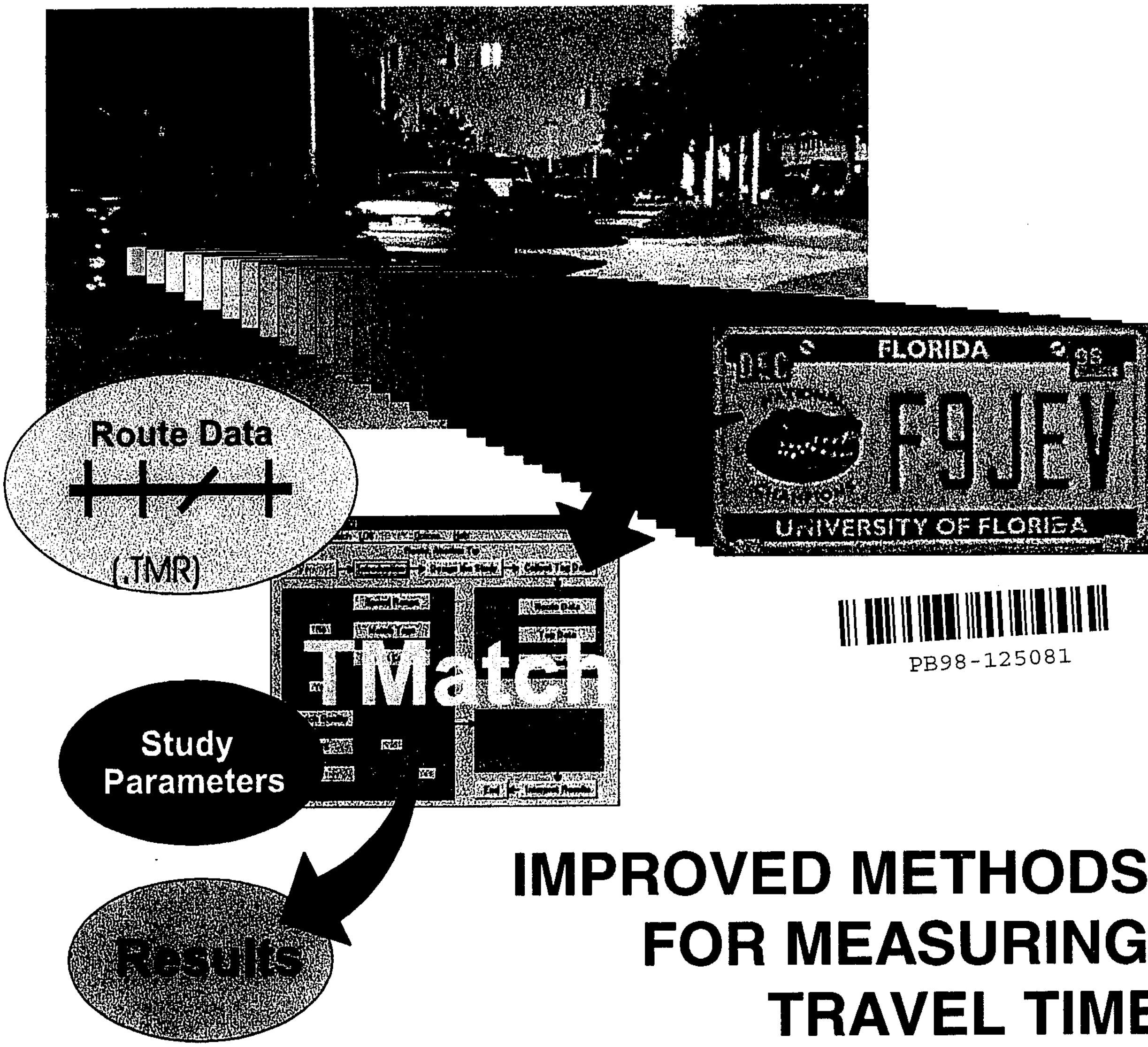


IMPROVED METHODS FOR MEASURING TRAVEL TIME ON ARTERIAL STREETS

Final Report

State Project 99700-3519-119
Contract B-8355
WPI 0510674
UF Project 4910450445312

UNIVERSITY OF FLORIDA
TRANSPORTATION RESEARCH CENTER

January 1998

EXECUTIVE SUMMARY

The State of Florida has invested a substantial amount of time and effort in its growth management policies and procedures. Current procedures require a periodic review of arterial levels of service (LOS) based on the average travel time of traffic on an arterial segment. Direct measurement of travel time is a useful method, but involves a high degree of statistical variation due to the nature of field measurements. Because of this variability, several observations are required to establish the required confidence in the data.

The time and expense involved in logging data using a moving vehicle study makes it impossible to obtain large data sample sizes economically. This causes an uncertainty as to the true travel time along an arterial route. To remedy this problem, it is necessary to develop new techniques for travel time and delay studies that will be more productive than moving vehicle studies.

The principal product of this project is an automated technique for collecting, adjusting, matching and analyzing license tag data to produce statistically valid estimates of travel time. Several independent investigations were carried out in support of the final product:

- A variety of travel-time data collection methods was considered, including some advanced technology applications such as global positioning systems (GPS), automated license tag scanning and on-board transponders. These applications are interesting in their own right, but they were not explored in depth under this project, because they offered no real productivity, cost or accuracy advantage for short-term studies of this nature.

- A Windows-based tag matching program called **"TMatch"** was developed. The primary advantage of the tag-matching technique is that it produces much larger sample sizes than the more conventional moving vehicle studies. This advantage is achieved at the cost of a somewhat more complex study that requires additional planning and organization that some agencies would prefer to avoid.

- Voice recognition technology was explored as a means of improving the data collection process. It was concluded that this technology is not yet ready for economical application in an outdoor environment with traffic noise. Direct entry of the tag data in the field using palmtop computers proved to be considerably more productive, and more accurate, than voice recognition.

- A sensitivity analysis was performed using simulation, with controlled errors introduced into the simulated tag data.

- An exponential approximation technique was proposed for adjusting measured travel times to reflect traffic volumes different than field study volumes. This technique proved to be useful within a limited range of traffic volume variation. The technique loses accuracy when volumes approach capacity and when green times are very long with respect to the cycle time.

- A pilot study was carried out to demonstrate the application of the TMatch program, and to compare the results with moving-vehicle travel time studies.

- Study guidelines were developed for sample selection. The human observer does not have the ability to record every license plate that passes his or her station in peak-period traffic. For this reason, it is necessary to develop selection criteria for the tag samples. The proposed selection criteria included recording of license tags from solid white passenger vehicles with standard State of Florida license tags only. The practice of recording only the last four digits of each tag was also recommended.

The **TMatch** program developed under this project is a tag-matching tool. That is, it takes lists of license tag numbers observed at a maximum of six stations along a roadway and compares them to find any matches (tags which were observed at both stations). This is made more interesting by recording not just the tags, but also the times at which they were observed. With this information, TMatch can compute a speed for each matched tag pair. Averaging these speeds gives an estimate of the average vehicle speed over the link of roadway between the stations.

When tags are observed at more than two stations, the matching can be performed on each pair of stations, and the results pooled to give a more robust estimate of the average speed between the first and last stations. This is the function of the TMatch software.

Several user-specified parameters may be entered to control the tag matching process, including minimum and maximum allowable speeds and the number of tag digits to be processed. An assessment of the overall confidence in the match results is also provided.

Compared to moving-vehicle travel time studies, enormous sample sizes may be obtained with license tag data matching. Nevertheless, moving-vehicle studies are still the most predominant method for collecting this type of data. The lack of a productive analysis program is one factor that has limited the use of tag-matching studies. That problem should be largely overcome by the results of this project.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3.
FL/DOT 99700-3519-119		PB98-125081

4. Title and Subtitle	5. Report Date
IMPROVED METHODS FOR MEASURING TRAVEL TIME ON ARTERIAL STREETS	January 1998
	6. Performing Organization Code
7. Author(s) Kenneth Courage, Randall Showers, James Harriot, William Schilling and Kevin Godbey	8. Performing Organization Report No. 4910450445312
9. Performing Organization Name and Address Transportation Research Center University of Florida 512 Weil Hall PO Box 116585 Gainesville, FL 32611-6585	10. Work Unit No. (TRAIS) 11. Contract or Grant No. B-8355
9. Sponsoring Agency Name and Address Florida Department of Transportation Research Center 605 Suwannee Street, M.S. 30 Tallahassee, FL 32301-8064	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code 99700-3519-119

15. Supplementary Notes

Prepared in cooperation with the U.S. Department of Transportation

16. Abstract

The time and expense involved in logging data using a moving vehicle study makes it impossible to obtain large data sample sizes economically. This causes an uncertainty as to the true travel time along an arterial route. To remedy this problem, it is necessary to develop new techniques for travel time and delay studies that will be more productive than moving vehicle studies.

The principal product of this project is an automated technique for collecting, adjusting, matching and analyzing license tag data to produce statistically valid estimates of travel time. The **TMatch** program was developed as a tag-matching tool. It takes lists of license tag numbers and times observed at a maximum of six stations along a roadway and compares them to find any matches (tags which were observed at both stations). With this information, **TMatch** computes a speed for each matched tag pair. Averaging these speeds gives an estimate of the average vehicle speed over the link of roadway between the stations. Several user-specified parameters may be entered to control the tag matching process, including minimum and maximum allowable speeds and the number of tag digits to be processed. An assessment of the overall confidence in the match results is also provided

Compared to moving-vehicle travel time studies, enormous sample sizes may be obtained with license tag data matching. Nevertheless, moving-vehicle studies are still the most predominant method for collecting this type of data. The lack of a productive analysis program is one factor that has limited the use of tag-matching studies. That problem should be largely overcome by the results of this project.

17. Key Words	18. Distribution Statement	
Travel time, License tag matching, Moving vehicle studies.	No restrictions.	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	98	

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

ACKNOWLEDGMENTS

This report was prepared by the University of Florida Transportation Research Center in cooperation with the State of Florida Department of Transportation and the U.S. Department of Transportation.

The project was coordinated by Prof. Kenneth G. Courage. Significant technical contributions were made by Randall Showers, James Harriot, Kevin Godbey and William Schilling. The technical coordinator for the Florida Department of Transportation was Lap Hoang, P.E.

DISCLAIMER

The opinions, findings and conclusions expressed in this report are those of the authors and not necessarily those of the Florida Department of Transportation or the U.S. Department of Transportation.

Table of Contents

Page

1. INTRODUCTION 1
    1.1. Problem Statement 1
    1.2. Project Objectives 1
    1.3. Summary of Activities and Results 2

2. METHODS OF MEASURING TRAVEL TIME 2
    2.1. Moving Vehicle Travel Time Studies 4
    2.2. License Tag Matching Studies 5

3. LICENSE TAG DATA COLLECTION 5
    3.1. License Tag Sampling Criteria 5
    3.2. Data Collection Techniques 6
    3.3. Limitations of Tag-Matching Studies 9
    3.4. Bias in Tag-Matching Studies 9

4. THE TMATCH PROGRAM 9
    4.1. TMatch File Structure and Information Flow 10
    4.2. TMatch Data Entry and Edit Screens 10
    4.3. TMatch Output Reports 10

5. TESTING OF TMATCH BY SIMULATION 12
    5.1. Effects of the Quality of Data 13
    5.2. Effects of Programmed Stops 13

6. FIELD TEST RESULTS 17

7. ADJUSTING TRAVEL TIMES FOR DIFFERENT VOLUME CONDITIONS . . 19

8. CONCLUSIONS AND RECOMMENDATIONS 19

APPENDIX A: Sample Size Requirements for Moving-Vehicle Travel Time Studies A-1

APPENDIX B: License Tag Matching with TMATCH B-1

APPENDIX C: Voice Recognition Studies C-1

APPENDIX D: Adjustment of Travel Times for Different Traffic Volume Conditions D-1

IMPROVED METHODS FOR MEASURING TRAVEL TIME ON ARTERIAL STREETS

1. INTRODUCTION

The State of Florida has invested a substantial amount of time and effort in its growth management policies and procedures. Included in this is a requirement for periodic review of arterial levels of service (LOS). The main LOS evaluation criterion is the average speed of traffic on an arterial segment. This can be estimated from the Highway Capacity Manual (HCM) model[1] based on measured arterial traffic volumes and signalized intersection capacities.

Accurate travel time estimates are critical for engineering studies and governmental comprehensive planning. It is often necessary to conduct travel time studies for such purposes as evaluating traffic improvements, verifying and calibrating the HCM estimation procedures, etc. Moving vehicle studies, as outlined in the Florida Department of Transportation's Manual on Uniform Traffic Studies (MUTS)[2], have proven to be an effective tool for obtaining such data.

1.1 Problem Statement

Direct measurement of travel time is a valuable tool in collecting travel time data, but does involve a degree of statistical variation due to the nature of field measurements. Because of this degree of variation, multiple observations are required to determine an interval of confidence in the data. The time and expense involved in logging data using a moving vehicle study makes it impossible to obtain large data sample sizes economically. This causes an uncertainty as to the true travel time along an arterial. To remedy this problem, it is necessary to explore new techniques for travel time and delay studies that will be more productive than moving vehicle studies. The new data collection methods must provide large samples of travel time data economically.

1.2 Project Objectives

The objectives of this project are as follows:

- Determine confidence intervals and sample size requirements for arterial segments from existing travel time and delay data from Dade and Broward Counties;

- Assess the feasibility of license tag matching as a more economical and productive alternative to moving vehicle studies and

- Assess the feasibility of voice recognition systems for improving the productivity of license tag data collection.

1.3 Summary of Activities and Results

The principal product of this project is an automated technique for collecting, adjusting, matching and analyzing license tag data to produce statistically valid estimates of travel time. Several independent investigations were carried out in support of the final product:

- A variety of travel-time data collection methods was examined.
- Sample size requirements were addressed through analysis of field data collected in Dade and Broward Counties. The results of this study are presented in Appendix A.
- A Windows-based tag matching program called **"TMatch"** was developed. Documentation for this program is presented in Appendix B.
- Voice recognition technology was explored as a means of improving the data collection process. It was concluded that this technology is not yet ready for economical application in an outdoor environment with traffic noise. The detailed studies are described in Appendix C.
- A technique was proposed for adjusting measured travel times to reflect traffic volumes different than field study volumes. This technique is examined in Appendix B.
- A pilot study was carried out to demonstrate the application of the TMatch program, and to compare the results with moving-vehicle travel time studies.

All of these activities and results will be summarized in the body of this report, which has been kept intentionally short to facilitate reading and comprehension of the major points. The study details are concentrated in the appendices for those who seek more depth.

2. METHODS OF MEASURING TRAVEL TIME

The measurement of travel time over an arterial segment involves a determination of when vehicles enter and leave the segment. It is neither necessary nor practical to collect this information for each vehicle using the segment. On the other hand, a single observation will not provide a valid representation of travel time for any useful purpose.

The number of required observations depends on the desired degree of precision and confidence in the results. As an example, Table 1 provides an estimate from the literature of the number of observations required to achieve a 95% level of confidence.

Table 1. Sample size to predict overall speed with 95% confidence.

TYPE OF FACILITY	NUMBER OF REQUIRED OBSERVATIONS
Signalized Urban Streets	
Two-lane, uncongested	32
Two-lane, congested	36
Multi-lane, uncongested	80
Multi-lane, congested	102
Rural Highways	
Two-lane, 1,130 vph	25
Two-lane, 1,440 vph	41
Four-lane uncongested	30

(Source: Reference 3)

Observations may be obtained either internally by stationing the observer inside of a test car, or externally by observing the times that vehicles enter and leave the study area. Moving vehicle studies are very easy to conduct, but offer very low productivity because of the time it takes one vehicle to traverse the study area. Such studies are also subject to the biases of the individual test drivers. The main advantage of moving vehicle studies is that they can produce more intensive measures of trip quality and cost, including number and duration of stops, stopped delay, and fuel consumption. Various comfort measures derived from speed and acceleration profiles have also been proposed.

External observation studies are more difficult to conduct because they require coordination of data from multiple points. On the other hand, they have the advantage of producing significantly more observations during a peak period, and this advantage is very important to the productivity of a travel time study. They do not offer the same assessment of trip cost and comfort as moving vehicle studies, but they add a potential byproduct in the form of origin-destination information if sufficient observers are placed in the field.

External observation studies have the advantage of eliminating the behavior of test vehicle drivers as a source of potential bias. On the other hand, a new source of bias is introduced due to programmed stops. Short duration stops (e.g., convenience store) will be included in the overall travel time of vehicles that are simply observed entering and leaving the study area. This can introduce significant problems if a large proportion of vehicles makes programmed stops.

2.1. Moving Vehicle Travel Time Studies

Moving vehicle studies are currently the most common way of estimating travel time on an arterial route. The techniques for conducting moving vehicle studies in Florida are prescribed in Reference 2. Moving vehicle studies will not be belabored in this report because the objective of the project described herein was to develop more productive alternatives to such studies.

In their simplest form, moving vehicle studies use only a driver and a stopwatch to measure travel time. It is common practice, however, to use some form of on-board instrumentation to automate the data collection process to increase some combination of productivity, accuracy or scope of the study. The Moving Vehicle Run Analysis Package (MVRAP) [4] is an example of a system that uses a transmission sensor to note the distance traveled by the vehicle during each second. This information is used to develop speed profiles from which several measures are derived. MVRAP also offers the productivity of a data base manager that facilitates the analysis of multiple runs to produce statistically-significant results.

Global Positioning System (GPS) technology has also been applied to travel time studies. Using GPS, a continuous (but approximate) record of the coordinates of a vehicle may be maintained and analyzed. A prototype GPS system has been developed by the Mitre Corporation for FHWA [5]. A practical field demonstration of GPS technology was carried out in three metropolitan areas of Louisiana by Quiroga and Bullock [6]. The same researchers also investigated the concept of combining GPS and Geographic Information System (GIS) technologies together in a spacial model [7].

Another application of instrumented vehicles may be found in the use of "probe" vehicles in advanced traffic management systems (ATMS). These vehicles are provided with the capability to advise a central location of their position periodically. The nature of the instrumentation and use of the information is specific to each system. While this technology may revolutionize travel time studies at some point in the future, it holds minimal interest at this time for the improvement of simple measurement techniques.

The advanced technology applications are interesting in their own right, but they were not explored in depth under this project, because they offer no productivity, cost or accuracy advantage compared to the simple on-board devices that use transmission sensors to measure elapsed distance directly along the route.

2.2. Licence Tag Matching Studies

License tag matching studies have been implemented for many years in obtaining origin/destination and travel time data. License tag studies are quite accurate for collecting travel time data due to the ability to obtain a large sample of travel times. This increase in accuracy is at the sacrifice of delay data along the study corridor. Typically, license tag data are collected at the beginning and end of the study corridor and possibly at a few critical intermediate locations. Usually, the last three or four digits of an automobile's license tag and a corresponding time are recorded at the specified nodes along the study.

The remainder of this discussion will focus on the developments of this project with respect to travel time estimation by license tag matching procedures.

3. LICENSE TAG DATA COLLECTION

The two main tasks in a license tag study are the collection of the tag data in the field and the processing of these data in the office. Since these tasks are more or less independent, they will be addressed separately. The data collection task will be covered in this setion.

3.1. License Tag Sampling Criteria

The human observer does not have the ability to record every license plate that passes his or her station. For this reason, it is necessary to develop selection criteria for the tag samples. If this is not done, and samples are collected at random, matching will be much less likely. For example, if one observer collects 10% of the license tags at random and the second observer (upstream) collects 10% of the tags at random, one would expect to match 1% of the tags. Had a selection criteria been set, and each observer collected the same 10% of data, then there should be a 10% match expected.

The most obvious way to select automobile samples is by color. People collecting data can determine a car's color from a long distance and do not have to make split second decisions as to whether or not a car's license plate should be sampled. From observation, white seemed to be a popular color for automobiles and easiest to distinguish because of fewer subtle shades, therefore it was chosen as the selection criterion for sampling.

To determine if this would provide an adequate sample, a preliminary field survey was conducted for an hour at an intersection in northwest Gainesville. During that hour, 648 cars were counted and of that amount, exactly 100 cars were white passenger vehicles. This roughly translates into an expected sampling equal to 15.4% of the traffic volume within a study corridor. During the preliminary survey, this turned out to be a reasonable sampling rate for a single observer recording license data, therefore the following selection criteria were adopted:

- Recording of license tags from solid white passenger vehicles. This includes solid white pickups, vans, and sport utilities that are not commercial vehicles.

- Recording of standard State of Florida license tags only. All out-of-state, vanity, and specialized plates are to be disregarded. (i.e. - University of Florida license plates were not recorded.)

- Recording of the last four digits of the license plate. If one of the digits was missed, a wild card character was entered in place of the missing data.

3.2. Data Collection Techniques

Several alternatives exist for collecting and recording the license tag data. The common requirement of all techniques is that the information (tag number and arrival time) must eventually appear in electronic format. For maximum flexibility, the format chosen for this project was a simple ASCII text file with a separate line for each tag entry. On each line, the tag number and entry time are separated by a comma. An example follows:

```
Q56U, 16:05:49
Q54U, 16:05:56
D61M, 16:06:09
I10X, 16:06:34
E42J, 16:08:07
Q10Q, 16:09:17
B49E, 16:10:49
A28W, 16:10:55
U94M, 16:10:59
Q72A, 16:11:03
```

There are several ways that this file may be produced, some of which are more automated than others. Figure 1 illustrates the options that were considered in connection with this study.

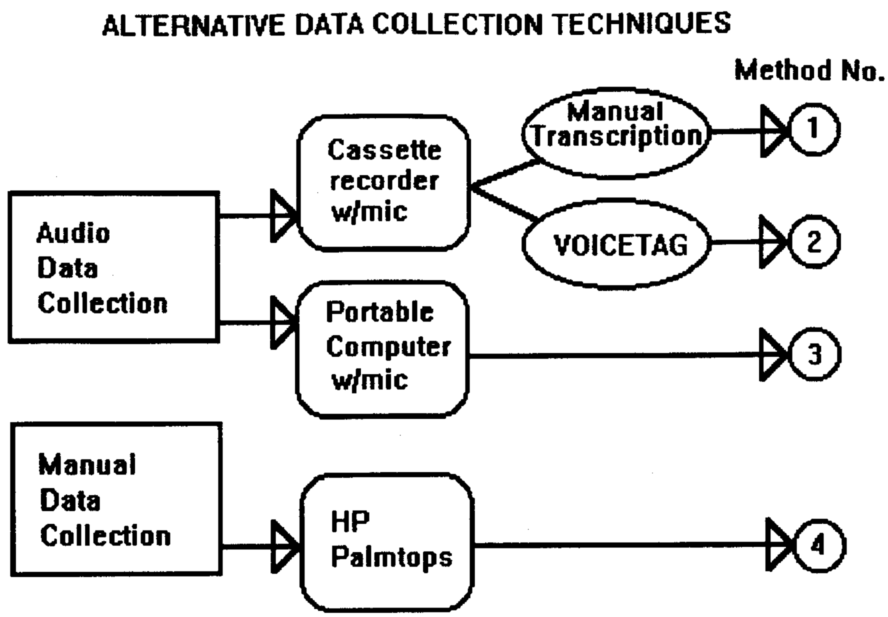


Figure 1. Alternative data collection methods used for field studies.

Note that none of the alternatives included full manual data entry and transcription of the data because of the laborious process involved. The two methods of collecting the basic tag data were manual and voice-recorded. The only option considered for the manual recording was a palmtop computer that stored the tag numbers directly as they were entered from the keyboard.

Three alternatives were examined for voice-recorded data:

- Transcription of the data into a personal computer during playback: In this case the tape was played continuously to preserve the original time base. Audible time checks were added to the tape at regular intervals to maintain synchronization with the computer.

- Playback of the tape units into a voice recognition program: This offers a higher degree of productivity, but, as explained in more detail in Appendix C, did not prove to be technically feasible.

- Recording of the voice data directly into a portable notebook computer: This avoids the signal degradation involved in the tape recording step, however other problems (also discussed in Appendix C) arose that eliminated this technique as a feasible alternative.

This left two candidate data collection techniques for further study: manual transcription of the tape recorded data, and direct entry of the tag numbers into a palmtop computer. A pilot study was conducted to determine the relative data capture rates. The results of this study, as illustrated in Table 2, indicated that both techniques offered approximately the same capture rates of approximately 400 tags per hour.

Table 2. Summary of maximum data capture rates.

Time Period (30 minutes)	Palmtops		Cassette Recorders	
	Observer 1	Observer 2	Observer 3	Observer 4
1	251	171	169	219
2	251	156	185	255
3	240	163	178	242
4	244	136	161	229
Subtotal	986	626	693	945
Total	1612		1638	
Tags/hour	**403**		**410**	

The advantage of the palmtop computers is one of productivity, because the tag data are placed directly into the text file for processing. The voice method, on the other hand, involves a lower equipment cost. Palmtop computers cost about $500 each, while an adequate portable tape recorder can be obtained for about $50.

Note that it is not necessary to use the same data collection technique at all stations because the file format required as an output from both techniques is identical. Palmtop computers may be used to the extent of their availability, with augmentation by audio recorders for multi-station studies.

This is another area that is served by advanced technology. Several vendors market video image processing devices that claim the ability to recognize license tags. It is quite conceivable that the tag data file could be produced in this manner. This alternative was not pursued as a study task because it offers no productivity advantage over a manual observer with a palmtop computer.

The deployment of video image processing for tag data collection would require a field crew at least as large, and probably larger, than the one-person crew required to collect the data manually.

3.3. Limitations of Tag-Matching Studies

Compared to moving-vehicle travel time studies, enormous sample sizes may be obtained with license tag data matching. Nevertheless, moving-vehicle studies are still the most predominant method for collecting this type of data. The lack of a productive analysis program is one factor that has limited the use of tag-matching studies. That problem should be largely overcome by the results of this project.

There are still other problems that discourage the use of tag studies. One is the sheer complexity and organization required to deploy a team of several observers to collect simultaneous data at different places. Another is the vulnerability of the study to the loss of data from one station. Still another is the sensitivity to tag errors and random matches that give erroneous results.

3.4. Bias in Tag-Matching Studies

Even a perfectly planned and executed study can produce inaccurate estimates of speed and travel time because of programmed stops that occur when commuters break their journey to perform errands of one kind or another (convenience store stops, school pickups and dropoffs, etc.). If the duration of a programed stop places the travel time beyond a reasonable range, the trip should be rejected by the matching software. However, if the duration is within the natural variability of trip times, the length of the programmed stop will be included in the observed travel time, resulting in a bias to the high side. An investigation of this phenomenon will be discussed later in this report.

4. THE TMATCH PROGRAM

The TMatch program is one of the principal products of this study. The complete documentation for the program is presented in Appendix B.

TMatch is a tag-matching tool. That is, it takes lists of license tag numbers observed at a maximum of six observation stations along a roadway and compares them to find any matches (tags which were observed at both stations). This is made more interesting by recording not just the tags, but also the times at which they were observed. With this information stored in files with the proper format, TMatch can compute a speed for each matched tag pair. Averaging these speeds gives an estimate of the average vehicle speed over the link of roadway between the stations.

When tags are observed at more than two stations, the matching can be performed on each pair of stations, and the results pooled to give a more robust estimate of the average speed between the first and last stations. This is the function of the TMatch software.

The estimated speed is of course statistical in nature, and thus naturally calls for associated estimates of error and reliability. TMatch output includes several such measures.

4.1. TMatch File Structure and Information Flow

The file structure and information flow is illustrated in Figure 2. Note that there are four files that contain information related to the tag matching process.

- Tag Data File (TMI): Contains a list of the tag numbers and observed times;
- Route Data File (TMR): Contains the lengths, speeds and other necessary characteristics of the route being studied;
- Study Parameter File (PAR): Contains the parameters that govern the matching process for a particular study; and
- Output File (TMO): Contains a summary of the results of the study.

4.2. TMatch Data Entry and Edit Screens

The main menu screen for TMatch is shown in Figure 3. The user interface for this program is unique in the sense that it provides a flow chart of the entire data analysis process, with each step selected by a mouse-click. Several screens are displayed for entry and editing of data, specification of tag-matching ranges and other parameters and for viewing of the results. Each of these screens is described in Appendix B.

4.3. TMatch Output Reports

TMatch produces a very detailed output report that describes the inputs, operating parameters, match confidence, etc. in addition to the travel time summary. All of the sections of the output report are presented in detail in Appendix B. Figure 4 shows a sample of the travel time summary results.

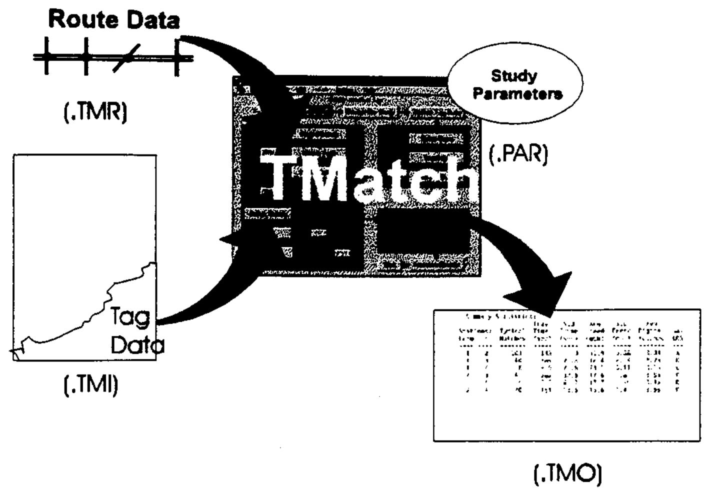


Figure 2. TMatch file structure and information flow.

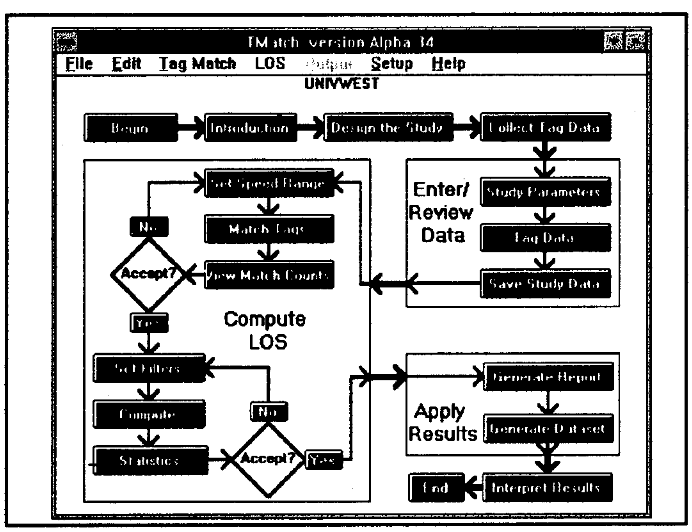


Figure 3. TMATCH Windows based Graphical User Interface Screen.

```
Study Results:
                        Adj.   Std    Adj    Std
Stations:     Adj   Time   Error  Speed  Error  Adj   LOS
From  To      N     (sec)  (sec)  (mph)  (mph)  LOS  Signif
----  ----  -----  -----  -----  -----  -----  ---  ------
  1     2    151    168    8.4   27.2   1.44    B   0.8344
  2     3    103    431   22.2   15.7   0.85    D   0.5597
  3     4     92    137   10.8   18.4   1.57    C   0.5117
  1     4    274    735   12.9   18.8   0.34    C   0.8585
```

Figure 4. Sample TMatch travel time summary report.

5. TESTING OF TMATCH BY SIMULATION

TMatch was tested extensively by simulation to ensure a workable product. Several sets of simulated tag data were generated by a program called TAGSIM that introduced random components into the essential parameters of a tag-matching study.

A sample TAGSIM parameter file follows:

```
TAGSIM Variables
Each line must contain a string , value
1. Starting time- 24 hr format- even 1/2 hour-  [TStart$],11:30
2. Period length- multiples of 1/2 hour -        [Period],1800
3. No of vehcles to generate                     [NumVeh], 100
4. Link length - same for all links- Feet      [LinkLen!], 2640
5. Running Speed - mph                         [RunSpeed], 50
6. Min cruise delay - sec                    [MinLinkDel], 0
7. Max cruise delay - sec                    [MaxLinkDel], 50
8. Random cruise delay Increment- sec               [LDI], 30
9. Probability of a programmed stop per link -  [PPStop!], .3
10. Min programmed stop time - sec-           [PStopTime], 100
11. Probability of completing trip on each link   [PCTL!], .9
12. Number of tag characters                     [NTChar], 4
13. Tag sequence specification - A # or B-      [TSpec$], BBBB
14. Probability of miss                          [PMiss!], .0
15. Probability of Wild Card                     [PWild!], .0
16. Probability of Error                          [PErr!], .00
17. Base file name for TMATCH data file *.TMI [BaseName$], chk1
18. Number of stations to create tag files    [NumOfStat],4
```

All of the parameters listed in this file may be changed to simulate a wide variety of randomized elements including:

- Variability of trip time and delay,
- Quality of input data (misses, errors, wild cards, etc.),
- Number of tag characters captured and
- Length and probability of programmed stops.

TAGSIM also generates its own output summary that represents the simulated measures for comparison with the TMatch output. An annotated example of the TAGSIM output report is presented in Figure 5.

5.1. Effects of the Quality of Data

The quality of the input data may be represented by the proportion of errors, misses and wildcards. A reduction in the quality of data should not bias the results, but it will reduce the confidence in the results, necessitating larger sample sizes.

The effects of each of these parameters was assessed independently by simulation. An index of match quality was first developed, with a value of 100 indicating a perfect match of all tags, and zero indicating no matches at all. The effect of each of the data quality parameters is shown in Figures 6, 7, and 8, which represent the proportion of errors, misses and wildcards, respectively.

Inspection of these figures indicates that errors have the greatest influence on match quality, followed closely by misses, and then wildcards.

5.2. Effects of Programmed Stops

It has already been pointed out that programmed stops can bias the travel times upwards, causing a tag-matching study to overestimate travel times and underestimate speeds. This effect was also investigated by simulation, with the results presented in Figures 9, 10 and 11. Figure 9 shows the bias effect, assuming that all of the vehicles with programmed stops were included in the tag-matching process. Three conditions are depicted, representing programmed stops of 1%, 5% and 10%. The length of programmed stop was varied from very short to 25% of the route travel time.

An effective matching program should be able to identify programmed stops and to reject tag matches from the associated vehicle. The ability of TMatch to accomplish this objective is demonstrated in Figure 10, which reflects 10% programmed stops, and Figure 11, which represents only 1% programmed stops.

```
=====================================================================
   SUMMARY         Link 1  Link 2  Link 3  Link 4  Link 5   Total
=====================================================================
Total TT sec/veh    374.6   367.5   386.4   363.6   358.9  1850.3
Overall Spd mph      19.2    19.6    18.6    19.8    20.1    19.5
Ave. Prg. stop      511.4   454.3   532.0   571.8   562.3   520.8
# of Prg. stops      26      24      29      16      17      112
Adj means programmed stops are excluded

'Adjustment excludes any tags with a programmed stop

Adj TT sec/veh      325.3   328.7   327.7   329.3   324.0  1635.1
Adj Speed mph        20.0    20.0    19.5    20.6    20.8    20.2
Varaince (sec)     1675.0  1709.0  1594.4  1706.2  1962.1 17854.0

'The varaince is link specific.  The total column is for trips traveling
'from link 1 to link 5 or the last link.  The same goes for the Standard
'Deviation.

Std. Deviation       40.9    41.3    39.9    41.3    44.3   133.6
Complete Link        270     281     263     267     274     179

'Number of vehicles completing the link.
'Total is usually less than probability of not completing a trip (NCT).
'Worst case is Number of links * NCT * Number of Vehicles.
'Once a tag leaves the route it never returns and since the probaility
'is checked for each link for each vehicle independently then chances of
'completing the entire route is far less than that stated in
'the input.

Missed Tag             61      62      61      57      51      54
'Link column: If either start or end station is missed then counted.
'Total column is counted if the first or last link is missed.

'If a tag is missed then it is excluded from the wild or error counts
'because it no longer exists as a tag.
Wild Card in Tag       49      51      52      49      34      36

'Link column: If either start or end station is Wild then counted.
'Total column is counted if the first or last link is Wild.

Error in Tag           50      44      50      51      47      49

'Link column: If either start or end station is Error then counted.
'Total column is counted if the first or last link is Error.
=====================================================================
```

Figure 5. Sample TAGSIM output with annotation.

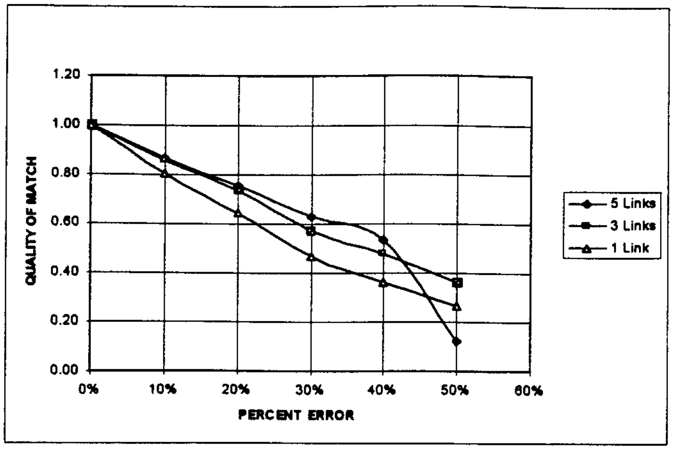


Figure 6. Effect of errors on match quality.

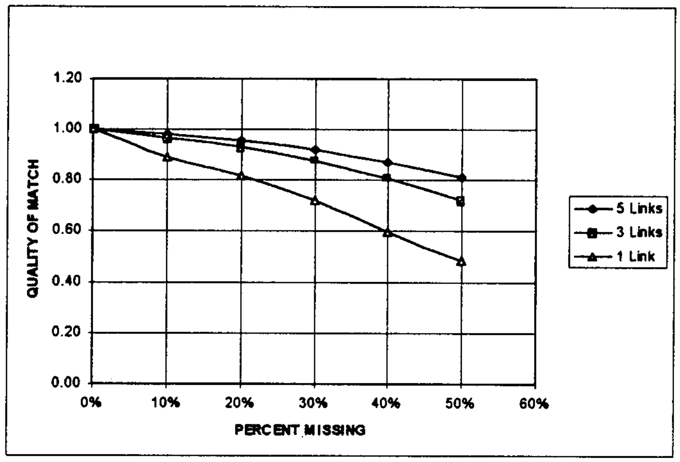


Figure 7. Effect of Missed tags on match quality.

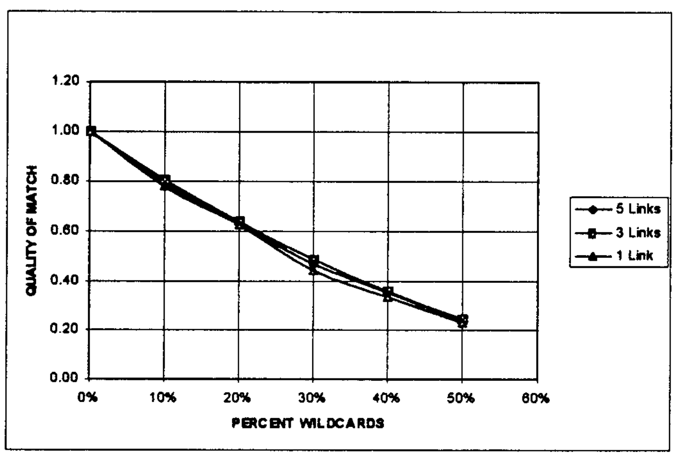


Figure 8. Effect of wildcards on match quality.

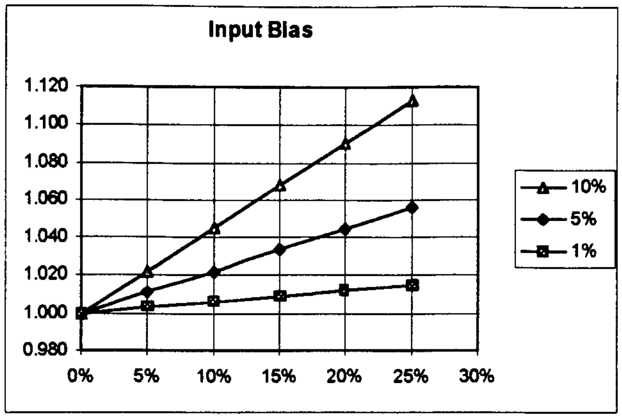


Figure 9. Effect of programmed stops on travel-time bias (raw data).

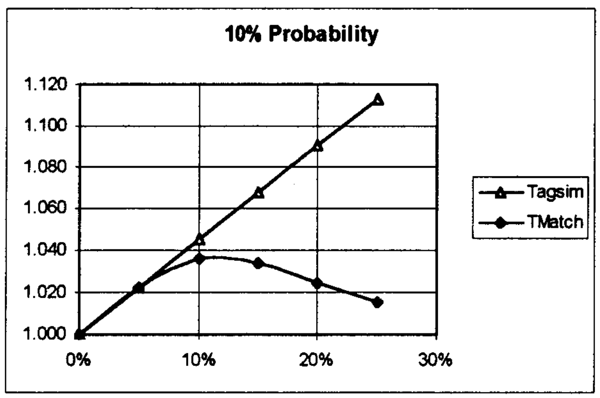


Figure 10. TMatch bias correction for 10% probability of a programmed stop.

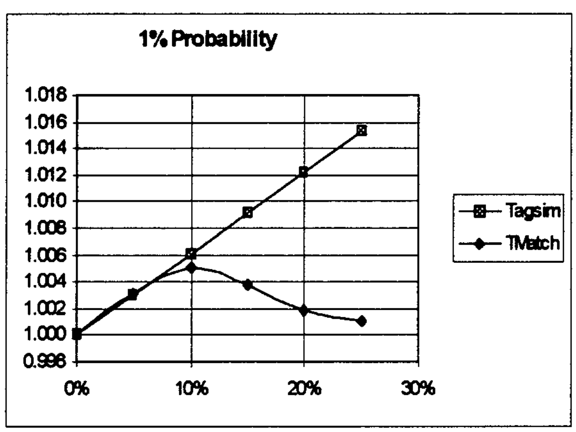


Figure 11. TMatch bias correction for 1% probability of a programmed stop.

Note that the shape of the curve and the maximum amount of bias is approximately the same for both proportions of programmed stops. The bias appears to reach a peak value of approximately four percent when the average programmed stop duration is approximately ten percent of the route travel time. Shorter programmed stops introduce less bias because they represent a proportionally smaller amount of added time. Longer programmed stops add more time, but they are more effectively rejected by the TMatch algorithms.

6. FIELD TEST RESULTS

A pilot field test of the entire data collection and analysis system was carried out in Gainesville during an afternoon peak period in November 1996. The route selected for the study was West University Avenue/Newberry Road from NW 22nd street to the Oaks Mall, a total distance of just under four miles. License tag data were collected for approximately two hours at four locations.

Simultaneous moving vehicle studies were performed for purposes of comparison. The moving vehicle study results are presented in Table 3. A total of six runs was made with an average travel time of 662 seconds per run, indicating an average speed of 20.2 mph.

The corresponding speed estimated by TMatch depends on the tag matching parameters that are specified. Several alternatives were therefore investigated. The number of license tag digits required in the match was set at 3 and 4. The upper speed range for a valid match was chosen as 60 and 80 mph. The minimum speed value was varied from zero to 10 mph.

The tag matching study results are presented in Table 4. Note that the average speed estimated with most of the combinations of tag-matching parameters was very close to the value measured in the moving vehicle study. There were no values significantly above the measured speed of 20.2 mph, but there was some serious underestimation of speed resulting from minimum speeds that were specified too low when only three characters were required for a match. It is clear that, at least in this study, allowing speeds below 5 mph with a three character match created too much potential for an erroneous match that biased the results as much as 50% low. With a four character match, the results were much less sensitive to the upper and lower speed ranges.

Table 3. MVRAP summary for the pilot license tag study.

MOVING VEHICLE RUN ANALYSIS PACKAGE
SUMMARY OF RUN DATA
License Tag Study

ROUTE NAME: SR 24 5 RUN(S) 3 LINKS

RUN#	DATE	TIME	TRAVEL TIME (sec)	DELAY (sec)	STOPS	FUEL	ACCELE-RATION NOISE
1	11-16-96	17:42	883	352	7	0.172	1.899
2	11-16-96	15:36	543	117	5	0.143	1.945
3	11-16-96	16:26	617	161	5	0.148	2.037
4	11-16-96	16:52	608	165	3	0.148	1.628
5	11-16-96	17:16	658	163	4	0.148	1.986

ARTERIAL SYSTEM PERFORMANCE EVALUATION
License Tag Study

ROUTE NAME: SR 24
STARTING AT NW 22ND AVE 5 RUN(S) 3 LINKS

END POINT	LINK LENGTH (ft)	TRAVEL TIME (sec)	DELAY (sec)	STOPS	ACCELE-RATION NOISE	SPEED (MPH) AVG.	RUN.
NW 34TH ST	6024	172.0	51.6	1.2	1.88	23.8	34.1
NW 55TH ST	9777	340.8	94.0	2.2	1.78	19.5	27.0
OAKS MALL	3865	149.0	46.0	1.4	2.16	17.6	25.5
TOTAL	19667	661.8	191.6	4.8	1.90	20.2	28.5

ROUTE TRAVEL TIME STD. DEV. WAS 130.34 SECONDS

Table 4. License tag speed computations using TMatch for a variety of parameter settings.

Number Digits Required in Matched Tag	Maximum Speed Range (mph)	Minimum speed Range (mph) 0.0	2.0	4.0	6.0	8.0	10.0
3	60	9.5	11.9	15.4	19.1	20.8	21.1
3	80	9.8	12.2	15.8	19.6	21.3	21.6
4	60	19.8	19.8	19.8	19.8	21.4	21.9
4	80	20.0	20.0	20.0	20.0	21.7	22.2

Note: All speeds are in miles per hour.

7. ADJUSTING TRAVEL TIMES FOR DIFFERENT VOLUME CONDITIONS

Direct field measurement is probably the most accurate method to obtain delay for any given intersection. However, measured delays can only reflect the traffic conditions that prevail at the time of the field studies. It is clearly not practical to obtain field studies under all traffic conditions, and an analytical adjustment procedure is the only economical means of projecting delay estimates to other traffic conditions. The development of such a procedure is described in Appendix.D

Analytical modeling is one way to compensate for the faults of direct field measurement. A model is inexpensive to use, readily available and allows for the determination of delay at different volumes. A widely used model describing signalized intersection delay is given in Chapter 9 of the HCM. The equation is dependent of the signal cycle length, green time, volume, and capacity. The model was developed empirically using data collected nationally.

Like the direct measurement method, the HCM delay equation has its faults. Because the model is based on empirical data, an intersection with a given cycle length, green time, volume, and capacity may have a delay much different from that which is directly measured there. This situation then leads back to the problem that there is no way to increase the volume in the model and calculate a delay that accurately represents the intersection. Therefore, the HCM equation becomes invalid for this condition and any effort to use it for this case would be futile.

The delay adjustment model described in Appendix D was compared to the HCM delay curve and the fit of the delay curve created by the delay adjustment model was found to approximate the HCM delay curve. The percent difference in the two curves was calculated and plotted in relation to the percent difference in the volumes. The results were consistent with what would be expected results of measurements of stopped delay at signalized intersections. The Delay Adjustment Model was not found to violate any key assumptions of stopped delay at signalized intersections.

8. CONCLUSIONS AND RECOMMENDATIONS

The principal product of this research effort has been the development of a practical technique for collecting license tag data in the field and analyzing the data in the office to produce valid estimates of travel time and speed on an arterial route. The primary advantage of the tag-matching technique is that it produces much larger sample sizes than the more conventional moving vehicle studies. This advantage is achieved at the cost of a somewhat more complex study that requires additional planning and organization that some agency agencies would prefer to avoid.

An attempt was made to exploit the advanced technology of computerized voice recognition for processing of license tag numbers to avoid the task of manual transcription of the field data. The results of the investigation showed that this is an idea that is ahead of its time. Direct entry of the tag data in the field using palmtop computers proved to be as productive as voice recording in terms of capture rates. Both techniques were considerably more productive, and more accurate, than voice recognition.

An attempt was also made to develop an adjustment model to extrapolate the results of a travel time study to different traffic volume conditions. Within the limits of the study parameters, it is suggested that the exponential approximation technique described in Appendix D can be applied with reasonable confidence over a ten percent volume variation under conditions of midrange v/c and g/C ratios. The technique begins to lose accuracy when volumes approach capacity and when green times are very long with respect to the cycle time.

References

1. *Highway Capacity Manual*, Special Report 209, Transportation Research Board, Washington, D.C. 1994.

2. The Florida Department of Transportation, Manual on Uniform Traffic Studies, 1997.

3. Berry, "Evaluation of Techniques for Determining Overall Travel Time", in Proc HRB, 1949.

4. University of Florida Transportation Research Center, A System for Evaluating Urban Congestion with Moving Vehicle Studies, 1990.

5. The GPS-based Odograph Prototype (GOP), Version 1.1 Users Guide, MITRE Corporation, 1995

6. Quiroga, Cesar A. and D. Bullock, "Travel Time Studies with GPS: An Improved Methodology.", Louisiana State University, 1996.

7. Darcy. Bullock and C. A. Quiroga, "Data Collection and Management of GPS Travel Time Data Using a GIS Model.", Louisiana State University, 1997.

Appendix A

Sample Size Requirements for Moving-Vehicle Travel Time Studies

Sample Size Requirements for Moving-Vehicle Travel Time Studies

Based on archived travel time field study data, guidelines have been explored to determine sample size requirements based on characteristics of the study corridor. Study length, number of links, average link spacing, critical volume to capacity ratio (V/c), and arterial classification are known values that could be used to predict a required sample size. The signalization characteristics are also looked at since the studies included in this report use signalized intersections to define link nodes. Therefore, the number of signals included within a corridor is equal to the number of links plus one. The average link spacing is also equal to the average signal spacing.

A.1 SUMMARY OF EXISTING SAMPLE SIZE GUIDELINES

Existing sample size calculation procedures require the sample standard deviation to compute a sufficient number of runs. Since the standard deviation is not known until after the runs have been made, it is difficult to predict the number of runs required in a study to obtain a predetermined level of confidence. The Traffic Analyzer-88 (TA88) Manual(1) suggests the assumption of normality of the travel time data runs. Based on the normal distribution, confidence lines can be plotted on a graph of required runs versus the computed standard deviation (as a percentage of the mean) to predict a required sample size (see Figure A-1). The confidence lines are based on the following equation:

$$\text{Confidence Bound} = \frac{Z * (\text{Standard Deviation})}{\text{Square root } (N)} \qquad \text{(Eq. 1)}$$

where: $Z = 1.96$ for 95% Confidence
$Z = 1.645$ for 90% Confidence

To use the graph, an observer would compute the standard deviation (as a percentage of the mean) of the collected runs. The user would find the corresponding standard deviation on the horizontal axis and draw a line straight up until it intersected the appropriate confidence line (based on the allowable study error). From that point, a horizontal line would be drawn to the vertical axis. This is the value of the required runs. If it is less than the number of runs already collected, the study is complete. If it is larger, more runs are required to obtain the proper confidence interval.

Sample size guidelines have also been created based on empirical research. From previously collected data, the method presented in Table A-1 can be used to predict a minimum sample size based on the type of facility being studied.

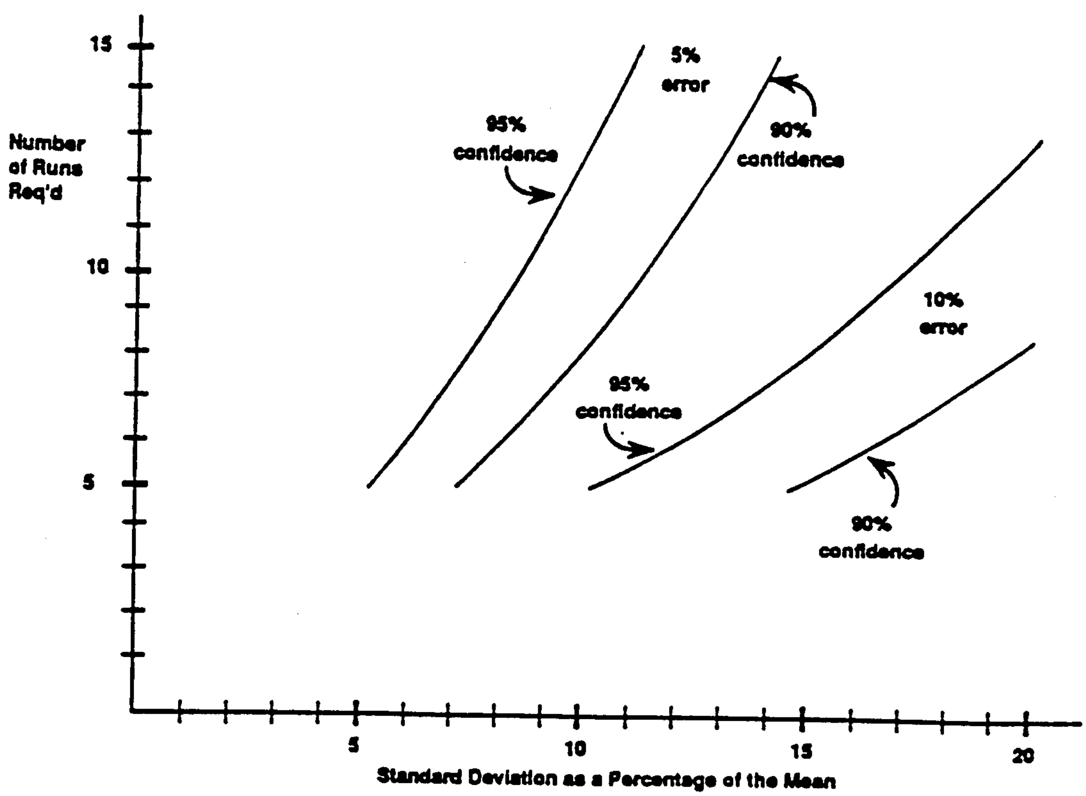


Figure A-1: TA88 Sample Size Requirements (Source: Reference 1).

Table A-1: Number of Test Runs to Predict Overall Speed with 95% Confidence.

TYPE OF FACILITY	NUMBER OF RUNS REQUIRED TO PRODUCE AN ACCURACY OF:		Implied Std Dev of Run Time (Assuming Normality)
	5%	10%	(%)
<u>Signalized Urban Streets</u>			
Two-lane, uncongested	30	8	14
Two-lane congested	40	10	16
Multi-lane uncongested	18	5	11
Multi-lane congested	50	13	18
<u>Rural Highways</u>			
Two-Lane, 1,130 vph	25	6	13
Two-lane, 1,440 vph	42	11	17

(Source: Reference 2)

The FDOT's Manual on Uniform Traffic Studies (MUTS) (3) addresses the issue with a prediction table based on a tabulated average range in the running speed (in miles per hour). This value, defined as R in the MUTS Manual, is calculated using Equation 2.

$$R = \frac{S}{N - 1} \qquad \text{(Eq. 2)}$$

where R = average range in the running speed
S = sum of absolute differences between individual run times and the average run time
N = number of completed test runs

Using the calculated R value, the required sample size can be retrieved from Table A-2.

Table A-2: Approximate Minimum Sample Size Requirements for Travel Time and Delay Studies with Confidence Level of 95.0 Percent.

AVERAGE RANGE IN RUNNING SPEED (mph)* R	MINIMUM NUMBER OF RUNS FOR SPECIFIED PERMITTED ERROR				
	+/- 1.0 mph	+/- 2.0 mph	+/- 3.0 mph	+/- 4.0 mph	+/- 5.0 mph
2.5	4	2	2	2	2
5.0	8	4	3	2	2
10.0	21	8	5	4	3
15.0	38	14	8	6	5
20.0	59	21	12	8	6

* Interpolation should be used when R is other than the numbers shown in column one.
(Source: Reference 3)

Unlike the previous table, this is not used as a predictor for sample size, but as a check for existing sample run adequacy.

A.2 RELATIONSHIP BETWEEN STANDARD DEVIATION AND STUDY LENGTH, NUMBER OF LINKS, AVERAGE LINK LENGTH, CRITICAL v/c RATIO, AND ARTERIAL CLASSIFICATION

Modeling a guideline for required sample runs based on known features of a study corridor would eliminate the cumbersome iterative process now used to check for sample adequacy. Using archived data from Dade and Broward counties (Districts 4 and 6), the study lengths, number of links (signals), average link length (signal spacing), and critical volume to capacity ratio (V/c) will be plotted versus the standard deviation (as a percentage of the mean) to determine if there is any correlation among these values. It should be noted that Dade and Broward County travel time data was used in this study because of the extensive databases that already exist in FDOT Districts 4 and 6. By doing so, the large expense of collecting new data for this study was avoided.

The archived data contains 89 travel time studies. Each study contains four to twenty-five sample runs. The study lengths range from approximately 4,800 feet to 28,200 feet (0.9 to 5.3 miles). The most links in any one study is 17 and the maximum average link length is approximately 7,500 (1.4 miles) long. The critical V/c ratios range from a low of 0.34 to a maximum value of

1.20 and all arterials are either class I or II. A summary of the study data is presented at the end of this appendix.

The travel time and distance information that was extracted from the travel time study database was used to calculate the average total distance, the average total travel time, and the standard deviation of the data. The number of links and runs were pulled directly from the database for use in the spreadsheet. Knowing the number of links and the average total distance, the average link lengths were computed. The largest V/c found within a given study was used as the critical V/c. The plots of the standard deviation versus the average study length, number of links, average link length, and critical V/c are Figure A-2, Figure A-3, Figure A-4, and Figure A-5, respectively.

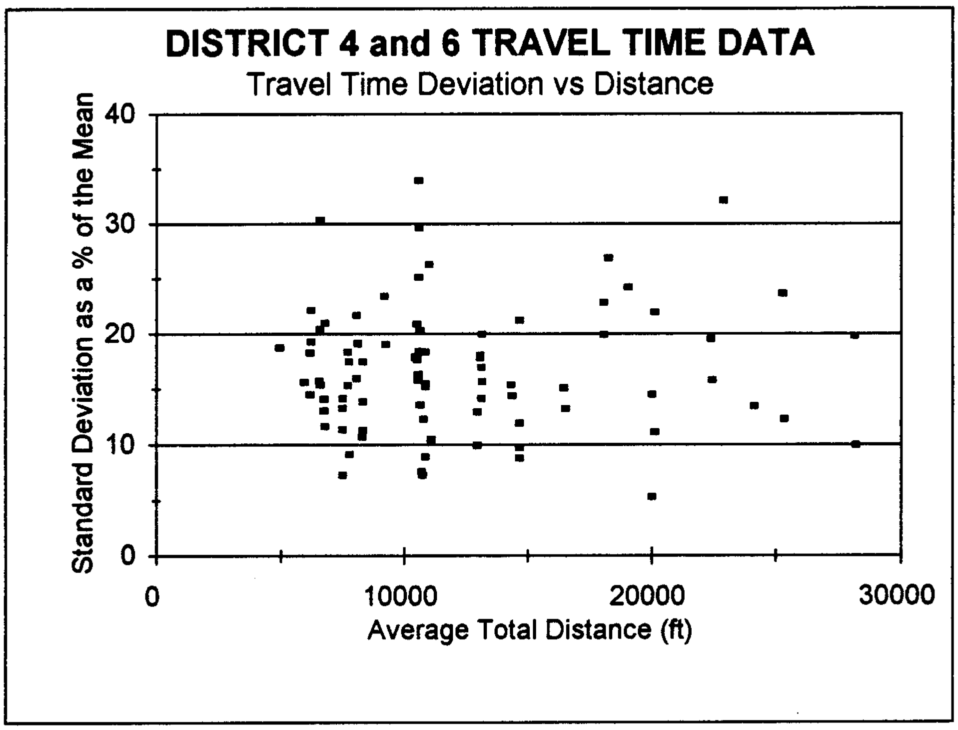


Figure A-2: Standard Deviation versus Average Total Distance.

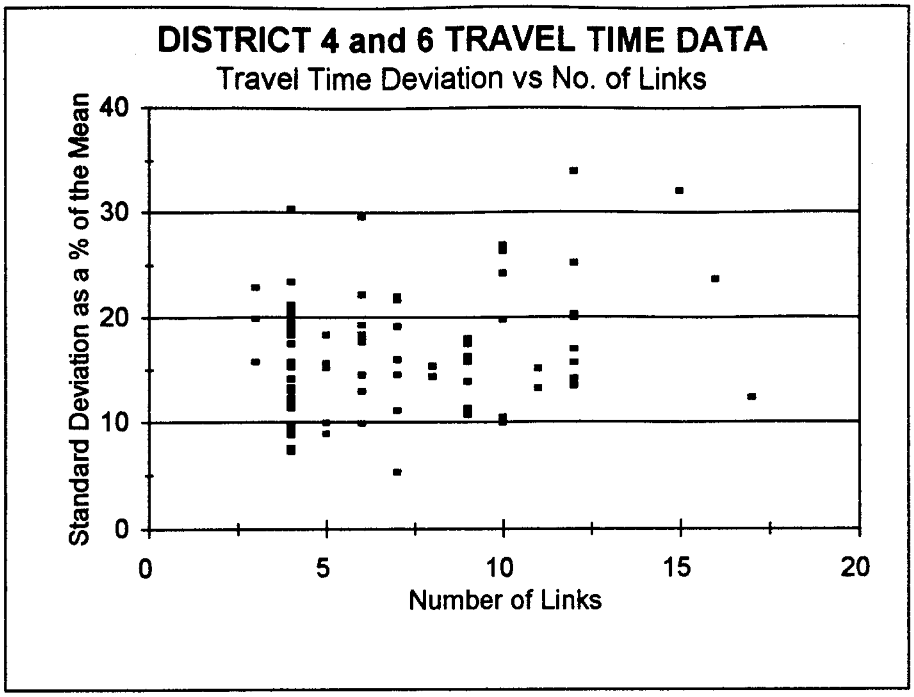


Figure A-3: Standard Deviation versus the Number of Links.

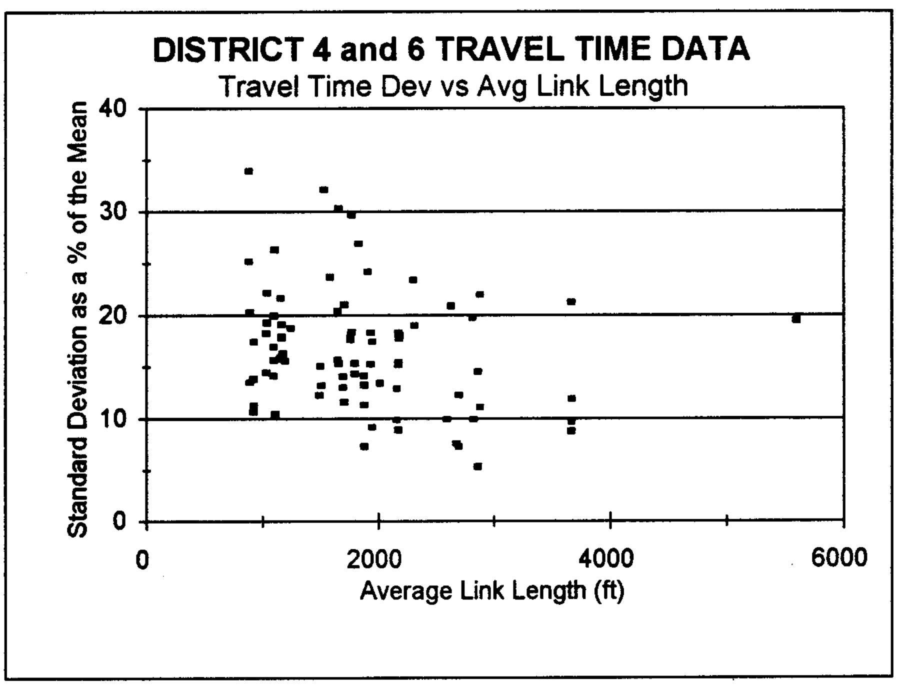


Figure A-4: Standard Deviation versus Average Link Length

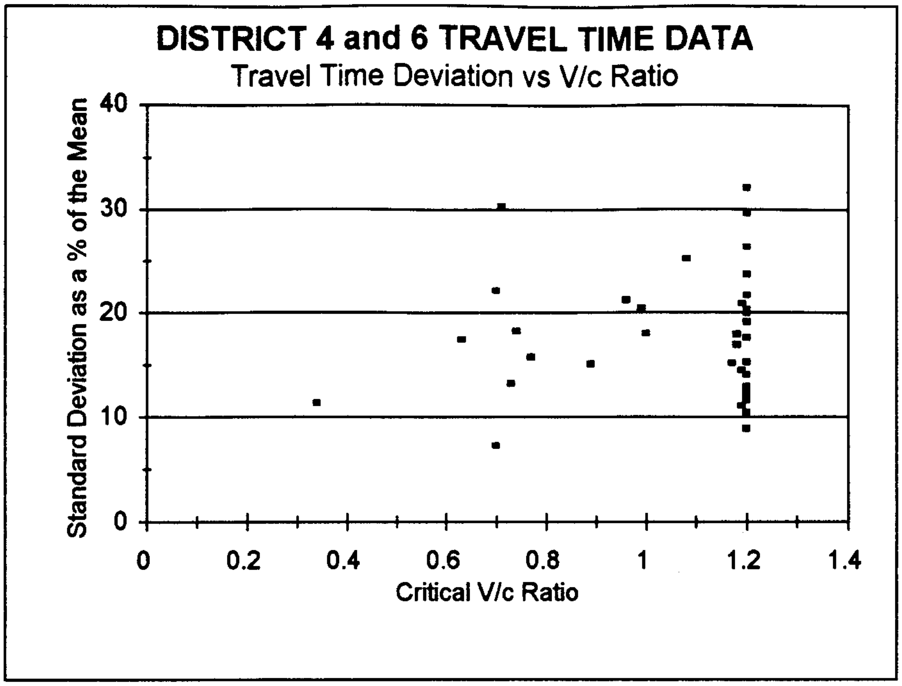


Figure A-5: Standard Deviation versus Critical V/c Ratio

The data scatter in Figures A-2, A-3, and A-4 indicate a lack of correlation between the study average total length, number of links, and link spacing versus the standard deviation. Unfortunately, most of the critical V/c ratios in the travel time studies were equal to 1.20. Therefore, there is a large data "clump" around V/c equal to 1.20 in Figure A-5. Because of this, it is difficult to draw any conclusions from the plot. In an attempt to use the V/c data, the travel time data was broken down into four categories considering arterial classification (based on free flow speed) and critical V/c. Using these categories, the average standard deviations were computed from the corresponding data and then used to determine expected sample sizes based on the TA88 graph (See Figure A-1). Table A-3 shows the resulting deviation values and the corresponding runs needed for a 95% confidence of a 10% error.

Table A-3: Predicted Sample Sizes for Dade and Broward County Travel Time
(Studies Based on Arterial Class and Critical v/c ratio)

	Class I Arterial		Class II Arterial	
	Avg Std Deviation (% mean)	Sample Size* Required (runs)	Avg Std Deviation (% mean)	Sample Size* Required (runs)
V/c less than or equal to 1.0	15.6	9	18.5	12
V/c greater than 1.0	16.8	10	23.4	21

* Sample size required for 95% confidence of a 10% error.

The table shows that the required sample sizes increase as both the arterial classification and critical V/c increase.

DADE AND BROWARD COUNTY TRAVEL TIME SUMMARY

The following table contains the travel time data taken from the District 4 and District 6 archives for use in this report. The first column displays the FDOT District in which the study was conducted. The second and third columns contain the route identification and the average total distance of the runs within the route. The number of links is displayed as well as the average link distance (number of links divided by the average total distance) in the next two columns. The number of data runs contained within the study and the route classification are shown. The corridor speed limit and critical V/c ratio are contained in the next two columns. The average travel times and standard deviations of the travel times are computed from the individual run data from the archives and displayed in columns ten and eleven. The final column contains the travel time standard deviation expressed as a percentage of the mean.

Table A-4: Summary of Dade/Broward Count Travel Time Data											
FDOT Dist.	Route ID	Avg Tot Distance (ft)	Links	Avg Link Spacing (ft)	Data Runs	Route Class	Speed Limit (mph)	Critical V/c	Avg Trav Time (s)	Std Dev (s)	Std Dev (% mean)
4	10OFFE	6,581.2	4	1,645.3	6				272.7	42.8	15.7
4	10OFFW	6,619.2	4	1,654.8	6				185.7	28.5	15.4
4	10PME	6,586.8	4	1,646.7	19	2	35	0.99	282.8	57.7	20.4
4	10PMW	6,615.6	4	1,653.9	18	2	35	0.71	230.6	70.0	30.3
4	110FFE	13,142.3	12	1,095.2	4				361.8	51.2	14.2
4	11OFFW	13,160.5	12	1,096.7	6				394.2	61.8	15.7
4	11PME	13,140.4	12	1,095.0	20	1	40	1.18	354.4	60.0	16.9
4	11PMW	13,164.0	12	1,097.0	22	1	40	1.20	355.1	70.9	20.0
4	12OFFN	10,451.2	9	1,161.2	5				370.8	66.0	17.8
4	12OFFS	10,573.8	9	1,174.9	5				366.8	59.7	16.3
4	12PMN	10,456.4	9	1,161.8	21	1	40	1.18	361.0	64.7	17.9
4	12PMS	10,582.2	9	1,175.8	21	1	40	0.77	365.1	57.5	15.8
4	13OFFE	8,128.6	7	1,161.2	5				397.2	190.2	47.9
4	13OFFW	8,081.3	7	1,154.5	6				312.8	49.8	15.9
4	13PME	8,141.0	7	1,163.0	24	2	35	1.20	272.6	52.1	19.1
4	13PMW	8,077.1	7	1,153.9	25	2	35	1.20	343.1	74.3	21.7
4	14OFFN	8,319.2	9	924.4	5				255.2	35.4	13.9
4	14OFFS	8,333.0	9	925.9	5				276.4	31.2	11.3
4	14PMN	8,315.1	9	923.9	21				260.0	27.8	10.7
4	14PMS	8,332.2	9	925.8	21				281.2	49.0	17.4
4	15OFFE	6,195.0	6	1,032.5	6				274.8	39.8	14.5
4	15OFFW	6,229.0	6	1,038.2	6				179.5	34.6	19.2
4	15PME	6,193.1	6	1,032.2	17	2	35	0.74	305.6	55.8	18.2
4	15PMW	6,227.9	6	1,038.0	18	2	35	0.70	181.3	40.1	22.1
4	16OFFE	10,597.9	12	883.2	11				468.0	158.9	33.9
4	16OFFW	10,653.2	12	887.8	11				397.1	53.7	13.5
4	16PME	10,597.2	12	883.1	22	1	40	1.08	455.0	114.6	25.2
4	16PMW	10,655.5	12	888.0	22	1	40	1.20	438.6	89.0	20.3
4	1OFFN	14,679.0	4	3,669.8	5				308.0	27.0	8.8
4	1OFFS	14,682.0	4	3,670.5	5				295.6	28.7	9.7
4	1PMN	14,677.3	4	3,669.3	15	1	45	0.96	351.9	74.6	21.2
4	1PMS	14,681.2	4	3,670.3	17	1	45	1.20	311.7	37.1	11.9
4	2OFFN	20,035.8	7	2,862.3	5				523.0	27.9	5.3
4	2OFFS	20,153.2	7	2,879.0	5				491.6	108.1	22.0
4	2PMN	20,032.5	7	2,861.8	12	1	40	1.19	454.8	66.0	14.5
4	2PMS	20,152.3	7	2,878.9	14	1	40	1.19	563.7	62.6	11.1
4	3OFFE	10,791.0	4	2,697.8	5				187.8	23.1	12.3
4	3OFFW	10,703.8	4	2,676.0	5				249.8	18.9	7.6
4	3PME	10,772.8	4	2,693.2	18	1	45	0.70	185.8	13.6	7.3
4	3PMW	10,514.1	4	2,628.5	23	1	45	1.19	259.7	54.2	20.9
4	4OFFN	10,860.4	5	2,172.1	8				224.4	34.6	15.4
4	4OFFS	10,863.1	5	2,172.6	8				249.4	45.7	18.3
4	4PMN	10,854.2	5	2,170.8	22	1	45	1.20	217.1	19.4	8.9
4	4PMS	10,860.3	5	2,172.1	24	1	45	1.17	273.1	41.5	15.2
4	5OFFN	13,082.0	6	2,180.3	5				259.2	46.1	17.8
4	50FFS	12,962.6	6	2,160.4	5				336.4	33.3	9.9
4	5PMN	13,089.3	6	2,181.5	15	1	45	1.00	286.3	51.7	18.0

Table A.4 (Continued)											
FDOT Dist.	Route ID	Avg Tot Distance (ft)	Links	Avg Link Spacing (ft)	Data Runs	Route Class	Speed Limit (mph)	Critical V/c	Avg Trav Time (s)	Std Dev (s)	Std Dev (% mean)
4	5PMS	12,973.8	6	2,162.3	14	1	45	1.20	403.1	52.1	12.9
4	6OFFE	7,788.6	4	1,947.2	5				191.8	17.6	9.2
4	6OFFW	7,728.0	4	1,932.0	5				216.0	39.6	18.3
4	6PME	7,785.5	4	1,946.4	18	2	40	0.63	175.1	30.5	17.4
4	6PMW	7,726.1	4	1,931.5	18	2	40	1.20	215.3	32.9	15.3
4	7OFFE	6,812.0	4	1,703.0	5				229.8	48.2	21.0
4	7OFFW	6,770.8	4	1,692.7	6				267.7	34.9	13.0
4	7PME	6,808.4	4	1,702.1	17	1	45	1.20	235.5	27.4	11.6
4	7PMW	6,767.5	4	1,691.9	20	1	45	1.20	230.6	32.5	14.1
4	8OFFE	10,534.4	6	1,755.7	5				388.4	69.1	17.8
4	8OFFW	10,611.0	6	1,768.5	5				306.0	56.1	18.3
4	8PME	10,533.8	6	1,755.6	16	1	35	1.20	393.0	69.3	17.6
4	8PMW	10,612.3	6	1,768.7	15	1	35	1.20	333.0	98.7	29.6
4	9OFFN	7,506.6	4	1,876.6	5				190.4	26.9	14.1
4	9OFFS	7,517.8	4	1,879.4	5				210.2	15.3	7.3
4	9PMN	7,508.5	4	1,877.1	15	2	35	0.34	195.4	22.2	11.4
4	9PMS	7,519.0	4	1,879.8	15	2	35	0.73	224.8	29.8	13.2
6	152ST-W	12,967.0	5	2,593.4	12				251.9	25.1	9.9
6	27AVE_N	14,375.2	8	1,796.9	14				426.0	61.1	14.3
6	27AVE_S	14,347.7	8	1,793.5	12				519.4	79.6	15.3
6	MGD_E	25,289.0	16	1,580.6	12	2	40	1.20	878.8	208.0	23.7
6	MGD_W	25,351.9	17	1,491.3	12	2	40	1.20	705.3	86.8	12.3
6	NW42A_N	4,845.5	4	1,211.4	12				264.3	150.5	57.0
6	NW42A_S	4,971.8	4	1,242.9	12				100.8	18.9	18.7
6	NW42N1	11,105.3	10	1,110.5	12	2	40	1.20	280.6	29.4	10.5
6	NW42S1	11,006.8	10	1,100.7	12	2	40	1.20	334.8	88.1	26.3
6	NW79A_E	28,159.6	10	2,816.0	12				1,029.7	203.3	19.7
6	NW79A_W	28,227.7	10	2,822.8	12				1,133.4	112.6	9.9
6	NW79B_E	22,936.3	15	1,529.1	13	2	40	1.20	790.8	254.0	32.1
6	SW40B_W	16,542.4	11	1,503.9	12				630.5	83.3	13.2
6	NW79B_W	24,150.9	12	2,012.6	12				777.7	104.5	13.4
6	SW40A_E	6,034.9	5	1,207.0	15	2	40	1.20	236.3	116.8	49.4
6	SW40A_W	5,971.0	5	1,194.2	15				177.3	27.6	15.6
6	SW40B_E	16,473.8	11	1,497.6	12	2	40	0.89	479.0	72.3	15.1
6	US1A_N	22,416.2	4	5,604.0	12				526.2	102.4	19.5
6	US1A_S	22,477.0	3	7,492.3	12				459.1	72.2	15.7
6	US1B_N	9,203.0	4	2,300.8	12				220.1	51.5	23.4
6	US1B_S	9,252.1	4	2,313.0	12				231.2	43.9	19.0
6	US1C_N	18,094.8	3	6,031.6	12				539.6	123.3	22.8
6	US1C_S	18,102.5	3	6,034.2	12				416.4	82.7	19.9
6	US1D_N	18,284.4	10	1,828.4	13				526.6	141.8	26.9
6	US1D_S	19,081.6	10	1,908.2	12				589.1	142.6	24.2

Appendix B

License Tag Matching with TMATCH

License Tag Matching with TMATCH

B.1. INTRODUCTION

TMatch is a software tool that assists in the estimation of a roadway's level of service (LOS) using a licence tag matching procedure. The basis of the tag-matching method is that the travel speed along a link may be estimated by observing vehicles at the ends of the link, and comparing the times at which they were observed. To uniquely identify the vehicles observed, their license tags are recorded, along with the observation time.

B.1.1. Introduction to Tag Matching

TMatch is a tag-matching tool. That is, it takes lists of license tag numbers observed at two observation stations along a roadway, and compares them to find any matches (tags which were observed at both stations). This is made more interesting by recording not just the tags, but also the times at which they were observed. With this information stored in files with the proper format, TMatch can compute a speed for each matched tag pair. Averaging these speeds gives an estimate of the average vehicle speed over the link of roadway between the stations, which can be used to estimate the LOS for the link.

When tags are observed at more than two stations, the matching can be performed on each pair of stations, and the results pooled to give a more robust estimate of the average speed between the first and last stations.

The estimated speed is of course statistical in nature, and thus naturally calls for associated estimates of error and reliability. TMatch output includes several such measures.

B.1.2. Introduction to Travel Time Studies Using Tag Matching

A travel-time study using TMatch may be broken into four tasks:

1. Design the study.
   Determine positions for the observation stations, and measure the distances between the stations and the other roadway characteristics.

2. Collect the data.
   Obtain the license tag numbers and times and store them in a text file with the proper format.

3. Match tags and Compute the LOS.
   Enter the roadway characteristics into a TMatch study, find tag matches between pairs of observation stations and use them to estimate average speed, travel time, and LOS.

4. Draw Conclusions.
   Evaluate the results and error/reliability measures, and draw conclusions.

TMatch automates Task 3, which is one of the more tedious parts of this process. As such, it is simply an automation tool and is not intended to replace the knowledge and experience of the investigator.

B.1.3. TMatch File Structure and Information Flow

The file structure and information flow is illustrated in Figure B-1. Note that there are four files that contain information related to the tag matching process.

- Tag Data File (TMI): Contains a list of the tag numbers and observed times;
- Route Data File (TMR): Contains the lengths, speeds and other necessary characteristics of the route being studied;
- Study Parameter File (PAR): Contains the parameters that govern the matching process for a particular study; and
- Output File (TMO): Contains a summary of the results of the study.

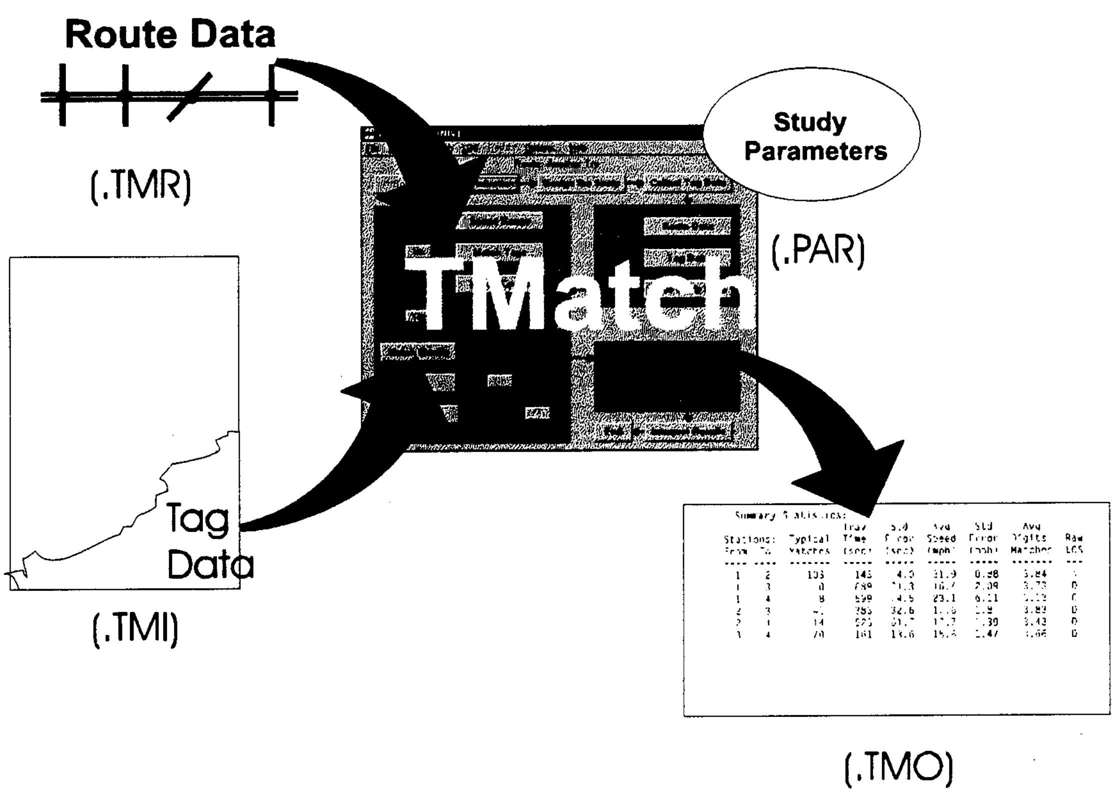


Figure B-1. TMatch file structure and information flow.

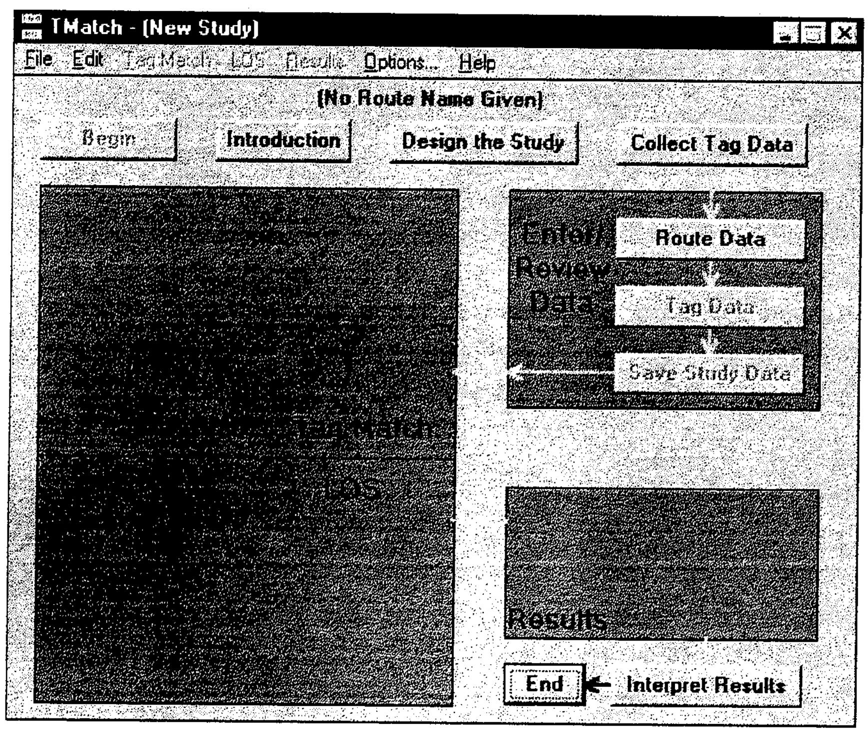
This document covers the essentials of using TMatch. If you have a lot of experience using computers in LOS studies, you will probably find what you need in the *Quick Start* section. The *Step-By-Step* section has more detailed instructions. Descriptions of the statistics reported in the TMatch output file are given in the section on *Interpreting the Results.*

B.2. QUICK START

Experienced computer users will find the program easy to install and run. The installation follows the standard Windows convention with a SETUP.EXE file on the installation disk. The setup program will install TMatch in any directory you specify, along with a complete set of sample data. When TMatch has been installed, you will normally launch the program from a desktop icon that looks like this:

Tmatch.lnk

When you double-click the icon, TMatch will execute and you will see the following main menu screen:



Note that there are three buttons across the top of the screen that will display an explanation of various aspects of tag matching studies.

- Introduction,
- Designing the Study and
- Collecting Tag Data.

The *Introduction* button will present the explanation given at the beginning of this document. The other two choices are summarized as follows:

B.2.1. Designing the Study

Computerized algorithms are fast, but are essentially mindless. This means that they will attempt to perform their tasks with whatever data they are given, regardless of whether the data are meaningful. Any meaningful conclusions, then, are as much a result of careful planning and execution as of software quality. For this reason, it is vital that any study be designed well. For travel-time studies using tag matching, this has three components:

A. Screening the roadway to be studied to ensure that tag-matching is a suitable method.
B. Laying out the observation stations at which tags will be observed.
C. Measuring the roadway characteristics of the stretches of roadway between the stations.

Generally, the only information you will need from the detailed sections will be the observation file format. The exceptions might be the sections on statistics (if you have knowledge of statistics and need more detail than is provided in the section on *Interpreting the Results*) and *File Formats* (if you are a programmer and want to use TMatch's efficient matching algorithm to generate match files, but use them for purposes other than TMatch's).

Screening the Roadway to be Studied

The roadway and study period will usually be dictated before the study is undertaken. Before approaching a travel-time study with a tag-matching method, however, you must determine whether the method is appropriate. The requirements of the tag-matching method are:

1. The roadway must be continuous (that is, a hypothetical asphalt worm which starts eating asphalt at one end of the study roadway must be able to eat its way to the other end, without leaving the study roadway).

2. The expected traffic volume during the study period must be fairly high (at least 900 vehicles per hour).

3. The lane structure of the roadway must allow the license tag observers to collect tags from all thru lanes (for example, a study of a six-lane highway would not be suited,

unless there were a median so that observers could be posted on both sides of one directions lanes).

4. There should not be off-roadway shortcuts for any sections of the roadway. (For example, consider a roadway making a large bend. A shortcut might be a side road which cuts straight across the bend, if a significant number of vehicles were expected to have a shorter travel time on the shortcut than on the study roadway.)

5. The expected number of programmed stops should be relatively small, no more than 10% of the total volume on any link. (A vehicle makes a programmed stop if it exits the roadway and returns at the same point after a delay, e.g., shopping).

6. There should be a reasonable number of vehicles expected to travel the entire length of roadway under study (at least 400).

7. Traffic flow over the study period should be fairly homogeneous.

Laying Out the Observation Stations

If the roadway has been found to be suited to a tag-matching approach, the next step is to decide where to position the observation stations along the roadway. This is largely at the discretion of the researcher; the only hard-and fast rule is the obvious one that there must be stations at the beginning and end of the study roadway.

Some rules of thumb are:

1. Position an observer downstream from an intersection where the traffic volume is expected to increase.

2. Position an observer upstream from an intersection where the traffic volume is expected to decrease.

3. Use at least three, preferably four or more, stations.

4. Position observers on either side of major traffic storage areas (such as malls).

5. Position observers where the roadway characteristics change.

These are only guidelines. Remember that ultimately you must judge the special characteristics of your study roadway and the needs of your study, and position observation stations accordingly. If these requirements are not met, it might be better to study the roadway using a different method.

Measuring the Roadway Characteristics

Positioning observation stations divides the roadway into simple links between adjacent stations. Accurately studying these links requires that we have some general information about them. In order to simplify the tag collection process, TMatch does not store these measurements in the station files. Rather, they are stored in the study file, where they are associated with the downstream station on the link. The first station in the study (the most upstream on the entire study roadway), therefore, has no associated characteristics.

The characteristics used by TMatch are:

- Link Length: The length in feet of the section of roadway between adjacent stations. This needs to be measured fairly carefully.
- Traffic Volume: Average number of vehicles per hour travelling the link during the study period. This can be estimated.
- Route Class: 1 (urban), 2 (transitional), or 3 (rural). Defaults to 2.
- Posted Speed: Speed limit (or average speed limit, rounded to a multiple of 5 mph).

(See *Required and Suggested Parameters* for more information.)

If these characteristics are not constant over a link, they should either be set to reflect the most important part of the link, or be averaged (weighted by length, for example).

B.2.2. Collecting the Data

Once the study has been designed, data collection is easy in theory (though sometimes tiresome in practice). It is important to remember that TMatch requires both license tags and the times at which they are observed. Here are some suggestions for obtaining high acquisition rates and data accuracy:

1. It is inefficient and inaccurate to have the observers enter both the tag number and the time. A much better approach is to have a program in which the observer enters the tag, and the software assigns the time (from the system clock).

2. Use a selection rule to select vehicles. Preferably, the rule should be one which the observer can apply as a vehicle approaches (e.g., grey passenger cars) rather than one which must be applied as it recedes (e.g., no bumper stickers).

3. Record an agreed-on tag fragment rather than the entire tag number. Recording the last four digits seems to be a good compromise between acquisiton speed and vehicle identification.
4. Don't blow off dense platoons of vehicles by collecting one tag from each. Try to record at least three (beginning, middle, and end) from a platoon of about 30 vehicles, more from larger.

5. Try to collect tags from all thru lanes (if there are multiple thru lanes).

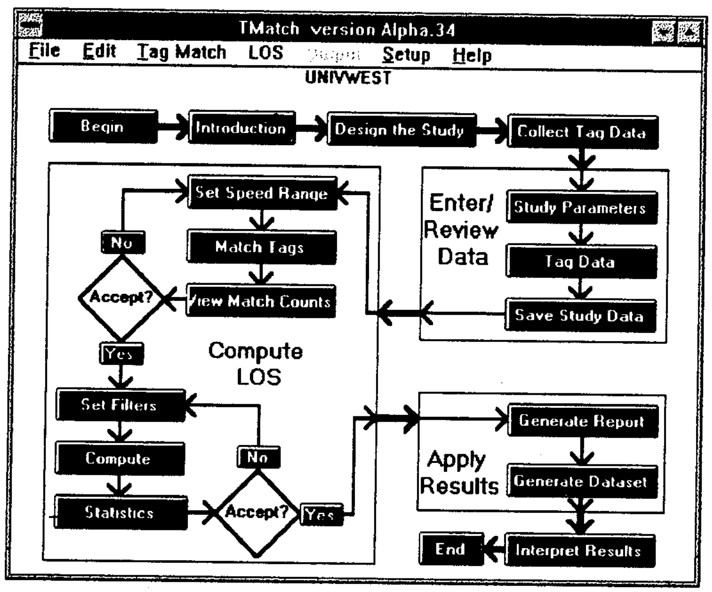
6. Don't collect tags from turning vehicles unless you have plenty of time.

7. Try to minimize the number of tags with missing digits. Missing digits can slow down the tag match, but more importantly can (if frequent) seriously damage the accuracy of the results.

B.2.3. Loading the Example Data

The example data set will facilitate the explanation of the various steps in the TMatch procedure. The balance of this discussion will therefore assume that you have loaded the data set that was included on the distribution disk. This is accomplished by clicking on the *File* menu item and choosing "Open." This will present you with a dialog box that only offers one data set, unless you have already entered some of your own data. The file name for this data set is **WUniv.TMR.** The data were collected on West University Avenue in Gainesville.

When the data set is loaded, the menu screen comes alive with more options, like this:



B.3. DATA ENTRY AND PROGRAM CONTROL DIALOG

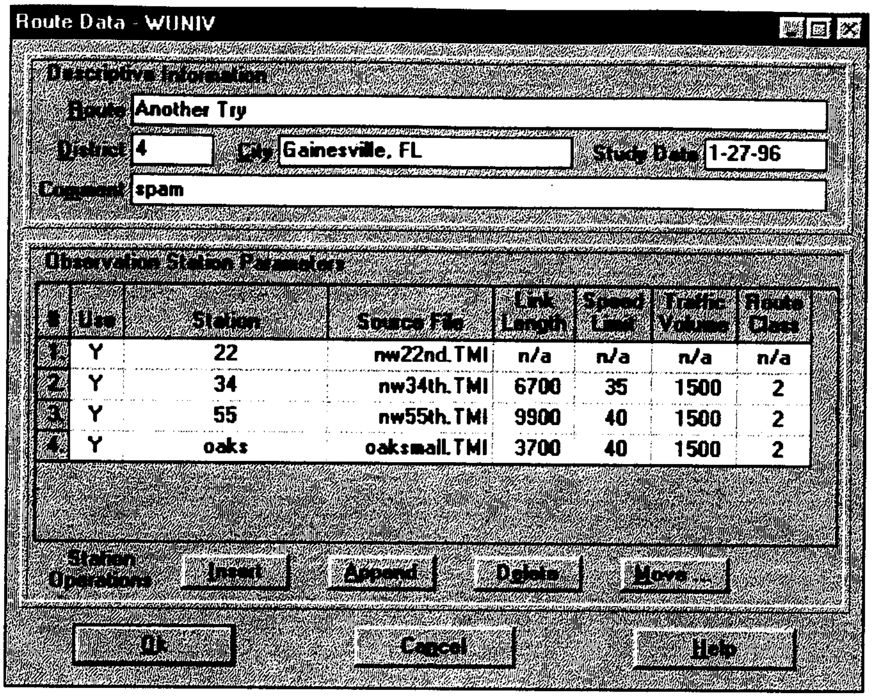
The menu options are presented in the form of a flowchart that depicts the steps in the tag matching procedure. Note that there are three general sections to the procedure as delineated by the rectangular boxes:

- Data entry and editing,
- Computation of LOS by tag matching and
- Review and interpretation of results.

The user interface screens associated with these steps will now be examined. Each screen will be presented on the following pages of this document.

B.3.1. Route Data Entry and Editing

There are two sections on this screen, as shown below.



The first contains the descriptive information, including:

- Route Name
- District
- City
- Date and
- Comment .

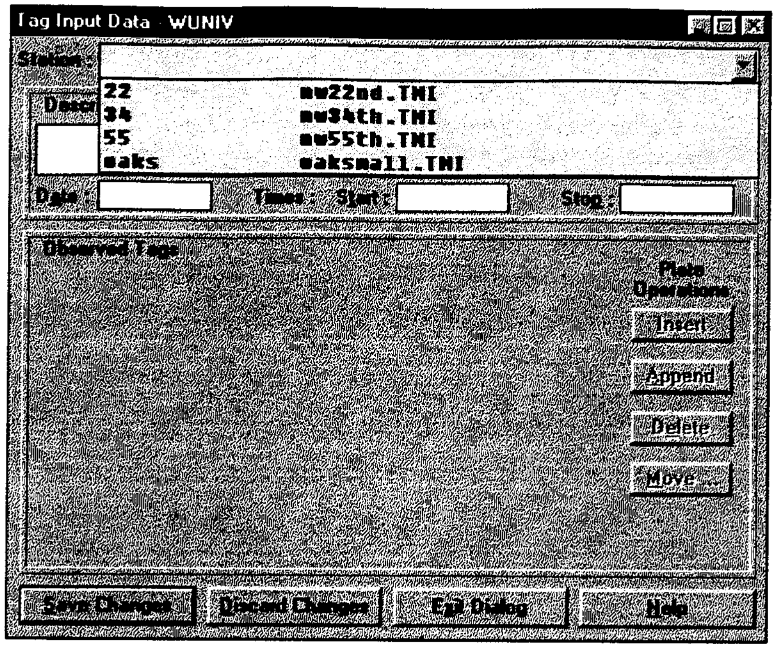
The second contains the data specific to each station, including:

- Use? (Y/N): Decides whether or not the station will be included in the analysis;
- Station identification (Usually cross street names);
- Input file name for the tag data
- Link length
- Posted speed
- Traffic volume
- Route class by the HCM definition

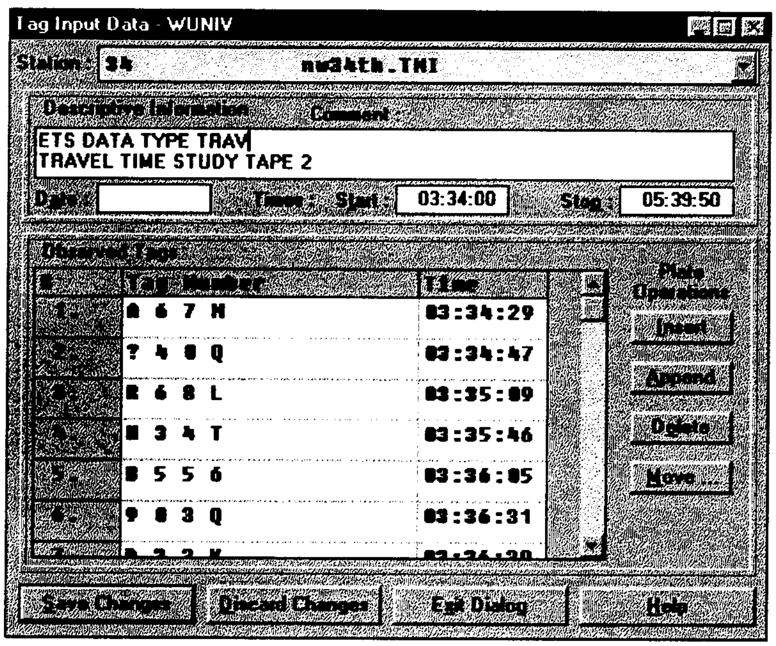
The standard “OK”, “Cancel” and “Help” buttons are also provided on this screen.

B.3.2. Tag Data Screen

The Tag Data Screen looks like this until you select a station from the list that is presented.

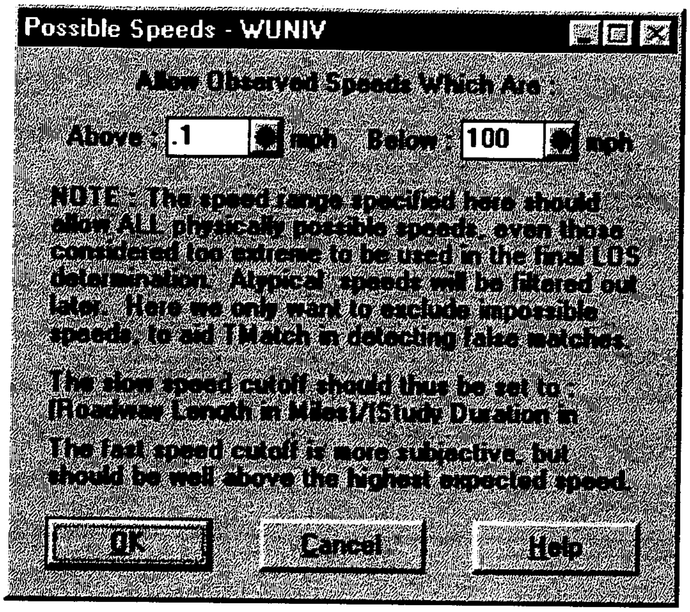


When the selection has been made, the tag data file (TMI) is retrieved and the data are displayed for review and editing.



B.3.3. "Set Speed Range" Dialog Box

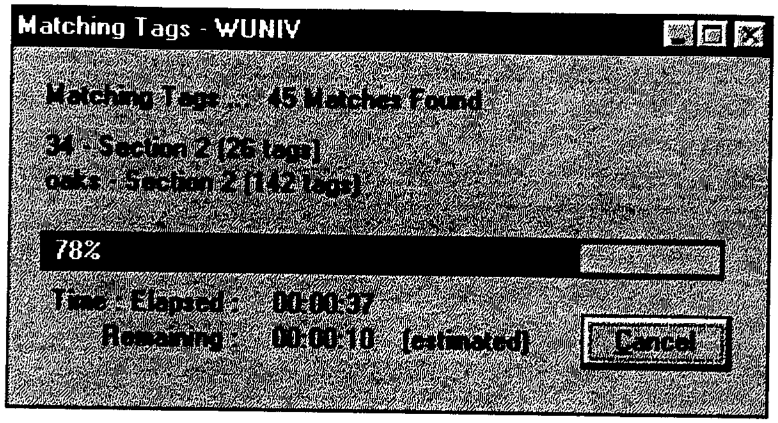
This dialog box is selected by clicking the box with the same name on the main screen flowchart. It is important that tag matches that imply speeds outside of a reasonable speed be rejected, because such matches generally represent erroneous data (either errors in data entry, or false matching of partial tag numbers.



It is equally important that the speed range encompass all reasonable speeds to avoid biasing the results one way or the other.

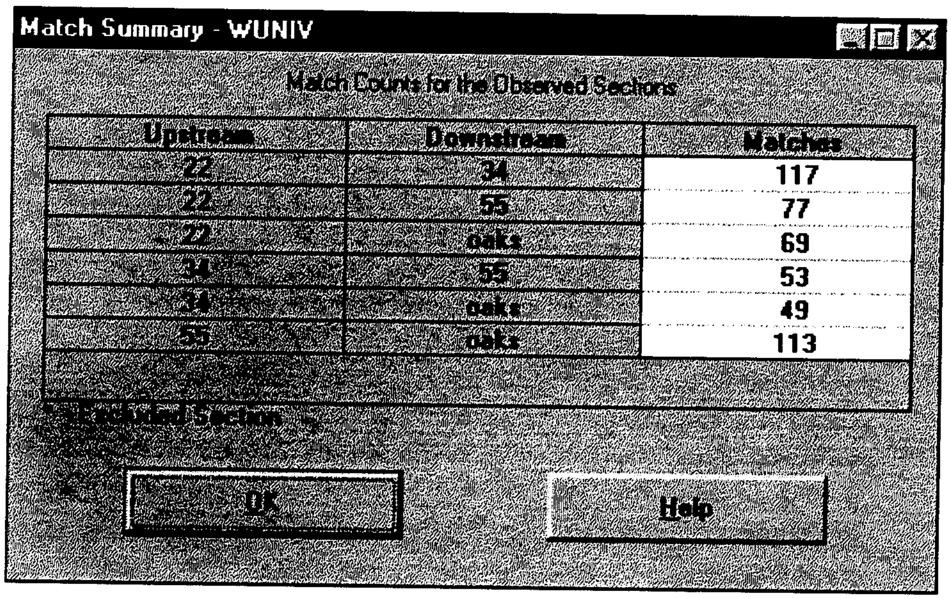
B.3.4. "Match Tags" Display

When you select the "Match Tags" box on the main screen flowchart, the matching process begins. This process can take several minutes, so a progress indicator display is presented. You may abort the process at any time with the "Cancel" button.



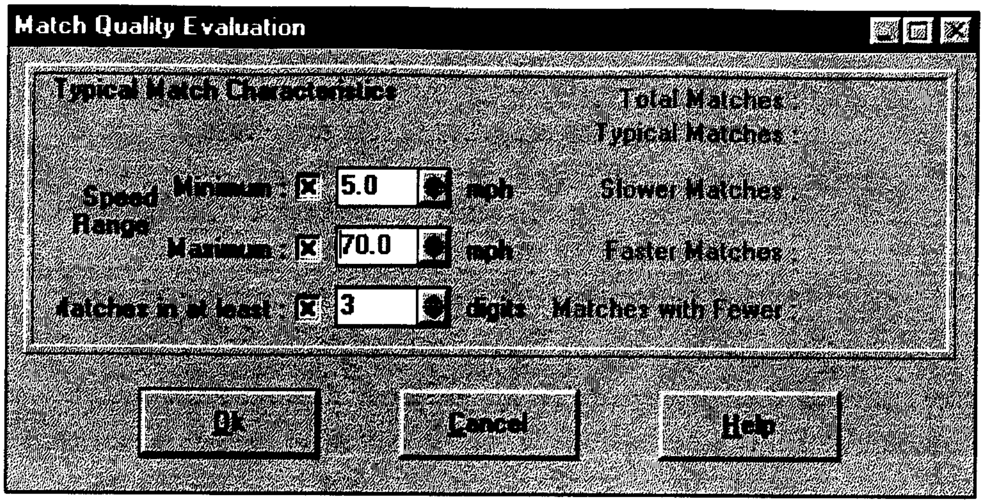
B.3.5 "View Match Counts" Display

This dialog box is selected by clicking the box with the same name on the main screen flowchart. The number of matches on each link is displayed here.



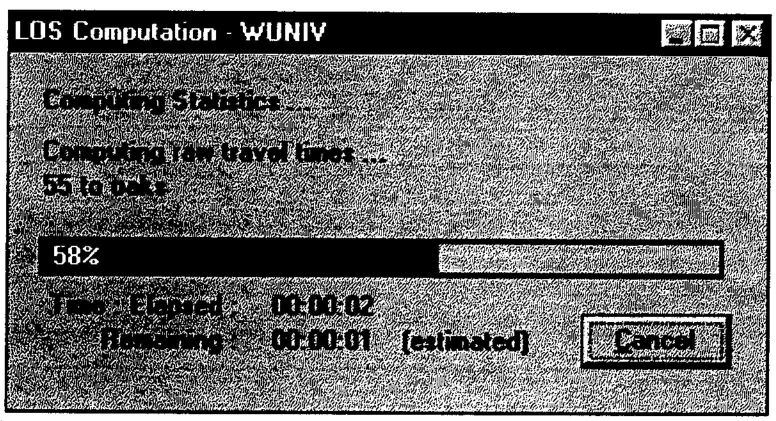
B.3.6. Match Quality Dialog Box

This dialog box is selected by clicking the box with the same name on the main screen flowchart.



B.3.7. LOS Computation Display

When you select the “Compute” box on the main screen flowchart, the matching process begins. During this process a progress indicator display is presented. You may abort the process at any time with the “Cancel” button.



B.3.8 Statistics Display

After the LOS computations have been completed, you may view the statistical display that presents sample sizes and standard errors for both speed and travel time. These figures are required to assess the statistical significance of the results.

Wuniv.tmo - Notepad

File Edit Search Help

Observation Stations:

Number	Name	Tags	Distance from Previous (ft)
1	22	400	0
2	34	426	6700
3	55	514	9900
4	oaks	328	3700

Study Results:

Stations: From	To	Adj N	Adj. Time (sec)	Std Error (sec)	Adj Speed (mph)	Std Error (mph)	Adj LOS	LOS Signif
1	2	144	172	8.8	26.6	1.42	B	0.6577
2	3	96	436	23.6	15.5	0.88	D	0.5159
3	4	85	146	11.5	17.3	1.47	D	0.963

B.4. SAMPLE TMATCH OUTPUT

A sample output including all of the data produced by TMatch is presented as follows:

```
LICENSE TAG TRAVEL TIME STUDY SUMMARY
Report Date:  07-30-1997

Study:         WUNIV
Study Date:    1-27-96
Study Period:  15:29:07 to 18:04:10

Route:         Another Try
District:      4 (Gainesville, FL)
Comment:       spam


Observation Stations:
                                          Distance from
Number          Name                Tags  Previous (ft)
------  --------------------        ----  -------------
   1     22                          400            0
   2     34                          426         6700
   3     55                          514         9900
   4     oaks                        328         3700


Study Results:
                        Adj.    Std     Adj     Std
Stations:    Adj       Time    Error   Speed   Error   Adj    LOS
From  To      N        (sec)   (sec)   (mph)   (mph)   LOS   Signif
---- ----   -----     -----   -----   -----   -----   ---   ------
  1    2     151        168     8.4    27.2    1.44    B    0.8344
  2    3     103        431    22.2    15.7    0.85    D    0.5597
  3    4      92        137    10.8    18.4    1.57    C    0.5117
  1    4     274        735    12.9    18.8    0.34    C    0.8585

NOTES: 1. The range of included speeds is 5.0 mph to 70.0 mph.
          Detectable LOS are ABCDEF.
       2. Power to detect LOS A was reduced by approximately 0.0%.
          Power to detect LOS F was reduced by approximately 0.0%.
```

Summary Statistics:

Stations: From	To	Typical Matches	Trav Time (sec)	Std Error (sec)	Avg Speed (mph)	Std Error (mph)	Avg Digits Matched	Raw LOS
1	2	103	143	4.0	31.9	0.88	3.84	A
1	3	40	689	71.3	16.4	2.09	3.73	D
1	4	8	599	74.5	23.1	6.11	3.13	C
2	3	41	383	32.6	17.6	1.8	3.83	D
2	4	14	525	61.7	17.7	1.39	3.43	D
3	4	70	161	13.6	15.6	1.47	3.66	D

Route Characteristics:

Stations: From	To	Link Length (ft)	Hourly Volume	Posted Speed (mph)	Route Class
1	2	6700	1500	35	2
1	3	16600	1500	40	2
1	4	20300	1500	40	2
2	3	9900	1500	40	2
2	4	13600	1500	40	2
3	4	3700	1500	40	2

NOTE: Volumes, Posted Speeds, and Route Classes for combined links are the averages for the simple links, weighted by link length.

Match Counts:

Stations: From	To	Matches: Total Count	Typ. (1)	Atyp. (2)	Atypical Match Breakdown: Outliers (3)	Fast (4)	Slow (5)	Weak (6)	Repeat (7)
1	2	117	103	14	13	1	11	2	0
1	3	77	40	37	8	0	8	4	32
1	4	69	8	61	3	0	20	20	45
2	3	53	41	12	7	1	7	4	0
2	4	49	14	35	0	1	18	15	15
3	4	113	70	43	5	7	34	14	0

```
NOTES: 1. TYPical matches are those which satisfy ALL the match
          selection criteria.  Statistics are computed using
          only typical matches; atypical matches are discarded.
       2. The ATYPical total may not equal the sum of the
          atypical breakdown counts.  This is because each
          atypical match may fail to satisfy several of the
          selection criteria.
       3. OUTLIER identification uses running box-and-whiskers
          with a symmetric sample of 39 elements,
          less 1 extreme observation(s) on each end.
          Whiskers are 3.0 * IQR.
       4. FAST matches are those with speeds above 70.0 mph.
       5. SLOW matches are those with speeds below 5.0 mph.
       6. WEAK matches are those which match fewer than 3 digits.
       7. REPEAT matches are those in which the upstream tag also
          appears in a match on a shorter link.
       8. Matches with speeds below 0.1 mph
          or above 100.0 mph were ignored in the tag match.
```

Utilizing the Results

The results of TMatch's computational analysis are stored in the text file project-name.TMO, in the Output Files directory.
This file is formatted using spaces for display in a non-proportional font. It can be imported into any word processing /desktop publishing software. Just keep in mind that the columnar displays will need to either be set in a non-proportional font, or realigned for a proportional font using your software's column-definition commands.

More importantly, you must interpret the results carefully to be sure that the conclusions you draw are justified by the analysis!

B.5. <u>MENU COMMANDS REFERENCE</u>

All the functions of TMatch are organized into six menus of related commands.

The File Menu lists commands for opening, closing, and saving TMatch projects.
The Edit Menu commands allow for editing the TMatch project data, tag data, and text files.
The Tag Match Menu commands control the tag match procedure.
The LOS Menu Commands for computing the estimated LOS.
The Output Menu Commands for generating the report and output datasets.
The Options Menu Allows customizing of user-controllable options.

B.5.1. File Menu Commands

File...New
: Asks whether to save any unsaved data, then initializes a new TMatch study by clearing all data from memory.

File...Open
: Opens an existing TMatch study, clearing any existing data (after asking whether to save changes).

File...Save .
: Saves the current study, prompting for a name if the study has not already been named.

File...Save As
: Prompts for a new filename for the current study, then saves the current data under the new name.

File...Close
: Clears current study data, asking whether to save any unsaved changes.

File...Delete
: Deletes a file or TMatch project.

File...Exit
: Asks whether to save any current data, then exits TMatch.

.B.5.2. Edit Menu Commands

Edit...Route Data
: Brings up the Route Data dialog to allow entry or modification of the general roadway characteristics.

Edit...Tag Data
Brings up the Tag Data dialog to allow checking or editing of tag data collected at the observation stations.

Edit...Report File
Starts the user-specified text editor with the report file for the current project.

Edit...Text File
Starts the user-specified text editor on a file selected by the user.

B.5.3. Tag Match Menu

Run...Set Possible Speeds
Brings up a dialog box to allow the user to set the range of speeds detected during the tag match. Setting this range assists TMatch in detecting incorrect matches during the tag match. This range should cover ALL possible vehicle speeds, even speeds considered too extreme to be included in the final LOS computation.

Run...Tag Match
Runs the tag match, comparing the tag lists for each upstream/downstream station pair to find matching license tags.

Run...View Match Counts
Displays a list of the number of tags matched between each pair of observation stations. Double-clicking on a links count excludes that link from the statistical analysis.

B.5.4. LOS Menu

LOS...Set Rejection Parameters
Run...Compute LOS
Run...Preview Statistics

B.5.5. Results Menu

Results...View Results
Results...Create Match Dataset
Results...Create Summary Dataset

B.5.5. Options Dialog

Storage Paths:

Project Files Path : This is the default path for storing study (.TMR) files.

Input Files Path : The default path for tag data files.

Text Files Path : This is the default path for all TMatch text files; in particular, the report and dataset files are created in this path.

Scratch Path : Storage path for temporary files created and deleted internally by TMatch.

Match Files Path: This is the base path for the matched-tag files. Each project has a subdirectory off this path in which its own match files are stored.

Operation Options:

Keep Match Files checkbox :
When this box is checked, match files are not deleted when a project is closed. This consumes some extra disk space, but saves the time of re-running tag matches.

Save Filter Settings checkbox :
Causes rejection filter settings to be saved and reloaded for each project. Otherwise, each loaded project starts with the default settings.

Discard All Tags with Missing Digits checkbox :
Causes all observed tags with missing digits to be excluded from the tag matching. This makes the tag match much faster, at the price of lost data.

Tag Digits Field :
Sets the number of digits in each license tag. Regardless of the number of digits recorded in the tag data files, TMatch formats all license tags to this number of digits. Extra digits are discarded from the left, or short tags are padded on the left with missing digits.

Text Editor Field :
Sets the application launched by TMatch to edit text files. The default is NotePad.

Report Options:

Include Peak Analysis checkbox :
Currently has no effect.

Include Match Counts checkbox : Causes the counts of matches used or discarded from the LOS computation to be displayed in a table in the report.

Include Station Comments checkbox :
Causes any comment lines in the tag data files to be appended to the report.

Include Matches checkbox :
Causes the actual matches observed to be appended to the report. Note : Checking this option tends to make the report too long for the default Text Editor, NotePad.EXE. If you check this box, you should probably change the Text Editor.

Color Options:

Double-click on these color blocks to cycle through the available colors.

Arrow Off Color Block : The color in which inactive arrows are displayed.

Non-Choice Arrow : The color in which the next step arrow is displayed, if it does not involve a choice.

Choice Arrow : The color of next step arrows when a choice must be made.

Control Buttons

OK Button : Accepts the current changes, exiting the dialog.

Cancel Button : Exits the dialog without saving any current changes to the options.

B.6. PROJECT / FILE FORMATS REFERENCE

Tag Data File

These are the text files which contain the license tags observed at each station.

Match File

These files are internally produced by TMatch, and contain special binary-format information on the detected matches. One file is produced for each pair of observation stations.

Report File

This is the space-formatted text file containing the results of the LOS computation.

Match Dataset File

This is a text file containing data for the matches observed for all the station pairs in a project. This file is suitable for use as a data file in statistical analysis packages.

Summary Dataset File

This is a comma-delimited file of the summary statistics for the observation station pairs.

Station Data File

- Comments
- Date line
- Start line
- Tag lines
- Stop line

Match File

- First Tag field
- First Time field
- Second Tag field
- Second Time Field
- Keep field
- Duplicate field
- Peak Period field

Report File

- Header
- Observation Stations section
- Study Results section
- Peak Analysis section
- Summary Statistics section
- Match Counts section
- Station Comments section
- Matches section

Match Dataset File

- Upstream station number
- Downstream station number
- Digits matched
- Time In
- Time Out
- Travel Time
- Travel Speed
- Filtered

Summary Dataset File

- Station Name
- Link Length
- Raw Travel Time
- Adjusted Travel Time
- Raw Speed
- Adjusted Speed
- Posted Speed
- Route Class

Appendix C

Voice Recognition Studies

VOICE RECOGNITION STUDIES

C.1. INTRODUCTION

Over the past five years, researchers have made remarkable progress in the development of continuous speech recognition. This involves the recognition of large vocabularies at rates equivalent to those of normal speech. As technology progresses, the hope is to provide a "seamless" means of data transfer. Seamless is meant to imply that the data can effortlessly be transferred from one media to another. For example, speech can easily be transferred to written text and written text can be converted back to speech.

The capability to deliver this type of service depends on the ability of researchers to solve existing signal degradation problems and to increase the robustness of the audio signals. Current problems with poor recognition due to background noise are hoped to be solved. Strides are also being made to solve the problems of extraneous speech rejection. Reducing the search time the speech software spends trying to match commands that do not exist is critical to continuous speech recognition. To compound the problem, these problems must be solved and implemented at a reasonable cost to the consumer(1).

Voice recognition has the potential to replace the labor intensive burden of transcribing audio tape license data for license matching studies. Voice recordings using cassette recorders offer inexpensive means of collecting and storing data, as well as, the ability to record larger sample sizes. During a cordon survey in the Boston area, data collectors achieved sampling rates of approximately 1,000 license tags per person per hour. Although voice recognition techniques were not used in the study, audio methods proved to provide larger samples than pencil and note pad or laptop computer (manual typing of data into computer)(2).

C.2. SELECTION OF VOICE RECOGNITION SOFTWARE

At the time this study was started, two major voice recognition programs were on the market. They were Verbex's Listen for Windows(3) and IBM's Continuous Speech Software (ICSS)(4). Both of theses programs were purchased and tested to determine which would be the most compatible with license tag studies.

C.2.1. Verbex Listen for Windows

"Listen" is a speaker dependent program, meaning that each person who will use the program must train the computer to understand his or her speech. Speaker dependent speech recognition programs offer more accuracy than speaker independent programs, but require considerable amounts of time for computer training. This means that every person intending to use the system would have to dedicate a lengthy amount of time to train the computer to recognize his or her own voice.

"Listen" and the required hardware were installed and two speech interfaces were trained to test the recognition capabilities of "Listen". The speech interfaces contained the commands for working in Windows and WordPerfect. Once it was established that "Listen" could accurately recognize live voice, an additional speech interface was trained using recorded voice. Recorded voice presented recognition difficulties for "Listen". In addition to lowered recognition, many inconsistencies were present, for example, when the same segment of recorded speech was played, the computer recognized it on the first and third presentations, but did not recognize it during the second presentation. Due to these problems, as well as the extensive voice training, the Verbex Listen system was determined incompatible for use with license tag studies.

C.2.2. IBM Continuous Speech Software (ICSS)

ICSS is a speaker independent voice recognition program. It is claimed by IBM that the software trains "itself" as a user speaks to the system. The ICSS software is very sensitive to the hardware that it is being run on. Because of this, it is important the recording gain control is set properly on the sound card and the threshold levels are set correctly in the ICSS system window. More information is presented in the next section on the actual field testing and accuracy rates of the ICSS Software.

C.2.3. ICSS Voice Recognition Rates

A program called VOICETAG was developed as an ICSS application to create a text file with license tag numbers and times in response to vocal input of the tag numbers. The use of VOICETAG in the field on a portable computer does not appear to be feasible at this time. Apparently the internal battery in the portable computer is not sufficient enough to operate the computer, sound card, and microphone. The portable computer on battery power recognized 0% of the VOICETAG commands issued to it.

The use of the software in an office desktop computer for conversion of audio tapes with tag data into TMI files does have some potential. Table C-1 and Figure C-1 summarize the recognition of license plates with a microphone in the laboratory under zero noise conditions versus those from audio cassettes recorded in the field with substantial background noise. Two different machines were tested to compare the compatibility of the different hardware with ICSS.

Table C-1. Summary of VOICETAG Recognition Rates

Trial	No. of Plates	Computer 1				Computer 2			
		Microphone in Laboratory		Field Audio Cassette		Microphone in Laboratory		Field Audio Cassette	
		No.	%	No.	%	No.	%	No.	%
1	169	58	34	96	57	100	59	77	46
2	185	60	32	89	48	116	63	77	42
3	178	15	8	99	56	88	49	91	51
4	161	24	15	72	45	88	55	81	50
5	219	30	14	20	9	119	54	52	24
6	255	45	18	26	10	126	49	49	19
7	242	34	14	30	12	111	46	42	17
8	229	42	18	28	12	112	49	39	17
Total	**1638**	**308**	**19%**	**460**	**28%**	**860**	**53%**	**508**	**31%**

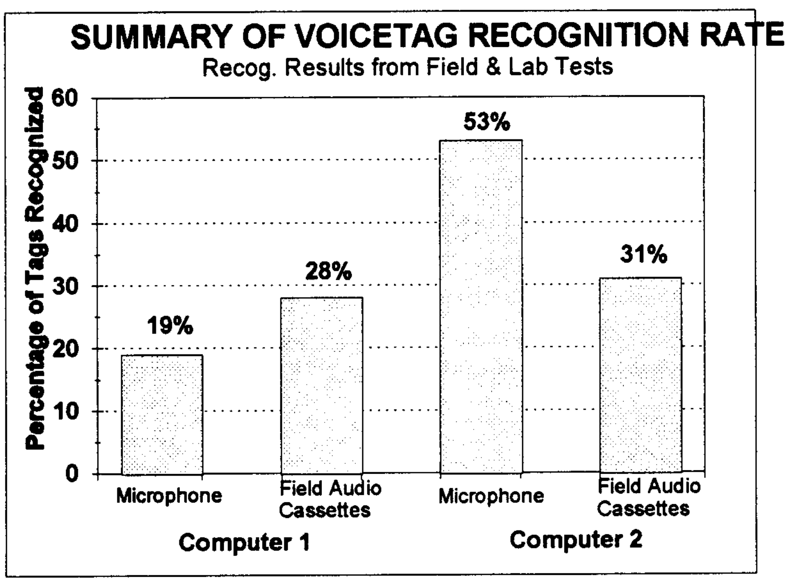


Figure C-1. Summary Graph of VOICETAG Recognition Rates.

The resulting low recognition rates are unexpected for computer 1. On the other hand, the computer 2 results were better than expected. Computer 1 actually had a lower recognition rate for the samples taken in the quiet laboratory conditions versus the playback of the cassette recorder into the microphone. Unfortunately, the sound card on computer 1 was difficult to set threshold values to because of a continuous "hum" that was picked up by ICSS. This may be the result of low quality sound card hardware.

The problem with the ICSS voice recognition software is its computer dependence. Under ideal conditions and pre-tuned threshold values the VOICETAG program will recognize large samples of plates. Under any other condition, the recognition is poor. The tapes presented problems due to the loud noises of trucks passing. When this would happen, the VOICETAG program would typically miss two or three license tags while it was trying to match the truck noise to a known sound in the vocabulary.

Computer 2 did an good job in the laboratory by recognizing 53% of the license plate samples. Both computers did about the same recognizing license plates from the cassette recorder playback. It appears that the recognition rate of audio cassettes for both computers is around 30%.

C.2.4. Expected Sample Sizes From Tag Matching Studies Using Voice Recognition
Based upon the findings of the voice recognition study, a prediction percentage can be calculated to estimate the expected number of matched samples from a study, knowing the traffic volume. The voice recognition study revealed that approximately 30% of the tags recorded to audio tape will be recognized by VOICETAG. The sampling study demonstrated that approximately 15.9% of the traffic volume's tags were sampled (based on a sample selection criteria of white passenger cars). If the same sampling criteria is used, the expected tag matching percentage of the total volume would be 4.8% of the through traffic volume during the study.

C.3. MAXIMUM POSSIBLE SAMPLE SIZES

During the travel time study it was evident that the people collecting data were not pushing their potential limits for entering tag samples. Since 34th Street is only two-lanes and only white passenger cars were being sampled, it should come as no surprise that more plates could have been collected.

To compare the maximum sampling rates of the palm-tops versus the voice recognition collection, a small study was set up along West University Avenue in front of the O'Dome (four-lane section with right turn lane at North-South Drive). Two cassette recorders and two palmtops were used to record license tag data as before, but this time the observers were instructed to collect as many license plates as they possibly could. The results of the study are presented in
Table C-2 and Figure C-2, below.

Table C-2. Summary of Maximum Sample Sizes

Time Period (30 minutes)	Palmtops		Cassette Recorders	
	Observer 1	Observer 2	Observer 3	Observer 4
1	251	171	169	219
2	251	156	185	255
3	240	163	178	242
4	244	136	161	229
Subtotal	986	626	693	945
Total	1612		1638	
Tags/hour	**403**		**410**	

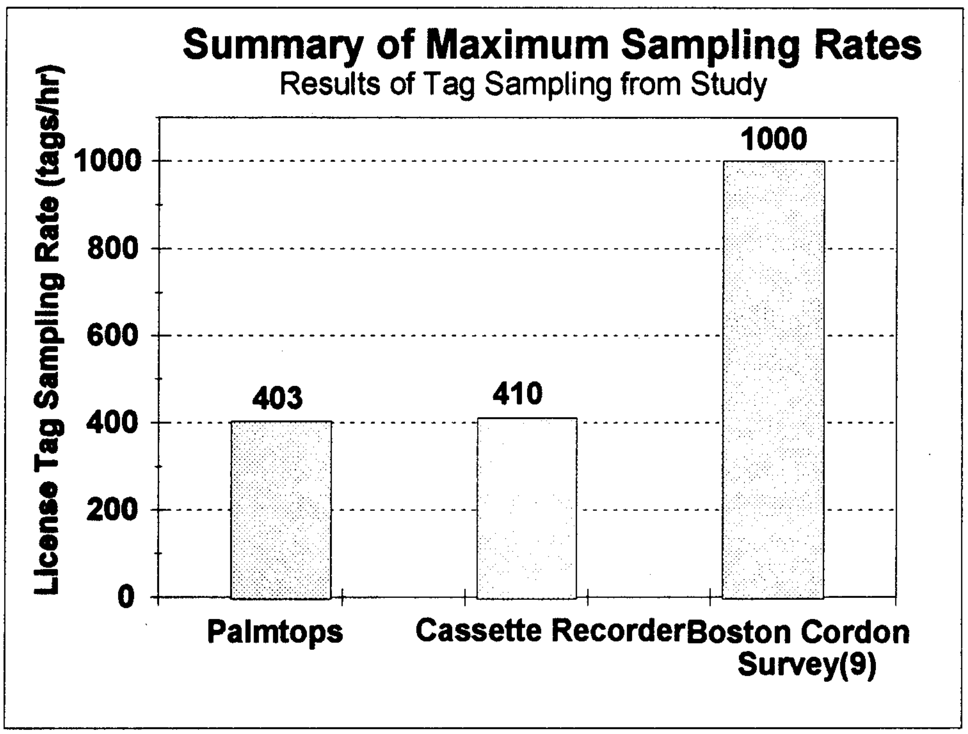


Figure C-2. Summary Graph of Maximum Sampling Rates from Field Study

It is clear that the sampling rate depends heavily upon the experience of the data collecting personnel. Observers 1 and 4 collected over 600 more license plates than observers 2 and 3 within a two-hour study. The sampling rates appear to be nearly identical for the palmtops and cassette recorders used in this study. It is somewhat surprising that the voice recorders did not show a distinct advantage over the palmtops. It must be remembered that tag entry into the cassette recorders is burdensome, since you have to say six voice commands for a single plate. This should also explain the difference in results between the cassette recorders used in this study and the cassette recorders used in the previously cited Boston study. During the Boston study voice commands were not used. Only the license letters and numbers were read into the recorders for later transcription.

C.4. APPENDIX C CONCLUSIONS

The overall conclusion of this report is that license tag matching does appear to be a feasible procedure for estimating travel times. The pilot studies described in this report indicated that adequate matching samples could be obtained. The software TMATCH demonstrated that it was capable of producing useful data. The voice recognition studies were not, however, as encouraging.

The results of this study indicate that voice recognition is indeed possible, but not necessarily the most desirable data acquisition method. The maximum recognition rates appear to be in the order of 50 percent of the total tag numbers observed. Many problems were experienced with fine-tuning of the equipment. It appears that the state of the art in this area has not reached the point where routine field studies could be carried out by traffic technicians with limited expertise in the specialized field of voice recognition.

It is interesting to note how the market perception of voice recognition has evolved in the past few years. At the time this study was initiated, there was a strong interest in voice recognition as a productivity aid in many areas of industry. The optimism of technological forecasters in this field has simply not materialized. A recent summary in the *Gainesville Sun*, reproduced below, suggests that there may be some hope for the future, but the use of voice recognition for field studies was an idea that was ahead of its time.

The option that proved the best in this study was the palmtop data collectors. They produced tag sampling rates equivalent, if not better than those achieved using cassette recorders and voice recognition. This is also the most direct means of data entry and is not subject to the uncertainties of voice transcription.

Gainesville Sun Clipping: September 23, 1996

IBM to market software that interprets speech

The New York Times

The long-sought goal of being able to speak directly to a computer rather than type on a keyboard took a step closer to reality this month when IBM announced new software that enables radiologists to dictate their reports directly into a personal computer.

The new offering, which results from years of speech-recognition research, is one of the first commercial products available that can recognize a large vocabulary spoken in a conversational tone. Other voice-recognition software that, for years, has been on the market from IBM and others requires the speaker to pause after each word to achieve a reasonable level of accuracy from a large vocabulary.

"What you see here is the first step toward the Holy Grail of continuous speech recognition," David Nahamoo of IBM said last week.

The technology is still not at the point where it can render any sentence, but it can be useful in specific disciplines like radiology, where many of the same technical phases are often repeated.

Radiologists spend most of their days interpreting images and sending reports of their findings to other doctors. The reports are often dictated into a recording machine, transcribed by a secretary and then corrected by the doctor — a process that can take days.

The new system recognizes about 25,000 English words and is about 95 percent accurate when first used, Nahamoo said. By spending perhaps 12 minutes training the software to recognize an individual's speech pattern, the accuracy can be improved. IBM plans to introduce similar products focused on other industries.

The software runs on the Microsoft Windows NT operating system and a personal computer powered by a 200-megahertz Pentium processor made by Intel Corp.

IBM said both the combined software and hardware would sell for $12,000 to $15,000 a unit. That is about equal to the amount that each of the country's 27,000 radiologists spends annually on transcription services, said David Wholley, the IBM executive who developed the product.

Appendix D

Adjustment Of Travel Times For Different Traffic Volume Conditions

ADJUSTMENT OF TRAVEL TIMES FOR DIFFERENT TRAFFIC VOLUME CONDITIONS

D.1. INTRODUCTION

The primary measure used to describe the performance of signalized intersections is delay. The Highway Capacity Manual (HCM) (Reference 1) provides a detailed procedure for estimating delay as a function of traffic volume and the operating parameters of an intersection. The planning and analysis process is complicated by the variations in delay that occur due to seasonal changes. While there exists a method for seasonal traffic volume adjustments, accurate estimates of delay are required to make the necessary decisions for planning purposes. The most accurate of these estimates would be direct field measurements, such as moving vehicle studies. Another method would be to use the HCM delay equation. Unfortunately, neither of these methods lends itself well to compensation for seasonal variations.

D.1.1. Problem Statement

Direct field measurement is probably the most accurate method to obtain delay for any given intersection. However, measured delays can only reflect the traffic conditions that prevail at the time of the field studies. It is clearly not practical to obtain field studies under all traffic conditions, and an analytical adjustment procedure is the only economical means of projecting delay estimates to other traffic conditions. The development of such a procedure is the subject of this appendix.

Analytical modeling is one way to compensate for the faults of direct field measurement. A model is inexpensive to use, readily available and allows for the determination of delay at different volumes. A widely used model describing signalized intersection delay is given in Chapter 9 of the HCM. The equation is dependent on the signal cycle length, green time, volume, and capacity. The model was developed empirically using data collected nationally. Like the direct measurement method, the HCM delay equation has its faults. Because the model is based on empirical data, an intersection with a given cycle length, green time, volume, and capacity may have a delay much different from that which is directly measured there. This situation then leads back to the problem that there is no way to increase the volume in the model and calculate a delay that accurately represents the intersection. Therefore, the HCM equation becomes invalid for this intersection and any effort to use it for this case would be futile.

This problem is illustrated in Figure D - 1. The measured data point of 48 sec/veh was obtained in a study in Miami, Florida. The intersection had a two lane approach with a signal cycle length of 120 seconds and a green time of 66 seconds. The volume measured at the intersection was 1158 veh/hr. The saturation flow rate was 1875 veh/hr. This gives a v/c=0.56. The curve in Figure D - 1 is from the model described in the 1994 HCM. As can be seen, if the delay at the intersection is desired for v/c=1.0 the HCM equation can not be used to determine the delay. If the HCM were used, a delay of 36.8 seconds would be calculated. When comparing this value with the measured delay at v/c=0.56 the problem is evident. The delay measured directly does fall on the HCM delay curve. Therefore, the HCM equation is invalid at this intersection.

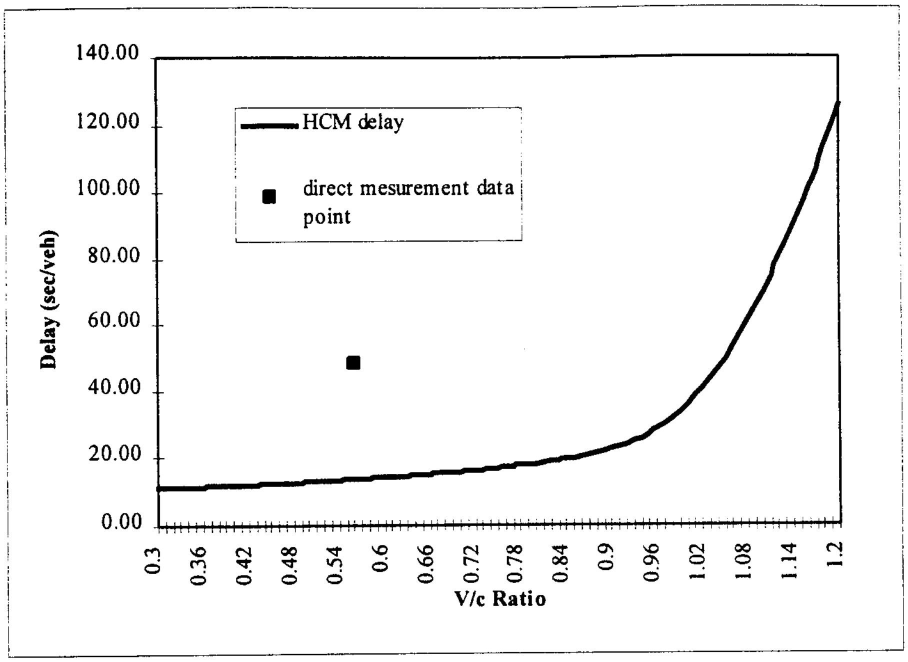


Figure D - 1. Illustration of potential incompatibility between the HCM delay estimation equation and field measurements.

It therefore is logical that an equation is needed to bridge the gap between the HCM delay equation and direct field measurement. This equation would allow the user to do a direct measurement study at an intersection and then calculate a new delay, for some increased volume, based on the measured delay and volume as well as other intersection characteristics.

D.1.2. Objectives

The objectives of the project described in this appendix are to develop a model that approximates the Highway Capacity Manual delay equation and to demonstrate that the model is a reasonable approximation of the HCM delay equation.

D.2. DEVELOPMENT OF THE DELAY ADJUSTMENT MODEL

Because the most common method of determining the delay at a signalized intersection is the procedure described in Chapter 9 of the Highway Capacity Manual, it stands to reason then that the model this paper proposes must be based on the HCM delay curve, yet still be general enough that it can be used at any intersection. The development of the model does follow this idea. The general shape of this new delay model curve must be consistent with the shape of the HCM delay curve. In addition to conforming in shape, the model must conform to general assumptions for both uniform delay regions and non-uniform delay regions. Essentially, the new model must not violate any assumptions or principles that are set by the HCM model.

D.2.1. The Highway Capacity Manual Delay Data Set

The Highway Capacity Manual delay equation is dependent on signal cycle length, lane group green time, approach volume, and approach capacity. In addition to the dependent variables two constants must be known; the delay adjustment factor (*DF*) and incremental delay calibration factor (*m*). Creation of the data set involved varying each independent variable with respect to the others. As for the two constants, coordinated semiactuated signals with arrival type 3 were assumed. The two assumptions are the default values recommended by the HCM. These values describe a common situation at signalized intersections. The independent variables were varied in order to develop a data set that would reflect most signal timing and approach volume situations. The ranges of the input parameters are shown in Table D - 1.

The traffic parameter values were needed in creating the data set based on the HCM delay equation. Using the ranges for each variable, and incrementing them one at a time created a five dimensional data set describing 72 individual intersections. For any given cycle length, green time ratio volume to capacity ratio, approach capacity, and volume to capacity ratio a delay was known and therefore, delay curves for 72 different signal situations were plotted.

Variable	Minimum	Maximum	Increment
Cycle Length	60	120	20
Green Time Ratio	0.20	0.80	0.12
Volume to Capacity Ratio	0.30	1.20	0.01
Number of Lanes	1	3	1

Table D - 1. Data Set Variable Parameters.

D.2.2. The Exponential Curve

A curve fitting the data set required the determination of what type curve would best fit the data, yet be simple enough to use and understand. It is known that any n^{th}-degree polynomial can be calibrated to fit a data set resulting in a smooth uniformly increasing curve. However, an n^{th}-degree polynomial would be difficult to explain in terms of how it relates to intersection delay. The only explanation would be that it fits the curve, which is what can be said about the HCM delay equation. The HCM delay equation was developed from a strictly empirical stand point. It is very difficult to explain why the equation is as complex as it is. By examining the data, an exponential function seems just as logical. The data appears to be roughly exponential. For approximation purposes, an exponential equation may be an adequate substitute to the

HCM equation. It certainly would be a much simpler equation to use, understand, and manipulate.

This is also confirmed by observing the shape of the two-way-stop intersection delay curve. This curve is explicitly defined by an exponential equation, Equation D - 1, and is shown in Figure D – 2 for a volume to capacity ratio range of 0.30 to 1.20.

$$D = e^{3.8(V/c)}$$

Equation D - 1

This delay curve follows the same trend as that for a signalized intersection shown in Figure D - 4, with the difference being the amount of delay observed at various volumes. The shape of an exponential curve is dependent on the constants. In the case of Figure D – 2, the 3.8 coefficient term defines the curvature, and by changing this constant the curve can be made deeper or shallower.

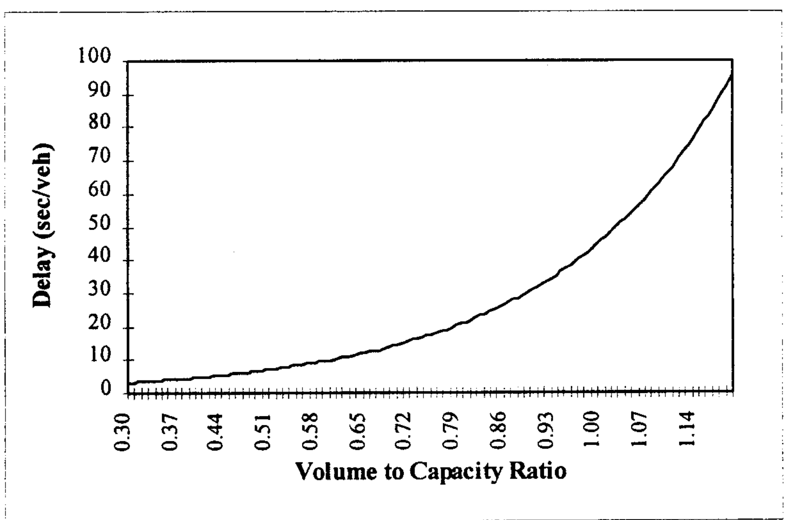


Figure D – 2. Two-Way Stop Intersection Delay Curve.

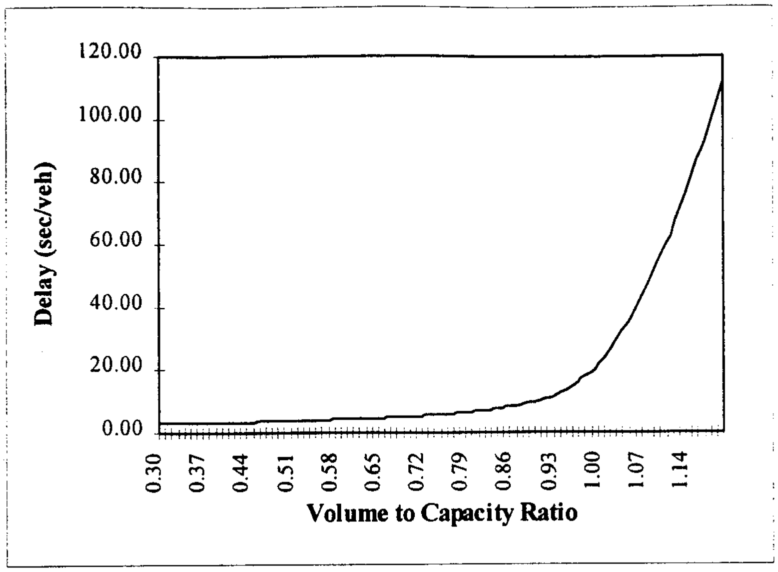


Figure D – 3. Signalized Intersection Delay Curve from HCM.

For the general exponential equation shown in Equation D - 2, where y and x are the independent and dependent variables, respectively, and the constants a, b, c, and d define curve.

$$y = a + be^{cx+d}$$

Equation D - 2

The constant a displaces the curve vertically; a change in a is proportional to a change in y. The constant b controls the magnitude of the exponential; a change in b causes a change in y as a function of the exponential. The exponent $cx+d$ changes the curvature of the function. An increase deepens the corner of the curve, while a decrease shallows the corner . Based on these facts, it seems logical that by manipulation of the constants any shape exponential curve can be created. Presumably even an exponential curve that approximates the HCM signalized intersection delay curve can be created.

The question then becomes just how precise an exponential approximation of the HCM delay curve will be. From observations it can be seen that the HCM delay curve approaches a line with definable slope in the $V/c>1.0$ range. An exponential curve approaches a vertical line as the independent variable increases. This obviously will create discrepancies for large V/c, but the exact degree of effect will have to be determined from more detailed analysis. For $V/c<1.0$ the exponential curve follows the HCM delay curve closely. Understanding that the HCM delay curve is least accurate for $V/c>1.0$, because of the simple fact that the least data was available for this range, some minimal error in this range may be more acceptable.

D.2.3. CURVE FITTING THE DATA SET

Fitting data to an exponential function is elementary. The natural log of each data point plotted will result in a straight line for data that is truly exponential. With this straight line, a linear regression can be performed to determine the coefficients. For the HCM delay data set that was generated, this process was attempted to find an approximating exponential function. Upon completing the first step of determining the natural log of each data point it was found that the delay data set was not linear in nature, and although the R^2 value for the corresponding linear regression was high, the resulting exponential equation produced approximations with unacceptable error. Visually, it could be determined that the curve resulting from the linear regression was not a good approximation of the HCM curve. The resulting equations did however provide a starting point for further manipulation.

The final manipulations were done manually. For these calculations Equation D - 2 was changed to Equation D - 3.

$$d = d_o + K_1 e^{K_2(V/c)}$$

Equation D - 3

For the conversion, $bx+c=K_2(V/c)$ and $a=d_o$, where V/c is the volume to capacity ratio and d_o is the vertical displacement. This change was made to minimize the number of constants and variables. All variables described for Equation D - 2 are represented, thus allowing for full manipulation of the exponential curve.

Assuming the delay, d_m, is measured for a given volume to capacity ratio, V_m/c, where V_m is the corresponding measured delay, Equation D - 3 becomes Equation D - 4.

$$d = d_m - K_1 e^{K_2(V_m/c)} + K_1 e^{K_2(V/c)}$$

Equation D - 4

Development of Equation D - 4 is illustrated in Figure D – 2. Equation D - 4 is defined by the independent variable V/c and the known values of d_m and V_m/c, yielding the dependent variable d. It can easily be seen that this is a single equation with a single unknown and two constants.

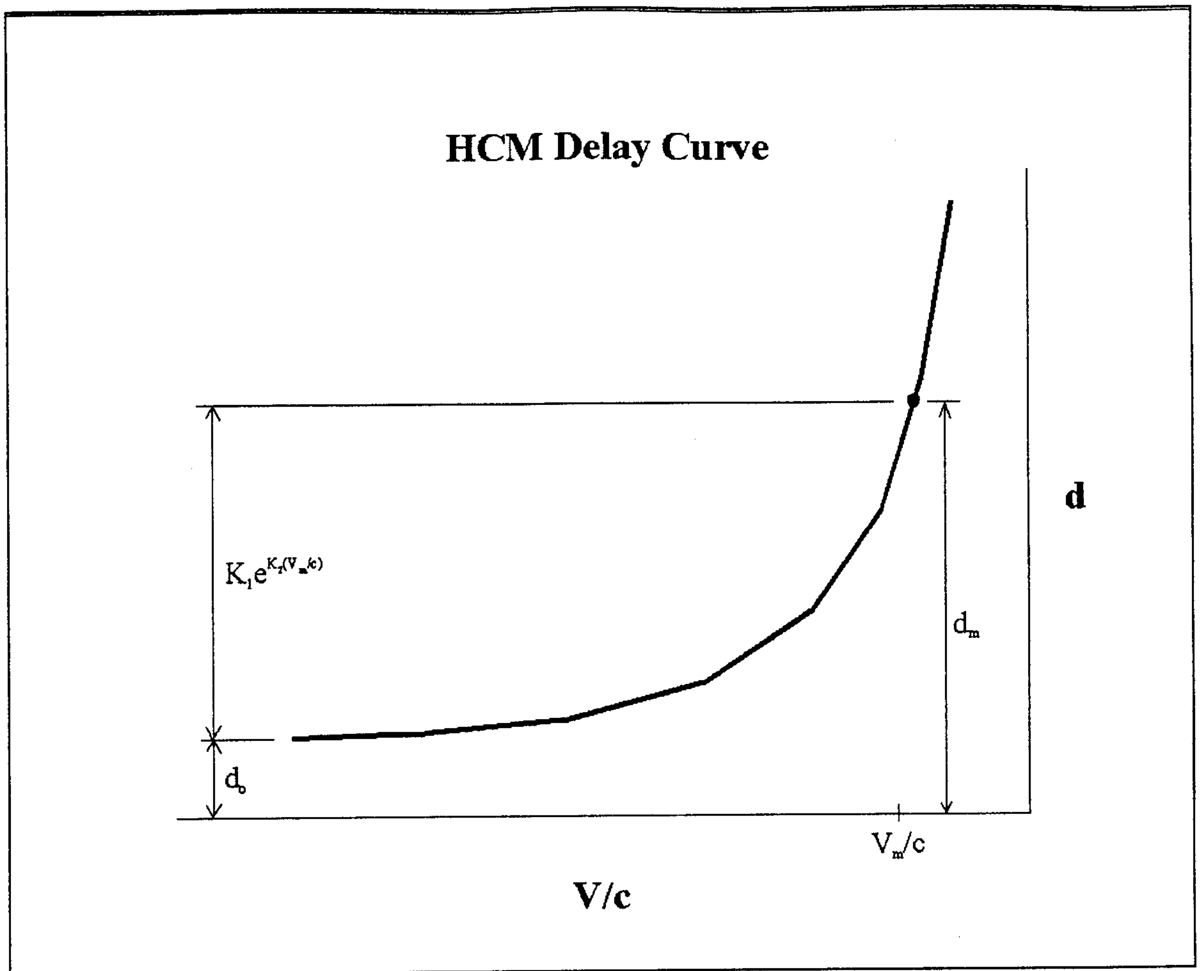


Figure D - 4. Exponential Delay Curve Derivation.

For Equation D - 4 to be useful, it would be highly desirable for the constants to be related to something that isolated each individual intersection. Therefore, determination of the constants K_1 and K_2 in terms of one or more intersection parameters is necessary. For each signalized intersection defined by the data set, Equation D - 3 was fitted manually. The manual fitting was performed in three steps. Step #1 was to set the approximating curve, defined by Equation D - 3, to the data set curve for each signalized intersection and adjust K_1 and K_2 to the best possible fit. This was done for all 72 signalized intersections, the values were recorded. Step #2 involved breaking down the values recorded in Step #1 by the principle parameters that describe each intersection; cycle length and capacity. A summary of the coefficients can be found in Appendix B. The plots of K_1 and K_2 are shown in Figure D – 5 and Figure D – 6, respectively. Step #3 involved determining equations that defined K_1 and K_2 in terms of capacity. From Figure D – 6, an exponential was fit using the method describe above in an attempt to find the approximating exponential function. The intermediate linear regression resulted in an R-square value of 0.910. From Figure D – 6, a linear function was fit. The linear regression resulted in an R-square value of 0.988. Appendix C provides information regarding the linear regression analysis for K_1 and K_2. From these results, equations for K_1 and K_2 were derived, defining K_1 and K_2 in terms of signalized intersection capacity (Equation D - 5 and Equation D - 6, respectively).

$$K_1 = 0.05796e^{-0.0011c}$$

Equation D - 5

$$K_2 = 6.4002 + 0.0009081c$$

Equation D - 6

The Highway Capacity Manual Signalized Intersection Delay approximation model is therefore defined by Equation D - 4, Equation D - 5, and Equation D - 6.

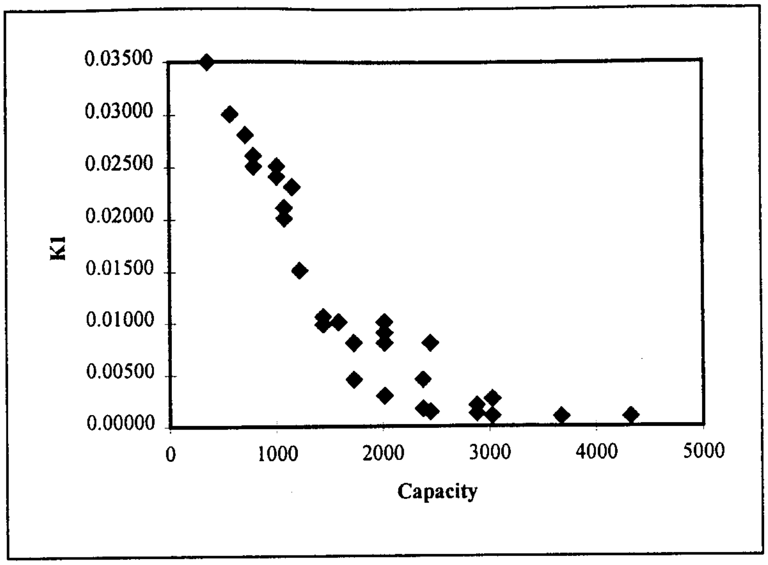


Figure D – 5. Capacity versus K1.

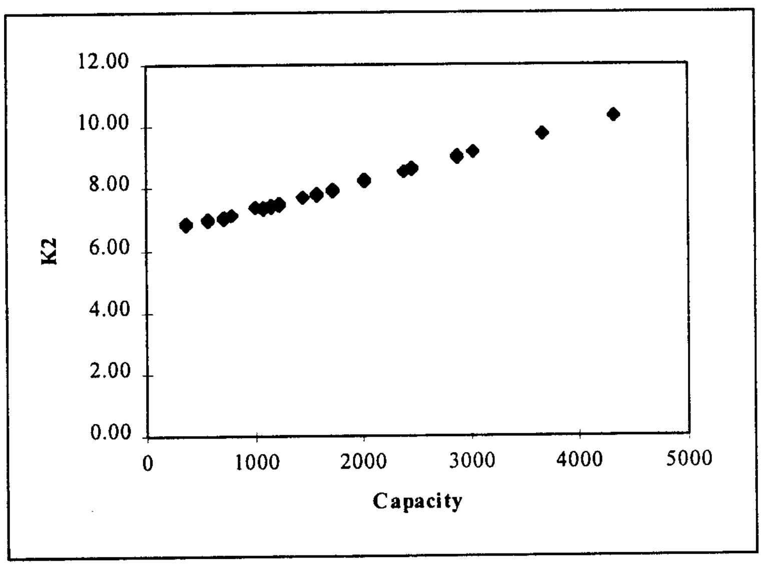


Figure D – 6. Capacity versus K2.

D.3. MODEL COMPARISION TO THE DELAY ADJUSTMENT MODEL

The development of the Delay Adjustment Model is directly based on data generated by the Highway Capacity Manual (HCM) Delay model. The Delay Adjustment Model was derived by curve fitting an exponential curve to the HCM data. The exponential equation defining the Delay Adjustment Model is a much simpler equation defining the HCM curve. Developing a simpler model was an objective of this research.

As with any approximation, a certain amount of error can be expected. Comparison of the Delay Adjustment Model to the HCM Delay Model is necessary to determine the quality of fit of the new model to the HCM model. Although the Delay Adjustment Model is developed from data produced by the HCM delay equation, it is necessary to determine the magnitude of the variance between the two models. This will provide information on the types of results and the reliability of the results to be expected when the Delay Adjustment Model is implemented. This section examines the shape of the curve of the Delay Adjustment Model compared to the curve produced with the HCM delay equation.

D.3.1. METHODOLOGY OF THE COMPARISON

The basic concept of this comparison was to place the Delay Adjustment Model on a point on the HCM delay curve and to determine the variation of the Delay Adjustment Model curve from the HCM delay curve. The variation was determined by calculating the absolute value of the difference between the Delay Adjustment Model value and the corresponding HCM equation value and dividing the result by the HCM equation value.

Three points of analysis, or central points, were selected: *V/c*=0.5, *V/c*=0.7, and *V/c*=0.9. Each was selected based on their respective locations on the delay curve. The delay curve has three distinct parts in which there is an obvious change in curvature. *V/c*=0.5 value was selected as representative of the horizontal or virtually horizontal portion of the delay curve; the uniform delay portion of the curve. *V/c*=0.9 was selected as representative of the virtually vertical portion of the delay curve, referred to as the incremental delay curve. The *V/c*=0.7 value represents transition from the uniform delay curve to the incremental delay.

The curves were analyzed for range 25% less than and greater than the central point. The range of ±25% was selected so that the range of one central point overlapped the range of the adjacent points. The delay was calculated from the Delay Adjustment Model and the HCM delay equation for each value of *V/c* from -25% to +25% of the central point. The value of *V/c* was incremented by 0.01 for each of the three ranges. The ranges are shown in Table D - 2.

Central Point	*V/c* Range
V/c=0.5	0.38 to 0.63
V/c=0.7	0.53 to 0.88
V/c=0.9	0.68 to 1.13

Table D - 2. *V/c* Ratio Ranges.

The percent difference between the HCM delay and the Delay Adjustment Model delay was calculated. The result was plotted against the percent spread from the central point. The resulting plots were visually inspected; the observations of each were recorded.

Each observation fell in one of five general categories that seemed to describe the fit of the Delay Adjustment Model to the HCM. For sorting and generalization purposes these five categories were assigned numeric values to summarize the quality of fit of the models. The five categories are described in Table D - 3 below.

Category 1	The percent difference between the HCM delay and the Delay Adjustment Model delay is less than 5%. The curve is below the unity slope line for all points.
Category 2	The percent difference falls between 5% and <10% and the curve is below the unity slope line for all points.
Category 3	The percent difference falls between 10% and <20% and the curve is below the unity slope line for all points.
Category 4	The percent difference is 20% or greater and the curve stays below or along the unity slope line in all areas.
Category 5	The curve is greater than the unity slope line in any or all areas.

Table D - 3. Descriptive Categories.

A sixth descriptive term was also defined: the unity slope line. The unity slope line is defined as the line on the % Difference versus % Variation plot where the percent difference is equal to the absolute value of the percent variation. For example, a point that has a +20% variation from the central point and a 20% difference falls on the unity slope line. Likewise, a point that has a -20% variation a 20% difference also falls on the unity slope line. The purpose of the unity slope line is to provide a frame of reference as to the shape of a curve.

The values of the percent difference defining the limits of each category one through five are the maximum value which is reached by the curve. Because of the methodology of the test, it is known that the difference in the two delay models will be zero at the central point. Therefore, the curve goes from zero to some maximum point on either side of the central point.

It should be noted that the curves are not necessarily uniform and may in fact oscillate over some ranges. The nonuniform and oscillatory nature of some of the curves is due to the shape of the two curves being compared. The HCM delay curve is essentially a complex polynomial curve. The Delay Adjustment Model is a natural log curve. In an effort to fit the natural log curve to the polynomial curve, the curves need to cross in more than one place. Typically the curves were set to intersect in two areas: the uniform delay portion of the curve and the incremental delay portion of the curve.

The analysis was performed for each of the central points specified in Table D - 2. The percent difference was then categorized by each case of signal characteristics, recorded, and sorted in reference to cycle length, *g/C* Ratio, and capacity. The purpose of the exercise was to examine trends in the percent difference between the two models over a wide range of signal characteristics.

D.3.2. ANALYSIS OF *V/c*=0.5 CENTRAL POINT

For the central point of *V/c*=0.5, the plot of the percent variation versus the percent difference falls within one of three categories: Category 1, Category 2, or Category 3. The plots do have a pattern. As the *g/C Ratio* increases from 0.2 to 0.8 the plots go from Category 1 to Category 3. The general area of transition from Category 1 to Category 2 occurs at a *g/C Ratio* of 0.44. The general area of transition from Category 2 to Category 3 occurs at a *g/c Ratio* of 0.68. The term "general area" of transition is used because the transition does not occur at an exact point. There are also several instances where there were specific signal characteristics which plot in a certain category but fall in an area dominated by another category. Specifically, all points with ±25% of the central point, *V/c*=0.5, have a difference of less than 20%. Also, only 11 curves fall in Category 3, all having a *g/C Ratio* greater than 0.68.

Most of the curves about the *V/c*=0.5 central point increase uniformly. In the area where the approach capacity is approximately 1400vph to 1500vph, there is a slight oscillation. However, the amplitude of the oscillation is no more than a 1% or 2 % difference.

Examples of the curves fitting Category 1, Category 2, and Category 3 for the V/c=0.5 central point are shown in Figure D - 7., Figure D - 8., and Figure D - 9., respectively. The curve in Figure D - 7. is defined by a capacity of 1728 vph, cycle length of 100 seconds, and *g/C Ratio* of 0.32. Figure D - 8. is defined by a capacity of

2376 vph, cycle length of 120 seconds, and *g/C Ratio* of 0.44. Figure D - 9. is defined by a capacity of 3672 vph, cycle length of 120 seconds and *g/C Ratio* of 0.68.

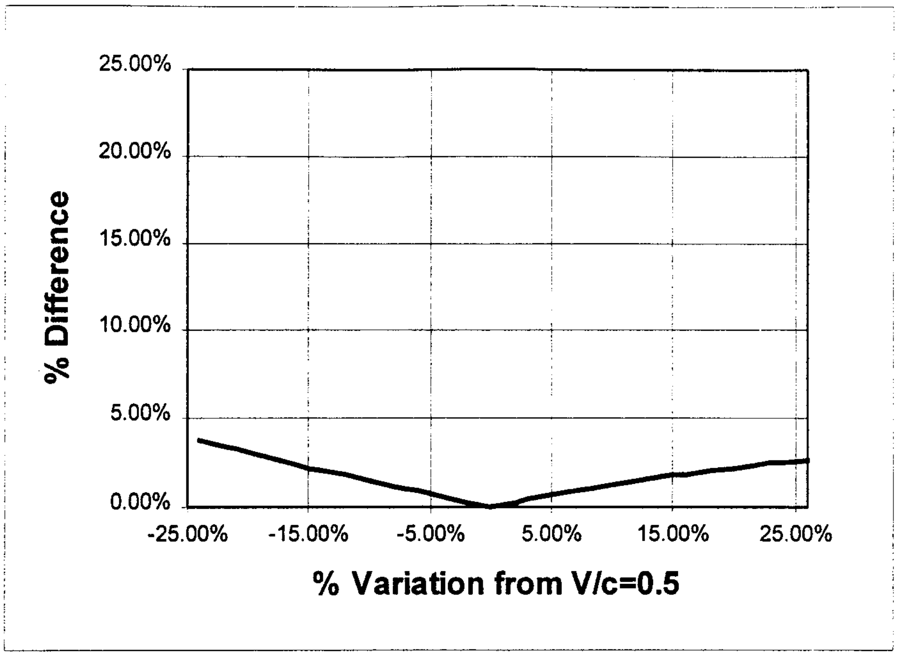


Figure D - 7. Category 1 volume variation effect.

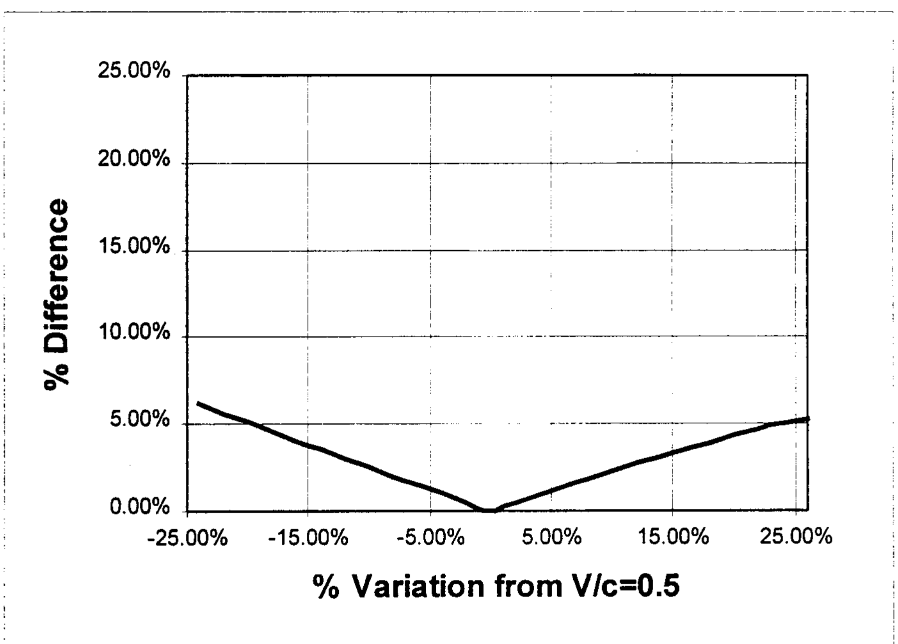


Figure D - 8. Category 2 volume variation effect.

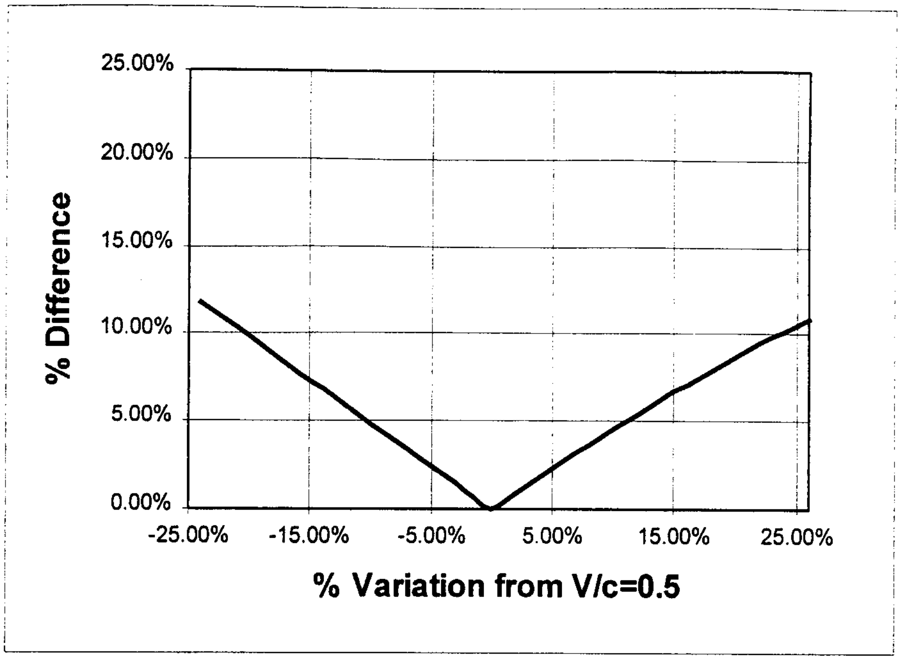


Figure D - 9. Category 3 volume variation effect.

D.3.3. ANALYSIS OF *V/c*=0.7 CENTRAL POINT

For the central point of V/c=0.7, the plot of the percent variation versus the percent difference falls within all of the five categories. The pattern is slightly different from the *V/c*=0.5 central point. As the g/C Ratio increases from 0.2 to 0.8 the plots go from Category 1 to Category 5. For the generated data set of signal characteristics there are a total of 10 that can be classified as either Category 4 or Category 5. These plots occur when the signal has a *V/c* Ratio of 0.8 and the cycle length is 60 or 80 seconds. The transitions are not as clear as those for the V/c=0.5 central point.

The curves for the V/c=0.7 central point have considerably more instances of oscillation than that described in the V/c=0.5 central point data set. The oscillations of the curves stem from the location of the delay curve on which the central point is found. V/c=0.7 falls in the area of transition between the uniform delay portion and the incremental delay portion. A goal of the derivation of the Delay Adjustment Model was to position the Model's curve so as to, as near as possible, approximate the HCM delay curve. To accomplish this the Delay Adjustment Model curve was allowed to intersect the HCM delay curve, sometimes at more than one point. These points were usually determined to be just to either side of the transition curve. The result of the curve may be a model that slightly underestimates the HCM model through the uniform delay and incremental delay portions of the curve while slightly overestimating the transition area between the two. For virtually all cases the oscillation amplitudes were less than 5%.

Examples of the curves fitting all five categories for the V/c=0.7 central point are shown in Figure D - 10. through Figure D - 14.. Figure D - 10. is a Category 1 example and is defined by a capacity of 576 vph, cycle length of 60 seconds, and *g/C Ratio* of 0.32. The slight oscillation discussed above can be seen for negative variation from the central point. The example of Category 2, shown in Figure D - 11. is defined by a capacity of 1584 vph, cycle length of 100 seconds, and a *g/C Ratio* of 0.44. The Category 3 curve is shown in Figure D - 12.. It is defined by a capacity of 2374 vph, cycle length of 80 seconds, and a *g/C Ratio* of 0.44. The Category 4 curve, shown in Figure D - 13., is defined by a capacity of 1224 vph, a cycle length of 60 seconds and a *g/C Ratio*=0.68. As can be seen from the figure, the curve follows the unity slope line for all points of percent variation from the central point. Figure D - 14. shows the Category 5 example. It is defined by a capacity of 1440 vph, a cycle length of 80 second, a *g/C Ratio* of 0.8.

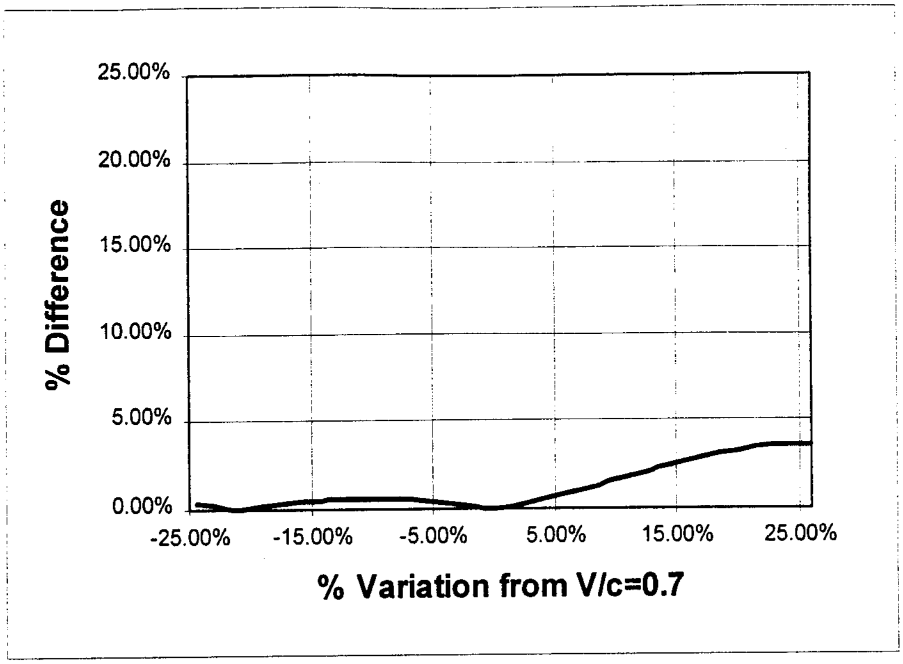


Figure D - 10. Category 1 volume variation effect.

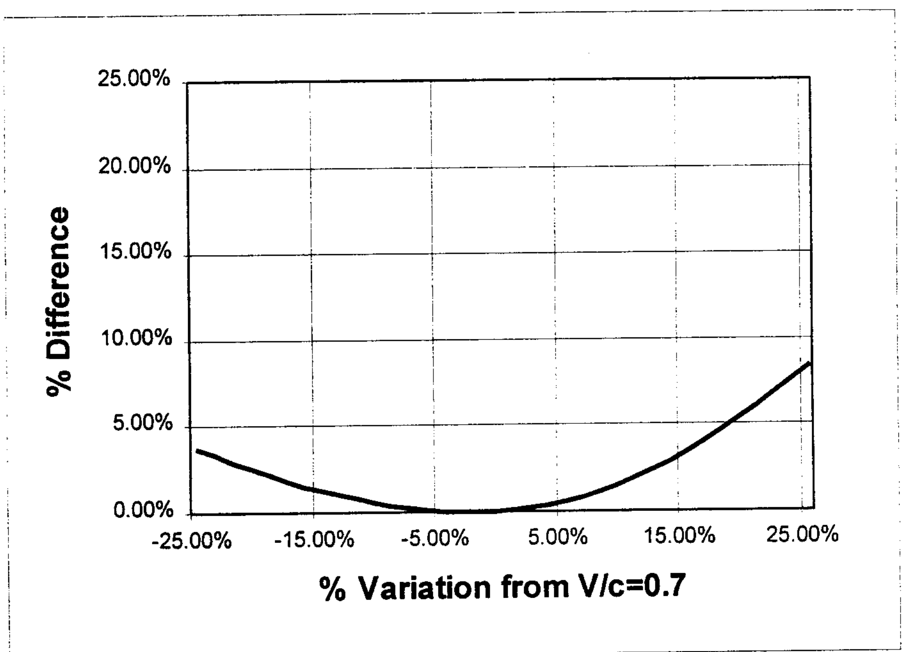


Figure D - 11. Category 2 volume variation effect.

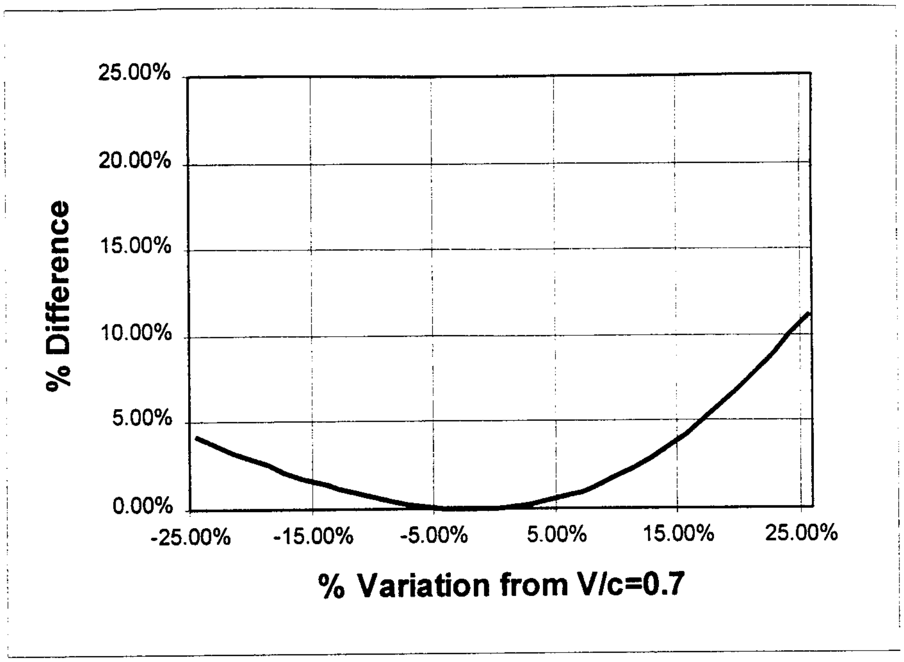


Figure D - 12. Category 3 volume variation effect.

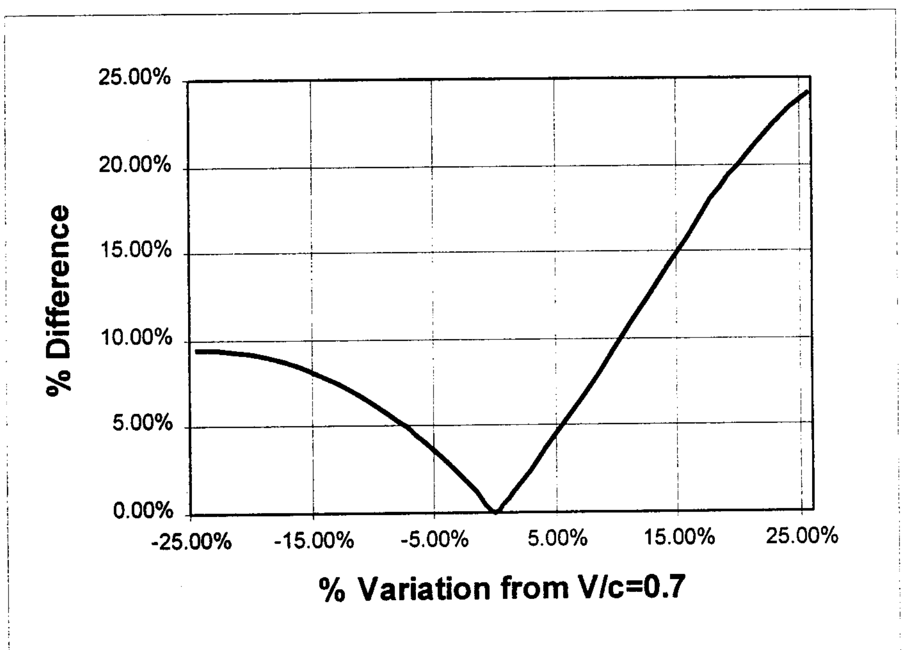


Figure D - 13. Category 4 volume variation effect.

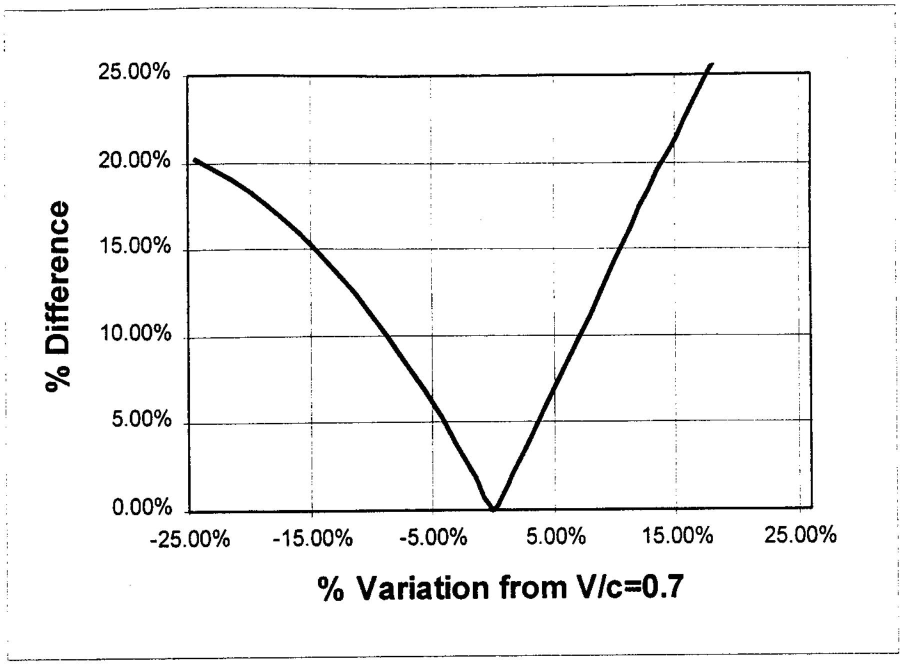


Figure D - 14. Category 5 volume variation effect.

D.3.4. ANALYSIS OF *V/c*=0.9 CENTRAL POINT

For central point *V/c*=0.9, the plots of percent variation versus percent difference fall within all five categories. As with the previous central points, the plots go from Category 1 to Category 5 as the *g/C Ratio* increases, although the changes are not as uniform as for central point *V/c*=0.5 and *V/c*=0.7. The plots seem to fall consistently into Category 5 for *g/C*=0.68 and *g/C*=0.8. There is only a single instance were a plot falls within Category 1. This occurs at *g/C*=0.2 and a cycle length of 120 seconds. A majority of the plots fall within Category 3 or Category 4.

All curves either oscillate or start to oscillate for points with positive variation from the central point. As with the oscillations around the central point of *V/c*=0.7, the cause is due to the fact that a natural log curve is being fit to a multinomial curve. For positive variation to the central point, the HCM curve is virtually linear, where in the same area the natural log curve still has curvature. As was discussed in the previous section the objective was to most reasonably fit the natural log curve to the HCM curve so that the result of the approximation was acceptable.

Examples of Category 2, Category 3, Category 4, and Category 5 curves are shown in Figure D - 15., Figure D - 16., Figure D - 17., and Figure D - 18., respectively. The Category 2 curve is defined by a *g/C Ratio* of 0.2, cycle length of 100 seconds and a capacity of 360 vph. The Category 3 curve is defined by a *g/C Ratio* of 0.44, cycle length of 60 seconds, and capacity of 792 seconds. The Category 4 curve is defined by a *g/C Ratio* of 0.56, cycle length of 80 seconds, and a capacity of 1,008 vph. The Category 5 curve is defined by a *g/C Ratio* of 0.8, cycle length of 120 seconds and a capacity of 4,320 vph.

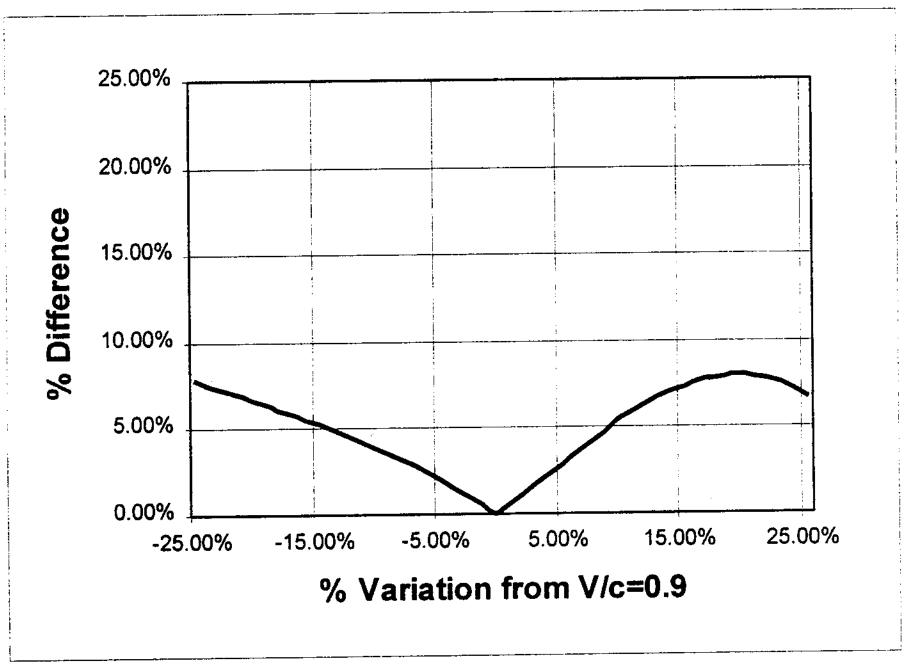


Figure D - 15. Category 2 volume variation effect.

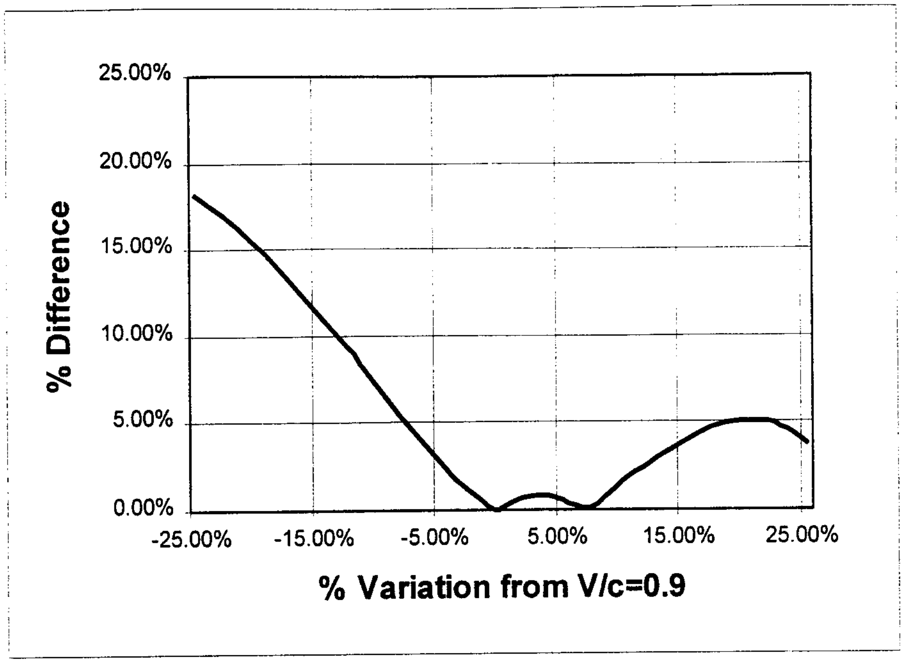


Figure D - 16. Category 3 volume variation effect.

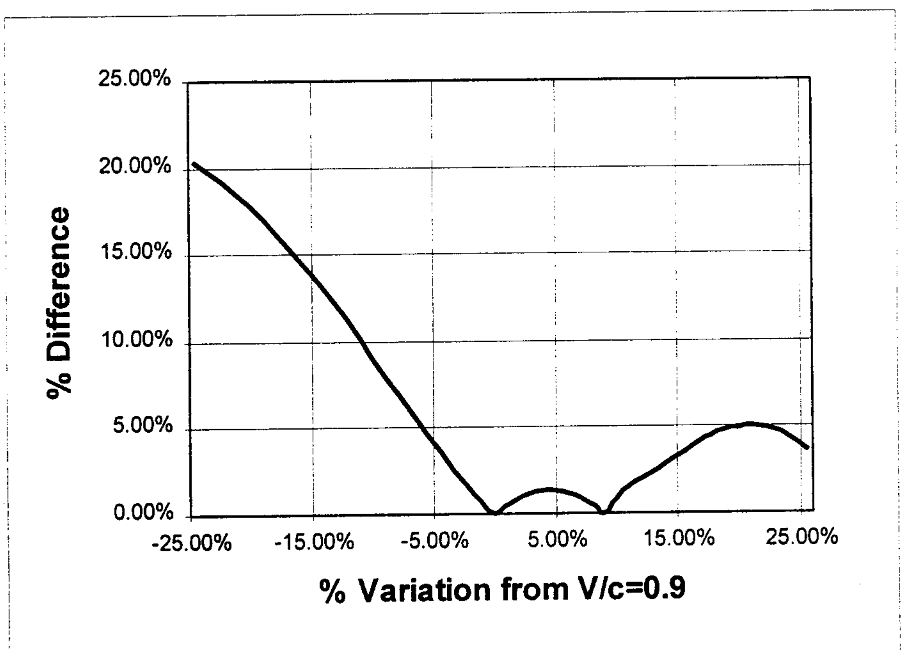


Figure D - 17. Category 4 volume variation effect.

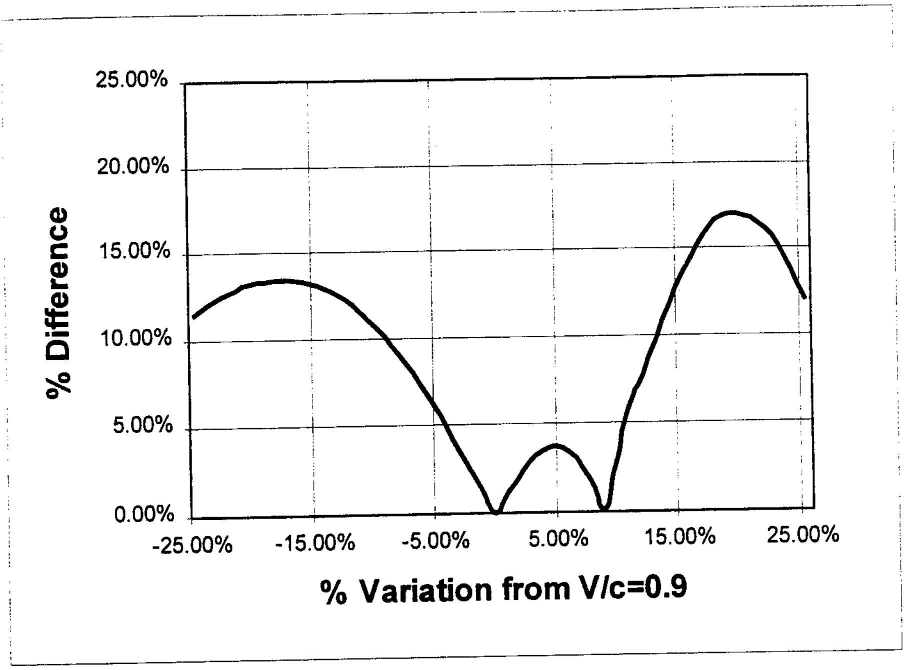


Figure D - 18. Category 5 volume variation effect.

D.4. CONCLUSIONS AND RECOMMENDATIONS

This project demonstrated that a simple approximation of the Highway Capacity Manual (HCM) delay equation could be developed. The approximation, referred to as the Delay Adjustment Model, is dependent on three variables: the approach capacity of the intersection, the existing volume of the intersection, and the stopped delay at the intersection corresponding to the existing volume. With this information, the stopped delay at the intersection can be projected for some increased volume. The limitation as to what increase in volume the delay could be projected was also presented.

The model was compared to the HCM delay curve and the fit of the delay curve created by the Delay Adjustment Model does approximate the HCM delay curve. The percent difference in the two curve was calculated and plotted in relation to the percent difference in the volumes. The results were consistent with what would be expected results of measurements of stopped delay at signalized intersections. The Delay Adjustment Model was not found to violate any principle assumptions of stopped delay at signalized intersections.

Within the limits of the study parameters, it is suggested that the exponential approximation technique described herein can be applied with reasonable confidence under conditions of midrange v/c and g/C ratios. The technique begins to lose accuracy when volumes approach capacity and when green times are very long with respect to the cycle time.