235 would not fit into any cluster.

Ideally, this clustering process would have provided groups of data with similar characteristics and narrow ranges of annual growth rates. Then, locations on Idaho highways, not located at ATR stations, would fit into one cluster based on their corresponding characteristics and could easily be assigned an annual growth rate based on the average rate of that cluster. Although this process did create clusters with characteristics in common, the method did not necessarily cluster the ATR stations in a way that most effectively reduced the range of growth rates in each cluster. Also, the gaps and overlaps in the characteristics of the clusters cause problems when placing data points into one, and only one, cluster. These results show that another method that also classified the ATR stations based on characteristics, but did so in a manner that always reduced the variability of the annual AADT growth rates in each subgroup and did not allow gaps and overlaps would be better suited for assigning annual AADT growth rates in the state of Idaho.

It was found, through investigation of existing methods, that a better method for classifying the ATR stations for forecasting Idaho AADTs is needed. This method needs to classify AADTs based on highway characteristics, reduce the variability in the annual AADT growth rate for each class, and not allow overlaps or gaps between classes. This seems like a logical way to classify and forecast the Idaho AADT data. Furthermore, such a classification method could also allow for implementation of some of the existing methods reviewed in this research if sufficient data were available. One classification method was found that could meet the needs of this research and it was the classification and regression tree method. The classification and regression tree (CART) method is readily implemented and the results are easy to explain. Also, this method allows for the addition of many variables to create subsets of data that have similar characteristics while reducing the variability in the dependent variable, annual AADT growth rate.

Chapter 4 ––Description and Justification of Methodology

Classification and regression trees are currently used in a variety of transportation applications, but have not been used to forecast AADTs. CART algorithms have been applied in the following areas:

- emissions estimates by classifying vehicles by type, year, engine type, among others [27];
- signal operations and timing to assist in implicating time-of-day timing plans as traffic conditions change [28];
- assigning quality ratings and methodology for rehabilitation procedures for distressed pavement [29]; and;
- activity generation in planning software from survey data to create subsets of people with similar travel patterns for use in assigning trips [30].

The CART algorithm creates regression trees using binary partitioning in which a data set is split by values in an independent variable to minimize the variation in the dependent variable in each of the two resulting sub groups, thus minimizing the deviance of the dependent variable in each sub group. The criterion used to determine the independent variable by which the split will be made is that which will result in the largest reduction in deviance of the dependent variable, which in this research was the AADT annual growth rate [26]. Research completed by Wolf, Guensler, Washington, and Bachman describes this process [27]. The method is generally described as answering two questions: First, which independent variable should be selected to create the greatest reduction in the variation of the dependent variable? And second, which value of the selected variable should be the breaking point to separate the two groups? Equation 4.1 is the objective function used for determining how to answer these two questions and it quantifies the total reduction in the deviance of the dependent variable created by a split into two new sub groups, which are referred to below as group *b* and group *c* [27].

$$\Delta = D_a - D_b - D_c \tag{4.1}$$

where:

Δ = deviance reduction after split;
D_a = deviance before the split;
D_b = deviance of subgroup b after the split; and
D_c = deviance of subgroup c after the split.

The deviances, also the sum of square errors, are calculated using Equation 4.2 [27].

$$D = \sum_{l=1}^{L} \left(Y_i - \mu\right)^2 \tag{4.2}$$

where

D = total deviance of **Y** ;
Y_l = lth observation in column vector **Y**;
μ = arithmetic mean of **Y**; and
L = sample size over which D is calculated.

This process is continued until the groups cannot be broken down any further either because the deviance could not be reduced or subsequent groups would be smaller than the minimum group size specified by the user [27]. The initial node of the tree, before any spit, is called the root. Points where splits occur are called nodes and any node that is not split further is called the termination node (TN). To automate this process, the CART algorithm is available in several statistics software packages including SAS, in the Enterprise Miner Addition, and in S-Plus, which was chosen for this project. The software allows the user to specify the minimum number of observations in a node and the minimum deviance within a node. Once either of these criterions is met a termination node is created.

For this research, the dependent variable was the annual AADT growth rate calculated using Equation 1.1 with the 1980 AADT as the past value and the 1990 AADT as the current value. Although several independent variables made up the dataset, only three independent variables were used in the regression tree analysis to explain the growth trend -- county population annual growth rate, functional class of the segment, and current AADT of the site [17, 20, 21]. These are independent variables with correlations of 0.2, which is important because the method assumes that the independent variables are not highly correlated with each other.

4.1 Sensitivity Analysis of CART Method

The CART method has not previously been used to forecast AADT. Therefore, it was important to verify the suitability of this application through a sensitivity analysis. Also, it was important to establish that the size of the calibration sample was adequate for the purpose of this research. The process for the sensitivity analysis was as follows:

- The ATR data set was split randomly into calibration and validation subsets of 42 stations and 10 stations, respectively.
- A regression tree was created using the calibration data set and average annual AADT growth rates for each terminal node were calculated.
- Annual AADT growth rates were assigned to the ATRs in the validation data set based on their characteristics, using the average annual AADT growth rates in the corresponding CART terminal nodes.

- The AADTs in the validation data set were forecasted from 1990 to 2000 with the assigned annual AADT growth rate using Equation 1.2.
- Forecast errors were then calculated by comparing the estimated 2000 AADT to the actual 2000 AADT given in the ATR database.

This process was repeated eight times so that a confidence interval of the mean error could be established to demonstrate the ability of the CART method to provide acceptable results. Ideally, this confidence interval would consistently be narrow to show that subsequent trials of this method would be expected to provide consistent accurate results. Two performance measures were used to analyze the errors: the absolute percent error and the absolute magnitude difference. The absolute percent error is calculated using Equation 4.3. This statistic depicted the percent error of the forecasted value and was used to reflect the magnitude of errors relative to the true values.

$$APE = \frac{abs(X_{CART} - X_{act})}{X_{act}} \times 100\% \qquad (4.3)$$

where

APE = absolute percent error;
X_{act} = actual value of AADT; and
X_{CART} = forecasted value of AADT

The mean absolute percent error (MAPE) is calculated by taking the average of the absolute percent error values. Trials were completed until the process depicted that additional trials of the CART method would likely provide similar results, proven when the results show a narrow confidence interval around the mean error. Eight trials were required to validate the CART method and the mean absolute percent error of all the trials had a 95 percent confidence interval less than 10 percent of the MAPE, which was 8.4 percent.

The regression trees for each trial were similar, which demonstrates that different trials of the CART method provide comparable results. However, differences did exist because of the small calibration sample sizes and the influence that one data point could have on the partitioning. Two of the eight trees are shown in Figure 4.1 to show that similarities exist. Presumably, there would have been fewer differences in trial trees had there been a larger calibration data set because the effect that one station could have on the partitioning would have been reduced. There were typically six or seven terminal nodes. Because each validation subset only included 10 ATR stations, all terminal nodes were not tested in each trial. However, this was necessary because there were few samples in the ATR data set and most were required to create, or calibrate, the regression tree. It was assumed that the limitations of a small validation sample size were effectively addressed by analyzing the results of eight iterations. The following trends were observed when comparing the trees resulting from the eight iterations.

- Several similar partitions occurred in most of the trial regression trees including partitions based on: county population growth, 1990 AADT, and functional class of the highway.
- Population growth was the most frequent first partition in the regression trees, dividing the high growth counties from the low to moderate growth counties.
- Generally, rural highways in the high growth counties have higher traffic growth than the low to moderate growth counties; therefore, this partition was expected. Notice how both trees were initially split into data sets with county population growth rates greater than or less than 0.7 percent. Also notice that the right hand side of the trees (the higher county population growth data set) have AADT annual rates greater than or equal to the left hand side (the lower county population growth data set).
- Low volume highways generally exhibit lower traffic annual growth rates than higher volume highways; thus, the next partitions were also anticipated. Notice how the second partitions split off the low volume highways. Also notice that the lower volume highways have AADT annual growth rates that are consistently lower than or equal to the higher volume highways.
- Because the partitions follow trends that were expected, this means that subsequent trials of the CART method would provide similar regression trees, even more similar if the calibration data set were larger.

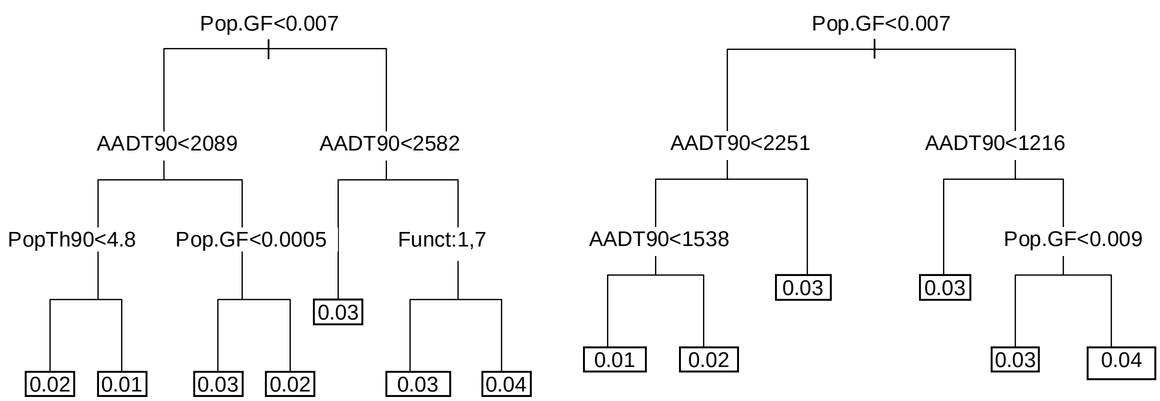

FIGURE 4.1 Typical Verification Trial Regression Trees

In summary, the sensitivity analysis demonstrated that the CART method provided similar regression trees with acceptable mean absolute percent errors. Because the trials provided consistently accurate results, this process established that the CART method would produce reliable regression trees for determining AADT annual growth rates using the Idaho data.

4.2 Final Calibrated Regression Tree

Data from the years 1980 and 1990 for all fifty-two ATR stations were used in the final calibration data set and the resulting regression tree, created using the CART method, is shown in Figure 4.2. The original regression tree consisted of eight terminal nodes with AADT growth rates ranging from less than one percent to over four percent. It was noted that terminal nodes 2 and 3 and 6 and 7 were very similar, respectively, and the final split did not add much to the accuracy of the AADT growth rate prediction. Therefore, the data points in terminal nodes 2 and 3 and 6 and 7, were combined into terminal nodes 2/3 and 6/7, respectively (see Appendix D). As a preliminary test, the AADT for the ATR stations used to calibrate the regression tree were forecasted from the year 1990 to the year 2000. As expected, the overall mean absolute percent error of 9.6 is within the 95 percent confidence interval that was created during the sensitivity analysis of the CART method. Based on Figure 4.2, a notable trend was found. One would expect that the terminal nodes with the highest deviances would have the highest MAPE and this is verified where terminal nodes one and eight have the highest deviances and MAPE. However, both MAPE and magnitude difference are required to adequately evaluate performance, where lower volume highways may have higher MAPEs, but lower magnitude differences and higher volume highways may have lower MAPEs and higher magnitude differences.

Similar to the trial regression trees developed as part of the sensitivity analysis, the county population annual growth rate was the most influential variable in the creation of the final regression tree, occurring in the tree in four different instances. Extremely high growth counties and extremely low growth counties were separated, terminal node eight and four, respectively. Terminal node 6/7 addresses the situation where AADT growth is sizable while population growth is small, which is why the regression tree works well for the Idaho data. Terminal node 6/7 is comprised of data points in counties that have many through routes, which could explain why the AADT growth tends to be inconsistently large in comparison to the population growth in these counties. The MAPE shown in figure 4.2 is calculated using validation data. This statistic, along with the magnitude percent difference for the same data, will show how well the growth rates forecast points that we not used to create the regression tree. This will be further addressed in the next section.

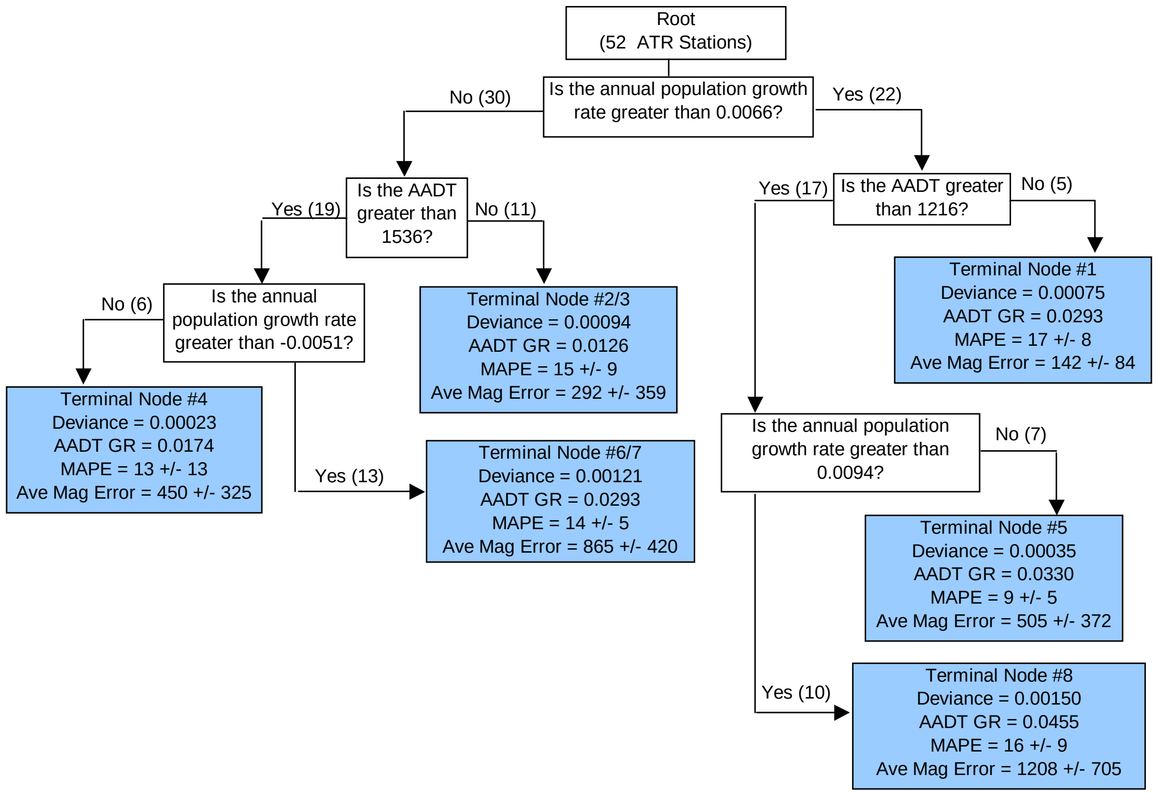

FIGURE 4.2 Final Calibrated Regression Tree

This CART method was also completed for the commercial AADT or CAADT. The data set for the CAADT was considerably smaller, however, with only 28 ATR stations having adequate data for inclusion in the CART analysis. It is not expected that the CAADT CART results represent the anticipated accuracies had the data set been much larger. Furthermore, the CAADT regression tree is provided here to show that this process would also work for predicting commercial traffic.

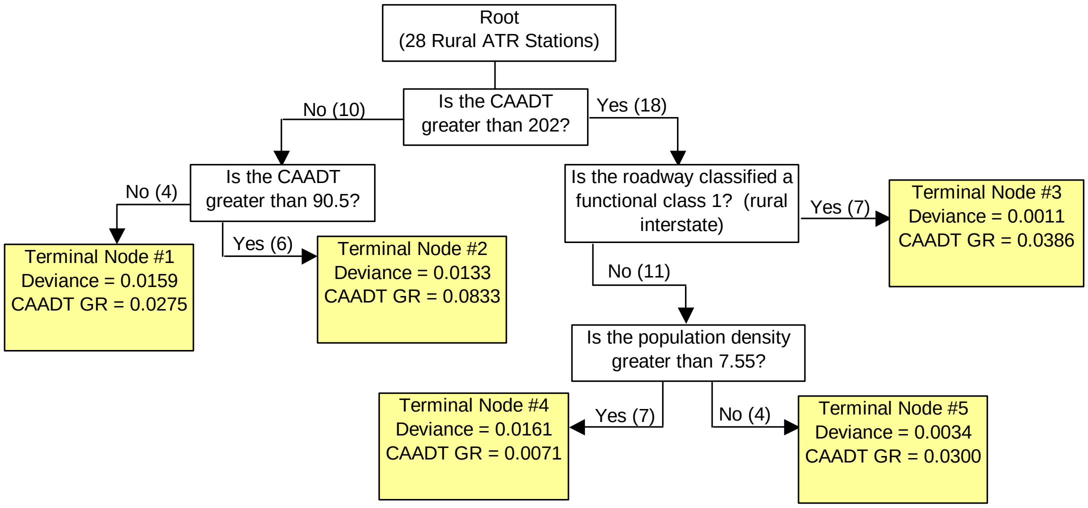

FIGURE 4.3 Calibrated CAADT Regression Tree

4.3 Validating the Regression Tree

Rural portable count locations were used to complete a validation test of the final regression tree. A stratified random sample, based on the terminal node, TN, characteristics, was used to ensure that all TN in the final tree were adequately tested. Portable counts were selected in order to make the number of samples in each TN equal to at least 10. This was not possible for every TN. The smallest sample size was 7 portable counts in terminal node 5. A validation sample size of 64 was created. The year 2000 AADT values were forecasted using the average annual AADT growth rates obtained from the final regression tree for each of the portable count locations in the validation sample. The details on the accuracy of each terminal node in the validation data set are shown in Table 4.1. The MAPE for the entire validation sample is 14.2 percent, which is consistent with the range of errors recommended by professionals in Idaho (i.e., 10 to 20 percent) and is comparable to or better than results reported in previous research. Saha and Fricker's elasticity-based model for Indiana produced a mean absolute percent error of 15.8 percent [10]. Memmott's method of fitting the best growth trend to historic data to produce 20-year forecasts had mean absolute errors of 28.7 percent for Dallas, Texas [11]. Of course, these studies did not consider conditions that were identical to those found in Idaho, but they do indicate that the proposed method is adequate. The MAPEs and average magnitude differences in Figure 4.2, as well as the 95 percent confidence interval, are based on the errors resulting from the validation data set. The existing growth factors from the ITD database were also selected for each portable count and used to forecast the 1990 data to the year 2000. This comparison shows that the CART method and the existing ITD method are comparable.

Overall, the ITD method had higher accuracy, but considering the following factors, the CART method is promising:

1. the ease of the CART method,
2. the small amount of data in the calibration data sets,
3. the potential to update ITD growth factors on a more frequent cycle.

TABLE 4.1 Validation Dataset Errors and Details

TN #	CART Growth Rate	ITD Growth Rate	Ave APE(CART)	Ave APE(ITD)	Ave Magnitude Difference (CART)	Ave Magnitude Difference (ITD)
1	0.0293	0.02	16.8	14.0	142	138
2/3	0.0126	0.02	14.8	15.9	292	298
4	0.0174	0.03	12.7	11.2	450	498
5	0.033	0.03	9.3	10.5	505	518
6/7	0.0293	0.03	13.9	12.5	865	833
8	0.0455	0.03	16.4	11.6	1208	1163
OVERALL AVERAGE =			**14.2**	**13.4**	**577**	**572**

Figure 4.4 is a scatter plot comparing the actual AADT against the forecasted AADT values for the validation sample. A line representing a perfect correspondence between the two is shown on the plot as well as two lines that represent a +/-10 percent deviance from that line. The +/- deviance from perfect correspondence was chosen because it represents the lower bound on the Idaho professional guidelines for AADT prediction accuracy. This figure shows that most data points fall close to the line of perfect correspondence, if not within the boundary. Figure 4.4 illustrates the occurrence of high percent errors in the low AADT region; therefore, the magnitude difference may be more useful to evaluate the accuracy of AADT forecasting in low volume locations. There were three obvious outliers identified using Figure 4.4. Two of these located are located in both terminal node 2/3 and Lincoln County. Both had much higher growth rates from 1990 – 2000 than predicted by the CART method. The AADT trends at these locations did not resemble the traffic growth from 1980 – 1990; therefore, such dramatic changes are difficult to forecast. Figure 4.5 shows the traffic trend of the portable count located on US-26 in Lincoln County. Notice the difference in growth between 1980 and 1990 versus 1990 and 2000. Figure 4.6 shows the AADT growth trend of the portable count on US-93 in Blaine County. This location has fairly consistent growth throughout the decades and is easier to forecast. The results have a R-squared value of 0.96, which indicates that the model accounts for 96 percent of the variation between the dependent and independent variables.

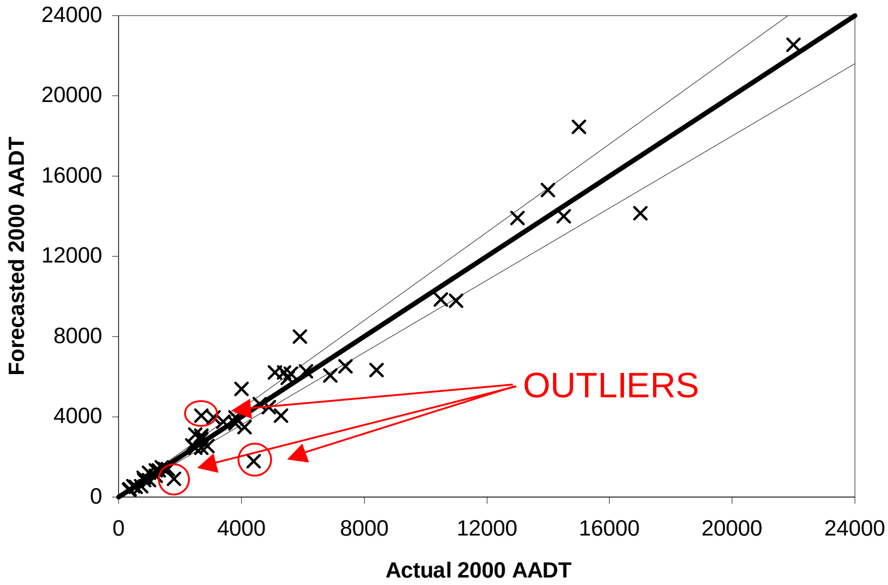

FIGURE 4.4 Scatter Plot of Actual AADT vs. Forecasted AADT

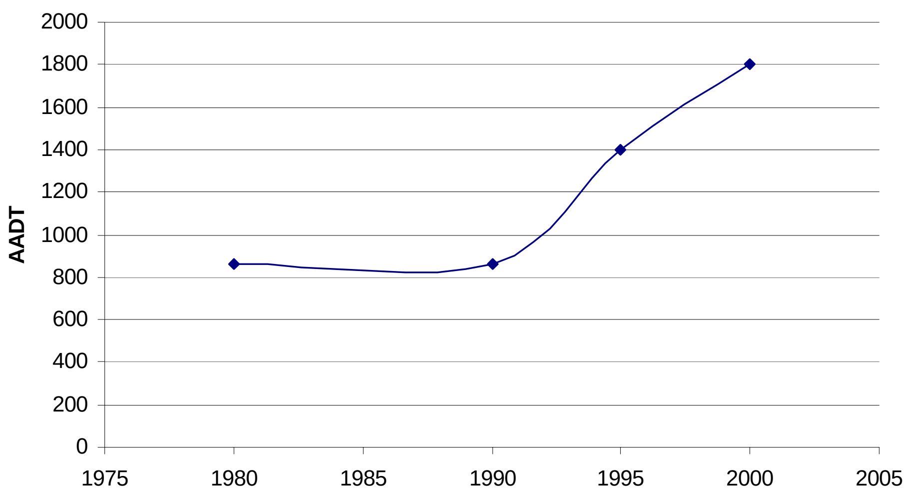

FIGURE 4.5 Portable Count on U.S. 26 in Lincoln County

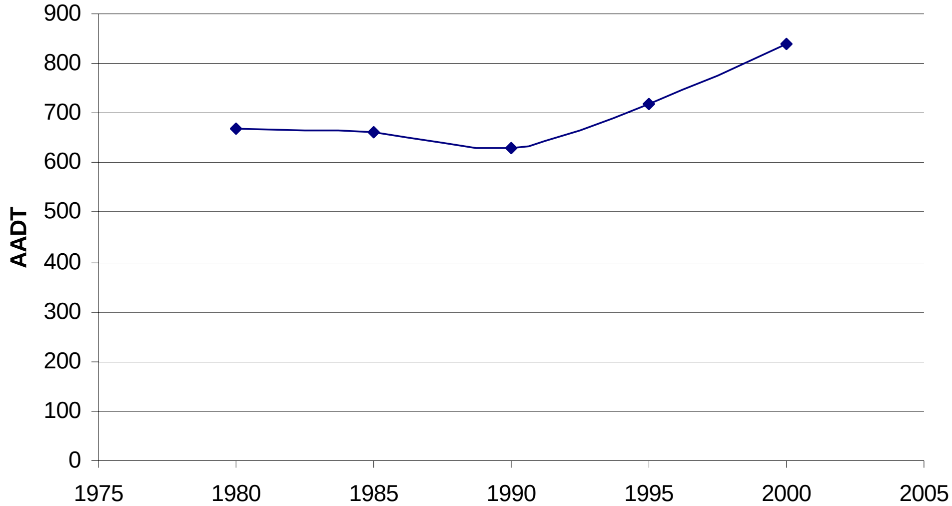

FIGURE 4.6 Portable Count on U.S. 93 in Blaine County

Several methods were examined in the early stages of this project such as linear regression, time-series, an elasticity-based method, and clustering. These methods did not provide promising results, but did indicate that a better method for classifying ATR data was needed. The classification and regression tree (CART) method is easily implemented and the results are easy to explain – professionals in Idaho that use the AADT forecasts set up both criteria. Also, this method allows for the addition of many variables to create subsets of data that have similar characteristics while reducing the variability in the dependent variable.

Chapter 5 ––Impacts of Errors

The importance of this project is to assess the impacts of AADT forecasting errors in transportation design and planning applications. Two applications were chosen for this assessment 1) the design of an overlay thickness and 2) a level of service analysis. Both applications were analyzed using both the actual and forecasted AADTs and differences were evaluated and explained.

5.1 Overlay Thickness

The equivalent single axle load, ESAL, is needed to calculate the overlay thickness required on rural roadways. ESALs are calculated using the current AADT, the growth factor, truck factors which are the average number of 18-kip single-axle load applications per truck for different roadways, and the percentage of the total traffic that are commercial vehicles [31].

$$ESAL = \left(AADT_{1990}\right)\left(T\right)\left(T_f\right)\left(G\right)\left(Y\right)\left(L\right)\left(365\right)\left(D\right) \tag{5.1}$$

where:

T	=	percent trucks in decimal format;
T_f	=	truck factor that varies by facility type;
G	=	growth factor;
Y	=	design period in years;
D	=	directional distribution factor; and
L		lane distribution factor.

The growth factor is found using the same equation that was used to find the annual AADT growth rates used in forecasting AADT [2].

$$G = \left(1 + g\right)^Y \tag{5.2}$$

where:

g = the annual AADT growth rate; and
Y = the design period in years.

Several assumptions were made when analyzing the pavement. The values were taken from acceptable ranges given in *Pavement Analysis and Design* by Huang and are shown in Table 5.1 [31].

TABLE 5.1 Assumptions for Overlay Thickness Design

Variable	Assumed Value
Truck factor (T_f) for:	
Rural interstate	0.52
Rural principle arterial	0.38
Rural minor arterial	0.21
Rural major collector	0.30
Directional distribution factor (*D*)	0.50
Design period (*Y*)	10 years
Modulus of Elasticity (ksi):	
Old asphalt layer	235
Base layer	25
Subbase layer	4
Subgrade layer	4
Overlay	350
Poisson Ratio:	
Old asphalt layer	0.35
Base layer	0.40
Subbase layer	0.40
Subgrade layer	0.45
Overlay	0.35
Existing Thickness (in):	
Old asphalt layer	3.5
Base layer	8.0
Subbase layer	16.0
Temperature at testing (degrees F):	
Pavement	91
Overlay	77
Minimum thickness for overlay (in)	0.5

In *Pavement Analysis and Design,* the minimum recommended design period for asphalt pavements is 15 years; however, the Idaho data only allowed for a 10-year design period [31]. Therefore, the ESALs would have been higher had a 15-year horizon been used. Additional assumptions include: 1) The base, subbase, and subgrade are all made up of linear granular material; 2) Failure could be caused by either fatigue or rutting in either the overlay or the existing pavement; and 3) The seasonal variation factors that modify the normal moduli values, based on seasonal temperatures and durations, are determined by locating the site in one of six Idaho climatic zones.

The WINFLEX 2000 software was used to determine two overlay thickness that were required using 1) the actual annual AADT growth rate and 2) the annual AADT growth rate found using the forecasting method [32]. These values and the differences were explained and the discrepancies noted later in this discussion.

5.2 Level of Service

The HCS2000 software was used to find the level of service for each highway using the forecasted AADT and the actual AADT values [33]. Similar to the overlay thickness application, a set of assumptions, based on acceptable ranges recommended in *Traffic Engineering,* was required to perform the analysis and they are shown in Table 5.2 [24].

TABLE 5.2 Assumptions for HCS2000 Capacity Analysis

Rural two-lane or multi-lane highways	Rural Interstates
K = 0.2 D = .70 Level Terrain PHF = 0.88 FFS = 60 mph Percent No Passing = 50 (only for 2-lane) 5 access points per mile	K = 0.15 D = .65 Level Terrain PHF = 0.85 FFS = 70 mph

For two-lane rural highways, the HCS2000 software determines the level of service based on the percent time spent following and the average travel speed. The LOS for multilane highways is determined using the average speed, density, volume to capacity ratio, and the service flow rate. The analysis of rural interstates uses density to determine the LOS. This software conforms to the guidelines set up in the most recent version of the Highway Capacity Manual, which describes these procedures in more detail [34].

5.3 Effects of Errors

Summaries of both transportation applications, overlay design and LOS analysis, are displayed in Tables 5.3 and 5.4. Table 5.3 shows that nearly half of the portable count locations have low ESALs only requiring the minimum overlay thickness of 0.5 inches. This means that the discrepancies in the overlay design were only apparent when the ESALs were high enough to require more than the minimum thickness. One of the locations in Lincoln County had an APE of 49.8 percent, but had a low AADT and ESAL. Therefore, because only the minimum was required, there was no difference in the required overlay thickness. Designers should not concern themselves with the accuracy of the AADT forecast in overlay design unless the calculated ESALs are sufficiently high to require a thickness greater than the minimum.

Of those locations that required more than the minimum thickness, 65 percent had less than a 0.3-inch difference between the actual overlay thickness and thickness calculated using the CART annual AADT growth rates. Forty-five percent of these locations had differences of less than 0.2 inches. The locations with great discrepancies in the overlay thickness design also had high forecasted AADT absolute percent errors. For instance, the portable count location in Lincoln County in terminal node three has an APE of 59.1 percent and an overlay thickness difference of 1.9 inches. Because the true annual growth rate was much greater than predicted, the actual thickness needed was greater than the thickness calculated using the CART annual growth rate. The other portable count locations with overlay thickness differences of greater than 0.3 inches also had APE of greater than 20 percent.

TABLE 5.3 Overlay Design

TN#	County	Route	MP	ESAL (act)	ESAL (CART)	Abs Difference between CART and Actual Overlay Thickness Requirements (in)
1	Oneida	SH-36	100.2	25,295	29,671	0
1	Madison	SH-22	68	15,768	12,716	0
1	Kootenai	SH-97	82.196	32,850	31,093	0
1	Custer	SH-75	191	45,530	30,694	0
1	Blaine	US-93	199.37	128,159	128,301	0
1	Oneida	SH-38	1.484	32,850	39,036	0
1	Jefferson	SH-33	59.062	57,488	50,517	0
1	Boundary	SH-1	0.1	76,285	91,643	0
1	Custer	US-93	160.2	61,028	76,277	0
2/3	Clearwater	SH-7	41	14,235	12,100	0
2/3	Latah	SH-8	36.81	31,025	31,027	0
2/3	Owyhee	SH-51	63.395	19,316	22,500	0
2/3	Camas	SH-46	42.952	23,393	24,821	0
2/3	Owyhee	SH-51	63.595	32,122	25,435	0
2/3	Power	SH-37	68.514	45,805	37,232	0
2/3	Lincoln	US-26	165.335	55,188	29,885	0
2/3	Caribou	SH-34	40	87,600	97,299	0
2/3	Bannock	US-91	22	94,663	83,356	0
2/3	Idaho	US-12	73.965	162,279	155,629	0
2/3	Clearwater	SH-11	0.2	95,564	99,285	0
2/3	Minidoka	SH-25	37.775	37,697	37,232	0
2/3	Gooding	US-30	173	49,056	45,870	0
2/3	Gooding	US-26	139	86,231	87,309	0
2/3	Adams	US-95	146	176,843	197,759	**0.2**
2/3	Lincoln	US-93	73.095	450,775	173,011	**1.9**
4	Lewis	US-12	52.389	319,010	312,355	**0.1**
4	Clark	I-15	166.352	526,752	465,989	**0.2**
4	Idaho	US-12	73.7	332,880	357,317	**0.1**
4	Idaho	US-95	240.04	270,465	299,646	**0.2**
4	Caribou	SH-34	59.252	82,125	123,249	0
4	Caribou	US-30	386.55	901,550	824,238	**0.2**
4	Idaho	SH-13	0.1	54,038	58,064	0
4	Shoshone	I-90	47.328	1,892,686	1,837,499	0
5	Oneida	SH-38	0.48	82,125	92,584	0
5	Jefferson	I-15	149.143	730,711	624,858	**0.3**
5	Valley	SH-55	121.75	221,920	214,930	0
5	Jefferson	I-15	134.89	812,837	818,139	0
5	Bonneville	US-26	343.5	374,490	358,857	0
5	Canyon	SH-55	8.718	343,283	370,467	**0.2**
5	Oneida	I-15	16.241	1,248,618	1,115,647	**0.2**
5	Canyon	SH-19	16.6	0	0	0
6/7	Benewah	US-95	389.84	298,205	260,942	**0.1**
6/7	Elmore	I-84 BUS	0.224	193,906	231,910	**0.3**
6/7	Owyhee	SH-55	0.1	205,970	242,346	**0.1**
6/7	Lincoln	SH-75	74.16	168,630	168,225	0
6/7	Adams	SH-55	155.852	228,855	280,843	**0.4**
6/7	Gooding	SH-46	10.702	161,184	120,984	0
6/7	Nez Perce	US-12	14.801	679,630	620,562	**0.3**
6/7	Franklin	US-91	7.074	141,803	161,902	0
6/7	Fremont	US-20	347.751	516,921	550,172	0
6/7	Bannock	I-15	40.108	1,419,704	1,150,122	**0.5**
6/7	Gooding	I-84	140.905	2,239,166	1,927,965	**0.2**
6/7	Power	I-86	52.875	1,958,736	1,907,371	**0.1**
6/7	Jerome	I-84	161.818	3,307,265	2,998,553	**0.3**
8	Custer	US-93	244.425	173,375	185,816	0
8	Boundary	US-95	522.829	319,010	338,909	**0.1**
8	Blaine	SH-75	102.224	306,600	413,441	**0.4**
8	Bonner	US-2	6.922	299,592	344,227	**0.1**
8	Boundary	US-2	64.45	381,425	465,197	**0.3**
8	Kootenai	SH-53	13.659	452,235	614,186	**0.5**
8	Blaine	SH-75	119.789	222,285	238,036	**0.7**
8	Kootenai	US-95	439.025	1,165,080	1,273,985	**0.1**
8	Kootenai	I-90	21.77	1,811,641	2,173,690	**0.3**
8	Bonner	US-95	475.4	1,373,130	1,406,962	0

Usually, designers would only be concerned with the results of LOS analysis, if the outcomes were unacceptable. For instance, a minimum LOS classification of C may be chosen for all rural highways. In this case, the designer would only be concerned if the analysis provided classifications of D or worse, when it actually should have been C or better using the correct AADT growth rate. For this reason, the AADT forecast errors on low volume highways do not hold much importance because the correct design decision would be determined regardless. Because terminal nodes one and two/three had low AADT values, all have future LOS classifications of A, B, or C. Therefore, these highways have acceptable service levels and would not need improvement despite forecast errors. In terminal node two/three, the location in Lincoln County was an exception. This portable count location had a high MAPE and magnitude difference errors meaning that the annual growth rate used to forecast was far different from the actual annual growth rate. As a result, the LOS classifications did not agree to the extent that an incorrect design decision would be made.

Only ten of the 64 portable count locations have differences between LOS classification using the actual and CART calculated forecasted design hour volumes, meaning that 84 percent of the locations in the validation data set had no discrepancy. Because four of these nine locations had LOS classifications on the C – D boarder, only eight percent of the validation data would result in an incorrect conclusion. For instance, in terminal node six/seven, the portable count location in Adams County actually has a year 2000 LOS of C, but an LOS of D was found using the annual growth rate from the CART forecasting method. In this case, funds may be used inefficiently to prematurely improve the quality of service on the highway.

TABLE 5.4 LOS Analysis

TN#	County	Route	MP	Lanes	%trucks	DHV 2000 Actual	DHV 2000 CART	Actual LOS	CART LOS
1	Boundary	SH-1	0.1	2	11	140	168	B	B
1	Madison	SH-22	68.0	2	6	67	54	A	A
1	Jefferson	SH-33	59.1	2	13	168	148	B	B
1	Oneida	SH-36	100.2	2	14	46	54	A	A
1	Oneida	SH-38	1.5	2	7	115	136	A	B
1	Custer	SH-75	191.0	2	12	139	93	**B**	**A**
1	Kootenai	SH-97	82.2	2	11	77	73	A	A
1	Blaine	US-93	199.4	2	22	118	118	A	A
1	Custer	US-93	160.2	2	8	154	192	B	B
2/3	Clearwater	SH-11	0.2	2	15	168	175	B	B
2/3	Minidoka	SH-25	37.8	2	5	196	194	B	B
2/3	Caribou	SH-34	40.0	2	16	140	156	B	B
2/3	Power	SH-37	68.5	2	8	139	113	B	B
2/3	Camas	SH-46	43.0	2	9	66	70	A	A
2/3	Owyhee	SH-51	63.4	2	14	50	59	A	A
2/3	Owyhee	SH-51	63.6	2	11	102	81	A	A
2/3	Clearwater	SH-7	41.0	2	13	28	24	A	A
2/3	Latah	SH-8	36.8	2	17	48	48	A	A
2/3	Idaho	US-12	74.0	2	18	182	175	B	B
2/3	Gooding	US-26	139.0	2	15	210	213	B	B
2/3	Lincoln	US-26	165.3	2	8	252	136	B	B
2/3	Gooding	US-30	173.0	2	8	224	209	B	B
2/3	Bannock	US-91	22.0	2	19	182	160	B	B
2/3	Lincoln	US-93	73.1	2	15	616	236	**D**	**C**
2/3	Adams	US-95	146.0	2	17	210	235	**B**	**C**
4	Clark	I-15	166.4	4	20	283	249	A	A
4	Shoshone	I-90	47.3	4	17	2030	1961	B	B
4	Idaho	SH-13	0.1	2	3	658	707	C	C
4	Caribou	SH-34	59.3	2	6	378	567	C	D
4	Lewis	US-12	52.4	2	18	350	343	C	C
4	Idaho	US-12	73.7	2	16	420	451	C	C
4	Caribou	US-30	386.6	2	27	686	627	**D**	**D**
4	Idaho	US-95	240.0	2	11	476	527	**C**	**D**
5	Oneida	I-15	16.2	4	20	722	635	A	A
5	Jefferson	I-15	134.9	4	20	449	452	A	A
5	Jefferson	I-15	149.1	4	20	400	339	A	A
5	Canyon	SH-19	16.6	4	6	1050	1242	**B**	**C**
5	Oneida	SH-38	0.5	2	6	378	426	C	C
5	Valley	SH-55	121.8	2	8	532	515	C	C
5	Canyon	SH-55	8.7	2	9	770	831	D	D
5	Bonneville	US-26	343.5	2	10	756	724	D	D
6/7	Bannock	I-15	40.1	4	20	819	642	A	A
6/7	Gooding	I-84	140.9	4	25	1073	915	A	A
6/7	Jerome	I-84	161.8	4	25	1658	1433	B	B
6/7	Elmore	I-84 BUS	0.2	2	14	350	419	C	C
6/7	Power	I-86	52.9	4	23	1024	997	A	A
6/7	Gooding	SH-46	10.7	4	6	742	546	A	A
6/7	Owyhee	SH-55	0.1	2	11	378	445	C	C
6/7	Adams	SH-55	155.9	2	11	434	533	C	D
6/7	Lincoln	SH-75	74.2	2	12	532	531	D	D
6/7	Nez Perce	US-12	14.8	2	14	966	882	**E**	**D**
6/7	Fremont	US-20	347.8	4	13	854	912	A	A
6/7	Franklin	US-91	7.1	2	7	784	895	D	D
6/7	Benewah	US-95	389.8	2	16	378	331	C	C
8	Kootenai	I-90	21.8	4	15	1463	1798	A	A
8	Kootenai	SH-53	13.7	2	20	826	1122	F	F
8	Blaine	SH-75	102.2	2	20	560	755	C	C
8	Blaine	SH-75	119.8	2	4	1820	1949	F	F
8	Boundary	US-2	64.5	2	11	714	871	D	D
8	Bonner	US-2	6.9	2	8	756	869	**E**	**F**
8	Custer	US-93	244.4	2	10	336	360	C	C
8	Boundary	US-95	522.8	2	17	378	402	C	C
8	Kootenai	US-95	439.0	2	12	1960	2143	D	D
8	Bonner	US-95	475.4	2	9	3080	3156	F	F

In this chapter, two design applications where AADT was a required input were examined by investigating the differences between the findings when using the actual AADT and the CART forecasted AADT. The first application was an overlay thickness design. The accuracy of the AADT forecast had no influence on overlay design unless the estimated ESALs were sufficiently large to require a thickness greater than the minimum. In most cases, large differences in the required overlay thickness did not result unless the APE of an AADT forecast was greater than 20 percent and the ESALs were large enough to require an overlay thickness greater than the minimum. These two conditions were only met in approximately 10 percent of the cases in the validation subset. Consequently, large errors in overlay design due to the inaccuracy of the AADT forecast should rarely occur in rural Idaho when using the CART method to assign annual growth rates. However, the frequency of large errors will increase as the number of years to the design year increase.

The other design application investigated in this project was an LOS analysis. Because designers would only be concerned with the LOS analysis results if the locations were predicted to provide unacceptable levels of service, the accuracy of the AADT forecast was less important on low volume roadways where traffic induced congestion was not an issue. In addition, discrepancies were only a problem if the forecast would result in an incorrect design decision. For instance, if the minimum LOS classification was set at C and the actual and the forecasted AADTs provided LOS classifications of A and B, respectively, the design decision that no improvements were necessary would be the correct result in each case. However, if the actual and forecasted AADTs provided LOS classifications of D and C, respectively, then the decision that no improvements were necessary would be incorrect. Therefore, design decisions based on LOS analysis would only provide an incorrect result when the LOS outcomes straddled the C – D boarder. Of the 64 portable count locations used in the validation data set, only ten had actual and forecasted AADT values that resulted in different LOS classifications. And of those, only in five cases would an incorrect design decision result, using the minimum acceptable LOS of C. As a result, an LOS analysis performed using AADT forecasts, computed using the CART annual growth rates, would seldom result in an incorrect design decision.

Chapter 6 ––Conclusions and Recommendations

This paper had two primary purposes. First, an acceptable method for forecasting AADT volumes was needed to assign more accurate annual growth rates to forecast AADTs on highways in rural Idaho. Second, a greater understanding of the impacts that AADT forecasting errors have in design applications was needed. The usefulness of several existing forecasting methods for this project was examined. It was concluded that a method was required that could classify the ATR stations while reducing the variability of the annual growth rates. The CART algorithm had been used in other transportation applications, but not for forecasting AADT specifically. This method worked well to classify the ATR stations into groups with similar characteristics while reducing the variability of the AADT annual growth rates. The method was validated using a stratified sample of portable count locations. This validation resulted in a mean absolute percent error of the entire validation data set that was 14.2 percent and nearly half of the portable count locations had percent errors of less than 10 percent. High percent errors at low traffic volume locations may have deceptive connotations. Instead, in these low-volume instances, the magnitude difference is a better measure of accuracy.

It must be noted that a ten-year forecast was used in this research and a 20-year forecast would usually be required for practical purposes. The restraints of the available data for this project did not allow for a 20-year forecast, however. Higher errors should be expected when 20-year forecasts are performed; although the magnitude of these errors is impossible to discern without more years of data. Research by Saha and Fricker forecasted up to twelve years with an MAPE of 15.8 percent [20]. Research by Memmott forecasted AADTs up to 20 years and had an MAPE of 28.7 percent [15]. The research shown in this report had errors similar to those accepted by Saha and Fricker and lower than those found acceptable by Memmott. Therefore, the CART method produces errors that should also be considered acceptable. Professionals in Idaho recommended that AADT 20-year forecasting errors be within 20 percent. Because nearly half of the data points in the validation set had errors less than 10 percent for the 10-year forecast, it is probable that most locations in practice would have 20-year forecasting errors less than 20 percent.

In the previous chapter, it was found that the design errors using AADT values forecasted using the CART method were acceptable. In the overlay thickness section, it was found that on low volume roads, which include many of the rural Idaho highways, the minimum overlay thickness was always required. In these cases, the forecasted value of the AADT had no effect on the design. For those highways with AADT volumes high enough to require more than the minimum overlay thickness, the difference between the thickness required from the forecasted AADT value and the actual thickness required was minimal in most cases. In the LOS analysis section, it was noted that a design error would only occur if the analysis provided a value close to the critical boundary. The C-D boundary was used as the critical boundary in this research. Of those close to this boundary, most locations in the validation data set matched LOS designation between the forecasted and actual AADT values. This means that most locations chosen in practice would also provide the correct design decision when the CART method is used for AADT forecasting.

The CART method created subsets of data that had similar characteristics and a low variability of annual AADT growth rates. If more data were available and the terminal nodes in the regression tree included a greater number of data points, the CART method could be used as a first step in the forecasting process. The following step would be compromised of other forecasting methods such as ARIMA time series or elasticity-based regression models, calibrated to each of the resulting CART subgroups. This was not performed in this project because the resulting terminal nodes contained no more than 16 ATR stations, which may be insufficient to create individual models by subgroup. A recommendation for future research is to investigate the use of the other forecasting techniques once the CART method classified the stations into similar groups.

The CART method was found to provide AADT forecasts with acceptable results using the rural Idaho data. It has accuracies similar to that of the existing ITD method and is much simpler to implement and update. The two design applications, overlay design and LOS analysis, were analyzed and it was found that the AADT forecasts rarely resulted in large design errors. Therefore, it is recommended that the CART method be implemented for forecasting AADT values.

References

1. *AASHTO Guidelines for Traffic Data Programs*. Joint Task Force on Traffic Monitoring Standards of the AASHTO Highway Subcommittee on Traffic Engineering. 1992: American Association of State Highway and Transportation Officials. 113.

2. Horowitz, A., *Guidebook on Statewide Travel Forecasting July 1999*. 1999, Federal Highway Administration. p. 139.

3. Wright, R. March 13, 2002. M. Hanenburg. email.

4. Richardson, C. March 12, 2002. M. Hanenburg. email.

5. Bushnell, D. May 14, 2002. M. Hanenburg. email.

6. Schuytema, P.L. May 24, 2002. M. Hanenburg. email.

7. Bisom, D. May 17, 2002. M. Hanenburg. email.

8. Parry, R. May 16, 2002. M. Hanenburg. email.

9. Robles, J. May 17, 2002. M. Hanenburg. email.

10. Manning, D. May 30, 2002. M. Hanenburg. email.

11. Martinez, J. May 21, 2002. M. Hanenburg. email.

12. Fogel, A.-M. May 31, 2002. M. Hanenburg. email.

13. Cipra, D. May 29, 2002. M. Hanenburg. email.

14. Horowitz, A. and D. Farmer, *Statewide Travel Forecasts: A Critical Review.* Transportation Research Record, 1999(1685): p. 13-20.

15. Memmott, J., *Factors that Affect Traffic Growth Rates and Projection of Traffic Volumes for Use in Highway Economic Models.* Transportation Research Record, 1983(912): p. 11-15.

16. Webpage. Highway Performance Monitoring System Field Book. http://www.fhwa.dot.gov/ohim/hpmsmanl/hpms.htm. accessed June 20, 2002.

17. Garber, N.J. and F. Bayat-Mokhari, *A Computerized Highway Link Classification System for Traffic Volume Counts.* Transportation Research Record, 1986(1090): p. 1-7.

18. *Traffic Monitoring Guide*. May 2001: FHWA, U.S. Department of Transportation.

19. Faghri, A. and P. Chakroborty, *Development and Evaluation of a Statistically Reliable Traffic Counting Program.* Transportation Planning and Technology, 1994. **18**: p. 223-237.

20. Saha, S. and J. Fricker, *Traffic Volumes Forecasting Methods for Rural State Highways.* Transportation Research Record, 1987(1203): p. 10-26.

21. Mohamad, D., K. Sinha, T. Kuczek and C. Scholer, *Annual Average Daily Traffic Prediction Models for County Roads.* Transportation Research Record, 1998(1617): p. 69-77.

22. Lam, W. and J. Xu, *Estimation of AADT from Short Period Counts in Hong-Kong -- A Comparison Between Neural Network Method and Regression Analysis.* Journal of Advanced Transportation, 2000. **34**(2): p. 249-268.

23. Clark, S., H. Chen and S. Grant-Muller, *Artificial Neural Network and Statistical Modeling of Traffic Flows -- The Best of Both Worlds.* World Conference on Transportation Research, 1999.

24. McShane, W.R., R.P. Roess and E.S. Prassas, *Traffic Engineering*. Second ed, ed. B. Stenquist. 1998, Upper River, NJ: Prentice-Hall. 714.

25. Webpage. SAS OnlineDoc. http://v8doc.sas.com/sashtml/. accessed July 26, 2002.

26. *S-Plus 2000 User's Guide*. 1999, Seattle, Washington: Data Analysis Products Division of MathSoft, Inc.

27. Wolf, J., R. Guensler, S. Washington and W. Bachman, *High-Emitting Vehicle Characterization Using Regression Tree Analysis.* Transportation Research Record. **1641**.

28. Smith, B.L., W. Scherer and T. Hauser, *Data-Mining Tools for the Support of Signal-Timing Plan Development.* Transportation Research Record, 2001. **1768**.

29. Saraf, C., *Review of PCR Methodology*. 1998, Federal Highway Administration.

30. *TRANSIMS: Transportation Analysis Simulation System Version: TRANSIMS 3.0*. 2002, Los Alamos National Laboratory: Los Alamos.

31. Huang, Y.H., *Pavement Analysis and Design*. 1993, Englewood Cliffs, New Jersey: Prentice Hall. 805.

32. *WINFLEX 2000*, National Institute for Advanced Transportation Technology, Idaho Transportation Department.

33. *HCS2000*, Transportation Research Board: Washington, D.C.

34. *Highway Capacity Manual*. 2000, Washington, D.C.: Transportation Research Board.

Appendix A – AADT Database

ITD provided their ATR database for rural highways, which includes location, AADT, county, and functional class for all available years. Years 1980, 1990, and 2000 are shown in Table A.1.

TABLE A.1 ATR AADT Database

ATR	Name	Route	Segment	Milepost	County	Year	Type	Funct	AADT
3	Twin Falls	US-30	2040	220.95	Twin Falls	1980	R	7	6943
3	Twin Falls	US-30	2040	220.95	Twin Falls	1990	R	7	7733
3	Twin Falls	US-30	2040	220.95	Twin Falls	2000	R	6	9090
4	S Pocatello	I-15	1330	61.87	Bannock	1980	R	1	8239
4	S Pocatello	I-15	1330	61.87	Bannock	1990	R	1	10303
4	S Pocatello	I-15	1330	61.87	Bannock	2000	R	1	15150
6	Lewiston	US-95	1540	305.1	Nez Perce	1980	R	2	6947
6	Lewiston	US-95	1540	305.1	Nez Perce	1990	R	2	8151
6	Lewiston	US-95	1540	305.1	Nez Perce	2000	R	2	10342
7	Jerome	I-84	1010	159.23	Gooding	1980	R	1	7477
7	Jerome	I-84	1010	159.23	Gooding	1990	R	1	11022
7	Jerome	I-84	1010	159.23	Gooding	2000	R	1	16821
8	Dudley	I-90	1660	35.59	Kootenai	1980	R	1	6316
8	Dudley	I-90	1660	35.59	Kootenai	1990	R	1	8134
8	Dudley	I-90	1660	35.59	Kootenai	2000	R	1	11050
9	Caldwell	SH-19	2050	15.42	Canyon	1980	R	7	4169
9	Caldwell	SH-19	2050	15.42	Canyon	1990	R	7	5311
9	Caldwell	SH-19	2050	15.42	Canyon	2000	R	6	7672
11	Paris	US-89	2380	13.946	Bear Lake	1980	R	6	1464
11	Paris	US-89	2380	13.946	Bear Lake	1990	R	6	1426
11	Paris	US-89	2380	13.946	Bear Lake	2000	R	2	1700
12	Ririe	US-26	2240	352.82	Bonneville	1980	R	2	1840
12	Ririe	US-26	2240	352.82	Bonneville	1990	R	2	2292
12	Ririe	US-26	2240	352.82	Bonneville	2000	R	2	3287
13	Salmon	US-93	2220	301.57	Lemhi	1980	R	2	1790
13	Salmon	US-93	2220	301.57	Lemhi	1990	R	2	2156
13	Salmon	US-93	2220	301.57	Lemhi	2000	R	2	2742
14	Shoshone	SH-75	2230	79.67	Lincoln	1980	R	6	1841
14	Shoshone	SH-75	2230	79.67	Lincoln	1990	R	6	2376
14	Shoshone	SH-75	2230	79.67	Lincoln	2000	R	6	3450
15	Potlatch	US-95	1540	363.89	Latah	1980	R	2	1456
15	Potlatch	US-95	1540	363.89	Latah	1990	R	2	1981
15	Potlatch	US-95	1540	363.89	Latah	2000	R	2	2547
17	Arco	US-20	2240	252.38	Butte	1980	R	2	1540
17	Arco	US-20	2240	252.38	Butte	1990	R	2	1870
17	Arco	US-20	2240	252.38	Butte	2000	R	2	2028
18	Raft River	I-86	1260	14.41	Cassia	1980	R	1	3387
18	Raft River	I-86	1260	14.41	Cassia	1990	R	1	4501
18	Raft River	I-86	1260	14.41	Cassia	2000	R	1	6193
19	Kamiah	US-12	1910	63.663	Lewis	1980	R	2	1614
19	Kamiah	US-12	1910	63.663	Lewis	1990	R	2	1824
19	Kamiah	US-12	1910	63.663	Lewis	2000	R	2	2121

ATR	Name	Route	Segment	Milepost	County	Year	Type	Funct	AADT
21	Chilco	US-95	1540	442.74	Kootenai	1980	R	2	4667
21	Chilco	US-95	1540	442.74	Kootenai	1990	R	2	8045
21	Chilco	US-95	1540	442.74	Kootenai	2000	R	2	12329
22	Malad	I-15	1330	1.965	Oneida	1980	R	1	3439
22	Malad	I-15	1330	1.965	Oneida	1990	R	1	5037
22	Malad	I-15	1330	1.965	Oneida	2000	R	1	8223
23	Council	US-95	1540	140.38	Adams	1980	R	2	1045
23	Council	US-95	1540	140.38	Adams	1990	R	2	1272
23	Council	US-95	1540	140.38	Adams	2000	R	2	1464
25	Sand Hollow	I-84	1010	19.1	Canyon	1980	R	1	7509
25	Sand Hollow	I-84	1010	19.1	Canyon	1990	R	1	11362
25	Sand Hollow	I-84	1010	19.1	Canyon	2000	R	1	17069
26	Kootenai	SH-200	1610	35.98	Bonner	1980	R	6	1717
26	Kootenai	SH-200	1610	35.98	Bonner	1990	R	6	2679
26	Kootenai	SH-200	1610	35.98	Bonner	2000	R	6	3633
27	St Maries	SH-3	1800	95.34	Kootenai	1980	R	6	906
27	St Maries	SH-3	1800	95.34	Kootenai	1990	R	6	1080
27	St Maries	SH-3	1800	95.34	Kootenai	2000	R	6	1503
28	Ketchum	SH-75	2230	135.95	Blaine	1980	R	6	952
28	Ketchum	SH-75	2230	135.95	Blaine	1990	R	6	1108
28	Ketchum	SH-75	2230	135.95	Blaine	2000	R	6	1278
29	Rogerson	US-93	2220	16.724	Twin Falls	1980	R	2	2097
29	Rogerson	US-93	2220	16.724	Twin Falls	1990	R	2	3211
29	Rogerson	US-93	2220	16.724	Twin Falls	2000	R	2	3803
30	Cotteral	I-84	1010	231	Cassia	1980	R	1	2836
30	Cotteral	I-84	1010	231	Cassia	1990	R	1	4279
30	Cotteral	I-84	1010	231	Cassia	2000	R	1	6077
31	Swan Valley	SH-31	2450	3.54	Bonneville	1980	R	7	648
31	Swan Valley	SH-31	2450	3.54	Bonneville	1990	R	7	912
31	Swan Valley	SH-31	2450	3.54	Bonneville	2000	R	7	1530
32	Ashton	US-20	2070	377.08	Fremont	1980	R	2	1752
32	Ashton	US-20	2070	377.08	Fremont	1990	R	2	2307
32	Ashton	US-20	2070	377.08	Fremont	2000	R	2	3069
34	Geneva	US-89	2380	38.51	Bear Lake	1980	R	6	585
34	Geneva	US-89	2380	38.51	Bear Lake	1990	R	6	613
34	Geneva	US-89	2380	38.51	Bear Lake	2000	R	2	646
35	Banida	US-91	2350	19.89	Franklin	1980	R	6	819
35	Banida	US-91	2350	19.89	Franklin	1990	R	6	1004
35	Banida	US-91	2350	19.89	Franklin	2000	R	6	1261
36	Border	US-30	2040	446.5	Bear Lake	1980	R	2	1026
36	Border	US-30	2040	446.5	Bear Lake	1990	R	2	1246
36	Border	US-30	2040	446.5	Bear Lake	2000	R	2	1507
38	Marsing	US-95	1540	22.72	Owyhee	1980	R	2	1151
38	Marsing	US-95	1540	22.72	Owyhee	1990	R	2	1364
38	Marsing	US-95	1540	22.72	Owyhee	2000	R	2	1400
39	Fenn	US-95	1540	247.03	Idaho	1980	R	2	1889
39	Fenn	US-95	1540	247.03	Idaho	1990	R	2	2194
39	Fenn	US-95	1540	247.03	Idaho	2000	R	2	2714

ATR	Name	Route	Segment	Milepost	County	Year	Type	Funct	AADT
40	Rathdrum	SH-53	1650	6.64	Kootenai	1980	R	6	2805
40	Rathdrum	SH-53	1650	6.64	Kootenai	1990	R	6	3896
40	Rathdrum	SH-53	1650	6.64	Kootenai	2000	R	2	6386
41	N Rathdrum	SH-41	1630	8.96	Kootenai	1980	R	7	3355
41	N Rathdrum	SH-41	1630	8.96	Kootenai	1990	R	7	5261
41	N Rathdrum	SH-41	1630	8.96	Kootenai	2000	R	6	8034
42	Athol	SH-54	1640	8.36	Kootenai	1980	R	7	854
42	Athol	SH-54	1640	8.36	Kootenai	1990	R	7	1682
42	Athol	SH-54	1640	8.36	Kootenai	2000	R	7	2203
43	Donnelly	SH-55	1990	127.72	Valley	1980	R	2	1844
43	Donnelly	SH-55	1990	127.72	Valley	1990	R	2	2484
43	Donnelly	SH-55	1990	127.72	Valley	2000	R	2	3079
44	Weiser	US-95	1540	77.96	Washington	1980	R	2	2800
44	Weiser	US-95	1540	77.96	Washington	1990	R	2	3616
44	Weiser	US-95	1540	77.96	Washington	2000	R	2	5294
45	Bovill	SH-3	1800	39.89	Latah	1980	R	6	540
45	Bovill	SH-3	1800	39.89	Latah	1990	R	6	568
45	Bovill	SH-3	1800	39.89	Latah	2000	R	6	503
46	Copeland	US-95	1540	527.28	Boundary	1980	R	2	573
46	Copeland	US-95	1540	527.28	Boundary	1990	R	2	925
46	Copeland	US-95	1540	527.28	Boundary	2000	R	2	999
47	Priest River	US-2	1590	2.64	Bonner	1980	R	2	4268
47	Priest River	US-2	1590	2.64	Bonner	1990	R	2	6235
47	Priest River	US-2	1590	2.64	Bonner	2000	R	2	7201
49	Riggins	US-95	1540	203.7	Idaho	1980	R	2	1214
49	Riggins	US-95	1540	203.7	Idaho	1990	R	2	1614
49	Riggins	US-95	1540	203.7	Idaho	2000	R	2	1783
50	Craters	US-93	2240	229.51	Butte	1980	R	2	836
50	Craters	US-93	2240	229.51	Butte	1990	R	2	1018
50	Craters	US-93	2240	229.51	Butte	2000	R	2	1113
51	Lorenzo	US-20	2070	325.74	Jefferson	1980	R	2	6735
51	Lorenzo	US-20	2070	325.74	Jefferson	1990	R	2	9775
51	Lorenzo	US-20	2070	325.74	Jefferson	2000	R	2	14535
53	Robie Creek	SH-21	2140	20.89	Boise	1980	R	6	1484
53	Robie Creek	SH-21	2140	20.89	Boise	1990	R	7	2238
53	Robie Creek	SH-21	2140	20.89	Boise	2000	R	6	3106
54	Mountain Home	US-20	2070	102.02	Elmore	1980	R	2	1263
54	Mountain Home	US-20	2070	102.02	Elmore	1990	R	2	1462
54	Mountain Home	US-20	2070	102.02	Elmore	2000	R	2	1898.8
55	Dickey	US-93	2220	129.08	Custer	1980	R	2	333
55	Dickey	US-93	2220	129.08	Custer	1990	R	2	445
55	Dickey	US-93	2220	129.08	Custer	2000	R	2	525
56	Howe	SH-33	2460	21.94	Butte	1980	R	7	344
56	Howe	SH-33	2460	21.94	Butte	1990	R	7	432
56	Howe	SH-33	2460	21.94	Butte	2000	R	6	540
58	Leadore	SH-28	2500	89.96	Lemhi	1980	R	6	446
58	Leadore	SH-28	2500	89.96	Lemhi	1990	R	6	434
58	Leadore	SH-28	2500	89.96	Lemhi	2000	R	6	540

ATR	Name	Route	Segment	Milepost	County	Year	Type	Funct	AADT
59	Newdale	SH-33	2460	112.05	Madison	1980	R	7	898
59	Newdale	SH-33	2460	112.05	Madison	1990	R	7	1324
59	Newdale	SH-33	2460	112.05	Madison	2000	R	6	1761
60	Alexander	US-30	2040	399.2	Caribou	1980	R	2	3418
60	Alexander	US-30	2040	399.2	Caribou	1990	R	2	3772
60	Alexander	US-30	2040	399.2	Caribou	2000	R	2	4889
61	Roberts	I-15	1330	132.78	Jefferson	1980	R	1	2422
61	Roberts	I-15	1330	132.78	Jefferson	1990	R	1	3419
61	Roberts	I-15	1330	132.78	Jefferson	2000	R	1	4505.5
67	Pocatello Air	I-86	1260	56.4	Power	1980	R	1	7027
67	Pocatello Air	I-86	1260	56.4	Power	1990	R	1	9026
67	Pocatello Air	I-86	1260	56.4	Power	2000	R	1	12154
68	Hailey	SH-75	2230	119.4	Blaine	1980	R	6	5106
68	Hailey	SH-75	2230	119.4	Blaine	1990	R	6	8931
68	Hailey	SH-75	2230	119.4	Blaine	2000	R	6	12687
71	Hammett	I-84	1010	114.5	Elmore	1980	R	1	5704
71	Hammett	I-84	1010	114.5	Elmore	1990	R	1	8501
71	Hammett	I-84	1010	114.5	Elmore	2000	R	1	12684

Appendix B ––Economic and Demographic Data by County

Economic and demographic data were required to perform different forecasting methods. This data are shown in Table B.1 by county.

TABLE B.1 Economic and Demographic Data

County	Year	Population (thousands)	Population Density (people per square mile)	Employment (thousands)	Income (per capita)
Adams	1980	3.3	2.4	1.4	16326
Adams	1990	3.3	2.4	1.6	12837
Adams	2000	3.5	2.5	2.0	13955
Bannock	1980	65.7	59.0	31.0	15789
Bannock	1990	66.2	59.5	31.1	16106
Bannock	2000	75.7	68.0	41.5	20090
Bear Lake	1980	7.0	7.2	2.4	13878
Bear Lake	1990	6.1	6.3	2.2	16542
Bear Lake	2000	6.4	6.6	3.0	19719
Blaine	1980	9.9	3.8	7.0	21224
Blaine	1990	13.8	5.2	12.1	12726
Blaine	2000	19.1	7.2	17.2	15641
Boise	1980	3.0	1.6	1.3	15508
Boise	1990	3.6	1.9	1.4	29202
Boise	2000	6.7	3.5	2.2	36906
Bonner	1980	24.3	14.0	9.6	13598
Bonner	1990	26.8	15.4	12.9	17068
Bonner	2000	37.0	21.3	18.8	16533
Bonneville	1980	66.3	35.5	31.8	16664
Bonneville	1990	72.6	38.9	39.5	15619
Bonneville	2000	82.7	44.3	52.7	18036
Boundary	1980	7.3	5.8	3.0	12527
Boundary	1990	8.4	6.6	3.8	20128
Boundary	2000	9.9	7.8	5.2	21859
Butte	1980	3.4	1.5	6.7	14196
Butte	1990	2.9	1.3	8.1	13379
Butte	2000	2.9	1.3	6.8	16777
Canyon	1980	84.0	142.4	37.9	13556
Canyon	1990	90.6	153.7	43.8	15231
Canyon	2000	132.4	224.5	63.3	19217
Caribou	1980	8.7	4.9	4.9	18517
Caribou	1990	7.0	3.9	4.0	16044
Caribou	2000	7.3	4.1	4.6	16857
Cassia	1980	19.5	7.6	10.2	14586
Cassia	1990	19.6	7.6	10.4	16799
Cassia	2000	21.4	8.3	12.7	19899
Custer	1980	3.5	0.7	1.7	15146
Custer	1990	4.2	0.8	2.6	19304
Custer	2000	4.3	0.9	2.6	20252
Elmore	1980	21.7	7.0	10.7	14912
Elmore	1990	21.2	6.9	10.9	17642
Elmore	2000	29.3	9.5	13.9	22119

County	Year	Population (thousands)	Population Density (people per square mile)	Employment (thousands)	Income (per capita)
Franklin	1980	9.0	13.5	3.5	11075
Franklin	1990	9.3	13.9	3.5	20320
Franklin	2000	11.3	17.0	4.7	18429
Fremont	1980	10.9	5.8	4.4	13525
Fremont	1990	10.9	5.9	4.3	12944
Fremont	2000	11.8	6.3	4.9	13922
Gooding	1980	11.9	16.3	5.4	14125
Gooding	1990	11.7	16.0	5.7	15117
Gooding	2000	14.2	19.4	7.8	15689
Idaho	1980	14.8	1.7	6.1	14623
Idaho	1990	13.8	1.6	6.7	18408
Idaho	2000	15.5	1.8	7.9	20198
Jefferson	1980	15.4	14.1	5.7	12393
Jefferson	1990	16.6	15.2	6.1	16536
Jefferson	2000	19.2	17.5	8.1	17282
Kootenai	1980	60.0	48.2	23.6	15800
Kootenai	1990	70.4	56.5	35.7	14887
Kootenai	2000	109.5	88.0	60.6	17049
Latah	1980	28.8	26.8	14.0	15143
Latah	1990	30.7	28.5	16.8	18700
Latah	2000	35.0	32.5	20.2	21678
Lemhi	1980	7.5	1.7	3.5	14572
Lemhi	1990	6.9	1.5	3.6	17209
Lemhi	2000	7.8	1.7	4.4	19662
Lewis	1980	4.1	8.6	1.8	14572
Lewis	1990	3.5	7.3	1.8	15368
Lewis	2000	3.7	7.8	2.1	19081
Lincoln	1980	3.5	2.9	1.9	15327
Lincoln	1990	3.3	2.8	1.8	19847
Lincoln	2000	4.0	3.4	2.0	20646
Madison	1980	19.7	41.7	8.4	14942
Madison	1990	23.8	50.4	10.7	17274
Madison	2000	27.5	58.4	14.8	17553
Nez Perce	1980	33.2	39.1	19.9	11993
Nez Perce	1990	33.8	39.9	21.6	11158
Nez Perce	2000	37.5	44.1	27.0	13156
Oneida	1980	3.3	2.7	1.4	17482
Oneida	1990	3.5	2.9	1.4	19930
Oneida	2000	4.1	3.4	1.9	23779
Owyhee	1980	8.4	1.1	3.5	13495
Owyhee	1990	8.4	1.1	3.2	13703
Owyhee	2000	10.6	1.4	3.9	15350
Power	1980	6.9	4.9	4.5	12518
Power	1990	7.1	5.0	4.6	14828
Power	2000	7.5	5.4	5.4	15313
Twin Falls	1980	53.1	27.6	28.0	17676
Twin Falls	1990	53.8	27.9	31.6	21066
Twin Falls	2000	64.5	33.5	41.1	20740
Valley	1980	5.7	1.5	3.1	16153
Valley	1990	6.1	1.7	4.1	17930
Valley	2000	7.6	2.1	5.7	20225
Washington	1980	8.8	6.1	3.8	18126
Washington	1990	8.6	5.9	3.9	20767
Washington	2000	10.0	6.8	5.1	24830

Appendix C ––S-Plus Code

S-Plus code was required for the classification and regression tree implementation. The code and the output are shown.

S-Plus CART Code:

```
tree(formula = AADT.GF ~ Funct + AADT90 + PopTh90 + Pop.GF, data = DS5,
        na.action = na.exclude, mincut = 5, minsize = 10, mindev = 0.005)
```

Output from each Verification Data Set:

Sample l:

```
Variables actually used in tree construction:
[1] "Pop.GF" "Funct" "AADT90"
Number of terminal nodes: 7
Residual mean deviance: 0.0001251 = 0.004378 / 35
Distribution of residuals:
   Min.  1st Qu.  Median    Mean 3rd Qu.  Max.
 -0.02625 -0.007252 0.0009352 5.782e-019 0.005674 0.02616
node), split, n, deviance, yval
   * denotes terminal node

 1) root 42 0.00981300 0.02917
  2) Pop.GF<0.00687839 25 0.00333500 0.02148
   4) Funct:2,6,7 20 0.00200300 0.01834
    8) AADT90<1125 5 0.00049620 0.01013 *
    9) AADT90>1125 15 0.00105800 0.02107
    18) Pop.GF<-0.00512501 5 0.00007003 0.01660 *
    19) Pop.GF>-0.00512501 10 0.00083750 0.02331
     38) AADT90<2341.5 5 0.00020170 0.02218 *
     39) AADT90>2341.5 5 0.00062290 0.02444 *
   5) Funct:1 5 0.00034440 0.03405 *
  3) Pop.GF>0.00687839 17 0.00282800 0.04047
   6) Funct:1,2 9 0.00094480 0.03736 *
   7) Funct:6,7 8 0.00169800 0.04397 *
```

Sample 2:

```
Variables actually used in tree construction:
[1] "Pop.GF" "Funct" "AADT90"
Number of terminal nodes: 6
Residual mean deviance: 0.00009463 = 0.003407 / 36
Distribution of residuals:
   Min.  1st Qu.   Median    Mean 3rd Qu.  Max.
 -0.01593 -0.007365 -0.0006057 3.304e-019 0.006372 0.01812
node), split, n, deviance, yval
   * denotes terminal node

 1) root 42 0.0073620 0.02715
  2) Pop.GF<0.00663673 24 0.0028430 0.02108
   4) Funct:2,6,7 18 0.0013040 0.01704
    8) Pop.GF<-0.0110589 8 0.0005583 0.01330 *
    9) Pop.GF>-0.0110589 10 0.0005443 0.02003
    18) AADT90<1538 5 0.0001567 0.01547 *
    19) AADT90>1538 5 0.0001797 0.02459 *
   5) Funct:1 6 0.0003670 0.03318 *
  3) Pop.GF>0.00663673 18 0.0024510 0.03526
   6) AADT90<2581.5 9 0.0010370 0.03114 *
   7) AADT90>2581.5 9 0.0011080 0.03938 *
```

Sample 3:

```
Variables actually used in tree construction:
[1] "Pop.GF" "Funct" "AADT90"
Number of terminal nodes: 6
Residual mean deviance: 0.0001021 = 0.003677 / 36
Distribution of residuals:
   Min.  1st Qu.  Median    Mean 3rd Qu.  Max.
 -0.01584 -0.006093 0.0001657 -3.304e-019 0.006926 0.02602
node), split, n, deviance, yval
   * denotes terminal node

 1) root 42 0.0092650 0.02835
  2) Pop.GF<-0.00771277 9 0.0007691 0.01140 *
  3) Pop.GF>-0.00771277 33 0.0052030 0.03298
   6) Pop.GF<0.00663673 17 0.0016110 0.02684
   12) Funct:2,6,7 12 0.0008754 0.02355
    24) AADT90<1797.5 5 0.0001150 0.02024 *
    25) AADT90>1797.5 7 0.0006664 0.02591 *
   13) Funct:1 5 0.0002935 0.03474 *
   7) Pop.GF>0.00663673 16 0.0022710 0.03950
   14) Funct:1,6 7 0.0004911 0.03356 *
   15) Funct:2,7 9 0.0013420 0.04411 *
```

Sample 4:

```
Variables actually used in tree construction:
[1] "Pop.GF" "AADT90"
Number of terminal nodes: 6
Residual mean deviance: 0.0001202 = 0.004327 / 36
Distribution of residuals:
   Min.  1st Qu.   Median    Mean 3rd Qu.  Max.
 -0.01864 -0.006888 7.506e-006 1.652e-019 0.006674 0.02587
node), split, n, deviance, yval
   * denotes terminal node

 1) root 42 0.0089670 0.02823
  2) Pop.GF<0.00663673 22 0.0029190 0.02055
   4) AADT90<2250.5 13 0.0008904 0.01535
    8) AADT90<1538 8 0.0006280 0.01312 *
    9) AADT90>1538 5 0.0001583 0.01893 *
   5) AADT90>2250.5 9 0.0011710 0.02805 *
  3) Pop.GF>0.00663673 20 0.0033250 0.03667
   6) AADT90<1216 5 0.0007505 0.02925 *
   7) AADT90>1216 15 0.0022070 0.03915
   14) Pop.GF<0.00938642 6 0.0002460 0.03148 *
   15) Pop.GF>0.00938642 9 0.0013740 0.04426 *
```

Sample 5:

```
Variables actually used in tree construction:
[1] "Pop.GF" "AADT90" "Funct"
Number of terminal nodes: 7
Residual mean deviance: 0.000118 = 0.004131 / 35
Distribution of residuals:
   Min.  1st Qu.   Median    Mean 3rd Qu.  Max.
 -0.01657 -0.007358 -0.0008131 4.956e-019 0.006398 0.02368
node), split, n, deviance, yval
   * denotes terminal node

 1) root 42 0.0099160 0.02770
  2) Pop.GF<-0.00479597 12 0.0010870 0.01387
   4) AADT90<1520 6 0.0007093 0.01031 *
   5) AADT90>1520 6 0.0002263 0.01743 *
  3) Pop.GF>-0.00479597 30 0.0056150 0.03323
   6) Pop.GF<0.00663673 13 0.0018380 0.02668
   12) Funct:2,6,7 8 0.0010160 0.02164 *
   13) Funct:1 5 0.0002935 0.03474 *
   7) Pop.GF>0.00663673 17 0.0027930 0.03824
   14) AADT90<1216 5 0.0007505 0.02925 *
   15) AADT90>1216 12 0.0014700 0.04199
    30) Pop.GF<0.00861409 5 0.0001844 0.03574 *
    31) Pop.GF>0.00861409 7 0.0009511 0.04645 *
```

Sample 6:
Number of terminal nodes: 7
Residual mean deviance: 0.0001018 = 0.003564 / 35
Distribution of residuals:
Min. 1st Qu. Median Mean 3rd Qu. Max.
-0.01585 -0.008212 0.0007899 -1.652e-019 0.006663 0.01914
node), split, n, deviance, yval
* denotes terminal node

1) root 42 0.00762800 0.02882
2) Pop.GF<0.00663673 22 0.00311300 0.02233
4) AADT90<2088.5 11 0.00089840 0.01554
8) PopTh90<4.8035 5 0.00006299 0.01894 *
9) PopTh90>4.8035 6 0.00072930 0.01270 *
5) AADT90>2088.5 11 0.00119900 0.02912
10) Pop.GF<0.000543431 5 0.00026880 0.03480 *
11) Pop.GF>0.000543431 6 0.00063500 0.02439 *
3) Pop.GF>0.00663673 20 0.00256800 0.03596
6) AADT90<2581.5 9 0.00103700 0.03114 *
7) AADT90>2581.5 11 0.00114900 0.03991
14) Funct:1,7 5 0.00032990 0.03402 *
15) Funct:2,6 6 0.00050140 0.04482 *

Sample 7:
Variables actually used in tree construction:
[1] "Pop.GF" "Funct" "AADT90"
Number of terminal nodes: 7
Residual mean deviance: 0.0001134 = 0.003968 / 35
Distribution of residuals:
Min. 1st Qu. Median Mean 3rd Qu. Max.
-0.02309 -0.007109 -3.689e-006 0 0.005591 0.02253
node), split, n, deviance, yval
* denotes terminal node

1) root 42 0.0097070 0.02713
2) Pop.GF<0.00663673 23 0.0028210 0.01877
4) Funct:6 6 0.0007225 0.00847 *
5) Funct:1,2,7 17 0.0012370 0.02241
10) AADT90<7942 12 0.0004722 0.01971
20) Pop.GF<-0.00479597 7 0.0001088 0.01799 *
21) Pop.GF>-0.00479597 5 0.0003133 0.02213 *
11) AADT90>7942 5 0.0004690 0.02887 *
3) Pop.GF>0.00663673 19 0.0033360 0.03724
6) AADT90<1503 5 0.0008223 0.03021 *
7) AADT90>1503 14 0.0021790 0.03975
14) Funct:1,2 8 0.0003542 0.03386 *
15) Funct:6,7 6 0.0011780 0.04760 *

The cluster analysis was implemented using S-Plus code. The code and the output are shown.

```
      *** Agglomerative Hierarchical Clustering ***
Call:
agnes(x = menuModelFrame(data = SDF13, variables = "AADT90,pop90,PopGR",
      subset = NULL, na.rm = T), diss = F, metric = "euclidean", stand = T,
      method = "average", save.x = T, save.diss = T)
Merge:
   [,1] [,2]
 [1,]  -5 -15
 [2,]  -7 -28
 [3,] -13 -23
 [4,] -12 -14
 [5,] -26 -45
 [6,]  2 -40
 [7,] -30 -39
 [8,] -20 -33
 [9,]  6  5
[10,] -27 -29
[11,]  -9  7
[12,]  9  4
[13,] -10 -17
[14,] -25  10
[15,] -11 -19
[16,] -31 -32
[17,]  -8 -24
[18,] -34 -49
[19,] -37 -44
[20,]  13  14
[21,]  3 -35
[22,]  11 -46
[23,]  19 -42
[24,] -41 -50
[25,]  -3 -52
[26,]  15 -36
[27,]  20 -43
[28,]  17  8
[29,]  12  22
[30,] -16  18
[31,]  23 -47
[32,]  -4  24
[33,]  27  21
```

```
[34,]  -1  -2
[35,]  25  32
[36,]  33  30
[37,]  28  16
[38,]  36  26
[39,]  29 -48
[40,]  35 -38
[41,]  -6  37
[42,]  41 -22
[43,] -21  31
[44,]  1 -18
[45,]  34  44
[46,]  39  38
[47,]  45  40
[48,]  46  43
[49,]  47  42
[50,]  49  48
[51,]  50 -51
Order of objects:
 [1] 1 2 5 15 18 3 52 4 41 50 38 6 8 24 20 33 31 32 22 7 28 40 26 45 12
[26] 14 9 30 39 46 48 10 17 25 27 29 43 13 23 35 16 34 49 11 19 36 21 37 44 42
[51] 47 51
Height:
 [1] 1.12309968 1.78761713 0.03380825 1.77698106 2.59419163 0.71395153
 [7] 1.16526691 1.07076311 0.71164390 1.59417309 3.08786946 1.64010527
[13] 0.52421785 0.79147125 0.22868054 1.28975077 0.51851983 1.71539966
[19] 3.45445723 0.06837624 0.19790969 0.29057116 0.16902693 0.39015625
[25] 0.15118978 0.82222018 0.38256820 0.22032344 0.66086095 1.48874851
[31] 1.85736291 0.44126165 0.61410470 0.47574606 0.30888267 0.75221004
[37] 1.08948611 0.08433070 0.65874168 1.22203318 0.91713196 0.59082044
[43] 1.32611431 0.48653484 0.71822632 2.72392880 1.77436542 0.60755315
[49] 0.67625240 0.95145495 4.24689916
Agglomerative coefficient:
[1] 0.8385112

Available arguments:
[1] "order"   "height"  "ac"    "merge"   "order.lab" "diss"
[7] "data"   "call"
```

Appendix D ––ATR Stations by Terminal Node

TN #	ATR
1	27
1	28
1	31
1	46
1	55
2/3	11
2/3	23
2/3	34
2/3	35
2/3	36
2/3	38
2/3	45
2/3	50
2/3	54
2/3	56
2/3	58
4	13
4	17
4	18
4	19
4	39
4	49
4	60

TN #	ATR
5	9
5	12
5	22
5	25
5	43
5	51
5	61
6/7	3
6/7	4
6/7	6
6/7	7
6/7	14
6/7	15
6/7	29
6/7	30
6/7	32
6/7	44
6/7	67
6/7	71
8	8
8	21
8	26
8	40
8	41
8	42
8	47
8	53
8	59
8	68