Report No. CDOT-DTD-R-2004-9
Final Report

USING RTD's TRANSIT VEHICLES TO DEVELOP FREEWAY SPEED MAPS FOR COLORADO

Sarosh Khan
Reema Kundu

June 2004

COLORADO DEPARTMENT OF TRANSPORTATION
RESEARCH BRANCH

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
CDOT-DTD-2004-9		

4. Title and Subtitle	5. Report Date
USING RTD's TRANSIT VEHICLES TO DEVELOP FREEWAY SPEED MAPS FOR COLORADO	June 2004
	6. Performing Organization Code
7. Author(s) Sarosh Khan Reema Kundu	8. Performing Organization Report No. CDOT-DTD-2004-9
9. Performing Organization Name and Address Colorado TransLab, Department of Civil Engineering University of Colorado at Denver Campus Box 113, P.O. Box 173364 Denver, CO 80217	10. Work Unit No. (TRAIS) 11. Contract or Grant No. 90.46
12. Sponsoring Agency Name and Address Colorado Department of Transportation - Research 4201 E. Arkansas Ave. Denver, CO 80222	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes
Prepared in cooperation with the US Department of Transportation, Federal Highway Administration

16. Abstract
The objective of this research was to develop an algorithm to estimate traffic speed for freeway sections in real-time, as part of a traveler information system. The location data for RTD buses were collected for several months to develop and test an algorithm to estimate traffic speed. The plan is to provide travelers traffic speed information, updated every few minutes, for freeway sections. Typically, this type of reporting is based on the data collected by fixed sensors such as detectors, video cameras and other sensors located on the freeway. However, CDOT is unable to report traffic information for freeway sections without such infrastructure in place. On the other hand, RTD buses traversing these same sections are equipped with GPS receivers, the data from which can be utilized in estimating traffic speed. As part of this project, a statistical model to estimate traffic speed from bus speed, geometric characteristics of freeway and weather conditions was developed. The model was developed and tested based on data collected for a 13-mile section of the Interstate 25 (I-25) freeway. The model's performance was further examined based on data collected for an 11-mile section of the Interstate 225 (I-225) freeway.

Implementation:
This research provides the methods and tools required to process RTD's AVL bus data to estimate traffic speed. Given that the data processing methods and the speed estimation algorithm have been developed, this model may now be implemented in the CDOT Transportation Management Center (CTMC) to estimate traffic speed. To significantly improve the availability of probe reports and to improve the performance of the model, it is suggested that CDOT and RTD consider exploring several options that include more frequent reporting of bus locations and improving the communication links to allow the data to be managed and processed for use by both agencies. As CDOT strives to provide real-time, accurate, reliable traveler information for major corridors, this research has shown that cooperation between agencies may allow them to leverage infrastructure investment dollars and develop strong partnerships of mutual benefit to serve both the traveling motorists and transit riders simultaneously.

17. Keywords	18. Distribution Statement
traffic speed, traveler information systems, GPS, least squares method, non-parametric regression, maximum likelihood estimation, logistic regression, proportional odds model	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	163	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGEMENTS

We would like to acknowledge the guidance of the project manager, Bruce Coltharp, and the study panel members, Frank Kinder, Bob Wycoff, Saeed Sobhi, Lou Lipp, Bob Tenney, William Johnson, Richard Griffin and Scott Sands.

The assistance of the following organizations and researchers is appreciated and gratefully acknowledged.
Susan Wayland and Tom Gardner, Regional Transportation District.
Gordy Hickman and Lou Lipp, Colorado Department of Transportation.

The research team lead by the principal investigator, Sarosh Khan included the following graduate and undergraduate students: Reema Kundu, Kittichai Thanasupsin, Jennifer Stumbo, Shadi Hakimi, Paison Areeprasertkul and Wesley Go, Colorado TransLab, Department of Civil Engineering, University of Colorado at Denver.

EXECUTIVE SUMMARY

The objective of this research was to develop an algorithm, to estimate traffic speed for freeway sections in real-time, as part of a traveler information system. The Colorado Department of Transportation (CDOT) currently monitors traffic conditions on limited sections of freeway based on fixed sensors installed on the mainline for ramp metering purposes. Based on these sensors, CDOT currently provides traveler information on its web site at www.cotrip.org and is extending its freeway surveillance coverage to provide traffic information to travelers. This research was funded to investigate the feasibility of using current infrastructure, Automatic Vehicle Location (AVL) system, in transit vehicles operated by the Regional Transportation District (RTD), to estimate traffic speed.

As part of the RTD's AVL system, all buses in Denver are equipped with global positioning systems (GPS) and dead reckoning (DR) sensors. The location data for all buses is reported every two minutes to the RTD Operation Center for fleet management. As part of this research, this data were collected for several months to develop and test an algorithm to estimate traffic speed.

The plan is to provide travelers traffic speed information, updated every few minutes, for freeway sections. Typically, this type of reporting is based on the data collected by fixed sensors such as detectors, video cameras and other sensors located on the freeway. However, CDOT is unable to report traffic information for freeway sections without such infrastructure in place. On the other hand, buses traversing these same sections are equipped with GPS receivers, the data from which can be utilized in estimating traffic speed. Several factors including, but not limited to, weather and freeway geometry may affect bus and traffic speed.

As part of this project, a statistical model to estimate traffic speed from bus speed, geometric characteristics of freeway and weather conditions was developed. The model was developed and tested based on data collected for a 13-mile section of the Interstate 25 (I-25) freeway. The model's performance was further examined based on data collected for an 11-mile section of the Interstate 225 (I-225) freeway. Least-squares method, non-parametric regression and maximum likelihood estimation method were used for model development. After a series of model

evaluations, logistic regression with proportional odds model was selected for the model development. The final model presented to estimate segment speed at 15-minute intervals is statistically significant with an R^2 of 94% and overall correct classification of 96% for both the calibration and the test set.

This research project has demonstrated the feasibility of using bus location data to estimate traffic speed at regular intervals for freeway sections. The model developed to estimate traffic speed performs well; however, it is conditional on the availability of bus reports from the AVL system. A non-linear model was developed to estimate speed, which exceeded expected performance. Test results show that the model performed equally well on both sections of freeway. The analysis indicates that the model would operate at an acceptable level for any freeway section with similar infrastructure. The model is statistically significant at 0.05 alpha level, according to the model chi square statistic. The model predicts 94% of the responses correctly. The results also indicate that the percent correct classification of the low (0-20 mph) traffic speed is lower compared to the other categories (20-40 mph and >40 mph), both for the calibration and the test set. Most importantly, the performance of the model depends on the number of bus reports available. The bus reports were available for 80% of the 15-minute time intervals based on the current AVL reporting interval of two minutes.

This research provides the methods and tools required to process RTD's AVL bus data to estimate traffic speed. Given that the data processing methods and the speed estimation algorithm have been developed, this model may now be implemented in the CDOT Transportation Management Center (CTMC) to estimate traffic speed. The CTMC is currently developing a GIS-based freeway speed map. The algorithm developed here may be used to develop a freeway speed map as envisioned in the project proposal.

To significantly improve the availability of probe reports and to improve the performance of the model, it is suggested that CDOT and RTD consider exploring several options that include more frequent reporting of bus locations and improving the communication links to allow the data to be managed and processed for use by both agencies. Current RTD operations for fleet management do not require more frequent reporting of bus location by the AVL system.

However, joint efforts are expected to yield unprecedented benefits to both agencies.

As CDOT strives to provide real-time, accurate, reliable traveler information for major corridors, this research has shown that cooperation between agencies may allow them to leverage infrastructure investment dollars and develop strong partnerships of mutual benefit to serve both the traveling motorists and transit riders simultaneously.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

AVL Data – Automated Vehicle Location system.

Bernoulli Distribution – This distribution describes the pattern of variation of a random variable that counts the number of "successes" in an occurrence in which one of the two possible outcomes can occur.

Bus Reports – The bus location data reported at regular intervals, e.g., every two minutes through the AVL system.

Calibration Set – The data set used to estimate the coefficients of the speed estimation model.

Continuous Variable – A variable that may take on any value, integer or fraction.

Correlation – It is a measure of the relationship between two variables. Here it was used to measure the relationship between traffic speed and bus speed.

CTMC – Colorado Transportation Management Center, an ITS (Intelligent Transportation Systems) center to plan, operate, manage and maintain the functions of advanced traffic management system (ATMS), advanced traveler information system (ATIS), and commercial vehicle operations (CVO) based on real-time information received from a transportation system.

Cumulative Logit Model – A model to estimate the likelihood of traffic speed for a freeway segment being in a particular category, dependent on a set of independent factors.

Data Pre-Processor Program – A program written to process the data available from the agencies to prepare to estimate bus speed and other variables serving as input to the traffic speed model.

Degree of Freedom – A mathematical concept, which indicates the number of observations or values in a distribution, that are independent of each other or are free to vary. They are used with various measures such as t-tests, analysis of variance, Chi-square, etc., to refine the results of treatments of probability or chance in determining statistical significance. The abbreviation for degrees of freedom is "DF" and appears routinely on many statistical reports.

Density – Number of vehicles per unit length (vehicles/mile).

Detector Station – The location of fixed sensors.

Deviance Statistic – It is a measure of the deviance of the estimated value from its true value.

Distance Weighted Speed – Average speed is estimated based on weighting individual speed observations on the distance traveled by the vehicle.

Dual Loop Installations – The detector installations on a freeway are said to be dual loop, when two loops per lane are at a particular station. These types of detectors are used to measure vehicle speed and arrival times at a single station.

Flow Rate – The rate in vehicles per hour at which traffic traverses a freeway segment.

Freeway Segments – A freeway section may be divided into segments based on either the position of on ramps or the posted speed limit or other criteria.

Generalized Additive Model – It is a regression technique used as scatter plot smoother. The algorithm in this model fits a smooth curve and displays the trend of the data points.

Geometric Characteristics – This includes the geometry of the freeway e.g., number of lanes, lane width, number of ramps and grade.

Global Positioning System – It is a worldwide satellite based navigation system formed from a constellation of 24 satellites and their ground stations.

Goodness of Fit Statistic – It is a statistical test in which the validity of one hypothesis is tested without specification of an alternate hypothesis.

Hetereoscedastic – The data set where the error variance related to an observation point is not constant. It is a non-uniform error variance.

High Occupancy Vehicle – Vehicles with more than one passenger.

Independent Variables – The variables used in regression analysis to explain the relationship with a dependent variable.

Instantaneous Speed – The speed at any given instant in time.

Interaction Variable – A variable used in regression analysis to represent the interaction between two or more independent variables.

Least Squares Method – It is a method used in regression analysis, that estimates the best fit curve based on minimizing the sum of the squared errors of a given set of data points.

Link Travel Time – The travel time of any vehicle for a segment of freeway.

Log Likelihood/LL – It is a function, which is a basis for deriving estimators for parameters of the given data. This function links the unknown model parameters to the assumptions and allows rigorous statistical inferences.

MLE – Maximum likelihood estimate of the unknown parameter in the model is that value that maximizes the log-likelihood.

Mobile Sensors – Sensors located within a vehicle to provide the location of the vehicle.

Model Chi-Square – It is a measure of overall fit of the model to the data.

NAD-27 – North American horizontal Datum of 1927 is a horizontal datum defining a relationship between physical earth and horizontal coordinates such as latitude and longitude.

Occupancy – It is usually expressed as a percentage; it is the percent time a detector is occupied by vehicles traveling over detectors.

Off Ramp – A ramp for the traffic to exit the freeway.

On Ramp – A ramp for the traffic to enter the freeway.

Ordinal or Nominal Variable – A categorical variable.

Outliers – The data points present in a data set due to an error.

Piece-Wise Generalized Linear Regression – It is a regression technique similar to linear regression but fits several straight lines to the data set.

P-Level – It is the probability level at which the null hypothesis in regression is accepted or rejected.

Polytomous Logistic Regression – It is an approach to predict variables just like ordinary least squares method but produces a polytomous outcome.

Prevailing Traffic Speed – The traffic speed at the prevailing traffic conditions.

Probe Vehicle Data – The data from probe vehicles or mobile sensors.

Proportional Odds Assumption – It is also called parallelism model, it tests whether the parameters are same across logits, simultaneously for all predictors.

Qualitative or Categorical – A variable whose values cannot be interpreted as numbers.

R^2 – It is an indicator of the percentage of variation in the data explained by the model.

Response Variable – The dependent variable estimated in regression analysis.

SAS – Statistical Analysis Software, provides extensive statistical capabilities with tools to handle a wide range of statistical analyses, including analysis of variance, regression, categorical data analysis, multivariate analysis, survival analysis, psychometric analysis, cluster analysis, and nonparametric analysis.

Section Density – The traffic density measured over a section of the freeway.

Skewness – It is a degree of asymmetry of a distribution.

Sky Conditions – The conditions of the sky during any hour of the day like, thick cloud, moderate cloudy and clear sky.

Space Mean Speed – The arithmetic mean of the speed of vehicles on a given segment at a given instant.

Spline Fit – It is a smoothing spline fit to a data set.

Spot Speed / Time Mean Speed – The speed of a vehicle measured at a particular spot or location; usually loop detectors provide this speed.

Standard Deviance – It is a measure of dispersion.

Standard Error – SE of the statistic is the standard deviation of the distribution of that statistic.

Standard Normal Deviate (SND) – It is the probability associated with an observation within a population. If the mean and standard deviation of the population is known then the SND can be computed as, $Z = (y-\mu) / \sigma$, where y is an observation, μ is the population mean and σ is the population standard deviation.

Time Weighted Speed – The average speed weighted based on the time a vehicle spends within a segment.

Visibility – It is expressed in meters; it is the measurement of visibility from the bumper of one vehicle to the bumper of the next vehicle.

Wald Statistic – This statistic is a test of significance of regression coefficient.

Weather Type or Weather Condition – It is code describing the weather condition as sunny or rainy.

Weaving Area– The area located within freeways, with conflicting movement of vehicles.

Wind Speed – It is a measure of the speed of the wind; it is expressed as miles per hour.

1.0 INTRODUCTION

The main goal of this project was to develop and test an algorithm to estimate traffic speed on freeway sections based on bus speed, geometric characteristics of freeway and weather conditions. The study area selected for this project consists of a 13-mile section of Interstate 25 (I-25) freeway, from the County Line Road to Colorado Boulevard, and an 11-mile section of Interstate 225 (I-225) freeway, from the I-25/I-225 interchange to 6th Avenue.

The algorithm was developed based on data from the Colorado Department of Transportation (CDOT), the Regional Transportation District (RTD) and the National Oceanic and Atmospheric Administration (NOAA). CDOT provided fixed sensor data for the I-25 and I-225 freeway sections, and RTD provided the AVL data for approximately 14 weeks. NOAA provided weather data for the same period. The sensor and AVL data were used to develop the speed model and to evaluate its performance based on statistical and operational performance measures. The consecutive location data from the RTD buses were used to estimate bus speed. Since the buses are equipped with GPS receivers they serve as mobile sensors. This study examines the feasibility of using bus AVL data to estimate traffic speed to develop a color-coded speed map of Colorado freeways. The weather was expected to affect the relationship between bus speed and traffic speed. Therefore, weather data including wind speed, visibility and type of weather were considered in developing the algorithm. Based on the algorithm developed, CDOT can develop speed maps for traveler information as part of the Advanced Traffic Management System and Advanced Traveler Information System being developed by the CTMC.

This chapter presents the purpose of the project, and a review of the literature in this area. Chapter 2 of this report presents details of the test networks selected. The model development and its testing are presented in Chapters 3 and 4, respectively. Chapter 5 summarizes the findings and the conclusions of the project.

1.1 Purpose

The main objective of this research was to develop a model to estimate traffic speed of freeway sections based on bus speed estimated from AVL in buses. Preprocessing procedures, tools and methods for the bus AVL data were also developed. The statistical goodness-of-fit analysis and the operational performance of the model for freeway sections were also examined.

1.2 Background

In the past, several research projects have been carried out to estimate freeway traffic speed or travel time based on data collected from fixed sensors such as loop detectors [1], video detection systems [2] and mobile sensors such as global positioning systems in vehicles or probe vehicles [3]. Most of these past studies attempted to estimate travel time based on data from selected automobiles serving as probe vehicles. In 1995, a research study compared link travel time and instantaneous speed from probe vehicles to freeway traffic-management system (FTMS) loop-detector data from loop detectors [4]. This research was carried out in Orlando, Florida on five different major arterials. A few probe vehicles were driven on these networks to obtain the probe data and the travel time of each link from this data. The estimated travel time was compared with the travel time estimated from detectors. Statistical analyses were conducted and high correlation was found between the detector data and the probe data. Finally, the probe speed estimates were compared with the odometer speed measurements for different driver characteristics and network configurations. The probe speed estimates were found to be within 10 km/hr of the actual speed estimates.

Another research project conducted in 1996 in Houston, Texas was aimed at estimating the number of probe vehicles required for the estimation of peak period travel speed [5]. The probe vehicle data from the Houston traffic monitoring system was obtained to statistically predict the travel speed variation and finally the required probe sample size. A regression equation developed for this purpose gave an R^2 value of 0.33, which concluded that there is enough statistical evidence that the speed information provided by the number of probe vehicles was reliable. In 1996 the Texas Transportation Institute in conjunction with the Texas Department of Transportation conducted a pilot study to check the feasibility of probe measured travel time to detect incidents on a freeway [6]. As a part of this study, 200 commuters were equipped with cellular telephones to collect travel time and incident data from three major facilities in Houston, Texas. The statistical principle of standard normal deviates (SND) was used to detect the major incidents. The expected travel time measurements derived from the probe vehicles were used to establish the average and standard deviation of travel times on each link of the system. Overall, the study indicated that the travel time provided by probe vehicles is feasible to achieve some level of incident detection.

A research project similar to the one mentioned above was conducted in 1997 at University of Illinois, Chicago to observe the effect of frequency of probe reports on the variance of link travel time estimates [7]. The analysis was done based on empirical data. This research showed that as probe vehicle speeds are correlated within a segment and a time interval, the standard error of the average speed estimates for a segment is a function of the variance and the covariance of the speeds. The research concluded that a small number of probes within a certain interval, say 5 minutes, yield a standard error that cannot be improved further by increasing the number of probes beyond a certain probe penetration level. Therefore, high levels of probe deployment are not necessary in order to obtain the link travel time as long as the links are covered by at least a few probes.

In 1999, a study conducted by Hellinga examined the sampling bias on accuracy of the probe estimates [8]. This research shows that when the probe vehicles represent a biased sample, the sample mean does not approach the population mean. This research was conducted on a typical link bounded by a signalized intersection at upstream and downstream ends, divided into several segments. The link data represented a biased sample when the probes varied between these nine subsections. Thus the degree to which the probe reports represent a biased sample is critical in assessing the reliability of sample mean as an estimate of population mean. This study shows that an effort should be made to identify any systematic bias in the probe sampling.

Finally, a paper presented at the 8th annual meeting of ITS America in Miami in 2001 showed that increasing the number of probes beyond a certain level does not improve the accuracy of the estimates [3]. The analysis of the standard error of the probe vehicle estimates of average link travel time and speed is shown in Figure 1.1. Here the curve flattens as the number of probes traveling a link increase, with very little marginal improvement in accuracy. In addition, the improvement in accuracy of the average link travel time estimate also varies based on flow conditions and link characteristics. Figure 1.1. shows the error analysis for a 2,735 ft link with a

3% grade between Bakerville and Herman Gulch in Colorado.

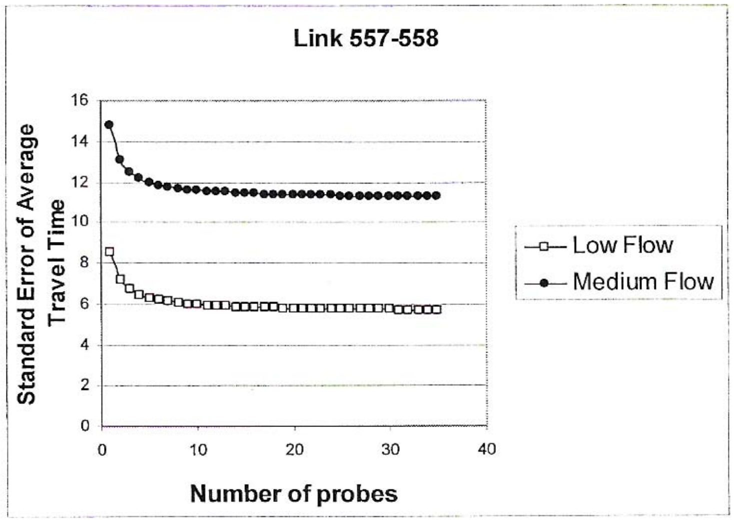


Figure 1.1. Standard Error of Average Link Travel Time Estimate for Link between Bakerville and Herman Gulch[3]

In summary, of all the studies reviewed, the main focus has been on estimating travel time or speed based on mobile sensors in automobiles. The proposed research attempts to examine the feasibility of using vehicle location data from buses to estimate traffic speed for traveler information. The following chapters present the methodology proposed to develop an algorithm to estimate traffic speed, the test networks used for the study and the statistical and operational performance of the model.

2.0 TEST NETWORKS AND DATA COLLECTION

Three test networks were selected to develop and test models to estimate traffic speed based on bus AVL data. This chapter contains five sections explaining in detail the type of data collected, the data processing procedures and tools that were developed. A description of the test networks, including the geometric characteristics, is provided. Details of the detector data, the AVL data and the weather data are also presented. The final section of this chapter summarizes the requirements for estimating traffic speed and calibrating the speed model.

2.1 Test Networks

The data for this project were collected from various sources including the Colorado Department of Transportation (CDOT), the Regional Transportation District (RTD) and the National Oceanic and Atmospheric Administration (NOAA). The main objective of the project was to develop an algorithm to estimate traffic speed from bus speed for a freeway section. A test section of Interstate 25 (I-25) for a distance of 13 miles was considered for model development and calibration. Figure 2.1 shows the study area. The study area includes 10 detector stations, one each located upstream of each on-ramp of the freeway. A detailed description of the sensors is presented in section 2.2 Detector Data.

The test network is located in the Denver Metro area where the highest traffic volumes in Denver are observed. Data were collected during the morning and evening peak hours, three hours (6AM-9AM) in the morning and three hours (3PM-6PM) in the evening. The data were collected for 14 weeks (five weeks in the months of April and May, 2001 and nine weeks from September to December, 2001). For model development and calibration purposes, data for three weeks during the April-May period and six weeks in the September-December period were used. The rest of the data were used to test the operational performance of the model. The model was also applied to another freeway section of Interstate 225 (I-225) to test its operational performance.

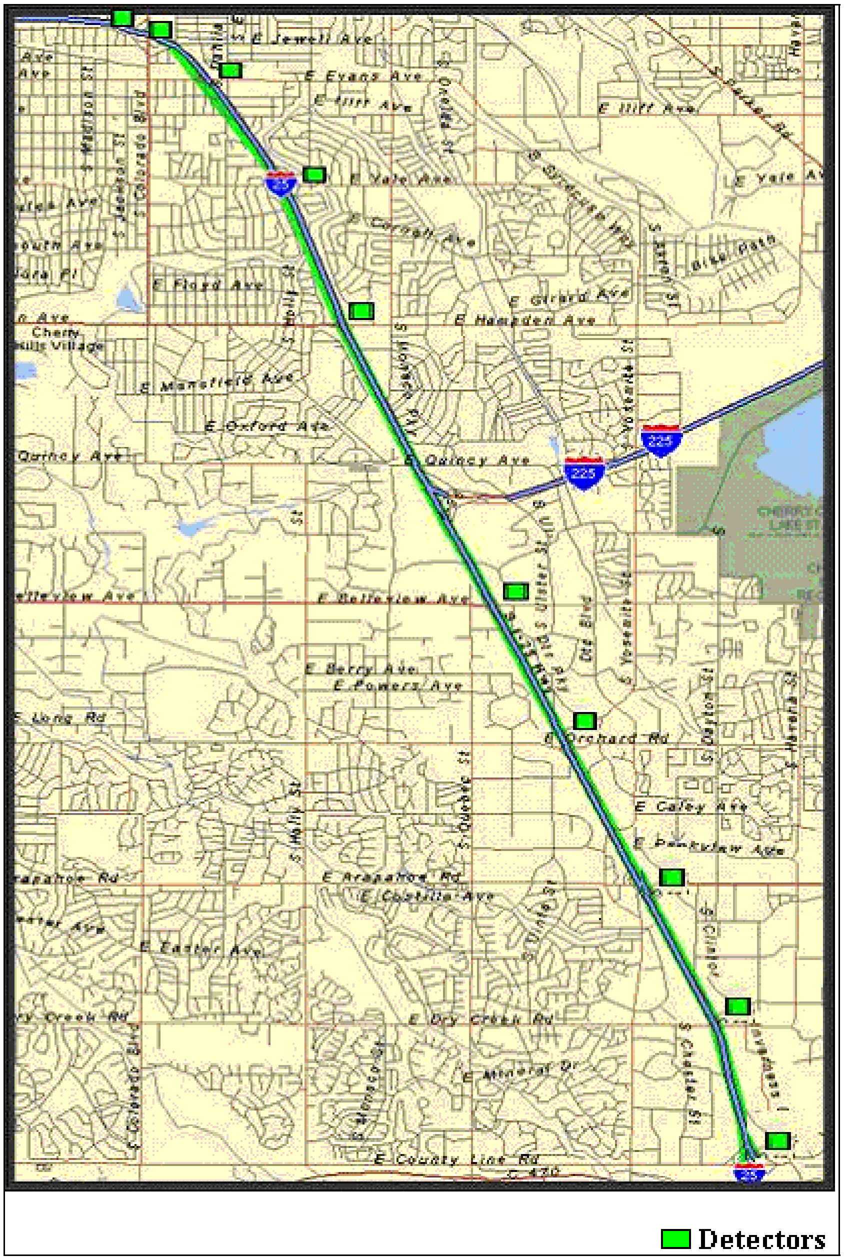


Figure 2.1. Schematic of the Interstate 25 Test Section.

Figure 2.2 shows a schematic of the I-225 section from the I-25 interchange to 6th Avenue. Data for I-225 were collected for five weeks over an 11-mile section of roadway. The geometric characteristics of I-225 are similar to those of I-25. The I-225 freeway section includes four fixed detector stations. The distance between the first and the second detector is five miles. The distance between the second and third and the third and fourth is two miles each.

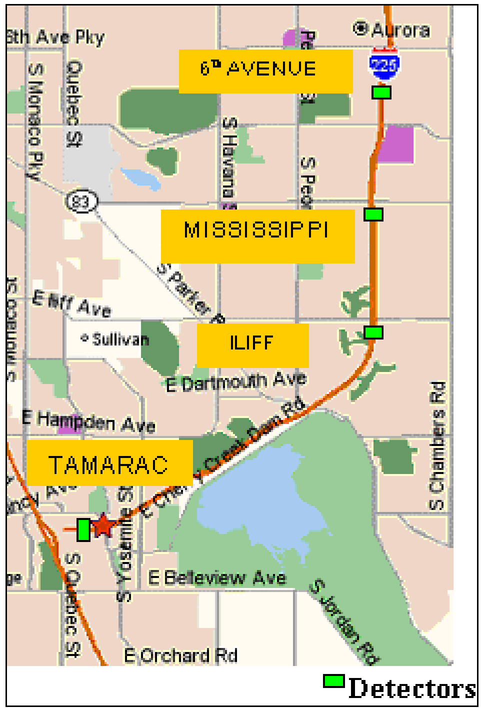


Figure 2.2. Schematic of the Interstate 225 Test Section.

Neither test network had any High Occupancy Vehicle (HOV) lanes at the time of the study; therefore, the buses traveled along with the mainline traffic. The I-25 test section generally has three lanes in each direction, with a number of on and off-ramps spaced approximately one mile apart. Figure 2.3 shows a picture of I-25 with three lanes in each direction near the off-ramp at Yale Avenue. Detectors are located on the freeway just upstream of each on-ramp. The data from the detectors were used to compute traffic speed. All RTD buses on this freeway section are equipped with Geographic Positioning System (GPS) receivers and dead-reckoning sensors, thus the bus location reports were obtained every two minutes.

Figure 2.3. A Section of Interstate 25 near Yale Avenue Showing 3 Lanes in each Direction

2.2 Detector Data

Fixed detector data were used to estimate traffic speed for the development and testing of the speed model in this project. As shown in Figure 2.1, there are 10 detectors located near on-ramps. Both sections of I-25 and I-225 have dual loop installations that report flow, occupancy and speed. Table 2.1 presents the location of the detectors on the I-25 freeway section. The average spacing of the detectors is approximately one mile. All the detectors are identified based on their proximity to the nearest on-ramp. Table 2.2 presents the detector locations for the I-225 freeway section. Except for the distance between the first two detectors (north bound direction) the rest of the detectors are approximately two miles apart. Some of the detectors malfunctioned during the data collection period. Therefore, limited data were available for the I-225 freeway section.

Table 2.1. Detector Locations on I-25.

Detector Location	Distance from Reference Point* (miles)
County Line Blvd.	1.82
Dry Creek Rd.	3.54
SE Arapahoe Rd.	4.60
NE Arapahoe Rd.	5.99
Orchard Rd.	7.15
Belleview Ave.	9.47
Hampden Ave.	10.48
Yale Ave	11.44
Evans Ave.	11.89
Colorado Blvd.	12.13

* The reference point is located about 1.82 miles south of the ramp at County Line Blvd. on I-25.

Table 2.2. Detector Locations on I-225.

Detector Location	Distance from Reference Point* (miles)
Tamarac Pky.	1.00
Iliff Ave.	5.84
Mississippi Ave.	7.34
6th Ave.	10.96

* The reference point is located about a mile south of the ramp at Tamarac Pky. on I-25.

2.3 AVL Data

The Regional Transportation District, the transit agency in the greater Denver Metro area, installed an Automatic Vehicle Location (AVL) system in 1993 to develop more efficient transit schedules, to improve the agency's on-street operations, and to increase safety through better management [9]. RTD’s AVL system components are shown in Figure 2.4. Each vehicle in the

RTD fleet is equipped with an Intelligent Vehicle Login Unit (IVLU) and a global positioning system (GPS) receiver capable of real-time differential correction. Although, a GPS receiver's signal may degrade due to obstructions in urban environments, RTD's AVL system integrates the GPS with inertial sensors or dead-reckoning (DR) sensors. The location accuracy of this type of GPS receiver is between 1-2 meters. However, specific information on the accuracy of RTD's integrated GPS-DR system is not available. The RTD fleet consists of 1,335 vehicles, including 935 fixed route buses. Bus location data are available every two minutes through this AVL system and was collected for this study.

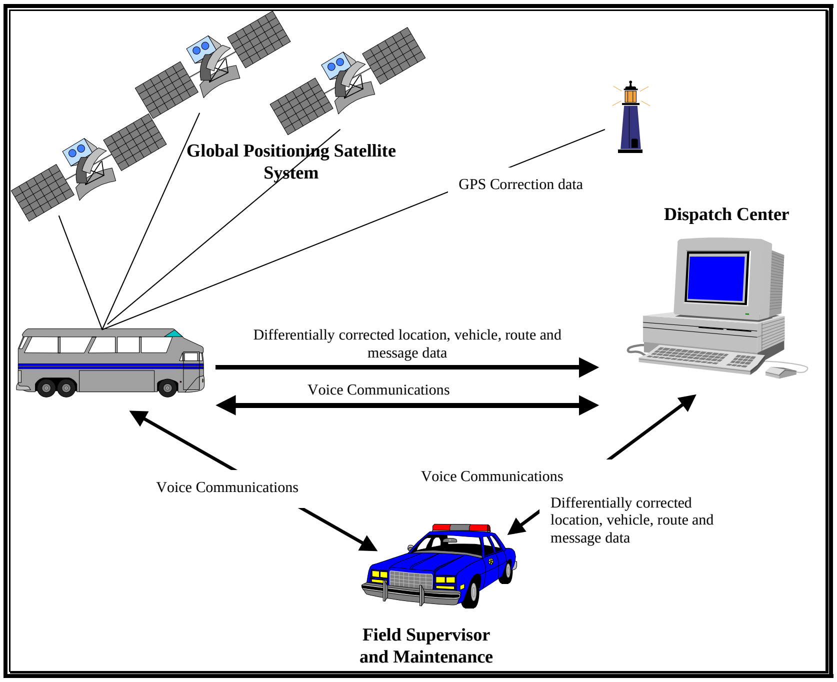


Figure 2.4. RTD's Automatic Vehicle Location System.

Typically, there are 18 bus routes on the northbound I-25 test section. Each route operates specific number of buses with different IVLUID, thus the number of bus reports directly depends upon the number of buses operating at certain hour of the day. During the morning peak period, the average bus flow rate is 14 buses per hour between County Line Road and the I-225 Interchange, and 23 buses per hour between the I-225 Interchange and Colorado Boulevard. During the afternoon peak period, the average bus flow rate is 4 buses per hour between County Line Road and the I-225 Interchange, and 11 buses per hour between the I-225 Interchange and

Colorado Boulevard. More buses operate on the northbound I-25 section of the test network during the morning peak period than during the afternoon peak period.

The Automated Vehicle Location (AVL) data obtained from the RTD included date, time, unique bus identification number (ID), route number, and location of buses in NAD 27 State Plane coordinate system. The unique bus ID appears on buses as shown in Figure 2.5. The bus ID, which uniquely identifies the buses, allows bus speed to be estimated based on data from the same bus at two consecutive locations.

Figure 2.5. Intelligent Vehicle Logic Unit ID Shown at Top Right Corner of RTD Bus.

2.4 Weather Data

The weather data were also obtained from NOAA for the morning and evening peak periods. The data included sky conditions, visibility, weather type, temperature, humidity, wind speed, pressure and precipitation. Based on the weather data, weather condition was determined and separated into five categories: sunny, cloudy, storm, rain and snow. Three variables, visibility, wind speed and weather conditions were expected to have an impact on traffic conditions.

2.5 Data for Model Development and Testing

The data used for model development included traffic speed, bus speed, geometric characteristics and weather condition. For calibration of the model a total of 9 weeks of data were used and 5 weeks of data were used for testing the speed model. The breakdown of the data set by calibration and test data set is summarized in Table 2.3.

Table 2.3. Data Used to Develop and Test the Speed Model.

Month	Data Available	Calibration Data Set	Test Data Set
April	1 week	1 week	
May	3 weeks	2 weeks	1 week
September	1 week	1 week	-
October	4 weeks	4 weeks	-
November	4 weeks	1 week	3 weeks
December	1 week	-	1 week
Total	**14 weeks**	**9 weeks**	**5 weeks**

A detailed explanation of the data processing procedures is presented in the next chapter and the calibration and testing of the model are presented in subsequent chapters.

3. 0 DATA PROCESSING

Data collected from AVL buses, detectors, geometric characteristics and weather must be preprocessed to estimate traffic speed for freeway sections. Data sets were developed based on this preprocessing procedure. In this chapter, the data processing methodology is presented in separate sections for detector data, bus data and weather and geometric data. The last section describes the data pre-processor program developed.

Figure 3.1 below illustrates the procedure adopted in developing the speed model. Initially the detector data, bus data and weather data were processed using a database program, described in Appendix C, to estimate traffic speed and bus speed for freeway segments. The output from this pre-processor was used to develop a statistical speed model. The following sections describe the computation of traffic speed, bus speed and some weather variables in detail.

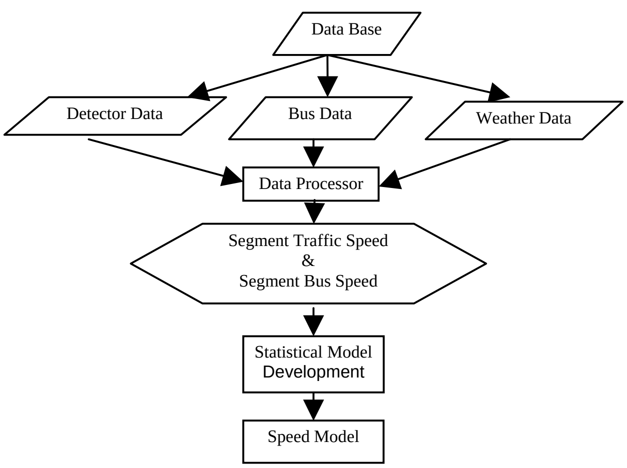


Figure 3.1. Model Development.

3.1 Segment Length Selection

To compute traffic speed and bus speed for freeway segments over fixed time intervals, segment length selection was examined. Various combinations of segment length and time intervals were used to develop the speed model. Fixed segment lengths of one and two miles, and variable segment lengths for five and 15 minute time intervals were considered. Given the current configuration that RTD's AVL system reports bus locations every two minutes, the five-minute time interval yielded very few bus reports per interval and therefore was not considered for further evaluation. Therefore, a 15-minute time interval to estimate traffic speed was considered appropriate for further analysis.

3.2 Detector Data

Detector data are available at the CDOT Transportation Management Center (CTMC). It is time stamped and includes the detector location, spot speed, occupancy and volume. It is updated every minute for all lanes of the selected freeway sections. To obtain traffic speed to develop the algorithm, space mean speed and the density were estimated from the detector data. Initially the time mean speed acquired from the detector data was converted to space mean speed to obtain average speed over a detector location. The average space mean speed was then utilized to compute density over the detector location, which is designated as lane density throughout the report. The lane density was then used in deriving traffic density for every segment of the freeway. The main objective behind all these exhaustive computations is to calculate the traffic speed, which will be used in the algorithm. The following compilation is a detailed description and derivation of the traffic characteristics described above.

The speed, flow and occupancy obtained from the detectors were measured every minute for the 10 detector stations on I-25. From the one-minute measured spot speeds, the average 15 minute time mean speed was computed [10] using the following equation,

$$u_t = \frac{1}{N}\sum_{i=1}^{N} u_i \tag{3.1}$$

where u_t is the 15-minute average spot speed and u_i is the 1-minute spot speed. Data for both morning and evening peak periods were analyzed. Space mean speed was estimated from time mean speed for 15-minute time intervals, for a section between detectors as follows [10], where u_s is the average space mean speed, u_t is the average time mean speed and σ_t^2 is the variance over time mean speed. The following equation illustrating the relationship between time mean speed and space mean speed was initially recognized by Wardrop [11],

$$u_s = u_t - \sigma_t^2 / u_t \tag{3.2}$$

The next macroscopic traffic parameter of interest, density (k) is estimated from flow (q) and average space mean speed (u_S) [12], For example for a given detector station at a particular time, the traffic density per lane or the lane density would be given by the following equation,

$$k = \frac{q}{u_s} \tag{3.3}$$

Based on the space mean speed estimated according to Eq (3.2) and the 15-minute flow rate estimated from one-minute volumes, Eq (3.3) was used to estimate traffic density from the average flow rate and the space mean speed for all lanes at each detector location.

Usually density can be computed from the occupancy provided by the detector data, but Zhou [13] has shown that density estimated from flow rate and space mean speed is more reliable [14]. Therefore, lane density was determined from the flow and the space mean speed and in turn used to compute the section density for each segment. Each freeway segment may include one or more detector stations. Figure 3.2 shows the detector stations for typical segments of varying length.

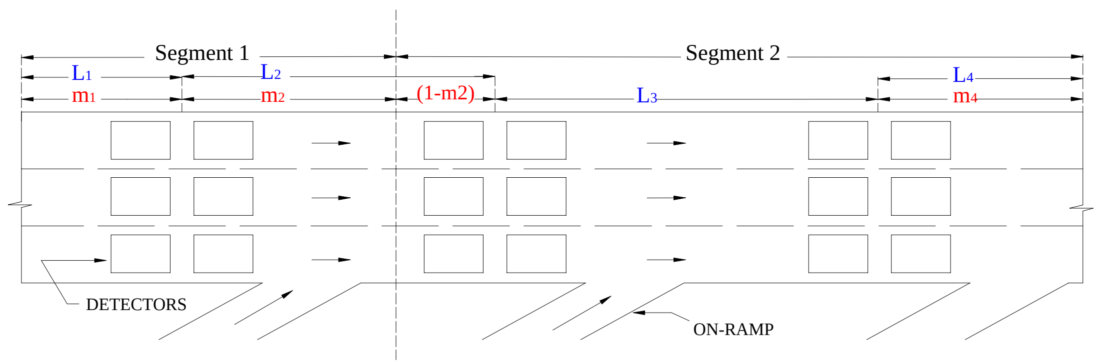


Figure 3.2. Freeway Segments of Varying Lengths and Detector Locations.

The segment density was estimated based on the density estimates at one or more detector stations within each segment as follows,

$$K_n = \frac{\left(k_q * L_q * m_q\right) + \left(\sum_{i=q}^{p} [k_i * m_{i+1} + k_{i+1}(1 - m_{i+1})] * L_{i+1}\right)}{\sum_{i=q}^{p+1} m_i * L_i + \sum_{i=q+1}^{p} (1 - m_i) * L_i} \tag{3.4}$$

This equation may be applied to estimate the density for each lane of the segment separately and in turn used to estimate segment density.

where,

n = segment number

L_i = distance between detector station *i-1* and *i*

m_i = fraction representing a detector's contribution to the density estimate of a segment based on the location of the detector with respect to the segment.

$$m_i \begin{cases} = \dfrac{\text{distance between segment's upstream boundary and first detector station}}{\text{distance between the upstream segment detector station and first detector station}} \text{....... } \textit{for first} \text{ detector} \\ = \dfrac{L_i}{2} \text{ .... } \textit{for} \text{ the section between detector stations} \\ = \dfrac{\text{distance between segment's downstream boundary and the last detector station}}{\text{distance between last detector and the next downstream detector}} \text{ .... } \textit{for} \text{ last detector} \end{cases} \tag{3.5}$$

k_i = density at detector station *i*

q = 1st detector location within segment *n*

p = the last detector location within segment *n*

For two special cases corresponding to *q* and *p* detector locations, L_1 = distance from the beginning of the network to the first detector and L_{11} = distance from the last detector to the end of network.

The above density equation may be explained in the following example. Figure 3.3 below shows the first, second and the last segment of a network. The first segment includes only one detector station, the second segment includes two detector stations and the last segment includes two detector stations.

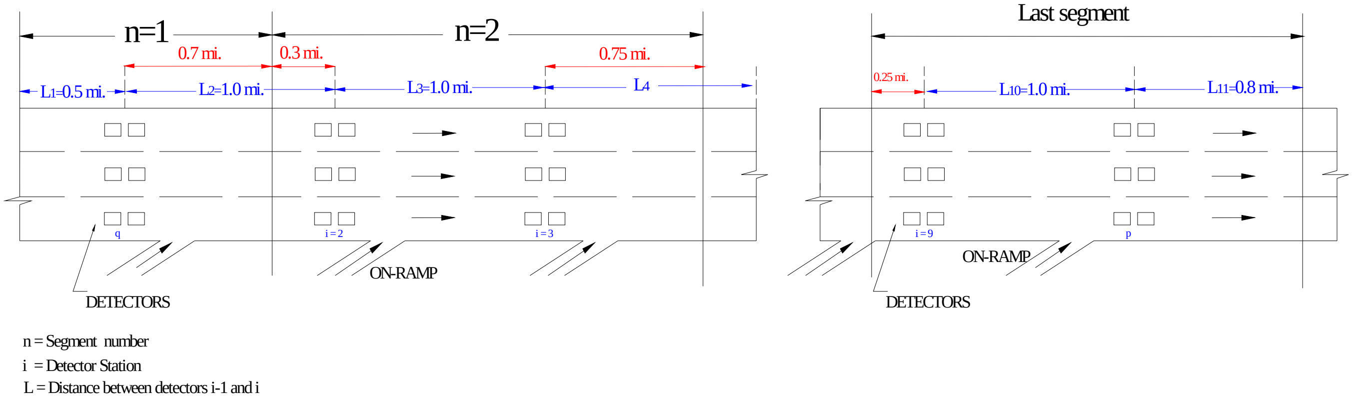


Figure 3.3. Example of Freeway Segments of Varying Lengths and Number of Detector Stations Within a Segment.

As shown in Figure 3.3, the first segment (n =1) includes only one detector station (i=1) located 0.5 miles from the upstream boundary of the segment. As the first segment is a special case, $m_1 = 1$ and therefore, $L_1=0.5$ miles. As the distance between the first detector station and the next is 1 mile and the distance between the first detector station and the downstream segment boundary is 0.7, according to Eq (3.5), m_2 corresponding to L_2 is 0.70 mile, thus $(1\text{-}m_2)$ = 0.30 mi. Let the density at the first be 28 veh/mile. Substituting the parameters in eq.(3.4), we get,

$$
\begin{aligned}
K_1 &= \frac{(k_1 * L_1 * m_1) + (k_1 * m_2 * L_2)}{m_1 * L_1 + (1 - m_2) * L_2} \\
&= \frac{(28*1*(0.50/1)) + (28*1*(0.70/1))}{0.50*1 + (1-0.70)*1} \\
&= 28\,veh/mi
\end{aligned}
\tag{3.6}
$$

Similarly, considering segment 2 in the Figure 3.3, the following parameters can be derived,

$(1\text{-}m_2)$ = 0.30 mi

m_3 = Li /2 = ½ = 0.5 mi

m_4 = 0.75 mi

L_2 = 1.0 mi

L_3 =1.0 mi

L_4 =1.0 mi

Let, k_2 = 68 veh/mi and k3 = 80 veh/mi. Substituting in eq. 3.2, we get,

$$K_2 = \frac{\left(k_2(1\text{-}m_2)*L_2\right)+\left(k_2*m_3*L_3\right)+\left(k_3*(1-m_3)*L_3\right)+\left(k_3*m_4*L_4\right)}{\left((1\text{-}m_2)*L_2\right)+\left(m_3*L_3\right)+\left((1\text{-}m_3)*L_3\right)+\left(m_4*L_4\right)}$$

$$= \frac{\left(68(1\text{-}0.70)*1\right)+\left(68*0.50*1\right)+\left(80*(1-0.50)*1\right)+\left(80(0.75)*1\right)}{\left((1\text{-}0.70)*1\right)+\left(0.50*1\right)+\left(0.50*1\right)+\left(0.75*1\right)} \tag{3.7}$$

$$= 75.32 veh/mi$$

3.3 Bus AVL Data

The Bus AVL (automated vehicle location) data obtained from RTD included location, date and time, and coordinates of the buses. The *x,y* coordinates recorded from the AVL data were projected in ArcView, and Network Analyst was used to estimate the distance each bus was from a fixed reference point on the test network. Currently, RTD's AVL system is configured to report the location of all buses every two minutes. Therefore, from two consecutive reports for the same bus, the distance traveled and time elapsed between reports was computed. For a group of reports within a segment (current segment) and a given time interval (current time interval), the average bus speed was computed. This is the simplest case. However, two consecutive reports from the same bus may not be dispatched from the current segment (or the segment for which the speed is being computed) and during the current time interval. However, even if both consecutive reports are not from the current segment and are not reported during the current time interval, these reports may be used to compute the average bus speed. Several different cases of bus reports were identified that could be used to compute the average bus speed within a segment. In a time-space diagram, Figure 3.4, Figure 3.5, Figure 3.6, and Figure 3.7 show the trajectory of eight buses to illustrate the eight cases that may arise based on the location of the bus and the time a bus report arrives. In these Figures, the X-axis represents time in 15-minute intervals and the Y-axis represents the distance along the freeway in consecutive segments.

Table 3.1 provides an explanation of the cases that may occur based on the location and the time that the reports arrive. Figure 3.4 through Figure 3.7 show the location and the time interval in which the reports arrive within a particular segment. The shaded area in all the figures show the current time interval and current segment, i.e., the segment and time interval for which average bus speed is being estimated. Since the bus speed is computed from a pair of reports it is important to note that only the second report that arrives in the current time interval are considered for bus speed estimation.

As mentioned earlier, the first case occurs when both the first and the second report arrives when the bus reports its two consecutive locations within the current time interval and the current segment. In this case, the bus speed for this segment may be estimated from the distance traveled and the time elapsed between two consecutive reports. For the second case, the two consecutive bus reports arrive when a bus is within the current segment. However, the first report arrives when the bus is in the previous time interval and the second report arrives during the current time interval. The third case occurs when both bus reports are within the current time interval, but the first report arrives when the bus is upstream of the current segment and the second report arrives when the bus is downstream of the current segment. For the fourth case, the first report is upstream of the current segment and the second report is downstream of the segment. These cases are then subdivided depending on when they reach the beginning of the current segment. The time interval that a pair of reports belongs to, or is used for, is determined based on the time the second report of the pair arrives. There were cases with no buses in a particular time interval for a certain segment. In this case the bus speed was not estimated. Once the records that belong to a particular time interval for a certain segment were grouped, then the bus speed was estimated. A weighted bus speed was estimated as,

$$V_{DWS} = \frac{\sum_{i=1}^{N} d_i^{'} * \frac{d_i}{t_i}}{\sum_{i=1}^{N} d_i^{'}} \tag{3.8}$$

Where,

d, t = distance and time between two reports,

d', t' = distance and time a bus spent in the current segment, and

N = number of pairs of reports.

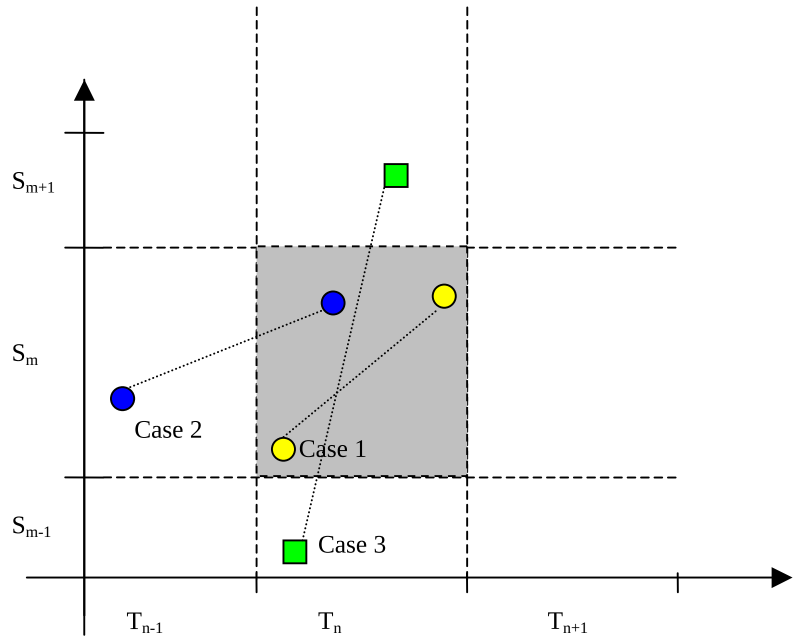


Figure 3.4. Trajectory of Buses Reporting: Location of Buses for Case 1, Case 2 and Case 3 Are Shown for the Current Segment and Current Time Interval (shaded area).

[Note: Sm = current segment, Sm-1 = upstream segment, Sm+1 = downstream segment, and Tn = current time interval, Tn-1 = previous time interval, Tn+1 next time interval]

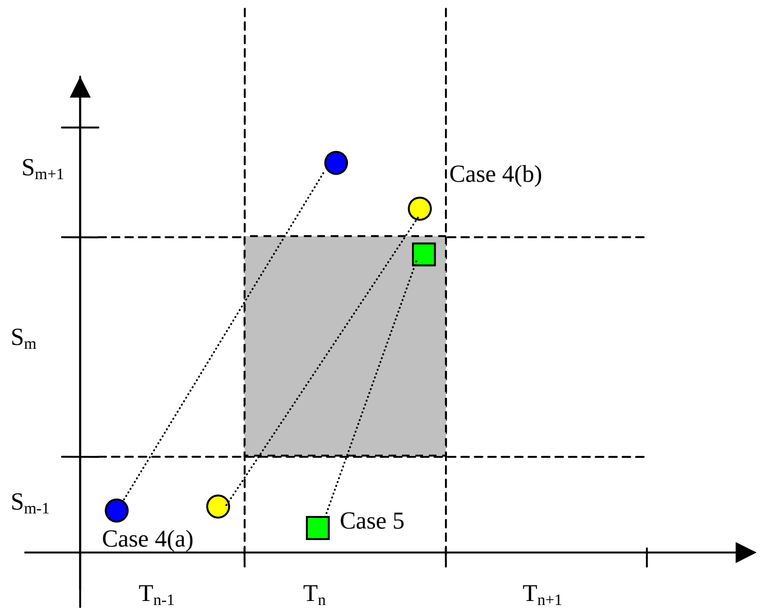


Figure 3.5. Trajectory of Buses Reporting: Location of Buses for Case 4(a), Case 4(b) and Case 5 Are Shown for the Current Segment and Current Time Interval (shaded area).

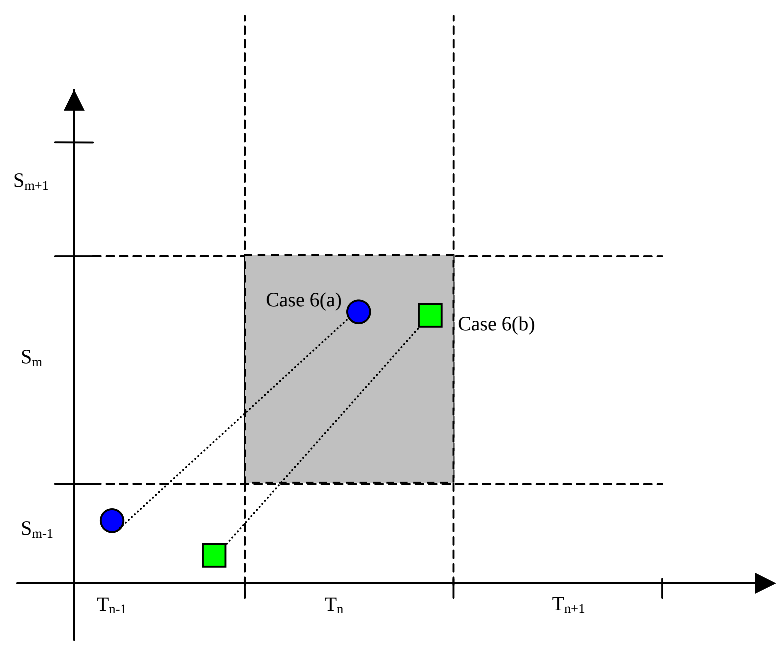


Figure 3.6. Trajectory of Buses Reporting: Location of Buses for Case 6(a)

and Case 6(b) Are Shown for the Current Segment and Current Time Interval (shaded area).

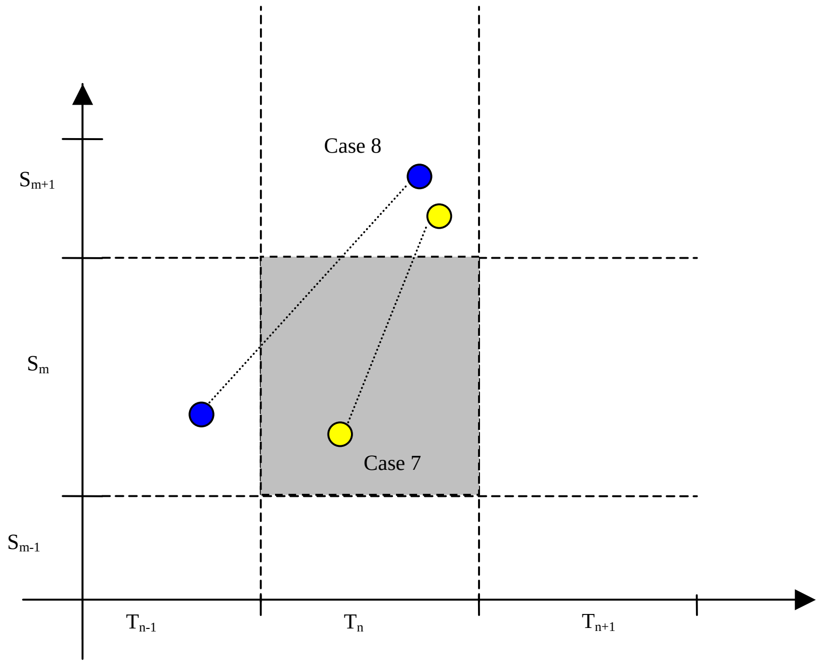


Figure 3.7. Trajectory of Buses Reporting: The Location of Buses for the Case 7 and Case 8 Are Shown for the Current Segment and Current Time Interval (shaded area).

Table 3.1. Eight Cases for Estimating Bus Speed Based on the Location and Time of Two Consecutive Bus Reports.

Case	Time Interval	Bus Location for Two Consecutive Reports
1	Both bus reports in the current time interval	Both bus reports in the current segment
2	The first report in the previous time interval and the second report in the current time interval	Both bus reports in the current segment
3	Both bus reports in the current time interval	First report in the upstream segment and second report in the downstream segment
4(a)	The first report in the previous time interval and the second report in the current time interval. The bus reaches the beginning of the current segment in the previous time interval	First report in the upstream segment and second report in the downstream segment
4(b)	The first report in the previous time interval and the second report in the current time interval. The bus reaches the beginning of the current segment in the current time interval	First report in the upstream segment and second report in the downstream segment
5	Both bus reports in the current time interval	First report in the upstream segment and second report in the current segment
6(a)	The first report in the previous time interval and the second report in the current time interval. The bus reaches the current segment in the previous time interval	First report in the upstream segment and second report in the current segment

6(b)	The first report in the previous time interval and the second in the current time interval. The bus reaches the current segment in the current time interval	First report in the upstream segment and second report in the current segment
7	Both bus reports in the current time interval	First report in the current segment and second report in the downstream segment
8	The first report in the previous time interval and the second report in current time interval	First report in current segment and second report in the downstream segment

3.4 Weather and Geometric Data

The effects of weather on the relationship between bus speed and traffic speed were taken into account based on wind speed, visibility and the type of weather. The data from NOAA consisted of wind speed, visibility and several other variables like, sky conditions, weather type and humidity. Based on weather type, humidity, sky conditions and precipitation records, weather condition was derived. Finally the weather information was summarized in five items: date, time, weather conditions, visibility and wind speed. Since weather conditions influences the traffic speed on a highway, this data were used in exploring the relationships between bus speed and traffic speed.

In addition to the weather information, geometric data were also used in estimating the traffic speed. The geometric data consisted of the number of on-ramps and off-ramps within each freeway segment. Previous research has shown that ramps have a significant influence on the average speed of the highway [15]. As the traffic from an on-ramp attempts to merge into the through traffic lanes, turbulence due to weaving affects the average segment speed. The turbulence or the weaving is high when the distance between an on-ramp and an off-ramp is less than 2500 feet. This condition exists in most of the segments of the test network. Thus these variables were included in the examination of the relationship between traffic speed and bus speed.

3.5 Data Preprocessor

The process of estimating segment density, segment speed and bus speed described in the previous section was automated using a processor program written in Microsoft Access, a database software. The source code for this program is included in Appendix A. This program is designed to work for any network with different geometric conditions. The program requires five different files: the detector and ramp details, weather, bus, and detector data. The detector details include the number of detectors present in the network, their distances from a reference point and a unique ID for every detector. The ramp information includes the number of on-ramps and off-ramps present on the highway, distances between on-ramps and their reference points. A description of the weather, bus and detector data was provided in Chapter 2. The five files mentioned are required to be placed in one folder to run the program. Therefore, the program takes complete information about the network and computes the necessary parameters required for estimating the traffic speed on the highway segment. The output from this program is used as input for the statistical model designed to estimate traffic speed.

The details of the program and the formatting requirements of the input data files and the output data file are provided in the Appendix C. The next chapter presents the methodology of the speed estimation model.

4.0 METHODOLOGY: ALGORITHM TO ESTIMATE TRAFFIC SPEED

The main objective of this research was to examine the feasibility of using AVL data from buses to estimate traffic speed to provide traveling motorists information on current traffic conditions. The effort in this research was focused on developing statistical models to estimate traffic speed as a discrete variable, for example in discrete ranges of less than 20 mph, 20-40 mph or >40 mph. Although traffic speed may also be estimated as a continuous variable, a preliminary analysis shows that a statistically significant relationship exists between bus speed and traffic speed and the model proposed demonstrates better operational performance as a categorical variable. The functional relationship is also shown to depend on the interaction between bus speed and freeway geometry.

This chapter presents descriptive statistics of the data used to examine the relationship between traffic speed and bus speed for a segment. The statistical techniques that were employed to explore and develop this relationship, as well as the model building process, are also presented. The data preprocessed as discussed in Chapter 3 were used to develop alternate models. The processed data included the following variables: time-interval stamped bus speed estimated from time-stamped AVL data for buses, traffic speed estimated from density and flow determined from detector data, number of on-ramps and off-ramps and weather. These variables were used in the developing the speed model. This chapter also presents the calibrated speed model in detail including a section that describes the methodology adopted in the development of the model, and describes how traffic speed was estimated. The first sub-section explains the basic concepts of regression analysis and descriptive analysis of data to develop an understanding of the underlying relationships in the data set. The second sub-section includes the step-by-step process used to develop the speed estimation model.

4.1 Descriptive Statistics of the Data

The data were analyzed to obtain descriptive statistics such as the mean, standard deviation, skewness, and the distribution of the individual variables of the data set. Table 4.1 presents the descriptive statistics of a few selected variables. While the mean provides a measure of the central tendency of the corresponding variable, one the most important statistics is the standard deviation and the standard error, which provide a measure of the dispersion, and hence an indication, about outliers in the data set. Since the skewness corresponding to most of the variables was very low, a least squares linear regression was initially applied to develop the model. Table 4.2 presents a few statistical elements of the variables, traffic speed and bus speed. An initial analysis of the correlation between bus speed and traffic speed for freeways segments showed a strong relationship (R=0.87). The descriptive analysis also showed a higher correlation between distance-weighted bus speed and traffic speed, thus distance-weighted bus speed was used for further analysis.

Table 4.1. Descriptive Statistics for Freeway Segments of Varying Lengths.

Variable	N	Mean	Std Dev	Std Error	Skewness
Traffic Speed (mph)	3424	50.28	17.240	0.295	1.355
Bus Speed (mph)	2350	40.76	15.690	0.324	-0.203
Number of on-ramps per unit length of freeway segment	3424	2.08	1.237	0.021	0.698
Number of off-ramps per unit length of freeway segment	3424	2.08	1.237	0.021	0.698

Table 4.2. Descriptive Statistics of Peak Period Traffic and Bus Speed on Freeway Segments.

Speed	Time	Mean (mph)	Standard Deviation (mph)	Skewness
Traffic Speed	6:00 A.M.-6:30 A.M.	55.619	2.919	-8.0E-01
Bus Speed	6:00 A.M.-6:30 A.M.	60.795	3.804	-1.32392
Traffic Speed	6:30 A.M.-7:00 A.M.	53.508	11.038	0.443158
Bus Speed	6:30 A.M.-7:00 A.M.	51.016	6.587	-1.26841
Traffic Speed	7:00 A.M.-7:30 A.M.	47.185	14.69	0.239189
Bus Speed	7:00 A.M.-7:30 A.M.	40.064	11.448	-7.20E-02
Traffic Speed	7:30 A.M.-8:00 A.M.	42.54	17.2	0.328586
Bus Speed	7:30 A.M-8:00 A.M.	31.888	11.38	0.600598
Traffic Speed	8:00 A.M.-8:30 A.M	45.77	15.96	0.14047
Bus Speed	8:00 A.M.-8:30 A.M	38.1	14.29	0.352615
Traffic Speed	8:30 A.M.-9:00 A.M	49.786	14.345	-2.20E-01
Bus Speed	8:30 A.M.-9:00 A.M	40.4	15.62	6.69E-04
Traffic Speed	3:00 P.M.-3:30 P.M	52.475	12.199	0.103301
Bus Speed	3:00 P.M.-3:30 P.M	46.54	14.01	-8.00E-01
Traffic Speed	3:30 P.M.-4:00 P.M	46.97	16.22	-1.40E-01
Bus Speed	3:30 P.M.-4:00 P.M	34.97	14.41	8.75E-02
Traffic Speed	4:00 P.M.-4:30 P.M	45.76	17.21	-1.40E-01
Bus Speed	4:00 P.M.-4:30 P.M	30.76	14.06	0.716196
Traffic Speed	4:30 P.M.-5:00 P.M	45.81	17.07	-1.20E-01
Bus Speed	4:30 P.M.-5:00 P.M	31.06	14.27	0.416891
Traffic Speed	5:00 P.M.-5:30 P.M	46.81	26.49	6.22859
Bus Speed	5:00 P.M.-5:30 P.M	30.85	15.46	0.446162
Traffic Speed	5:30 P.M.-6:00 P.M	45.78	18.09	-3.80E-01
Bus Speed	5:30 P.M.-6:00 P.M	34.08	15.35	0.275262

The predictor variables that were examined include: bus speed as a continuous variable, number of on-ramps, number of off-ramps, total number of ramps, number of ramps per unit length of Freeway segment, and weather variables such as; wind speed, visibility and weather conditions.

The dependent variable, traffic speed, is estimated to provide travelers with a quantitative measure of traffic condition. Therefore, estimating it as a categorical variable was explored. For a qualitative or categorical dependent variable, methods such as logit or probit analysis are most appropriate.

4.2 Statistical Methods for Model Development

The generalized linear modeling provides a framework for examining a logit or logistic regression. To explore any non-linearity in the relationships, generalized additive models were also applied. This section provides an overview of generalized linear models and generalized additive models.

4.2.1 General Linear Model

Linear regression explains the relationship between traffic speed and bus speed based on a straight line fit to the data. The linear regression model postulates that $y = \alpha + \beta x + e$, where the residual e is a random variable with mean zero. The coefficients α and β may be determined by minimizing the sum of the square residuals.

The general linear model (GLM) uses the method of least squares to fit linear models. The GLM procedure can be used for statistical analyses such as: simple regression, multiple regression, analysis of variance (ANOVA), especially for unbalanced data, analysis of covariance, response-surface models, weighted regression, polynomial regression and multivariate analysis of variance (MANOVA).

The basic statistical assumption underlying the least-squares approach is that the observed values of each dependent variable can be written as the sum of three parts: a fixed component βx, which is a linear function of the independent coefficients, and a random error, e.

The least-squares approach provides estimates of the linear parameters that are unbiased and have minimum variance among linear estimators. Under the assumption that the errors have a normal (or Gaussian) distribution, the least-squares estimates are the maximum likelihood estimates and their distribution is known. All of the significance levels ("*p* values") and confidence limits calculated by the GLM procedure require this assumption of normality in order to be exactly valid, although they are good approximations in many other cases.

4.2.2 Generalized Additive Model

The Generalized Additive Model (GAM) procedure is a non-parametric regression technique. The non-parametric regression or local regression is used when the functional form is not known and thus cannot be parameterized in terms of any functions. The basic idea behind this kind of regression is predicting a data point, for example y, by fitting a parametric function in the neighborhood of the data point y.

Compared to General Linear Model, non-parametric regression enables us to explore the data more flexibly, uncovering structure in the data that might otherwise be missed. Let Y be a response variable and X_1, X_2, ... , X_p be a set of predictor variables. A regression procedure can be viewed as a method for estimating the expected value of Y given the values of X_1, X_2, ... , X_p. The standard linear regression model assumes a linear form for the conditional expectation

$$E\left(Y \mid X_1, X_2, ..., X_p\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...\beta_p x_p \tag{4.1}$$

Given a sample, estimates of β_0, β_1,…, β_p are obtained by the least-squares method.
The additive model generalizes the linear model by modeling the conditional expectation as:

$$E\left(Y \mid X_1, X_2, ..., X_p\right) = S_0 + S_1(X_1) + S_2(X_2) + ... + S_p(X_p) \tag{4.2}$$

where $S_i(X)$, $i = 1,2, ... , p$ are smooth functions.

Generalized Additive Models consist of a random component, an additive component, and a link function relating the two components. The response Y, the random component, is assumed to have exponential family density, which can be represented as follows,

$$f_Y\left(y;\theta;\phi\right) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right] \tag{4.3}$$

where θ is called the natural parameter and φ is the scale parameter. The mean of the response variable μ is related to the set of covariates X_1, X_2, ... , X_p by a link function g. The quantity

$$\eta = s_0 + \sum_{i=1}^{p} s_i(X_i) \tag{4.4}$$

where $s_1(\cdot)$, ... , $s_p(\cdot)$ are smooth functions defines the additive component, and the relationship between μ and η is defined by g(μ)= η. The most commonly used link function is the canonical link, for which η=θ.

4.2.3 Generalized Linear Model: Logistic Regression

For an ordinal or nominal dependent or response variable, logistic regression may be applied using the generalized linear modeling framework with a logit link function. Thus, logistic regression predicts the probability of the response variable being in a particular category. Generally, polytomous logistic regression models cumulative logits of the response variable with proportional odds model. A logit or the log odds ratio is ln(p/(1-p)), where p is the probability that the event y occurs and (1-p) is the probability that the event y does not occur. The proportional odds model assumes that cumulative logits of response variable are parallel linear functions of independent variables. In this case the model assumes a linear relationship for all of the logits. If the predicted probabilities of the traffic speed in three categories are p1, p2 and p3 then the equation for the parallel regression lines can be written as,

$$\log\left(\frac{p_1}{1-p_1}\right) = \alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots \beta_n x_n \tag{4.5}$$

and

$$\log\left(\frac{p_1+p_2}{p_3}\right) = \alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots \beta_n x_n \tag{4.6}$$

From the equations it can be clearly seen that except for the intercept, the rest of the model parameters are the same for all the categories and the odds, $p_1/(1-p_1)$ and $p_1 + p_2/(p_3)$ are proportional. Logistic regression restricts the probabilities of the response variable to lie between *0* and *1*. The probability p_i is estimated by the logit link function, where $p = 1 / [1 + exp (\alpha + \beta x)]$, that is as $\alpha + \beta x$ decreases, p approaches *1* and as $\alpha + \beta x$ increases, p approaches zero.

For the logit model, $\log\left(\frac{p}{1-p}\right) = \alpha + \beta x + e$, the Maximum Likelihood Estimation (MLE) method is used to estimate the logit coefficients of the model parameters. The least-square method for linear regression reduces the sum of the squared distance of the data points from the regression line whereas the MLE maximizes the log likelihood (LL), which infers how likely it is to predict the observed response variable from the observed predictor variable. MLE follows an iterative algorithm assuming an initial value of logit coefficients to maximize the LL. This algorithm predicts the magnitude of the logit coefficients until a maximum value of log likelihood is reached. Following the initial estimation of logit coefficients, the residuals are tested and the function is improved to meet the convergence criteria. The logit coefficients are re-estimated until the convergence criterion is met.

Based on the procedures outlined in this section, the traffic speed model was developed and is described in the next chapter of this report.

5.0 MODEL DEVELOPMENT

The previous chapter provides an overview of the modeling framework that was applied to develop the traffic speed estimation model. This chapter presents a summary of the model development procedure, the statistical tests performed to examine the goodness-of-fit of the models evaluated, and finally the operational performance of the model. The model was developed based on data collected for a freeway section of Interstate 25, and tested on the same network based on an independent test data set. In addition, the model was also tested based on data from another freeway section of Interstate 225. The overall model developed by logistic regression provided a good fit at the p-level less than 0.0001. To further analyze the reliability of the model it is necessary to evaluate the goodness-of-fit statistic. This chapter includes a detailed analysis of the goodness-of-fit statistic and also operational tests of the model for the Interstate 25 and Interstate 225 freeway sections.

5.1 Model Building Procedure

The speed model was developed, using the statistical software package SAS, and involved a step-by-step process. This section summarizes the process.

Initially, a General Linear Model (GLM) was developed based on data from the I-25 freeway network. The speed estimates from the GLM correctly classified 70% of the observations. To improve the model further, a Generalized Additive Modeling technique was applied non-parametrically to determine the appropriate non-linear functional forms of the independent variables. The results showed that for bus speed less than 20 mph, a non-linear transformation of bus speed and for bus speed greater than 20 mph a linear model provided a statistically significant estimate of traffic speed.

The GAM results showed that the distribution between bus speed and traffic speed is not completely linear. Therefore, a bus speed of 20 mph was considered to be a breakpoint and a piecewise generalized linear regression procedure was taken into account to further improve the model. The predictions from this procedure gave an 83 % correct classification. Since there are some drawbacks in considering non-linear terms to fit a linear model, a more robust statistic technique was considered.

Logistic regression is a type of statistical technique that uses maximum likelihood estimation technique, a robust variance estimator, to predict more accurate variance estimates and confidence intervals for this problematic case of a misspecified model. The logistic regression technique correctly classified 96% of the data, and therefore was the statistical technique used to compute the estimates for the final model.

The statistical procedures to develop the model described above are explained in detail further.

5.1.1 General Linear Model

Initially, a general linear model (GLM) was developed based on data from the I-25 freeway network. The dependent variable for this linear regression procedure was traffic speed and the independent variables were bus speed, number of on-ramps and off-ramps within a freeway segment, wind speed, visibility and weather condition. An interaction variable, bus speed and number of ramps per unit length, was also considered in the model development. All the variables included in this linear model were continuous variables. A step-wise linear regression procedure was used to choose the appropriate independent variable. The model obtained by the GLM procedure included the independent variables bus speed and the interaction variable. The R^2 value of 0.79 for the final model, indicates that 79% of the variability in the model was accounted for, by the dependant variables. All the parameters in the model were statistically significant at 0.05 alpha level. The overall percent correct classification was found to be 82.30%. It was also observed that the correct classification of lower traffic speed was lower. This observation prompted an exploration of the appropriate functional transformation of the independent variables to further improve the model performance. Therefore, a statistical procedure, Generalized Additive Model, described in Chapter 4 was applied and is presented in the next section.

5.1.2 Generalized Additive Model

The Generalized Additive Model (GAM) procedure was used to explore the relationship between bus speed and the interaction variable (bus speed and number of ramps per unit length of a freeway segment) on the response variable, traffic speed, in a non-parametrical way. A non-linear relationship between bus speed less than 20 mph and traffic speed, and a linear relationship between bus speed greater than 20 mph and traffic speed was observed. A similar relationship was observed for the interaction variable. Figure 5.1, Figure 5.2, and Figure 5.3 portray the distribution between the bus speed and the traffic speed. Figure 5.1 illustrates both the non-linear and linear distribution between the two variables. Figure 5.2 shows a non-linear nature of the relationship for bus speed less than 20 mph and the corresponding traffic speed. Figure 5.3 shows a linear distribution of bus speed greater than 20 mph and the corresponding traffic speed. The results from the GAM procedure lead to the choice of piecewise generalized linear regression for further development of the model.

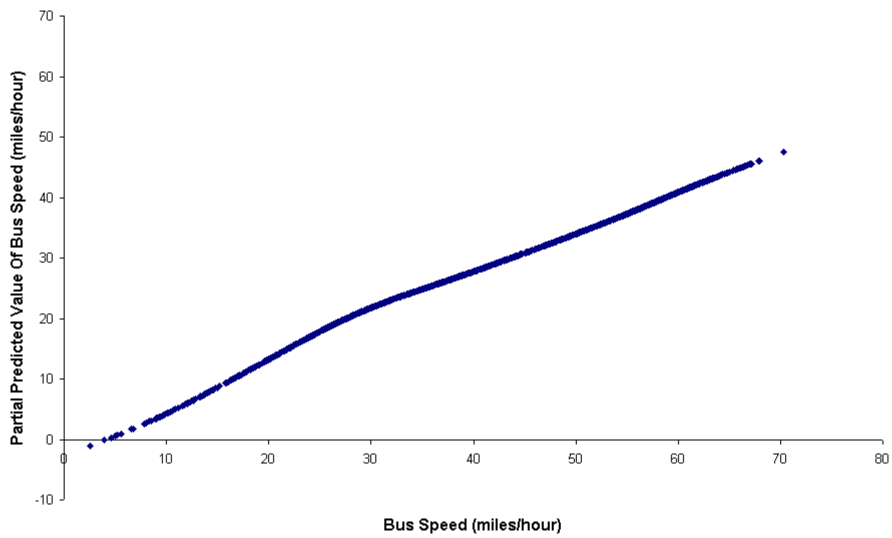


Figure 5.1. GAM for Total Data.

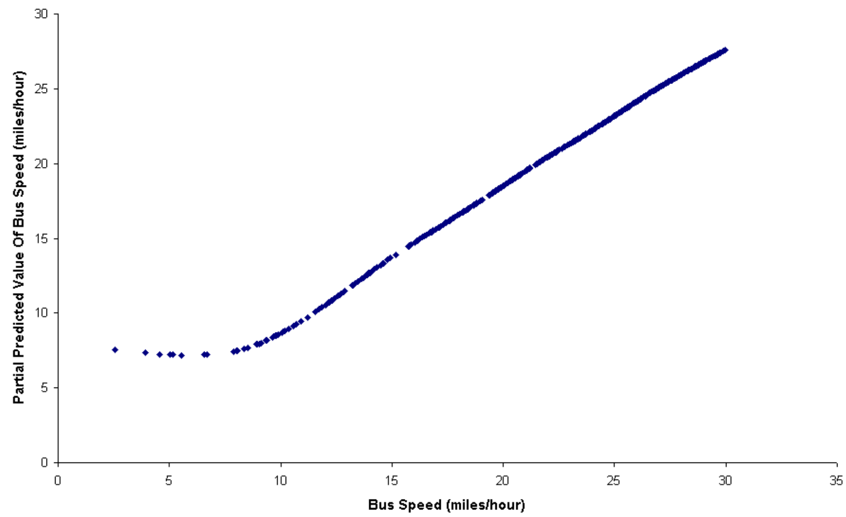


Figure 5.2. GAM for Bus Speed < 20 mph.

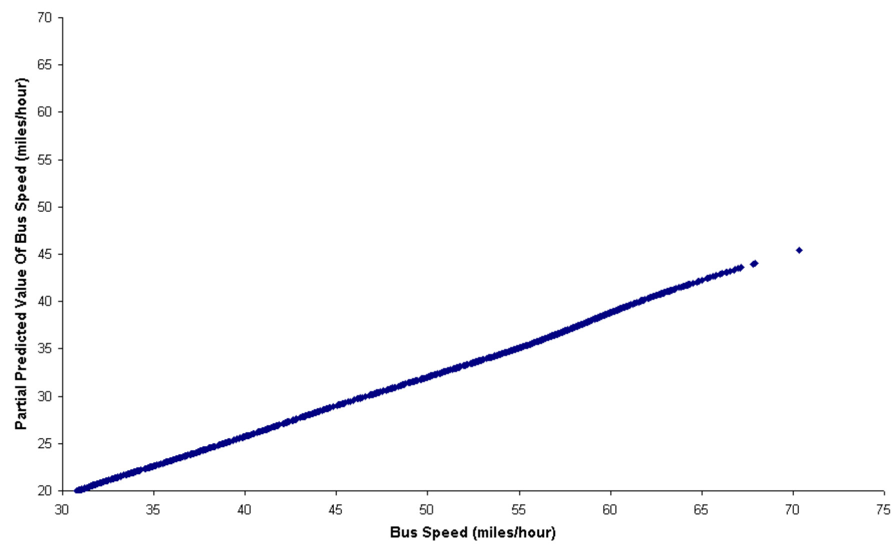


Figure 5.3. GAM for Bus Speed > 20 mph.

The theory behind the piecewise regression is that the independent variable, bus speed, specified as two independent variables based on a break point of 20 mph, behaves differently for bus speed less than and greater than 20 mph. The independent variables for this model were bus speed and the interaction variable (bus speed and the number of ramps per unit length of a freeway segment) and the dependent variable was traffic speed. The R^2 obtained from this model was 0.83; explaining 83% of the variability in the model is contributed by the independent variable. The overall percent correct classification was found to be 83.1% for the model. Compared to the previous GLM model, this model showed slight improvement. However, since bus speed less than 20 mph has a non-linear behavior, the linear least-square method is not appropriate for the model development. There are also other disadvantages of a linear least- squares procedure such as limitations in the shapes that linear models can assume over long ranges, poor extrapolation properties and sensitivity to outliers. Linear models with non-linear parameters like bus speed and the interaction variable often give optimal estimates of the unknown parameters. Outliers may skew the results of a least-square analysis. Thus the model validation with outliers becomes critical in obtaining sound answers to the questions motivating the construction of the model. There is always a possibility that there may be problems due to hetereoscedasticity. This problem is eliminated by the logistic regression procedure, which adopts a robust methodology to predict the outcome.

5.1.3 Logistic Regression

The logistic regression technique was applied to developing the relationship between traffic speed and bus speed. The response variable, traffic speed, was specified as three categories, 0-20 mph, 20-40 mph, and >40 mph. Since the response variable (traffic speed) has more than two categories, Bernoulli distribution would be inappropriate. Therefore, the polytomous logistic regression procedure was adopted for the model development.

A logistic regression model, with four independent variables was selected following a model development effort. The four independent variables used in this model are: bus speed less than 20 mph, bus speed greater than 20 mph, an interaction variable1 defined as the product of bus speed less than 20 mph and the number of ramps per unit length of a freeway segment, and a second interaction variable2 defined as the product of bus speed greater than 20 mph and the number of ramps per unit length of a freeway segment. Table 5.1 presents the coefficient of the transformed parameters and the corresponding statistics estimate by the MLE method. As the cumulative logit model predicts cumulative probabilities, therefore only two intercepts are estimated. The model was specified to predict the higher category first. Thus the Intercept 3 corresponding to the third category (> 60 mph) of traffic speed is predicted. Similarly the Intercept 2 corresponding to the second category (40 mph - 60 mph) and third category of traffic speed (< 40 mph) were predicted.

Table 5.1. Parameters and Estimates for the Model Based on the Calibration Data Set.

Dependent Variable = Traffic Speed			
Variable	Coefficient	Wald Statistic	Significance Level
Intercept 3	-13.1300	135.5826	<0.0001
Intercept 2	-4.8024	33.1882	<0.0001
X_1 = (Log(Bus Speed < 20))^6	-0.6592	61.4540	<0.0001
X_2 = (Log(Bus Speed > 20))^2	0.7979	90.3403	<0.0001
Interaction 1= X_1*Number of ramps per unit length	7.7300	125.1063	<0.0001
Interaction 2= X_2*number of ramps per unit length	-0.00065	21.7170	<0.0001
Model Chi Square = 32.3109			

[Where, interaction 1 = Number of ramps per unit length of a freeway segment corresponding to Bus Speed < 20 mph, interaction 2 = Number of ramps per unit length of a freeway segment corresponding to Bus Speed > 20 mph]

As seen from the GAM results, the bus speed less than 20 mph includes a spline fit. Therefore, the bus speed less than 20 mph is transformed to the 6th power of logarithmic function, and bus speed greater than 20 mph includes a linear fit, thus the bus speed was transformed to a 2nd power of logarithmic function. Similarly the interaction variable at the break point, 20 mph was transformed to a logarithmic function and 2nd power function respectively. The dependent variable, traffic speed, is discrete in nature with three categories, speed ranging from < 20 mph, 20 to 40 mph, and greater than 40 mph. The probabilities of the higher categories of traffic speed (2nd and 3rd category) were predicted and thus the corresponding intercepts.

Table 5.1 shows that all the parameter coefficients are statistically significant. The *–2 log likelihood* statistic estimated by the model is 266.63, which is significant at the 0.05 alpha level. This means that the null hypothesis that all the explanatory variables are equal to zero can be rejected. Another statistic, the *Wald statistic* estimated as [estimate / standard error] ^{2}, is chi square distributed with one degree of freedom, which is significant for all the coefficients at 0.05 alpha level. The logistic coefficient obtained by the model can be interpreted as the change in the dependent variable, logit (Traffic Speed), associated with unit change in the independent variable. Since the predictions are not a linear function of the independent variable, bus speed and the interaction variable, the slope of the curve depends on the value of the independent variables. Figure 5.4 and Figure 5.5 show the plot of the logits verses the bus speed (< 20 mph), and bus speed (> 20 mph) respectively. While the curve for the bus speed greater than 20 mph is linear, the relationship between the logit and the bus speed less than 20 mph is clearly non linear.

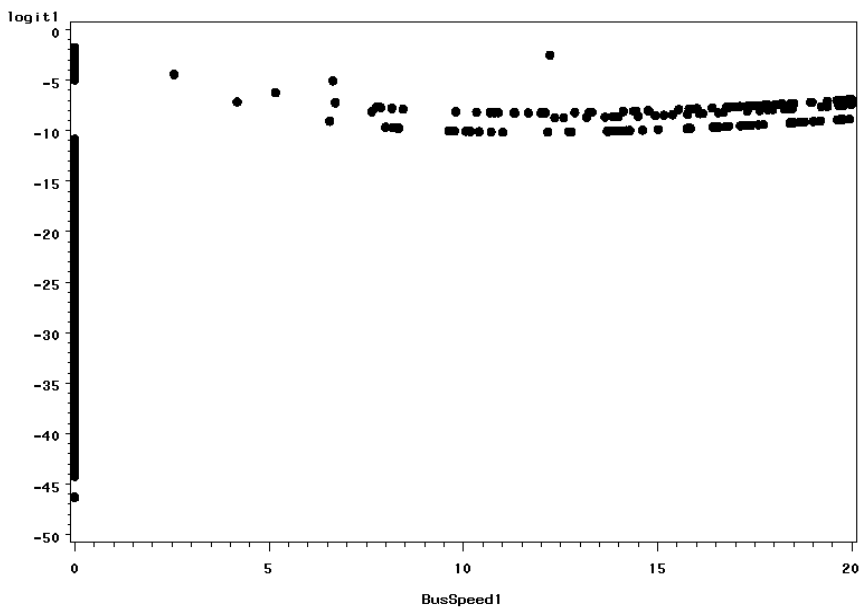


Figure 5.4. Plot of Logit Verses Bus Speed < 20 mph.

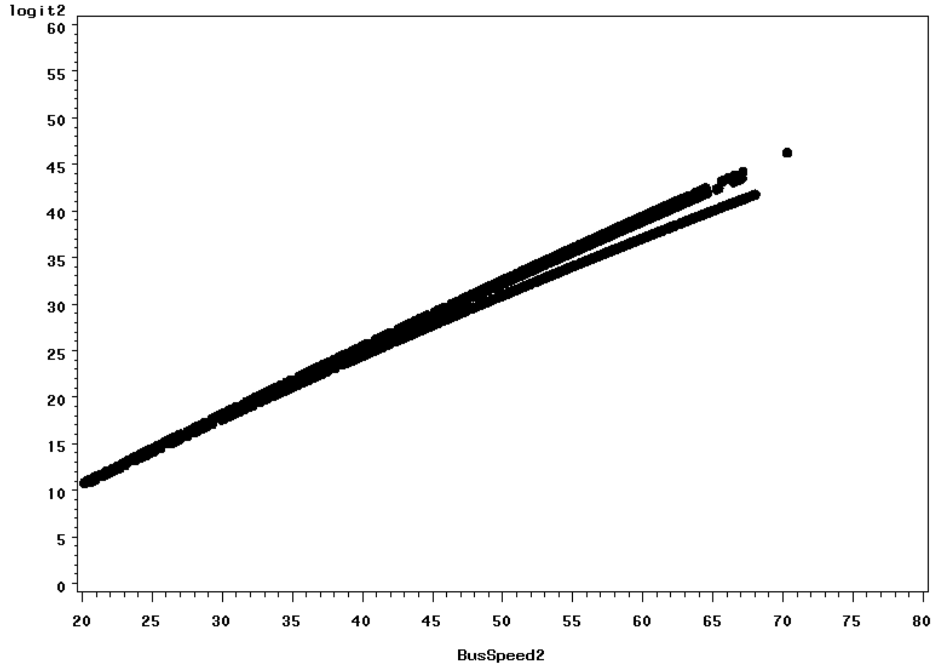


Figure 5.5. Plot of Logit Verses Bus Speed > 20 mph.

Since the traffic speed has three categories, the equations for the cumulative logit model can be written as:

$$\log\left(\frac{p_3}{(1-p_3)}\right) = (-13.13) + (-06592) * \left(\log(busspeed < 20mph)\right)^6 + (0.7979) * \left(\log((busspeed > 20mph))\right)^2 + (7.73) * interaction1 + (-0.00065) * interaction2 \qquad (1.1)$$

$$log\left(\frac{p_2 + p_3}{p_3)}\right) = (-4.8024) + (-06592) * \left(\log(busspeed < 20mph)\right)^6 + (0.7979) * \left(\log((busspeed > 20mph))\right)^2 + (7.73) * interaction1 + (-0.00065) * interaction2 \qquad (1.2)$$

Where p_1, p_2, p_3 are the probabilities of the traffic speed being in the respective categories and *interaction 1* is the product of the number of ramps per unit length of freeway segment and the transformation of bus speed < 20 mph and *interaction 2* is the product of the number of ramps per unit length of a freeway segment and a transformation of the bus speed > 20 mph. From these equations the probability of all three categories of the traffic speed can be computed as,

$$p_1 = 1 - (p_3 + p_2) \tag{1.3}$$

$$Let \;\; \alpha = \exp((-13.13) + (-06592) * (\log(busspeed<20mph))^6 + (0.7979) * (\log((busspeed>20mph)))^2 + (7.73) * interaction1 + (-0.00065) * interaction2 \;\;) \tag{1.4}$$

$$Let \beta = \exp((-4.8024) + (-06592) * (\log(busspeed<20mph))^6 + (0.7979) * (\log((busspeed>20mph)))^2 + (7.73) * interaction1 + (-0.00065) * interaction2 \;\;) \tag{1.5}$$

From equation (1.4) and (1.5);

$$p_3 = \frac{\alpha}{(1+\alpha)} \tag{1.6}$$

$$p_2 = \frac{\alpha * (\beta - 1)}{(1+\alpha)} \tag{1.7}$$

The logistic model has a good fit at 0.0001 p-level. Since the sample size for this model is large and there are many independent variables, the p value is usually smaller than 0.05. Several goodness-of-fit statistics, model chi-square, deviance statistic, and R^2 and operational measures such as percent correct predictions may be used for model evaluation. The next section provides a detailed description of goodness-of-fit statistics.

5.2 Calibrated Model

The stepwise logistic regression used in the development of the model predicted the traffic speed using four independent variables. At each stage of model development, the goodness-of-fit measures were compared. After the final model was developed, the model's goodness-of-fit measure was evaluated in predicting the dependent variable. The performance of the model was evaluated based on the model chi-square value of 32.31, significant at the 0.0001 p-level. The proportional odds assumption (POM) for the model is not fulfilled. However, the cumulative logit model may not be rejected. For larger samples, usually the POM does not hold well.

Therefore they may be ignored as there are other statistics that can be used to judge the fitness of the model [16]. Table5.2. displays the model chi-square and the log likelihood statistic that were also considered in the evaluation of the model.

Table 5.2. Logistic Regression Results for the Calibration Data Set.

-2 Log Likelihood = 489.910			
	Chi Square	**DF**	**Significance**
Model	32.3109	4	<0.0001
Re-Scaled R^2 = 0.94			

The deviance is considered to be one goodness-of-fit statistic. It compares the fitted model with the saturated model. The deviance statistic tests whether this difference is obtained by chance. In this case the deviance was 489.91 at 1.00 p-level, which indicates a very good fit. The R^2 obtained in this model was 0.94, showing that 94% of the variance of traffic speed was accounted by the logistic regression equation and thus this value is substantive enough to consider the equation significant. All the predicted coefficients of the parameters were also found to be significant at 0.05 alpha level.

Another performance measure of the model is based on its operational performance. This examines the correct classification rate or the classification table. Table 5.3 shows the classification table of the calibration data set. The overall classification is 95.36%. The percent correct classification for the category 1 (< 20 mph), category 2 (20-40 mph) and category 3 (> 40 mph) of traffic speed was 76%, 95% and 96% respectively.

Table 5.3. Percent Correct Classifications for the Calibration Set.

Observed	**Traffic Speed Categories (mph)**	**Predicted**			**% Correct**
		1	2	3	
1	<20	93	30	0	76
2	20-40	0	655	32	95
3	>40	0	47	1496	97

5.3 Testing the Model

Since the model performed well on the calibration data set it was tested on a test data set for the same freeway section, and also for a different freeway section. The two freeway sections used for testing the speed model were Interstate 25 (for different days) and Interstate 225. Data for Interstate 25 were collected for 14 weeks, e.g., 4 weeks in the months of April and May, and 10 weeks from September to December 2001. For test purposes, 3 weeks from the November data and 1 week each from the May and December data were used. The rest of the data were used previously for calibrating the model. Another freeway section, Interstate 225, was also considered to test the model. The data for Interstate 225 were collected for 5 weeks on an 11-mile section of roadway.

On the test section of the Interstate 225 there are detectors located at Tamarac Parkway, Iliff Avenue, Mississippi Avenue and the 6th Avenue on-ramps. Traffic data were collected for both the northbound and southbound direction. However, only the northbound data were used for the prediction of traffic speed as most of the detectors on the southbound direction were malfunctioning. The basic freeway geometry such as the number of lanes and the position of ramps are similar to Interstate 25. The spacing between on-ramps and the total length of the highway considered for this study is summarized in Table 2.2. Figure 2.2 illustrates the schematic of Interstate 225. Traffic data utilized for the prediction of traffic speed were collected for a five-week period.

The results of the test set for the I-25 freeway is presented in Table 5.4. The overall correct classification was 96%. Categories 1, 2 and 3 were correctly classified at 83%, 95%, and 97% respectively. A similar test was conducted for the I-225 freeway section.

Table 5.4. Percent Correct Classifications for the I-25 Freeway Test Data Set.

Observed	Traffic Speed Categories (mph)	Predicted			% Correct
		1	2	3	
1	<20	38	8	0	83
2	20-40	0	346	20	95
3	>40	0	28	913	97

Data from the I-225 freeway were also used to test the probability equation obtained from the logistic model. Table 5.4 summarizes the percent correct classification for the predicted traffic speed, obtained from bus speeds. As can be seen in the Table, category 3 (> 40 mph) was predicted 91% correctly. There were no speed categories less than 40 mph due to the absence of detector data on this freeway during the data collection period.

Table 5.5. Percent Correct Classification for the I-225 Freeway Test Data Set .

Observed	**Traffic Speed Categories (mph)**	**Predicted**			**% Correct**
		1	2	3	
3	>40	0	17	168	91
Overall % Correct for the Calibration set and Test set = 96 %					

The results of the model illustrate that the speed model proposed performs well in correctly classifying traffic speed based on bus speed estimated from the AVL data from buses and freeway geometry data. It may be mentioned that only limited testing could be performed due to less availability of data from a second test section. During the proposal development stage, it was envisioned that fixed detector data would be available for two test networks – the 6th Avenue and the I-225 networks. However, during the course of this research project, fixed detector data were not available for the 6th Avenue network to allow further testing of the proposed model.

5.4 System to Develop Speed Maps Based on Bus AVL Data

The main objective of this project was to examine the feasibility of estimating traffic speed based on data collected by the RTD's Bus AVL system to develop a model and report on its performance.

In an effort to address these objectives, data preprocessing procedures, tools and methods were developed. In this section, a system is proposed that would allow a traffic management center to use these tools to estimate and display traffic speed for freeway segments. It may be mentioned that this algorithm may not be used where buses are not available, or where buses travel in HOV lanes. In addition, as outlined in the proposal, the tools/computer programs developed as part of this project were designed for off-line testing of these concepts. Only off-line data (not real-time data) were provided by RTD and on-line data (real-time data) from RTD were not available as part of this project.

Each bus report generated by the AVL system may be entered into a database (e.g. Oracle or Access) to create live data tables. Individual bus reports may be paired to determine when and where the last probe report arrived for this bus. Based on the cases outlined in Table 3.1, for a pair of bus reports and its corresponding case number, the relevant segment (S_m) and time interval (T_n) may be determined. This process is repeated for all buses in the database for the speed update time interval (e.g. 15 minutes). At the end of the interval, a weighted average bus speed may be estimated based on Eq. 3.5 for segment S_m for interval T_n. Based on the bus speed estimate, the speed estimate algorithm may be applied. This model outputs the probability of speed in any given speed category.

Finally, the speed category for each segment may be estimated for all segments for time interval T_n. This system is summarized in a simple flow diagram shown in Figure 5.6. The Figure also shows the two programs that were developed: a data processing program and the speed estimation model. These programs may be implemented in any GIS based graphical display tool such as ArcIMS to develop freeway speed maps.

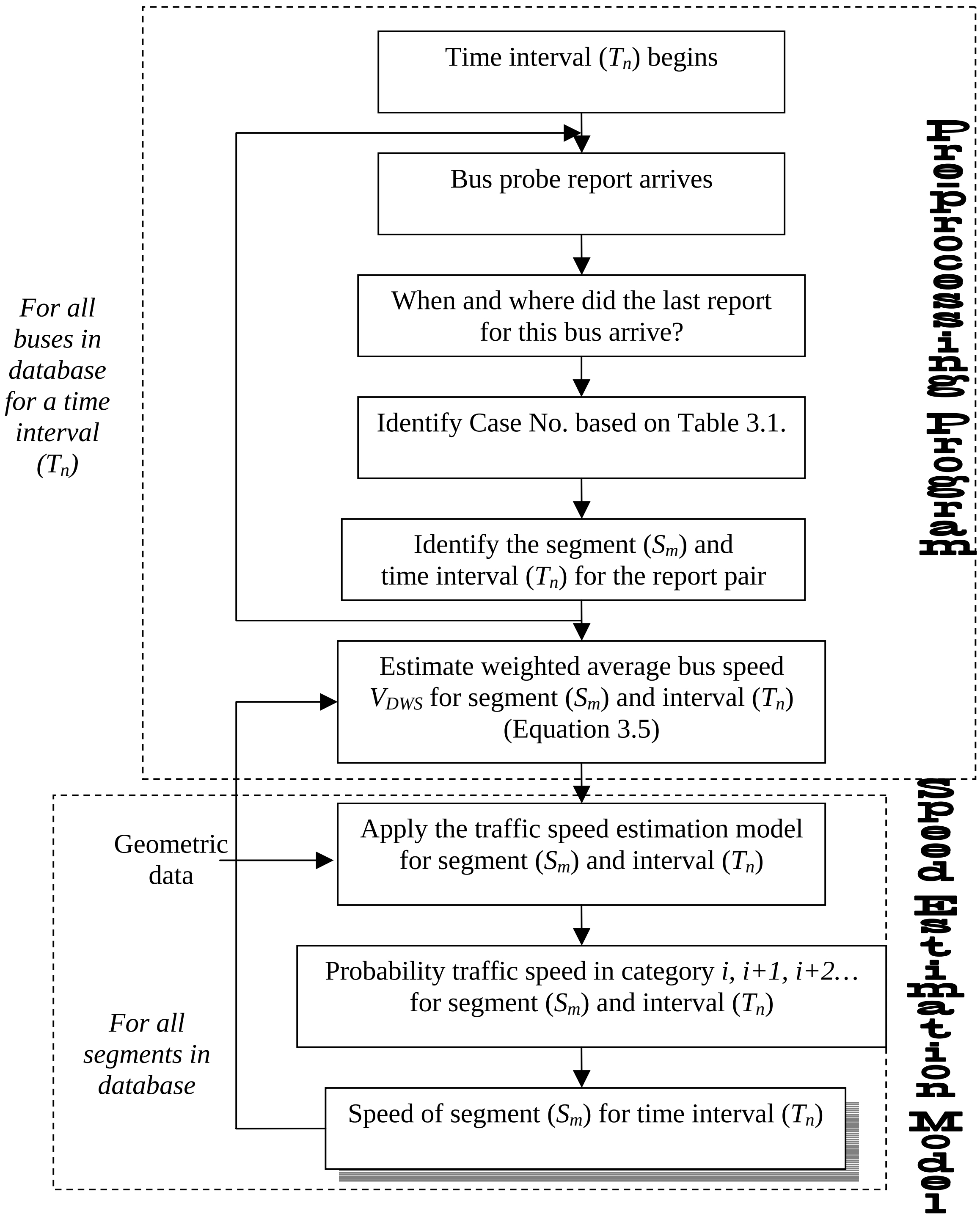


Figure 5.6. System to Develop Traffic Speed Estimates for Freeway Segments Based on Bus AVL Data

6.0 CONCLUSIONS AND RECOMMENDATIONS

The objective of this research was to develop an algorithm to estimate real-time traffic speed for freeway sections, as part of a traveler information system. The Colorado Department of Transportation (CDOT) currently monitors traffic conditions on limited sections of freeway based on fixed sensors installed on the mainline for ramp metering purposes. Based on these sensors, CDOT currently provides traveler information on its web site at www.cotrip.org and is extending its freeway surveillance coverage to provide traffic information to travelers. This research was funded to investigate the feasibility of using current infrastructure, Automatic Vehicle Location (AVL) system, in transit vehicles operated by the Regional Transportation District (RTD), to estimate traffic speed.

The plan is to provide travelers traffic speed information, updated every few minutes, for freeway sections. Typically, this type of reporting is based on the data collected by fixed sensors such as detectors, video cameras and other sensors located on the freeway. However, CDOT is unable to report traffic information for freeway sections without such infrastructure in place. On the other hand, buses traversing these same sections are equipped with GPS receivers, the data from which can be utilized in estimating traffic speed. Several factors including, but not limited to, weather and freeway geometry may affect bus and traffic speed.

As part of this project, a statistical model to estimate traffic speed from bus speed, geometric characteristics of freeway and weather conditions was developed. The model was developed and tested based on data collected for a 13-mile section of the Interstate 25 (I-25) freeway. The model's performance was further examined based on data collected for an 11-mile section of the Interstate 225 (I-225) freeway. Least squares method, non-parametric regression and maximum likelihood estimation method were used for model development. After a series of model evaluations, logistic regression with proportional odds model was selected for the model development. The final model presented is statistically significant with an overall correct classification of 96% for both the calibration and the test set.

This research project has demonstrated the feasibility of using bus location data to estimate traffic speed at regular intervals for freeway sections. The model developed to estimate traffic speed performs well; however, it is conditional on the availability of bus reports from the AVL system. A non-linear model was developed to estimate speed, which exceeded expected performance. Test results show that the model performed equally well on both sections of freeway. The analysis indicates that the model would operate at an acceptable level for any freeway section with similar infrastructure. The model is statistically significant at 0.05 alpha level, according to the model chi square statistic. The model predicts 94% of the responses correctly. The results also indicate that the percent correct classification of the low (0-20 mph) traffic speed is lower compared to the other categories, both for the calibration and the test set. Most importantly, the performance of the model depends on the number of bus reports available. The bus reports were available for 80% of the 15-minute time intervals based on the current AVL reporting interval of two minutes.

Further testing is recommended for lower traffic speeds on the freeway. The model illustrates that for low average speed of traffic, bus speed is higher as the buses tend to travel in the left lane of the freeway. The number of ramps in the model is a very significant variable and influences the traffic speed. There are other geometric features such as; number of lanes, gradient and super elevation, that were not considered in the model as they were not prominent on the freeway section used in this project. As these geometric characteristics of freeway may also affect traffic speed, future testing on additional freeway segments including more variation in these characteristics is recommended.

This research provides the methods and tools required to process RTD’s AVL bus data to estimate traffic speed. Given that the data processing methods and the speed estimation algorithm has been developed, this model may now be implemented in the CDOT Transportation Management Center (CTMC) to estimate traffic speed. The TOC is currently developing a GIS-based freeway speed map. The algorithm developed here may be used to develop a freeway speed map as envisioned in the project proposal.

To significantly improve the availability of probe reports and to improve the performance of the model, it is suggested that CDOT and RTD consider exploring several options that include more frequent reporting of bus locations, and improving the communication links to allow the data to be managed and processed for use by both agencies. Current RTD operations for fleet management do not require more frequent reporting of bus location by the AVL system. However, joint efforts are expected to yield unprecedented benefits to both agencies.

As CDOT strives to provide real-time, accurate, reliable traveler information for major corridors, this research has shown that cooperation between agencies may allow them to leverage infrastructure investment dollars and develop strong partnerships of mutual benefit to serve both the traveling motorists and transit riders simultaneously.

REFERENCES

1. Navarro, M.E. and N.M. Rouphail, Analysis of Alternative Service Measures for Freeway Facilities, in Transportation Research Circular E-C018, Fourth International Symposium on Highway Capacity. 2000. p. pp. 73-83.
2. Dods, J.S., Traffic Incident Detection using a Selective Video Technique, in Proceedings of the Fifteenth Australian Road Research Board Conference. 1990: Darwin. p. 1-15.
3. Khan, S. and K. Thanasupsin, Estimating Link Travel time on I-70 Corridor: A Real-Time Demonstration Prototype. 2000, Colorado Department of Transportation.
4. Rakha, H. and M. Van Aerde, Accuracy of vehicle-probe estimates of link-travel time and instantaneous speed, in The proceedings of the 1995 annual meeting of ITS America. 1995.
5. Doughlas, J.H. and M.T. Shawn, *Probe vehicle sample sizes for real-time information: The Houston experience*. The proceedings of the 1996 annual meeting of ITS America, 1996. **Volume 1**.
6. Blake, K.N., C.L. Dudek, and C.E. Mountain, Using probe-measured travel times to detect major freeway incidents in Houston, Texas, in Transportation Research Record. 1996. p. pp. 213-220.
7. Sen, A., et al., *Frequency of probe reports and variance of travel time estimates*. Journal of Transportation Engineering, 1997(July/August): p. pp. 290-297.
8. Hellinga, B. and L. Fu, *Assesing expected accuracy of probe vehicle travel time reports*. Journal of Transportation Engineering, 1999(November/ December): p. 524-530.
9. Castle Rock Consultants, C., Denver regional transportation district automatic vehicle location system, in Evaluation Final Report Prepared for the Federal Transit Administration. 1998.
10. Gerlough, D.L. and M.J. Huber, *Traffic Flow Theory*. Transportation Research Board, 1975.
11. Gerlough, D.L. and M.J. Huber, *Traffic Flow Theory*. Transportation Research Board, 1975: p. 199-201.
12. May, D.A., *Traffic Flow Fundamentals*. 1990, Englewood Cliffs, New Jersey: Prentice Hall.
13. Zhou, M. and F.L. Hall, *Investigation of Speed-Flow Relationship under Congested*

Conditions on a Freeway. Transportation Research Record, 1999. **1678**: p. pp. 64-72.

14. http://www-cta.ornl.gov/cta/research/trb/tft.html.

15. *Highway Capacity Manual*. Transportation Research Board
National Research Council. 2000: Washington D.C.

16. Rudolf, J.F. and C.L. Ramon, *SAS System for Regression*. 1995, Cary, NC: SAS Institute Inc.

APPENDIX A

MOBILE SENSORS

This appendix provides information related to the operations of the mobile sensors used as part of the automatic vehicle location system (AVL). The need for the AVL system for RTD was determined in 1989 when RTD conducted a cost benefit study regarding improved communication system alternatives. The study showed that RTD required an increase in radio channels, greater data transmission and the latest in Computer Assisted Dispatching (CAD) and AVL technology. The main purpose of the AVL system is; to improve the ability of dispatchers to adjust on-street operations, to provide accurate and real time information to the riders, to increase safety and to develop efficient schedules.

The pictures in this appendix show RTD's Operation Center (OC). Dispatch consoles and computer workstations are located in the OC, and fleet activity is controlled from there. The CAD feature updates the dispatcher on all the activities on the street, and notifies the dispatcher to take action to rectify any anomalies. Monitors in the OC receive the location information of each vehicle. The vehicle (bus) is displayed on a map of the Denver area. The vehicle location is updated every two minutes under normal conditions. Other data such as schedule and incidents are reported on a separate monitor.

Each RTD vehicle is equipped with an AVL package, which consists of a mobile radio, onboard processor, driver interface and GPS antenna. The vehicle odometer is connected to the onboard processor. The following pictures show the OC the dispatcher's monitor, the display of the RTD vehicle and the AVL package in the RTD vehicle.

Figure A-1 Mobile Data Terminal

Figure A-2 Transit Control Head

Figure A-3 Dispatcher's Console

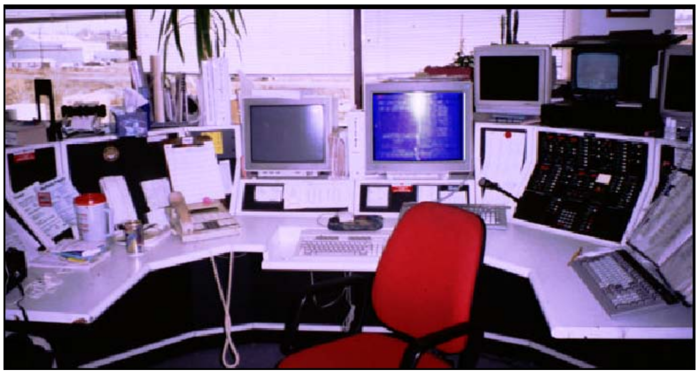


Figure A-4 CAD, AVL and Legacy VAX Equipment at a Dispatcher's Console

Figure A-5 Automatic Vehicle Location Screen

Figure A-6 Computer Aided Dispatch Screen

Figure A-7 CAD Screen Vehicle Identification Menu

Figure A-8 CAD Screen Incident Queue

Figure A-9 New Dispatcher's Console

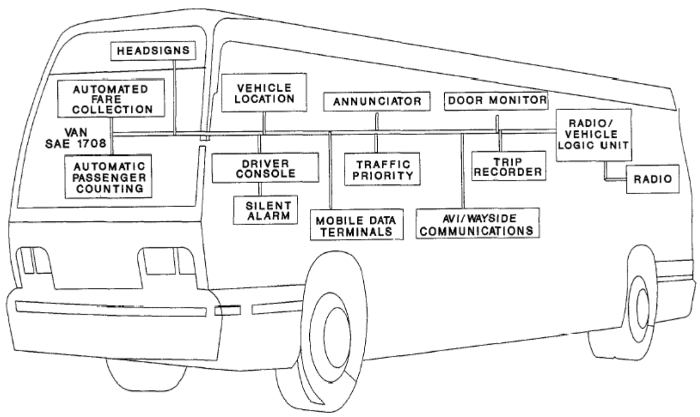


Figure A-10 AVL System Components in RTD Bus

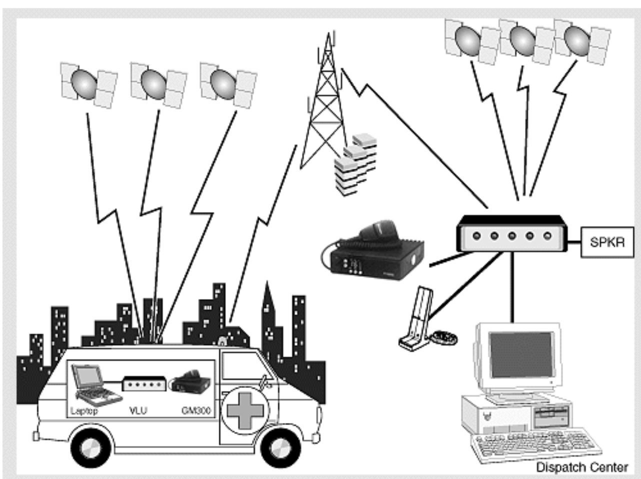


Figure A-11 Denver's AVL System

Figure A-12 RTD Bus Displaying Intelligent Vehicle Login Unit ID(1802)

APPENDIX B

STUDY AREA

This appendix contains pictures of the Interstate 25 study area of this research project. The study area is 13 miles long and consists of 3 to 5 lanes. No HOV lanes were available at the time that this research study was conducted. It includes 12 on-ramps at: (1) E. County Line Road, (2) E. Dry Creek Road, (3) WB Arapahoe Road, (4) EB Arapahoe Road, (5) E. Orchard Road, (6) E. Belleview Avenue, (7) I-225 Interchange, (8) E. Hampden Avenue, (9) E. Yale Avenue, (10) E. Evans Avenue, (11) SB Colorado Boulevard and off-ramps at: (1) E. Dry Creek Road, (2) E. Arapahoe Road, (3) E. Orchard Road, (4) E. Belleview Avenue, (5) I-225 Interchange, (6) E. Hampden Avenue, (7) E. Yale Avenue, (8) E. Evans Avenue and (9) SB Colorado Boulevard. All the pictures show the highway before the commencement of the TREX project.

Figure B-1 Northbound Interstate 25 approaching Orchard Road (Exit 198)

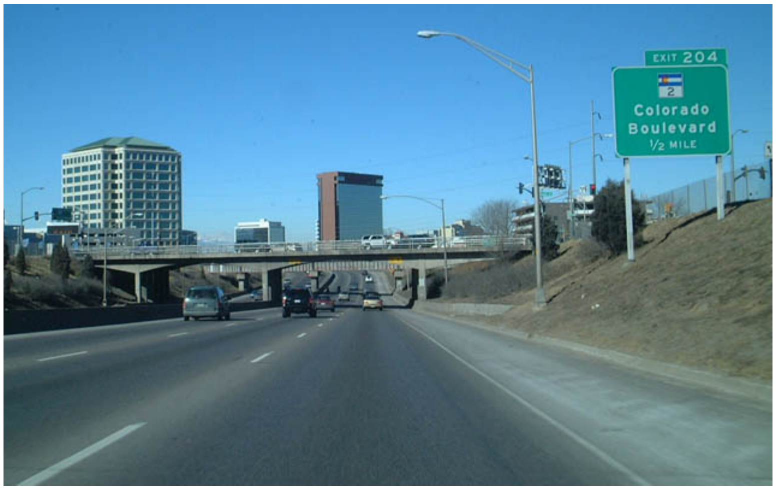


Figure B-2 Northbound Interstate 25 approaching Junction Colorado 2 (Colorado Boulevard), Exit 204

Figure B-3 Northbound Interstate 25 approaching Dry Creek Road (Exit 196)

Figure B-4 Northbound Interstate 25 at Orchard Road (Exit 198) approaching Colorado 88 (Belleview Avenue, Exit 199).

Figure B-5 Northbound Interstate 25 showing mileage sign providing the distance to Belleview Avenue (Colorado 88), Interstate 25, and Hampden Avenue (U.S. 285 and Colorado 30)

APPENDIX C

CODE WRITTEN IN VISUAL BASIC ACCESS TO PREPROCESS DATA FILE FOR MODEL DEVELOPMENT

This appendix includes the source code to generate a data file for the model development. The major inputs to the program include the detector, ramp, bus, and detector data. It also includes a means of specifying the segment lengths and the time interval for traffic speed reporting. The detector data may be provided as an EXCEL file describing the location of detectors in the freeway, the controller ID and the distance of the detectors from a reference point. The ramp data may also be provided as an EXCEL file including the location of ramps, i.e. the distance of the ramps from a reference point.

Initially, the detector and ramp data may be entered by specifying appropriate EXCEL files. The bus data may be specified as a *.csv file. The format for all the files is also included in this appendix. The segment lengths can be specified and also the time period can be selected. After specifying the proper inputs, the detector file is automatically loaded into the program for the rest of the computation. To run this program efficiently, all the input files should be placed within the same source folder. After the program runs successfully, the output can be exported to a separate folder. The total operation for a particular data set may be executed within a minute.

```
Option Explicit
Option Compare Database
Public Const dblHour As Double = 1 / 24
Public Const dblMin As Double = dblHour / 60
Public Const dblSec As Double = dblMin / 60
Public Const MinTime As Date = #12:01:30 AM#
Public Const MaxTime As Date = #12:04:10 AM#
```

```
Public Function GetPath(str As String) As String
'extracts directory path from import string
Dim I As Integer
 For I = Len(str) To 1 Step -1
    If Mid(str, I, 1) = "\" Then
       GetPath = Mid(str, 1, I - 1)
       Exit Function
    End If
  Next I
End Function
Public Function GetName(str As String) As String
'extracts file name and extension from import string
Dim I As Integer
  For I = Len(str) To 1 Step -1
    If Mid(str, I, 1) = "\" Then
       GetName = Mid(str, I + 1, Len(str))
       Exit Function
    End If
  Next I
End Function
Public Function FillTripDir()
'determine trips and directions and populate tblImport
Dim db As DAO.Database
Dim rstCombo, rstImport As DAO.Recordset
Dim strSQL, strDir1, strDir2 As String
Dim iTrip, iLeg As Integer
Dim Dist1, Dist2 As Double
Dim Time1, Time2 As Date
  Set db = CurrentDb
  Set rstCombo = db.OpenRecordset("tblCombo")
  rstCombo.MoveFirst
  iTrip = 1
  'for each combination of Route and IVLUID
  Do Until rstCombo.EOF
    strSQL = "SELECT tblImport.* FROM tblImport " & _
     "WHERE tblImport.Date = #" & rstCombo.Fields("Date") & "# " & _
     "AND tblImport.Route = """ & rstCombo.Fields("Route") & """ " & _
```

```
 "AND tblImport.IVLUID = " & rstCombo.Fields("IVLUID") & ";"
 Set rstImport = db.OpenRecordset(strSQL)
With rstImport
    .MoveFirst
    strDir1 = "X"
    iLeg = 1
    .MoveNext
    'for each pair of reports
    Do Until .EOF
        'get distance and time for first report
        .MovePrevious
        Dist1 = .Fields("Distance")
        Time1 = .Fields("Time")
        'get distance and time for second report
        .MoveNext
        Dist2 = .Fields("Distance")
        Time2 = .Fields("Time")
        'move back to first report
        .MovePrevious
        'establish direction of travel for pair
        If Dist2 > Dist1 Then
            strDir2 = "N"
        Else
            strDir2 = "S"
        End If
        'if pair is first leg of trip, then...
        If iLeg = 1 Then
            'if report times are greater than 4 minutes apart
            'or less than 1.5 minutes apart,
            'then increment trip and reset leg
            If Time2 - Time1 > MaxTime _
             Or Time2 - Time1 < MinTime Then
                .Edit
                .Fields("TripID") = iTrip
                .Fields("Direction") = "X"
                .Update
                iTrip = iTrip + 1
```

```
        iLeg = 1
      'else increment leg
      Else
        .Edit
        .Fields("TripID") = iTrip
        .Fields("Direction") = strDir2
        .Update
        strDir1 = strDir2
        iLeg = iLeg + 1
      End If
'if pair is subsequent leg of trip, then...
  Else 'If iLeg > 1 Then
      'if report times are greater than 4 minutes apart
      'or less than 1.5 minutes apart,
      'then increment trip and reset leg
      If Time2 - Time1 > MaxTime _
       Or Time2 - Time1 < MinTime Then
        .Edit
        .Fields("TripID") = iTrip
        .Fields("Direction") = strDir1
        .Update
        strDir1 = "X"
        iTrip = iTrip + 1
        iLeg = 1
      'if direction of travel changes,
      'then increment trip and reset leg
      ElseIf strDir2 <> strDir1 Then
        .Edit
        .Fields("TripID") = iTrip
        .Fields("Direction") = strDir1
        .Update
        strDir1 = "X"
        iTrip = iTrip + 1
        iLeg = 1
        'else increment leg
      Else 'If strDir2 = strDir1 Then
        .Edit
```

```
                .Fields("TripID") = iTrip
                .Fields("Direction") = strDir2
                .Update
                strDir1 = strDir2
                iLeg = iLeg + 1
              End If
            End If
            .MoveNext
            .MoveNext
          Loop
           'if last report, then increment trip and reset leg
          .MoveLast
          .Edit
          .Fields("TripID") = iTrip
          .Fields("Direction") = strDir1
          .Update
          iTrip = iTrip + 1
          iLeg = 1
        End With
        rstCombo.MoveNext
      Loop
End Function
Public Function FillSegmentStart(dist As Double) As Long
'looks up SegmentID for dist as first report of pair
Dim db As DAO.Database
Dim rst As DAO.Recordset
Set db = CurrentDb
    Set rst = db.OpenRecordset("tblSegment")
    With rst
        .MoveFirst
        Do Until .EOF
          'find SegmentID within which dist falls
          If .Fields("StartDistance") <= dist And _
           .Fields("EndDistance") > dist Then
            FillSegmentStart = .Fields("SegmentID")
            Exit Function
          Else
```

```
            .MoveNext
        End If
    Loop
    FillSegmentStart = 0
  End With
End Function
Public Function FillSegmentEnd(dist As Double) As Long
'looks up SegmentID for dist as second report of pair
Dim db As DAO.Database
Dim rst As DAO.Recordset
  Set db = CurrentDb
  Set rst = db.OpenRecordset("tblSegment")
  With rst
    .MoveFirst
    Do Until .EOF
        'find SegmentID within which dist falls
        If .Fields("StartDistance") < dist And _
         .Fields("EndDistance") >= dist Then
            FillSegmentEnd = .Fields("SegmentID")
            Exit Function
        Else
            .MoveNext
        End If
    Loop
    FillSegmentEnd = 0
  End With
End Function
Public Function FillIntervalStart(time As Double) As Long
'looks up IntervalID for time as first report of pair
Dim db As DAO.Database
Dim rst As DAO.Recordset
  Set db = CurrentDb
  Set rst = db.OpenRecordset("tblInterval")
  With rst
    .MoveFirst
    Do Until .EOF
        'find IntervalID within which time falls
```

```
            If .Fields("StartTime") <= time And _
             .Fields("EndTime") > time Then
                FillIntervalStart = .Fields("IntervalID")
                Exit Function
            Else
                .MoveNext
            End If
        Loop
        FillIntervalStart = 0
    End With
End Function
Public Function FillIntervalEnd(time As Double) As Long
'looks up IntervalID for time as second report of pair
Dim db As DAO.Database
Dim rst As DAO.Recordset
    Set db = CurrentDb
    Set rst = db.OpenRecordset("tblInterval")
        With rst
        .MoveFirst
        Do Until .EOF
            'find IntervalID within with time falls
            If .Fields("StartTime") < time And _
             .Fields("EndTime") >= time Then
                FillIntervalEnd = .Fields("IntervalID")
                Exit Function
            Else
                .MoveNext
            End If
        Loop
        FillIntervalEnd = 0
    End With

End Function
Public Function FillPairs()
'pair up trips from qryNorth and populate tblPairs
Dim db As DAO.Database
Dim rstCalcs As DAO.Recordset
```

```
Dim Pair, Trip1, Trip2 As Long
Dim Seg1, Seg2, Int1, Int2 As Long
Dim Time1, Dist1, Time2, Dist2 As Double
Dim strSQL As String
 Set db = CurrentDb
    Set rstCalcs = db.OpenRecordset("qryNorth")
        'clear tblPairs
    strSQL = "DELETE * FROM tblPairs;"
    DoCmd.SetWarnings False
    DoCmd.RunSQL strSQL
    DoCmd.SetWarnings True
        Pair = 1
    With rstCalcs
        'get info for first report of pair
        .MoveFirst
        Trip1 = .Fields("TripID")
        Seg1 = .Fields("SegmentStartID")
        Int1 = .Fields("IntervalStartID")
        Time1 = .Fields("Time")
        Dist1 = .Fields("Distance")
        .MoveNext
                Do Until .EOF
            'get info for second report of pair
            Trip2 = .Fields("TripID")
            Seg2 = .Fields("SegmentEndID")
            Int2 = .Fields("IntervalEndID")
            Time2 = .Fields("Time")
            Dist2 = .Fields("Distance")
            'if reports are for same trip,
            'then populate tblPairs
            If Trip2 = Trip1 Then
                'if second report falls outside of valid range
                'then skip this pair
                If Int2 = 0 Then
                    'do not consider this pair
                Else
                    DoCmd.SetWarnings False
```

```
                strSQL = "INSERT INTO tblPairs " & _
                 "VALUES ( " & Pair & ", " & _
                 Trip1 & ", " & _
                 Seg1 & ", " & _
                 Int1 & ", #" & _
                 Time1 & "#, " & _
                 Dist1 & ", " & _
                 Seg2 & ", " & _
                 Int2 & ", #" & _
                 Time2 & "#, " & _
                 Dist2 & ", " & _
                 CDbl(Time2 - Time1) & ", " & _
                 Dist2 - Dist1 & " );"
                DoCmd.RunSQL strSQL
                DoCmd.SetWarnings True
                Pair = Pair + 1
            End If
        End If
                'carry over info for second report
        'to first report of new pair
        Trip1 = .Fields("TripID")
        Seg1 = .Fields("SegmentStartID")
        Int1 = .Fields("IntervalStartID")
        Time1 = .Fields("Time")
        Dist1 = .Fields("Distance")
MoveNext
    Loop
  End With
End Function
Public Function FillCalcs()
'calculate DPrime and TPrime of pairs and populate tblCalculations
Dim db As DAO.Database
Dim rstPairs As DAO.Recordset
Dim rstSegs As DAO.Recordset
Dim rstInts As DAO.Recordset
Dim iInt As Integer
Dim Pair As Long
```

```
Dim Int1, Int2, Seg1, Seg2 As Long
Dim SegBreakStart, SegBreakEnd As Long
Dim Dist1, Dist2, PairDist As Double
Dim Dprime, Dsegment, DistBreak As Double
Dim PairTime, Tprime As Double
Dim Time1, Time2, TimeBreak As Date
Dim strSQL As String

    Set db = CurrentDb
    Set rstPairs = db.OpenRecordset("tblPairs")
    strSQL = "SELECT * FROM tblSegment"
    Set rstSegs = db.OpenRecordset(strSQL)
    strSQL = "SELECT * FROM tblInterval"
    Set rstInts = db.OpenRecordset(strSQL)
        'clear tblCalculations
    strSQL = "DELETE * FROM tblCalculations;"
    DoCmd.SetWarnings False
    DoCmd.RunSQL strSQL
    DoCmd.SetWarnings True
        With rstPairs
        .MoveFirst
                'calculate for each pair
        Do Until .EOF
            'get information for each pair
            Pair = .Fields("PairID")
            Seg1 = .Fields("Segment1ID")
            Int1 = .Fields("Interval1ID")
            Time1 = .Fields("Time1")
            Dist1 = .Fields("Distance1")
            Seg2 = .Fields("Segment2ID")
            Int2 = .Fields("Interval2ID")
            Time2 = .Fields("Time2")
            Dist2 = .Fields("Distance2")
            PairTime = .Fields("PairTime")
            PairDist = .Fields("PairDistance")
                        'if report pair straddles 2 intervals,
            'then get time, distance, and SegmentID for break
```

```
If Int1 <> Int2 Then
    If Int1 = 0 Then
        'move to interval of second report and get start time
        rstInts.FindFirst "[IntervalID] = " & Int2
        TimeBreak = rstInts.Fields("StartTime")
    Else 'If Int1 > 0 Then
        'move to interval of first report and get end time
        rstInts.FindFirst "[IntervalID] = " & Int1
        TimeBreak = rstInts.Fields("EndTime")
    End If
    'calculate distance at break
    DistBreak = Dist1 + CDbl(TimeBreak - Time1) / PairTime * PairDist
    'calculate segment at break as start and end of pair
    SegBreakStart = FillSegmentStart(DistBreak)
    SegBreakEnd = FillSegmentEnd(DistBreak)

    'pair break with second report
    'and move to segment of break
    rstSegs.FindFirst "[SegmentID] = " & SegBreakStart
                'calculate for each segment covered by pair
    Do Until rstSegs.EOF
        'calculate Dsegment of each segment
        Dsegment = rstSegs.Fields("EndDistance") _
         - rstSegs.Fields("StartDistance")
                    'calculate Dprime of each pair
        'if first report is in this segment
        If SegBreakStart = rstSegs.Fields("SegmentID") Then
            'if second report is in this segment
            If Seg2 = rstSegs.Fields("SegmentID") Then
                Dprime = Dist2 - DistBreak
            'if second report is in later segment
            Else 'If Seg2 > rstSegs.Fields("SegmentID") Then
                Dprime = rstSegs.Fields("EndDistance") - DistBreak
            End If
                    'if first report is in earlier segment
        Else 'If SegBreakStart < rstSegs.Fields("SegmentID") Then
            'if second report is in this segment
```

```
            If Seg2 = rstSegs.Fields("SegmentID") Then
               Dprime = Dist2 - rstSegs.Fields("StartDistance")
            'if second report is in later segment
            Else 'If Seg2 > rstSegs.Fields("SegmentID") Then
               Dprime = Dsegment
            End If
         End If
         'calculate Tprime of each pair
         Tprime = Dprime / PairDist * PairTime
          'populate tblCalculations
         strSQL = "INSERT INTO tblCalculations VALUES ( " & _
          .Fields("PairID") & ", " & _
          rstSegs.Fields("SegmentID") & ", " & _
          Int2 & ", " & _
          Dprime & ", " & _
          Tprime & " );"
         DoCmd.SetWarnings False
         DoCmd.RunSQL strSQL
         DoCmd.SetWarnings True
                        'exit loop when last segment is reached
         If Seg2 = rstSegs.Fields("SegmentID") Then Exit Do
         'otherwise, move to next segment
         rstSegs.MoveNext
      Loop
               'else if report pair falls within same interval,
   'then calculate for each segment covered by pair
   Else 'If Int1 = Int2
      'move to segment of first report
      rstSegs.FindFirst "[SegmentID] = " & Seg1
    'calculate for each segment covered by pair
      Do Until rstSegs.EOF
         'calculate Dsegment of each segment
         Dsegment = (rstSegs.Fields("EndDistance") _
          - rstSegs.Fields("StartDistance"))
        'if first report is in this segment
         If Seg1 = rstSegs.Fields("SegmentID") Then
            'if second report is in this segment
```

```
            If Seg2 = rstSegs.Fields("SegmentID") Then
                Dprime = PairDist
            'if second report is in later segment
            Else 'If Seg2 > rstSegs.Fields("SegmentID") Then
                Dprime = rstSegs.Fields("EndDistance") - Dist1
            End If
        'if first report is in earlier segment
        Else 'If Seg1 < rstSegs.Fields("SegmentID") Then
            'if second report is in this segment
            If Seg2 = rstSegs.Fields("SegmentID") Then
                Dprime = Dist2 - rstSegs.Fields("StartDistance")
            'if second report is in later segment
            Else 'If Seg2 > rstSegs.Fields("SegmentID") Then
                Dprime = Dsegment
            End If
        End If
        'calculate Tprime of each pair
        Tprime = Dprime / PairDist * PairTime
         'populate tblCalculations
        strSQL = "INSERT INTO tblCalculations VALUES ( " & _
         .Fields("PairID") & ", " & _
         rstSegs.Fields("SegmentID") & ", " & _
         Int1 & ", " & _
         Dprime & ", " & _
         Tprime & " );"
        DoCmd.SetWarnings False
        DoCmd.RunSQL strSQL
        DoCmd.SetWarnings True

        'exit loop when last segment is reached
        If Seg2 = rstSegs.Fields("SegmentID") Then Exit Do
        'otherwise, move to next segment
        rstSegs.MoveNext
    Loop
End If
'move to next pair
.MoveNext
```

```
        Loop
    End With
End Function
Public Function FillExport()
'populate tblExport from tblPairs and tblCalculations
Dim db As DAO.Database
Dim rstPairCalcs As DAO.Recordset
Dim rstExport As DAO.Recordset
Dim strSQL As String
Dim nTrip, nPair As Integer
Dim VTWS, VDWS, VSMS As Double
Dim sumVTWSnum, sumVTWSden As Double
Dim sumVDWSnum, sumVDWSden As Double
Dim sumVSMSden As Double
    Set db = CurrentDb
    'clear tblExport and populate with intervals and segments
    strSQL = "DELETE * FROM tblExport;"
    DoCmd.SetWarnings False
    DoCmd.RunSQL strSQL
    DoCmd.OpenQuery "qryPreExport"
    DoCmd.SetWarnings True
    Set rstExport = db.OpenRecordset("tblExport")
    With rstExport
        .MoveFirst
        Do Until .EOF
            'build calculations recordset for each export record
            strSQL = "SELECT * FROM qrySpeed WHERE " & _
             "SegmentID = " & .Fields("SegmentID") & " AND " & _
             "IntervalID = " & .Fields("IntervalID") & ";"
            Set rstPairCalcs = db.OpenRecordset(strSQL)
            'initialize average speed equation numerators and denominators
'            nTrip = 0
            nPair = 0
            sumVTWSnum = 0
            sumVTWSden = 0
            sumVDWSnum = 0
            sumVDWSden = 0
```

```
        sumVSMSden = 0
        'if there are no pairs, do not calculate speed
        If rstPairCalcs.RecordCount = 0 Then
'            .Edit
'            .Fields("NPairs") = nPair
'            .Update
        'else calculate 3 different average speeds
        Else
            rstPairCalcs.MoveFirst
            Do Until rstPairCalcs.EOF
                nPair = nPair + 1
                'summate VTWS numerator and denominator
                sumVTWSnum = sumVTWSnum + _
                 rstPairCalcs.Fields("Dprime")
                sumVTWSden = sumVTWSden + _
                 rstPairCalcs.Fields("Tprime")
                                'summate VDWS numerator and denominator
                sumVDWSnum = sumVDWSnum + _
                 rstPairCalcs.Fields("Dprime") * _
                 rstPairCalcs.Fields("PairDistance") / _
                 rstPairCalcs.Fields("PairTime")
                sumVDWSden = sumVDWSden + _
                 rstPairCalcs.Fields("Dprime")
                'summate VSMS denominator
                sumVSMSden = sumVSMSden + _
                 rstPairCalcs.Fields("PairTime") / _
                 rstPairCalcs.Fields("PairDistance")
                rstPairCalcs.MoveNext
            Loop
                        'calculate the 3 average speeds and populate export
            VTWS = sumVTWSnum / sumVTWSden / 5280 / 24
            VDWS = sumVDWSnum / sumVDWSden / 5280 / 24
            VSMS = 1 / (1 / nPair * sumVSMSden) / 5280 / 24
                    .Edit
'            .Fields("NTrips") = nTrip
            .Fields("NPairs") = nPair
            .Fields("VTWS") = VTWS
```

```
            .Fields("VDWS") = VDWS
            .Fields("VSMS") = VSMS
            .Update
         End If
               .MoveNext
      Loop
   End With
   'calculate number of trips (buses) and populate export
   DoCmd.SetWarnings False
   DoCmd.OpenQuery "qryCountTrips"
   DoCmd.OpenQuery "qryUpdateTrips"
   DoCmd.SetWarnings True
End Function
Public Function FillDaily()
'populate tblDaily from qryDaily, tblPairs, and tblCalculations
Dim db As DAO.Database
Dim rstDailyPairCalcs As DAO.Recordset
Dim rstDaily As DAO.Recordset
Dim strSQL As String
Dim nPair As Integer
Dim VTWS, VDWS, VSMS As Double
Dim sumVTWSnum, sumVTWSden As Double
Dim sumVDWSnum, sumVDWSden As Double
Dim sumVSMSden As Double

   Set db = CurrentDb
      'clear tblDaily and populate with intervals, segments, and dates
   strSQL = "DELETE * FROM tblDaily;"
   DoCmd.SetWarnings False
   DoCmd.RunSQL strSQL
   DoCmd.OpenQuery "qryPreDaily"
   DoCmd.SetWarnings True
   Set rstDaily = db.OpenRecordset("tblDaily")
      With rstDaily
      .MoveFirst
            Do Until .EOF
         'build calculations recordset for each export record
```

```
        strSQL = "SELECT * FROM qryDailySpeed WHERE " & _
         "[Date] = #" & .Fields("Date") & "# AND " & _
         "SegmentID = " & .Fields("SegmentID") & " AND " & _
         "IntervalID = " & .Fields("IntervalID") & ";"
        Set rstDailyPairCalcs = db.OpenRecordset(strSQL)
                'initialize average speed equation numerators and denominators
        nPair = 0
        sumVTWSnum = 0
        sumVTWSden = 0
        sumVDWSnum = 0
        sumVDWSden = 0
        sumVSMSden = 0
                'if there are no pairs, do not calculate speed
        If rstDailyPairCalcs.RecordCount = 0 Then
'           .Edit
'           .Fields("NPairs") = nPair
'           .Update
        'else calculate 3 different average speeds
        Else
            rstDailyPairCalcs.MoveFirst
            Do Until rstDailyPairCalcs.EOF
                nPair = nPair + 1

                'summate VTWS numerator and denominator
                sumVTWSnum = sumVTWSnum + _
                 rstDailyPairCalcs.Fields("Dprime")
                sumVTWSden = sumVTWSden + _
                 rstDailyPairCalcs.Fields("Tprime")
                'summate VDWS numerator and denominator
                sumVDWSnum = sumVDWSnum + _
                 rstDailyPairCalcs.Fields("Dprime") * _
                 rstDailyPairCalcs.Fields("PairDistance") / _
                 rstDailyPairCalcs.Fields("PairTime")
                sumVDWSden = sumVDWSden + _
                 rstDailyPairCalcs.Fields("Dprime")
                  'summate VSMS denominator
                sumVSMSden = sumVSMSden + _
```

```
            rstDailyPairCalcs.Fields("PairTime") / _
            rstDailyPairCalcs.Fields("PairDistance")
           rstDailyPairCalcs.MoveNext
         Loop
         'calculate the 3 average speeds and populate tblDaily
         VTWS = sumVTWSnum / sumVTWSden / 5280 / 24
         VDWS = sumVDWSnum / sumVDWSden / 5280 / 24
         VSMS = 1 / (1 / nPair * sumVSMSden) / 5280 / 24
         .Edit
         .Fields("NPairs") = nPair
         .Fields("VTWS") = VTWS
         .Fields("VDWS") = VDWS
         .Fields("VSMS") = VSMS
         .Update
       End If
     .MoveNext
     Loop
   End With
   'calculate number of trips (buses) and populate tblDaily
   DoCmd.SetWarnings False
   DoCmd.OpenQuery "qryCountDailyTrips"
   DoCmd.OpenQuery "qryUpdateDailyTrips"
   DoCmd.SetWarnings True
End Function
Public Function UndoHourglass()
   DoCmd.Hourglass (False)
End Function
Public Sub ImportDetector()
' Version 1:
' Version 2: Add 3 more columns (P1Vol,P2Vol,P3Vol) in tblDetectorImp for density calculation 04/26/2002
' Version 3: Add WHERE clause to check if empty record. This can happen while importing
'           the Excel file into Temp table (05/10/2002)\
Dim strSQL As String
Dim I As Long
Dim db As DAO.Database
Dim rstDetectorID As DAO.Recordset
'Import data into temporary table
```

```
    DoCmd.TransferSpreadsheet acImport, acSpreadsheetTypeExcel9, "Temp", "detector.xls", True

'Copy data into tblDetectorImp sorted accordingly and delete temporary table
    DoCmd.SetWarnings False
    strSQL = "DELETE * FROM tblDetectorImp;"
    DoCmd.RunSQL strSQL
    strSQL = _
     "INSERT INTO tblDetectorImp ( GroupID, ControllerID, Location, [Date], [Time], P1S1, P2S2, P3S3, P4S4,
P5S5, P1Vol, P2Vol, P3Vol ) " & _
     "SELECT Temp.GroupID, Temp.ControllerID, Temp.Location, " & _
     "DateValue(Temp.DateTime), TimeValue(Temp.DateTime), " & _
     "Temp.P1S1, Temp.P2S2, Temp.P3S3, Temp.P4S4, Temp.P5S5, " & _
     "Temp.P1Vol, Temp.P2Vol, Temp.P3Vol " & _
     "FROM Temp WHERE Temp.ControllerID <> NULL ORDER BY Temp.GroupID, Temp.ControllerID,
DateValue(Temp.DateTime), TimeValue(Temp.DateTime);"
    DoCmd.RunSQL strSQL
    DoCmd.DeleteObject acTable, "Temp"
    DoCmd.SetWarnings True
'ID of each detector
    'DoCmd.SetWarnings False
    'strSQL = "DELETE * FROM tblDetectorID;"
    'DoCmd.RunSQL strSQL
    'strSQL = "INSERT INTO tblDetectorID ( ControllerID, Location ) " & _
    ' "SELECT DISTINCT tblDetectorImp.ControllerID, tblDetectorImp.Location FROM tblDetectorImp;"
    'DoCmd.RunSQL strSQL
    'DoCmd.SetWarnings True
End Sub
Public Sub FillDetector()
'Version 1: Compute Average Speed for Each Detector
'Version 2: Compute also Average Density for Each Detector (04/26/2002)
Dim db As DAO.Database
Dim strSQL As String
Dim rstDetectorExp As DAO.Recordset
Dim rstDetectorID As DAO.Recordset
Dim rstInterval As DAO.Recordset
Dim rstData As DAO.Recordset
Dim rstDate As DAO.Recordset
```

```
Dim rstVariance As DAO.Recordset
Dim sumSpeed As Double, denomenator As Integer
'Version 2:
Dim Count_P1Vol As Integer, Count_P2Vol As Integer, Count_P3Vol As Integer
Dim sumP1Vol As Long, sumP2Vol As Long, sumP3Vol As Long
Dim Count_P1S1 As Integer, Count_P2S2 As Integer, Count_P3S3 As Integer
Dim sumSpeed_P1S1 As Double, sumSpeed_P2S2 As Double, sumSpeed_P3S3 As Double
'End Version 2:
    Set db = CurrentDb

    'Add Date to tblDetectorDate - These are Date of Detector Data imported
    strSQL = "DELETE * FROM tblDetectorDate;"
    DoCmd.SetWarnings False
    DoCmd.RunSQL strSQL
    strSQL = "INSERT INTO tblDetectorDate " & _
            "SELECT DISTINCT Date FROM tblDetectorImp " & _
            "ORDER BY Date;"
    DoCmd.RunSQL strSQL
    'End Add Date
    'Clear tblDetectorExp
    strSQL = "DELETE * FROM tblDetectorExp;"
    DoCmd.RunSQL strSQL
    Set rstInterval = db.OpenRecordset("tblInterval")
    Set rstDetectorID = db.OpenRecordset("tblDetectorID")
    Set rstDate = db.OpenRecordset("tblDetectorDate")
    Set rstDetectorExp = db.OpenRecordset("tblDetectorExp")
    'Initialization
    sumSpeed = 0
    denomenator = 0
    sumP1Vol = 0: sumSpeed_P1S1 = 0
    sumP2Vol = 0: sumSpeed_P2S2 = 0
    sumP3Vol = 0: sumSpeed_P3S3 = 0
    Count_P1Vol = 0: Count_P1S1 = 0
    Count_P2Vol = 0: Count_P2S2 = 0
    Count_P3Vol = 0: Count_P3S3 = 0

    With rstDetectorID
```

```
.MoveFirst
rstInterval.MoveFirst
rstDate.MoveFirst
Do Until .EOF
   Do Until rstDate.EOF
      Do Until rstInterval.EOF
      'Get Variances From "tblDetectorImp"
      strSQL = "SELECT VAR(P1S1) AS Var1, VAR(P2S2) AS Var2, VAR(P3S3) AS Var3 FROM tblDetectorImp WHERE GroupID = " & .Fields("GroupID") & " AND ControllerID = " & _
            .Fields("ControllerID") & _
            "AND [Date] = #" & rstDate.Fields("Date") & "# " & _
            "AND [Time] >= #" & rstInterval.Fields("StartTime") & "# AND [Time] < #" & _
            rstInterval.Fields("EndTime") & "#;"
      Set rstVariance = db.OpenRecordset(strSQL)

      strSQL = "SELECT * FROM tblDetectorImp WHERE GroupID = " & .Fields("GroupID") & " AND ControllerID = " & _
            .Fields("ControllerID") & _
            "AND [Date] = #" & rstDate.Fields("Date") & "# " & _
            "AND [Time] >= #" & rstInterval.Fields("StartTime") & "# AND [Time] < #" & _
            rstInterval.Fields("EndTime") & "#;"
      Set rstData = db.OpenRecordset(strSQL)

      If rstData.RecordCount < 1 Then 'No record
         rstDetectorExp.AddNew
         rstDetectorExp.Fields("DetectorNewID") = .Fields("DetectorNewID")
         rstDetectorExp.Fields("Date") = rstDate.Fields("Date")
         rstDetectorExp.Fields("IntervalID") = rstInterval.Fields("IntervalID")
         rstDetectorExp.Fields("AVGSPEED") = -1
         rstDetectorExp.Update
         rstInterval.MoveNext
         denomenator = 0
         sumSpeed = 0
         'Version 2:
         sumP1Vol = 0: sumSpeed_P1S1 = 0
         sumP2Vol = 0: sumSpeed_P2S2 = 0
         sumP3Vol = 0: sumSpeed_P3S3 = 0
```

```
    Count_P1Vol = 0: Count_P1S1 = 0
    Count_P2Vol = 0: Count_P2S2 = 0
    Count_P3Vol = 0: Count_P3S3 = 0
    'End Version 2:
Else
    rstData.MoveFirst
    Do Until rstData.EOF
        'Version 1: Average Speed
        If rstData.Fields("P1S1") >= 0 Then
            sumSpeed = sumSpeed + rstData.Fields("P1S1")
            denomenator = denomenator + 1
            sumSpeed_P1S1 = sumSpeed_P1S1 + rstData.Fields("P1S1")
            Count_P1S1 = Count_P1S1 + 1     'Counter
        End If
        If rstData.Fields("P2S2") >= 0 Then
            sumSpeed = sumSpeed + rstData.Fields("P2S2")
            denomenator = denomenator + 1
            sumSpeed_P2S2 = sumSpeed_P2S2 + rstData.Fields("P2S2")
            Count_P2S2 = Count_P2S2 + 1     'Counter
        End If
        If rstData.Fields("P3S3") >= 0 Then
            sumSpeed = sumSpeed + rstData.Fields("P3S3")
            denomenator = denomenator + 1
            sumSpeed_P3S3 = sumSpeed_P3S3 + rstData.Fields("P3S3")
            Count_P3S3 = Count_P3S3 + 1     'Counter
        End If

        'Version 2: Average Density
        'Volumn
        If rstData.Fields("P1Vol") >= 0 Then                'Lane 1
            sumP1Vol = sumP1Vol + rstData.Fields("P1Vol")
            Count_P1Vol = Count_P1Vol + 1
        End If
        If rstData.Fields("P2Vol") >= 0 Then                'Lane 2
            sumP2Vol = sumP2Vol + rstData.Fields("P2Vol")
            Count_P2Vol = Count_P2Vol + 1
        End If
```

```
    If rstData.Fields("P3Vol") >= 0 Then                'Lane 3
        sumP3Vol = sumP3Vol + rstData.Fields("P3Vol")
        Count_P3Vol = Count_P3Vol + 1
    End If
    'End Version 2:

    rstData.MoveNext
Loop
rstData.MoveFirst
rstDetectorExp.AddNew
rstDetectorExp.Fields("DetectorNewID") = .Fields("DetectorNewID")
rstDetectorExp.Fields("Date") = rstDate.Fields("Date")
rstDetectorExp.Fields("IntervalID") = rstInterval.Fields("IntervalID")
If denomenator > 0 Then
    rstDetectorExp.Fields("AVGSPEED") = sumSpeed / denomenator
Else
    rstDetectorExp.Fields("AVGSPEED") = -2  'All are negative values
End If

'Version 2:
With rstDetectorExp
    'Average Volumn for Each Lane
    If Count_P1Vol = 0 Then 'No valid input
        Count_P1Vol = 1
    End If
    If Count_P2Vol = 0 Then 'No valid input
        Count_P2Vol = 1
    End If
    If Count_P3Vol = 0 Then 'No valid input
        Count_P3Vol = 1
    End If
    .Fields("AvgQ1") = sumP1Vol / Count_P1Vol     'Lane1
    .Fields("AvgQ2") = sumP2Vol / Count_P2Vol     'Lane2
    .Fields("AvgQ3") = sumP3Vol / Count_P3Vol     'Lane3

    'Average Speed for Each Lane
    If Count_P1S1 = 0 Then 'No valid input
```

```
   Count_P1S1 = 1
End If
If Count_P2S2 = 0 Then 'No valid input
   Count_P2S2 = 1
End If
If Count_P3S3 = 0 Then 'No valid input
   Count_P3S3 = 1
End If
.Fields("TMS1") = sumSpeed_P1S1 / Count_P1S1   'Lane1
.Fields("TMS2") = sumSpeed_P2S2 / Count_P2S2   'Lane2
.Fields("TMS3") = sumSpeed_P3S3 / Count_P3S3   'Lane3

'Variance for Each Lane
.Fields("Var1") = rstVariance.Fields("Var1")     'Lane1
.Fields("Var2") = rstVariance.Fields("Var2")     'Lane2
.Fields("Var3") = rstVariance.Fields("Var3")     'Lane3

'SMS for Each Lane
If .Fields("TMS1") = 0 Then 'No valid input
   .Fields("SMS1") = 0
Else
   .Fields("SMS1") = .Fields("TMS1") - (.Fields("Var1") / .Fields("TMS1")) 'Lane1
End If
If .Fields("TMS2") = 0 Then 'No valid input
   .Fields("SMS2") = 0
Else
   .Fields("SMS2") = .Fields("TMS2") - (.Fields("Var2") / .Fields("TMS2")) 'Lane2
End If
If .Fields("TMS3") = 0 Then 'No valid input
   .Fields("SMS3") = 0
Else
   .Fields("SMS3") = .Fields("TMS3") - (.Fields("Var3") / .Fields("TMS3")) 'Lane3
End If

'Density for Each Lane
If .Fields("SMS1") = 0 Then 'No valid input
   .Fields("K1") = 0
```

```
                Else
                    .Fields("K1") = .Fields("AvgQ1") / .Fields("SMS1")   'Lane1
                End If
                If .Fields("SMS2") = 0 Then 'No valid input
                    .Fields("K2") = 0
                Else
                    .Fields("K2") = .Fields("AvgQ2") / .Fields("SMS2")   'Lane2
                End If
                If .Fields("SMS3") = 0 Then 'No valid input
                    .Fields("K3") = 0
                Else
                    .Fields("K3") = .Fields("AvgQ3") / .Fields("SMS3")   'Lane3
                End If

                'Average From 3 Lanes
                .Fields("Density") = (.Fields("K1") + .Fields("K2") + .Fields("K3")) / 3   'All lanes
            End With
            rstDetectorExp.Update
            rstInterval.MoveNext

            'Initialization for Next Interval
            sumP1Vol = 0: sumSpeed_P1S1 = 0
            sumP2Vol = 0: sumSpeed_P2S2 = 0
            sumP3Vol = 0: sumSpeed_P3S3 = 0
            Count_P1Vol = 0: Count_P1S1 = 0
            Count_P2Vol = 0: Count_P2S2 = 0
            Count_P3Vol = 0: Count_P3S3 = 0
            denomenator = 0
            sumSpeed = 0
        End If
        Loop 'Next Interval
        rstInterval.MoveFirst
        rstDate.MoveNext
    Loop 'Next Date
    rstDate.MoveFirst
    .MoveNext
Loop 'Next Detector ID
```

```
    End With

    'Clean Up Memory
    DoCmd.SetWarnings True
    Set rstDetectorID = Nothing
    Set rstDate = Nothing
    Set rstInterval = Nothing
    Set rstData = Nothing
    Set rstDetectorExp = Nothing
    Set rstVariance = Nothing

End Sub

Public Sub ImportDensity()
'Not actually Import, But Prepare All Values for the Equation
'Fill those in "tblDensityImp"

Dim db As DAO.Database
Dim rstDetectorID As DAO.Recordset
Dim strSQL As String
Dim I As Long, NextDetectorID As Integer

    Set db = CurrentDb
    Set rstDetectorID = db.OpenRecordset("tblDetectorID")
    NextDetectorID = 0

    DoCmd.SetWarnings False
    strSQL = "DELETE * FROM tblDensityImp;"
    DoCmd.RunSQL strSQL

    'Step 1: Get Detector Density From "tblDetectorExp"
    strSQL = "INSERT INTO tblDensityImp (DetectorNewID, [Date], IntervalID, K) " & _
     "SELECT DetectorNewID, Date, IntervalID, Density FROM tblDetectorExp;"
    DoCmd.RunSQL strSQL

    'Step 2: Get SegmentID and Length
```

```
    With rstDetectorID
        .MoveLast
        For I = 1 To rstDetectorID.RecordCount - 1
            strSQL = "UPDATE tblDensityImp SET SegmentID = " & .Fields("SegmentID") & _
                    ", L = " & .Fields("Length") & _
                    ", M = " & .Fields("M") & _
                    ", Mnext = " & .Fields("Mnext") & _
                    " WHERE DetectorNewID = " & .Fields("DetectorNewID") & ";"
            DoCmd.RunSQL strSQL

            'We are not using last detector
            If I <> rstDetectorID.RecordCount Then 'Set Lnext (Li+1)
                .MovePrevious
                strSQL = "UPDATE tblDensityImp SET Lnext = " & .Fields("Length") & _
                        " WHERE DetectorNewID = " & .Fields("DetectorNewID") - 1 & ";"
                DoCmd.RunSQL strSQL
            End If
        Next
    End With

    DoCmd.SetWarnings True
    Set rstDetectorID = Nothing

End Sub

Public Sub FillDensity()
'Version 1: Calculation of Segment Density Equation
'Version 1: Fill in tblDensityExp
'Version 2: Add Number of On/Off Ramps in Each Segment (05/10/2002)
'Version 3: Add Weather Information (05/17/2002)

Dim db As DAO.Database
Dim strSQL As String
Dim rstDensityExp As DAO.Recordset
Dim rstSegment As DAO.Recordset
Dim rstInterval As DAO.Recordset
Dim rstDate As DAO.Recordset
```

```
Dim rstData As DAO.Recordset

'Expression(s)
Dim Exp1 As Double, Exp2 As Double
Dim Exp3 As Double, Exp4 As Double
Dim Knext As Double, Lnext As Double, Mnext As Double
Dim Kn As Double

'Number of Detectors
Dim NumOfDetectors As Integer, I As Integer
Dim NoOnRamp As Integer, NoOffRamp As Integer

'Weather Information
Dim WeathCode As Integer, WeathID As Integer
Dim Visibility As Double, WindSpeed As Double

    Set db = CurrentDb
    'Clear "tblDensityExp"
    strSQL = "DELETE * FROM tblDensityExp;"
    DoCmd.SetWarnings False
    DoCmd.RunSQL strSQL
    DoCmd.SetWarnings True

    Set rstInterval = db.OpenRecordset("tblInterval")
    Set rstSegment = db.OpenRecordset("tblSegment")
    Set rstDate = db.OpenRecordset("tblDetectorDate")
    Set rstDensityExp = db.OpenRecordset("tblDensityExp")
    Kn = 0
    NumOfDetectors = 0
    WeathCode = Visibility = WindSpeed = WeathID = 0

    With rstSegment 'Each Segment
        .MoveFirst
        Do Until .EOF
        'Version 2: Add Number of On/Off Ramps in Each Segment
        NoOnRamp = OnRamp(.Fields("SegmentID"))
        NoOffRamp = OffRamp(.Fields("SegmentID"))
```

```
    rstDate.MoveFirst
    Do Until rstDate.EOF    'Each Date
      rstInterval.MoveFirst
      Do Until rstInterval.EOF    'Each Time Interval
        strSQL = "SELECT * FROM tblDensityImp WHERE (SegmentID = " & .Fields("SegmentID") & _
          " AND [Date] = #" & rstDate.Fields("Date") & "#" & _
          " AND IntervalID = " & rstInterval.Fields("IntervalID") & ") ORDER BY DetectorNewID;"

        Set rstData = db.OpenRecordset(strSQL)
        rstData.MoveFirst
'Calculation of Segment Density
'EXPRESSION 1 :
        Exp1 = rstData.Fields("K") * rstData.Fields("L") * rstData.Fields("M")
        NumOfDetectors = rstData.RecordCount
'EXPRESSION 2 :
        Exp2 = 0
        For I = 1 To NumOfDetectors
          Lnext = rstData.Fields("Lnext")
          Mnext = rstData.Fields("Mnext")
          If I = NumOfDetectors Then 'Last detector in this segment
            Knext = 0
          Else
            rstData.MoveNext
            Knext = rstData.Fields("K")
            rstData.MovePrevious
          End If

          Exp2 = Exp2 + ((rstData.Fields("K") * Mnext) + (Knext * (1 - Mnext))) * Lnext
          If I < NumOfDetectors Then 'Last detector in this segment
            rstData.MoveNext
          End If
        Next I
'EXPRESSION 3 :
        Exp3 = 0
        rstData.MoveFirst
        I = 1
        While (I < NumOfDetectors)
```

```
        Exp3 = Exp3 + (rstData.Fields("M") * rstData.Fields("L"))
        rstData.MoveNext
        I = I + 1
    Wend
    Exp3 = Exp3 + (rstData.Fields("M") * rstData.Fields("L"))   'p
    Exp3 = Exp3 + (rstData.Fields("Mnext") * rstData.Fields("Lnext")) 'p+1
'EXPRESSION 4 :
    Exp4 = 0
    rstData.MoveFirst
    rstData.MoveNext 'i=q+1
    For I = 2 To NumOfDetectors 'q+1 to p
        Exp4 = Exp4 + (rstData.Fields("L") * (1 - rstData.Fields("M")))
        If I < NumOfDetectors Then
            rstData.MoveNext
        End If
    Next I

    Kn = (Exp1 + Exp2) / (Exp3 + Exp4)
    'End Calculation

    'Add to Table "tblDensityExp"
    rstDensityExp.AddNew
    On Error Resume Next
    rstDensityExp.Fields("SegmentID") = rstSegment.Fields("SegmentID")
    rstDensityExp.Fields("Date") = rstDate.Fields("Date")
    rstDensityExp.Fields("IntervalID") = rstInterval.Fields("IntervalID")
    If Kn > 0 Then
        rstDensityExp.Fields("Kn") = Kn
        rstDensityExp.Fields("AvgSpeed") = 72.0934 * Exp(1.8034 * Log(1 - (Kn / 178.1)))
        rstDensityExp.Fields("SpdRange") = SpdRange(rstDensityExp.Fields("AvgSpeed"))
    Else
        rstDensityExp.Fields("Kn") = Null
        rstDensityExp.Fields("AvgSpeed") = Null
        rstDensityExp.Fields("SpdRange") = Null
    End If
    rstDensityExp.Fields("OnRamp") = NoOnRamp
    rstDensityExp.Fields("OffRamp") = NoOffRamp
```

```
            'Weather Info.
            WeathID = WeatherID(rstDate.Fields("Date"), rstInterval.Fields("StartTime"), WeathCode, Visibility,
WindSpeed)
            rstDensityExp.Fields("WeathCode") = WeathCode
            rstDensityExp.Fields("Visibility") = Visibility
            rstDensityExp.Fields("WindSpeed") = WindSpeed
            rstDensityExp.Update
            Kn = 0
          rstInterval.MoveNext    'Next Interval
          Loop
        rstDate.MoveNext          'Next Date
        Loop
      .MoveNext                   'Next Segment
      Loop
  End With

  'Clean Up Memory
  Set rstSegment = Nothing
  Set rstDate = Nothing
  Set rstInterval = Nothing
  Set rstData = Nothing
  Set rstDensityExp = Nothing

End Sub

Function SpdRange(avgSpeed As Double) As Integer
'Version 3: Classify the Average Speeds into Range 1-4 (05/10/2002)
'       0-20  -> 1
'       20-40 -> 2
'       40-60 -> 3
'       > 60  -> 4

  If (0# <= avgSpeed) And (avgSpeed <= 20#) Then
    SpdRange = 1
  ElseIf (20# < avgSpeed) And (avgSpeed <= 40#) Then
    SpdRange = 2
  ElseIf (40# < avgSpeed) And (avgSpeed <= 60#) Then
```

```
        SpdRange = 3
    ElseIf (avgSpeed > 60#) Then
        SpdRange = 4
    Else
        SpdRange = 0    'Not Defined
    End If
End Function

Function OnRamp(SegmentID As Integer) As Integer
'Version 1: Number of On Ramps in particular Segment (05/10/2002)
Dim db As DAO.Database
Dim strSQL As String
Dim rstSegment As DAO.Recordset
Dim rstData As DAO.Recordset

    Set db = CurrentDb
    strSQL = "SELECT ID FROM tblRamp WHERE SegmentID=" & SegmentID & " AND OnOff=1;"
    Set rstData = db.OpenRecordset(strSQL)

    On Error GoTo ERROR
    rstData.MoveFirst
    Do Until rstData.EOF    'This is required to have a correct Return Value from RecordCount
        rstData.MoveNext
    Loop
    OnRamp = rstData.RecordCount
    Set rstData = Nothing
    Set rstSegment = Nothing
    Exit Function

ERROR:
    OnRamp = 0

End Function

Function OffRamp(SegmentID As Integer) As Integer
'Version 1: Number of Off Ramps in particular Segment (05/10/2002)
Dim db As DAO.Database
```

```
Dim strSQL As String
Dim rstSegment As DAO.Recordset
Dim rstData As DAO.Recordset

    Set db = CurrentDb
    strSQL = "SELECT * FROM tblRamp WHERE SegmentID = " & SegmentID & " AND OnOff = 2;"
    Set rstData = db.OpenRecordset(strSQL)

    On Error GoTo ERROR
    rstData.MoveFirst
    Do Until rstData.EOF    'This is required to have a correct Return Value from RecordCount
        rstData.MoveNext
    Loop
    OffRamp = rstData.RecordCount
    Set rstData = Nothing
    Set rstSegment = Nothing
    Exit Function

ERROR:
    OffRamp = 0

End Function

Sub ImportWeather()
'Version 1: Import file "weather.xls" into tblWeather (05/17/2002)
Dim strSQL As String

'Import data into temporary table
    DoCmd.TransferSpreadsheet acImport, acSpreadsheetTypeExcel9, "WeathTemp", "weather.xls", True

'Copy data into tblWeather sorted accordingly and delete temporary table
    DoCmd.SetWarnings False
    strSQL = "DELETE * FROM tblWeathImp;"
    DoCmd.RunSQL strSQL
    strSQL = _
     "INSERT INTO tblWeathImp ( [Date], [Time], WeathCode, Visibility, WindSpeed ) " & _
     "SELECT DateValue(WeathTemp.Date), TimeValue(WeathTemp.Time), " & _
```

```
 "WeathTemp.WeathCode, WeathTemp.Visibility, WeathTemp.WindSpeed " & _
 "FROM WeathTemp WHERE WeathTemp.WeathCode <> NULL ORDER BY DateValue(WeathTemp.Date), TimeValue(WeathTemp.Time);"
   DoCmd.RunSQL strSQL
   DoCmd.DeleteObject acTable, "WeathTemp"
   DoCmd.SetWarnings True

End Sub

Function WeatherID(WDate As Date, WTime As Date, ByRef WeathCode As Integer _
            , ByRef Visibility As Double, ByRef WindSpeed As Double) As Integer
'Version 1: Return WeatherID as specified Date and StartTime
'       Also return WeathCode, Visibility, and WindSpeed
'       for that particular WeatherID
'       (05/17/2002)

Dim db As DAO.Database
Dim strSQL As String
Dim rstData As DAO.Recordset
Dim sMin As String
Dim ETime As Date

   sMin = Mid(CStr(WTime), 3, 2)
   WTime = WTime - CInt(sMin) * dblMin
   ETime = WTime + 60 * dblMin

   Set db = CurrentDb

   strSQL = "SELECT * FROM tblWeathImp WHERE [Date] = #" & WDate & "# AND [Time] >= #" & WTime & "# AND [Time] < #" & ETime & "#;"
   Set rstData = db.OpenRecordset(strSQL)
   rstData.MoveFirst

   WeathCode = rstData.Fields("WeathCode")
   Visibility = rstData.Fields("Visibility")
   WindSpeed = rstData.Fields("WindSpeed")
```

```
    WeatherID = rstData.Fields("WeatherID")
    Set rstData = Nothing
    Return
End Function
```

APPENDIX D

SEGMENT LENGTH SELECTION

This appendix describes the various models developed before the selection of the final model. The following table describes the various segment lengths considered with different time intervals. Fixed length, and one and two mile segments, for 5 and 15-minute time intervals were selected for developing a general linear model (GLM). Segment lengths based on the location of on-ramps for 15-minute time intervals were also evaluated. The best results were obtained for varying segment lengths.

For all the different segment lengths selection options considered, some of the statistical results are presented in this section. For all the models developed, traffic speed was considered the dependent variable. The independent variables were bus speed, number of on-ramps per unit length of a freeway segment, total number of ramps, and the interaction variable were bus speed and number of on-ramps per unit length of a freeway segment, and wind speed.

Table D-1. Models Developed Using General Linear Regression Technique.

Cases	Models	DF	R^2	RMSE
1	2 mile segment, 15-minute time interval	4	0.70	6.811
2	2 mile segment, 5-minute time interval	2	0.68	13.465
3	1 mile segment, 15-minute time interval	4	0.71	8.47
4	Segments at On-Ramp, 15-minute time interval	4	0.62	9.034
5	Varying Segment Length, 15-minute time interval	2	0.79	5.748

```
15 min, 2 mi

              Parameter           Estimate     Standard Error    t Value    Pr > |t|

              Intercept        20.29593119      0.53733382       37.77      <.0001
              VDWS              0.40998117      0.01258530       32.58      <.0001
              OnRamp           -4.89454752      0.32281612      -15.16      <.0001
              VDWS*OnRamp       0.17847232      0.00809860       22.04      <.0001
              WindSpeed        -0.04781443      0.01245765       -3.84      0.0001

5 min, 2 mi


              Parameter           Estimate     Standard Error    t Value    Pr > |t|

              Intercept        12.43548880      0.17695081       70.28      <.0001
              VDWS              0.58252734      0.00681687       85.45      <.0001
              VDWS*OnRamp       0.06564101      0.00380773       17.24      <.0001


15 min, 1 mi

              Parameter           Estimate     Standard Error    t Value    Pr > |t|

              Intercept        23.49752325      0.50725485       46.32      <.0001
              VDWS              0.33324626      0.01505495       22.14      <.0001
              OnRamp          -25.22639272      0.48187846      -52.35      <.0001
              VDWS*OnRamp       0.57080321      0.01624984       35.13      <.0001
              WindSpeed        -0.22508718      0.02072798      -10.86      <.0001

15 min, On ramp

              Parameter            Estimate     Standard Error    t Value    Pr > |t|

              Intercept        -10.19321255      0.57325085      -17.78      <.0001
              VDWS               0.86524370      0.01337961       64.67      <.0001
              OffRamp           21.67681932      0.46815145       46.30      <.0001
              VDWS*OffRamp      -0.24227394      0.01088784      -22.25      <.0001
              WindSpeed         -0.04900896      0.01694786       -2.89      0.0038
```

15 min, varying segment

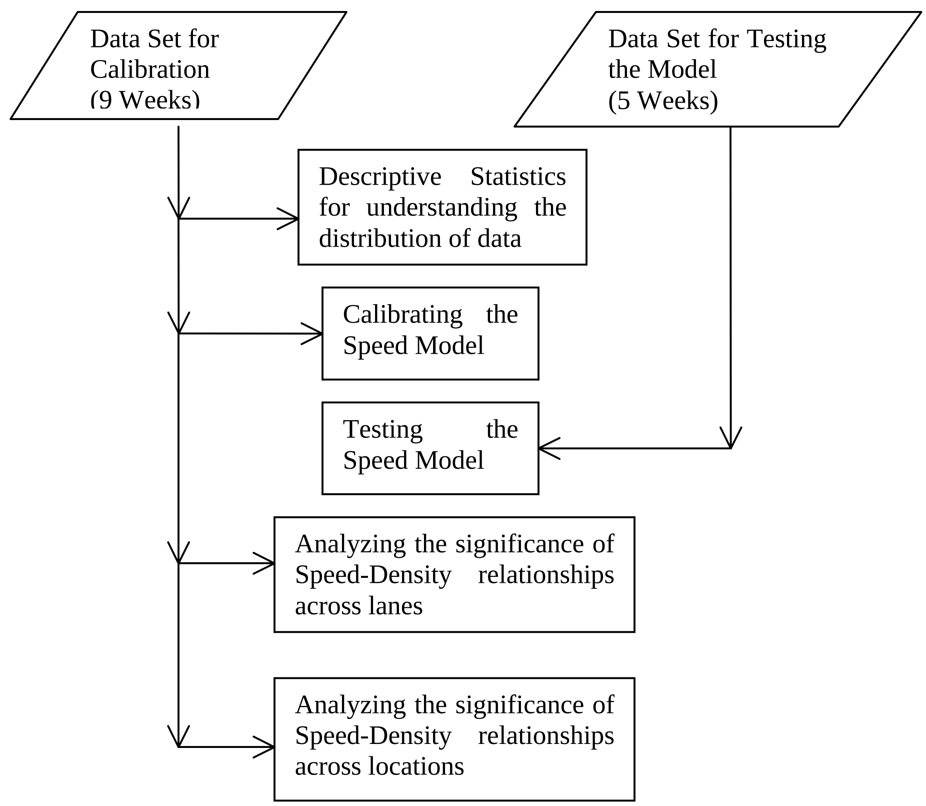
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	-10.19321255	0.57325085	-17.78	<.0001
VDWS	0.86524370	0.01337961	64.67	<.0001
OffRamp	21.67681932	0.46815145	46.30	<.0001
VDWS*OffRamp	-0.24227394	0.01088784	-22.25	<.0001
WindSpeed	-0.04900896	0.01694786	-2.89	0.0038

APPENDIX E

PROGRAM FOR LOGISTIC REGRESSION

This section describes the approach used in developing the model in a statistical package, SAS. In this section the source code used in SAS for the model development is provided.

Approach Used For Developing Speed Model Using SAS



SAS CODE FOR DESCRIPTIVE STATISTICS

```
PROC IMPORT OUT= WORK.SpeedModel1
            DATAFILE= "D:\My Documents\varysegment\outputvarysegment_summer.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
run;
data newspeedmodel;
set speedmodel1;
If Kn=0 then AvgSpeed=0;
If AvgSpeed=0 then delete;
run;
proc means data = newspeedmodel maxdec=3 nmiss range skewness
kurtosis cv css uss t std stderr mean min max n;
title ' Assorted Statistics for 15 minutes Varying Segment';
run;
proc corr data=newspeedmodel;
title ' Correlation for 15 minutes Varying Segment';
run;
```

SAS CODE FOR CALIBRATING THE MODEL

```
PROC IMPORT OUT= WORK.SpeedModel1
            DATAFILE= "D:\My Documents\varysegment\outputvarysegment.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
run;
data newspeedmodel1;
set speedmodel1;
If AvgSpeed =72.0934 then delete;
if vdws=. then delete;
if vdws<=20 then dws1=vdws;
if vdws>20 then dws1=0;
if vdws>20 then dws2=vdws;
if vdws<=20 then dws2=0;
if (dws1<20 and avgspeed<20) then vdwsinv=(log(dws1))**6;else vdwsinv=dws1;
if (dws2<=40 and avgspeed<=40) then vdwsinv1=(log(dws2))**2;else
if (dws2>40 and avgspeed<=40) then vdwsinv1=(log(dws2))**2;else vdwsinv1=dws2;
if vdwsinv1=. then vdwsinv1=0;
ramp = onramp + offramp;
length=length/5280;
rampunitlength= ramp/length;
if vdwsinv1=0 then ramp1=0;else ramp1=ramp;
if vdwsinv1=0 then length1=0;else length1=length;
if vdwsinv1=0 then rampunitlength1=0;else rampunitlength1=rampunitlength;
/*if vdwsinv=0 then rampunitlength2=0;else rampunitlength2=rampunitlength;*/
run;
data newspeedmodel1;
set newspeedmodel1;
If avgspeed >= 0 and avgspeed<=20 then detectorlevel= 1;else
If avgspeed > 20 and avgspeed<=40 then detectorlevel= 2;else
if avgspeed > 40 then detectorlevel=3;
run;
```

```
proc sort data=newspeedmodel1;by descending vdwsinv vdwsinv1;run;
proc logistic desc data= newspeedmodel1;
class detectorlevel;
model detectorlevel = vdwsinv vdwsinv1 vdwsinv1*rampunitlength1/aggregate scale=n rsq;
output out=results predprobs=(i c);
title 'Ordered Logit Analysis / Distance Weighted Speed';
run;
proc print data=results /noprint noobs;
var vdws ip_1 ip_2 ip_3;
run;
data newspeedmodel2;
set results;
p1=ip_1;
p2=ip_2;
p3=ip_3;
;
if (p1>p2 and p1>p3)then predictedTrafficSpeed=1;else
if (p2>p1 and p2>p3)then predictedTrafficSpeed=2;else
if (p3>p1 and p3>p2)then predictedTrafficSpeed=3;
run;
proc freq data=newspeedmodel2;
tables detectorlevel*predictedTrafficSpeed;
title '2-WAY CONTINGENCY TABLE';
run;
```

SAS CODE FOR TESTING THE MODEL

```
PROC IMPORT OUT= WORK.SpeedModel1
            DATAFILE= "D:\My Documents\TestFiles\output_testset_5Weeks.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
run;
data modifiedspeedmodel;
set speedmodel1;
If AvgSpeed =72.0934 then delete;
if vdws=. then delete;
if vdws<=20 then dws1=vdws;else dws1=0;
if vdws>20 then dws2=vdws;else dws2=0;
if (dws1<20 and avgspeed<20) then vdwsinv=(log(dws1))**6;else vdwsinv=dws1;
if (dws2<=40 and avgspeed<=40) then vdwsinv1=(log(dws2))**2;else
if (dws2>40 and avgspeed<=40) then vdwsinv1=(log(dws2))**2;else vdwsinv1=dws2;
if vdwsinv1=. then vdwsinv1=0;
if vdwsinv=. then vdwsinv=0;
ramp = onramp + offramp;
length=length/5280;
rampunitlength =(ramp/length);
if vdws<=20 then ramp1=rampunitlength;else ramp1=0;
if vdws>20 then ramp2=rampunitlength;else ramp2=0;
interaction1=log(dws1*ramp1);
interaction2=(dws2*ramp2)**2;
if interaction1=. then interaction1=0;
if interaction2=. then interaction2=0;
interaction=interaction1 + interaction2;
If avgspeed >= 0 and avgspeed<=20 then detectorlevel= 1;else
If avgspeed > 20 and avgspeed<=40 then detectorlevel= 2;else
if avgspeed > 40 then detectorlevel=3;
run;
data predicted;
```

```
set modifiedspeedmodel;
cal3=-13.13 -0.6592*vdwsinv +0.7979*vdwsinv1 +7.73*interaction1 -0.00065*interaction2;
cal2=-4.8024 -0.6592*vdwsinv +0.7979*vdwsinv1 +7.73*interaction1 -0.00065*interaction2;
p3 = exp(cal3)/(1 + exp(cal3));
pcombo = exp(cal2)/(1 + exp(cal2));
p2 = pcombo - p3;
p1= 1 - (p2 + p3);
run;
data testspeedmodel;
set predicted;
if (p1>p2 and p1>p3)then predictedTrafficSpeed=1;else
if (p2>p1 and p2>p3)then predictedTrafficSpeed=2;else
if (p3>p1 and p3>p2)then predictedTrafficSpeed=3;
run;
proc freq data=testspeedmodel;
tables detectorlevel*predictedTrafficSpeed;
title '2-WAY CONTINGENCY TABLE';
run;
```

SAS CODE FOR PLOTTING THE LOGITS ESTIMATED FORM THE SPEED MODEL

```
PROC IMPORT OUT= WORK.SpeedModel1
DATAFILE= "D:\My Documents\varysegment\outputvarysegment_summer.xls"
DBMS=EXCEL2000 REPLACE;
GETNAMES=YES;
run;
data modifiedspeedmodel;
set speedmodel1;
If AvgSpeed =72.0934 then delete;
if vdws=. then delete;
if vdws<=20 then dws1=vdws;else dws1=0;
if vdws>20 then dws2=vdws;else dws2=0;
if (dws1<20 and avgspeed<20) then vdwsinv=(log(dws1))**6;else vdwsinv=dws1;
if (dws2<=40 and avgspeed<=40) then vdwsinv1=(log(dws2))**2;else
if (dws2>40 and avgspeed<=40) then vdwsinv1=(log(dws2))**2;else vdwsinv1=dws2;
if vdwsinv1=. then vdwsinv1=0;
if vdwsinv=. then vdwsinv=0;
ramp = onramp + offramp;
length=length/5280;
rampunitlength =(ramp/length);
if vdws<=20 then ramp1=rampunitlength;else ramp1=0;
if vdws>20 then ramp2=rampunitlength;else ramp2=0;
interaction1=log(dws1*ramp1);
interaction2=(dws2*ramp2)**2;
if interaction1=. then interaction1=0;
if interaction2=. then interaction2=0;
interaction=interaction1 + interaction2;
If avgspeed >= 0 and avgspeed<=20 then detectorlevel= 1;else
If avgspeed > 20 and avgspeed<=40 then detectorlevel= 2;else
if avgspeed > 40 then detectorlevel=3;
run;
data predicted;
```

```
set modifiedspeedmodel;
cal3=-13.13 -0.6592*vdwsinv +0.7979*vdwsinv1 +7.73*interaction1 -0.00065*interaction2;
cal2=-4.8024 -0.6592*vdwsinv +0.7979*vdwsinv1 +7.73*interaction1 -0.00065*interaction2;
cal1= 4.8024 +0.6592*vdwsinv -0.7979*vdwsinv1 -7.73*interaction1 +0.00065*interaction2;
p3 = exp(cal3)/(1 + exp(cal3));
pcombo = exp(cal2)/(1 + exp(cal2));
p2 = pcombo - p3;
p1= 1 - (p2 + p3);
run;
data modifiedpredicted;
set predicted;
logit1=cal1;
logit2=cal2;
BusSpeed1=vdwsinv;
BusSpeed2=vdwsinv1;
run;
*plotting logit vs busspeed<20mph;
symbol1 v=dot i=none c=black;
proc gplot data=modifiedpredicted;
plot logit1*BusSpeed1/frame haxis=0 to 20 by 5 vaxis=-50 to 0 by 5;
run;
proc gplot data=modifiedpredicted;
plot logit2*BusSpeed2/frame haxis=20 to 80 by 5 vaxis=0 to 60 by 5;
run;
```

SAS CODE FOR TESTING THE SIGNIFICANCE OF SPEED-DENSITY RELATIONSHIPS ACROSS LANES

```
PROC IMPORT OUT= WORK.ANOVAUK

DATAFILE="D:\MyDocuments\ProjDocuments\Paper1\LaneWiseRelationshipsForAllLocation
s.xls"
            DBMS=EXCEL2000 REPLACE;
    GETNAMES=YES;
run;
Proc anova data=anovauk;
class density2;
model speed2 = density2;
title 'Lane wise relationships for Location2';
run;
Proc anova data=anovauk;
class density3;
model speed3 = density3;
title 'Lane wise relationships for Location3';
run;
Proc anova data=anovauk;
class density4;
model speed4 = density4;
title 'Lane wise relationships for Location4';
run;
Proc anova data=anovauk;
class density5;
model speed5 = density5;
title 'Lane wise relationships for Location5';
run;
Proc anova data=anovauk;
class density6;
model speed6 = density6;
title 'Lane wise relationships for Location6';
```

```
run;
Proc anova data=anovauk;
class density7;
model speed7 = density7;
title 'Lane wise relationships for Location7';
run;
Proc anova data=anovauk;
class density8;
model speed8 = density8;
title 'Lane wise relationships for Location8';
run;
Proc anova data=anovauk;
class density9;
model speed9=density9;
title 'Lane wise relationships for Location9';
run;
Proc anova data=anovauk;
class density10;
model speed10 = density10;
title 'Lane wise relationships for Location10';
run;
Proc anova data=anovauk;
class density11;
model speed11 = density11;
title 'Lane wise relationships for Location11';
run;
```

SAS CODE FOR TESTING THE SIGNIFICANCE OF SPEED-DENSITY RELATIONSHIPS ACROSS LOCATIONS

```
PROC IMPORT OUT= WORK.ANOVAUK
            DATAFILE= "D:\My Documents\ProjDocuments\Paper 1\
LaneWiseRelationshipsForAllLocations.xls"
            DBMS=EXCEL2000 REPLACE;
     GETNAMES=YES;
run;
proc anova data=anovauk;
class density1to6;
model speed1to6=density1to6;
run;
proc anova data=anovauk;
class density8to10;
model speed8to10=density8to10;
run;
```

APPENDIX F

RELATIONSHIPS BETWEEN TIME AND SPACE SPEED STATISTIC

Time Mean Speed from Space Mean Speed

Consider that the total stream is made up of *m* substreams, with each substream having its own speed. Let

$\bar{u}_t$ = average time mean speed

$\bar{u}_s$ =average space mean speed

q_i = flow of the ith substream

u_i = speed of the ith substream

k_i = concentration of the ith substream

q = flow of the total traffic stream

u = mean speed of the total traffic stream

k = concentration of the total traffic stream

σ_s^2 = variance about space mean speed

σ_t^2 = variance about time mean speed

Using the method of Wardrop, segregate total flow into m subflows by speed. Define

$$\bar{u}_t = \frac{\sum_{i=1}^{m} q_i u_i}{\sum_{i=1}^{m} q_i} = \frac{\sum q_i u_i}{q} \qquad \textbf{(1.1)}$$

and

$$\bar{u}_s = \frac{\sum_{i=1}^{m} k_i u_i}{\sum_{i=1}^{m} k_i} = \frac{\sum k_i u_i}{k} \qquad \textbf{(1.2)}$$

But

$$q_i = k_i u_i \qquad \textbf{(1.3)}$$

Thus,

$$\bar{u}_t = \frac{\sum (k_i u_i) u_i}{q} = k \frac{\sum k_i u_i^2}{kq} = k \sum \frac{f_i^{'} u_i^2}{q} \qquad \textbf{(1.4)}$$

where

$$f_i^{'} = \frac{k_i}{k} \qquad \textbf{(1.5)}$$

But

$$q = k\,\bar{u}_s \qquad \textbf{(1.6)}$$

Thus

$$\bar{u}_t = k \sum \frac{f_i^{'} u_i^2}{k\,\bar{u}_s} = \frac{1}{\bar{u}_s} \sum f_i^{'} \left[\bar{u}_s + \left(u_i - \bar{u}_s \right) \right]^2$$

$$= \frac{1}{\bar{u}_s} \left[\sum f_i^{'} \bar{u}_s^{\,2} + 2\bar{u}_s \sum f_i^{'} \left(u_i - \bar{u}_s \right) + \sum f_i^{'} \left(u_i - \bar{u}_s \right)^2 \right] \qquad \textbf{(1.7)}$$

But

$$\sum f_i^{'}\left(u_i - \bar{u}_s\right) = 0 \qquad \textbf{(1.8)}$$

by definition of mean and

$$\sum f_i^{'}\left(u_i - \bar{u}_s\right)^2 = \sigma_s^2 \qquad \textbf{(1.9)}$$

$\sum f_i^{'}\left(u_i - \bar{u}_s\right)^2 = \sigma_s^2$ by definition of variance about the space mean speed; therefore

$$\bar{u}_t = \frac{1}{\bar{u}_s}\left[\bar{u}_t^{\,2} + 0 + \sigma_s^2\right]$$

$$= \bar{u}_s + \frac{\sigma_s^2}{\bar{u}_s} \qquad \textbf{(1.10)}$$

Derivation in the Continuous Case

Define

$$\bar{u}_s = \int_0^\infty u f_s(u)\,du \qquad \textbf{(1.11)}$$

$$\bar{u}_t = \int_0^\infty u f_t(u)\,du \qquad \textbf{(1.12)}$$

where

$f_t(u)$ = speed density in time and

$f_s(u)$ = speed density in space.

An important relationship given in Haight and Mosher but proved by Breiman is:

$$\bar{u}_s\ f_t(u) = u f_s(u) \qquad \textbf{(1.13)}$$

After multiplying both sides by *u* and integrating over the entire range of *u*:

$$\bar{u}_s = \int_0^\infty u f_t(u)\,du = \int_0^\infty u^2 f_s(u)\,du$$

$$\bar{u}_s \bar{u}_t = \int_0^\infty u^2 f_s(u)\,du \qquad \textbf{(1.14)}$$

Define

$$\sigma_s^2 = \int_0^\infty u f_s(u)\,du - \left(\bar{u}_s\right)^2 \qquad \textbf{(1.15)}$$

Substituting Eq.(1.14) in Eq. (1.15) gives

$$\sigma_s^2 = \bar{u}_s \bar{u}_t - \left(\bar{u}_s\right)^2 \qquad \textbf{(1.16)}$$

and

$$\bar{u}_t = \bar{u}_s + \frac{\sigma_s^2}{\bar{u}_s} \qquad \textbf{(1.17)}$$

Relationship Between Arithmetic Mean and Harmonic Mean

To examine this computation, consider the arithmetic and harmonic means without any context of traffic. Define

$M = \frac{1}{N}\sum_{i=1}^{N} X_i$ = arithmetic mean

$V_m = \frac{1}{N}\sum_{i=j}^{N} \left(X_i - M\right)^2$ = variance about arithmetic mean

$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{X_i}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{X_i}}$ = harmonic mean

Expand in Taylor series:

$$\frac{1}{X_i} = A_0 + A_1\left(X_i - M\right) + A_2\left(X_i - M\right)^2 + A_3\left(X_i - M\right)^3 + \ldots$$

Evaluate constants by differentiating,

$$A_J = \frac{1}{j!}\frac{d^j}{dX^j}\left(\frac{1}{X_i}\right)| M$$

for

j = 0,1, …

$$\frac{1}{X_i} = \frac{1}{M} - \frac{1}{M^2}(X_i - M) + \frac{1}{M^3}(X_i - M)^2 - \frac{1}{M^4}(X_i - M)^3 + \ldots \qquad \textbf{(1.18)}$$

(Note: eq. (1.18) converges for 0<x<2M, which is usually the caser for traffic.)

Then

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{X_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M} - \frac{1}{NM^2}\sum(X_i - M)$$
$$+\frac{1}{NM^3}\sum(X_i - M)^2 - \frac{1}{NM^4}\sum(X_i - M)^3 + \frac{1}{NM^5}\sum(X_i - M)^4 + \ldots$$

By the definition of the arithmetic mean

$$\sum_{i=1}^{N}(X_i - M) = 0\,.$$

Similarly, for any distribution that is approximately symmetrical

$\sum_{i=1}^{N}(X_i - M)^a \approx 0$ for odd values of a.

Thus

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{X_i} = \frac{1}{NM}(N) - \frac{1}{NM^2}(0)$$
$$+\frac{1}{M^3}(V_m) - \left(\frac{1}{M^4 N}\right)(0) + \frac{1}{M^5 N}\sum(X_i - M)^4 \ldots$$

It can be assumed that

$M^5 N >> \sum(X_i - M)^4$ since eq. (1.18) is converging.

Thus the last term can be neglected, as can all later terms in the expansion.

Then

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{X_i} = \frac{1}{M} + \frac{V_m}{M^3} = \frac{M^2 V_m}{M^3}$$

$$H = \frac{1}{\left(M^2 + V\right)/M^3} = \frac{M^3}{M^2 + V_m} = \frac{M}{1 + \frac{V_m}{M^2}}$$

$$= \frac{M\left(1 - \frac{V_m}{M^2}\right)}{\left(1 + \frac{V_m}{M^2}\right)\left(1 - \frac{V_m}{M^2}\right)}$$

$$= \frac{M\left(1 - \frac{V_m}{M^2}\right)}{\left(1 - \frac{V_m{}^2}{M^4}\right)} \approx M\left(1 - \frac{V_m}{M^2}\right) = M - \frac{V_m}{M}$$

Converting to traffic notation gives:

$$\bar{u}_s = \bar{u}_t - \sigma_t^2 / \bar{u}_t \qquad \textbf{(1.19)}$$

as an approximate method for use in traffic engineering practice. Note that this relationship when combined with eq. (1.19) implies that

$$u_s / u_t = \sigma_s^2 / \sigma_{ts}^2$$

Thus, in using eq (1.19) one must be willing to accept this assumption.