APPLYING THE TRANSIMS MODELING PARADIGM
TO THE
SIMULATION AND ANALYSIS OF
TRANSPORTATION AND TRAFFIC CONTROL SYSTEMS

Final Report
KLK231
N07-05

National Institute for Advanced Transportation Technology
University of Idaho

Michael Dixon; Karl Chang
Sanjeev Keecheril; Brent Orton
March 2007
DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.
Applying the TRANSIMS Modeling Paradigm to the Simulation and Analysis of Transportation and Traffic Control Systems

Michael Dixon; Karl Chang; Sajeev Keecheril; Brent Orton

National Institute for Advanced Transportation Technology
University of Idaho
PO Box 440901; 115 Engineering Physics Building
Moscow, ID 838440901

US Department of Transportation
Research and Special Programs Administration
400 7th Street SW
Washington, DC 20509-0001

This report focuses on improving the TRANSIMS transportation planning model. TRANSIMS has been validated to a limited degree for traffic operations due to the lack of readily available standard performance measures and operations data output. This study developed a Traffic Data Extractor Tool (TDET) that provides more useful performance measures than those of TRANSIMS. Validation results for unsignalized intersections concluded that TRANSIMS tends to overestimate control delays and major street left turns are not modeled accurately. For signalized intersections, TRANSIMS performed very well. Compared with the field data, in some cases TRANSIMS surpassed Highway Capacity Software.

Transportation planning models are also reliant upon demographic knowledge associated with traffic analysis zones (TAZs). Census data is an obvious choice for “loading” TAZs with demographic data. These demographic input data are limited since TAZs must either be of a size equal to the smallest census data level or composed of combined elements of the Census data, forming larger TAZs. Part 2 provides a methodology for incorporating remotely sensed data and local land use zoning data to disaggregate the Census block demographics to smaller, user defined TAZs thus providing a higher spatial resolution to the planning model process and increasing model accuracy.

Key Words
Intersections; travel demand; traffic delay; traffic data; traffic simulation; Geographic information systems;

16. Abstract
17. Key Words
18. Distribution Statement
19. Security Classif. (of this report)
20. Security Classif. (of this page)
21. No. of Pages
22. Price

Unclassified

Unclassified

105

...
TABLE OF CONTENTS

LIST OF FIGURES.................................................................................................................V

LIST OF TABLES ....................................................................................................................VII

EXECUTIVE SUMMARY .......................................................................................................1

INTRODUCTION .....................................................................................................................3

DESCRIPTION OF PROBLEM .................................................................................................4

APPROACH .............................................................................................................................6

PART 1: AN IMPROVED TECHNIQUE FOR THE VALIDATION OF TRANSIMS FOR
TRAFFIC OPERATIONS .........................................................................................................8

METHODOLOGY ......................................................................................................................8

OVERVIEW OF TRANSIMS ..................................................................................................8
  Introduction to TRANSIMS ..................................................................................................8
  Purpose of TRANSIMS .......................................................................................................9
  TRANSIMS Architecture ..................................................................................................11

TRANSIMS Output Files.......................................................................................................20

Summary ................................................................................................................................21

METHODOLOGY ...................................................................................................................23

Intersections and Field Data Collection ..............................................................................23

TRANSIMS Modeling ..........................................................................................................23

Traffic Data Extractor Tool (TDET) .....................................................................................25

Experiment Measures.........................................................................................................25
  Field Delay ........................................................................................................................26
  Highway Capacity Software Delay ..................................................................................26
  TRANSIMS Delay ..............................................................................................................29
  TDET Delay ......................................................................................................................30
Applying the TRANSIMS Modeling Paradigm to the Simulation and Analysis of Transportation and Traffic Control Systems
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRANSIMS architecture</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Output from TRANSIMS using the signal.csh script</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>Network intersections</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>White/Styner and SH 8 intersection</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>Intersection of Jackson &amp; A Street</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>TRANSIMS Model for SH8 &amp; White Ave.</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>Nodes-A: Street &amp; Jackson Street</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>Links: A Street &amp; Jackson Street</td>
<td>39</td>
</tr>
<tr>
<td>9</td>
<td>Activity Locations: A Street &amp; Jackson Street</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>Flowchart for TDET</td>
<td>46</td>
</tr>
<tr>
<td>11</td>
<td>Entry Form</td>
<td>47</td>
</tr>
<tr>
<td>12</td>
<td>Instructions form</td>
<td>47</td>
</tr>
<tr>
<td>13</td>
<td>Delay form</td>
<td>48</td>
</tr>
<tr>
<td>14</td>
<td>Link length file</td>
<td>49</td>
</tr>
<tr>
<td>15</td>
<td>Query form</td>
<td>50</td>
</tr>
<tr>
<td>16</td>
<td>TDET output</td>
<td>51</td>
</tr>
<tr>
<td>17</td>
<td>Input form for query</td>
<td>53</td>
</tr>
<tr>
<td>18</td>
<td>Output from the query</td>
<td>53</td>
</tr>
<tr>
<td>19</td>
<td>Manual and TDET demand control delay</td>
<td>58</td>
</tr>
<tr>
<td>20</td>
<td>Manual and TDET stopped delay</td>
<td>59</td>
</tr>
<tr>
<td>21</td>
<td>TRANSIMS and field delay values</td>
<td>62</td>
</tr>
<tr>
<td>22</td>
<td>TRANSIMS and field delay values</td>
<td>62</td>
</tr>
<tr>
<td>23</td>
<td>Predicted values and actual values for minor street left turns</td>
<td>62</td>
</tr>
<tr>
<td>24</td>
<td>Predicted values and actual values for minor street through</td>
<td>65</td>
</tr>
<tr>
<td>25</td>
<td>Predicted values and actual values for minor street right</td>
<td>66</td>
</tr>
<tr>
<td>26</td>
<td>Signalized intersection TRANSIMS and field values</td>
<td>68</td>
</tr>
<tr>
<td>27</td>
<td>TRANSIMS and field delay values</td>
<td>69</td>
</tr>
</tbody>
</table>
Figure 28 Field and TRANSIMS delay values for left turn movements for signalized intersections. ............................................................ 71
Figure 29 Field and TRANSIMS delay values for through movements for signalized intersections. .......................................................... 71
Figure 30 Field and TRANSIMS delay values for right turn movements for signalized intersections. ......................................................... 72
Figure 31 Trajectory of vehicle movement on lane number 2 of the signalized intersection for one cycle length. ........................................... 74
Figure 32 Example for Area-Proportional Estimation............................................. 83
Figure 33 Feature Extraction Error by Experiment for Each Visual Region............... 99
LIST OF TABLES

Table 1 Traffic Control Status. .................................................................21
Table 2 Traffic Volumes at the Intersection of SH 8 & White Avenue ..........33
Table 3 Traffic Volumes at the Intersection of Jackson & A Streets ..........33
Table 4 Field Control Delay at the Intersection of SH 8 & White Avenue ....35
Table 5 Field Queue Delay at the Intersection of A Street & Jackson Street ....35
Table 6 Signal Timing at the Intersection of Jackson & A Streets .............36
Table 7 Required Number of Runs for Each Interval and Movement ..........40
Table 8 Output File Description ...............................................................52
Table 9 Query-Output File Description .....................................................54
Table 10 Vehicle Snapshot Data for Vehicle Number 101848 ..................56
Table 11 Intersection Data for 101848 .....................................................57
Table 12 Mean Delay Values of TRANSIMS and Field for Unsignalized Intersections ..........................63
Table 13 Error Statistics for Unsignalized Intersection ..........................63
Table 14 Movement Volume and Control Delay from Field and TRANSIMS .64
Table 15 Error Statistics by Unsignalized Intersection Turning Movement ....67
Table 16 Field, HCS and TRANSIMS Control Delays and Error Statistics for Unsignalized Intersections. ..................................................................................67
Table 17 Mean Delay Values of TRANSIMS and Field for Unsignalized Intersections. ..........................69
Table 18 Error Statistics for Signalized and Unsignalized Intersections .........70
Table 19 Average Control Delay and Traffic Volumes by Turning Movement from Field and TRANSIMS for Signalized Intersections. ........................................70
Table 20 Error Statistics--Signalized Intersection Turning Movements .......73
Table 21 Signalized Intersection Control Delays and Error Statistics for HCS and TRANSIMS. ..................................................................................73
Table 22 Types and combinations of experimental input data ...................90
Table 23 Accuracy Assessment by Experiment ..........................................94
Table 24 Percentile Ranking of Correctly Classified Percentages Minus Error 97
Table 25 TAZ Household estimation RMSE and MAER Error Rates ..........97
EXECUTIVE SUMMARY

The National Institute of Advanced Transportation Technology (NIATT) at the University of Idaho has completed research into two methods for improving the new transportation planning model, labeled TRANSIMS, developed by the US Department of Transportation and the Los Alamos National Laboratories. TRANSIMS is a transportation planning micro simulation model that can model individual demand responses to changes in transportation network costs. The new model, while providing more sensitive transportation planning information, also requires improved data manipulation and data input. The two parts of this NIATT report address this important requirement and provides improved methodologies for transportation modeling agencies and professionals.

Part 1 of this paper develops a new data extraction tool called Traffic Data Extractor Tool (TDET) for the TRANSIMS model and used actual field data to validate TRANSIMS output for signalized and unsignalized intersections. The study validated that TDET is efficient in extracting the performance measures from TRANSIMS and can be used as a tool to validate TRANSIMS. TDET provides demand stopped delays, supply stopped delays, demand control delays, supply control delays. In addition, it creates a database which includes the vehicle snapshot file and intersection snapshot output files and link file and link length input files from the TRANSIMS in MS Access tables. Data provided by the TDET query form can be used to extract vehicle trajectory data that prove useful for validating additional micro-simulation features in TRANSIMS.

Results of validation tests, it can be concluded that TRANSIMS does not model unsignalized intersections very well, tending to over estimate control delays. For signalized intersections, TRANSIMS performed very well, in some case surpassing Highway Capacity Software (HCS). In fact, TRANSIMS models right turn movements more accurately than HCS. However, care must be taken to work with replicate runs to produce more reliable averages, because individual run results have high variability.
It is recommended from this study that TDET can be used for the validation of TRANSIMS with field or with other transportation software. It is recommended that the modeling concept used in TRANSIMS has to be updated to accommodate major street left turn movements because the router module in TRANSIMS uses the delay at the intersection to determine least-cost routes on which to assign traffic. More validation has to be done to ascertain the efficiency of TRANSIMS in the modeling of signalized intersections, focusing on actuated isolated intersections and actuated coordinated intersections.

Part 2 of this report proposes a methodology for incorporating remotely sensed data as well as local land use zoning data to disaggregate the Census block demographics to smaller, user defined traffic analysis zones (TAZs); thus providing a higher spatial resolution to the planning model process and increasing model accuracy. Procedures fulfilling these two objectives would greatly enhance travel demand modeling efforts for two reasons. First, traffic assignment results would tend to be more accurate [7]. Second, facilitating TAZ size reduction would result in more accurate representation of land use and its access to transportation. A more representative picture of land use variations leads to the improved location of individual travelers, further supporting recent efforts with micro-simulation planning models [8, 9]. Various types of remotely sensed data prototype tested are discussed and recommended for application.
INTRODUCTION

The purpose of this report is to improve transportation engineers and planners ability to utilize the TRANSIMS model recently developed by the US Department of Transportation and the Los Alamos National Laboratories. Two parts of this report address differing aspects of the model usage. The first part develops a Traffic Data Extractor Tool (TDET) that would provide more useful performance measures than those currently provided by TRANSIMS. This part also validates TDET and the utility of using TDET by using it to validate TRANSIMS’ ability to model traffic operations at isolated intersections.

The second part of this report examines new methods for providing employment and household data that are heavily relied on as inputs in the modeling process. In particular, these data describe the location and type of households and the location and type of activities from which travel purpose is derived by aggregate area. Specifically, the intent of this research was to propose and test computational procedures, by which demographic and socioeconomic data can be systematically assigned to Traffic Analysis Zones (TAZ), whose areas are much smaller than the areas to which observed or sampled demographic and socioeconomic data are currently associated. Objectives for these computational procedures are (1) to accurately associate the demographic and socioeconomic data to each TAZ for entire modeled regions and (2) to reduce the human effort required to make these associations. In concept, the proposed procedures are applicable to determining a variety of household and employment socioeconomic variables for TAZs. Nevertheless, this research focused on estimating the number of households in TAZs.

Both Parts 1 and 2 are the result of research funded by the National Institute for Advanced Transportation Technology (NIATT) at the University of Idaho with funds from the U.S. Department of Transportation (DTRS98-G-0027). They represent NIATT’s strong commitment to undergraduate and graduate student education in the classroom and in research. This research has also resulted in high resolution data for local, state, and national uses and improves transportation engineers and planners ability to utilize TRANSIMS for transportation planning.
DESCRIPTION OF PROBLEM

Analysts are unable to analyze many of the issues in transportation planning such as sustainable development, environmental impacts of proposed projects and intelligent transportation systems (ITS), using the traditional deterministic four step transportation planning model. The USDOT and Los Alamos National Laboratories developed TRANSIMS in 1996 as part of the Travel Model Improvement Program (TMIP). TRANSIMS is a transportation planning micro simulation model that can model individual demand responses to changes in transportation network costs. The current dichotomy in transportation engineering, traffic operations on one hand and transportation planning on other, may be well addressed in TRANSIMS, because it models the behavior of individual vehicles and how they interact with each other and traffic control. Some research has been done to determine the capability of TRANSIMS in bridging the gap between traffic operations and transportation planning [1] [2].

To date, TRANSIMS has been validated to a limited degree for traffic operations. This limited validation is primarily due to the difficulty in obtaining industry standard performance measures and operations data. Data that are readily available from TRANSIMS are not directly comparable to measures and data that can be readily collected in the field and included in established traffic analysis models.

The first problem addressed in Part 1 is the development and validation of a traffic data extractor tool (TDET) to provide more useful performance measures than those currently provided by TRANSIMS. Finally, the utility of using TDET was demonstrated by using it to validate TRANSIMS’ ability to model traffic operations at isolated intersections.

The study objectives are to

- Automate reading in raw vehicle data files output by TRANSIMS,
- Determine TRANSIMS generated volumes by turn movement,
- Determine TRANSIMS generated delays by turn movement,
- Create database of TRANSIMS raw vehicle data that enables queries, and
- Validate software capabilities to meet the above objectives.
The problem addressed in Part 2 deals with the fact that transportation planning models rely heavily on employment and household data as inputs in the modeling process. In particular, these data describe the location and type of households and the location and type of activities from which travel purpose is derived by aggregate area. Traditional travel demand models employ these data to describe the trip ends produced and attracted at subdivisions or traffic analysis zones (TAZs) of the modeled region. Trips are then distributed between the TAZs and assigned to the transportation network in a manner usually satisfying or approximating user equilibrium. Given accurate trip distribution, the quality of traffic assignment is directly related to the detail with which the TAZs represent the demographic and socioeconomic characteristics of the modeled region. Horowitz tested traffic assignment results for traditional TAZs and TAZs with subzoning for each link. His test confirmed that traffic assignment is more realistic, but does increase computational demands [7]. Ideally, travel demand models would represent each household, or activity location with their own TAZ. Unfortunately, computational and source data constraints make this ideal condition very difficult to achieve.
APPROACH

The main object of Part 1 has been to develop a Traffic Data Extractor Tool (TDET) that would provide more useful performance measures than those currently provided by TRANSIMS. A validation of TRANSIMS is performed, demonstrating the usefulness of TDET with field data and Highway Capacity Software.

As a part of the development, TRANSIMS architecture, simulation logic, and output file contents were reviewed. Finally, a software tool that can provide the delay measurements and volumes for each movement at an intersection was developed using Visual Basic 6.0, MS Access and MS Excel. TDET’s capability of querying and developing a database allows the user to conduct many traffic operational analyses to further validate TRANSIMS. A validation of TDET, using manually extracted data from TRANSIMS, proved that this tool accomplishes its primary functions of extracting performance measures.

Part 2 examines the use of Graphic Information Systems (GIS) to provide the necessary demographic and socioeconomic data for transportation models. This research examines a methodology for accurately distributing Census data needed for transportation models to higher resolution or smaller TAZs. The method for determining the socioeconomic variables for TAZs on a finer level than Census blocks but using Census demographic data is described herein as the area-proportional TAZ estimation approach. The essence of this technique is to estimate the proportion of a Census block’s quantitative demographic measures belonging to a smaller subsection of the Census block. These subsections become the new, disaggregate TAZs. The quantitative demographic measures are estimated by comparing the sum of the building footprint area of a particular land use within the sub-Census-block TAZs to the total building footprint areas of the same use in TAZs’ encompassing Census block. Quantitative census demographics which can reasonably be correlated to building areas can then be estimated for each TAZ as scaled by the proportion described. Remote sensing techniques for automated feature extraction (AFE) of building footprints from various types of imagery and spatial data constitute the means for determining such building footprint areas and proportions.
The area-proportional TAZ estimation technique is examined, followed by a detailed discussion of the underlying techniques and procedures used to generate the supporting data and inputs. Each section will be concluded with a description of the approach used to validate and quantify error for demographic estimation and input data creation.
PART 1: AN IMPROVED TECHNIQUE FOR THE VALIDATION OF TRANSIMS FOR TRAFFIC OPERATIONS

METHODOLOGY

OVERVIEW OF TRANSIMS

This section first compares TRANSIMS with other simulation software. The architecture of TRANSIMS is then described along with the input data required for the TRANSIMS model. Then a description of how the TRANSIMS micro-simulation model works is given. Finally, the output data generated by TRANSIMS are described.

Introduction to TRANSIMS

Macro-simulation and micro-simulation models are the two primary types of simulation models used in the field of transportation engineering. While the macro-simulation models consider traffic as an aggregate flow using continuum equations, micro-simulation models incorporate specific car-following, vehicle performance and lane changing behavior algorithms to model individual vehicles in the traffic stream. Compared to traditional macroscopic models like Synchro, Transyt-7F, Highway Capacity Software (HCS), and Passer, microscopic simulation models, such as CORSIM, TRANSIMS, VISSIM, and Paramics, are better able to model the inherently complex stochastic and dynamic nature of transportation systems. Many of the micro-simulation models currently available in the transportation engineering field are not capable of bridging the gap between transportation planning and operations

TRANSIMS attempts to address these two subclasses in transportation engineering. The TRANSIMS micro-simulator is a modified car following model: although the vehicle movements occur based on the gap ahead and the current velocity, the leader car’s velocity is not used in the calculations of the follower as in other car following simulations. The TRANSIMS micro-simulation model is based on particle hopping or cellular automata (CA) theory and it models each vehicle and driver in a discrete manner [3].
TRANSIMS combines demand modeling and flow behavior on streets, attempting to describe the whole traffic system behavior within a common analysis environment. It consists of mutually supporting simulations, models, and databases that employ advanced computational and analytical techniques to create an integrated regional transportation system analysis environment. By applying advanced technologies and methods, it simulates the dynamic details that contribute to the complexity inherent in transportation issues.

Purely in terms of network traffic assignment independent of trip end and services (access), TRANSIMS has the potential to more realistic representation of travel demand (TD) service by time-of day (TOD) than other stochastic models. This is because it explicitly models intersection congestion and congestion propagation through time and space.

However, an investigation of intersection congestion and congestion propagation is required to establish an acceptable level of confidence for TRANSIMS and its basic modeling capabilities. This can be accomplished through sensitivity analysis and comparison with industry accepted models, practices, and field data.

For TRANSIMS to fully bridge the gap between planning and operations, it must provide intersection performance measures that are commonly used to assess traffic operations quality, such as control delay or stopped delay. While TRANSIMS output provides travel time, it does not provide the control delay or stopped delay. Consequently, the control delay must be extracted from TRANSIM’s vehicle event output files before they can be compared with the industry accepted models and field delay data. Therefore, to allow more useful operations analysis and comparisons with field data and industry models, a Traffic Data Extractor Tool (TDET) has to be developed to extract approach and stopped delay from the TRANSIMS vehicle event files (vehicle snapshot files).

**Purpose of TRANSIMS**

The traditional quantitative method for transportation planning, the “four-step process,” consists of trip generation, trip distribution, mode choice, and traffic assignment.
The first step, trip generation, estimates the number of trip generated to and from the different zones and trip distribution determines where the generated trips are destined, thus it connects these origins and destinations via traffic streams. Mode choice and traffic assignment determine the modes and routes for these traffic streams. The outcome of using the four-step process is the origin-destination (OD) matrix, which gives the number of trips traveling from one zone to another in a given time step.

Because groups of people are served in the “four-step process,” the four-step process does not track behavior of individuals and makes any analysis based on individual traveler characteristics impossible. For example, when modeling mode choices, individual traveler characteristics such as income or distances to or from a bus stop can affect their decision of whether or not to ride the bus. But it is not possible to include these individual traveler characteristics in the modeling of mode choice in the four step model.

Another disadvantage of the four step process is that it is valid only for time-independent problems. In the trip assignment step, the supply to a link is dependent on the static link performance functions and does not represent the time-dependent queue buildup from bottlenecks or sequential chaining of trips. For example, due to the time dependent queue buildup, the number of trips in a specific path would be reduced. In traffic assignment, no information on demographic characteristics of travelers available; therefore analysis by sub-populations is also not possible.

The disadvantages in the four step process and the enormously important role of individual values and preferences in transportation, lead to an approach to the problem with an agent-based micro-simulation. TRANSIMS maintains the individuality of travelers throughout the modeling process and the location of every traveler is known at any time in this model. In TRANSIMS, travelers are represented as individual “agents” who make independent decisions about their actions, including short-term decisions about acceleration, lane changing, and planning daily activities. As an agent-based simulation model, TRANSIMS has the flexibility to more accurately model the options and choice mechanisms of travel.
TRANSIMS Architecture

Figure 1 shows the architecture of TRANSIMS [4]. The five modules summarize the simulation approach in TRANSIMS as follows: (1) Population synthesizer; (2) Activity generator; (3) Route planner; (4) Traffic micro-simulator, and (5) Emissions estimator.

Figure 1 TRANSIMS architecture.
Each of these modules requires extensive discussion to understand their corresponding methodologies. However, in this research, we are only concerned with the fourth module, Traffic micro-simulator. As a result, the following description of TRANSIMS is limited to the micro-simulator methodology.

**TRANSIMS Micro-Simulator**

The micro-simulation uses route plans from Module 3 as input and simulates the transportation network at a microscopic level of detail. This module simulates the interaction between demand and supply. In this module, the travelers execute the travel plans. Movement of vehicles and the interactions of the travelers in the transportation system of the study area are simulated second by second for a user defined time period. The aggregation of these interactions produce performance measure quantities like volume, flow, link travel time, etc. The micro simulator can produce highly detailed information about individual vehicles, sections of highways, intersections and other elements of the transportation network. A detailed description of TRANSIMS micro-simulation logic is given in the following discussion.

The TRANSIMS traffic highway micro-simulation model is based on hopping or cellular automata (CA) theory [1]. In the cellular automata technique, the road link, the basic entity of traffic simulation, is cut into cells 7.5m in length that can contain zero or one car at any given time. For each lane, an array of cells is modeled. In the cellular automata technique, each vehicle has an integer velocity with values between zero and \( V_{\text{max}} \). For an arbitrary configuration, one update of the system consists of four consecutive steps: acceleration, deceleration, randomization and car motion, which are performed in parallel for all vehicles.

**Vehicle Motion**

Movement of vehicles is performed by jumping one cell to another in a time step where the new position of the vehicle is determined by a set of “driving rules.” The acceleration and deceleration of a vehicle depends on its current speed, and the gap between it and the immediate vehicle ahead in the same lane. If the gap, distance between a vehicle and the leading vehicle, is greater than the distance which would be traveled in one time step then the vehicle tries to accelerate up to a desired speed. This desired speed is limited to the posted speed limit on each
link and the maximum speed for the vehicle type and subtype. The conditions behind the acceleration and deceleration are described in the list below. Keep in mind that TRANSIMS defines gaps as the number of cells to the lead vehicle and speeds as number of cells per time step. For a given time step, gap units are comparable to speed units and as a result gaps and speeds can be compared:

- Deceleration: If the next car ahead is too close (gap < \(v\); where gap is the number of empty cells between a following car and a leading car); then the velocity of the following car is reduced to \(v = \text{gap}\). [3]

- Acceleration: If, however; the gap is sufficiently large (gap \(\geq v+1\)) and if the velocity is smaller than the maximum velocity (\(v < v_{\text{max}}\)) then the velocity is increased by one: \(v = v + 1\). [3]

Keep in mind that TRANSIMS defines gaps as the number of cells to the lead vehicle and speeds as number of cells per time step. For any given time step, gap units are comparable to speed units and as a result gaps and speeds can be compared: The natural fluctuations of driver behavior are modeled using randomization that introduces speed fluctuations for free driving. These reactions result in over-reactions that may occur for decelerations and retardations for acceleration.

A randomization parameter is used to reduce the velocity by 1 (\(P_{\text{noise}}\)) when the velocity is 1 or larger (\(v \geq 1\)). This parameter is the probability that a vehicle ends up slower than the velocity that was calculated using the conditions for acceleration and deceleration. Three effects of the randomization are shown below:

1. **Free Driving:** If there are no other vehicles nearby then the speed of the vehicle fluctuates between \(V_{\text{max}-1}\) and \(V_{\text{max}}\). After the randomization, the velocity of the vehicle will be \(V_{\text{max}-1}\) with \(P_{\text{noise}}\) or \(V_{\text{max}}\) with probability \(1 - P_{\text{noise}}\). The resulting average speed of a freely driving vehicle will be \(V_{\text{max}} - P_{\text{noise}}\). [5]

2. **Over Reactions during Deceleration or Car Following:** If a vehicle ahead is too close, the vehicle itself attempts to adjust its velocity such that it would, in the next time-step, reach...
a position just behind the vehicle ahead. [4]. In other words, assume a vehicle with velocity $V_i$ at a time step $i$, approaches a slow moving vehicle then $V_{i+1}$ will be equal to the gap. But with $P_{\text{noise}}$ the velocity may be one smaller than the gap.

3. Retardation during Acceleration: Essentially, the acceleration is linear, but with $P_{\text{noise}}$ it is possible that no acceleration happens in the current step [4]. Therefore, with $P_{\text{noise}}$ the acceleration pattern of a vehicle starting from zero would be

$$0 \rightarrow 0 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 3 \ldots \ldots \ldots \text{instead of } 0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \ldots \ldots .$$

While $P_{\text{noise}}$ is often set to 0.5, for theoretical work 0.2 is used for realistic traffic modeling. [3].

**Lane Changing**

TRANSIMS has a set of rules to examine each vehicle and determine if it will change lanes during the time step. A vehicle changes lanes for passing a slower vehicle in the current lane or to make turns at intersections to follow its planned route. The neighboring position in the target lane ($x_o$), gap forward to the vehicle in front in the current lane ($G_c$), gap forward to the vehicle in front in the target lane ($G_o$), and the gap backward in the target lane ($G_b$) are used to make a passing lane changing decision. The vehicle will be marked for lane changing if the neighboring position ($x_o$) in the target lane is vacant and the forward gap ($G_o$) on the target lane is larger than the gap on the current lane.

If a vehicle ahead in the current lane prevents the acceleration of the subject vehicle having speed $V$, i.e., $V + 1 > G_c$, then the following conditions should be satisfied for a lane change:

1. The gap forward to the vehicle in front in the target lane (the lane to which the subject vehicle intends to move) is greater than the gap in the current lane, i.e., $G_o > G_c$
2. The gap forward to the vehicle in front in the target lane is large enough to maintain the subject vehicle’s current speed, i.e., $G_o \geq V$.
3. The gap backward in the other lane, the gap between the subject vehicle and the vehicle behind it in the other lane, is greater than or equal to the maximum speed of a vehicle in the simulation, i.e., $G_b \geq V_{\text{Globalmax}}$. 
The lane changing for route plan following is controlled by another parameter (W), which represents the urgency of vehicle to make a lane change based on the plan following as it nears an intersection, along with the above mentioned constraints. This parameter is used to override constraints three or four, depending on the situation as a vehicle approaches an intersection at which it must turn for it to follow its planned route. This parameter is estimated using the following equation:

\[ W = V_{\text{max}} - \frac{(V_{\text{max}} - 1)}{n} \cdot D_{\text{pf}}, \]

where

- \( V_{\text{max}} \) = the maximum speed attainable by the vehicle on that link
- \( n \) = the number of lane changes required to reach the preferred lane
- \( D_{\text{pf}} \) = the set distance from the intersection from where the subject vehicle considers a lane change to follow its route plan. This value will be entered in the configuration file for the entire simulation.
- \( D_i \) = the current distance of the vehicle from the intersection.

The value of W is set to zero initially. When W is zero then it does not influence the lane changing decision. During simulation, the value of W ranges from one to \( V_{\text{max}} \). When the value of W reaches the vehicle velocity (\( W \geq V \)), then the third constraint is relieved. When velocity reaches \( V_{\text{max}} \), then third and fourth constraints are relieved. The vehicle then makes the lane change even if the only one neighboring cell on the target lane is free.

To avoid collisions, TRANSIMS allows lane changes in a certain direction in each time step, left to right, during odd time steps and right to left in even time steps.

**Fixed-Time Signalized Intersections**

In TRANSIMS, a simulated vehicle approaching a signal has the potential to leave the link if its approaching speed is greater than or equal to the remaining number of cells on the link. A vehicle is allowed to enter an intersection if the following conditions are satisfied:
• A vehicle has a speed equal to the number of empty cells between the vehicle and the end of the link.

• A vehicle is the last vehicle on the link in the current lane going towards the intersection.

• A vehicle should satisfy the traffic control at the intersection. The traffic control indicates a permitted or protected movement of a vehicle to enter the intersection. If the vehicle encounters a protected movement, for example a protected left turn movement with a green arrow the vehicle will be allowed to enter the intersection. If it is a permitted movement, for example a permitted left turn movement with a green ball signal then the vehicle checks opposing movements for a gap to enter into an intersection. Other examples of unprotected movements are at yield signs, right turn on red, stop, and on ramp to freeways. All of them work in TRANSIMS as explained earlier. More discussion about the protected and unprotected movements is included in the later part of this section.

• An acceptable gap should be available between the vehicle and conflicting traffic with the right-of-way.

• Once the vehicle is allowed to enter the intersection, the vehicle stays in an intersection queue buffer for the time required to traverse the intersection, instead of passing on to the next link. For a vehicle to proceed into the buffer, the buffer cannot be full. For a vehicle to proceed from the buffer to the outbound link, the destination cell on the outbound link should not be occupied.

Protected Turning and Unprotected Movements

If the vehicle wants to leave the link, the algorithm checks the traffic control to determine if the vehicle can leave the link. If it encounters a protected green or yellow signal, it allows the vehicle to enter the intersection, and if it meets a red light, the vehicle is not allowed to leave the link.

If it is a permitted green signal, the vehicle checks for all opposing vehicles for a gap which is larger or equal to the gap multiplied by a user defined factor known as the Gap Velocity Factor
(GVF). The gap multiplied by the GVF is a minimum time headway, which means that the current gap should be equal to or greater than the GVF times the speed of the oncoming vehicle. In all signalized intersection cases, the vehicle is placed in a queued buffer when it enters an intersection, and it resides in the intersection buffer for a user-specified time before exiting to the destination link. If the traffic control indicates a permitted or protected movement for a vehicle, it enters into the intersection buffer if the buffer is not full. The time spent in the intersection queue buffer is dwelling time in addition to any delay time incurred while waiting for a gap in case of permitted movements. The vehicle will exit from the buffer, after the dwelling time period has expired, to the first cell on the destination link if this cell is vacant and the gap in the opposing traffic stream is acceptable. If the first cell is not vacant, the vehicle waits in the buffer until the cell becomes vacant. If the intersection buffer is full, no vehicle is allowed to enter the buffer. Each intersection has one queue buffer for each incoming lane. The vehicle velocity or speed does not change during this operation; therefore the exiting vehicle will be placed in the destination cell with the same velocity that they entered in the buffer. [5]

**Actuated Signalized Intersections**

TRANSIMS includes one actuated signal control algorithm and uses detectors to provide measurements of vehicle demand in order to model actuated signals. Detectors record vehicle detections and provide a time series of vehicle detections to the signal control algorithm. The signal control algorithm calculates traffic flow characteristics for each phase, including flow and density. Finally, the control algorithm uses a stochastic approach to select the next phase, where the next phase is chosen based on phase probabilities that give preference to phases with a high density but a low flow [1].

The probability of a phase being selected is proportional to the value calculated in Eq. 2:

$$
\pi_p \propto \prod_{m \in M^{(p)}} \frac{1}{D_{(p,m)}} \sum_{q \in D_{(p,m)}} \exp \left( \beta \frac{D^d + \rho_0}{q^d + q_0} \right),
$$

where

- $M^{(p)}$ = the set of movements possible during phase p,
- $D_{(p,m)}$ = the set of detectors associated with movement m of phase p,
\( \beta, \rho_0, q_0 \) are fixed user-defined parameters,

\( \rho^d \) = density over detector \( d \), and

\( q^d \) = flow over detectors over detector \( d \).

The “optimal” parameter values for the actuated signal algorithm were found to be as follows:

[1] \[
\begin{align*}
\beta &= 1.0 \text{ meter per second} \\
\rho_0 &= 0 \text{ per meter} \\
q_0 &= 0.1 \text{ per second}
\end{align*}
\]

The following performance measures are measured by TRANSIMS simulation:

- Flow (veh/hr) :
- Travel time (sec/veh)
- Time stopped (sec/veh)
- Accelerations from stop (#/veh)

The script signal.csh in the directory \$TRANSIMS_HOME/data/gensig/scripts runs to produce these performance measures for each movement at the intersection. [2] The output will have the following format as shown in Figure 2 [1].
Unsignalized Intersections

The movements from the minor streets at two-way stop and yield controlled intersections will go through a “Gap Acceptance” mechanism, as explained in the signalized section in this section, before moving to the destination link. The traffic flow characteristics of unsignalized intersections in TRANSIMS are similar to the signalized intersections except for the following:

- Unsignalized intersections have no intersection queue buffer. Therefore vehicles go right through the intersection if the opposing gap is acceptable [5].
- Vehicles leaving an unsignalized intersection approach go down to their destination link as far as prescribed by their velocity, not just to the first cell of the link [5].

When a simulated vehicle approaches an unsignalized intersection, if the approaching speed is greater than or equal to the remaining number of cells on the link, it has the potential to leave the link. If a vehicle has the potential to leave the link, the algorithm checks for the “traffic control.” The traffic controls implemented in TRANSIMS are no control, yield and stop.

If there is no control, the vehicle moves to the cell that it can reach at its current speed in one time step. If the vehicle encounters a yield sign, the vehicle checks all the opposing lanes for a gap that is larger or equal to GVF times the speed of the opposing vehicle. Once the gap is
accepted, the vehicle moves to the destination cell on its exiting link. If the vehicle encounters a stop sign, the vehicle stops for a minimum of one time step. The vehicle checks the gaps until one is accepted, and the vehicle goes to the first cell of the destination link, with a velocity of one.

**TRANSIMS Output Files**

The output files provided by TRANSIMS include

- **Snapshot Data**: Snapshot data provide three types of information about vehicles and traffic control for each second of simulation. These files contain detailed information about the vehicles, intersections, and traffic control for each time step and they are described below:

- **Vehicle Snapshot file**: This file provides the information about the vehicles traveling on a link for every time step. This output file includes the vehicle identification number, time, link, lane in which the vehicle is traveling, node identification number from which the vehicle entered the link, velocity of the vehicle, and distance from the node at which the vehicle entered the link.

- **Intersection Snapshot Data**: This file provides information about the vehicles in the intersection buffers. This file includes the information of vehicle identification number, time, node identification number, link identification number, lane in which the vehicle entered the intersection, and the vehicle position in the intersection buffer (it can contain a vertical queue).

- **Traffic Control Snapshot data**: This file provides a description of the signal state at a signalized intersection node for each time step. It provides the node identification number, time, link, lane identification number and the type of signal indication. Traffic control data reports the current status of the control at the intersection. It is represented by integers from 0 to 7 as shown in Table 1.1.

- **Summary Data**: There are a variety of summary data files all of which provide summaries of different simulation output. The summary data file of particular interest to this research is the link travel time summary data file and it can be described as follows:

- **Link Travel Time Summary Data File**: This data file provides link and lane identification numbers, identification number of the node from which the vehicles are traveling.
simulation time, the number of vehicles leaving the link, the sum of the vehicle travel times of the vehicles leaving the link, the sum of squares of the travel time, vehicle turning movement when exiting the link, and number of vehicles on the link. It should be noted that the sum of the travel times provided by this file includes the time vehicles spent in the intersection buffer of the upstream intersection.

Table 1 Traffic Control Status.

<table>
<thead>
<tr>
<th>Integer</th>
<th>Status of Traffic Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None</td>
</tr>
<tr>
<td>1</td>
<td>Stop</td>
</tr>
<tr>
<td>2</td>
<td>Yield</td>
</tr>
<tr>
<td>3</td>
<td>Wait</td>
</tr>
<tr>
<td>4</td>
<td>Caution</td>
</tr>
<tr>
<td>5</td>
<td>Permitted</td>
</tr>
<tr>
<td>6</td>
<td>Protected</td>
</tr>
<tr>
<td>7</td>
<td>Permitted after stop</td>
</tr>
</tbody>
</table>

**Summary**

TRANSIMS does not provide traffic control delay, which is a standard performance measure for assessing the level of service of an intersection, but does provides link travel time. In TRANSIMS, the link travel time for vehicles leaving a link includes the time spent in the upstream signalized intersection queue buffer. The difference between the actual travel time and the free-flow travel time for all vehicles on a link can be calculated to get the delay on that link [4]. Actual travel time is provided by the Summary File, and free flow travel time can be obtained from link data by dividing the length of the link by the Free Flow Speed and length of the link. It should be keep in mind that the estimation of delay by this method will not be equal to control delay because the actual and free-flow travel time does not include the acceleration time from the stop bar at an intersection. Also the actual travel time contains buffer time while the free flow travel time does not include it.
According to the *Highway Capacity Manual (HCM)*, the control delay at an intersection is equal to the stopped delay multiplied by 1.3 [6]. Control delay includes the initial deceleration delay, queue move-up time, stopped delay and final acceleration delay. The delay provided by TRANSIMS can be called link delay, which includes time spent in the upstream intersection queue buffer, the delay on the link due to interaction with other vehicles before reaching the intersection, deceleration delay before joining the queue, queue move-up time and stopped delay. As the length of the link increases, the difference between the delay from TRANSIMS and the *HCM* control delay at an intersection may increase.

Because traffic engineers require control delay for the assessment of intersection operations, it is necessary to develop a tool to extract control delay from TRANSIMS. TRANSIMS provides the vehicle movement details in the Vehicle Snapshot file for each time step, and control delay can be measured using this data. This research focuses on developing a software delay extraction tool and researching the capability of TRANSIMS to model intersection operations. By understanding TRANSIMS’ ability to model intersection operations, the understanding of how well TRANSIMS bridges the gap between planning and operations will significantly increase.
Methodology

As per current industry standards, control delay and stop delay are the performance measures used to measure the level of service of an intersection. Although TRANSIMS provides link travel time, it does not provide control delay or stop delay. Therefore it is necessary to estimate the control delay and stop delay from the raw vehicle data files provided by the TRANSIMS output. A Traffic Data Extraction Tool (TDET) was developed. TDET capabilities were validated by comparing TDET results to results obtained by manually extracting information from TRANSIMS. Control delay and stop delay from the Field, from TRANSIMS, from Highway Capacity Software (HCS) and from TDET are compared to validate the TRANSIMS model utility in traffic operations. This section discusses the methodology of the study with a brief description of the intersections used for testing and analysis, development of the TDET and the method for obtaining the delay measure from field data, TRANSIMS, HCS and TDET.

Intersections and Field Data Collection

A fixed time signalized intersection and a two-way stop controlled intersection were used in this analysis. While the fixed time signalized intersection is used to research the capability of TRANSIMS to model signalized intersections, the two-way stop controlled intersection is used to determine how well TRANSIMS models gap acceptance and two-way stop controlled intersection operations. The study of both signalized and unsignalized intersections is necessary to demonstrate TDET and to further validate TRANSIMS. It should be noted that TDET can be used to extract data for any portion of a TRANSIMS network and is not limited to intersections.

TRANSIMS Modeling

This research used the TRANSIMS network developed by NIATT for the City of Moscow. To streamline the model runs, the unsignalized and signalized intersections were analyzed using different network files. In the case of the analysis of the unsignalized intersection, the network was comprised of only one intersection. For the signalized intersection analysis, a larger network was used to better model the effect of nearby signalized intersections. Only one signalized intersection, the intersection of D and Main Streets, affects vehicle arrivals at A and Jackson.
Streets. As a result, the analysis network for the signalized intersection was expanded to include both intersections. However, data were only extracted for A Street and Jackson. Figure 3 shows the location of these adjacent intersections.

![Network intersections](image)

Figure 3 Network intersections.

An existing model of Moscow was used that included all of the signalized intersections in the city. However, to reduce the computational burden and to simplify the O-D matrix input into TRANSIMS, travel demand was limited to those OD pairs that would contribute to turning movements at the two intersections included in the analysis network. The O-D matrix volumes were modified to reproduce intersection turning movement volumes at A Street and Jackson to within +/- 99 percent of the observed volume. Twelve sample runs were made to find the number of runs required to get a 90 percent confidence interval. Eighty-eight sample runs were required
to get the 90 percent confidence interval and were made for the analysis. A detailed discussion of TRANSIMS modeling is included. The output files, vehicle snapshot file, summary file, signal snapshot file and intersection snapshot file were yielded by TRANSIMS.

**Traffic Data Extractor Tool (TDET)**

TDET was developed using Visual Basic 6.0 to get the control delay and stopped delay from TRANSIMS. This program provides a database containing all of the vehicle snapshot data, stopped delay and approach delay performance measures for direct comparison with field data and indirect comparison with traffic models. In addition, this program provides an efficient way to query the database by time, node and distance from intersection. This database query feature facilitates the efficient extraction of vehicle events by points and times of interest.

TDET uses the link data, node data, vehicle snapshot data, intersection snapshot data, signal data and a link length file created by the user. In the link length file, the user can input the distance from the intersection stop bar in which the delay will be calculated.

The user defines the start time, end time and interval size by which to summarize the output. For example, if the user has three hours of data and wants to get delay during one hour, then the user can enter the start time and end time for that one-hour duration. Interval size defines the lengths of the intervals used to summarize the delay in that one hour period. The output from this program is produced as an Excel file.

**Experiment Measures**

Delay measures were obtained using four methods. This section will describe how we obtained delay measures. Four methods were used:

1. Field delay (stopped delay)
2. TRANSIMS summary output based delay (link delay)
3. TDET extracted delay (stopped delay and approach delay)
4. HCS based HCM delay (control delay)
Field Delay
In this research, time stamps of vehicles joining the queue and departing the stop bar were extracted from the video tapes using TRAFFIC TRACKER. For each movement, the difference between these time stamps was used to calculate delay for each vehicle. Total queue delay for each movement for a 15-minute time period was calculated by adding up the delay of all vehicles departed from the stop bar in that particular interval. Average queue delay for each movement was estimated by dividing the total delay with the total number of vehicles.

Control delay for signalized intersections was estimated from queue delay using the methodology discussed in Appendix A of the HCM. According to the HCM methodology, an acceleration/deceleration correction delay should be added with queue delay to obtain the control delay at a signalized intersection. The acceleration/deceleration correction delay was obtained by using the following equation:

$$d_{ad} = FVS \times CF$$

where

$$d_{ad} = \text{Acceleration/Deceleration correction delay},$$

$$FVS = \frac{V_{stop}}{V_{tot}}$$

$$V_{stop} = \text{Number of vehicles stopped in 15-minute time period}$$

$$V_{tot} = \text{Total number of vehicles arrived in 15-minute time period},$$

$$CF = \text{Acceleration/deceleration correction factor [6].}$$

For unsignalized intersections, a constant value of five sec/veh was added to the queue delay to account for the delay due to deceleration of vehicles from free-flow speed to join the queue and acceleration of vehicles from stop bar to free-flow speed [6]. Stopped delay for signalized and unsignalized intersections was estimated by dividing the control delay by 1.3.

Highway Capacity Software Delay
HCS is a deterministic model based on HCM methodology to estimate intersection control delay. The HCM describes two different methodologies to estimate control delay for stop controlled
intersections and signalized intersections. The manual defines the control delay at a two-way stop controlled intersection as the total elapsed time from the time a vehicle stops at the end of the queue to the time the vehicle departs from the stop bar. The HCM uses the following equation to estimate the control delay for movements at a two-way stop controlled intersection:

\[
d = \frac{3600}{C_{m,x}} + 900T \left[ \frac{V_x}{C_{m,x}} - 1 + \sqrt{\left( \frac{V_x}{C_{m,x}} - 1 \right)^2 + \frac{3600}{450T} \left( \frac{V_x}{C_{m,x}} \right)} \right] + 5,
\]

where

\[
d = \text{control Delay (s/veh)},
\]

\[
V_x = \text{flow rate for the movement x (veh/hr)},
\]

\[
C_{m,x} = \text{capacity of the movement x (veh/hr), and}
\]

\[
T = \text{analysis time period (h)}.
\]

The constant five sec/veh is used to account for the deceleration and acceleration of vehicles before joining the queue and after leaving the stop bar.

For signalized intersections, the HCM methodology uses the following equation to find the control delay.

\[
d = d_1 (PF) + d_2 + d_3,
\]

where

\[
d = \text{control delay per vehicle (sec/veh)},
\]

\[
d_1 = \text{uniform control delay assuming uniform arrival of vehicles (s/veh)},
\]

\[
PF = \text{progression adjustment factor},
\]

\[
d_2 = \text{incremental delay to account for random arrivals and oversaturated queues (sec/veh)}
\]

\[
d_3 = \text{initial queue delay (sec/veh)}.
\]

Uniform delay \(d_1\) is the delay estimate for uniform arrivals of vehicles and stable flow. It is estimated by using Eq.6:
Applying the TRANSIMS Modeling Paradigm to the Simulation and Analysis of Transportation and Traffic Control Systems

\[
d_t = \frac{0.5C \left( 1 - \frac{g}{C} \right)^2}{1 - \left( \min(1, x) \frac{g}{C} \right)},
\]

(6)

where

- \( C \) = Cycle length (sec),
- \( g \) = effective green time (sec), and
- \( x = \frac{v}{c} \) ratio or degree of saturation for lane group.

Progression Adjustment Factor (PF) indicates the proportion of vehicles arriving on green at the signal and is computed using Eq. 7:

\[
PF = \frac{(1 - P) f_{PA}}{1 - \left( \frac{g}{C} \right)},
\]

(7)

where

- \( P \) = proportion of vehicles arriving on green,
- \( \frac{g}{C} \) = proportion of green time available, and
- \( f_{PA} \) = supplemental adjustment factor for when the platoon arrives during green.

Incremental delay is the delay estimate due to non-uniform arrivals and temporary cycle failures as well as those caused by sustained periods of over saturation. The incremental delay is estimated using Eq. 8:

\[
d_2 = 900T \left[ X - 1 + \sqrt{(X - 1)^2 + \frac{8kIX}{cT}} \right],
\]

(8)

where

- \( T \) = duration of analysis period (hr),
- \( k \) = incremental delay factor that is dependent on controller settings,
\[ l = \text{upstream filtering/metering adjustment factor}, \]
\[ c = \text{lane group capacity (veh/h), and} \]
\[ X = \text{lane group v/c ratio or the degree of saturation}. \]

Initial Queue delay is due to the delay caused by the residual queue, which has not been serviced by the intersection in the previous cycle. This can be estimated by using the following Eq. 9 (HCM):

\[ d_s = \frac{1800Q_b(1 + u)t}{cT}, \tag{9} \]

where

\[ Q_b = \text{initial queue at the start of period T (veh)}, \]
\[ c = \text{lane group capacity (veh/h)}, \]
\[ T = \text{duration of analysis period (hr)}, \]
\[ t = \text{duration of unmet demand in analysis period T (hr)}, \] and
\[ u = \text{delay parameter} \]

The control delay for this research is calculated using the HCS, which is based on the current HCM methodology.

**TRANSIMS Delay**

TRANSIMS does not provide control delay, but it provides the link travel time. The delay nearest to the control delay can be obtained by finding the difference between the link travel time and the free flow travel time, which would be the approach delay. The real travel time on a link includes the time taken by the vehicle to traverse the upstream intersection. The delay obtained by finding this difference is not consistent with the field or HCM control delay values. Even though this delay and others are inconsistent, comparison of these delay values will provide valuable insight into the intersection modeling capabilities of TRANSIMS, keeping in mind their differences.

In this research the link travel time summary data and the link data obtained from TRANSIMS is used to get delay. The summary file provided the real travel time on the link and the link data
provides the link length and free flow speed on the link. The link delay is obtained by the following equation:

\[
\text{Link Delay} = \text{Real Travel Time} - \left(\frac{\text{Link Length}}{\text{Free Flow Speed}}\right).
\]

**TDET Delay**

TDET estimates the approach and stopped delay for a movement for a user specified distance from the intersection stop bar and time interval. TDET estimates the delay by adding, the difference between the time taken by the vehicle to travel from the user defined distance to the next link and the free flow travel time and the time spent in the intersection buffer, together for a particular movement. This delay will be equal to control delay if the vehicle has accelerated by the time that it gets through the intersection (buffer zone). From this point forward in this thesis, the delay obtained from TDET will be labeled as control delay.

**Field Data Collection**

Geometry, traffic control, traffic volume, time of arrival and departure of vehicles and other signal parameters such as phasing, green time, yellow time and red time of each phase at the signalized intersection were collected from the field. The following sections of this section discuss how the data were collected from the field and extracted from the field data.

Two intersections in Moscow, Idaho, were considered for the analysis. The intersection of SH8 and White Avenue was selected as the two-way stop-controlled intersection. SH 8 is the major street and White Avenue is the stop-controlled minor street. The fixed time signalized intersection of A and Jackson Streets was selected for the signalized intersection. The selection criteria for the intersections were

- Low pedestrian Traffic.
- TRANSIMS does not model pedestrian traffic, a major factor that influences the traffic operation of an intersection. In order to avoid the influence of pedestrian traffic in the delay, intersections having a very small number of pedestrian traffic were selected for the analysis,
- Level terrains with zero grades, because grading affects the delay at an intersection,
- Good sight lines for videotaping the intersections, and
- Sufficiently complex to allow a full testing of the TRANSIMS simulation logic.

Video cameras were used at these intersections to record the traffic data. The recorded data were extracted from the video camera using the Traffic Tracker tool. The methodology used to measure field delay is discussed in the experiment measures section.

**Intersections Characteristics**
Intersection geometries and camera placements at SH8/White Avenue and A/Jackson Streets intersections are shown in Figures 4 and 5. Two video cameras at each intersection were deployed to provide continuous coverage of the intersection for the peak two-hour period, with one camera’s surveillance region overlapping the next.

![Figure 4 White/Styner and SH 8 intersection.](image)

A Street/ Jackson Street is a two phase, fixed time signalized intersection with a permitted left turn for westbound left turn traffic.
Field Data Collection
Traffic volume and delay data were collected in July 2004 at the intersections of White Avenue and SH8 and Jackson Street and A Street using video cameras. The following data were collected at these intersections:

- Traffic volume
- Time of arrival and departure of the vehicles
- Signal parameters (phasing, Green, Yellow and Red time for each phase)

Traffic Volume
Traffic volumes, the time stamp of vehicles joining the queue, and the time stamp of vehicles leaving the stop bar (the queue) were extracted from the video tapes using Traffic Tracker and were processed in 15-minute time intervals. Traffic volumes extracted from the videotapes for each intersection are shown in Tables 2 and 3.
Table 2 illustrates that the north and southbound approaches on SH 8 have higher volumes than the east and westbound approaches. In addition, the eastbound right turn and westbound right turn have higher traffic volumes than the other minor street movements.

Table 3 Traffic Volumes at the Intersection of Jackson & A Streets
Table 3 shows that southbound traffic on Jackson Street, a one-way street, has the highest volume compared to the westbound and eastbound traffic, in all the intervals. The right turn movement on southbound and eastbound approaches service a fair amount of traffic and the westbound traffic on A Street carries the lowest volume at this intersection.

**Field Control Delay**

The Traffic Tracker tool was used to get the time stamp of a vehicle joining the queue and departing from the stop bar. The difference between time stamps is the queue delay experienced by the vehicle. For example, if a vehicle joins the queue at 4:30:12 and leaves the stop bar at 4:30:35, the queue delay for that vehicle is 23 sec. The estimation of control delay from this queue delay is explained in the “Methodology” subsection above. For a time interval, the average control delay (veh/sec) is estimated by dividing the sum of all vehicle delays by the number of vehicles. In this research, the average control delay was calculated for 15-minute intervals. The time stamps of vehicles entering the queue and departing from the stop bar and the delay of each vehicle are included on a CD as part of this thesis. The field control delay at SH8 and White Avenue intersection and field queue delay at the intersection of A Street and Jackson Street are given below in Tables 4 and 5 respectively.

From Table 4, it can be concluded that the minor street movements, eastbound left, eastbound through and westbound through movements, have higher delays. Eastbound through movement has the highest delay of 97 seconds in the fourth interval. Even though the eastbound right turn has higher volume than the westbound right turn, the control delay is found to be smaller. This is due to the heavy traffic volume on southbound movement, which reduces the number of acceptable gaps for westbound right turn.
Table 4 Field Control Delay at the Intersection of SH 8 & White Avenue

<table>
<thead>
<tr>
<th>Movement</th>
<th>15-minute time Interval Delay (veh/sec)</th>
<th>4:23:00</th>
<th>4:38:00</th>
<th>4:53:00</th>
<th>5:08:00</th>
<th>5:33:00</th>
<th>5:48:00</th>
<th>6:03:00</th>
<th>6:18:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBLT</td>
<td></td>
<td>12</td>
<td>36</td>
<td>41</td>
<td>54</td>
<td>28</td>
<td>17</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>EBTH</td>
<td></td>
<td>29</td>
<td>27</td>
<td>20</td>
<td>97</td>
<td>15</td>
<td>17</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>EBRT</td>
<td></td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>33</td>
<td>14</td>
<td>12</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>WBLT</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WBTH</td>
<td></td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>27</td>
<td>26</td>
<td>7</td>
<td>19</td>
<td>45</td>
</tr>
<tr>
<td>WBRT</td>
<td></td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>NBLT</td>
<td></td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>NBTH</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NBRT</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SBLT</td>
<td></td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>SBTH</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SBRT</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5 Field Queue Delay at the Intersection of A Street & Jackson Street

<table>
<thead>
<tr>
<th>Movement</th>
<th>15-minute time Interval Delay (veh/sec)</th>
<th>4:00:00</th>
<th>4:15:00</th>
<th>4:30:00</th>
<th>4:45:00</th>
<th>5:00:00</th>
<th>5:15:00</th>
<th>5:30:00</th>
<th>5:45:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBTH</td>
<td></td>
<td>22</td>
<td>15</td>
<td>18</td>
<td>24</td>
<td>18</td>
<td>22</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>EBRT</td>
<td></td>
<td>8</td>
<td>5</td>
<td>14</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>WBLT</td>
<td></td>
<td>11</td>
<td>14</td>
<td>17</td>
<td>11</td>
<td>23</td>
<td>9</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>WBTH</td>
<td></td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SBLT</td>
<td></td>
<td>4</td>
<td>20</td>
<td>0</td>
<td>9</td>
<td>23</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SBTH</td>
<td></td>
<td>30</td>
<td>27</td>
<td>30</td>
<td>29</td>
<td>26</td>
<td>26</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>SBRT</td>
<td></td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5 shows that the southbound through movement on Jackson Street and eastbound through movement on A Street has higher delays. As shown in Table 5, these movements carry a larger traffic volume to increase the delay.
Signal Timing

Signal timing for the fixed-time signal at the intersection of A and Jackson Streets was collected from the field and is given below in Table 6.

Table 6 Signal Timing at the Intersection of Jackson & A Streets

<table>
<thead>
<tr>
<th>Movement</th>
<th>Green Time (sec)</th>
<th>Yellow Time (sec)</th>
<th>Red Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EB/WB</td>
<td>30.0</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>SB</td>
<td>40.0</td>
<td>3.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

This intersection has two phases. Jackson Street serves the maximum number of vehicles; therefore it has a higher green time than the eastbound and westbound movements.

Summary

The field data described in this section were used in TRANSIMS and in the HCM delay equations. Field volume was used for calibration purposes in TRANSIMS and field delay was used for comparison purposes.

TRANSIMS Modeling

This section explains how the intersections were modeled using TRANSIMS. Unsignalized and signalized intersections were modeled in TRANSIMS in two different network files. The following steps were involved in the modeling:

- Creating network nodes, links and activity locations.
- Inputting lane geometry and traffic control details.
- Inputting O-D volumes.
- Calibration/validation for the traffic volumes at the intersection.
- Experiment simulation runs for testing TDET software and TRANSIMS outputs

The following section contains a detailed discussion of the modeling of these two intersections.
Modeling of Intersections

The unsignalized intersection of SH 8 & White Avenue was represented by a network of links and nodes in TRANSIMS, as shown in Figure 6. The node numbers 1, 2, 3, 4, and 5 were connected using links 12, 13, 14, and 15. Node one represents the intersection of the two streets. The remaining links are dummy links required to load the network from the four activity locations A2, A3, A4, and A5. Dummy links are necessary in this instance because TRANSIMS output summary files will not include vehicles that do not travel the entire link length. If the activity location were located on links 12, 13, 14, and 15, no vehicles would be included in the measures used in the TRANSIMS summary file. Each activity location represents an origin and/or destination for the OD volumes entered into TRANSIMS, where these OD volumes closely reproduce the field turning movement volumes.

Figure 6 TRANSIMS Model for SH8 & White Ave.
The signalized intersection of A Street and Jackson Street was modeled in TRANSIMS using links and nodes as shown Figures 7 and 8. Node 276 represents the subject intersection and the node 190 is the upstream intersection of D Street and US 95, which affects the operation of the subject intersection. To be consistent with the field conditions, this upstream intersection of D Street and US 95 was included in the model. Activity locations 401, 402, 1101, 502, 503, 601, 602 and 603 that generate traffic volumes are shown in Figure 9.

Figure 7 Nodes-A: Street & Jackson Street.
Figure 8 Links: A Street & Jackson Street.

Figure 9 Activity Locations: A Street & Jackson Street.
Determining sample size

Number of runs required to get a 90 percent confidence interval was estimated by Eq. 11:

\[ N = \left( \frac{t_{\alpha/2}}{\mu e} \right)^2, \]  

(11)

where

- \( \mu \) = mean
- \( d \) = standard deviation
- \( e \) = allowable error specified as a fraction of the mean, and
- \( t_{\alpha/2} \) = the critical value of the t-distribution at the confidence interval of 1-\( \alpha \).

For this research, the level of significance was assumed to be \( \alpha = 0.10 \), resulting in a 90 percent confidence interval. To determine the number of replicate simulation runs, a single TRANSIMS run was made for a three-hour period. The three-hour period was broken into 15-minute time intervals to get 12 observations. The turning movement approach delays obtained from those 12 observations were used to get the required number of runs. Table 7 shows the number of runs required for each interval. The number of runs required was found given a 90 percent confidence interval and error of the mean (e) of 10 percent. The following table shows the number of runs required for achieving a 90 percent confidence interval.

**Table 7 Required Number of Runs for Each Interval and Movement**

<table>
<thead>
<tr>
<th>Node From</th>
<th>Movement</th>
<th>Volume</th>
<th>Control Delay/veh</th>
<th>Number of Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stdev</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>RIGHT</td>
<td>40</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.38</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>THROUGH</td>
<td>8</td>
<td>44.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>83.49</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>LEFT</td>
<td>1</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>464.06</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>THROUGH</td>
<td>6</td>
<td>36.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.17</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>RIGHT</td>
<td>31</td>
<td>18.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.68</td>
</tr>
</tbody>
</table>

As shown in the Table 7, the left turn movement from node number 5 to node number 1 required the maximum number of runs—464—to get a 90 percent confidence interval equivalent to +/- 10
percent of the mean. However, the traffic volume is one in that interval. The next highest number of 83 runs was used in this research. Although the minimum number of runs was 83 runs, to be conservative at least 88 runs were made for each intersection.

**Traffic Data Extraction Tool (TDET)**
TRANSIMS does not provide control delay or stopped delay as output. Instead, it provides a link travel time summary data file, which provides the sum of the vehicle travel times for the vehicles leaving the link. TRANSIMS snapshot data provides detailed information about simulation at a point in time, which collectively can give a complete trajectory of each vehicle in simulation. Therefore, the snapshot data file along with the TRANSIMS input data files, such as the link data file and the node data file, can be used to acquire more accurate values for delay than the TRANSIMS summary output file. In addition, the delay values obtained can be extracted in a manner that is consistent with how field delay data were collected.

The purpose of this research was to develop a software tool to extract approach delay and stopped delay from TRANSIMS simulation output files. A secondary objective was to be able to query data from the database created by the program for delay to get the information describing the operations of an intersection for a particular interval. The following sections discuss the development of the TDET including the methodology, instructions for using it, and its advantages over the TRANSIMS summary output.

This program was developed using Visual Basic 6.0 and MS Access database. The output of the program is a populated MS Access database and an Excel file containing aggregated data such as traffic volumes and average delays.

**Methodology**
TDET has two modules. One module develops and populates a database from TRANSIMS output, and the other estimates delay by querying the database developed by the first module. The delay is estimated using the following files:

- Node
- Link
- Vehicle-snapshot
- Intersection-snapshot, and
- Link length.

For a vehicle, the total delay is the sum of link delay and the delay experienced in the buffer zone. The difference between the actual travel time experienced by the vehicle and the free flow travel time is the link delay. Free flow travel time is calculated using the free flow speed from the link file and the link length from the link length file.

A vehicle’s link travel time is found by using its trajectory, which is developed using the vehicle-snapshot file, node file, and link file. The intersection snapshot file is used to get the time spent by a vehicle in the intersection in the case of signalized intersections. It should be noted that the time taken by the vehicle to accelerate from the stop bar included in this delay. This delay is called control delay.

TDET estimates the demand delay, supply delay, and stopped delay. Demand delay is delay experienced by vehicles that arrived in the given interval until their demand for service has been met. In the case of demand delay, if a vehicle arrived during interval t then its corresponding delay is aggregated with all of the other vehicles that arrived during interval t. Supply delay is delay experienced by vehicles that departed in the given interval, since they arrived. In the case of supply delay, if a vehicle departed during interval t, its corresponding delay is aggregated with all of the other vehicles that arrived during interval t. TDET estimates the stopped delay by counting the number of vehicles with zero velocity at regular time steps and using the ITE stopped delay procedure to calculate stopped delay.

The second module of the program uses the output from the first module, a database that has the following tables:
- Vehicle-snapshot table,
- Intersection-snapshot table,
- Link table,
Steps involved in the development of this program are outlined below.

Algorithm: Estimation of Approach Delay and Stopped Delay

Step 1: Get text file paths and names

- Input:
  - Link Data File
  - Node Data File
  - Link Length File
  - Vehicle Snapshot File
  - Intersection Snapshot File
  - Duration
  - Interval

Step 2: Read Network Files

- Read Link Data File and Node Data File; find out the type of movement from one link to another using the easting and northing of the nodes; store the results in an array.

Step 3: Create Database Tables

- Create a database to store the files.

Step 4: Read Link Length File

- Read the link length file and store in a table as well as in an array.

Step 5: Read Vehicle Snapshot File

- Read vehicle snapshot data and store into a table as well as in an array.

Step 6: Read Intersection Data File

- Read intersection data and store into a table as well as in an array.

Step 7: Link Traveltime Estimation

- Extract the travel time for each link in the link length file using the following equation:

\[ R_x = t(x) - t(x+1) \]  

(12)
where

\[ R_x = \text{travel time for a length of } x \text{ from the intersection stop bar}, \]

\[ t_{(x)} = \text{time when the vehicle is first seen on the link at a distance } x \text{ from the intersection stop bar. If the vehicle does not have a recorded event at the specified location, extrapolate to estimate the time.} \]

\[ t_{(x+1)} = \text{time when the vehicle traverses the stop bar. Again, if the vehicle does not have a recorded event at the stop bar, extrapolate to estimate the time.} \]

Step 8: Estimation of Free-Flow travel time

\[ \Rightarrow \text{Estimate the free flow time for each link in the link length file using the length } (x) \text{ and free flow speed. The free flow travel time is calculated using the Eq. 13:} \]

\[ FR_{(x)} = l_{(x)}/ FV_{(x)}, \]  

(13)

where

\[ FR_{(x)} = \text{freeflow travel time over the distance } x \text{ to the stop bar}, \]

\[ l_{(x)} = \text{length of } x \text{ from the intersection stop bar on the link}, \]

\[ FV_{(x)} = \text{free-flow speed of the link}. \]

Step 9: Extracting Control Delay and Stopped Delay

\[ \Rightarrow \text{Equations 14 and 15 define how delay is calculated in both cases.} \]

\[ D = R(x) - FR(x), \]  

(14)

where

\[ Rx = \text{travel time for a length of } x \text{ from the intersection stop bar and} \]

\[ FR(x) = \text{freeflow travel time over the distance } x \text{ to the stop bar,} \]

and

\[ D_{\text{stop}} = \sum t_{(x)} , \]  

(15)

where

\[ t_{x} = \text{timestamps with zero velocity for a length of } x \text{ from the intersection stop bar.} \]

Step 10: Provide interval index and estimate supply delay and demand delay

\[ \Rightarrow \text{Calculate the interval indices for the arrival time and departure time. Sum up the delays experienced by the vehicles arrived in a particular interval using the interval index for the arrival time to aggregate the demand delay. Sum up delays} \]
experienced by the vehicles departed in the particular interval using the interval index for departure time to aggregate the supply delay.

Step 11: Acquire Simulation Output of Intersection Buffer Delay

→ Acquire the time spent in the intersection buffer in case of signalized intersection. This information is extracted from the intersection snapshots file.

Step 12: Acquire Total Delay

\[ D^{\text{TOTAL}} = D + \text{Intersection Buffer Delay} \]  

(16)

Step 13: Output

→ Provide output in an Excel sheet
→ Provide all the data in tables in a database.

The flowchart for TDET is illustrated in Figure 10.
User Interface and Input Files

By clicking the “enter” button as shown in Figure 11, the user can enter into the program, which leads to the form shown in Figure 12. This form has three tabs: Directions, Control Delay, and Intersection Details.

The Directions tab (Figure 12) gives the instructions to the user for using the software. The Control Delay tab allows the user to input the files necessary for the calculation of approach...
delay (Figure 13); and the Intersection Details tab allows the user to enter the data for querying the database (Figure 15).

![TRANSIMS DATA EXTRACTOR](image)

**Figure 11 Entry Form.**

Following are the steps involved in running the software:

**Step 1.** Select 'Run' tab
- Load TRANSIM output file name 'Link' by clicking the 'Load Link data File' button.
- Load TRANSIM output file name 'Node' by clicking the 'Node Data File' button.
- Click the 'Link length' button to load the file which contains the lengths of the links to be considered for the estimation of delay.

For example, let us assume a link 501 has a link length of 1200m. However, the user wants to get the delay for a distance of the 50cm from the upstream intersection. Therefore user has to make a test file which has a format as given below:

<table>
<thead>
<tr>
<th>Columns represent linkname, from node, to node and linklength. It should be noted that the one link will be having two link lengths. The file can be created using 'link' output file from TRANSIMS using an excel sheet.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Load 'transmapshot' TRANSIMS output files. This file will have a &quot;*txt&quot; extension.</td>
</tr>
<tr>
<td>- Load 'Intersection Snapshot' TRANSIMS output files. This file will have a &quot;*int&quot; extension.</td>
</tr>
<tr>
<td>- Input the start time, end time and interval. For example, if the user has three hours data and wants to get delay for one hour data, then user has to enter the start time and end time for that one hour duration. Interval time gives the delay for an interval such as 15 min delay in that one hour.</td>
</tr>
</tbody>
</table>

**Step 3.** Click 'run' to run the program.

**Step 4.** While each file is processed, the program will ask the user to input the output file name and location. The files will be saved as excel files.

Also the input data will be stored in MS Access database named 'transmap (run no)'.

**Figure 12 Instructions form.**
The input files required for the estimation of approach delay and stopped delay are as follows:

- Link File
- Node File
- Link Length File
- Vehicle Snapshot File, and
- Intersection Snapshot File.

Link and Node data files are the network data files input into TRANSIMS. The link data file contains all the data that specifies the property of the links, which include the link name, length of the link, connecting node numbers, speed limit, free-flow speed and capacity. The node file describes the nodes with the node number, coordinates, and elevation.

![Figure 13 Delay form.](image-url)
The link length data file is a file that has to be created by the user as a text file. It allows the user to specify the distance \((x)\) from the intersection stop bar at which delay will begin to be measured. For example, if the link length is 1000 meters and the user wants delay from 50 meters upstream of the stop bar up to the stop bar then the user should specify the distance as 50 meters. The link length file has the following fields: link number, from node, to node, and the length of the link that used to measure the delay. It should be noted that the user has to specify the link distance in both the directions. The format of this file that has to be followed is given below. The link name is in the first column, while the second, third, and fourth columns contain the from-node, to-node, and the link length, respectively.

![Link length file](image)

Vehicle and intersection snapshot data files are output by TRANSIMS. The formats of these files were discussed earlier.

Other inputs required from the user are the start time, end time and interval size, as shown in Figure 13. Start time and end time are the starting time and ending time of the data to be considered for the delay results. For example, let us assume that the user has vehicle and intersection snapshot files for a three hour period from 7:30 to 10:30 and needs delay from 8:30 to 9:30. Then the start time should be input as 8:30 and the end time as 9:30. Alternatively, if the user wants data covering the entire three-hour period then the start time would be 7:30 and the
end time is 10:30. The interval size text box allows the user to specify the time interval size over which data will be aggregated. For example, if the user wants delay for 15-min time intervals from a one-hour period then a value of 15 should be entered. The result is an output of aggregated data in four intervals, from 7:30-7:45, 7:45-8:00, 8:00-8:15 and 8:15-8:30.

This tool can be used to get delays for a maximum of ten multiple runs, where the user enters the vehicle snapshot and intersection snapshot data files for each run to get the delay output. Output of this form creates a database, in the directory where the program is located, which will be used by the query form later in the program. Once the user interface fields are entered, click the “Run” button to get the approach delay and stopped delay.

Figure 15 shows the query form that is used to extract details regarding the traffic operation of an intersection for a given interval of time. This part of the program produces an output that can be
used to get the details of vehicles using the intersection, information about the control of the intersection, such as the beginning and ending times of a phase and signal indications. To do a query on the database, the user inputs the node number of the intersection, the start time and end time of the interval, the signal-snapshot data file from TRANSIMS, and selects the database produced by executing the delay extraction function of the program. The user browses to the signal data file and the database by pressing the “signal” and “database” buttons, respectively. It should be noted that the delay part of the program generates a database in the directory where this program is located. The database for each run will be created in this directory.

Output Files

The TDET delay extraction function generates an MS Excel output as shown in Figure 16 and an MS Access database in the program directory containing the tables from which the data were extracted to determine the different forms of delay.

![Figure 16 TDET output.](image-url)
The format of the Excel output file is described in Table 8. The heading for each column in the Excel file is given in Table 8 with its corresponding description.

Table 8 Output File Description

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>From_Link</td>
<td>link number from which the vehicles leave the intersection</td>
</tr>
<tr>
<td>To_Link</td>
<td>link number to where the vehicles traveling from the intersection</td>
</tr>
<tr>
<td>From_Node</td>
<td>Start node of the link from where the vehicle leaves the intersection.</td>
</tr>
<tr>
<td>To_Node</td>
<td>End node of the link from where the vehicle leaves the intersection.</td>
</tr>
<tr>
<td>Supply_Volume</td>
<td>Traffic volume departed in the interval</td>
</tr>
<tr>
<td>Supply_Total Stopped Delay</td>
<td>Total stopped delay of the departed vehicles in the interval</td>
</tr>
<tr>
<td>SupplyStopped Delay/Veh</td>
<td>Stopped delay per vehicle of the departed vehicles in the interval</td>
</tr>
<tr>
<td>Movement</td>
<td>Type of movement</td>
</tr>
<tr>
<td>Interval</td>
<td>Represents the interval for which the delay is estimated</td>
</tr>
<tr>
<td>Demand_Volume</td>
<td>Total traffic volume arrived in the interval</td>
</tr>
<tr>
<td>Demand_Total Control Delay</td>
<td>Total control delay of the arrived vehicles in the interval</td>
</tr>
<tr>
<td>Demand_Control Delay/Veh</td>
<td>Control Delay per vehicle of the arrived vehicles in the interval</td>
</tr>
</tbody>
</table>

A query was completed to illustrate how to use the query function. For this illustration the purpose of the data was to output the data for all vehicles that arrived during a time period at the signalized intersection represented by the node 276. The appearance of the form after specifying all of the input values necessary to accomplish the query is shown in Figure 17 and the resulting output in Figure 18.
Figure 17 Input form for query.

Figure 18 Output from the query.
The format of the Excel output file is described in Table 9. The heading for each column in the Excel file is given in Table 9 with its corresponding description.

Table 9 Query-Output File Description

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle_ID</td>
<td>Vehicle number</td>
</tr>
<tr>
<td>Time</td>
<td>Time stamp of the vehicle</td>
</tr>
<tr>
<td>Node</td>
<td>User specified node</td>
</tr>
<tr>
<td>Link</td>
<td>Link from which the vehicle entering the intersection</td>
</tr>
<tr>
<td>Lane</td>
<td>The number of lane entering the intersection</td>
</tr>
<tr>
<td>Signal</td>
<td>Type of control</td>
</tr>
<tr>
<td>Distance</td>
<td>Distance from the intersection</td>
</tr>
<tr>
<td>Velocity</td>
<td>Velocity of the vehicle</td>
</tr>
</tbody>
</table>

**Advantages of TDET**

TRANSIMS output provides travel time but TDET provides supply control delay and demand control delay along with supply volume and demand volume. Travel time provided by TRANSIMS includes the time spent by a vehicle on the upstream intersection, therefore the delay estimated from TRANSIMS output is not precise. TDET delay includes the delay experienced in the intersection buffer in the case of a signalized intersection to provide more precise approach delay. In addition to the delay and query features, TDET creates a MS Access database which includes the vehicle snapshot table, link table, node table, intersection snapshot table, link length table and signal snapshot table. Additional types of queries can be executed directly in MS Access. When additional queries are found that would be useful then they can be added as an additional TDET query.
ANALYSIS

The main objectives of this project were the development of TDET and the validation of TRANSIMS. TDET is validated by comparing its delay measures to delay measures that were manually calculated from the TRANSIMS snapshot files, after it was stored in an Access database. Validation of TRANSIMS was conducted using the field data, TDET, and HCS. Even though the main objective of this thesis was to develop a tool for the validation of TRANSIMS, this research will give some insight into the capabilities and flaws in TRANSIMS modeling of traffic operations.

Validation of TDET

In order to make sure that TDET follows the algorithm, as mentioned in the previous discussion on the TDET, it was validated with manually extracted values. This validation lends credibility to TDET as a tool to validate TRANSIMS and its ability to efficiently process TRANSIMS output. Two measures were selected to validate TDET. One was traffic demand control delay and the other was traffic demand stopped delay. Characterizing the delay measures by the term “traffic demand” emphasizes the fact that the delay measures were aggregated by vehicle arrival time intervals, not departure time intervals.

TDET Vs Manual Calculation

For the signalized intersection of A Street and Jackson Avenue, output data files obtained from TRANSIMS simulation for the second field interval, 4:15 to 4:30, was selected for the analysis. Delay was determined manually and using TDET for a five-minute period in the middle of the overall three-hour simulation period.

To determine the delay values manually, delay of vehicles moving from link 360 to 392 (southbound through movement) was calculated. Link delay and intersection buffer delay for each vehicle arriving during the five-minute period were summed to determine the total delay. The following example explains how the delay was manually determined for an arriving vehicle. Data from the vehicle snapshot file for the vehicle number 101848 is shown in Table 10. In order
to reduce the size of the table, only the required fields for determining the demand control delay and stopped delay are shown in Table 10.

Table 10 Vehicle Snapshot Data for Vehicle Number 101848

<table>
<thead>
<tr>
<th>DISTANCE</th>
<th>DRIVER</th>
<th>LINK</th>
<th>NODE</th>
<th>TIME</th>
<th>VEHICLE</th>
<th>VELOCITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30748</td>
<td>101848</td>
<td>7.5</td>
</tr>
<tr>
<td>22.5</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30749</td>
<td>101848</td>
<td>15</td>
</tr>
<tr>
<td>37.5</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30750</td>
<td>101848</td>
<td>15</td>
</tr>
<tr>
<td>45</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30751</td>
<td>101848</td>
<td>7.5</td>
</tr>
<tr>
<td>60</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30752</td>
<td>101848</td>
<td>15</td>
</tr>
<tr>
<td>67.5</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30753</td>
<td>101848</td>
<td>7.5</td>
</tr>
<tr>
<td>75</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30754</td>
<td>101848</td>
<td>7.5</td>
</tr>
<tr>
<td>90</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30755</td>
<td>101848</td>
<td>15</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30756</td>
<td>101848</td>
<td>15</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30757</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30758</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30759</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30760</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30761</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30762</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30763</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30764</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30765</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30766</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30767</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30768</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30769</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30770</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30771</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30772</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30773</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30774</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30775</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30776</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30777</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30778</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30779</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30780</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30781</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30782</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30783</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30784</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30785</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30786</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30787</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>101848</td>
<td>360</td>
<td>239</td>
<td>30788</td>
<td>101848</td>
<td>0</td>
</tr>
<tr>
<td>7.5</td>
<td>101848</td>
<td>352</td>
<td>276</td>
<td>30790</td>
<td>101848</td>
<td>7.5</td>
</tr>
<tr>
<td>22.5</td>
<td>101848</td>
<td>352</td>
<td>276</td>
<td>30791</td>
<td>101848</td>
<td>15</td>
</tr>
<tr>
<td>37.5</td>
<td>101848</td>
<td>352</td>
<td>276</td>
<td>30792</td>
<td>101848</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 10 shows that the vehicle first appeared at a distance of 7.5 meters from the starting of the link with a velocity of 7.5 meter/second and last appeared at a distance of 105 meter with a velocity of 0 meter/second. The time stamps of the first appearance and last appearance of the vehicle were 30748 and 30788 respectively. The time of the vehicle entering the link at a
distance of 0 meters can be determined by extrapolation, using the velocity and distance of when the vehicle first appeared. For vehicle 101848, the extrapolated time it first appeared is 30748 - (7.5/7.5) = 30747.

Similar calculations were used to determine the departure time of vehicles. In the case of vehicle 101848, the velocity was 0; therefore the time at the end of the link stamp is 30788. The following items detail the calculations necessary to determine the delay on the link:

- Travel time on link = 30788-30747 = 41 second,
- Length of the link 360 from link file = 118.2 meters,
- Free flow speed = 13.4 meters/second,
- Free flow travel time = 118.2/13.4 = 8.8 second, and
- Link delay = actual travel time - true travel time = 41 - 8.8 = 32.1 seconds

Table 11 shows a portion of the intersection snapshot file. Each row of data in the table describes the vehicle occupying it, where the buffer can only be occupied by one vehicle at a time. Vehicles occupying the intersection buffer are listed in time sequential order. When the vehicle ID changes, this can be interpreted to mean that the previous vehicle left the buffer and in so doing exited the intersection.

Table 11 Intersection Data for 101848

<table>
<thead>
<tr>
<th>LANE</th>
<th>LINK</th>
<th>NODE</th>
<th>QINDEX</th>
<th>TIME</th>
<th>VEHICLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30789</td>
<td>101848</td>
</tr>
<tr>
<td>3</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30931</td>
<td>101941</td>
</tr>
<tr>
<td>3</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30668</td>
<td>101945</td>
</tr>
<tr>
<td>2</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30797</td>
<td>101996</td>
</tr>
<tr>
<td>1</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30793</td>
<td>101998</td>
</tr>
<tr>
<td>2</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30656</td>
<td>102006</td>
</tr>
<tr>
<td>1</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30791</td>
<td>102048</td>
</tr>
<tr>
<td>2</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30610</td>
<td>102129</td>
</tr>
<tr>
<td>1</td>
<td>360</td>
<td>276</td>
<td>1</td>
<td>30931</td>
<td>102131</td>
</tr>
</tbody>
</table>
Vehicle 101848 is the first vehicle listed in Table 11. The vehicle first appeared in the intersection buffer at time 30789 and it no longer appears in the intersection buffer. Therefore, the time that vehicle 101848 spent in the intersection buffer was 1 second. These last two findings of link delay (32.2 sec) and time in the intersection buffer (1 sec) give a total delay experienced by vehicle 101848 of 33.2 seconds (rounded).

A scatter plot comparing all of the manually extracted demand control delay and TDET are given in Figure 19. It is evident from Figure 19 that the demand delay values from TDET and manual extracted values are equal for all the vehicles. This suggests that TDET correctly calculates traffic demand control delay, according to the procedure described in the discussion of the TDET previously.

![Manual Estimation Vs TDET](image)

Figure 19 Manual and TDET demand control delay.

The vehicle snapshot file from TRANSIMS provides the time stamps for each simulation second. Therefore, the stopped delay for a given vehicle is the total number of time stamps for that vehicle with a velocity of zero. From Table 10, the total number of timestamps with a velocity of zero is 32 and from Table 11 it is one second. Therefore the stopped delay of vehicle 101848 is 33-sec. Stopped delays from TDET and manual estimation for all the vehicles are shown in Figure 20.
Figures 19 and 20 show that the delay estimated by the manual method and TDET are the same. Therefore it can be concluded that TDET functions correctly and can be used for the extraction of delay from TRANSIMS output files and a tool to validate TRANSIMS.

**Validation of TRANSIMS**

A validation of TRANSIMS conducted to determine the extent to which the TRANSIMS’ modeling logic could portray the actual traffic operations at unsignalized and signalized intersections. In this research, validation of TRANSIMS is conducted using the field data and HCS. TDET gives delay values from TRANSIMS, which are then compared with field values and HCS output. The primary purpose of the validation exercise is to demonstrate TDET’s usefulness as a tool for facilitating the analysis of TRANSIMS traffic output.

For traffic models, one important attribute of TRANSIMS is the ability to accurately predict performance measures, such as control delay. To assess TRANSIMS’ ability to replicate the actual values from the field and from the Highway Capacity Software (HCS), a prevalent software used in the transportation engineering industry four measures are used:

- Mean Error (ME),
- Mean Absolute Error (MAE),
- Mean Percentage Error (MPE), and
• Mean Absolute Percentage Error (MAPE).

Error in the predicted delay from TRANSIMS and the actual delay from field and HCS output is obtained by subtracting the predicted delay value from the actual delay values, as shown in Eq. 16.

\[ e_i = D_{ai} - D_{pi}, \]  

(16)

where

- \( e_i \) = error,
- \( D_{ai} \) = actual delay from field or HCS,
- \( D_{pi} \) = predicted delay from TRANSIMS.

Mean error (ME) is the average of errors for all the predictions.

\[ ME = \frac{1}{n} \sum_{i=1}^{n} e_i, \]  

(17)

where

- \( ME \) = mean error and
- \( n \) = number of predicted values.

Mean error does assess the magnitude of the error bias, but does not adequately define the average magnitude of the errors because the negative and positive error values cancel each other out. In order to overcome this problem, Mean Absolute Error (MAE), which is estimated from the absolute error, is used:

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |e_i|, \]  

(18)

Bias and precision can also be measured using the percentage of errors. Similar to the ME measure, error bias can be measured using the mean percentage error (MPE). The percent magnitude can be measured using the mean absolute percentage error (MAPE):

\[ MPE = \frac{1}{n} \sum_{i=1}^{n} \frac{D_{ai} - D_{pi}}{D_{ai}}, \]  

(19)
Validation of Unsignalized Intersections Traffic Operations

For the unsignalized intersections, the benchmark control delay is based on field measured queue delay. According to HCM methodology, the control delay is calculated from queue delay by adding five seconds to the queue delay. These five seconds account for the deceleration of vehicles to join the queue and the acceleration from the stop bar to the free flow speed.

TDET provides the stopped delay of vehicles and adjusts this to control delay by adding five seconds. HCM methodology-based control delay is obtained from the HCS software for each 15-minute interval. At this point, it should be noted that the definitions of queue delay and stopped delay are not the same. The point on which they differ is the inclusion of the queue move-up time. Recognizing this difference, the actual difference in the resulting values will be sufficiently small to allow a comparison between (TDET stopped delay + 5 seconds) and (Field queue delay + 5 seconds) to assess the accuracy of TRANSIMS. This slight compromise of data consistency was necessary because the field video tapes upon which this validation is based were misplaced by the individuals that extracted the data.

Comparison of TRANSIMS and Field

Predicted TRANSIMS delay values and actual field delay values are plotted in Figure 21, and the mean delay of the predicted and actual values are shown in Table 12. Mean error, mean absolute error, mean percentage error and mean absolute percentage errors are shown in Table 13 for the purpose of comparison.

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{D_{ai} - D_{pi}}{D_{ai}} \right|
\]  

(20)
Figure 21 TRANSIMS and field delay values.

From Figure 21, it is evident that most of the predicted values are above the actual values. For example, a five-second field delay value has a corresponding 216-second TRANSIMS delay value. This suggests that TRANSIMS overestimates delay.

Figure 22 TRANSIMS and field delay values.
Figure 22 confirms that the variance or standard deviation of the field values is less than the TRANSIMS delay values.

Table 12 Mean Delay Values of TRANSIMS and Field for Unsignalized Intersections.

<table>
<thead>
<tr>
<th></th>
<th>Mean Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRANSIMS</td>
<td>13.8</td>
</tr>
<tr>
<td>Field</td>
<td>18.7</td>
</tr>
</tbody>
</table>

Table 12 shows that the difference in the mean delay values from TRANSIMS and the Field is approximately 5.0 seconds. However, it is not possible to assert whether TRANSIMS is adequate for modeling unsignalized intersection traffic operations.

Table 13 Error Statistics for Unsignalized Intersection.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Error</td>
<td>-4.88</td>
</tr>
<tr>
<td>MAE</td>
<td>10.03</td>
</tr>
<tr>
<td>MPE</td>
<td>-35.6</td>
</tr>
<tr>
<td>MAPE</td>
<td>75.3</td>
</tr>
</tbody>
</table>

Good models give smaller values for MAE and MAPE and the values of ME and MPE will be closer to zero. Table 13 illustrates that the measures of bias (ME and MPE) have large negative values, which suggests that TRANSIMS tends to overestimate delay. The TRANSIMS model predicts the control delay with a MAE of ten seconds and a MAPE of 75.3 percent. These statistics, together with the others shown in Table 13 suggest that TRANSIMS has some flaws in the modeling of unsignalized intersection operations that either need to be recalibrated or undergo a change in logic.

In the case of two-way stop controlled intersections, delay is estimated for the minor street turning movements and major street left turn movements. The average field measured volumes and control delay from the field and TRANSIMS estimated delay are shown in Table 14 for each
movement. In this table, the eastbound and westbound approaches are the minor street approaches.

Table 14 Movement Volume and Control Delay from Field and TRANSIMS.

<table>
<thead>
<tr>
<th>Movement</th>
<th>FIELD</th>
<th>TRANSIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volume</td>
<td>Delay</td>
</tr>
<tr>
<td>EBLT</td>
<td>5.4</td>
<td>31.4</td>
</tr>
<tr>
<td>EBTH</td>
<td>6.3</td>
<td>35.4</td>
</tr>
<tr>
<td>EBRT</td>
<td>25.3</td>
<td>13.8</td>
</tr>
<tr>
<td>WBLT</td>
<td>1.0</td>
<td>12.5</td>
</tr>
<tr>
<td>WBTH</td>
<td>8.5</td>
<td>29.1</td>
</tr>
<tr>
<td>WBRT</td>
<td>37.3</td>
<td>12.3</td>
</tr>
<tr>
<td>NBLT</td>
<td>17.6</td>
<td>9.3</td>
</tr>
<tr>
<td>NBTH</td>
<td>76.3</td>
<td>0.0</td>
</tr>
<tr>
<td>NBRT</td>
<td>1.5</td>
<td>0.3</td>
</tr>
<tr>
<td>SBLT</td>
<td>45.8</td>
<td>10.1</td>
</tr>
<tr>
<td>SBTH</td>
<td>112.0</td>
<td>0.6</td>
</tr>
<tr>
<td>SBRT</td>
<td>8.3</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 14 shows that TRANSIMS systematically predicts delay values that are much larger than the field delays for the eastbound and westbound lefts turn movements. The same can be said for the west and eastbound through movements. However, the east and westbound right turn movement delay values are approximately equal. The large error statistics for the minor street left and through movements for this two-way stop controlled intersection may be attributable to TRANSIMS inability to model gap acceptance and/or the need for additional calibration.

From Table 14, it is evident that TRANSIMS underestimates the major street left turns. For instance, the TRANSIMS northbound left turn delay is 5.1 seconds, while the field delay is 9.3 seconds. When considering that five seconds was added to the TRANSIMS delay to approximate control delay, it is clear that few, if any, vehicles are stopping. After further investigation, it was
found that major street left turn vehicles are not stopping for the major street opposing traffic to accept a gap.

A more in depth analysis of the minor street movement error statistics is given in the following discussion. Figures 23, 24 and 25 show the predicted and actual values for the minor street left, through, and right turns.

Figure 23 Predicted values and actual values for minor street left turns.
Figure 24 Predicted values and actual values for minor street through.

Figure 25 Predicted values and actual values for minor street right.
Figures 23 and 24 show that the predicted values are much larger than the field values. This shows TRANSIMS’ inability to model minor street left turn and through movement operations and its ability to model minor street right turn movement operations. Error statistics for the minor street movements are given in Table 15 support this conclusion, where the right turn movement error statistics are much lower than those of the other two movements.

Table 15 Error Statistics by Unsignalized Intersection Turning Movement.

<table>
<thead>
<tr>
<th>Movement</th>
<th>LEFT TURN MOVEMENTS</th>
<th>THROUGH MOVEMENTS</th>
<th>RIGHT TURN MOVEMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Error</td>
<td>-24.1603</td>
<td>-13.5558</td>
<td>-1.09808</td>
</tr>
<tr>
<td>MAE</td>
<td>30.41009</td>
<td>25.47071</td>
<td>3.333197</td>
</tr>
<tr>
<td>MPE</td>
<td>167.1861</td>
<td>96.7336</td>
<td>30.84059</td>
</tr>
<tr>
<td>MAPE</td>
<td>-145.911</td>
<td>-70.0501</td>
<td>-14.8698</td>
</tr>
</tbody>
</table>

Comparison of TRANSIMS, HCS and Field

HCM analysis is done for the first four 15-minute intervals of the field data and the average results from TRANSIMS, field studies and HCS are shown in Table 16. Because field measured critical gap and follow-up time values are unavailable, default values are used in HCM analysis. HCS provides the lane group control delays. To be consistent, the TRANSIMS and field delay values were converted to lane group control delays.

Table 16 Field, HCS and TRANSIMS Control Delays and Error Statistics for Unsignalized Intersections.

<table>
<thead>
<tr>
<th>Mvmt</th>
<th>Control Delay</th>
<th>Error from Field</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Field</td>
<td>HCS</td>
<td>TRANSIMS</td>
</tr>
<tr>
<td>EBLTTH</td>
<td>47.5</td>
<td>95.2</td>
<td>59.1</td>
</tr>
<tr>
<td>EBRT</td>
<td>16.5</td>
<td>12.9</td>
<td>18.6</td>
</tr>
<tr>
<td>WBLTTH</td>
<td>28.6</td>
<td>46.9</td>
<td>44.9</td>
</tr>
<tr>
<td>WBRT</td>
<td>13.1</td>
<td>11.8</td>
<td>13.2</td>
</tr>
<tr>
<td>NBLT</td>
<td>8.8</td>
<td>8.7</td>
<td>0.1</td>
</tr>
<tr>
<td>SBLT</td>
<td>10.5</td>
<td>8.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Mean</td>
<td>-9.8</td>
<td>-1.9</td>
<td>36.1</td>
</tr>
</tbody>
</table>
Table 16 shows that when compared to TRANSIMS, HCS has better overall results. On the other hand, TRANSIMS does perform better for the minor street movements.

Table 16 shows that TRANSIMS tends to overestimate the minor street delays and the difference in control delay from field and TRANSIMS would be large if the major street left turn operations were more realistic. If major street left turn operations were improved to reflect realistic gap acceptance behavior, then there would be queuing of the major street left turn. This major street left turn queuing would reduce the capacity of the minor street left and through movements in the model, further increasing the TRANSIMS minor street left and through movement delays, which are already too high.

**Signalized Intersections**

Validation of the signalized intersection was carried out by comparing control delay from the field, TDET and the HCS.

**Comparison of TRANSIMS with Field Data**

TRANSIMS delay values and field values are plotted in Figure 26 and the mean delay values from the field and from TRANSIMS are shown in Table 17.

![Figure 26 Signalized intersection TRANSIMS and field values.](image-url)
Table 17 Mean Delay Values of TRANSIMS and Field for Unsignalized Intersections.

<table>
<thead>
<tr>
<th></th>
<th>Mean Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRANSIMS</td>
<td>12.11</td>
</tr>
<tr>
<td>Field</td>
<td>11.27</td>
</tr>
</tbody>
</table>

It is evident from Figure 25 that estimates vary widely from the field values. However, Table 15, which contains the average delay values, averaged over all time intervals, shows that overall the estimates are distributed along the line. Table 17 shows that the average difference in mean delay values is approximately one second. Figure 25 and Table 17 show that TRANSIMS modeling of fixed time signalized intersections seems to be good.

Figure 27 TRANSIMS and field delay values.

Figure 27 confirms that the variance or standard deviation of the field values is less than the TRANSIMS delay values.
Table 18 Error Statistics for Signalized and Unsignalized Intersections.

<table>
<thead>
<tr>
<th></th>
<th>Signalized Intersection</th>
<th>Unsignalized Intersection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Error</td>
<td>-0.84</td>
<td>-4.88</td>
</tr>
<tr>
<td>MAE</td>
<td>5.26</td>
<td>10.03</td>
</tr>
<tr>
<td>MPE</td>
<td>-12.09</td>
<td>-35.6</td>
</tr>
<tr>
<td>MAPE</td>
<td>47.93</td>
<td>75.3</td>
</tr>
</tbody>
</table>

From Table 18 it can be seen that the measures of bias (ME and MPE) have small values relative to those shown for the unsignalized intersection. This relationship is consistent with the other two measures, where the TRANSIMS MAE and MAPE error statistics are much smaller at approximately 5 seconds and 48 percent, respectively.

Table 19 illustrates that the average control delay values from TRANSIMS are close to the average field values, which strongly suggests that TRANSIMS models fixed time signalized intersections well. A turning movement wise analysis was conducted to find out how TRANSIMS models individual turning movement delays and to visualize the corresponding variability. Figures 28, 29, and 30 show the predicted values for left, through and right turn movements, where each point represents one observation.

Table 19 Average Control Delay and Traffic Volumes by Turning Movement from Field and TRANSIMS for Signalized Intersections.

<table>
<thead>
<tr>
<th>Movement</th>
<th>FIELD</th>
<th>TRANSIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Volume</td>
<td>Delay</td>
</tr>
<tr>
<td>EBTH</td>
<td>57</td>
<td>19.4</td>
</tr>
<tr>
<td>EBRT</td>
<td>31</td>
<td>7.8</td>
</tr>
<tr>
<td>WBLT</td>
<td>2</td>
<td>18.6</td>
</tr>
<tr>
<td>WBTH</td>
<td>10</td>
<td>15.0</td>
</tr>
<tr>
<td>SBLT</td>
<td>3</td>
<td>9.2</td>
</tr>
<tr>
<td>SBTH</td>
<td>125</td>
<td>11.7</td>
</tr>
<tr>
<td>SBRT</td>
<td>47</td>
<td>2.1</td>
</tr>
</tbody>
</table>
Figure 28 Field and TRANSIMS delay values for left turn movements for signalized intersections.

Figure 29 Field and TRANSIMS delay values for through movements for signalized intersections.
Figure 30 Field and TRANSIMS delay values for right turn movements for signalized intersections.

In all cases, results for individual 15-minute periods vary substantially. Figures 28 and 30 show that the left and right turn movement predicted delay values do not fall around the line, which shows that the mean predicted values are away from the line. This high variability in the left and right turn movement delay may be an artifact TRANSIMS shortcoming with modeling gap acceptance behavior. However, Figure 29 shows that the through delay values are clustered around the line, suggesting that on average TRANSIMS models through vehicle signalized operations more consistent with field operations.

Error statistics for left, through and right turn movements are shown in Table 20. The MAE and MAPE error statistics for the left turn movement are approximately 10 seconds and 107 percent, respectively. Table 20 also shows that the left turn movement ME value is very small while MAE and MAPE are large. This large variability that are centered around the field delay value can be explained by recalling that the largest left turn volume is three vehicles, which is very low for a 15-minute period. Increasing the simulation time would have reduced this variability.
Table 20 Error Statistics--Signalized Intersection Turning Movements.

<table>
<thead>
<tr>
<th>Movement</th>
<th>LEFT TURN MOVEMENTS</th>
<th>THROUGH MOVEMENTS</th>
<th>RIGHT TURN MOVEMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Error</td>
<td>-1.19</td>
<td>-2.61</td>
<td>1.55</td>
</tr>
<tr>
<td>MAE</td>
<td>10.50</td>
<td>4.72</td>
<td>2.49</td>
</tr>
<tr>
<td>MPE</td>
<td>-45.74</td>
<td>-20.30</td>
<td>-10.47</td>
</tr>
<tr>
<td>MAPE</td>
<td>107.27</td>
<td>32.81</td>
<td>61.06</td>
</tr>
</tbody>
</table>

Comparison of Data from TRANSIMS, HCS and the Field

HCM analysis was done for the peak hour. Four 15-minute intervals of the field data from 4:30 to 5:30 PM were used. The average results from TRANSIMS, Field and HCS are shown in Table 19. Because HCS provides the lane group control delays, consistency in delay measures was maintained by converting the TRANSIMS and field delay values to lane group control delays.

Table 21 Signalized Intersection Control Delays and Error Statistics for HCS and TRANSIMS.

<table>
<thead>
<tr>
<th>Mvmt</th>
<th>Control Delay</th>
<th>Error from Field</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Field</td>
<td>HCS</td>
<td>TRANSIMS</td>
</tr>
<tr>
<td>EBTH</td>
<td>20.9</td>
<td>19.7</td>
<td>22.3</td>
</tr>
<tr>
<td>EBRT</td>
<td>8.6</td>
<td>16.9</td>
<td>3.4</td>
</tr>
<tr>
<td>WB</td>
<td>14.6</td>
<td>16.3</td>
<td>17.3</td>
</tr>
<tr>
<td>NBLTTH</td>
<td>11.6</td>
<td>11.3</td>
<td>11.3</td>
</tr>
<tr>
<td>NBRT</td>
<td>2.0</td>
<td>11.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Mean</td>
<td>-3.0</td>
<td>0.2</td>
<td>97.1</td>
</tr>
</tbody>
</table>

In Table 21, the mean row shows that HCS has better overall results. On the other hand, TRANSIMS does perform better for the right turn movements. In addition, TRANSIMS error statistics are comparable to those of HCS, suggesting that stop bar operations at signalized intersections are fairly consistent with expected levels of accuracy in the transportation engineering industry.
Vehicle Trajectory using TDET

Vehicle trajectories can also be obtained from TDET using the TDET query form. The query output gives vehicle positions and time stamps on approaches to an intersection with the traffic control. To demonstrate this query function, a query demonstration was made for a time period covering more than one cycle length (70 seconds), starting at the beginning of green and ending at 80 seconds for the southbound through movement. This analysis queried for vehicles in the second through lane next to the shared left and through lane on the southbound approach to the A and Jackson Streets intersection. The table of TDET results is given on the CD attached to this thesis. Figure 31 illustrates the results in the form of a time–space plot.

![Figure 31 Trajectory of vehicle movement on lane number 2 of the signalized intersection for one cycle length.](image)

From Figure 31, it can be seen that the vehicle numbers 1, 7, 8, 10 and 11 arrived at the intersection during the red time. Vehicle number 2 changed lanes from lane 3, to lane 2 and then to lane 1. As a result, vehicle 2 can be seen on lane two for 4 seconds. Vehicle number 3 changed
from lane 2 to lane 1, and for this reason it is visible in lane two for 7 seconds. Vehicle 10 changed from lane 2 to lane 1 after it joined the queue. From this output, it appears that lane changes frequently occur near the intersection, which is not very representative of field conditions. Two factors may contribute to this behavior. One is vehicle routing requires the vehicles to make lane changes within short distances. The other is that the TRANSIMS driver behavior algorithms insufficiently model driver anticipation of upstream turns and driver propensity to change lanes upstream of the intersection.

It can be concluded from this section that TDET can be used to get the vehicle trajectory for the study of vehicle following and driver behavior characteristics that emerge from TRANSIMS micro-simulation.

**Summary**

This section shows that TDET efficiently extracts data from the TRANSIMS output files and provides useful performance measures in the form of delay for analyzing traffic operations. Also, this tool can be used for the validation of TRANSIMS for traffic operations. Validation of TRANSIMS with Field and HCS using TDET shows that TRANSIMS has serious short comings when modeling unsignalized intersections but models signalized intersections at acceptable levels of accuracy, when compared to HCS results.
CONCLUSIONS AND RECOMMENDATIONS

Conclusions

With regards to the objective of this study, it can be concluded from the validation of TDET tool that it is efficient in extracting the performance measures from TRANSIMS and can be used as a tool to validate TRANSIMS. TDET provides demand stopped delays, supply stopped delays, demand control delays, supply control delays. In addition, it creates a database which includes the vehicle snapshot file and intersection snapshot output files and link file and link length input files from the TRANSIMS in MS Access tables. Data provided by the TDET query form can be used to extract vehicle trajectory data that prove useful for validating additional micro-simulation features in TRANSIMS.

Based on the validation of TRANSIMS for unsignalized intersections, it can be concluded that TRANSIMS does not model unsignalized intersections very well, tending to over estimate control delays. If the major street left turn operations were improved to reflect realistic gap acceptance behavior then the delays on the minor street would increase even more.

For signalized intersections, TRANSIMS performed very well, in some case surpassing HCS. In fact, TRANSIMS models right turn movements more accurately than HCS. However, care must be taken to work with replicate runs to produce more reliable averages, because individual run results have high variability.

HCS performs better in overall modeling, but this is expected because it has been created and refined based on extensive model development and calibration. There is no cause for concern when modeling signalized intersections. However unsignalized intersections are not modeled well, especially when considering the fact that TRANSIMS does not model any control on major street left turns, making the travel costs of this turn much lower than reality. As a result, this low major street left turn travel cost may result in assigning unrealistically high numbers of vehicles to paths that include major street left turns. TRANSIMS model minor street left turn movements much more efficiently. It may be due to the reason cited above.
Recommendations

It is recommended from this study that TDET can be used for the validation of TRANSIMS with field or with other transportation software. It is recommended that the modeling concept used in TRANSIMS has to be updated to accommodate major street left turn movements because the router module in TRANSIMS uses the delay at the intersection to determine least-cost routes on which to assign traffic. More validation has to be done to ascertain the efficiency of TRANSIMS in the modeling of signalized intersections, focusing on actuated isolated intersections and actuated coordinated intersections.

At this time, TDET creates a rich database in MS Access that can be queried in many ways. While this is a valuable contribution to TRANSIMS validation, it would be very useful to add some automated performance calculations, in addition to delay. For example, queue length and green time arrival rates should be included in the set of automatically extracted performance measures.
PART 2: GIS APPLICATIONS GIS INPUT DATA PROCESSING METHODOLOGIES FOR TRANSPORTATION PLANNING MODELS

METHODOLOGY

Background

Current transportation planning practice, with respect to associating demographic and socioeconomic data to TAZs, is usually composed of tasks where planners break the modeled region down to small areas, the TAZs. As a rule of thumb, TAZ boundaries follow US Census boundaries [10]. However, modelers frequently reduce the TAZ size, in an effort to more accurately model land use variations. Modified TAZ boundaries are often implemented using boundaries manually established by the modeler and the coinciding socioeconomic and demographic variables are determined through a tedious process, such as dividing Census data from a block or block group to respective, smaller TAZs.

A procedure was developed for estimating the number of households by socioeconomic categories for TAZs in an effort to more accurately represent the spatial variation of household characteristics for micro-simulation [11]. This procedure is based on a multi-proportional technique that assumes initial values and TAZ totals are available. As a result, it relies on socioeconomic data that may not exist for each TAZ. For example, if TAZs were smaller than their respective Census block groups or blocks then the TAZ data, such as number of households, may not be available.

 Developments in GIS have led to increasingly computerized methods for obtaining, processing, and managing TAZ data. Researchers working with TRANSIMS divided a TAZ into sub-TAZ areas, called activity locations, and used GIS based techniques to associate socioeconomic data to the sub-TAZ areas. For a given TAZ, trip ends were distributed based on the weights of the sub-TAZs, where each sub-TAZ is weighted by a measure of the relevant land use or activity such as the number of households or employees. Geocoding, a GIS technique, was suggested to
measure the number of households in each sub-TAZ. With geocoding, individual households are located on a TIGER Line network, based on their respective addresses [12]. The researchers did not implement geocoding in their reported work. Errors associated with this suggested approach remain to be determined, especially in light of unlisted phone numbers and P.O. Box addresses [10].

In an effort to measure land use mix, researchers have developed indices, utilizing land use zoning data and transportation infrastructure data in a GIS environment [13, 14]. These measures potentially contribute to a better understanding of traveler behavior and how it relates to land use and the transportation system.

Recent research in Florida suggests that using GIS to combine TAZ, land use, and aerial imagery can increase understanding of land use trends [15]. The researchers integrated data from a number of data sources within a GIS environment. Relevant data that were included were aerial photographs, population data, land use zoning data, employment data, and traffic analysis zones. TAZ socioeconomic data were given from an existing database, so the primary contribution relevant to this paper is the demonstration of the capability and value of integrating the described datasets to measure and understand land use in the context of transportation systems. No procedures were developed for associating data from aggregate areas to smaller TAZs.

Research using GIS techniques to classify land use from aerial photos, Landsat Thematic Map satellite data, and land use zoning data proved promising [16]. Accuracy ranges for properly classifying residential, commercial, and industrial land uses ranged from 70 to 90%. Residential, commercial, and industrial classification accuracies were 90%, 83%, and 70%, respectively.

Recent conclusions from a research consortium using remote sensing for environmental assessment and planning of transportation projects strongly suggests that GIS techniques can be used to classify land use features. This research was limited to corridor alternative analysis and did not consider measuring land use in the context of TAZs as input for travel demand modeling [17].
Some have proposed and implemented the use of sub-TAZ, or smaller TAZ, areas and found that their use is favorable and major metropolitan areas have undertaken the task of increasing the number of TAZs to increase spatial resolution [7]. However, in order to reduce TAZ size, modelers need to know socioeconomic data that are usually not available at that size.

Computational procedures do not exist for systematically associating socioeconomic data to TAZs when the totals are not known because their areas vary from the areas in which the data are available. Fortunately, relevant GIS tools, techniques, and databases are available to transportation planners and engineers. A computational procedure that fuses existing data to obtain more accurate TAZ demographic and socioeconomic variable values is needed to increase the ease and flexibility with which modelers can increase spatial resolution through smaller TAZs.

**Methodology**

Transportation planning models estimate the number and nature of trips generated within a geographical area or TAZ using demographic measures known for such TAZs. Typical measures needed for estimating trips and their purposes are the number of households, number of employees, and other demographics. The process of determining these quantities for each TAZ can be difficult, especially if the data from which these measures are taken are not aggregated into geographical boundaries equivalent to those of the desired TAZ. Because such demographics are generally known for Census blocks, the smallest Census division, they are a likely candidate for higher level planning models for this reason. The purpose of this research is to establish a methodology for accurately distributing Census data needed for transportation models to higher resolution or smaller TAZs. The method for determining the socioeconomic variables for TAZs on a finer level than Census blocks but using Census demographic data is described herein as the area-proportional TAZ estimation approach.

The essence of this technique is to estimate the proportion of a Census block’s quantitative demographic measures belonging to a smaller subsection of the Census block. These subsections become the new, disaggregate TAZs. The quantitative demographic measures are estimated by
comparing the sum of the building footprint area of a particular land use within the sub-Census-block TAZs to the total building footprint areas of the same use in TAZs’ encompassing Census block. Quantitative census demographics which can reasonably be correlated to building areas can then be estimated for each TAZ as scaled by the proportion described.

Remote sensing techniques for automated feature extraction (AFE) of building footprints from various types of imagery and spatial data constitute the means for determining such building footprint areas and proportions. The term remotely sensed is used herein to describe images and spatial data observed from a remote position.

The efficacy of area-proportional TAZ estimation relies upon the quality of the input data. A major objective of this research was to evaluate this technique’s performance as a function of input data used as well as to evaluate the quality of the feature extraction given the input data used. Building footprint areas automatically observed from the input images are a primary input used to calculate the proportions mentioned. The images from which buildings were observed and extracted are secondary input data. Various combinations of remotely sensed secondary data were tested for the quality of the building footprint areas they generated as well as the success of the area proportional method using these input data and combinations of input data. Extracted building footprints, in particular are sensitive to the variation of these combinations.

Because it is the central focus of this research, the area-proportional TAZ estimation technique will be presented first, followed by a detailed discussion of the underlying techniques and procedures used to generate the supporting data and inputs. Each section will be concluded with a description of the approach used to validate and quantify error for demographic estimation and input data creation.

Area-Proportional TAZ Estimation

This method assumes that the following information is given:

1. Socioeconomic data areas (SDA)
2. Socioeconomic data (SD) that can logically be scaled or distributed
3. TAZ boundaries (developed by the user; cannot straddle multiple SDAs)
4. Location, land use type, and footprint area of buildings
Given the above information, the task is to estimate the socioeconomic variables for TAZs based on the same known variables for the SDAs. Equation 21 gives computation by which the socioeconomic data belonging to socioeconomic data areas may be scaled to estimate the proportion present in a TAZ.

\[ I_{ij}^k = L_j \times \frac{a_{mj}^k}{A_{mj}} \]  

(21),

where,

- \( amjk \) building footprint area within TAZ \( k \) in SDA \( j \) given to land use type \( m \) associated with SD variable \( i \) (must be logically scaleable),
- \( Amj \) total building footprint area in SDA \( j \) given to land use type \( m \), equal to \( \sum_k a_{mj}^k \),
- \( lijk \) estimate of SD variable \( i \) for TAZ \( k \) within data area \( j \), and
- \( Lij \) given value of SD variable \( i \) in SDA \( j \).

For example, consider establishing the number of households in the TAZs shown in Figure 32. Socioeconomic data areas (SDA) A to C have boundaries that are different from any of the TAZs, indicated by numbers (note that each TAZ created must be completely within an SDA). Assume that residential land use (\( m = \) residential) can be associated with the SD variable of interest, in this case, number-of-households (\( i = \) No. Households). The number of households within each SDA, \( Lij \), can then be distributed to each TAZ by taking the sum of the building footprint areas of land use \( m \), within TAZ \( k \) (a sub-part of SDA \( j \)) divided by the sum of the building footprint area of the same land use \( m \) for the subject SDA \( j \) multiplied by \( Lij \). Again referring to Figure 32, if the total building footprint area within the residential land use in data area A, \( A_{residential} \), were 100,000 m\(^2\) and the corresponding area in TAZ 3 of data area A, \( a_{residential, A3} \), were 25,000 m\(^2\), then the proportion would be 0.25. If \( L_{number of household, A} \) were 200 households, then \( l_{number of household, A3} \) would be 50 households. In short, the SD variable of choice is distributed to the TAZs within each SDA by using the proportion of building footprint areas with consistent land use.
Data Types and Preparation Procedures

The area proportional estimation method assumes the availability of various source data and resulting information types used as input variable values. This section explains the source data used in this research and the preparation process that results in the information required for input variable values to estimate TAZ socioeconomic data. Specifically, the information assumed available by this method for establishing input variable values can be listed as follows:

1. TAZ boundaries
2. Location, land use type, and footprint area of buildings
3. Socioeconomic data areas (SDA)
4. Socioeconomic data (SD)
5. Direct association between land use and SD variables (i.e., residential land use and number of households)

First, a discussion of the different data types and sources is given. Then each of the five listed items of information assumed available is discussed.

Data Types

Diverse data types could be used to support the estimation of SD variable values for TAZs. Some variables that are commonly used in transportation planning are directly related to residential land uses. This research utilizes the land use measure of number of households. Fortunately, the method presented in this paper could be applied to other measures, such as the number of individuals employed by employment type and the number of households by household type.
The city of Moscow, Idaho was used as the test bed for the proposed methodology. Data types used in this research and their corresponding sources are as follows:

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Use Zoning Map for City of Moscow Idaho</td>
<td>Digitized land use map where land use areas are represented by polygons stored in a GIS database</td>
</tr>
<tr>
<td>US Census 2000 TIGER Line file: US Census Bureau</td>
<td></td>
</tr>
<tr>
<td>Year 2000 US Census Bureau Number of Households</td>
<td>Block and block group Census data for number of households</td>
</tr>
<tr>
<td>Year 2003 aerial photographs &amp; LiDAR Elevation Data: Avista Utilities</td>
<td></td>
</tr>
<tr>
<td>Year 2004 Satellite imagery: QuickBird</td>
<td></td>
</tr>
</tbody>
</table>

**TAZ boundaries**

Understandably, the proposed method requires TAZ boundaries. TAZ boundaries are defined by the user with only two constraints: first, the boundaries of a given TAZ must be within a single SDA; second, TAZ boundaries should not bisect any building. TAZ boundaries can be established manually and/or through computational processes. For this research, TAZs were computationally generated to reflect natural breaks formed by the intersection of municipal land use zoning data and US Census blocks. Further divisions could be derived using features such as topographical breaks and water formations. One advantage of using natural breaks in the creation of TAZs is the resulting simplification of the computational process for the area of proportional estimation method. This simplification is brought about by creating traffic analysis zones of homogenous land use type eliminating the need to differentiate the land use of building footprint.
areas. This means that all building footprint areas within a TAZ may be collectively summed to obtain \( amjk \) for use in the area proportion method.

In using natural breaks for traffic analysis zones, it is important to note that correction of the SDA boundaries may be necessary to create consistency with aerial photography and other remotely sensed data used. This consistency will prevent the creation of TAZs that bisect buildings and erroneously assign building footprint area to multiple centers of traffic activity.

**Location, land use type, and footprint area of buildings**
The objective here is to determine the location, land use type, and footprint area of all buildings in the city. Clearly, manually performing the task would be time consuming. Fortunately, computational methods for obtaining building footprints are available through remote sensing technology. This developing technology can provide imperfect but useful information about a remotely sensed area. In the context of this research, remote sensing describes the capture of urban features from a detached or remote source. This feature extraction occurs in two steps: first the acquisition of digital imagery, and second, processing of that imagery in which desired features (in this case building footprints) are identified and extracted. The quality of feature extraction results obtained using remote sensing techniques can vary significantly with various types of image data. Satellite data, for instance, has the advantage of providing coverage over a large area of a single image but must be of sufficient spatial resolution for image processing algorithms to recognize the features of interest. Arial photography is generally of sufficient spatial resolution but most often offers less in terms of spectral resolution. Another type of remotely sensed data called LiDAR, or Light Detection and Ranging, provides information about feature elevation but is generally less widely available. One objective of this research was to identify those remotely sensed data, or combinations of data, from which the best urban extraction results can be obtained. A discussion of this exploration and its structure follows in the Testing and Experimental Design section.

**Footprint Area of Buildings:**
The process of obtaining building footprint areas and locations through image processing and feature extraction entails three major steps: first, the creation of training data; second, feature
extraction; and third, cleaning. Training data are manually digitized examples of the desired feature of interest which are then used by image processing algorithms to identify similar features throughout the image. In this research, the best feature extraction results have been obtained using training data that are randomly distributed across the variety of building footprints within the image. The number of training data used was approximately ten percent of the total number of expected building features.

A variety of image processing software perform feature extraction functions. For this research, Visual Learning System's Feature Analyst Extension to Environmental Systems Research Institute’s ArcGIS geographical information systems package was used. The quantity of training data and the manner in which they are distributed across the image bears a large influence on the results of the feature extraction process. This research found the approach to training data mentioned above to produce the best results.

Image feature extraction or classification produces three things:

1. Correctly classified or extracted features coincident with the same feature on a reference feature extraction or map,
2. Commission error or the classification of an undesired feature by mistake, and
3. Omission error, the absence of classification or feature extraction on a feature that actually exists but was not identified as such by the image processing algorithm.

The concept of cleaning feature extraction results is not widespread and was only found in the Feature Analyst extension mentioned. There are three possible avenues for enhancing image feature extraction results: 1) aggregation of pixels to a selected threshold, 2) pixel edge smoothing of extracted features, and 3) retro-enhancing training data by identifying correctly classified and commission error regions. These three procedures are described below.

Aggregation of pixels
Aggregation of pixels is a process in which the user defines the number of pixels in the digital image that the typical feature of interest would be equal to or greater than. This is basically the same as specifying an area threshold eliminating all classified features that are not at least as large as the area chosen for the smallest acceptable building footprint. Aggregation of pixels was applied to all features extracted in the experiments behind this research.
Pixel edge smoothing
Because image processing algorithms must “decide” on the inclusion or exclusion of each pixel in an image, extracted features tend to have jagged edges on close examination. This is also due to the mixed pixel concept where a portion of the pixel contains the feature of interest and a portion does not. The final cause of this jagged edge is that features are extracted from grid based pixels so that all lines diagonal to the orientation of the grid are actually zigzags. In vectorizing extracted features, i.e. converting them from a grid base to vector referenced data, each jag becomes a vector referenced vertex. A line may contain this zigzag pattern under close examination so that where two vertices could suffice to represent the line, thousands are present. Memory and digital storage requirements are obviously higher for the latter case. Smoothing is the process of removing all vertices within a tolerance from the straight line. It is useful for reducing storage requirements for extracted features and enhancing aesthetics. However, the process is computationally intensive and for this reason it was not applied in this research.

Retro-enhancing training data
This involves the user identification of correctly and incorrectly classified features after a classification, followed by a reclassification [23]. This cleaning procedure was not used in this research because of the introduction of bias through user reevaluation of results. The exclusion of bias was of particular importance in order to fairly evaluate the input data experiments.

Location and Landuse
Feature location is a natural result of the classification process. The result of the classification process is a map, image, or collection of geographic vector data conforming to the same geographic projection as its source image.

Features extracted as buildings were equated to building foot prints. The land use type of the buildings foot prints was determined by overlaying the extracted building foot prints in the building footprint map with the digitized land use map. If a building footprint polygon fell within a land use zone polygon then the land use type field of the building footprint polygon was updated with the land use type of the land use polygon.
As usual, the accuracy of this method is dependent on the quality of the data. This is especially true for mixed land use situations. Conditions of mixed land use occur because of …

**Socioeconomic data areas (SDA)**
Census boundaries frequently define the areas for which socioeconomic data (SD) are available. In this research, Census 2000 block and block group Census areas were used. Only those blocks and block groups that covered areas of the City of Moscow were included in this research.

**Socioeconomic data (SD)**
Socioeconomic data corresponding to each SDA were also used in this research. One table of Census data was used and it was for the number of households, by Census block. Values in this table were used as the aggregate SD variable for Equation 1 to estimate TAZ level SD variable values.

**Direct association between land use and SD variables**
In order to use Equation 1, an association between land use and the SD variables must be drawn. This is necessary in order to assume that the SD variable is proportional to the geographic area of a land use in TAZs, $a_{mj}^k$, with respect to the geographic area of the same land use in their corresponding data areas, $A_{mj}$, as shown in Equation 22. For example

$$\frac{l_{mj}^k}{L_{mj}} \propto \frac{a_{mj}^k}{A_{mj}}$$  \hspace{1cm} (22)

In this research, the above association was drawn between residential land use and number of households by Census block.

**Testing and Experiment Design**
Recommendation of the most apt data for building footprint feature extraction is one of the primary objectives of this research. In order to accommodate this aim, four remotely sensed data types and 14 major experiments were planned. The first four of these experiments are the
extraction of building footprints using one of each of the four data types. The balance of the experiments is based on promising combinations of the data.

Two prominent factors noted were considered in designing the evaluation of the feature extraction results. The primary factor is the input data set or combination of datasets used. The second is the natural separation of sensed regions as observed by the naked eye. Experiments were evaluated for extraction results overall and as a function of four sub-regions into which the benchmark evaluation data were divided. These are 1) a commercial industrial region with less vegetation, larger buildings and closer spacing; 2) a large multi-family residential area on the University of Idaho Campus with low-lying vegetation and continuous terraced buildings; 3) a new residential region with little mature vegetation, fairly consistent building sizes, and greater building spacing; and 4) an older residential region containing more mature vegetation and tree occlusion, greater variety in building sizes, and generally more dense building spacing. In both cases standard remote sensing accuracy assessment concepts were used.

For remote sensing applications, a common method for quantifying the effectiveness of image processing and feature extraction is through commission errors and omission errors. In this research, a commission error describes a feature misclassified as a building and an omission error indicates failure to classify an actual building as such. These errors were quantified in terms of percentages of the correct building footprint area recorded in the benchmark data. Commission error can, therefore, be greater than 100%, while the omission error and correctly classified percentages must always sum to 100%. These errors were extracted for all buildings within the benchmark dataset for each experiment.

Four types of input data form the basis for the experiments performed. These input data and the combinations analyzed are shown in Table 22. Several variations including alternative arrangements of the data were used for additional experiments, increasing the number of experiments from the core 14 to 22. Because the 15th combination including all four types of data was unmanageably large, it was excluded.
Table 22 Types and combinations of experimental input data

<table>
<thead>
<tr>
<th>Number</th>
<th>Experiment Name</th>
<th>RGB Data</th>
<th>DEM 10ft</th>
<th>Quickbird Multispectral</th>
<th>Quickbird Panchromatic 1m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Avista RGB .5ft</td>
<td>USDA RGB 1 m</td>
<td>NIR 8 ft</td>
<td>RGB 8 ft</td>
</tr>
<tr>
<td>19</td>
<td>AvistaM_7Ag900RoBld</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>AvistaBE2_23Ag900RoBld</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>USDADE3_7Ag90RoBld</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>USDAM_7Ag99RoBld</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>AvistaDEMM_5Ag59RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>QbirdM_7Ag99RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>QbirdPanM_7Ag196RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>AvistaRGB&amp;DEMM_7Ag900RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>AvistaRGB&amp;QbirdMulM_7Ag90RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>USDA&amp;QbirdNIRM_7Ag49RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>USDA&amp;QbirdMulM_7Ag99RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>QbirdMul&amp;AvistaIDMEM_7Ag99RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>QbirdMul&amp;AvistaIDMSPm_7Ag99RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>QbirdPan&amp;NIRM_7Ag196RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>QbirdPan&amp;MulM_7Ag99RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>AvistaRGB&amp;QbirdNIRM_7Ag900RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AvistaRGB&amp;DEMM&amp;QbirdMulM_7Ag900RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>AvistaRGB&amp;DEMM&amp;QbirdPANm_7Ag900RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>AvistaRGB&amp;QbirdPANm&amp;MulM_7Ag900RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>AvistaIDMSPm&amp;QbirdPan&amp;MulM_7Ag196RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>AvistaDEMM&amp;QbirdPan&amp;MulM_7Ag196RoBld</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

- a Non-standard image processing kernal tried
- b 1 meter RGB used instead of 0.5 foot
- c NIR band used alone rather than full Multispectral
- d DEM Slope used rather than DEM

Benchmark Dataset for Validation

To quantify method degree of accuracy, a benchmark data set was used. The purpose of this benchmark data set was to establish the true values for \( lijk \). This validation data set is a map of manually digitized building footprints whose land use types were verified in the field.

Digital maps of building footprints were created for four 0.33 square mile regions of the City of Moscow that represent the spectrum of land use in small cities and suburban areas. These maps were created by manually digitizing all building footprints in the selected regions. Then the land use type for each building was verified based on the following land use type categories:

1. Commercial/industrial
2. School
3. Government
4. Single-family residential
5. Multi-family residential

In addition, the number of land use units residing in each building was also field verified and recorded.
While a more detailed list of land use types could have been used, the five items listed above were sufficient to validate the accuracy of the proposed methods. This sufficiency is based on the primary objective of the methods in this paper, which is to locate places of residence. If type of employment was also considered in the methods then certainly a more detailed list would be needed.

After completing the digital map of building footprints and establishing their respective land use types, the building footprint and TAZ maps were overlaid. Then, the validation SD variables for each TAZ fully encompassed by the building footprint map were measured. In this research, the variables measured were number of households.

Two error measurement equations were used to quantify the error between the estimated $l_{ij}$ values and the benchmark values. Equation 23 is the equation for the root mean squared error (RMSE) for experiment $s$ and benchmark $r$.

$$RMSE^s_r = \sqrt{\frac{\sum \sum k \left(l^k_{ij} - \bar{l}^k_{ij} \right)^2}{n_{TAZ}}}$$  \hspace{1cm}  (23)

Where,

$s$    index of the application for which the RMSE is being calculated (1, 2, ...,23),

$n_{TAZ}$ number of TAZs.

Another equation that was used to measure errors is shown in Equation 24 and is the mean absolute error ratio (MAER) for experiment $s$.

$$MAER^s_r = \frac{1}{n_{TAZ}} \sum \sum k \left| \frac{l^k_{ij} - \bar{l}^k_{ij}}{\bar{l}^k_{ij}} \right|$$  \hspace{1cm}  (24)
RESULTS

Results of the experiments performed are twofold in their practical use speaking to the twofold objective of this research. First, the data processing results illustrate the success rates for correct feature extraction as a function of the input data used, and second, the data analysis results which demonstrate the accuracy of demographic estimates obtained through the area-proportional TAZ estimation methodology for each combination of input data.

Data Processing Results

The objective to recommend more functional data types and combinations of data types for the purpose of sensing and extracting building footprint areas has been implemented as described in the Testing and Experiment Design portion of the Methodology section. The results of these various experiments are available in the aggregate overall form as seen in Table 23, and a disaggregate form by visual region in the graphs of

Table 25 TAZ Household estimation RMSE and MAER Error Rates
Table 23 shows the overall results in terms of omission and commission error and percent correctly classified. Variations on the 23 experiments shown in Table 22 are indicated by a superscript lowercase letter and the legend below the table.
Table 23 Accuracy Assessment by Experiment

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Correctly Classified</th>
<th>Commission Error</th>
<th>Omission Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AvistaRGB&amp;DEM&amp;QbirdMulM_7Ag900RoBld</td>
<td>73%</td>
<td>59%</td>
<td>27%</td>
</tr>
<tr>
<td>2 AvistaRGB&amp;DEM&amp;QbirdPANIM_7Ag900RoBld</td>
<td>71%</td>
<td>58%</td>
<td>29%</td>
</tr>
<tr>
<td>3 AvistaRGBQbirdPANIM&amp;MulIM_7Ag900RoBld</td>
<td>70%</td>
<td>59%</td>
<td>30%</td>
</tr>
<tr>
<td>4 AvistDE&amp;M&amp;QbirdPAN&amp;MulIM_7Ag196RoBld</td>
<td>69%</td>
<td>91%</td>
<td>31%</td>
</tr>
<tr>
<td>5 AvistDEMSlp&amp;MuIM_7Ag5RoBld</td>
<td>52%</td>
<td>213%</td>
<td>48%</td>
</tr>
<tr>
<td>6 AvistDEMSlp&amp;QbirdPan&amp;MuIM_7Ag196RoBld</td>
<td>70%</td>
<td>90%</td>
<td>30%</td>
</tr>
<tr>
<td>7 AvistRGB&amp;QbirdPANM_7Ag900RoBld</td>
<td>68%</td>
<td>40%</td>
<td>32%</td>
</tr>
<tr>
<td>8 QbirdMul&amp;AvistDEMM_7Ag9RoBld</td>
<td>75%</td>
<td>178%</td>
<td>25%</td>
</tr>
<tr>
<td>9 QbirdMul&amp;AvistDEMSlpM_7Ag9RoBld</td>
<td>73%</td>
<td>125%</td>
<td>27%</td>
</tr>
<tr>
<td>10 QbirdMuIM_7Ag9RoBld</td>
<td>67%</td>
<td>133%</td>
<td>34%</td>
</tr>
<tr>
<td>11 QbirdPan&amp;MuIM_7Ag9RoBld</td>
<td>69%</td>
<td>91%</td>
<td>31%</td>
</tr>
<tr>
<td>12 QbirdPan&amp;NIRM_7Ag196RoBld</td>
<td>45%</td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td>13 QbirdPanM_7Ag196RoBld</td>
<td>40%</td>
<td>38%</td>
<td>60%</td>
</tr>
<tr>
<td>14 USDA&amp;QbirdMulIM_7Ag49RoBld</td>
<td>62%</td>
<td>84%</td>
<td>38%</td>
</tr>
<tr>
<td>15 USDA&amp;QbirdNIRM_7Ag49RoBld</td>
<td>77%</td>
<td>227%</td>
<td>23%</td>
</tr>
<tr>
<td>16 USDBE3_7Ag49RoBld</td>
<td>55%</td>
<td>111%</td>
<td>45%</td>
</tr>
<tr>
<td>17 USDM_7Ag49RoBld</td>
<td>53%</td>
<td>92%</td>
<td>47%</td>
</tr>
<tr>
<td>18 USDARGB_AvistDEMM_7Ag49RoBld</td>
<td>51%</td>
<td>99%</td>
<td>49%</td>
</tr>
<tr>
<td>19 AvistM_7Ag900RoBld</td>
<td>71%</td>
<td>53%</td>
<td>29%</td>
</tr>
<tr>
<td>20 AvistRGB&amp;QbirdMulIM_7Ag900RoBld</td>
<td>71%</td>
<td>71%</td>
<td>29%</td>
</tr>
<tr>
<td>21 AvistBE2_23Ag900RoBld</td>
<td>71%</td>
<td>71%</td>
<td>29%</td>
</tr>
<tr>
<td>22 AvistRGB&amp;DEMM_7Ag900RoBld</td>
<td>73%</td>
<td>73%</td>
<td>27%</td>
</tr>
<tr>
<td>23 AvistaRGB&amp;DE&amp;M&amp;QbirdNIRM_7Ag900RoBld</td>
<td>76%</td>
<td>76%</td>
<td>25%</td>
</tr>
</tbody>
</table>

a Non-standard image processing kernal tried  
b 1 meter RGB used instead of 0.5 foot  
c NIR band used alone rather than full Multispectral  
d DEM Slope used rather than DEM

Experiment 7 resulted in the best feature extraction in terms of lower error rates and higher classification accuracy based on the percentiles given in Table . Experiments 1, and 19 were next in line, also producing above average feature extractions. In all three cases, the highest resolution imagery is a part (or all) of the feature extraction input. This implies that higher resolution imagery produces better results. Although feature extraction results may be refined to obtain better error rates [23], experiment outputs were analyzed in their raw form to avoid introducing bias.

A disaggregation of the results presented in Table 24 by four visual regions is shown in
Table 25 TAZ Household estimation RMSE and MAER Error Rates

. The experiment number is on the abscissa of each graph with each trend representing one of the four visual classes including 1) a commercial industrial region with less vegetation, larger
buildings and closer spacing; 2) a large multi-family residential area on the University of Idaho Campus with low-lying vegetation and continuous terraced buildings; 3) a new residential region with little mature vegetation, fairly consistent building sizes, and greater building spacing; and 4) an older residential region containing more mature vegetation and tree occlusion, greater variety in building sizes, and generally more dense building spacing.

From

Table 25 TAZ Household estimation RMSE and MAER Error Rates
Applying the TRANSIMS Modeling Paradigm to the Simulation and Analysis of Transportation and Traffic Control Systems, it is apparent that most data combinations used experience the best feature extraction in the commercial industrial visual region. Not surprisingly, the newer residential visual region is generally more correctly classified than the older residential and campus housing. Campus housing and older residential however, appear to do better or worse than one another depending on the input data used.

Table 24 Percentile Ranking of Correctly Classified Percentages Minus Error

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Overall Correctly Classified minus Commission and Omission Error</th>
<th>Rank</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 AvistRGB&amp;QBirdPanM_7Ag900RoBld</td>
<td>-0.035</td>
<td>1</td>
<td>100.00%</td>
</tr>
<tr>
<td>19 AvistM_7Ag900RoBld</td>
<td>-0.108</td>
<td>2</td>
<td>95.40%</td>
</tr>
<tr>
<td>1 AvistaRGB&amp;DEM&amp;QbirdMulM_7Ag900RoBld</td>
<td>-0.120</td>
<td>3</td>
<td>90.90%</td>
</tr>
<tr>
<td>2 AvistaRGB&amp;DEM&amp;QbirdPANIM_7Ag900RoBld</td>
<td>-0.162</td>
<td>4</td>
<td>86.30%</td>
</tr>
<tr>
<td>3 AvistaRGBQbirdPAN&amp;MulM_7Ag900RoBld</td>
<td>-0.183</td>
<td>5</td>
<td>81.80%</td>
</tr>
<tr>
<td>23 AvistaRGB&amp;DEM&amp;QbirdNIRM_7Ag900RoBld</td>
<td>-0.245</td>
<td>6</td>
<td>77.20%</td>
</tr>
<tr>
<td>22 AvistaRGB&amp;DEMM_7Ag900RoBld</td>
<td>-0.267</td>
<td>7</td>
<td>72.70%</td>
</tr>
<tr>
<td>21 AvistBE2_23Ag900RoBld</td>
<td>-0.288</td>
<td>8</td>
<td>68.10%</td>
</tr>
<tr>
<td>20 AvistaRGB&amp;QbirdMulM_7Ag900RoBld</td>
<td>-0.295</td>
<td>9</td>
<td>63.60%</td>
</tr>
<tr>
<td>6 AvistaDEMSlP&amp;QbirdPan&amp;MulM_7Ag196RoBld</td>
<td>-0.508</td>
<td>10</td>
<td>59.00%</td>
</tr>
<tr>
<td>11 QbirdPan&amp;MulM_7Ag900RoBld</td>
<td>-0.529</td>
<td>11</td>
<td>54.50%</td>
</tr>
<tr>
<td>4 AvistaDEMQbirdPan&amp;MulM_7Ag196RoBld</td>
<td>-0.531</td>
<td>12</td>
<td>50.00%</td>
</tr>
<tr>
<td>13 QbirdPanM_7Ag196RoBld</td>
<td>-0.577</td>
<td>13</td>
<td>45.40%</td>
</tr>
<tr>
<td>14 USDA&amp;QbirdMulM_7Ag49RoBlbd</td>
<td>-0.595</td>
<td>14</td>
<td>40.90%</td>
</tr>
<tr>
<td>12 QbirdPan&amp;NIRM_7Ag196RoBld</td>
<td>-0.644</td>
<td>15</td>
<td>36.30%</td>
</tr>
<tr>
<td>9 QbirdMul&amp;AvistaDEMSlPM_7Ag9RoBld</td>
<td>-0.801</td>
<td>16</td>
<td>31.80%</td>
</tr>
<tr>
<td>17 USDA_7Ag49RoBld</td>
<td>-0.866</td>
<td>17</td>
<td>27.20%</td>
</tr>
<tr>
<td>18 USDARGB_AvistaDEMM_7Ag49RoBld</td>
<td>-0.971</td>
<td>18</td>
<td>22.70%</td>
</tr>
<tr>
<td>10 QbirdMulM_7Ag9RoBld</td>
<td>-0.998</td>
<td>19</td>
<td>18.10%</td>
</tr>
<tr>
<td>16 USDABE3_7Ag49RoBld</td>
<td>-1.015</td>
<td>20</td>
<td>13.60%</td>
</tr>
<tr>
<td>8 QbirdMul&amp;AvistaDEMM_7Ag9RoBld</td>
<td>-1.289</td>
<td>21</td>
<td>9.00%</td>
</tr>
<tr>
<td>15 USDA&amp;QbirdNIRM_7Ag49RoBld</td>
<td>-1.719</td>
<td>22</td>
<td>4.50%</td>
</tr>
<tr>
<td>5 AvistaDEMM_5Ag5RoBld</td>
<td>-2.091</td>
<td>23</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 25 TAZ Household estimation RMSE and MAER Error Rates

<table>
<thead>
<tr>
<th>Experiment</th>
<th>RMSE</th>
<th>MAER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 AvistaRGB&amp;DEM&amp;QbirdMulM_7Ag900RoBld</td>
<td>0.4807514</td>
<td>0.4862526</td>
</tr>
<tr>
<td>2 AvistaRGB&amp;DEM&amp;QbirdPANIM_7Ag900RoBld</td>
<td>0.4347072</td>
<td>0.437072</td>
</tr>
<tr>
<td>3 AvistaRGBQbirdPAN&amp;MulM_7Ag900RoBld</td>
<td>0.5364010</td>
<td>0.536401</td>
</tr>
<tr>
<td>4 AvistaDEMQbirdPan&amp;MulM_7Ag196RoBld</td>
<td>1.0547475</td>
<td>1.054747</td>
</tr>
<tr>
<td>5 AvistaDEMM_5Ag5RoBld</td>
<td>0.5187122</td>
<td>0.518712</td>
</tr>
</tbody>
</table>

Experiment 2 applies the TRANSIMS Modeling Paradigm to the Simulation and Analysis of Transportation and Traffic Control Systems, it is apparent that most data combinations used experience the best feature extraction in the commercial industrial visual region. Not surprisingly, the newer residential visual region is generally more correctly classified than the older residential and campus housing. Campus housing and older residential however, appear to do better or worse than one another depending on the input data used.
Figure 33 Feature Extraction Error by Experiment for Each Visual Region

**Data Analysis Results**

Table 25 below, gives the overall error rates for number of households estimated using the area proportional method compared to the Benchmark validation data collected in the field. Experiment 5, a feature extraction experiment based on digital elevation data alone, showed the highest root mean square error. The extraction resulted in much of the lower lying areas being identified as buildings. Much better results were obtained from using slope data derived from the digital elevation data used in Experiment 5 as in Experiment 9 which demonstrated the lowest root mean square error. This is most likely because of the sharp elevation change or “slope” from grade elevation to building tops. The mean absolute error ratio was also lowest for Experiment 9. This is significant since the Experiment 9’s feature extraction was not the best considering error. Its performance in the household estimation may be due to areas of commission error compensating for omission error, resulting in building footprint area proportions quantitatively...
close to the validation building footprint areas. In any case the performance of the highest resolution red-green-blue data, Avista Utilities RGB in the feature extraction and the Slope data derived from the Avista Digital elevation data suggests that an experiment including the combination of the two may produce highly favorable results. Such an experiment was not a part of this study and is recommended for future research.

Experiment 12 exhibited the largest mean absolute error ratio by a wide margin. This experiment included the use of Near Infrared satellite data. Feature extraction using the Near Infrared data also performed more poorly than average. The Near Infrared spectra is commonly used for sensing the health and presence of vegetation. Because vegetation is among the principal occlusions limiting the aerial view of building rooftops, it was thought that the Near Infrared data would prove very useful. As in the case of the digital elevation data, it is possible that a derived product from the NIR data such as a vegetation index by pixel might prove a better integration of this type of data.
CONCLUSIONS AND RECOMMENDATIONS

**Recommended Input Data and Data Combinations**

From the feature extraction results it is apparent that combinations of two datasets bring about the best extraction results. The high resolution (6-inch) Red-Green-Blue combined with the .68 meter Panchromatic satellite imagery or the 8ft resolution Multispectral imagery including Red, Green, Blue, and Near Infrared bands, produced the best feature extraction. Commercial-Industrial areas have the best classification rates with relatively low error rates. The older residential visual region, campus housing visual regions, and similar areas exhibit more data dependent qualities than the “cleaner” commercial industrial and new residential regions. This is because of the increased variety of reflectances that occur in the more visually busy areas. Under these conditions more spectral coverage or a greater number of bands can help to differentiate building rooftops from more similar and nearby surroundings. Increased vegetation produces more shadowed areas so that remote sensing instruments receive more erratic or no signal.

**Traffic Analysis Zone Dissaggregation**

Household estimates for disaggregate TAZs derived using the Area-Proportion methodology described generate reasonable estimates with moderate error. Minor discrepancies between the Census demographic measurements and the field collected validation data form an error baseline. Without interviews at each location, as occur in Census data collection, field researchers must rely on observation and estimation. In this research the results showed significant improvement with corrections to the validation data. The households measured for validation purposes and the Census household measurements are generally very close. Other techniques mentioned in the Future Research section may provide further insight to the evaluation of the area proportional approach. Notwithstanding, the Area-Proportional technique could also be applied to provide disaggregate estimates of other types of demographic data such as employment data. See the Future Research section for further discussion.
Future Research

As the Methodology section suggests, principles of the area proportion approach could be used to distribute various demographic data to smaller traffic analysis zones. A possible avenue for future exploration would be to distribute employment and/or household demographic data to a general network including regional employment data and area proportional distribution of Census employment data to TAZs in commercial and industrial land-use zone areas.

Other areas for further research include exploration of validation data collection techniques and alternative data inputs as a means for evaluating remotely sensed features and household estimates.
REFERENCES


