Akcelik & Associates Pty Ltd
PO Box 1075G, Greythorn, Vic 3104 AUSTRALIA
info@sidrasolutions.com

Management Systems Registered to ISO 9001
ABN 79 088 889 687

REPRINT

Field Assessment of the Performance of Computer-Based Signal Timing Models at Individual Intersections in North Carolina

STEVEN M. CLICK, and NAGUI M. ROUPHAIL

REFERENCE:

CLICK, S.M. and ROUPHAIL, N.M. (2000). *Field Assessment of the Performance of Computer-Based Signal Timing Models at Individual Intersections in North Carolina*. Final Report for the North Carolina Department of Transportation. Center for Transportation Engineering Studies, Department of Civil Engineering, North Carolina State University.

NOTE:

This document has been made available with permission from its authors and the North Carolina Department of Transportation. Appendices of the report are not included in this reprint.

PO Box 1075G, Greythorn Victoria 3104, Australia
Email: info@sidrasolutions.com

Field Assessment of the Performance of Computer-Based Signal Timing Models at Individual Intersections in North Carolina

FINAL REPORT

Mr Steven M. Click
Dr Nagui M. Rouphail

Submitted to

The North Carolina Department of Transportation

April 2000

FIELD ASSESSMENT OF THE PERFORMANCE OF COMPUTER-BASED SIGNAL TIMING MODELS AT INDIVIDUAL INTERSECTIONS IN NORTH CAROLINA

FINAL REPORT

Submitted to

The North Carolina Department of Transportation

by

Nagui M. Rouphail, P.I.
North Carolina State University

Steven M. Click, R.A.
North Carolina State University

April 2000

Table of Contents

1. Introduction 1
1.1 About this Document 1
1.2 Purpose and Objectives 1
1.3 Scope 2
1.4 Computer-based Signal Timing Models 2
1.5 Final Report Organization 3
1.6 Model Selection 4
1.7 Naming Conventions 5

2. Model Overview and Characteristics 6
2.1 Introduction 6
2.2 Model Introduction 6
2.3 Model Inputs 8
2.3.1 Introduction 8
2.3.2 Input Interfaces 8
2.3.3 Input Parameters: Geometry 9
2.3.4 Input Parameters: Volumes and Flow Characteristics 10
2.3.5 Input Parameters: Signal Settings 11
2.3.6 Model Ease of Use 12
2.4 Model Optimization Procedures 14
2.4.1 Phase Optimization 14
2.4.2 Optimization Parameters 15
2.4.3 Overall Optimization Capabilities 17
2.5 Model Outputs 18
2.5.1 Introduction 18
2.5.2 On-Screen Outputs 18
2.5.3 File and Printed Outputs 20
2.5.4 Output Summary 22
2.6 Conclusions And Recommendations 23
2.6.1 Introduction 23
2.6.2 Overall Model Comparison 23

3. Data Collection Planning 25
3.1 Introduction 25
3.2 Data Collection Planning 25
3.2.1 Step 1: Define the Experiment's Purpose, Importance, and Scope 25
3.2.2 Step 2: Determine the Experimental Unit 26
3.2.3 Step 3: Define the "Total" Population 26
3.2.4 Step 4: Select an Experimental Population 27
3.2.5 Step 5: Select an Analytical Method 27
3.2.6 Step 6: Search for Existing Data 28
3.2.7 Step 7: Data Collection Planning 28
3.3 Purpose, Relevance and Scope 28

3.4 Selecting An Experimental Unit....29
3.4.1 Introduction 29
3.4.2 The Intersection Unit 29
3.4.3 The Approach Unit....30
3.4.4 The Lane Group Unit....30
3.4.5 The Lane Unit 31
3.4.6 Chosen Experimental Unit....31
3.5 Potential Site Characteristics....32
3.5.1 Introduction 32
3.5.2 Factors....32
3.5.3 Conclusions....34
3.6 Identification of Key Factors....35
3.6.1 Lane Group Types....35
3.6.2 Left Turn Treatment....35
3.6.3 Controller Type....35
3.6.4 Volume to Capacity Ratio 35
3.7 Summary....36

4. Data Collection and Reduction 37
4.1 Introduction 37
4.2 Site Selection....37
4.3 Data Collection....39
4.3.1 Queue Data 39
4.3.2 Flow Data 39
4.3.3 Geometric Data....39
4.3.4 Timing Data 40
4.4 Data Reduction....40
4.4.1 Geometric Data....40
4.4.2 Demand Data....40
4.4.3 Timing Data 41
4.4.4 Delay Data....43
4.5 Summary....46

5. Model Evaluation 47
5.1 Introduction 47
5.2 Evaluation Methodology 47
5.2.1 Introduction 47
5.2.2 Modeling 47
5.3 Comparisons....49
5.4 Intersection Level Comparisons....49
5.4.1 Visual Comparisons....49
5.4.2 Regression Comparisons 53
5.4.3 Stepwise Inclusion of Variables....54
5.5 Lane Group Comparisons....57
5.5.1 HCM Evaluation....58
5.5.2 EVIPAS Evaluation....59

5.5.3 SIDRA Evaluation 60
5.5.4 SIGNAL94-Alt Evaluation 61
5.5.5 TRANSYT-7F Evaluation 63
5.6 Summary 64
5.7 Model Rating 66

6. Optimization 67
6.1 Introduction 67
6.2 Methodology 67
6.2.1 Optimization Intervals 67
6.2.2 Phasing Plans 68
6.2.3 Comparison Methodology 68
6.3 Optimization Results 69
6.3.1 EVIPAS 69
6.3.2 SigCinema 72
6.3.3 Signal94 75
6.3.4 Sidra 79
6.3.5 Transyt 81
6.4 Summary 83

7. Summary, Conclusions and Recommendations 85
7.1 Project Rationale 85
7.2 Summary 85
7.3 Conclusions 87
7.3.1 Model Usability 87
7.3.2 Model Evaluation 88
7.3.3 Model Optimization 89
7.4 Study Recommendations 90

8. Appendix: Map of Data Collection Sites 8-1

9. Appendix: Data Collection Site Intersection Schematics 9-1

10. Appendix: Example Data Reduction - Site 1 AM Data 10-1

11. Appendix: Copy of Data Reduction Methodology (from the Draft 1997 HCM) 11-1

12. Appendix: Graphs of Field vs. Model Delay by Lane Group and by Model 12-1

Table of Figures

Figure 1. Organization of the Document 4
Figure 2. Model Naming Conventions used in the Document 5
Figure 3. Geometric Inputs 10
Figure 4. Volume and Flow Inputs 11
Figure 5. Signal Setting Inputs 12
Figure 6. Summary of Phase Optimization Capabilities....15
Figure 7. Optimization Options 17
Figure 8. On-screen Outputs....20
Figure 9. File and Printed Output 22
Figure 10. Initial Model Rating 24
Figure 11. Factors and Their Levels....33
Figure 12. Summary of Which Lane Group the Factors Affect 34
Figure 13. Summary of Characteristics of Data Collection Sites....38
Figure 14. Sample Calculation of Green Time for a Phase 41
Figure 15. Determination of Phasing Plan from Average Movement Green Times 42
Figure 16. Sample Calculation of Queue Delay for a 15 minute Period 44
Figure 17. Graph of Intersection Level Field Delay vs EVIPAS Estimate 50
Figure 18. Graph of Intersection Level Field Delay vs HCM Estimate 50
Figure 19. Graph of Intersection Level Field Delay vs SIDRA Estimate 51
Figure 20. Graph of Intersection Level Field Delay vs SIGNAL94-Alt Estimate 51
Figure 21. Graph of Intersection Level Field Delay vs. TRANSYT Estimate....52
Figure 22. Intersection Level Regression Results (All V/C)....54
Figure 23. Intersection Level Regression Results (V/C < 1.0)....54
Figure 24. Variables included in Stepwise Regression Analysis 55
Figure 25. Results of Stepwise Regression Analysis for All VC 56
Figure 26. Results of Stepwise Regression Analysis when VC < 1.0 56
Figure 27. HCM-based Stepwise Regression Results 58
Figure 28. EVIPAS Stepwise Regression Results....59
Figure 29. SIDRA Stepwise Regression Results....60
Figure 30. SIGNAL94-Alt Stepwise Regression Results....61
Figure 31. TRANSYT-7F Stepwise Regression Results....63
Figure 32. Summary of Stepwise Regression Results 64
Figure 33. Model Evaluation Rating (1-5, 5 best) 66
Figure 34. Key to Summary Tables....70
Figure 35. Summary of EVIPAS Optimization Results 71
Figure 36. Summary of SigCinema Optimization Results....73
Figure 37. Summary of Signal94 Optimization Results....76
Figure 38. Summary of Signal94 Alternate Optimization Results....78
Figure 39. Summary of SIDRA Optimization Results 80
Figure 40. Summary of Transyt Optimization Results....82
Figure 41. Graphical Summary of Optimization Results from All Models....83
Figure 42. Summary of Rankings of Model Optimization Results (1-5, 5 best)....84
Figure 43. Summary of Model Usability Ratings (0-5)....87

Figure 44. Stepwise Regression Results- Intersection Level..88
Figure 45. R-Square Regression Results at the Lane Group Level 89
Figure 46. Summary of Model Optimization Results..90

1. Introduction

1.1 About this Document

This document is the final report for the research project entitled “Field Assessment of the Performance of Computer-Based Signal Timing Models at Individual Intersections in North Carolina.” This two-year research project was sponsored by the North Carolina Department of Transportation (NCDOT) and supervised by the Signals and Geometrics Section of the Department.

1.2 Purpose and Objectives

This project was designed to give valuable, timely information to NCDOT about the quality of results from currently available computer-based isolated intersection signal timing models. Because the outputs of such models are used to make operational decisions, the results from this research should become part of day-to-day engineering consideration at DOT’s across the country. If a particular model is inaccurate in its estimation of intersection performance parameters, its use may lead to poor engineering decisions, and thus poor field operation.

The objectives of this research can be summarized as follows:

1. Review computer-based signal timing procedures for isolated intersections currently in use for pretimed or actuated operations.
2. Identify desirable model attributes for inclusion in signal timing decisions. This may include engineering attributes such as the ability to handle certain intersection conditions or more subjective attributes such as user friendliness.
3. Evaluate a select set of signal timing models that meet Objective 2 (above). This set should include current models routinely used by NCDOT.
4. Calibrate and validate the selected models in Objective 3 (above) using North Carolina data collection sites covering the range of geometric and traffic conditions encountered in the state.

5. Recommend signal timing changes for the data collection sites based on the optimization features of the selected models. Evaluate intersection performance after the proposed timings are implemented.
6. Develop recommendations for NCDOT about which models to use for given traffic and geometric conditions. This process could involve the use of more than one model for a particular site or configuration.

1.3 Scope

This study is limited to isolated intersections in North Carolina. No attempt was made to evaluate model capabilities for signal systems. No field data were gathered outside the Research Triangle area of North Carolina. While the findings of this research may be applicable to other areas, no attempt has been made to carry out a global evaluation of the models.

Also, the study was limited to a finite number of computer-based signal timing models. As computer software is almost continually being upgraded and revised, the study team was limited to a single version of each software package and used it throughout the study. Therefore, the results of this study are not necessarily applicable to different versions of the software packages. Nor are the results applicable to “clones” of any piece of software considered. Further information about the specific software versions used in this study are given in Chapter 2.

1.4 Computer-based Signal Timing Models

This study focused on computer based signal timing models implemented via computer software using PC-based platforms. In addition, the majority of the models considered during the study had an “optimization” feature. This was necessary to fulfill the stated objectives of the project.

An “optimization” feature is a capability on the part of the software package to suggest what the signal timing parameters should be for a given set of geometric and traffic conditions. Signal parameters include such things as cycle length, phase selection and sequence, left turn treatments, and green time allocation.

1.5 Final Report Organization

During the course of the study, the study team naturally subdivided the research project into three natural divisions or phases. As the sections in this report are parallel to the project phases, a short description of each of the phases follows.

- **Phase 1: Model Overview**. The purpose of this phase was to compare and contrast the models based on their stated capabilities and their user-friendliness. Stated capabilities included available and required inputs, simulation and optimization capabilities, and output parameters given both on screen and in printed or file form. User friendliness evaluations covered input modules and output navigation. This phase of the project consumed about six months of the project and its results were documented in a working paper. A by-product of this step was the final selection of models to be included in the rest of the study.
- **Phase 2: Model Evaluation**. The purpose of this phase was to see how well each of the models simulated field conditions. The study team gathered field data on intersection geometrics, traffic conditions, and delay. The geometric and traffic condition data were entered, and the model predictions of delay were compared with field delays at both the intersection and lane group level. This phase of the project consumed about a year and generated results which were presented at two separate one-day workshops for NCDOT personnel. The results presented at those workshops are reported here.
- **Phase 3: Optimization**. The purpose of this phase of the project was to evaluate the optimization capabilities of each of the models. The models optimized phasing for a subset of the data from Phase 2 and the results from each model were compared. This phase of the project consumed about six months of the total project time. A small portion of the results from this phase were presented at the second one-day workshop mentioned above. Those plus the remainder of the results are presented in this document.

This report contains complete documentation of all three phases of the project. The remainder of this report is organized as shown in Figure 1.

Figure 1. Organization of the Document

Chapter 2	Model Overview and Program Characteristics	(Phase 1)
Chapter 3	Data Collection Planning	(Phase 2)
Chapter 4	Data Collection and Reduction	(Phase 2)
Chapter 5	Model Evaluation	(Phase 2)
Chapter 6	Optimization Experiments	(Phase 3)
Chapter 7	Conclusions and Recommendations	

1.6 Model Selection

As noted before, one of the results of Phase 1 of the project was a final selection of models for use in the project. A summary of the steps taken to make that selection follows.

1. **Identification of available models**. The study team, in conjunction with the project's Technical Advisory Committee (TAC), identified all the models currently in use by NCDOT or by consultants which have submitted work to NCDOT.
2. **Initial cut**. The TAC made some initial recommendations about which models to include based solely on frequency of use.
3. **Initial evaluation**. The study team made an initial evaluation of the remaining models. This evaluation was based on several criteria, including required inputs, optimization capabilities, flexibility, model outputs, and user friendliness. The results of this evaluation were presented to the TAC.
4. **Final cut**. Based on the initial evaluation, the TAC and study team chose five (5) models for investigation during the project. Those models were: EVIPAS, HCS, SIG/CINEMA, SIGNAL94, SIDRA, and TRANSYT-7F.

1.7 Naming Conventions

In order to improve the readability of this document, a streamlined method of referring to each of the models and sub-models has been used in this document. Use of the abbreviations shown in Figure 2 should make the results documented in this document clear and easily understood by the reader.

Figure 2. Model Naming Conventions used in the Document

HCS	The Highway Capacity Software
SigCinema	The SIG/CINEMA package
Sig94-HCM	The "Analyze" module of SIGNAL94 (HCM analysis)
Sig94-Alt	The "Evaluate" module of SIGNAL94 (Proprietary analysis)
Sidra	The SIDRA package
Transyt	The TRANSYT-7F package
Evipas	The EVIPAS package

2. Model Overview and Characteristics

2.1 Introduction

The purpose of this chapter is to give the reader an idea of the basic capabilities of each of the models included in the study. The results presented here are a summary of the results presented to the TAC after the initial evaluation by the study team. This chapter is a comprehensive summary of the working paper developed by the study team at the end of the first phase of the project (Model Overview).

2.2 Model Introduction

Each of the six models is described below based on exerts from their individual model user guides or on-line help features.

- **EVIPAS (Version 7.0)** EVIPAS is an optimization/simulation model for actuated controlled, isolated intersections. It is capable of analyzing and developing the optimal timing plan for a wide range of geometric configurations, detector layouts, and almost any phasing pattern available in a standard dual ring NEMA or Type 170 controller. The user can select from a variety of measures of effectiveness including delay, vehicle operating cost, vehicle depreciation cost, fuel consumption or pollutant emissions to determine the optimum settings for pretimed, semi-actuated, fully-actuated, or volume-density controls with or without pedestrian actuation. Optimized timing can be developed for any combination of maximum green, minimum green, vehicle extension, added time per actuation, maximum initial time, minimum gap, time before reduction, and time to reduce minimum gap for any or all phases of a dual ring controller. (EzVIPAS 1.0 User Guide, Viggen Corporation, 1993)

- **HCS (Release 2.4D)** The Highway Capacity Software (HCS) is developed and maintained by McTrans as part of its user-supported software maintenance as a faithful implementation of the Highway Capacity Manual procedures. Since its initial issue to McTrans, additional revisions have been made to the computational code and the user interface. The urban

Signals module of HCS represents the implementation of Chapter 9 of the 1994 Highway Capacity Manual.

- **SIG/Cinema (Version 1.11)** SIG/Cinema is an extension of HCM/Cinema which analyzes signalized intersections according to the 1994 update to chapter 9 of the Highway Capacity Manual. SIG/Cinema features: 1) Friendly user interface to ease data entry, 2) immediate pictorial feedback as data items are entered, 3) step-by-step on-line guidance by the software, 4) menu format eliminates need for expertise with computers, 5) hard copy of any screen display, 6) TRAF-NETSIM simulation of traffic operations, 7) animation of simulated traffic movements, 8) batch mode option to greatly enhance user productivity, and 9) hot-line telephone support. SIG/Cinema integrates a general purpose signal phasing and timing optimization algorithm with the latest HCM Chapter 9 capacity and Level of Service procedures. You can choose from five signal optimization objectives. (SIG/Cinema User Guide, KLD Associates, Inc, 1996)

- **SIGNAL94 (Version 1.20)** SIGNAL94 is one of the programs within the TEAPAC program package. SIGNAL94 is designed to aid in the analysis and optimized design of isolated intersection control based on factors such as approach capacity, lane usage, phasing, and pedestrian constraints. The methodology uses the capacity analysis procedures documented in the 1994 update to the HCM. The program can be used to analyze existing conditions or to design for future conditions. (SIGNAL94 Tutorial / Reference Manual, Strong Concepts, 1995)

- **SIDRA (Version 5.02A)** The SIDRA package has been developed by AARB Transport Research Limited as an aid for design and evaluation of signalized intersections (fixed-timed / pretimed and actuated), roundabouts, two-way stop control, all-way stop control, and yield control intersections. SIDRA uses detailed analytical traffic models coupled with an iterative approximation method to provide estimates of capacity and performance statistics (delay, queue length, stops, etc.). SIDRA traffic models can be calibrated for local conditions. The US HCM version of SIDRA is based on the calibration of model parameters against the 1994 HCM. (SIDRA 5 User Guide, AARB Transport Research, 1996)

- **TRANSYT-7F (Release 7.2)** One of the two main functions of the TRANSYT-7F model is to simulate the flow of traffic in a signalized network. Simulation is an analytical process that attempts to represent real world conditions, in this case traffic flowing through the network, being stopped at intersections by red signals and soon thereafter moving during green signals. The second main function of the TRANSYT-7F model is to develop optimized traffic signal timing plans. It is important to note that TRANSYT-7F deals in pretimed operations. Since only one signal cycle is simulated, this cycle and its timing and resulting traffic operations are assumed to repeat as "average" conditions throughout the a period of analysis. (TRANSYT-7F User Guide, USDOT & FHWA, 1991)

2.3 Model Inputs

2.3.1 Introduction

This section details the input modules of each of the six models under comparison. First, there is a brief description of the input interface and how information is managed. This is followed by a series of comparative tables which highlight the available inputs of the different models. Where necessary, text explanations are given to describe special considerations. The final section contains a comparison based on overall input "friendliness", a combination of graphical quality, error checking, and ease-of-use.

2.3.2 Input Interfaces

- **EVIPAS** The EVIPAS model does not come with a data input manager. The program uses formatted FORTRAN card-style input. For the purposes of the study, however, the EZVIPAS data input manager was used to simplify the process of entering data for EVIPAS. The EZVIPAS data input manager moves the user through several screens to enter data. Two different files must be coded to run EVIPAS.

- **HCS** HCS comes with a built-in data input manager. While the data input manager does guide the user through the input process, it is very busy and does not identify which fields are required or optional.

- **SIG/Cinema** SIG/Cinema also comes with its own data editor with built in steps to guide the user throughout the entire process. The module is highly graphical in nature, showing the user step-by-step what they have entered.

- **SIGNAL94** SIGNAL94 provides a data input manager as part of the main program. Not a graphical interface, the input manager is very spreadsheet oriented, with labeled rows and boxes for input. There are actually two different ways to input the same information, a quick reference manager and a more detailed module.

- **SIDRA** The SIDRA model has a subsection called "RIDES" which acts as a data input processor. Using a series of data input screens, the RIDES module guides the user through basic data input, using graphic pictures of the intersection to help the user check their input for accuracy. The program also uses a hierarchical input style, so that intersection-wide parameters can be left as default or changed at the approach, the lane or even the movement level.

- **TRANSYT-7F** TRANSYT-7F provides a basic data manager which has the least available user support. The manager is entirely text-based, using drop down text helps after the user has begun the input process. There is no system of checks or any assurance that the user has entered all necessary data. For the purposes of this study, a data input manager called EZ-7 Plus was used for creating data sets.

2.3.3 Input Parameters: Geometry

The first input consideration is that of the allowable geometry. Questions about number of approaches, number of lanes, available grades, short lanes, etc. can help to determine the value of a model by its applicability. Figure 3 compares the geometry inputs from the models.

Figure 3. Geometric Inputs

		EVIPAS	HCS	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Approaches							
	Number of Approaches	5	4	4	4	8	6
	Non Square Alignment	N	N	N	N	Y	Y
	Grades	Y	Y	Y	Y	Y	Y
	Turn Radius	Y	N	N	N	Y	N
	Approach Length	Y	N	N	N	Y	Y
	Approach Speed	Y	N	N	N	Y	Y
	Exit Length	N	N	N	N	Y	Y
	Exit Speed	N	N	N	N	Y	Y
	Number of Exit Lanes	N	N	N	N	Y	Y
	Median Width	N	N	Y	N	Y	N
Lanes							
	Shared Lane Analysis	Y	Y	Y	Y	Y	Y
	Exclusive Turn Lanes	Y	Y	Y	Y	Y	Y
	Turn Lane Lengths	Y	N	Y	Y	Y	Y
	Downstream Short Lanes	N	N	N	N	Y	Y
	Lane Width Adjustment	Y	Y	Y	Y	Y	Y
	Lane Utilization Factors	Y	Y	N	Y	Y	N
	Detailed Lane Configuration	Y	Y	Y	Y	Y	Y
	User-assignemd Movements in Lanes	Y	Y	Y	Y	Y	Y
	Pedestrian Walkways	N	N	N	N	Y	N
Detectors							
	Specific Location	Y	N	N	N	N	N
	Detector Type	Y	N	N	N	N	N
	General Actuation	Y	Y	Y	Y	Y	Y
Other							
	Bus Traffic	Y	Y	Y	Y	Y	N
	On-street Parking	Y	Y	Y	Y	Y	N
	Pedestrian Call Button	N	Y	Y	N	N	N
	Location (CBD, Non-CBD, Rural, etc.)	N	N	Y	Y	N	Y

2.3.4 Input Parameters: Volumes and Flow Characteristics

The next input consideration is that of volume and flow characteristics. Questions about flow rates, practical degree of saturation, and saturation flows can help determine the viability of a model's results. Figure 4 compares the volume and flow characteristic inputs from the 6 models.

Figure 4. Volume and Flow Inputs

		EVIPAS	HCS	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Base Volumes							
	By Specific O/D	N	N	Y	Y	Y	Y
	By Turning Movement	Y	Y	Y	Y	Y	Y
	By Turning Percentages	Y	N	N	N	N	Y
	Specific Pedestrian Volumes	N	Y	Y	Y	Y	N
	Pedestrian Crossing Time	N	N	Y	N	Y	N
Heavy Vehicles							
	Percentage	Y	Y	Y	Y	Y	Y
	By Movement	Y	Y	N	Y	Y	Y
	Multiple Types	Y	N	N	N	N	N
Multiple Volumes							
	By Growth Factor	N	N	Y	N	Y	N
	Multiple Loading Periods	Y	N	N	N	N	Y
	Midblock Flow Entry	N	N	N	N	N	Y
Flow Characteristics							
	Arrival Type	N	Y	Y	Y	Y	N
	Saturation Flow	Y	Y	Y	Y	Y	Y
	Practical Degree of Saturation	N	N	N	N	Y	Y
	Right Turns on Red	Y	Y	Y	Y	Y	N
	Pre-Green Sneakers	Y	N	N	N	N	N
	Post-Green Sneakers	N	N	Y	N	N	Y
	Permissive Left Turn Model	Y	Y	Y	Y	Y	Y
	Opposed Turn Model	N	N	N	N	Y	N
Flow Period							
	User-defined Time Period	Y	N	N	N	Y	Y
	Peak Hour Factor	N	Y	Y	Y	Y	N

2.3.5 Input Parameters: Signal Settings

The ability of a model to handle various signal settings while making an optimization or evaluation run is critical. Questions about the model's ability to test multiple phasings, to handle changes in detector settings, and to include start loss and end gain times in the analysis can make or break a model's accuracy and utility. The following table compares the signal setting options from the six models.

Figure 5. Signal Setting Inputs

		EVIPAS	HCS*	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Minimums and Maximums							
	Minimum Cycle Lenth	Y	-	Y	Y	Y	Y
	Maximum Cycle Length	Y	-	Y	Y	Y	Y
	Minimum Displayed Green (or phase)	Y	-	Y	Y	Y	Y
	Maximum Displayed Green (or phase)	Y	-	N	N	N	Y
	Minimum Gap	Y	-	N	N	Y	N
Phase Selection							
	Program Chooses Phasing	N	-	Y	Y	N	N
	Multiple Phasing Options in One Run	N	-	Y	Y	N	N
	Phase Elimination (removed from plan)	N	-	Y	Y	Y	N
	Skip Phasing (removed from one or more cycles)	Y	-	N	N	N	N
	Lead / Lag Phasing	Y	-	Y	Y	Y	Y
Controller							
	2-Ring Controller	Y	N	N	N	N	Y
	Pretimed	Y	Y	Y	Y	Y	Y
	Semi-Actuated	Y	N	Y	N	Y	N
	Fully-Actuated	Y	N	Y	N	Y	N
	Volume-Density	Y	N	N	N	N	N
	Implicit Actuated (reduces delay, no explicit modeling)	N	Y	N	Y	N	N
	User-selected Rest Phase	Y	N	N	N	N	N
	Green Split Priority	N	N	N	N	Y	N
Times							
	Startup Lost Time	Y	N	N	N	Y	Y
	Effective End Gain	Y	N	N	N	Y	Y
	Cycle Lost Time	Y	Y	Y	Y	Y	Y
	Yellow Phase	Y	Y	Y	Y	Y	Y
	All Red Phases	Y	Y	Y	Y	Y	Y
	Pedestrian Only Phase	Y	Y	N	N	Y	Y
	User-Specified Increment in Optimization	N	N	Y	Y	Y	Y
Simulation							
	User-given Specific Timings	Y	Y	Y	Y	Y	Y

* HCS has no optimization capabilities.

2.3.6 Model Ease of Use

Of the six models, two have very friendly input managers, two have an acceptable data managers, and two have a difficult input manager. The friendly data managers are associated with SIDRA, and SIG/Cinema. The acceptable data managers are associated with HCS and SIGNAL94. The most difficult input managers are associated with TRANSYT-7F and EVIPAS, although there are other commercially available software packages which generate both TRANSYT-7F and EVIPAS input data sets. As noted before, however, both EVIPAS and TRANSYT-7F were coded using separate data input managers. Both of these managers raise the overall ease of use from difficult to acceptable.

SIDRA and SIG/Cinema are very similar in terms of ease of data entry. They both make use of graphical views to help the user "see" what they are entering. There is a slight edge to the SIG/Cinema model because of mouse support which SIDRA does not have. SIG/Cinema also requires less data than SIDRA.

The EVIPAS input manager (called EZVIPAS) virtually forces the user to input every piece of necessary data. The processor moves from one screen to the next, placing the cursor in the proper location at each step in the process. While there are no graphical representations to help the user check their entries, a user who has access to all the necessary input data will have little trouble preparing the files needed to execute EVIPAS. One drawback of the EVIPAS input manager, however, is the need to create two different input files in order to make a single run.

The HCS data input manager has only two text-based screens, and virtually every field is a required one. The manager is very busy-looking and can confuse the unfamiliar user. Overall, though, it is adequate to the task.

The SIGNAL94 data input manager is a text-based input manager with s spreadsheet look. There is a quick-entry screen which allows the user to go to a single screen to enter all needed data or a more detailed sub-process which allows for more data input over several screens. Some of the data entry labels are not fully intuitive, but some text descriptions are available.

The TRANSYT-7F data manager (EZ-7 Plus) is difficult for a novice to navigate. In addition to using unusual keys for navigation, it also provides no mouse support. It does provide an error-checking option when generating input files; however, the input manager does not provide the user with all the options available when coding TRNASYT directly.

2.4 Model Optimization Procedures

2.4.1 Phase Optimization

The capabilities of each model are discussed in text below and then tabulated for comparison.

- **EVIPAS** EVIPAS requires the user to input a phase pattern which is available on a standard dual-ring controller. Since EVIPAS is primarily designed to handle actuated and volume-density controllers, the model does not have the ability to choose its own phasing pattern. It does, however, have the ability to skip a phase in the pattern if no traffic is there for that phase. This simulates an actual actuated controller which would not present a phase without a call for that phase. (Note: If EVIPAS is used to design a fixed-time controller, it will not skip phases.)

- **HCS** HCS has no optimization capabilities.

- **SIG/Cinema** SIG/Cinema will choose the best phasing as well as the best splits. The user is allowed to select "available" phase patterns, from which SIG/Cinema then chooses the best overall phase pattern. Available patterns are chosen for each street (lead, lag, dual left, etc.) and do not have to be the same for both streets (i.e. the user can require one street to use a particular phase pattern while letting SIG/Cinema choose the pattern for the other street).

- **SIGNAL94** SIGNAL94 will also choose the best phasing and best splits. The model will inform the user as to what phase patterns are possible based on the given input data. The user may then select any subset of those potential patterns for the software to evaluate. SIGNAL94 then determines the best overall intersection performance settings from the subset of potential phase patterns defined by the user.

- **SIDRA** The SIDRA model also requires the user to input a phase sequence, but SIDRA does have the ability to drop phases from a phase plan as needed. For example, an intersection might have both a protected left turn phase and permitted left turns during the green through phase. If the intersection will operate more effectively without the protected left turn phase (by requiring all left turns to be taken via the permissive model) then the software will remove that phase from the plan and advise the user of that change.

- **TRANSYT-7F** TRANSYT-7F will not determine the best phasing pattern. The user is required to enter a phasing pattern and the model will provide the best splits it can find for that pattern. It also cannot drop phases from the timing plan.

Figure 6. Summary of Phase Optimization Capabilities

	EVIPAS	HCS*	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Able to Select Phasing Itself	N	-	Y	Y	N	N
Able to Drop Phases from a Plan	N	-	Y	Y	Y	N
Able to Drop Phases from Individual Cycles	Y	-	N	N	N	N
Requires Starting Point Phase Times from User	N	-	N	N	N	Y

* HCS has no optimization capabilities

2.4.2 Optimization Parameters

The capabilities of each model are discussed in text below and then summarized in a table.

- **EVIPAS** EVIPAS allows the user to select from two basic categories of optimization: cost or emissions. If cost is chosen, the user may then select a vehicle depreciation cost, a delay cost, or a fuel consumption cost. Only one of the costs may be used as the optimization criterion. If the emission option is chosen, the user may then choose to optimize based on Carbon Monoxide, Nitrous Oxide, or Hydrocarbons as the evaluation emission. In addition, the user may select a predefined combination of the three emissions as the evaluation criteria.

- **HCS** HCS has no optimization capabilities.

- **SIG/Cinema** SIG/Cinema allows the user four choices for optimization. The user may balance critical volume to capacity ratios, equalize delay per vehicle, minimize delay per vehicle, or minimize delay per vehicle with level-of-service constraints.

- **SIGNAL94** SIGNAL94 has an interesting feature. While it will only optimize based on overall intersection delay, it allows the user to include "acceptable" or "target" levels of service for different approaches. This would allow a user to penalize a minor street in the optimization process so as to improve overall operating conditions on a major street.

- **SIDRA** SIDRA offers the user a wide range of operational parameters for optimization. The model can optimize for delay, queues (either average or back-of-queue), stop rate, a weighted performance index, balanced degree of saturation, balanced capacity, spare capacity, fuel, cost, or emissions. While the program will only optimize for one of these values at a time, it will give output statistics on all of them, so that the user can compare the resulting plans based on each criteria.

- **TRANSYT-7F** TRANSYT-7F gives the user several options in terms of optimization. The user may optimize based on stops, stops and delay, delay, stops and delay and queues, total cost, and total progression opportunities. TRANSYT-7F also offers what it calls a "quick optimization" which is less useful in the isolated signal environment than in the system environment which TRANSYT-7F can handle.

Figure 7. Optimization Options

		EVIPAS	HCS*	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Minimize							
	Delay	Y	-	Y	Y	Y	Y
	Stops or Stop Rate	N	-	N	N	Y	Y
	Average Queue	N	-	N	N	Y	Y
	Back of Queue	N	-	N	N	Y	N
	Fuel Consumption	Y	-	N	N	Y	N
	Vehicle Depreciation	Y	-	N	N	N	N
	Vehicle Emissions	Y	-	N	N	Y	N
	User Cost	N	-	N	N	Y	Y
	Weighted Performance Index (different in each model)	N	-	N	N	Y	Y
Maximize							
	Intersection Capacity	N	-	N	N	Y	N
	Spare Intersection Capacity	N	-	N	N	Y	N
Equalize							
	Critical V/C ratios (or degree of saturation)	N	-	Y	N	Y	N
	Delay Per Vehicle	N	-	Y	N	N	N
Other Options							
	Allow LOS Constraints	N	-	Y	N	N	N
	Allow Target LOS	N	-	N	Y	N	N

* HCS has no optimization capabilities

2.4.3 Overall Optimization Capabilities

It is very difficult to choose an overall "best" because of the different capabilities of each of the models. Of this set, however, it seems that SIG/Cinema has a slight edge over the rest because it can select the best phasing time and provides a fairly broad range of optimization methods. SIDRA clearly provides the greatest number of options in terms of means of optimization; however, it requires the user to select a phasing plan at the start.

The other models do not really display any great advantage or disadvantage. EVIPAS won't choose phasing, but it does offer options in terms of optimizing max greens and other actuated signal parameters. There is no clear way to compare these differences.

2.5 Model Outputs

2.5.1 Introduction

The output provided by the model is as critical as the optimization power of a model. Even if a software package generates excellent timing plans, if the information is not available to the user in a clear, concise format, it will be useless. Output should be both complete and understandable. This chapter looks at the information available from each of the programs compared. Considered first is the on-screen output which is available. Later, the printed information is discussed.

It should be noted that most programs of any type will generate an input echo in the output files. This is not considered to be a highly significant feature, so information about the nature of input echoing is not included. This chapter focuses on the output information determined by each model.

2.5.2 On-Screen Outputs

- **EVIPAS** EVIPAS provides virtually no on-screen data. The only information available is that the simulation has proceeded successfully. No information about timing or signal settings is available on the screen. A data file is created, but it must be viewed from another text-based application.

- **HCS** HCS provides on-screen output from each step in the HCM process. These screens are provided both during the running of the model and afterward. The user can make modifications to some of the values displayed during an analysis and see the effects of their changes.

- **SIG/Cinema** One of the interesting aspects of the SIG/Cinema analysis process is its interactive use of HCM Chapter 9. While the program is running, the user is shown screens which have detailed HCM analysis output. If the user wants to make a change to any of these input worksheets, they may. After the program has chosen the best phasing and cycle length, the user is shown a simple screen with signal phases, green, yellow, and red times. In addition, the user is shown a worksheet with delay and LOS calculations, giving the user quick access to degree of saturation, delay, and LOS for lane groups, movements, and the entire intersection.

- **SIGNAL94** Once SIGNAL94 has run, the user can navigate a series of screens to access specific pieces of output. The first is phase information, which gives green, yellow, and red times for all phases. The capacity analysis information includes a full HCM analysis of the intersection, including lane group, approach, and intersection g/C, v/c, delay, LOS, and back of queue information.

- **SIDRA** SIDRA provides a significant amount of on-screen output information. There are two types of output data available: graphical data and text data. The graphical data is available in a program module called "GOSID" which enables the user to bring up several graphical representations of the output results including the sensitivity of performance measures to cycle length. Examples include delay, LOS, queues, stops, degree of saturation, capacity, flows, and average speed.

  The text output includes a series of tables summarizing the results of the run. Information available includes traffic flow data, intersection parameters, movement capacities, phase information, signal settings, lane performance information, delay studies, signal timing data, intersection performance, queues, speeds, and stop data. There is also a full set of HCM-style analysis worksheets. The program offers a “clickable” table navigation feature which makes it very easy to find specific information.

- **TRANSYT-7F** TRANSYT-7F provides the user with the choice of on-screen or in-file output. The on-screen option simply causes all the output to scroll by the user so fast as to be unreadable. Therefore, TRANSYT-7F essentially has no available on-screen output. When run in screen-output mode, the only available pieces of information are the actual phase splits.

Figure 8. On-screen Outputs

		EVIPAS	HCS	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Delays and LOS							
	By Movement	N	Y	Y	Y	Y	N
	By Approach	N	Y	Y	Y	Y	N
	Entire Intersection	N	Y	Y	Y	Y	N
Flow Characteristics							
	Queues	N	N	N	N	Y	N
	Capacity	N	Y	Y	N	Y	N
	Spare Capacity	N	N	N	N	Y	N
	Degree of Saturation	N	Y	Y	Y	Y	N
	Stop Rates	N	N	N	N	Y	N
Signal Timings							
	Phase Lengths	N	Y	Y	Y	Y	N
	Green Times	N	Y	Y	Y	Y	N
	Yellow Times	N	Y	Y	Y	Y	N
	All Red Times	N	N	Y	Y	Y	N
	Cycle Length	N	Y	Y	Y	Y	N
	Cycle Length Comparisons	N	N	N	N	Y	N

2.5.3 File and Printed Outputs

- **EVIPAS** The EVIPAS output file contains a concise summary of all program determined information. Beyond the input data echo, EVIPAS reports the quality of each of its simulation runs and the bounds for each of the variables it tries to optimize. The final results of the program are summarized in two tables and one set of text notes. The first table gives average stopped delay, average total delay, and maximum delay for each lane in the system. The second table gives the phase time for each phase from the timing plan and the cycle length. The text notes give the average total and stopped delay for the intersection.

- **HCS** HCS is not designed to give specific "file output." In fact, the "input file" will contain the output results so that they can be recalled at a later date. The output fields are the same as noted above. The user can print any of the output worksheets. In the event that the user desires a separate output file, there is a "print to file" option.

- **SIG/Cinema** SIG/Cinema allows the user the option to print each of the on-screen output screens. This gives the user access to signal phases, green, yellow, and red times, and a worksheet with delay an LOS calculations, giving the user quick access to degree of saturation, delay, and LOS for lane groups, movements, and the entire intersection.

- **SIGNAL94** Like SIG/Cinema, SIGNAL94 gives the user the option to transfer the on-screen information to a printout style. One of the nicer features of the SIGNAL94 model is its ability to construct various file types. It can generate a basic text file like the other models, but it also has the ability to generate a spreadsheet-style output. This file is in the *.wk3 format, designed for Lotus 123, but also readable by Excel and similar softwares. The software can export phase information (green, yellow, and red times for all phases) and capacity analysis information (including a full HCM analysis of the intersection, with lane group, approach, and intersection g/C, v/c, delay, and LOS) plus back of queue information.

- **SIDRA** SIDRA's printed output is exactly the same as the textual output that is available on-screen. It includes a series of tables summarizing the results of the run. Information available includes traffic flow data, intersection parameters, movement capacities, phase information, signal settings, lane performance information, delay studies, signal timing data, intersection performance, queues, speeds, and stop data. There is also a full HCM analysis worksheet within the information.

- **TRANSYT-7F** TRANSYT-7F provides almost as detailed an output as SIDRA, though less of the information is useful. After the data input echo, TRANSYT-7F provides information on the relative improvement of each consecutive iteration. After determining the best cycle, TRANSYT-7F provides the user with two intersection-based information tables and with one system-wide information table. The system-wide information table provides little of interest to the user. The first individual intersection table gives an expanded HCM analysis. This table contains flow, saturation flow, degree of saturation, uniform delay, random delay, total delay, average delay, stops, back of queue, queue capacity, queue time, fuel consumption, effective green, and LOS. The second table provides all the signal setting information including splits, percent splits, and cycle length information. It also provides pin settings, though these are seldom used anymore.

Figure 9. File and Printed Output

		EVIPAS	HCS	SIG/Cinema	SIGNAL94	SIDRA	TRANSYT-7F
Delays and LOS							
	Stopped Delay	Y	Y	N	N	Y	Y
	Average Total Delay	Y	N	Y	Y	Y	Y
	Maximum Delay	Y	N	N	N	Y	N
	By Lane	Y	N	N	N	Y	Y
	By Movement	N	Y	Y	Y	Y	Y
	By Approach	N	Y	Y	Y	Y	Y
	Total Intersection	Y	Y	Y	Y	Y	Y
Flow Characteristics							
	Queues	N	N	N	N	Y	Y
	Capacity	N	Y	N	Y	Y	Y
	Spare Capacity	N	N	N	N	Y	Y
	Degree of Saturation	N	Y	Y	Y	Y	Y
	Stop Rates	N	N	N	N	Y	Y
	Flow	Y	Y	Y	Y	Y	Y
Signal Timings							
	Phase Lengths	Y	Y	Y	Y	Y	Y
	Green Times	Y	Y	Y	Y	Y	Y
	Yellow Times	Y	Y	Y	Y	Y	Y
	All Red Times	N	N	Y	Y	Y	Y
	Cycle Length	Y	Y	Y	Y	Y	Y
	Cycle Length Comparisons	N	N	N	N	Y	N
	Old Style Pin Settings	N	N	N	N	N	Y

2.5.4 Output Summary

Overall, SIDRA is the model with the most complete output. It's highly organized on-screen data managers allow the user to see a variety of charts, graphs, and tables which more than completely describe the performance of the signal under a variety of settings. The manager also allows easy access to information, without the need to weed through unwanted information.

The HCS, TRANSYT-7F, SIG/Cinema, and SIGNAL94 models each provide about the same information. All three of these models allow the user to see completed HCM type worksheets, giving the user a familiar feel to output data as well as complete data.

The EVIPAS output clearly leaves the user without critical information. Two of the more significant missing pieces of information are capacities and degrees of saturation.

2.6 Conclusions and Recommendations

2.6.1 Introduction

This paper has provided a comprehensive review of six traffic signal timing models for use at isolated intersections. The review encompassed the required model inputs, the optimization capabilities and the provided outputs (both printed, and on-screen). In addition, a numerical comparison of the models' results is given for a sample problem taken from the 1994 HCM.

Our review indicates that most models have some common capabilities, for example optimizing for cycle length and splits under a given phase plan. However, the models varied significantly in terms of a) explicitly handling actuated control, b) optimization capabilities, c) geometric configurations that can be entered d) production of on-screen inputs and outputs and e) general ease of use and results interpretation.

2.6.2 Overall Model Comparison

While a final decision on which models should be further pursued for field evaluation in North Carolina is not made at this time, we offer the following ratings based on our experience with the models' use thus far. The criteria listed represent the team's collective view of the major areas deserving consideration in the model comparison process. The rating scale ranges from 1 (poor) to 5 (excellent). The technical committee is encouraged to review these ratings and be prepared to discuss them prior to the development of the field data collection plan. A particular item for the committee's consideration is the designation of weights to be used for each criteria, which would assist the research team to focus on a reduced set of models.

Figure 10. Initial Model Rating

	EVIPAS	HCS	SIG/CINEMA	SIGNAL94	SIDRA	TRANSYT-7F
1. Consideration of Actuated Control	5	2	2	2	3	2
2. Phasing Plan Flexibility	4	5	4	5	3	3
3. Evaluation of Existing Timing	2	4	5	4	5	5
4. Considers Impact of Signal Coordination	1	3	3	3	3	5*
5. Multiple Criteria for Optimization	4	0	2	3	5	4
6. Diversity of Input Conditions	4	3	3	3	5	4
7. Ease of Data Entry	2	3	5	3	4	2
8. Diversity of Model Output	2	3	3	3	5	4
9. Ease of Output Interpretation	2	4	5	4	4	4
Total	**26**	**27**	**32**	**30**	**37**	**28**

* For a single intersection analysis, TRANSYT will always assume random arrivals. Signal coordination effects can be included by coding additional intersections upstream of each approach at the subject intersection

3. Data Collection Planning

3.1 Introduction

The purpose of this chapter is to acquaint the user with the data collection issues which had to be considered by the study team for successful completion of this project. First in the chapter will be a description of the general data collection planning process. This will be followed by results from or descriptions of selected steps in the process. This represents the planning aspects of the second phase of the project (Model Evaluation).

3.2 Data Collection Planning

The Manual of Traffic Engineering Studies (ITE) details seven steps that should be taken to plan for any data collection effort. Those steps are summarized below.

3.2.1 Step 1: Define the Experiment's Purpose, Importance, and Scope

This step lays the keystone for all following steps. Defining a purpose is the first step in laying an experimental foundation. A clearly defined purpose statement gives both focus to the experiment and allows for the evaluation and selection of options which arise during experimental design. A clearly stated importance for the experiment helps to ensure that the experiment is of value when completed. In short, if no importance can be assigned to an experiment, there is no reason for it to be done.

The scope of the experiment is important for establishing limits. If an experiment is too broad in scope, it becomes impossible to complete with any significant findings. If an experiment is too narrow in scope, its results are of limited value because of their lack of applicability to general cases. Usually, it is necessary to balance the desire to gain all knowledge with the constraints of time and budget.

3.2.2 Step 2: Determine the Experimental Unit

The experimental unit is the base level at which data is to be collected. The unit should be selected with the experiment's purpose in mind, so that the data taken help to give answers to purpose-related questions.

As an example, consider an interstate safety study. There are several possible experimental units. Researchers could select an individual accident as the study unit. They might also select a particular stretch of highway as an experimental unit. The types of information available to use in analysis would be very different depending on which unit was selected.

3.2.3 Step 3: Define the "Total" Population

Once the experimental unit has been selected, it is important to define the population of units - that is, what are all the possible types of units which may be encountered during the course of the experiment. During this step, units are described using factors and levels. A factor is a particular type of influence on the unit, and a level is one specific influence for a given factor. By definition, each factor must have two or more levels (or else there is nothing to vary across the experiment).

Referring back to our example of a highway safety study, assume for a moment that the chosen experimental unit was a stretch of highway. One factor might be the number of lanes. The levels could be 2, 3, and 4 or more. This is a factor with three levels. Another factor could be the average daily traffic on the freeway section. Levels could be defined as large, medium, and small traffic volumes. A third potential factor could be the width of cracks on an asphalt pavement.

During this process, the goal is to identify all the factors which may have a significant effect on the experimental unit. Unfortunately, it is not always possible to do so, as some influences are either so subtle or so seemingly unconnected that they escape notice. Additionally, once a population has clearly grown beyond the scope of the experiment, the focus shifts from finding all the possible factors to finding the most important ones.

3.2.4 Step 4: Select an Experimental Population

After the entire population has been defined, the experimenters must decide what portion of the population is to be tested in the experiment. As mentioned before, the scope of an experiment balances universal applicability with governing constraints. Segments of the population are included or excluded based on initial findings, frequency, importance, and professional judgment.

One way to reduce the population is by combining levels of a particular factor. For example, instead of considering large medium, and small traffic volumes, the researchers could eliminate the medium category, and classify everything as small or as large.

The other way to reduce the size of an experimental population is to assert that a particular factor has no significant impact and remove it from consideration. From our freeway safety example, researchers may conclude that the degree to which an asphalt pavement exhibits cracks is unlikely to have significant effect on accidents, and therefore drop it from consideration altogether.

3.2.5 Step 5: Select an Analytical Method

After the experimental population has been defined, what remains is to determine which method of analysis will best suit the purpose of the experiment. Analytical methods usually involve statistical tests, each of which are particularly suited to a certain type of experimental unit and which are suited to giving a particular type of information. Selecting the proper analytical method goes a long way toward building a successful experiment.

Again consider the highway safety example. If the purpose of the experiment is to determine which factors contribute to safety concerns so that new freeways can be built for higher levels of safety, an ANOVA analysis, which considers each factor as a treatment, would be appropriate. However, if the goal is to develop a means of predicting the number of accidents which can be expected on a section of freeway, a regression analysis would be much more appropriate.

3.2.6 Step 6: Search for Existing Data

Once both the experimental unit and the analytical method have been chosen, the data requirements are known. If an acceptable data set containing the needed information already exists, the experiment can proceed using this data. If the search turns up insufficient existing data, it becomes necessary to collect data for the experiment.

3.2.7 Step 7: Data Collection Planning

If it is necessary to collect data for a particular experiment, planning for that data collection becomes essential for the success of the experiment. The data collection plan should ensure success not by selecting data collection sites that force a particular experimental result, but which cover the full range of the experimental population equally.

3.3 Purpose, Relevance and Scope

During the first step of the data collection process, the study team, in conjunction with the TAC, devised the following purpose, relevance, and scope for the evaluation phase of this project.

The purpose of this experiment is to determine how well each of the computer models estimates delay at isolated intersections in North Carolina. This experiment should also attempt to determine what conditions contribute significantly to the error in field delay estimation.

This study is relevant because it will allow NCDOT to select an appropriate computer model as determined by the conditions present at a given intersection. Knowledge about what characteristics cause errors in delay and queue length estimation will allow users to avoid the use of a particular model under conditions which contribute significantly to errors in delay and queue length estimation. NCDOT can also inform subcontractors about the acceptability of results based on the information gained during this study. In the final analysis, the beneficiaries of this study are the driving population of the state of North Carolina.

The results of this study may also be used by the study's advisory committee in the process of selecting which models will be under consideration during the optimization experiments of this project, when the optimization capabilities of the models will be compared.

Some limitations to the scope have already been implied by the purpose. For one, this study is limited to conditions in North Carolina. NCDOT is the funding agency, and so all results should be directly applicable to North Carolina. Second, it is limited to isolated, signalized intersections. This limitation was indicated in the proposal, which clearly stated the consideration of only isolated, signalized intersections. Third, it is limited to delay comparisons.

3.4 Selecting An Experimental Unit

3.4.1 Introduction

This section discusses the potential experimental units considered by the study team. As noted earlier, an experimental unit is the base level item at which data is to be collected. Each unit is considered with respect to its availability from both field measurements and model output. The pros and cons of each are also discussed, and the study team's recommendation is made. It should be noted that the contents of the rest of this working paper are based on the study team's decisions here.

3.4.2 The Intersection Unit

The first potential experimental unit considered by the study team was the entire intersection. As all of the models produce a statistic of total intersection delay, and since it is possible to compute a total intersection delay value based on field data collection, the delay aspect of the experimental purpose could be met.

The biggest problem with choosing the intersection as the experimental unit is the sample size. Each different intersection layout would require replication within the study design. This would provide an extremely large population for consideration, requiring a sample size beyond the time and budget constraints for this project. Because of this, the study team is not recommending the intersection as the experimental unit.

3.4.3 The Approach Unit

The study team next considered an individual approach as a potential experimental unit. All but one of the computer models gives approach delay information, and the one which does not specifically report approach delay information does report information which can be used to calculate approach delay. The calculation of approach delay from field data is one step in the process of calculating total intersection delay, so the delay comparison required by the purpose statement is certainly possible at the approach level.

In addition, the population of approaches is necessarily smaller than the population of intersections (since intersections contain at least 3 approaches). Thus, the sample size for approaches would not be as daunting as for intersections. Admittedly, if the approach was selected as the experimental unit, the study team would have to "select out" some approach configurations which are less prevalent to meet time and budget constraints.

Unfortunately, the study team felt that the process of determining an average approach delay, which involves combining the delays of all lane groups present on that approach, might make it more difficult to determine which of the conditions at the approach cause the error. As a result, the study team did not select the approach as the experimental unit.

3.4.4 The Lane Group Unit

The next step down the hierarchy of the intersection is the lane group. As with the approach, all of the computer models either directly report lane group delays or give information which would allow the study team to calculate lane group delay. Field collection of delay information provides lane group delay, so the purpose statement requirement of delay comparison can be met with the lane group unit. In addition, use of the lane group as the experimental unit would keep the experiment in line with the Highway Capacity Manual, which also uses the lane group as the basis for intersection analysis.

The shift from approach to lane group, like the shift from intersection to approach, again decreases the population size (since approaches are a collection of one or more lane groups). In fact, there are basically seven different lane group types: left (L), through (T), right (R), left-through (LT), through-right (TR), left-right (LR), and left-through-right (LTR). The LR and LTR lane groups can be easily combined during analysis to reduce number of types to six. This keeps the population to a more manageable level, helping to ensure a broader application of results.

In addition to overcoming problems associated with sample size, use of the lane groups unit also reduces problems in identifying causes of error. Thus, the lane group appears to be a favorable candidate for the experimental unit.

3.4.5 The Lane Unit

Stepping down from the lane group brought the study team to the individual lane as a potential experimental unit. The lane has the same population size as the lane group (in fact, it has the same 7 possible types). This means that the population should be manageable.

The problem with using the lane as an experimental unit was that two of the six models (Sig-Cinema and Signal94) do not give output statistics at the lane level. The selection of the lane as the experimental unit would preclude the study of these two models. Thus the study team rejected the lane as an experimental unit.

3.4.6 Chosen Experimental Unit

Based on the considerations listed above, the study team chose the lane group as the experimental unit. This will allow the comparison of both delay and queue length values given by the computer models with field data.

3.5 Potential Site Characteristics

3.5.1 Introduction

The purpose of this section is to define the complete lane group population. It is from this "total" population that the study team selected the experimental population. The first step in defining the complete lane group population was to determine all the factors which might affect the observations of experimental units. The final step was to determine the levels that each factor might take on.

As mentioned before, there are seven possible lane groups - thus seven experimental units. They are the left turn lane group (L), the through lane group (T), the right turn lane group (R), the left-through lane group (LT), the left-right lane group (LR), the through-right lane group (TR), and the left-through-right lane group (LTR).

3.5.2 Factors

Figure 11 lists the population of factors, their levels, and which lane groups they apply to. It should be noted that some treatment levels are naturally occurring (such as ranges of v/c ratio and % right turns) and some were selected by the study team (such as type of control and left turn).

Figure 11. Factors and Their Levels

Factor (Relevant Lane Groups)	# of Levels	List of Levels (Ranges)
1. Controller Type (all)	3	Pretimed
		Actuated
		Volume-Density
2. Left Turn Treatment (L, LT, LR, LTR)	3	Protected
		Protected-Permissive
		Permissive
3. Right Turn Treatment (R, TR, LR, LTR)	5	Slip / Continuous
		Protected without RTOR
		Permissive (pedestrians) with RTOR
		Permissive without RTOR
		2-Phase Protected-Permissive
4. Intersection Location (all)	3	Urban
		Suburban
		Rural
5. Left Turn Lanes (L, LT, LR, LTR)	2	Single
		Multiple
6. Right Turn Lanes (R, LR, TR, LTR)	3	De-Facto
		Single
		Multiple
7. Approach Speed (all)	3	High (>45 mph)
		Moderate (35-45)
		Low (<35)
8. Crossing Distance (all)	3	Long (>75 ft)
		Medium (40-75)
		Short (<45)
9. Delineation (all)	3	Marked
		Unmarked
10. Queue Fills Pocket (L, R)	3	Frequently (>50 % of time)
		Infrequently (1-50%)
		Never (0%)
11. Lane Group Share of Cycle Length (all)	3	Dominant Share (>50% of C)
		Equal Share (10-50%)
		Minor Share (<10%)
12. Total Cycle Length (all)	3	Long (>105 sec)
		Moderate (70-105)
		Short (<70)
13. Volume to Capacity Ratio (all)	3	Above Capacity (>1.05)
		At Capacity (.95-1.05)
		Below Capacity (<.95)
14. Initial Queue (all)	3	Present and Large (large is >5 veh/lane)
		Present and Small (small is < 5 veh/lane)
		Not Present
15. Pedestrian Volume (R, LR, TR, LTR)	3	High (>100 peds/hr)
		Moderate (25-100)
		Low (<25)
16. % Turning Vehicles (L, R, LT, LR, TR, LTR)	3	Most (>40%)
		Equal Share (20-40%)
		Minor Share (<20%)
17. Opposing V/C Ratio (L, LT, LR, LTR)	3	High (>1.05)
		Moderate (.95-1.05)
		Low (<.95)
18. Opposing Lanes (L, LT, LR, LTR)	2	Single
		Multiple

Key to this description of the population is that some factors did not apply to all of the lane groups. Based on the numbering of the different factors as presented in Figure 11, Figure 12 presents a summary of the factors affecting each lane group.

Figure 12. Summary of Which Lane Group the Factors Affect

Factors Affecting (X = May Affect)							
	Lane Group						
Factor	**L**	**LT**	**LR**	**LTR**	**T**	**TR**	**R**
1	X	X	X	X	X	X	X
2	X	X	X	X			
3			X	X		X	X
4	X	X	X	X	X	X	X
5	X	X	X	X			
6			X	X		X	X
7	X	X	X	X	X	X	X
8	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X
10	X						X
11	X	X	X	X	X	X	X
12	X	X	X	X	X	X	X
13	X	X	X	X	X	X	X
14	X	X	X	X	X	X	X
15			X	X		X	X
16	X	X	X	X		X	X
17	X	X	X	X			
18	X	X	X	X			

3.5.3 Conclusions

This list of factors and their levels is not meant to be comprehensive, but rather to begin to define the total population. Each of the factors listed has the potential to contribute significantly to the error in any of the computer program's computation of delay or queue length. However, it should be emphasized that the study team had no intention of attempting to quantify the impacts of all of the factors listed here. To attempt to do so would require data collection from 2,066,242,608 lane groups, a number which exceeds the population of lane groups in the state. It was therefore necessary to select the most critical elements for evaluation. The results of this process are discussed in the next section.

3.6 Identification of Key Factors

Because of the significant number of factors identified, the study team felt it critical to focus on a few key factors. With the help of the TAC, the study team developed a list of key factors for inclusion in the study.

3.6.1 Lane Group Types

The TAC was interested in virtually all lane group types. They were interested in the L, LTR, T, TR and R lane groups because of their frequent occurrence. In addition, they had a particular interest in what the study team came to call the 2LTR lane group - a LTR lane group that is spread over 2 lanes. There was some consideration of removing this configuration from the list of "allowables" and so data on it was considered significant.

3.6.2 Left Turn Treatment

The TAC was particularly interested in each of the model's ability to handle different left turn treatments. In many publications to date, left turn treatment is listed as a weak point in traffic models.

3.6.3 Controller Type

The TAC was predominantly interested in actuated and volume-density controllers. Most of the signal timing models currently available (and 5 of the 6 included in the study) were primarily designed for fixed-time signals. These models generally apply a multiplicative factor to pretimed results and report the result as actuated or volume-density results.

3.6.4 Volume to Capacity Ratio

The TAC was interested in model performance over a range of volume to capacity ratios. Of special interest was near-capacity performance.

3.7 Summary

Based on the data collection planning steps described herein, the study team developed a data collection plan to meet the study's needs. The next chapter will describe the sites selected for study and describe both the data collection and data reduction processes.

4. Data Collection and Reduction

4.1 Introduction

This chapter summarizes the data collection and reduction steps taken during this research project. Sample calculations are included in the text.

4.2 Site Selection

Instead of pre-selecting all sites before data collection began, the study team and advisory committee decided upon a continuing process of site selection. After data had been collected and reduced from a few sites, the study team would present a report to the advisory committee, letting them know what data had been collected and what data were still needed. The advisory committee would then suggest additional sites to help fill in the gaps in the dataset.

As detailed in the previous chapter, the study team was interested in varying a few key factors at the collection sites. Control type, lane group type (including left turn treatment), and v/c ratio were the factors of primary interest. Figure 13 summarizes the eight sites chosen by the advisory committee and highlights their respective features.

Figure 13. Summary of Characteristics of Data Collection Sites

Site	Intersection	Approach	Lane Groups	Controller Type	V/C Ratio	Left Turn Treatment
1	NC55 and Sedwick	SB	L, TR	fully actuated	under capacity	protected only
	15-Apr-97	WB	LTR			permissive only
	7:00 to 9:00 am	NB	L, TR			protected only
	4:00 to 6:00 pm	EB	LTR			permissive only
2	Duke St. (one-way) and Chapel Hill	WB	TR	pretimed	under capacity	no left turns
	1-May-97	NB	LT, R			no opposing traffic
	7:00 to 9:00 am, 4:00 to 6:00 pm	EB	L, T			protected-permissive
3	Walnut and Tryon	SB	TR, L	fully actuated	at / over capacity	protected-permissive
	5-Jun-97	WB	L, T, R			protected-permissive
	6:30 to 8:30 am	NB	L, T, R			permissive only
	3:30 to 5:30 pm	EB	L, T, R			permissive only
4	US 70 and Ebenezer Church Rd.	SB	L, T, R	volume - density	over capacity	permissive only
	11-Jun-97	WB	L, T, R			protected only
	7:00 to 9:00 am	NB	L, T, R			permissive only
	4:30 to 6:30 pm	EB	L, T, R			protected only
5	Duke University Dr. and Anderson	SB	L, TR	fully actuated	under capacity	permissive only
	18-Jun-97	WB	LT, TR			permissive only
	7:00 to 9:00 am	NB	L, TR			permissive only
	4:00 to 6:00 pm	EB	LT, TR			permissive only
6	US 98 and Lynn Road	SB	LTR	fully actuated	under capacity	permissive only
	2-Jul-97	WB	LT, TR			permissive only
	7:00 to 9:00 am	NB	LTR			permissive only
	4:00 to 6:00 pm	EB	LT, TR			permissive only
7	US 70 and Ebenezer Church Rd.	SB	L, T, R	volume - density	at capacity	permissive only
	9-Jul-97	WB	L, T, R			protected only
	*9:00 to 11:00 am	NB	L, T, R			permissive only
		EB	L, T, R			protected only
8	NC55 and Sedwick	SB	L, TR	fully actuated	under capacity	permissive-protected
	After Left turn treatment change	WB	LTR			permissive only
	17-Sep-97	NB	L, TR			permissive-protected
	4:30 to 6:30 PM	EB	LTR			permissive only

Review of this table indicates that the term “site” does not necessarily indicate a different location, rather a difference in conditions. There were, in fact, only six different locations from which data were collected. Site 7 and Site 4 were at the same location, but at different time periods to generate different v/c ratios. Site 8 and Site 1 were at the same location, but had a different left turn treatment in effect.

For more detailed information on the specific locations of each of the sites, the reader may refer to the map located in Appendix A.

4.3 Data Collection

The data collection process was geared specifically to obtain the required elements for both model evaluation and model optimization. This included queue length, geometrics, traffic flow, and signal timing conditions. The specific data collected on-site are described below.

4.3.1 Queue Data

Queue data were collected from each lane group on each approach every 15 seconds. When queue lengths were short enough to be managed easily, the specific number of vehicles between the stop bar and the last queued vehicle were specifically recorded. When queue lengths became less manageable, distance markers and a site-specific jam density were used to estimate the number of vehicles in the queue. At a few sites, it was necessary to define a “maximum back of queue.” This was due to the excessive queue lengths observed (usually) on through movements at sites where the long queues went beyond an observer’s sight distance.

4.3.2 Flow Data

Individual turning movement volumes were recorded in 15 minute intervals for all approaches at the intersection. At Sites 2 through 8, right-turn-on-red data were specifically recorded as well (and an estimate made for Site 1). None of the sites had significant pedestrian activity, so no pedestrian data was collected.

No attempt was made to collect saturation flow data for the sites. The study team and advisory committee determined that the normal use of the models would not involve a saturation flow study; rather, the model estimate of saturation flow would be used.

4.3.3 Geometric Data

The majority of the geometric data used in the study came from intersection plans rather than from actual field measurements. The plans provided speed limits, detector locations, slopes, and other required or optional data. The one key datum that was collected in the field was lane width. This was due to the difficulty in determining the exact lane width from some plans.

4.3.4 Timing Data

In the field, the study team collected all the phase times displayed during the data collection period. This was done by recording the time at which a particular phase turned green and the time it turned red. In many cases, manpower shortages were overcome by videotaping the signal heads for selected approaches, allowing the determination of phase times to be completed at a later date.

4.4 Data Reduction

After collecting data from a particular site, the next step was to reduce it to the required format for data entry or for comparison purposes. For data entry into the models, the information needed included geometric data, demand data, and timing data. For comparison purposes, delay data was necessary. The following sections describe the reduction process used for each data type.

4.4.1 Geometric Data

The geometric data required no reduction whatsoever. This information was simply entered into the models. In some cases, a spreadsheet was used to organize the data for ease of data entry, but no changes were made to the data.

4.4.2 Demand Data

Demand was derived from two types of data collected in the field, flows and queues. The demand for a given time period can be defined as:

$$\text{Demand} = \text{Discharge Flow} + (\text{Final Queue} - \text{Initial Queue})$$

In most cases, all of this information was available and the calculation was therefore direct. However, when sight distance problems required the use of a maximum back of queue, the exact queue values were not known. When this occurred, the study team used the “Apparent Demand” instead of the true demand as the input to the model. The Apparent Demand can be calculated by:

$$\text{Apparent Demand} = \text{Discharge Flow} + (\text{Final Visible Queue} - \text{Initial Visible Queue})$$

This allowed for a demand estimation even when the exact queue length was not know.

4.4.3 Timing Data

The determination of average phase times was necessary for simulation of field conditions by the models. While very straightforward, this process did involve more steps than the previous data types mentioned, and so a sample calculation will accompany the text descriptions.

Recall that the 15 minute time period had been selected for use during the study. The first step, then, was to determine for each phase how much green was displayed during a particular 15 minute period. Figure 14 shows a sample of this calculation. In this and every other example in this chapter, the data shown are from the AM peak data collection at Site 1. A full copy of the Excel spreadsheet calculations for this time period can be found in Appendix B.

Figure 14. Sample Calculation of Green Time for a Phase

	Change to Green					Change to Red					Yellow			Green Time
	HH	MM	SS			HH	MM	SS			sec			sec
7:00 - 7:15														
	07	00	00	AM		07	00	19	AM		4.5			15
	07	00	32	AM		07	02	48	AM		4.5			132
	07	03	08	AM		07	04	10	AM		4.5			58
	07	04	34	AM		07	05	11	AM		4.5			33
	07	05	37	AM		07	05	54	AM		4.5			13
	07	06	15	AM		07	06	29	AM		4.5			10
	07	06	46	AM		07	07	03	AM		4.5			13
	07	07	19	AM		07	07	36	AM		4.5			13
	07	08	00	AM		07	08	31	AM		4.5			27
	07	08	50	AM		07	09	50	AM		4.5			56
	07	10	18	AM		07	11	57	AM		4.5			95
	07	12	19	AM		07	12	33	AM		4.5			10
	07	13	00	AM		07	13	23	AM		4.5			19
	07	13	36	AM		07	14	07	AM		4.5			27
	07	14	27	AM		07	15	00	AM		0.0			33
Total Green Time in 15 Minute Period														**547**

To determine the total green time in the 15 minute period, the green time for each time the phase was displayed is first calculated. Subsequently, the totals for each phase are added together. The formula for green time during a phase is:

Phase Green = Change to Red Time – Change to Green Time – Yellow Time

The yellow time is constant (4.5 sec) except for the final entry, where it is 0.0. This is because the phase actually remained green well into the next 15 minute time period. However, the green displayed in the next time period should not be included in this time period, so the green is "ended" at the end of the 15 minute period.

After totaling the green time in a 15 minute period, the average green per cycle is determined by dividing the total green in the period by the number of cycles in the 15 minute period. As the observed number of cycles may be different for different phases, the main street through was used to determine the number of phases in a 15 minute period.

After the average green time for each phase was determined, they were combined to determine a phasing plan. Figure 15 shows 2 examples of this process.

Figure 15. Determination of Phasing Plan from Average Movement Green Times

Time		Movement Green Time				
From	**To**	**E&WB A**	**NB L**	**NB TR**	**SB L**	**SB TR**
7:00	7:15	10.9	1.5	39.1	1.5	39.1
7:15	7:30	11.6	0.4	25.6	0.8	24.8

Time		Phasing Plan				
From	**To**	E&WB A	N&SB L	NB LTR	SB LTR	N&SB TR
7:00	7:15	10.9	1.5	0.0	0.0	39.1
7:15	7:30	11.6	0.4	0.0	0.4	25.0

Some of the conversions are very simple and straightforward. For example, consider the E&WB approach phasing, which is fully permitted (green ball to both approaches). It's average movement green time is moved directly to the phasing plan green time.

The N&SB timing is a bit more involved. Consider the 7:15-7:30 time frame. The NB L has 0.4 seconds of average green per cycle while the SB L has 0.8 seconds. This is converted into 0.4 seconds of N&SB L green plus 0.4 seconds of SB LTR green in the phasing plan. Unfortunately, the NB TR average green time is slightly longer than the SB TR average green time. Because of this, the value for N&SB TR green time in the phasing plan is an average of the different values. (Note that the SB TR average green time must be reduced by the 0.4 seconds which has been displayed in the SB LTR phase before the average is computed to determine the N&SB TR phase green time)

Minor differences in average green times, such as the one discussed above, occurred occasionally and were handled by the averaging process described above. These errors are likely due to slight differences in the timing device used or in the record vs play speed of the video cameras and video players used in data collection and reduction. No significant differences were encountered in the entire data reduction process.

4.4.4 Delay Data

The determination of field delay involved the most complex data reduction process of all. Delay was necessary, not as an input to the models, but for comparison with model estimates for model evaluation purposes. The process is heavily based on the methodology outlined in the draft version of the 1997 HCM for collection of control delay at signalized intersections. A copy of this method is reproduced in Appendix C. While this method is designed to collect average control delay instead of average stopped delay, the study team chose it because 1) the average control delay could be easily converted to average stopped delay, and 2) the data would be of more use for later research if it were in the new (HCM 1997) format.

4.4.4.1 Step 1

The first step is to calculate what is termed the “time in queue delay” for a 15 minute period. This simply involves multiplying the number of vehicles queued by the time they were queued and summing over the 15 minute period. Recall that queue measurements were taken every 15 seconds, so the number of vehicles queued at each 15 second interval was multiplied by 15 seconds. A sample of this calculation for one 15 minute period and one lane group is shown below in Figure 16.

Figure 16. Sample Calculation of Queue Delay for a 15 minute Period

Queues (TR Lane Group)							
	Second				**Queues**	**Total Delay**	
Hour: Minute	:00	:15	:30	:45	(sum)	(sec)	(min)
7:00	0	0	0	0	0	0	0
7:01	0	0	0	0	0	0	0
7:02	0	0	0	0	0	0	0
7:03	0	3	0	0	3	45	0.75
7:04	0	0	0	0	0	0	0
7:05	0	0	0	0	0	0	0
7:06	0	0	1	0	1	15	0.25
7:07	0	0	0	0	0	0	0
7:08	1	0	0	0	1	15	0.25
7:09	0	0	1	0	1	15	0.25
7:10	0	1	3	0	4	60	1
7:11	0	0	0	0	0	0	0
7:12	0	0	1	0	1	15	0.25
7:13	2	2	0	0	4	60	1
7:14	2	0	1	0	3	45	0.75
15 min Total	**5**	**6**	**7**	**0**	**18**	**270**	**4.5**

This calculation was simple and straightforward.

4.4.4.2 Step 2

The next step was to convert from total delay for a lane group to the average queued delay per vehicle for that lane group. This was done by dividing the total for the lane group by the demand for that 15 minute period in that lane group. No sample of this calculation is shown.

4.4.4.3 Step 3

The next step was to apply a correction factor to change from “average time in queue delay” to “average control delay”. This was a very involved process which involves tabulated values which are included in Appendix C. While a full sample calculation is not shown, some explanation of the particular implementation used by the study team follows.

The applied correction factor was based on certain flow characteristics, some of which were estimated rather than directly collected. One of the parameters is average speed in the lane group, for which the study team used the posted speed limit.

The other parameter needed for selecting a correction factor is the number of vehicles in the lane group which stopped at least once during the 15 minute time period. As this value was not directly collected in the field, an estimate was necessary. A stop rate was estimated using the formula:

$$\% \text{ Stopping} = (1 - g/C) / (1 - D/S)$$

Where:
g/C is the green split to cycle length ratio,
D is the demand, and
S is the saturation flow rate.

Values for green split and cycle length were taken from the average phasing plan. The demand was known (or the apparent demand was used), and the saturation flow rate was taken from the HCS estimate of saturation flow for that lane group.

While these estimates were clearly not as accurate as field measurements (especially for saturation flow rate), the study team did not feel that they would significantly jeopardize the conclusions of this study.

4.4.4.4 Step 4

The final step in delay determination was to convert from “average control delay” to “average stopped delay”. This conversion was made by dividing the average control delay value by 1.3, as indicated in the 1994 HCM.

4.5 Summary

The data reduction process, while long and involved, was designed to produce the best possible data for model input or for comparison of results. The reduction process, once designed, was highly automated, making use of spreadsheet capabilities to perform the calculations quickly and accurately.

There was, in a sense, one more “data reduction” task during the process – the reduction of Transyt output data. Unlike the other models in the study, the Transyt model does not directly estimate stopped delay. To allow for consistent comparisons with other models, the Transyt estimate of average total delay was divided by 1.3 to convert to average stopped delay.

5. Model Evaluation

5.1 Introduction

This chapter details both the methodology for model evaluation and the results of that evaluation. Several of the appendices will be referenced to show detail or samples of what is summarized in this chapter. The contents of this chapter are the results from the second phase of the project.

As noted before, all the data collected was reduced over 15 minute time periods. All the comparisons of data points herein are for 15 minute data.

5.2 Evaluation Methodology

5.2.1 Introduction

This, the "result gathering" step of the second phase of the project, was of critical importance to the study team. This step would help to determine which, if any, of the models were able to accurately depict field conditions. If a model is unable to accurately predict field conditions, then even if it has the most robust optimization module, its results are of no value to the practitioners.

5.2.2 Modeling

The first step in the evaluation process was to model the field conditions in each of the models in the study. All of the 15 minute time periods were used for this step. Both geometric and traffic conditions were entered into the models. It should be noted that when a particular piece of data was not specifically collected, each model's individual defaults were used.

For example, the study team did not collect data on lost time. Therefore, each model's default value of lost time was used, even though they were not all the same. In both planning and operational analysis, the user is often missing some types of data and is forced to rely on the defaults in the model being used.

After the modeling was completed, the delays and v/c ratios (where available) were recorded for each lane group. In the case of Evipas, lane-by-lane outputs were combined into lane group results for use in the comparison phase.

Of note is the treatment of HCS, SIGCINEMA, and the HCM module in SIGNAL94. These three models should produce exactly the same results, as all are computerized versions of the HCM. The study team verified that the models were, in fact, giving the same results from the same conditions by modeling two complete sites plus two points from each of the remaining sites in all three models.

This comparison showed that while the different models may give different results, those differences are on the order of +/- 0.1 sec/veh of stopped delay. The study team concluded that rounding was likely the cause of these differences, and so all the comparisons in this section for these three models are actually based on the results from the HCS software package.

Also of note is the fact that there were occasions when a model simply failed to provide an analysis. While these were rare, they did occur. The model which had the highest occurrence of this problem was the Evipas model, which at times would simply not generate appropriate traffic volumes. As Evipas is a microsimulation model, it randomly generates traffic to apply to the intersection. On some occasions, especially with Site 4, the model would generate what the study team considered to be insufficient traffic for the modeling to be accurate, and so those time periods were not considered during evaluation.

There were no “close calls” in this process of eliminating data points. All of the points used in the study had simulated approach volumes between 90% and 110% of the field volumes. The eliminated points had simulated approach volumes below 20% of the field volumes. The model never failed by a significant over application of volumes; failures were always underapplications.

5.3 Comparisons

After the modeling was completed, each model's results were compared with the field data. As noted before, the only datum common to all models is delay. Because of its use as the measure of LOS in the HCM, the type of delay used in comparisons is average stopped delay per vehicle. The necessary conversions were discussed previously.

During the process of model evaluation, the study team realized that their sample of pretimed intersections should probably be dropped from the study. For one, there was only one pretimed site, so there was not a sufficient sample for determining effects due to pretimed controllers. Second, the pretimed site where data had been taken had at least one approach which was coordinated in a system, and while some of the models provide an arrival type adjustment, not all of them can handle this situation. As noted in the scope, this study is focused on isolated intersections. For these reasons, these points have been removed from the numerical analyses which follow.

5.4 Intersection Level Comparisons

The study team first decided to look at average intersection stopped delay comparisons. Since this value is used to determine intersection LOS, its accuracy can figure heavily in operational decision-making.

5.4.1 Visual Comparisons

The first step the study team took in evaluating the models was to visually inspect graphs of Field vs. Model delay. This would help the team to determine if there were any visible trends to the data. The graphs that were developed by the team are shown in Figure 17 to Figure 21. Each graph contains data for one model from all 8 sites.

Figure 17. Graph of Intersection Level Field Delay vs EVIPAS Estimate

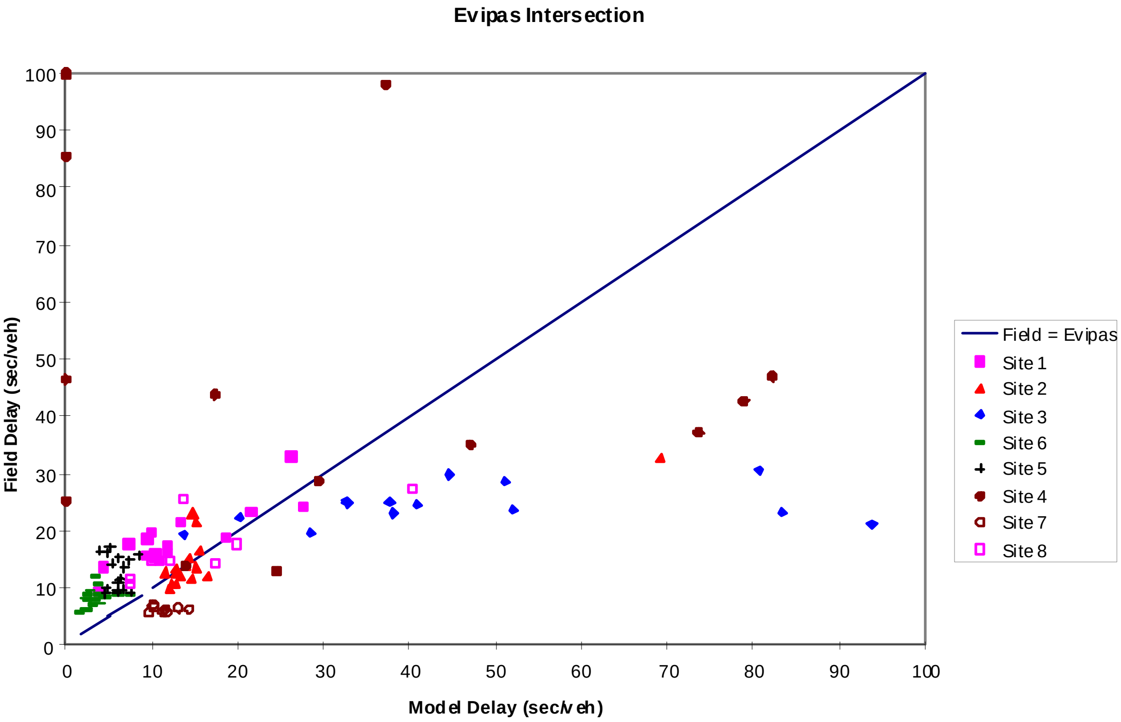


Figure 18. Graph of Intersection Level Field Delay vs HCM Estimate

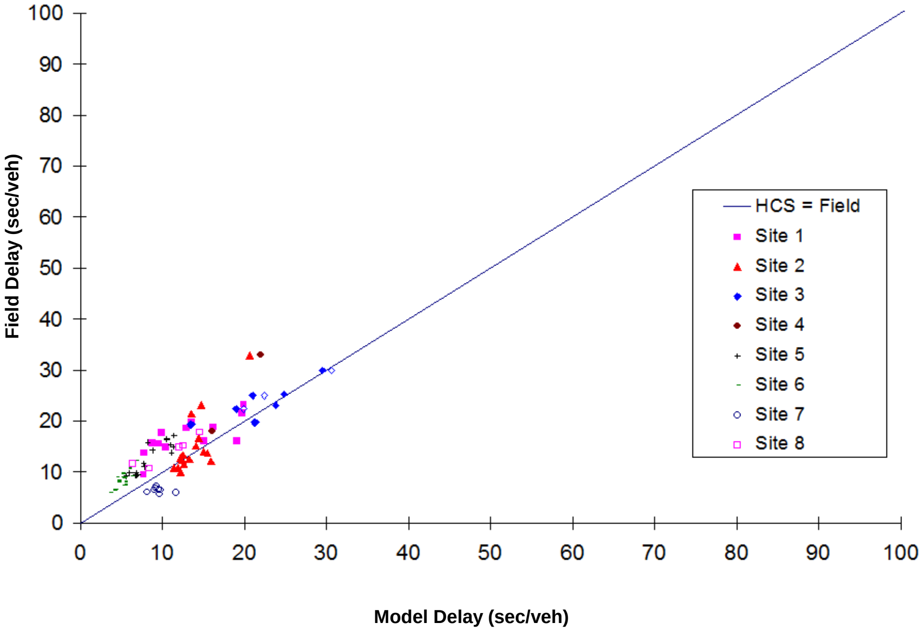


Figure 19. Graph of Intersection Level Field Delay vs SIDRA Estimate

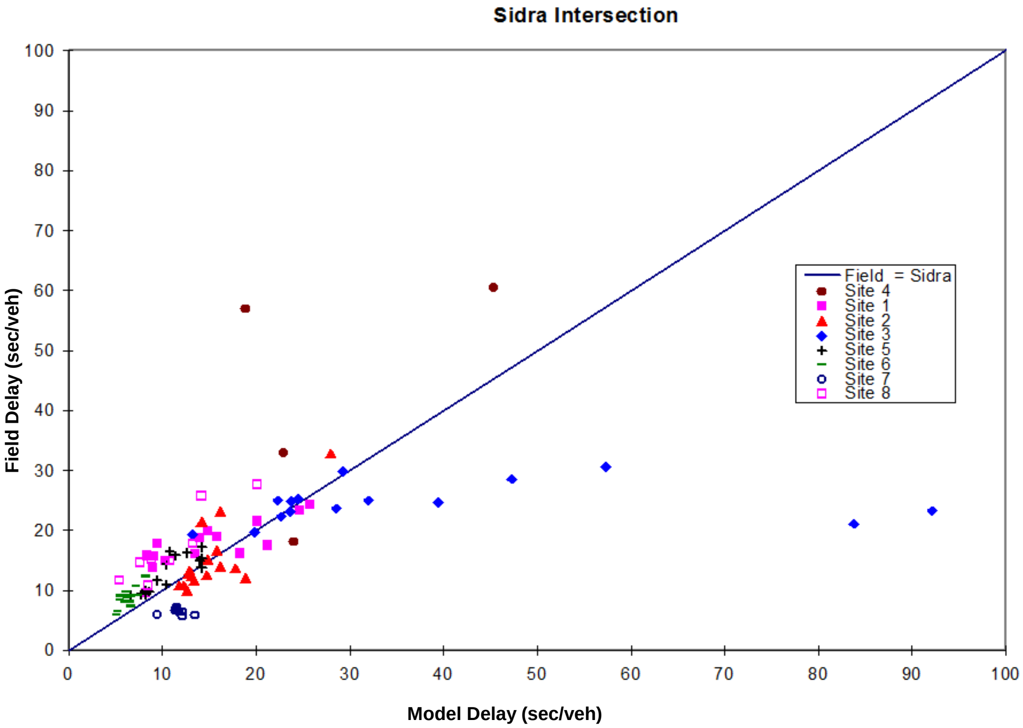


Figure 20. Graph of Intersection Level Field Delay vs SIGNAL94-Alt Estimate

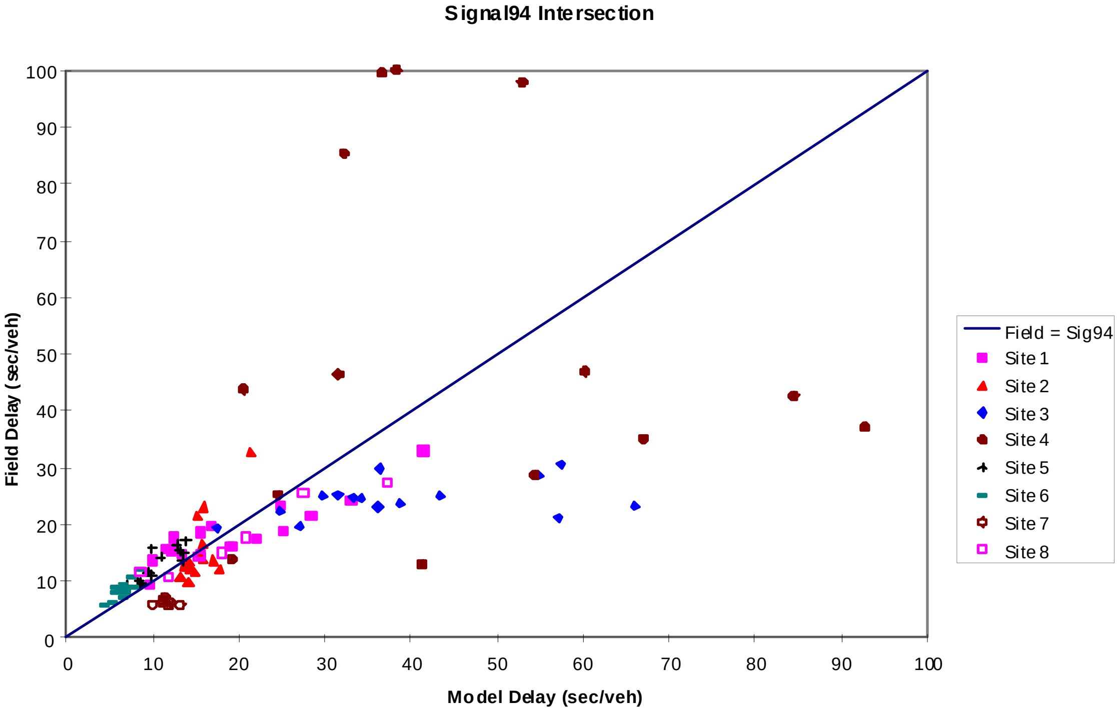
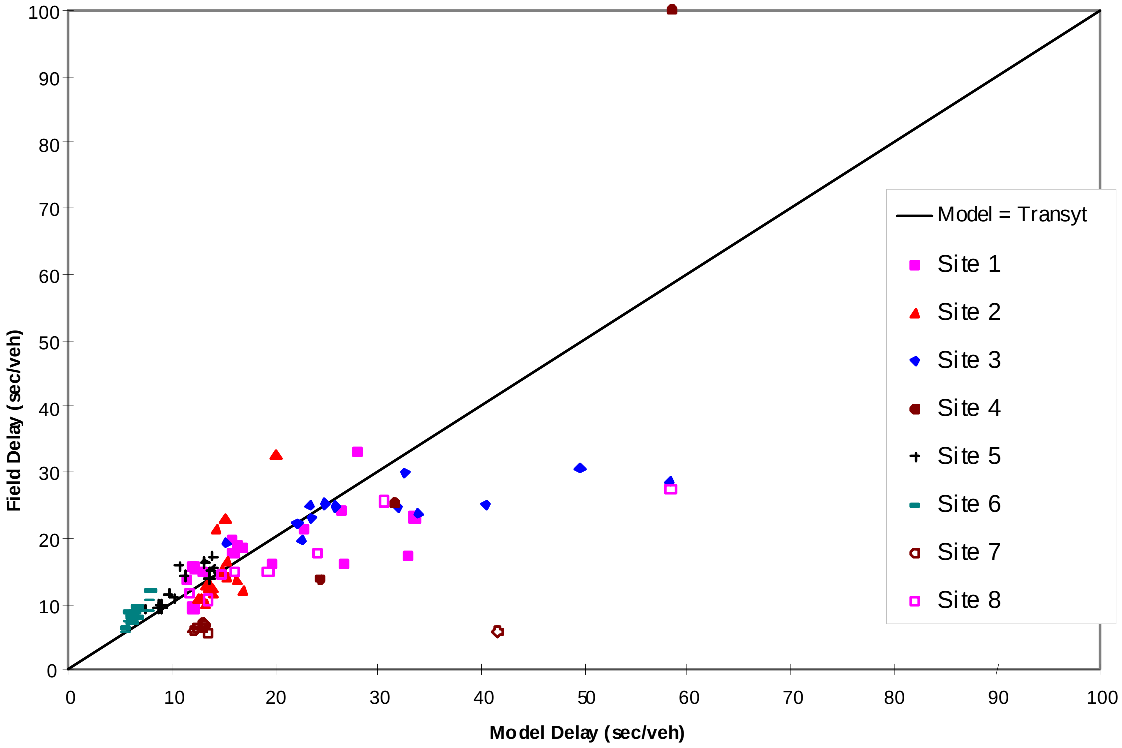


Figure 21. Graph of Intersection Level Field Delay vs. TRANSYT Estimate



When inspecting the graphs, one must keep in mind the following information. Each data point represents a 15 minute average intersection stopped delay. There are up to 110 such points on each graph. Some graphs have fewer because of excessive delays (off the scale) or because of modeling failure (for some reason, the model did not generate a usable delay estimate). Points are sorted by site. The "ideal" line is included on each graph to help in visualization of trends. A point above the line means that the field delay is higher than the model estimation, so the model underpredicted delay. Likewise, a point below the line means the model overpredicted delay.

It should be noted that the scale on the graphs has been chosen intentionally. The selection of 100 seconds as the maximum for both the horizontal and vertical axes does not imply that there were no points with coordinates beyond 100 seconds. A common scale was necessary to enable comparison between models without scale-based bias. The study team chose the 100 second limit because it contained the majority of points (>95%) in a reasonable range.

As expected, visual inspection of the data does uncover trends. In the HCS-SigCin-Sig94-HCS plot, there are no points where delay exceeds 35 seconds, most likely because of these models' inability to handle oversaturated conditions. The delay data on the graph, however, seems to track the ideal line nicely, suggesting that these models simulate field conditions accurately.

The SIDRA plot shows notably more points, including points beyond delays of 35 seconds. The points below that 35 second threshold seem to track the ideal line almost as well as the previous set. Beyond 35 seconds, however, there is much less visible accuracy.

In the Sig94-Alt plot, the points seem to track the ideal line only up to about the 25 second area. Beyond this there is a clear trend of points under the line (which means the model overpredicts delay). There is also a smattering of points up above the line which seem to follow no pattern.

The Transyt graph seem to have a tighter tracking of points to the ideal line than the Sig94-Alt, but also experiences a definite trend of overprediction after about 25 seconds. The Evipas plot shows not only an overprediction trend beyond 25 seconds but quite possibly an underprediction trend before 10 seconds.

5.4.2 Regression Comparisons

The study team next made simple regression comparisons between field and model intersection delays. Two different types of regressions were performed: one in which all points were included and one where points with v/c ratios > 1.0 were excluded. In the second group, a point was excluded if any lane group at the intersection had a v/c >1.0.

These two different regressions were developed because of the results seen in visual inspections. Of course, the HCM-based models are incapable of developing intersection delays if any of the lane groups in their analysis are over capacity. In addition, many of the models which can estimate delay for over capacity conditions showed a "threshold" where trends changed during visual inspection. Thus, the study team decided to perform both types of regression comparisons.

Figure 22 and Figure 23 summarize the results of the regression analyses. Please note that there is only one set of results for both the HCS-SigCin-Sig94 (no over capacity points for inclusion in first regression) model and the Evipas model (no v/c ratio provided by Evipas to allow for sorting). The models are presented in order of decreasing quality of fit (R^2).

Figure 22. Intersection Level Regression Results (All V/C)

Model	Intercept	Coefficient	R-Square
Transyt	18.532	0.0093	0.3381
Sig94-Alt	7.1539	0.6264	0.2975
Evipas	17.121	0.1962	0.0399
Sidra	21.987	0.0047	0.0384

Figure 23. Intersection Level Regression Results (V/C < 1.0)

Model	Intercept	Coefficient	R-Square
HCS	3.1262	0.9419	0.6939
Transyt	2.3282	0.8425	0.6759
Sidra	2.929	0.88	0.6216
Sig94-Alt	5.8956	0.684	0.2883

These regression analyses seem to validate the visual inspection conclusions. There is a significantly better quality of fit in the regressions where the v/c ratio was limited. The HCM-based and Sidra models both show coefficients close to 1.0 and intercepts close to 0, which suggests that the models represent field conditions well. Evipas and to a lesser extent Transyt show a coefficient further away from 1.0, reinforcing the visually noted trend of model overprediction.

5.4.3 Stepwise Inclusion of Variables

After completing the simple regression analysis, the study team moved on to a stepwise regression analysis. This process involves inserting and removing variables from a regression analysis based on their contribution to the resulting model accuracy. If a variable does not make a significant contribution to the accuracy of the model, it will not be included. This allows a determination of what factors caused the difference between model prediction and field delay.

While this process can create a complicated model, it allows the user to easily select those variables which are of statistical significance in the problem. In this particular problem, there were three key variables plus their interactions that the study team wished to evaluate for significance. To complete this analysis, the study team used the SAS package's stepwise regression analysis module (PROC STEPWISE). The default values for variable inclusion and removal were used.

As with the regression analysis discussed previously, separate analyses were made for under capacity and oversaturated conditions. Figure 24, Figure 25 and Figure 26 summarize the variables included in the stepwise regression analysis and the result of those analyses.

Figure 24. Variables included in Stepwise Regression Analysis

D	The estimated stopped delay as determined by the individual models
VC	The intersection-level volume to capacity ratio reported by the model. In some cases this is Xc, in others it is the highest v/c ratio for any lane group or lane.
ST	The signal controller type. The pretimed site was removed from the study. Thus, ST is either actuated or volume-density.
D*VC	The interaction between Delay and VC.
D*ST	The interaction between Delay and Signal Controller Type.
VC*ST	The interaction between VC and Signal Controller Type.
D*VC*ST	The interaction between Delay, VC, and Signal Controller Type.

Figure 25. Results of Stepwise Regression Analysis for All VC

Model	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
Sig94-Alt	0.64	94	X	X		X			X	X
Transyt	0.57	94	X		X					
Sidra	0.32	94	X		X	X		X		
Evipas	0.28	94	X	X	-		-		-	-

Figure 26. Results of Stepwise Regression Analysis when VC < 1.0

Model	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
Transyt	0.91	67	X	X		X	X			
Sidra	0.85	67	X	X	X					X
HCS	0.83	70	X	X	X	X	X			X
Sig94-Alt	0.67	91	X	X		X	X	X	X	X

From these tables, we can clearly see that none of the models performed well when oversaturated data points were included in the analysis. Limiting the dataset to those points where all the lane groups were under capacity, however, gives dramatically improved results for Transyt and Sidra. While there is no way to discuss “improvement” of HCS, it also performed very well. The only model which does not seem to perform well is the Sig94-Alt model. One likely cause of this lack of improvement is the number of oversaturated points as determined by the model. While both Sidra and Transyt generated 27 oversaturated data points, Sig94-Alt only generated 3.

A more subtle point to consider is the number of variables entered into the model. A higher number of variables will always improve the R^2 of that particular model. In this, the Sidra and Transyt models have an edge over the HCS and Sig94-Alt models, as Sidra and Transyt tie for the minimum number of variables included during the stepwise analysis.

5.5 Lane Group Comparisons

After completing the intersection level comparisons, the study team turned their attention to the lane group data. As with the intersection data, the first step was a visual analysis of the data. While this involved a considerable number of graphs (9 lane groups x 5 models = 45 graphs of data), it produced disappointingly few trends. The majority of the lane group graphs showed no discernible trend to the data. Because of this, the study team decided to skip the simple regression step of analysis and move directly to the stepwise regression analysis. For reference, the graphs produced for lane groups are included in the Appendix D.

The lane group results will be presented model-by-model. For each model, the results from all lane groups included in the study are shown. The lane groups are labeled by turning movement (L for left, T for through, and R for right) and are presented in alphabetical order. Because of the significant difference between permitted and protected left turn movements, these were considered as separate lane groups, indicated by a -Per or -Pro, respectively. A -Mixed extension indicates protective and permissive phasing are present. Finally, because of specific interest by NCDOT, the LTR lane group was divided into two types: the single lane LTR (denoted LTR) and the multi-lane LTR (denoted 2LTR). All the multi-lane LTRs in this study had two lanes.

Results are shown for each model; however, they are labeled in a distinct manner. Results for all of the HCM-based models (HCS, SIG/CINEMA, and the HCM module in SIGNAL94) are reported as “HCM.” There were minor differences between the three different implementations of the HCM, but nothing statistically significant was discovered. The results shown are based on the HCS delay estimations.

The results reported under SIGNAL94-Alt are for the alternative analysis module of SIGNAL94. All the other results are given by specific model name.

Overall, the sample sizes used in these analyses were very good. Most of the lane groups averaged near 100 data points. All of them have over 50 except for the L-Mixed lane group, which dips into the upper thirties when only under capacity points are considered.

5.5.1 HCM Evaluation

Figure 27 gives the results from stepwise regression analysis of the HCM-based models. Recall that there is only an under capacity model for the HCM-based models.

Figure 27. HCM-based Stepwise Regression Results

Including Points with v/c < 1.0

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.81	64	X	X	X		X			
L-Mixed	0.10	40	X			X				
L-Per	0.42	87	X					X		
L-Pro	0.55	55	X	X		X	X		X	
LTR-Per	0.28	75	X				X			
R	0.50	112	X	X	X	X				
T	0.00	97								
TR	0.56	116	X	X		X		X	X	

Clearly these results are less promising than the intersection results discussed above. While the 2LTR-Per lane group showed an excellent fit, only about half of the remaining lane groups showed a good fit, the L-Pro, R, and TR. The rest are poor to moderate.

The T lane group performed so poorly that no variables were entered into a model. This was unexpected, as the T lane group is usually the cleanest at the intersection. However, the T lane group was the only lane group where observers’ sight distance was a problem in measuring queues. At two sites in particular, the queuing was well past what could be measured, even though the lane group was under capacity. This fact can cause significant under prediction of delay, likely the reason for the poor fit. If the problem is data-related, it should be expected to show up in other models.

5.5.2 EVIPAS Evaluation

Figure 28 gives the results from stepwise regression analysis of the EVIPAS model. Recall that there is only an over capacity model for EVIPAS, as it does not provide a v/c ratio.

Figure 28. EVIPAS Stepwise Regression Results

Including All Points

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.40	64	X	X	-		-		-	-
L-Mixed	0.27	46	X	X	-		-		-	-
L-Per	0.44	91	X	X	-		-		-	-
L-Pro	0.25	63	X	X	-	X	-		-	-
LTR-Per	0.42	76	X	X	-		-		-	-
R	0.00	114			-		-		-	-
T	0.11	96	X	X	-		-	X	-	-
TR	0.35	112	X	X	-	X	-	X	-	-

These results contain no lane groups with a good quality of fit. This is most likely due to the lack of a true “capacity” model. Unlike the HCM-based models, EVIPAS was able to generate a fit for the T lane group. However, the fit was very poor and does not refute the conclusions discussed above.

EVIPAS was able to generate moderately good fits for the 2LTR-Per, L-Per, and the LTR-Per lane groups. This suggests that the permissive left turning model in EVIPAS is fairly accurate.

5.5.3 SIDRA Evaluation

Figure 29 gives the results from stepwise regression analysis of the SIDRA model. Section A contains results when all points are included; Section B includes only those with v/c < 1.0.

Figure 29. SIDRA Stepwise Regression Results

A. Including All Points

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.75	64	X	X	X		X			
L-Mixed	0.50	46	X	X	X		X			
L-Per	0.34	106	X			X				
L-Pro	0.34	78	X			X			X	
LTR-Per	0.10	80	X		X					
R	0.54	114	X			X	X	X		X
T	0.20	114	X			X	X		X	X
TR	0.70	118	X	X			X	X	X	X

B. Including Points with v/c < 1.0

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.75	64	X	X	X		X			
L-Mixed	0.53	38	X	X	X		X			
L-Per	0.50	85	X	X						X
L-Pro	0.75	65	X			X		X		X
LTR-Per	0.13	79	X		X					
R	0.59	111	X			X	X	X		X
T	0.21	97	X	X		X		X		
TR	0.73	115	X	X		X		X		

SIDRA performed very well, especially in under capacity conditions. In those analyses, only two of the eight lane groups did not show at least a 0.5 R-square value. In fact, both the 2LTR-Per and the L-Pro lane groups had an excellent fit.

The T lane group again showed a very poor fit. Especially considering the quality of fit among the other lane groups, this is further evidence that the problem is with the data and not with the model. The other poorly fit lane group was the LTR-Per, which is perhaps the most complicated to model of all.

5.5.4 SIGNAL94-Alt Evaluation

Figure 30 shows the results from the evaluation of the alternative analysis module in SIGNAL94.

Figure 30. SIGNAL94-Alt Stepwise Regression Results

A. Including All Points

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.81	64	X	X	X		X			
L-Mixed	0.27	46	X	X			X			
L-Per	0.31	108	X			X				
L-Pro	0.46	78	X			X	X		X	
LTR-Per	0.31	80	X				X			
R	0.49	114	X	X		X				
T	0.29	114	X	X		X	X	X		X
TR	0.54	117	X	X				X		X

B. Including Points with v/c < 1.0

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.81	64	X	X	X		X			
L-Mixed	0.36	39	X	X			X			
L-Per	0.41	87	X					X		
L-Pro	0.67	70	X			X		X		X
LTR-Per	0.32	75	X				X			
R	0.50	113	X	X	X	X				
T	0.29	101	X	X		X	X		X	X
TR	0.59	115	X	X		X		X	X	X

SIGNAL94-Alt performed well in many of the lane groups. The 2LTR-Per, L-Pro, R, and TR lane groups all show a good fit. None of the lane groups performed poorly. The T lane group model in SIGNAL94-Alt showed the best fit to field data yet, though it was still not a good fit.

The unique feature of these models is the lack of difference between the “all points” models and the “points with v/c < 1.0” models. This is most likely due to the exclusion of only a small number of points based on v/c ratio. In fact, in only one lane group does the number of over capacity data points exceed 10.

5.5.5 TRANSYT-7F Evaluation

Figure 31 shows the results from the evaluation of TRANSYT-7F.

Figure 31. TRANSYT-7F Stepwise Regression Results

A. Including All Points

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.79	64	X	X	X		X			
L-Mixed	0.38	46	X	X			X			
L-Per	0.31	106	X			X			X	
L-Pro	0.30	77	X			X			X	
LTR-Per	0.45	80	X				X			
R	0.42	114	X	X		X				
T	0.23	114	X			X		X	X	X
TR	0.36	118	X		X	X				

B. Including Points with v/c < 1.0

Lane Group	R-Square	N	Variables Included							
			Int	D	VC	ST	D*VC	D*ST	VC*ST	D*VC*ST
2LTR-Per	0.79	64	X	X	X		X			
L-Mixed	0.52	43	X	X			X			
L-Per	0.40	89	X			X			X	
L-Pro	0.58	63	X	X				X		X
LTR-Per	0.25	79	X		X		X			
R	0.42	114	X	X		X				
T	0.15	101	X	X						X
TR	0.51	113	X	X						X

TRANSYT performed almost as well as SIDRA. While only one lane group had an R-square above 0.7 (2LTR-Per), six of the eight lane groups had an R-square above 0.4. Once again, the T lane group had a poor fit, continuing evidence that there is a problem in the field data.

The other poorly-fit lane group was the LTR-Per. As noted before, this is perhaps the most complicated lane group to model. Not only does it include a permissive left turn model, but also with mixed movements in a single lane.

5.6 Summary

Figure 32 contains a summary of the R-square values reported in the previous tables.

Figure 32. Summary of Stepwise Regression Results

A. Including All Points

Lane Group	EVIPAS	SIDRA	SIGNAL94-Alt	TRANSYT-7F		Average by Lane Group
2LTR-Per	0.40	0.75	0.81	0.79		**0.69**
L-Mixed	0.27	0.50	0.27	0.38		**0.36**
L-Per	0.44	0.34	0.31	0.31		**0.35**
L-Pro	0.25	0.34	0.46	0.30		**0.34**
LTR-Per	0.42	0.10	0.31	0.45		**0.32**
R	0.00	0.54	0.49	0.42		**0.36**
T	0.11	0.20	0.29	0.23		**0.21**
TR	0.35	0.70	0.54	0.36		**0.49**
Average by Model	**0.28**	**0.43**	**0.44**	**0.41**		

B. Including Points with v/c < 1.0

Lane Group	HCM-Based	SIDRA	SIGNAL94-Alt	TRANSYT-7F		Average by Lane Group
2LTR-Per	0.81	0.75	0.81	0.79		**0.79**
L-Mixed	0.10	0.53	0.36	0.52		**0.38**
L-Per	0.42	0.50	0.41	0.40		**0.43**
L-Pro	0.55	0.75	0.67	0.58		**0.64**
LTR-Per	0.28	0.13	0.32	0.25		**0.25**
R	0.50	0.59	0.50	0.42		**0.50**
T	0.00	0.21	0.29	0.15		**0.16**
TR	0.56	0.73	0.59	0.51		**0.60**
Average by Model	**0.40**	**0.52**	**0.49**	**0.45**		

First, consider the averages by lane group. These values give indications of difficult modeling tasks. Regardless of which set of points are used, the T lane group has the worst average R-square value. As noted above, the study team believes that observers' sight distance problems led to the difficulty in producing a good model to field fit.

Another possible explanation for the poor model performance lies with the estimate of saturation flow. All the models began with an ideal saturation flow rate of 1900 vphpl. Review of the field data suggests that this may have been a significant underestimation for some of the T lane groups. In fact, review of the peak 15 minutes at Site 4 suggest that the actual saturation flow rate was in excess of 2300 vphpl, a difference which would certainly cause poor prediction by the model.

The next-to-worst lane group was the LTR-Per lane group. Several factors likely produce this result. For one, this lane group involved a permissive turning model that required the inclusion of opposing traffic effects. For another, it involved a mixed traffic stream. In addition, many times in the study an LTR-Per lane group was opposed by another LTR-Per lane group. In this situation, there were times when one permissive left could assist another, adding to the complications in attempting to model this lane group's performance.

At the other end of the spectrum, the 2LTR lane group seems to have been the easiest to model. This seems unusual, as this lane group had many of the same characteristics as the LTR lane group. The improved fit was likely due to the v/c ratio of these data points. None of the data points were determined to be over capacity by any of the computer models. In fact, review of the data indicates that not only are these points under saturated, they are often very under saturated ($v/c < 0.5$).

In the under capacity table, the next three highest lane groups are the L-Pro, the TR, and the R. All of these lane groups are unopposed movements. It seems logical that unopposed movements would be easier to model than opposed movements.

Now, consider the averages by model. For one, there are no tremendous differences between the models. With all points included, three of the models differ in R-square by only 0.03. When only under capacity points are included, the gap between all four is only 0.12.

If the two averages are considered together, the SIDRA model has the best combined score. The SIGNAL94-Alt model is a close second, and TRANSYT-7F a close third. As neither HCS nor EVIPAS has more than one score, there is no way to discern their place in this order.

A closer look at Figure 32 reveals some additional points. Consider the under capacity section of the table. Some lane groups perform similarly in all models. For example, the 2LTR-Per lane group performs very well in all models. The LTR-Per performs poorly in all. Some lane groups, however, are modeled particularly well by a portion of the models. Note the L-Mixed lane group. Both SIDRA and TRANSYT performed significantly better than the other models.

This phenomenon is not consistently attributable to any particular model. SIDRA had a notably higher score for the TR lane group than any other model. It also had a notably lower score for the LTR-Per lane group. SIDRA is again higher in the L-Pro. This suggests that certain intersection types might be best modeled with a particular piece of software.

5.7 Model Rating

Figure 33 contains the model rating based on the evaluation discussed above. The individual scores (for “all points” and for “v/c<1.0”) are based on the quality of fit discussed in the stepwise regression analysis above. The “Composite” score is an average of the individual scores.

Figure 33. Model Evaluation Rating (1-5, 5 best)

	Model				
Evaluation	**HCM**	**Evipas**	**Sidra**	**Sig94-Alt**	**Transyt**
Including all points		2	4	4	4
Including poitns with v/c<1.0	3		5	4	4
Composite	**3**	**2**	**4.5**	**4**	**4**

6. Optimization

6.1 Introduction

The next phase of the project was to evaluate the optimization capabilities of each of the models. The purpose of this phase was to determine which model provided the most significant improvement in performance after optimization. This phase was of significant importance to the project because NCDOT uses computer-based signal timing models for optimization much more often than for simple simulation of current conditions.

After the model evaluation phase of the project was complete, the study team, in conjunction with the advisory committee, began to consider which models to include in the optimization phase. It was decided to pursue optimization with all the models except, of course, for HCS, which has no optimization capabilities.

6.2 Methodology

6.2.1 Optimization Intervals

The first step in the methodology was to determine which time periods should be optimized and evaluated. As noted earlier, when the study team collected data for the evaluation phase of the project, the decision was made to collect sufficient information to allow the use of the same data in the optimization phase of the project.

The question then arose as to which of the time periods should be optimized. The study team deemed that peak periods from each two hour block of data collection were the most significant. This is due to the fact that the determination of signal timing plans are based on peak flows. Thus, it would not be feasible to optimize every 15-minute time interval, given the likelihood that field controllers do not have the capability to implement so many interval based settings. This gave a total of 14 optimization intervals for comparison.

6.2.2 Phasing Plans

Since not all of the models under investigation were able to select a phasing plan as part of their optimization module, the question arose as to which phasing plans should be evaluated in the models. Three different phasing plans were chosen. The first is based on the displayed (field) phasing plan. The other two were extracted from two models which are able to select a phasing plan as part of their optimization procedures (SigCinema and Signal94).

Optimization was carried out as follows. Each model would first be used to optimize the cycle length and splits using the field phasing plan. Then, each model optimized cycle length and splits using a SigCinema and a Signal94 phasing plan. This methodology was applicable to all models except EVIPAS, which has no average phase structure. The methodology used for optimization of the EVIPAS model is discussed later (in 6.3.1).

6.2.3 Comparison Methodology

After completing the optimization runs discussed above, the study team began the task of comparing the results obtained from the different models. The first question centered around what the optimization results were to be compared with. Since the task was to evaluate the level of improvement from optimization, optimization results were compared to simulation results from the same model and time period.

The next questions centered around what measure to compare. Several delay measures at lane group and intersection level were considered. In the end, the study team chose the intersection LOS as the best comparative statistic, as it could be determined for all sites in all models, regardless of v/c conditions.

In summary, LOS of each optimization were compared with the simulation LOS results, and an average LOS improvement was determined across all the sites. The same process was applied to a comparison between field delay and model simulation delay. While this comparison is not as rigorous as that presented in the previous chapter, it does provide a good assessment of a model’s ability to improve on current signal timing.

6.3 Optimization Results

6.3.1 EVIPAS

As noted before, the nature of the Evipas optimization module is different from those of the other models. As Evipas is truly an actuated signal model, it is designed to optimize actuated signal parameters, such as max green, directly.

Unfortunately, the complex nature of such an optimization leads to one of the most frustrating features of Evipas. The model does not conduct an exhaustive search. Instead, it makes use of a hill climbing search procedure. While this does allow Evipas to deal with multivariate optimization, it leads to two key problems.

One, there is no guarantee that the solution given is the global optimal solution. The model may very well get caught at a local minimum and produce no better solutions. Two, there is no way to estimate how much longer the search process will continue. Most of the models give some indication of how far along in the optimization process they are, either by providing the percentage of the optimization complete or by reporting which cycle length is currently under scrutiny. The Evipas model can do neither.

Tests of the optimization module in Evipas determined that, as suspected, the more variables that must be optimized, the longer the optimization process will take. Some of the more complex optimizations took unacceptably long to run. In fact, one optimization was terminated after 5 consecutive days of computation without completion.

Reasonable optimization times (always less than 15 minutes, usually less than 5) were found to be possible if the optimization focused only on max greens. This, then, became the process for optimization in Evipas. Minimum greens, extensions, and gaps were left as in the field, and maximum greens were optimized by Evipas.

Consider Figure 35, which contains a summary of EVIPAS intersection -level LOS results. These include its ability to replicate field LOS (columns 2&3) and to improve on it by optimization (columns 3&4).

Note: Figure 35, and those like it for the other models, all use a common layout. A key to that layout is provided in Figure 34.

Figure 34. Key to Summary Tables

Model Delay
(Sec/veh)

Model LOS
(A-F)

LOS Difference
Model - Field

Figure 35. Summary of EVIPAS Optimization Results

Avg Int Delay **LOS** Score†	**Field Data**	**Simulation of Field Settings**	**Optimization of Field Settings Sep. 15 min Opts**
Site 1 AM Peak	16.0 **C**	10.4 **B** -1	7.44 **B** 0
PM Peak	23.2 **C**	21.6 **C** 0	29.8 **D** 1
Site 2 AM Peak	13.3 **B**	12.8 **B** 0	7.4 **B** 0
PM Peak	12.0 **B**	16.6 **C** 1	12.87 **B** -1
Site 3 AM Peak	23.3 **C**	83.4 **F** 3	79.49 **F** 0
PM Peak	28.6 **D**	51.2 **E** 1	40.29 **E** 0
Site 4 AM Peak	47.2 **E**	78.9 **F** 1	44.29 **E** -1
PM Peak	118.5 **F**	* **F** 0	* **F** 0
Site 5 AM Peak	13.7 **B**	6.8 **B** 0	4.97 **A** -1
PM Peak	11.0 **B**	6.1 **B** 0	5.05 **B** 0
Site 6 AM Peak	7.4 **B**	4.0 **A** -1	2.34 **A** 0
PM Peak	12.2 **B**	3.5 **A** -1	2.52 **A** 0
Site 7 AM Peak	6.7 **B**	10.3 **B** 0	4.19 **A** -1
Site 8 PM Peak	27.5 **D**	40.2 **E** 1	12.82 **B** -3
Average Score		**+ 0.3**	**- 0.4**

This table helps to demonstrate why the study team chose LOS instead of average delay as the statistic for comparison. While Evipas showed an improved delay in 12 of the 14 cases, there were only 5 cases where the LOS showed an improvement. Overall, Evipas fared well. In only one case did the model generate a timing plan which worsened the LOS. The average effect of Evipas optimization was an improvement of about half a LOS.

The simulation results indicate that EVIPAS generally overestimates intersection delay and LOS.

6.3.2 SigCinema

The SigCinema model was used both to optimize the field phasing plan and to select a new phasing plan. When selecting a new phasing plan, the model was allowed to select from all 121 of the default phasing plans included. Figure 36 contains a summary of the results of both of the optimizations in addition to a comparison of model vs. field LOS.

Figure 36. Summary of SigCinema Optimization Results

Avg Int Delay **LOS** Score†		**Field Data**	**Simulation of Field Plan (HCS Module)**	**Optimization of Field Plan**	**Optimization of SigCin Plan**
Site 1	AM Peak	16.0 **C**	19.1 **C** 0	17.2 **C** 0	16.8 **C** 0
	PM Peak	23.2 **C**	19.9 **C** 0	17.8 **C** 0	17.8 **C** 0
Site 2	AM Peak	13.3 **B**	12.5 **B** 0	7.2 **B** 0	5 **A** -1
	PM Peak	12.0 **B**	15.9 **C** 1	9.9 **B** -1	8.4 **B** -1
Site 3	AM Peak	23.3 **C**	* **F** 3	* **F** 0	* **F** 0
	PM Peak	28.6 **D**	* **F** 2	* **F** 0	* **F** 0
Site 4	AM Peak	47.2 **E**	* **F** 1	* **F** 0	* **F** 0
	PM Peak	118.5 **F**	* **F** 0	* **F** 0	* **F** 0
Site 5	AM Peak	13.7 **B**	11.1 **B** 0	4.5 **A** -1	4.5 **A** -1
	PM Peak	11.0 **B**	7.8 **B** 0	5.1 **B** 0	5.1 **B** 0
Site 6	AM Peak	7.4 **B**	5.4 **B** 0	2.7 **A** -1	2.7 **A** -1
	PM Peak	12.2 **B**	6.7 **B** 0	3.9 **A** -1	3.9 **A** -1
Site 7	AM Peak	6.7 **B**	9.2 **B** 0	6.6 **B** 0	3.1 **A** -1
Site 8	PM Peak	27.5 **D**	* **F** 2	26.5 **D** -2	9.2 **B** -4
Average Score			**+ 0.6**	**- 0.4**	**- 0.7**

The model performed generally as expected. Its estimations of field delay when simulating field conditions were generally accurate. Optimizing the field plan generated an overall improvement in intersection LOS, and selection of a new phasing plan with optimized splits generated greater improvement in intersection LOS.

A closer look at the table gives insight into the average scores. First, consider the simulation score. Earlier in this document, the HCS module was touted as a highly accurate model, yet here the average score suggests that it regularly overestimates LOS. However, the earlier results were tempered by the inability of this HCM-based model to estimate delay when $v/c > 1.0$. If this comparison is limited by that same constraint, the average difference between field and model LOS drops to about 0.1, paralleling the results discussed earlier.

The optimization of field plan average score indicates that the model was able to improve LOS by about half a level by optimizing the cycle length and splits. This ability is seen at a wide range of LOS scores, with an F improved to D, C to B, and B to A. The model was not, however, able to improve on the extremely oversaturated conditions at Sites 3 and 4. This indicates a need for additional capacity at intersections which cannot be provided simply by improved timing plans.

The same was true when SigCinema was allowed to select a new phasing plan. On a few occasions, the model selected the same plan as was in the field. In these cases, the delays in the upper left hand corners of boxes in the two optimization columns are the same. In most cases, this process generated only slightly better results than optimizing the field phasing plan. This suggests that the field plans were well developed.

There was one case, however, where the improvement was more dramatic. The Site 8 PM peak LOS improved from F to B when SigCinema selected and optimized the phasing plan. This site was marked as a leading candidate for future field validation.

6.3.3 Signal94

The SigCinema model was used both to optimize the field phasing plan and to select a new phasing plan. When selecting a new phasing plan, the model was allowed to select from all 49 default phasing plans included. The LOS results were determined using both the HCM module and the alternative analysis module.

Figure 37 and Figure 38 contain a summary of the results of both of the optimizations in addition to the simulation vs field data. Figure 37 contains results when the HCM module is applied. Figure-38 shows results obtained using the alternative analysis module.

Figure 37. Summary of Signal94 Optimization Results

Avg Int Delay **LOS** Score†	**Field Data**	**Simulation of Field Plan (HCS Module)**	**Optimization of Field Plan (HCS Module)**	**Optimization of Sig94 Plan (HCS Module)**
Site 1 AM Peak	16.0 **C**	19.1 **C** 0	16.6 **C** 0	10.3 **B** -1
PM Peak	23.2 **C**	19.9 **C** 0	19.8 **C** 0	8.0 **B** -1
Site 2 AM Peak	13.3 **B**	12.5 **B** 0	6.8 **B** 0	4.2 **A** -1
PM Peak	12.0 **B**	15.9 **C** 1	9.8 **B** -1	4.5 **A** -2
Site 3 AM Peak	23.3 **C**	* **F** 3	* **F** 0	* **F** 0
PM Peak	28.6 **D**	* **F** 2	* **F** 0	* **F** 0
Site 4 AM Peak	47.2 **E**	* **F** 1	* **F** 0	* **F** 0
PM Peak	118.5 **F**	* **F** 0	* **F** 0	* **F** 0
Site 5 AM Peak	13.7 **B**	11.1 **B** 0	3.8 **A** -1	3.8 **A** -1
PM Peak	11.0 **B**	7.8 **B** 0	4.8 **A** -1	4.8 **A** -1
Site 6 AM Peak	7.4 **B**	5.4 **B** 0	4.6 **A** -1	4.6 **A** -1
PM Peak	12.2 **B**	6.7 **B** 0	4.7 **A** -1	4.7 **A** -1
Site 7 AM Peak	6.7 **B**	9.2 **B** 0	6.3 **B** 0	3.4 **A** -1
Site 8 PM Peak	27.5 **D**	* **F** 2	17 **C** -3	9.8 **B** -4
Average Score		**+ 0.6**	**- 0.6**	**- 1.0**

The simulation vs. field results from this module are identical to the SigCinema model. The observations made about SigCinema's results are thus applicable to Signal94.

Signal94 seemed slightly superior to SigCinema at improving simulated field conditions both when optimizing the current phasing plan and when selecting a new plan. Again, as with the SigCinema model, Signal94 demonstrated its capabilities across a wide range of LOS conditions.

An item not indicated in the tables is the type of signal timing plan that was selected by Signal94. In all cases but one, the recommended phasing plan was a 2-phase signal, with mostly very short (on the order of 20-30 second) cycle lengths. In fact, a 20-second 2-phase plan was recommended for both the AM and PM peaks at Site 4, a significantly over capacity intersection. This suggests that a significant portion of the capacity was accounted for during yellow time on the signal. While purely numerical analysis of the results suggests that Signal94 performed better than SigCinema, these unusual phasing plans and cycle lengths obviously suggest otherwise.

Figure 38. Summary of Signal94 Alternate Optimization Results

Avg Int Delay **LOS** Score†	**Field Data**	**Simulation of Field Plan (Alt. Module)**	**Optimization of Field Plan (Alt. Module)**	**Optimization of Sig94 Plan (Alt. Module)**
Site 1 AM Peak	16.0 **C**	19.0 **C** 0	22.3 **C** 0	12.8 **B** -1
PM Peak	23.2 **C**	24.8 **C** 0	21.8 **C** 0	13.6 **B** -1
Site 2 AM Peak	13.3 **B**	14.3 **B** 0	7.6 **B** 0	4.9 **A** -1
PM Peak	12.0 **B**	17.8 **C** 1	10.9 **B** -1	5.6 **B** -1
Site 3 AM Peak	23.3 **C**	66.1 **F** 3	81.6 **F** 0	63.8 **F** 0
PM Peak	28.6 **D**	54.8 **E** 1	41.7 **E** 0	73.2 **F** 1
Site 4 AM Peak	47.2 **E**	84.5 **F** 1	137.5 **F** 0	132.6 **F** 0
PM Peak	118.5 **F**	35.5 **D** -2	121.5 **F** 2	91.9 **F** 2
Site 5 AM Peak	13.7 **B**	13.4 **B** 0	5.2 **B** 0	5.2 **B** 0
PM Peak	11.0 **B**	9.8 **B** 0	6.4 **B** 0	6.4 **B** 0
Site 6 AM Peak	7.4 **B**	6.7 **B** 0	3.5 **A** -1	5.5 **B** 0
PM Peak	12.2 **B**	8.4 **B** 0	6.1 **B** 0	6.1 **B** 0
Site 7 AM Peak	6.7 **B**	11.2 **B** 0	7.6 **B** 0	4.4 **A** -1
Site 8 PM Peak	27.5 **D**	37.3 **D** 0	23.8 **C** -1	7.5 **B** -2
Average Score	25.8	28.8 **+ 0.3**	35.5 **- 0.1**	31.0 **- 0.3**

The analysis of same splits by the alternative analysis module in Signal94 gave significantly different results. Comparing the results in Figure 37 and Figure 38, we see that the alternative analysis module indicates significantly reduced improvements in intersection LOS. In fact, these results are not as good as those of SigCinema.

Where the HCM module always showed improvement when optimization was performed, the alternative analysis method indicates that some of the “optimized” signal timing plans actually worsen the intersection LOS. The contradictory results of the two modules of Signal94 along with the unusual phasing plan selections make it difficult to determine the true quality of its optimization module.

6.3.4 Sidra

Sidra was the first model to utilize the full methodology as described above (6.2.3). Both Sidra and Transyt, which will be discussed next, were used to optimize three different phasing plans, the field plan, the plan recommended by SigCinema, and the plan recommended by Signal94. Figure 39 gives the results for SIDRA.

Figure 39. Summary of SIDRA Optimization Results

Avg Int Delay **LOS** Score†	**Field Data**	**Simulation of Field Plan**	**Optimization of Field Plan**	**Optimization of SigCin Plan**	**Optimization of Sig94 Plan**
Site 1 AM Peak	16.0 **C**	18.3 **C** 0	15.2 **C** 0	20.7 **C** 0	13.9 **B** -1
PM Peak	23.2 **C**	24.6 **C** 0	14.5 **B** -1	14.5 **B** -1	5 **A** -2
Site 2 AM Peak	13.3 **B**	12.9 **B** 0	12.1 **B** 0	6.5 **B** 0	6.5 **B** 0
PM Peak	12.0 **B**	18.9 **C** 1	13.5 **B** -1	9.3 **B** -1	9.3 **B** -1
Site 3 AM Peak	23.3 **C**	92.2 **F** 3	85.2 **F** 0	90.1 **F** 0	60.5 **F** 0
PM Peak	28.6 **D**	47.3 **E** 1	48 **E** 0	56.5 **E** 0	42 **E** 0
Site 4 AM Peak	47.2 **E**	5648.9 **F** 1	247.7 **F** 0	118.5 **F** 0	200.8 **F** 0
PM Peak	118.5 **F**	33.2 **D** -2	31.8 **D** 0	38.5 **D** 0	23.6 **C** -1
Site 5 AM Peak	13.7 **B**	14.2 **B** 0	5.7 **B** 0	5.7 **B** 0	5.8 **B** 0
PM Peak	11.0 **B**	10.4 **B** 0	7.3 **B** 0	9.4 **B** 0	9.4 **B** 0
Site 6 AM Peak	7.4 **B**	6.6 **B** 0	4.1 **A** -1	4.1 **A** -1	4.3 **A** -1
PM Peak	12.2 **B**	8.2 **B** 0	6.1 **B** 0	6.1 **B** 0	6.1 **B** 0
Site 7 AM Peak	6.7 **B**	11.6 **B** 0	9.2 **B** 0	4.5 **A** -1	4.7 **A** -1
Site 8 PM Peak	27.5 **D**	20.1 **C** -1	14.9 **B** -1	7.0 **B** -1	6.9 **B** -1
Average Score	25.8	426.2 **+ 0.2**	36.8 **- 0.3**	28.0 **- 0.4**	28.5 **- 0.6**

First of all, the simulation average score reasserts what was discussed in the previous chapter, that the Sidra model is highly accurate at estimating field conditions. This fact lends more credence to the results of the optimization runs made by Sidra.

Like both SigCinema and Signal94, Sidra shows improved LOS conditions when given the opportunity to optimize green splits and cycle length, regardless of which of the three timing plans are considered. Unlike the other models, however, Sidra did not show improvement when LOS conditions were initially poor. Sidra was able to show improvement only when the initial LOS was C or better. This seems reasonable, as an intersection which is truly oversaturated generally requires additional capacity in terms of lanes, not simple adjustments to green splits and cycle lengths.

The timing plans suggested by SigCinema and Signal94 performed slightly better than the field plans, and those of Signal94 performed slightly better than those of SigCinema. As noted before, the phasing plans and cycle lengths suggested by Signal94 were almost always 2-phase and very short. When Sidra optimized the cycle length for these 2-phase phasing plans, the cycle lengths were usually above 200 seconds in length, casting additional suspicion on the Signal94 results.

6.3.5 Transyt

As noted before, the Transyt model was used to optimize green splits for three different phasing plans, the field plan, the SigCinema plan, and the Signal94 plan. The results of these optimizations are summarized in Figure 40.

Figure 40. Summary of Transyt Optimization Results

Avg Int Delay LOS Score†	Field Data	Simulation of Field Plan	Optimization of Field Plan	Optimization of SigCin Plan	Optimization of Sig94 Plan
Site 1 AM Peak	16.0 **C**	26.6 **D** 1	28.2 **D** 0	27.3 **D** 0	22.5 **C** -1
PM Peak	23.2 **C**	33.5 **D** 1	31.4 **D** 0	31.1 **D** 0	20.8 **C** -1
Site 2 AM Peak	13.3 **B**	13.5 **B** 0	13.2 **B** 0	11.2 **B** 0	11.2 **B** 0
PM Peak	12.0 **B**	16.9 **C** 1	17.5 **C** 0	14.2 **B** -1	14.2 **B** -1
Site 3 AM Peak	23.3 **C**	95.2 **F** 3	109.4 **F** 0	123.8 **F** 0	120.8 **F** 0
PM Peak	28.6 **D**	58.2 **E** 1	76.8 **F** 1	86.3 **F** 1	94.2 **F** 1
Site 4 AM Peak	47.2 **E**	1475.5 **F** 1	215.4 **F** 0	212.5 **F** 0	149.1 **F** 0
PM Peak	118.5 **F**	674.0 **F** 0	78.0 **F** 0	64.8 **F** 0	53.5 **E** -1
Site 5 AM Peak	13.7 **B**	13.6 **B** 0	12.3 **B** 0	12.3 **B** 0	12.3 **B** 0
PM Peak	11.0 **B**	10.2 **B** 0	10.2 **B** 0	10.2 **B** 0	10.2 **B** 0
Site 6 AM Peak	7.4 **B**	5.8 **B** 0	8.4 **B** 0	8.4 **B** 0	8.4 **B** 0
PM Peak	12.2 **B**	7.9 **B** 0	11.0 **B** 0	11.0 **B** 0	11.0 **B** 0
Site 7 AM Peak	6.7 **B**	13.2 **B** 0	17.5 **C** 1	10.5 **B** 0	9.4 **B** 0
Site 8 PM Peak	27.5 **D**	58.3 **E** 1	31.6 **D** -1	24.6 **C** -2	24.6 **C** -2
Average Score	25.8	178.7 **+ 0.6**	47.2 **+ 0.1**	46.3 **- 0.1**	40.2 **- 0.4**

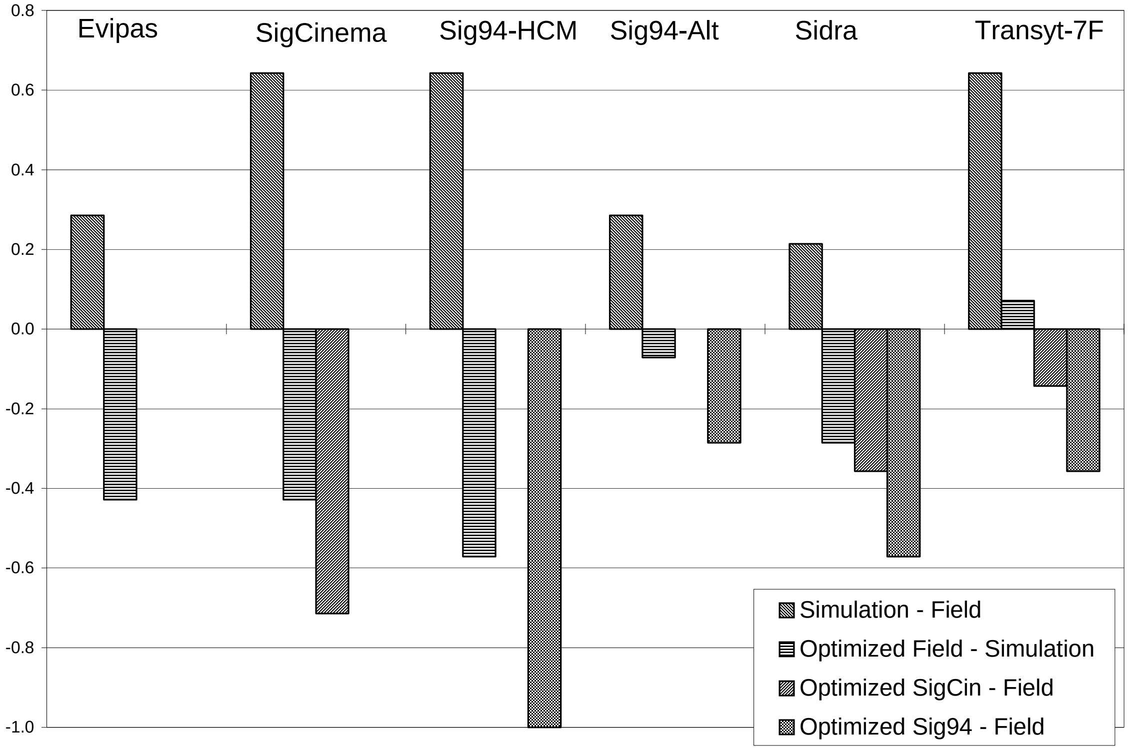
Except perhaps for the Signal94 model, the results of the Transyt model optimizations show the least improvements. First, consider the results of optimizing the field phasing plans. Transyt was the only model in the study which actually had an average score that indicated poorer LOS conditions after optimization. In addition, while conditions did improve when optimizing with the SigCinema and Signal94 plans, the improvements were less than those generated by the other models.

These poor results may all have the same root cause. As noted early in this paper, Transyt was selected for inclusion in this study because of its use by NCDOT to determine cycle lengths and splits for intersections in North Carolina. The odds are that it was used to develop the field timing plans which were studied at the 8 sites in this study. This implementation of previously optimal Transyt results may have biased the study against this model.

6.4 Summary

To better visualize the results from each of the models, Figure 41 contains a graphical representation of the average improvement in LOS obtained form each of the optimization experiments discussed above. The average simulation scores are also included.

Figure 41. Graphical Summary of Optimization Results from All Models



In Figure 41, the ideal model should have an average score of 0.0 in the simulation column. This would indicate that the model generated, on average, exactly the same LOS conditions as were measured in the field. Then, in each of the optimization columns, the lowest possible (highest negative value) indicates the greatest improvement in conditions by the optimization method.

The Sidra model has the best simulation result, and the Signal94-Alt has the best combination of optimization results, though problems with the optimization results from the Signal94-HCM model mentioned before cast doubt on the accuracy of this result. The Signal94-Alt results, which from this table seem to better estimate field conditions, suggest that the improvements are overestimated by the Sig94-HCM module.

Based on the results discussed in this chapter, the study team determined relative scores for each model’s optimization capabilities, which are shown in Figure 42.

Figure 42. Summary of Rankings of Model Optimization Results (1-5, 5 best)

Optimization using...	**EVIPAS***	**SIG/CINEMA****	**SIGNAL94-HCM***	**SIGNAL94-Alt****	**SIDRA**	**TRANSYT-7F**
...Field Plan	3	3	3	2	3	1
...SigCinema Plan	-	4	-	-	4	2
...Signal94 Plan	-	-	3	3	5	3
Composite Ranking	**3**	**3.5**	**3**	**2.5**	**4.0**	**2**
* Evipas was only used to optimize field max greens, so there is only one ranking						
** SigCinema and Signal94 were used to select and optimize their own phasing plan						
but not used to optimize each other's plan, so there are only 2 rankings.						

7. Summary, Conclusions and Recommendations

7.1 Project Rationale

This research project was initiated to provide assistance to NCDOT regarding the selection of computer-based signal timing models at isolated intersections in the state. With the accelerated urbanization of fringe areas surrounding the major metropolitan centers in the state, and the subsequent increase in the sheer number of traffic signal installations statewide, the issue of deciding which model—or series of models-- is appropriate for statewide use becomes very critical. Further compounding the problem is the fact that there are currently no national guidelines on signal timing, compared, say, with signal installation warrants. Therefore, there was considerable interest at NCDOT in producing information on the utility of a number of alternative signal timing models with a view towards standardization of signal timing and operational analysis for isolated intersections in the state. This study was intended to address that very problem.

7.2 Summary

The study began by formulating a number of key issues, such as
What should be the scope of the evaluation method ?
Which models should be considered for evaluation ? and
What criteria should be used for evaluation ?

With respect to item (a), the scope of the evaluation method, the research team along with the NCDOT project technical committee, agreed that the study should be limited to “isolated intersections” which are not part of a signal system; that the evaluation should encompass pre-timed, actuated and volume-density controlled intersections; and that the study sites should be limited to a 50-mile radius around the Research Triangle Region. The latter restriction was dictated by budget considerations.

Regarding item (b), the consideration of models for inclusion in the evaluation, two steps were taken. First, a large pool of models were evaluated based primarily on a review of the user manuals and experience in using the model software on simple, sample problems. Subsequently, a limited number of models were selected for further field validation. In all, six models were carried over to the second phase:

Highway Capacity Software (HCS) (1994 version; release 2.4 d);
EVIPAS for actuated control (1997);
SIDRA (version 5.02 ; 1996);
SIG/CINEMA (version 1.11; 1996);
SIGNAL94 (version 1.20; 1997) and
TRANSYT-7F (release 7.2, 1996).

Detailed descriptions of the models' capabilities are given in Chapter 2 of this document.

There were three basic criteria upon which these models were evaluated. The first is the model "usability", a general term which reflects its ease of use, diversity of applications to various intersection configurations and control methods, clarity and extent of text and graphical output, etc. This criterion, while somewhat subjective, is important to NCDOT from the perspective of training their personnel in the use of the models.

The second criterion was the model's accuracy in predicting field conditions. This was evaluated by comparing delays measured in the field (based on 28 hours of data collection at eight sites) and the predicted model delays when the field conditions were simulated in each model. Delays were compared for lane groups and intersection-wide. In all cases, most of the models' default calibration parameters were retained, as this would most likely resemble the typical use of the models by NCDOT personnel. Statistical regression analysis tools available in the SAS computer package were used extensively to determine the significance of the results obtained in this phase.

The final criterion relates to the effectiveness of the model optimization rules. Here, the team looked at the capabilities of each model in selecting optimal phase plans, cycle lengths and green splits. Further, the team evaluated the degree to which the simulated field performance could be improved through optimization of peak period settings. The original intent of this analysis was to recommend and then field-evaluate proposed signal timing changes to field settings at selected sites. Unfortunately, due to constraints on the phase sequences implemented by NCDOT, the field assessment of proposed optimal settings could not be carried out within the scope and budget of this study (since a phase plan change may involve the addition of signal heads to add a protected phase). However, the recommendations derived from the optimization portion of the study stand, and could be implemented at a later date.

7.3 Conclusions

7.3.1 Model Usability

As explained in Section 7.2, the first set of results pertains to the models' usability. A summary of these findings is depicted in Figure 43 below. Ratings varied from zero (worst) to five (best), based on the set of criteria indicated in the figure. As shown, both SIDRA and SIG/CINEMA had the highest usability ratings, followed in order by SIGNAL94, TRANSYT, HCS and EVIPAS.

Figure 43. Summary of Model Usability Ratings (0-5)

Usability Criterion	Hcs	Sig/Cin	Signal94	Sidra	Transyt	Evipas
Actuated Control	2	2	2	4	2	4
Flexible Phasing	3	4	5	3	3	4
Existing Timing	5	5	4	5	5	2
Coordination	5	5	5	5	2	1
2+ Optimization	0	2	3	4	4	4
Input Diversity	3	3	3	4	5	4
Ease of Input	4	5	2	3	1	3
Output Diversity	2	3	3	5	4	3
Understandable	4	5	4	4	4	2
TOTALS	**3.2**	**3.8**	**3.4**	**4.1**	**3.3**	**3.1**

7.3.2 Model Evaluation

The second set of results pertains to the degree to which the model delay results match field delays at the intersection level. Since all HCS-based models (HCS, Sig/Cinema and Signal 94) can generate intersection level delays only when all lane groups are undersaturated, the comparison below is limited to v/c <1.0. Figure 44 gives the R-square value for the stepwise regression models between field intersection delay and model delay prediction. Note that all HCS-based models are represented by HCS. The regression model explanatory variables are highlighted by (x). The sample size refers to the number of 15-minute periods in which intersection level delays were calculated for each model. That number varies according to each model's estimation of v/c ratio. Overall, the results were very encouraging, given the fact that no model calibration was undertaken. Both SIDRA and TRANSYT performed well both in terms of the R-square value and in having the fewest number of explanatory variables. This is important since ideally, we should be able to explain the field delay by the model delay estimate only. The fact that additional explanatory variables are needed indicates that the model estimate must be adjusted using those variables in order to match the field delay. Notably absent from the list is the EVIPAS model, which cannot generate a v/c ratio. However, this model performed very poorly (R^2= 0.28) when all data were considered.

Figure 44. Intersection Level Stepwise Regression Results

	Hcs-Type	Sidra	Signal-94Alt	Transyt-7F
R-square	0.83	0.85	0.67	0.91
Sample Size	70	67	91	67
Variables in Model				
Intercept (INT)	X	X	X	X
Model delay (D)	X	X	X	X
Model v/c Ratio (VC)	X	X		
Signal Type (ST)	X		X	X
D * VC	X		X	X
D * ST			X	
VC * ST			X	
D * VC * ST	X	X	X	

The next set of evaluation tests pertains to the matching of field and model delays at the lane group level. Using the same reasoning as in the previous section, Figure 45 below summarizes the obtained model R^2 values for all under-saturated lane groups. The R^2 given in the table are also averaged by model and lane group to give an indication of which model is generally superior across all lane groups and which lane group delay is better estimated across all models. Not surprisingly, the lane group level fit is somewhat inferior to the intersection level fit for all models. This is to be expected since the intersection-level delays may mask a number of deficiencies that exist at the lane group level. Overall, SIDRA and Signal94-Alt methods fared best, followed by TRANSYT and the HCM-based models. The best delay estimates came from the 2LT-per , L-Pro, TR and R lane groups. The worst performance occurred on the T lane group. Further investigation of this unexpected result indicated that this was more a reflection of the inability to properly measure long queue lengths in the field, which occurred predominantly for the “T” lane group. The LTR-per lane group did not fare well either, but that was not unexpected due to the complexity of traffic behavior in a single lane, permissive lane group.

Figure 45. R-Square Regression Results at the Lane Group Level

Lane Group	Hcs-Type	Sidra	Signal-94Alt	Transyt-7F	Lane Group Avg.
2LTR-Per	0.81	0.75	0.81	0.79	**0.79**
L-Mixed	0.1	0.53	0.36	0.52	**0.38**
L-Per	0.42	0.5	0.41	0.4	**0.43**
L-Pro	0.55	0.75	0.67	0.58	**0.64**
LTR-Per	0.28	0.13	0.32	0.25	**0.25**
R	0.5	0.59	0.5	0.42	**0.5**
T	0	0.21	0.29	0.15	**0.16**
TR	0.56	0.73	0.59	0.51	**0.6**
Model Avg.	**0.4**	**0.52**	**0.49**	**0.45**	

7.3.3 Model Optimization

This last set of results pertains to the quality of the signal optimization plan for each model. As stated in Chapter 6, this was done by comparing the average intersection delay in the peak period at each site before and after signal optimization. In all cases, the field phasing plans were used, and in some cases plans derived by SigCinema and Signal94 were also evaluated. Figure 46, reproduced below from Figure 42 gives a summary of model optimization ratings. Once again, both Sidra and SigCinema rated highest on these criteria, with TRANSYT surprisingly not performing as expected.

Figure 46. Summary of Model Optimization Results

Phase Optimization	Evipas	SigCin	Sidra	Signal-94HCM	Signal-94Alt	Transyt-7F
Field Plans	3	3	3	3	2	1
SigCinema Plans	-	4	4	-	-	2
Signal94 Plans	-	-	5	3	3	3
Composite Rating	3	3.5	4	3	2.5	2

7.4 Study Recommendations

Before recommending a particular model, it is important to indicate that, in general, all models except perhaps for EVIPAS fared rather well in this study. Considering the fact that no model was calibrated—the team used the models' defaults for saturation flows, lost time, sneakers, etc—the levels of correlation between field and model delays were quite good, all exceeding 67% at the intersection level. By eliminating HCS from consideration since it cannot (and should not) be used to generate optimal signal plans, the team was left with four viable candidates: SigCinema; Sidra; Transyt and Signal94.

We have further determined that most Signal94 phasing plans generated in the optimization phase were two phases with very low cycle lengths, which cast some doubt on the validity of the model. In addition, an unusual feature of the Signal 94 model output can lead to misunderstandings about intersection delays and LOS. When using the HCM, if a particular lane group is over capacity, no delay estimate is produced for that lane group, its approach, or the intersection. Both HCS and SigCinema reproduce this feature, but Signal94's HCM module does not. In Signal94, if a lane group is overcapacity, its delay is calculated as if the v/c ratio was 0.999, and the resulting delay is used to determin approach and intersection level average delays and LOS. While these occurrances are noted with an @ symbol, the very fact that an intersection level delay and LOS are reported can easily mislead the user.

The remaining models: SigCinema, Sidra and Transyt, were then evaluated on the basis of how well they simulated the field conditions. At the intersection level, Transyt (R^2= 0.91) was superior to both Sidra (R^2= 0.85) and SigCinema (R^2= 0.83). However, at the lane group level, Sidra (R^2= 0.52) outperformed both SigCinema (R^2= 0.45) and Transyt (R^2= 0.40). Considering that it is more relevant to determine which lane group is critical (as opposed to an average intersection performance, which could be satisfactory), more weight was given to the lane group analysis.

On the basis of the results obtained in this study, the team recommends to NCDOT that the use of Transyt-7F for timing analysis of isolated intersections in the state should cease. Because of this, the study team also highly reccomends that NCDOT investigate the accuracy of the Transyt-7F's evaluation of signal systems. Lack of a quality isolated intersection model calls into question the quality of Transyt's system evaluation.

It is further recommended that SIDRA 5.02A (and subsequent releases) be the standard model where such analyses should be conducted statewide. SIDRA's powerful combination of high model usability, superior matching of field delays at the lane group level, versatile optimization features, and graphical user interface makes it the model of choice from this study. Sidra is also unique among the models studied in its ability to incorporate many geometric features of the intersection into the analysis, such as length of turn bays, turning radii, etc. These are very appealing features that can truly integrate Signals and Geometrics. The only known Sidra deficiency compared to SigCinema and Signal94 is the lack of phase plan optimization capability. Further discussion of this issue with the NCDOT project technical committee indicated that phase plans are more or less standardized across the state and, therefore, would not be considered as an option in using signal timing models.

8. Appendix: Map of Data Collection Sites

Not included in this reprint

9. Appendix: Data Collection Site Intersection Schematics

Not included in this reprint

10. Appendix: Example Data Reduction - Site 1 AM Data

Not included in this reprint

11. Appendix: Copy of Data Reduction Methodology (from the Draft 1997 HCM)

Not included in this reprint

12. Appendix: Graphs of Field vs. Model Delay by Lane Group and by Model

Not included in this reprint