USE OF WIDE-AREA MOTION IMAGERY (WAMI) FOR TRANSPORTATION PLANNING AND OPERATIONS

Morgan State University
The Pennsylvania State University
University of Maryland
University of Virginia
Virginia Polytechnic Institute & State University
West Virginia University

The Pennsylvania State University
The Thomas D. Larson Pennsylvania Transportation Institute
Transportation Research Building  University Park, PA 16802-4710
Phone: 814-865-1891  Fax: 814-863-3707
www.mautc.psu.edu
FINAL REPORT

USE OF WIDE-AREA MOTION IMAGERY (WAMI) FOR TRANSPORTATION PLANNING AND OPERATIONS

Kathleen L. Hancock, PE, PhD
Associate Professor
The Charles Via, Jr. Department of Civil and Environmental Engineering
Systems Urban Mobility Laboratory, National Capital Region
Virginia Tech

Md Rauful Islam
Graduate Research Assistant
The Charles Via, Jr. Department of Civil and Environmental Engineering
Systems Urban Mobility Laboratory, National Capital Region
Virginia Tech

February 2016
VT-2013-03
### 16. Abstract

Wide-area motion imagery (WAMI), in combination with PVLabs’ integrated Tactical Content Management System spatio-temporal capability, automatically identifies and captures every vehicle in the video view frame, storing each vehicle with a discrete ID, track ID, and time-stamped location. This unique data capture provides comprehensive vehicle trajectory information for an extended period of time, approximately three continuous hours, over a relatively large area, approximately four square miles. This report presents the results of an initial exploration of the use of this data to support transportation planning and operation activities. A subset of the data was extracted, cleaned, validated, and processed for use in calibrating car-following submodels for use in microscopic simulations. A flexible multi-dimensional filtering method was developed to evaluate and filter the data provided by PVLabs to extract valid vehicle tracks. A 10-minute sample of tracks was validated using imagery frames from the video. Resulting tracks were map-matched to roads and individual lanes to support macro and microscopic traffic characteristic extraction. A spatio-temporal trajectory data model was developed to efficiently store vehicle tracks along with the traffic model attributes required for modeling car following behavior. The final processed dataset includes all vehicles and their trajectories for an area of approximately 4-square miles that includes a dense and complex urban network of roads over a three-hour period. Several car-following and lane-changing models were reviewed in detail as were calibration and validation of these models. Both a global deterministic and stochastic calibration process were applied to a 15-minute period for the portions of tracks that occurred on two principle one-way arterials. The preliminary results indicated that the global deterministic calibration produced results similar to previous work. Results of the stochastic calibration showed a significant reduction in the variance from the initially assumed value indicating that the parameters are converging to an expected distribution and have the potential to improve model performance for the test conditions.
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 U.S. Department of Transportation’s University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This project was supported in part by the Virginia Center for Transportation Innovation and Research (VCTIR) and PVLabs. The authors would like to thank Declan Keogh and PV Labs for providing the WAMI data and for their support throughout this project and Michael Fontaine for his input and guidance.

Some of the material presented in this report was provided as an unpublished interim report to VCTIR.
ABSTRACT

Wide-area motion imagery (WAMI), in combination with PVLabs’ integrated Tactical Content Management System spatio-temporal capability, automatically identifies and captures every vehicle in the video view frame, storing each vehicle with a discrete ID, track ID, and time-stamped location. This unique data capture provides comprehensive vehicle trajectory information for an extended period of time, approximately three continuous hours, over a relatively large area, approximately four square miles. This report presents the results of an initial exploration of the use of this data to support transportation planning and operation activities.

A subset of the data was extracted, cleaned, validated, and processed for use in calibrating car-following submodels for use in microscopic simulations. A flexible multi-dimensional filtering method was developed to evaluate and filter the data provided by PVLabs to extract valid vehicle tracks. A 10-minute sample of tracks was validated using imagery frames from the video. Resulting tracks were map-matched to roads and individual lanes to support macro and microscopic traffic characteristic extraction. A spatio-temporal trajectory data model was developed to efficiently store vehicle tracks along with the traffic model attributes required for modeling car following behavior. The final processed dataset includes all vehicles and their trajectories for an area of approximately 4-square miles that includes a dense and complex urban network of roads over a three-hour period.

Several car-following and lane-changing models were reviewed in detail as were calibration and validation of these models. Both a global deterministic and stochastic calibration process were applied to a 15-minute period for the portions of tracks that occurred on two principle one-way arterials. The preliminary results indicated that the global deterministic calibration produced results similar to previous work. Results of the stochastic calibration showed a significant reduction in the variance from the initially assumed value indicating that the parameters are converging to an expected distribution and have the potential to improve model performance for the test conditions.
# TABLE OF CONTENTS

DECLARATION .............................................................................................................................. III  
ACKNOWLEDGMENTS ................................................................................................................. III  
ABSTRACT .................................................................................................................................. IV  
TABLE OF CONTENTS ............................................................................................................... V  
LIST OF FIGURES ..................................................................................................................... VII  
LIST OF TABLES ....................................................................................................................... VII  
LIST OF ACRONYMS ............................................................................................................. VIII  
INTRODUCTION .......................................................................................................................... 1  
  Study Area and WAMI Data Overview ................................................................. 2  
  Vehicle Trajectory Extraction and Validation .................................................. 3  
  Modeling Vehicle Movements in Microscopic Simulation ........................ 4  
  Calibration and Validation of Existing Simulation Models ......................... 4  
WIDE AREA MOTION IMAGERY (WAMI) DATA ................................................................. 7  
  Background of Using Video and Trajectory Data in Microscopic Model Development .... 7  
  Description of WAMI Data ...................................................................................... 9  
    Data Attributes ....................................................................................................... 11  
  Trajectory Data ........................................................................................................... 12  
  Vehicle Track Validation ......................................................................................... 12  
    Filtering ................................................................................................................... 12  
    Map Matching ....................................................................................................... 15  
TRACK DATA MODEL .............................................................................................................. 18  
  Background of Existing Data Models for Storing Trajectory Data ................... 18  
  Traffic Data Extraction ............................................................................................... 21  
    Macroscopic Data ................................................................................................. 22  
    Microscopic Data ................................................................................................. 24  
MICROSCOPIC TRAFFIC SUB-MODELS ............................................................................... 26  
  Background ............................................................................................................... 26  
    Car-Following Models .......................................................................................... 26  
    Lane Changing Model ......................................................................................... 33  
  Car Following Model Calibration and Validation ............................................ 46  
    Data Sampling and Smoothing ....................................................................... 47  
    Model Calibration ............................................................................................... 48
Preliminary Results ................................................................................................................... 52
  Deterministic Calibration ..................................................................................................... 53
  Stochastic Calibration ......................................................................................................... 53

CONCLUSIONS........................................................................................................................... 56

RECOMMENDATIONS .............................................................................................................. 57

REFERENCES ............................................................................................................................. 59
LIST OF FIGURES

Figure 1. Location of data collect in Hamilton, Ontario .......................................................... 2
Figure 2. Example of Vehicle Tracks with Internal Origins and Destinations ....................... 3
Figure 3: Aircraft with Mounted Camera Platform – platform outlined. ............................... 10
Figure 4: Tracking Point Data within the Collect Area in Downtown Hamilton, Ontario ......... 11
Figure 5: Temporal Trajectory Bins ...................................................................................... 14
Figure 6: Histogram of Track Duration .................................................................................. 15
Figure 7: Map Matching Algorithm ....................................................................................... 16
Figure 8: Illustration of Map Matching Algorithm .................................................................. 17
Figure 9. 3-D Representation of Trajectory Data. .................................................................. 19
Figure 10: Network Data Schema of Trajectory Data ............................................................. 20
Figure 11: Data Schema for Storing tracking data modified from Brakatsoulas et al. ............ 21
Figure 12: Origin and Destination of Individual Tracks.......................................................... 22
Figure 13: Mean Speeds and Standard Deviations on Arterials .............................................. 23
Figure 14: 15-minute Traffic Flow on Hamilton Arterials ....................................................... 24
Figure 15: Microscopic Time Series Traffic Flow Measures for a Sample Leader-Follower Vehicle Pair. ...................................................................................................................... 25
Figure 16: Classification of lane-changing models ................................................................. 34
Figure 17: Calibration and Validation Steps of a Car-Following Model using WAMI data ....... 48
Figure 19: Definition of Inverse Problem ............................................................................. 50
Figure 18: Framework for Calibration ................................................................................... 50
Figure 20: Bayesian estimation process using Markov Chain Monte Carlo Method .......... 51
Figure 21: Statistics Assoc with Coefficients Resulting from Stochastic Calibration Process .... 54

LIST OF TABLES

Table 1. Rankings for use in Track Filter .................................................................................. 13
Table 2: Distribution of Critical Distance ............................................................................... 37
Table 3. Lane Choice Utility Function .................................................................................... 38
Table 4: Path Influence Factor ............................................................................................... 39
Table 5: Lane Choice Utility Function .................................................................................... 40
Table 6: Non Linear Gap acceptance Parameter used in TransModeler ............................... 42
Table 7: Linear Gap acceptance Parameter used in TransModeler ........................................ 42
Table 8: NGSIM Gap acceptance Parameter used in TransModeler ....................................... 42
Table 9: Linear Helly Car Following Model Coefficient Estimates ......................................... 53
Table 10: Coefficients from Stochastic Calibration of the Helly Car Following Model .......... 53
# LIST OF ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>Action Point model</td>
</tr>
<tr>
<td>CA</td>
<td>Collision Avoidance model</td>
</tr>
<tr>
<td>CBD</td>
<td>Central Business District</td>
</tr>
<tr>
<td>DBMS</td>
<td>Database management systems</td>
</tr>
<tr>
<td>DGPS</td>
<td>Differential Global Positioning Systems</td>
</tr>
<tr>
<td>DLC</td>
<td>Discretionary Lane Change</td>
</tr>
<tr>
<td>FHWA</td>
<td>Federal Highway Administration</td>
</tr>
<tr>
<td>FLC</td>
<td>Forced Lane Change</td>
</tr>
<tr>
<td>GF</td>
<td>Generalized Force model</td>
</tr>
<tr>
<td>GHR</td>
<td>Gazis-Herman_Rothery car-following model</td>
</tr>
<tr>
<td>GM</td>
<td>General Motors model</td>
</tr>
<tr>
<td>GOF</td>
<td>Goodness of Fit</td>
</tr>
<tr>
<td>HOV</td>
<td>High Occupancy Vehicle</td>
</tr>
<tr>
<td>IDM</td>
<td>Intelligent Driver Model</td>
</tr>
<tr>
<td>IMSE</td>
<td>Integrated Mean Square Error</td>
</tr>
<tr>
<td>JND</td>
<td>Just Noticeable Distance</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>MLC</td>
<td>Mandatory Lane Change</td>
</tr>
<tr>
<td>MoP</td>
<td>Measure of Performance</td>
</tr>
<tr>
<td>NAD</td>
<td>North American Datum</td>
</tr>
<tr>
<td>NDS</td>
<td>Naturalistic Driving Study</td>
</tr>
<tr>
<td>NGSIM</td>
<td>Next Generation SIMulation</td>
</tr>
<tr>
<td>OD</td>
<td>Origin-destination</td>
</tr>
<tr>
<td>OV</td>
<td>Optimal Velocity model</td>
</tr>
<tr>
<td>PPI</td>
<td>Path Plan Impact</td>
</tr>
<tr>
<td>SHRP</td>
<td>Strategic Highway Research Program</td>
</tr>
<tr>
<td>TCMS</td>
<td>Tactical Content Management System</td>
</tr>
<tr>
<td>UTM</td>
<td>Universal Transvers Mercator coordinate system</td>
</tr>
<tr>
<td>WAMI</td>
<td>Wide Area Motion Imagery</td>
</tr>
<tr>
<td>WGS84</td>
<td>World Geodetic System 1984</td>
</tr>
</tbody>
</table>
INTRODUCTION

As transportation agencies face increased demands for limited resources, more reliance is placed on advanced technologies and computational tools to support decision making for planning, evaluating and operating transportation facilities. Many of these tools rely on representative models of aggregate traffic flow measures and/or microscopic measures of vehicle or driver behaviors. Typically, these were developed, calibrated and validated using sample data which were generally fixed in time or space or both. Although new data sources, such as mobile sensors and naturalistic driver data, are available, capturing information about every individual vehicle track over a large area across an extended period of time has not previously been readily available to transportation researchers or practitioners.

The ability to collect airborne persistent wide-area motion imagery (WAMI) data for use in intelligence and reconnaissance was originally developed to support military activities. Enhanced stabilization techniques, advanced camera technologies, and longer battery life have improved the successful detection and tracking of stationary and moving objects over large areas for extended periods of time (Puckrin et al. 2012). Use of these spectrometric tools is becoming more common in civil applications including search and rescue, geological survey, pollution monitoring, forest fire detection and monitoring, and combustion studies (Puckrin et al. 2012).

Current examples of public sector transportation information derived from WAMI include intersection counts, vehicle classification, and origin-destination counts, all of which consist of counts of vehicles at fixed locations in the view frame. A powerful potential, which the results of this project begins to leverage, is the capture and use of vehicle trajectory data, which provides paths of individual vehicles linked to corresponding spatio-temporal derived characteristics. Because of the comprehensive nature of WAMI data for capturing microscopic vehicle behavior, it has the potential for use in calibrating and validating existing microscopic vehicle models as well as developing new models. Several possibilities were explored including:

- extracting vehicle behavior characteristics for use in calibrating and validating micro-simulation models
- understanding and modeling vehicle behavior related to microscopic traffic flow characteristics
- understanding and modeling vehicle movements with internal origins and/or destinations
- generating link volumes that are explicitly associated with specific O-D pairs for use in improving and validating traffic assignment and simulation
- associating vehicle trips with lanes
- identifying lane changes by number and location on a road segment
• calculating traffic characteristics along a road segment or route, such as speed or gap size as a function of location and time for individual vehicles, platoons, vehicle types, or all vehicles for understanding, establishing and measuring urban traffic performance measures

Given the comprehensive nature and size of the data resulting from WAMI, this research narrowed the scope to the first two areas for more detailed exploration, extraction and representation of vehicle movements.

**Study Area and WAMI Data Overview**

The data for this project was collected and extracted by PVLabs using their airborne data collection hardware and proprietary computer vision extraction and processing system. The data collect consisted of two three-hour periods, one during the morning commute and one during the evening commute, covering approximately four square miles of downtown Hamilton, Ontario, Canada shown in Figure 2. Hamilton is a midsize city on the coast of Lake Ontario and the study area covers the central business district as well as several nearby neighborhoods, the port facility, and a large freight rail terminal and switching yard. The street network consists of a dense grid of urban arterials, collectors and local roads.

The processed data consist of point locations of every vehicle detected from each frame of the high-resolution video at a frequency of 1 Hz or 1 frame per second. During the processing, each point or vehicle was associated with its unique track across frames in which that point or vehicle appeared. Figure 1 shows an example of the PVLabs processed trajectory data overlaid on imagery showing individual vehicles and their corresponding tracks generated from the previous frames. More detail about the data is provided in the next section.

In addition to the extracted data, PVLabs provided frame-by-frame images from the video for a 15-minute period, resulting in 900 jpgs, to support validation of vehicle tracks and map matching as discussed in the subsection on Vehicle Track Validation.

![Figure 1. Example of Vehicle Tracks with Internal Origins and Destinations (source: PVLabs)](image)
Vehicle Trajectory Extraction and Validation

Because one of the intended uses of the data is to validate and calibrate existing microscopic traffic behavior models used in micro-simulation, having an accurate and complete data set of vehicle tracks is critical. Based on an initial review of the original dataset, it was clear that a large number of false positives were included. After overlaying the data on the
imagery, several additional issues were identified including dropped vehicles in urban canyons and under tree canopies, split trajectories when vehicles were lost, multiple points for vehicles stopped at intersections resulting in a “bumblebee” pattern in the trajectories, and very short trajectories. To ensure the data were appropriate for this research, a comprehensive filtering process was applied to the data which was then validated using the imagery as discussed by Islam et al (2016) and summarized in the subsection on Vehicle Track Validation.

Modeling Vehicle Movements in Microscopic Simulation

Microscopic traffic simulation is a commonly used tool for planning and evaluating improvements to transportation facilities. Its use varies from basic traffic-signal timing improvements to evaluating impacts of state-of-the-art ITS implementation to analyzing system wide variations resulting from changes to facility and/or traffic characteristics.

A micro-simulation model consists of several sub-models intended to predict different driver/vehicle characteristics. Two important sub-models are the car-following model and the lane-changing model which describe the longitudinal and lateral movements of individual vehicles both independently and with respect to each other. These models calculate the movement of a vehicle on the network at discrete time intervals based on a given set of driver/vehicle and network characteristics. Traditionally, driver/vehicle characteristics are based on human factors which are difficult to extrapolate from limited controlled experiments. The data collected by PVLabs captures the track of every vehicle on the network over an extended period of time which can be used to quantify and contextualize this behavior. The focus of this phase of the exploratory research is to understand current models and identify options for improving or modifying these models using WAMI data.

Calibration and Validation of Existing Simulation Models

Because of the increasing reliance on simulation, improving calibration and validation of models is gaining importance. Difficulty in obtaining comprehensive microscopic behavior data has been considered a major drawback in effectively understanding and simulating traffic (Punzo, Formisano et al. 2005). Traditionally, calibration and validation efforts have concentrated on matching macroscopic measures (flow, speed, and capacity) either from field experiments or generated by the micro-simulation models with representative field data. Much of the literature related to car following and lane changing is concentrated on theoretical properties of the models and their relationship to macroscopic flow characteristics (Brackstone et al. 1999; Punzo et al. 2005) while empirical verification of the underlying assumptions have not been extensively researched due to lack of accurate and unbiased time series data (Brackstone et al. 1999).
There continues to be a lack of established methods for calibration of model parameters. Because of limitations in earlier data collection technologies, parameters of traffic simulation models could not be measured directly. Traditionally, their estimation has been indirectly determined by setting up an appropriate optimization framework that attempts to reproduce a real transportation system under prescribed conditions. Additionally, concerns have been raised over the impact of measurement error on parameter estimation and the requirement for and limitations of data smoothing of time series data (Ossen et al. 2008). It has also been suggested that trajectory data gathered during one type of traffic condition may not provide enough information to reliably calibrate the behavioral parameters for other conditions (Ossen et al. 2009). Estimated parameters using optimization methodologies have also been shown to produce significantly different results (Punzo et al. 2012).

**Car following.** The first documented car following model was developed in the 1950’s using speed and distance data collected by hard-linked vehicles on a test track (Chandler et al. 1958). More recently, speed, acceleration and inter-vehicle gap data between leader and follower cars were obtained on a test track in Japan using differential global positioning systems (DGPS) (Brockfeld et al. 2004). These types of experiments use pattern matching techniques to establish deterministic driver behavior which may not be a good representation of real-world behavior. In an attempt to improve upon this estimation, a probabilistic approach has also been explored. Calibration of the car following model using a Bayesian belief network has been proposed to more accurately represent the probabilistic nature of driver behavior (Miska et al. 2006). Miska et al. used microscopic data from equipped vehicles and remote sensing. Although this remotely sensed data is similar to the WAMI data, its geographic extent, duration, and comprehensiveness is much less. Other methodologies used for calibrating the car following model include the use of loop detector data of speed, density and occupancy (Rakha et al. 2007, Rakha and Wang 2009), radar instrumented vehicles (Kesting et al. 2008) and time series data from GPS instrumented vehicles (Montanino et al. 2012).

**Lane changing.** Several methods using trajectory data have been proposed for development and calibration of lane-change models. The Next Generation SIMulation (NGSIM) program used terrestrial-mounted cameras to collect video of vehicles on a three individual road segments and produced vehicle trajectory data used to develop and calibrate the target lane model (Loveland et al. 2008). A primary limitation of this program is its limited coverage. Daamen et al. (2010) used a helicopter to collect vehicle trajectory data at two locations in the Netherlands. The same data collection technique was used by Hoogendoorn et al. (Schreuder et al. 2003). Knoop et al. (2012) used microscopic vehicle trajectory data collected from A270 near Eindhoven, Netherlands, and M42 near Birmingham, U.K., for their study to quantify the number of lane changes. The M42 study used loop detector data to estimate the number of lane changes using a splash-over effect of loop detector counts.
Irrespective of differences in outcome, these calibration studies are important because they provide the opportunity to assess the performance of existing lane-changing and car-following models. The outcomes of these studies are used to obtain empirical insights into lane-changing and car-following behavior. Insights can include the extent of heterogeneity between drivers (Ossen et al. 2005) and the degree of multi- or spatio-temporal variability of driver anticipation (Hoogendoorn et al. 2006). Each methodology has its own strengths and weakness, but a common weakness is that none cover a relatively large geographical area consisting of different types of transportation facilities over an extended period of time and all use samples instead of the entire population.

While numerous advances have been made in the field of traffic simulation, a lack of consensus about the underlying sub-models continues to result in different approaches and solutions to modeling traffic facilities and conditions. The objectives of this study were to identify what macroscopic and microscopic traffic characteristics could be extracted or calculated from WAMI data and how these data could be used in micro-simulation model development, calibration, and validation. This research analyzed the state of the practice of the use of WAMI data and considered methods for using WAMI data to improve micro-simulation models for use in traffic operation and management. Preliminary calibrations were performed on a small subset of data for a car-following submodel.
This research explores the use of WAMI data for calibrating and validating microscopic traffic models. A summary of the state of the practice in video and trajectory data is presented followed by a detailed description of the data that was used for this research. This is followed by a summary of the filtering and validation that was performed on the WAMI data.

Background of Using Video and Trajectory Data in Microscopic Model Development

The methodology of estimating trajectory information can be broadly classified into two categories; i) post processing of vehicle positional data collected from on-board sensors, usually GPS devices (instrumented vehicle) and ii) post processing of video data (remote sensing). Processing of vehicle positional information obtained from on-board sensors is relatively straightforward due to availability of GPS devices and geospatial tools. Sensor tracking data suffers from processing challenges including measurement errors caused by inherent limitations of the sensors and limitations of the geospatial tools which are not designed for mining large temporal datasets requiring complex data filtering and data smoothing (Punzo et al. 2009, Punzo et al. 2012). Although video data is capable of providing more comprehensive information about vehicle movement and driver behavior, video data processing is challenging and the quality of processed data can be problematic. The majority of video data is also spatially restricted, typically to a particular stretch of a road, because the cameras are usually mounted on fixed structures such as a signal support, signal gantry or tall building adjacent to the road. Each of these categories are discussed in more detail in the following paragraphs.

Trajectory data for use in model development, calibration, and/or validation using on-board sensors are mostly collected under controlled test conditions using a limited number of vehicles. Gurusinghe et al. (2002) collected GPS data from a platoon of ten vehicles on two parallel sections. With a sampling rate of data collection of 0.1 second, the authors determined that this high resolution data can facilitate accurate car following analysis. The same dataset was then used by Ranjitkar et al. (2003) to examine the stability of car-following models with increased platoon size and found large intra and inter-vehicle variation in reaction time. The authors concluded that the stimulus-response based car following model might be too simple to explain car following behavior. Brockfeld et al (2004) collected GPS track data from nine vehicles on a test track in Japan for calibration and validation of ten different car-following models. The authors concluded that the differences between individual drivers are larger than the differences between different models and that because these models were developed for specific conditions, they are not capable of generalizing to other situations. Ranjitkar et al (2004) used a similar data collection method to measure performance of six different car-following models and found different driving conditions have less impact on model parameter estimation than an optimization approach to calibrating the model. The authors also found that headway
calculations from position data might be noisier than GPS-derived speed data. The authors then used the same data to calibrate the GHR model using a genetic algorithm based optimization technique (Ranjitkar et al. 2005). Because these data were collected on a test track, they do not reflect complex driving conditions that occur on actual roadways. To overcome this shortcoming, Punzo and Simonelli collected GPS data from instrumented vehicles under real traffic conditions in Naples, Italy (Punzo et al. 2005). Four microscopic car-following models were calibrated for individual drivers then subsequently validated by time series trajectory output from the follower car which was estimated by the car following models using trajectory input of the leader car. The authors concluded that choice of the performance measure is very important to the calibration and that selection of a different performance measure may lead to different parameter estimation. The authors also reported validation effort results similar to those reported by Ranjitkar et al (2004) and Brockfeld et al (2004) where intra-personal and inter-personal variability was greater than variability of different model output.

Vehicle trajectory data obtained from analyzing video has primarily been used for the microscopic model development. The method of video data collection can be broadly classified into three categories: i) terrestrial mounted cameras (those mounted on stationary roadside structures, ii) cameras mounted on airborne platforms and iii) cameras mounted on vehicles.

The most noteworthy study and example of terrestrial video data collection was done in the Next Generation Simulation (NGSIM) program (Loveland et al. 2008). Video was collected at three locations in California; I-80 in the San Francisco Bay area in Emeryville, US-101 in Los Angeles and Lankershim Boulevard in Los Angeles. The length of each roadway segment was between 500 and 640 meters. Data was collected for different durations and times of the day. A customized software application, NG-VIDEO, was used to track the position of every vehicle at a frequency of 0.1 second. Some analysis of NGSIM data have shown processing issues such as vehicle overlaps. The NGSIM data have been used for calibration and validation of many microscopic traffic flow models including the Target lane-changing model (Choudhury 2005), cooperative/forced freeway merging behavior (Choudhury et al. 2009, Wan et al. 2014), arterial lane selection model for urban intersections (Choudhury et al. 2008), and lane position and modeling lane changing execution behavior for different vehicle classes (Moridpour et al. 2010, Aghabayk et al. 2011). This data has also been used in calibrating existing models (Monteil et al. 2014), as well as determining the effect of correlation between model parameters (Kim et al. 2011).

Ozaki (Ozaki 1993) used video data collected from the 32nd floor of a city office building and found that driver’s reaction times were not stable over time, space, or driving condition. Ahn et al (2004) used video data collected from a building near two signalized intersections in Oakland, California, for verification of Newell’s simplified car-following theory.
Treiterer and Myers used an aerial camera mounted on a helicopter that captured images at a mean interval of 1.0 second and observed a hysteresis phenomenon between traffic flow and density for individual lanes (Treiterer 1974). Ossen et al (2005) used a similar approach to collect data from Freeway A2 in Utrecht, in the Netherlands to estimate the parameters for the Gazis–Herman–Rothery (GHR) car-following model and found that car following behavior varies widely within the observed driver population. The same dataset was used in calibration and parameter estimation of other car following models, (Hoogendoorn et al. 2010, Hoogendoorn et al. 2011) and for determining heterogeneity in car following characteristics (Ossen et al. 2011). A similar approach was used for the development of the multi-anticipated car following model by Hoogendoorn et al (2006).

A primary advantage of terrestrial and airborne video data collection is the ability to observe vehicles without influencing drivers. Another method of collecting video data is by installing multiple video cameras inside a vehicle to collect driver characteristic data (e.g., driver’s facial data, interactions with the dashboard, road sign, traffic signal other conditions in the vehicle) along with the surrounding driving conditions (e.g., forward, rear, and right-side views). This method is used in the Strategic Highway Research Program (SHRP) 2-Naturalistic Driving Study (NDS) (Hanowski et al. 2006, Board 2013). Disadvantages of this method include the possibility that drivers may be influenced by the cameras and that cameras are only installed in a small sample of the vehicle population. Because the data from this study is just becoming accessible, few results are available in the literature at this time.

Data from high resolution video and the subsequent knowledge discovery process is limited to academia and are mostly used for research and development purposes. This type of data is shifting from research and development to private sector deliverables accessible to transportation professionals. Skycomp indicates that it provides its customers with traffic flow characteristics including OD matrices, link level volumes, level of service, travel times, speeds, bottleneck characteristics, lane use compliance, parking entry/exit information and other mobility and performance measures (Skycomp 2014).

**Description of WAMI Data**

PVLabs is an advanced imaging solutions company specializing in the design and development of turnkey aerial imaging systems. The basic principle behind PV Labs wide area motion imagery data collection system is to capture continuous high resolution video imagery of a wide area using an airborne platform, either fixed-wing or rotor as shown in Figure 3. PVLabs uses Gen-V Technology for steering and stabilization of a camera fixed to a Gyrostabilized Gimbal Airborne Platform, or gimbal, mounted on the aircraft. The WAMI data used in this study was collected in July, 2013 by PVLabs for an area of approximately four square miles covering the central business district of Hamilton, Ontario, Canada. This area consists of
10 different types of roadways including major and minor arterials, collectors, and local streets as well as both one and two way facilities. The area also includes a railroad and shipping yard along Lake Ontario. The collection occurred over two three-hour periods, one during the morning peak and one during the afternoon peak. The frequency of image capture is 1 Hz or one frame per second. The processed WAMI data consists of one record per vehicle per second which resulted in 6.58 gigabytes for the morning peak data collect and 6.92 gigabytes for the afternoon peak. Figure 4 shows individual time-stamped vehicle points for a 15-minute interval during the evening peak period.

Once the video is collected, the WAMI data processing system consists of geospatially referencing the images, and detecting and classifying vehicles. This image registration and the detection algorithms are proprietary to PVLabs and details are beyond the scope of this study. The positional accuracy of tracked vehicles is dependent on the image registration system and errors in either the registration or the detection algorithm result in errors in the tracks and resulting microscopic parameters. As a result, special attention is required in evaluating data accuracy and data smoothing which is one of the considerations of this study (Ossen et al. 2008).
In addition to the tagged vehicle location and associated trajectory, several other attributes related to vehicle trajectory data are provided by PVLabs. A detailed list includes:

1. **Frame No**: Unique identification number for every instance the image is taken
2. **Frame Time**: Time that the image was taken. This is also the time of detection of a vehicle at particular location.
3. **Vehicle ID**: Unique identification number for every vehicle identified.
4. **Track ID**: Unique track number associated with each vehicle.
5. **Vehicle Size in Pixel**: Number of Pixels the vehicle occupies in the image.
6. **Speed in Pixel**: Instantaneous speed of the vehicle calculated by comparing the position of the vehicle across two consecutive frames.
7. **Target centroid Latitude**: Latitude of the centroid of the vehicle. The centroid is the geometric centroid of the pixels identified as a vehicle.
8. **Target Centroid Longitude**: Longitude of the centroid of the vehicle. The centroid is the geometric centroid of the pixels identified as a vehicle.

Figure 4: Tracking Point Data within the Collect Area in Downtown Hamilton, Ontario (source: Islam et al, 2016).
9. **Target Centroid Speed:** Speed in Kilometers per Hour of the centroid of the vehicle.

10. **Target Centroid Heading:** Bearing of the vehicle’s heading.

Although Vehicle Size is provided with the data, efforts to classify vehicles without manual review have been unsuccessful due to the large variation in pixel size even for the same vehicle along an individual track. This variation in pixel size occurs due to changing vehicle and camera positions at each image capture and to the location of the pixel with respect to the focal centroid of the image.

**Trajectory Data**

Vehicle tracks make up the underlying data required to calculate microscopic traffic characteristics as discussed in the section on Traffic Data Extraction. Therefore, it is important to ensure that the tracks included in any further development are valid and accurately represent vehicle paths and vehicle locations with respect to each other. This includes filtering the data to remove (1) false positives, i.e. points and tracks that are not representative of vehicles, and (2) vehicles or track segments that should not be included in the analysis, i.e. vehicles in parking lots.

Another issue is related to vehicles that are stopped at traffic control devices or in the traffic flow. Although stable, the camera continues to move as it records imagery and the resulting centroid of stationary objects shifts from frame to frame resulting in a “bumble bee” path. Although not part of the track validation, additional processing is required to accurately reflect stopped vehicle behavior in the resulting track data.

**Vehicle Track Validation**

Track validation for this study consists of two steps, (1) filtering the data and (2) ensuring the resulting tracks occur on the road network, commonly referred to as map-matching. For microscopic measures, the tracks need to be matched to the actual lane of travel not just to the link representation as is done with standard map-matching techniques. Map matching also allows for calculation of several macroscopic measures as discussed in the subsection on Traffic Data Extraction. Each process is summarized in the following sections. Details are included in Islam et al (2016).

**Filtering**

Filtering is used to remove false positives and vehicle tracks that should not be used for determining the specified microscopic traffic characteristics. This research developed a filtering technique that considers both the spacio-temporal and trajectory characteristics of individual
vehicles and uses a constraint-based, multi-dimensional query-driven process which is easy to understand and apply.

The dimensions used in the filter include temporal, spatial location, and track sinuosity. Each category is evaluated according to rule-based queries and assigned an ordinal ranking from 1 to 4 as shown in Table 1 for this research and discussed in the corresponding paragraphs below. The rankings are then weighted using an absolute or relative scale and combined to obtain the final filter value as shown in Equation 0.

\[ A_t = w_{temp} \times R_{temp} + w_{spt} \times R_{spt} + w_{sin} \times R_{sin} \]  

(0)

Where,

\[ w_{temp} = \text{Weight of temporal attribute} \]
\[ R_{temp} = \text{Ranking of temporal attribute} \]
\[ w_{spt} = \text{Weight of spatial attribute} \]
\[ R_{spt} = \text{Ranking of spatial attribute} \]
\[ w_{sin} = \text{Weight of sinuosity attribute} \]
\[ R_{sin} = \text{Ranking of sinuosity attribute} \]

An advantage to this approach is its flexibility and simplicity. The ranking categories, ordinal rankings, and weighting can be easily adjusted based on characteristics of the data and purpose of the analysis.

Temporal Filtering. Tracks with different travel times are assigned an ordinal value based on temporal binning. Tracks of shorter duration have a higher probability of being a false detection and are assigned a lower rank. The bins and ranks used for this research are provided in Table 1. Figure 5 shows the unfiltered tracks in each bin.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal</td>
<td>Length of track in seconds</td>
<td>&lt; 10 sec</td>
<td>11 – 30 sec</td>
<td>31 – 60 sec</td>
<td>&gt; 60 sec</td>
</tr>
<tr>
<td>Spatial</td>
<td>Percent of points in road polygon boundaries</td>
<td>&lt; 25 %</td>
<td>25% – 50%</td>
<td>51% – 75%</td>
<td>&gt; 76%</td>
</tr>
<tr>
<td>Sinuosity</td>
<td>Length of curve / Euclidean distance from O to D</td>
<td>&gt; 2.0</td>
<td>1.51 – 2.0</td>
<td>1.26 – 1.5</td>
<td>0.75 – 1.0</td>
</tr>
</tbody>
</table>

Table 1. Rankings for use in Track Filter (source: Islam et al, 2016)
In the exploratory phase of the research, filtering was based only on time. During the initial validation of the resulting tracks, the results did not adequately represent actual vehicle tracks and the additional components were added to address issues that were identified.

**Spatial Filtering.** This component is based on the position of every point associated with a trajectory and is highly dependent on the positional accuracy of the tracked vehicle and of the geospatial polygon representation of the underlying roadway network. This approach cannot be used with a centerline representation and should be used with extreme caution with polygons that have been generated from centerlines. For this research, the polygon representation was
manually constructed for the analysis area using ortho-imagery from ESRI’s World Imagery (2016) which is independent of the validation imagery. The ranking is based on the percent of points for a given track that are located within the street polygon and are provided in Table 1.

**Track Sinuosity Filtering.** Sinuosity of a curve is defined as the curvilinear length divided by the Euclidean distance between the end points of the curve and has a value from one for a straight line to infinity for a closed loop. This measure assists in identifying stationary objects including stopped vehicles. Interestingly, the WAMI data appeared to consistently identify false positives on top of high-rise buildings. Rankings are applied to the sinuosity value of each track and are provided in Table 1.

**Validation of Data Filtering.** For this research, the filtering process was validated by comparing the resulting filtered data for approximately 30 blocks of the two major one-way arterials, King St and Main St, for the first 10 minutes of the afternoon peak to the corresponding imagery provided by PVLabs. The review process was applied to every 5th image or for 5-second intervals. The number of vehicles on the travel lanes of the segments were manually counted on each of the 120 images and matched with the vehicle counts for varying parameter cutoff values in Equation 1.

A final weighting of 20%, 40% and 40% for the temporal, spatial, and sinuosity dimension, respectively, and a cut off value of 2.2 for $A_t$ were identified as the most effective filter for identifying valid tracks. Figure 6 shows the final histogram of track duration compared to the original unfiltered data for the same location and period. More detail on the validation is provided in Islam et al (2016).

**Map Matching**

The purpose of map matching is to match the points in the track to the correct link on the road network. After review of the literature (White et al, 2000; Greenfeld, 2002; Najjar et al, 2005; Quddus, 2006; Quddus et al, 2007; Velaga et al, 2009; Miwa et al, 2012), the approach used in this

![Figure 6: Histogram of Track Duration (source: Islam et al, 2016)](source: Islam et al, 2016)
research consists of a combination of geometric and topologic analysis as outlined in Figure 7.

The geometric component of this process uses the point-to-curve method to select a set of candidate arcs for each trajectory point. An arc is selected as a candidate if it is within the maximum search distance. The topology component addresses the relationship between adjacent objects. This research uses the network topology – which link is connected to which link in order, and the heading of links and tracks including turn movements. Figure 8 shows an example of results for track points at an intersection. More details of the procedure are provided in Islam et al (2016).

**Accuracy of the Algorithm.** 100 vehicle tracks were visually matched with vehicle locations in the imagery provided by PVLabs. On one-way arterials, the match rate was 97.87% while for two-way arterials and local streets, the rate dropped to 96.53%. Both of these are very high and indicate that the algorithm used works well for this type of data on dense urban road networks.
It should be noted that throughout the data analysis, accuracy of image registration and the resulting positional accuracy is not assessed.

Figure 8: Illustration of Map Matching Algorithm (source: Islam et al, 2016)
TRACK DATA MODEL

Traditional data models do not provide the necessary framework to support the use of 3-dimentional tracks, where time is the third dimension, within a network-based geospatial framework and where the total population of tracks is available for calculating microscopic traffic characteristics. The state of the practice in data models that most closely address these issues is presented followed by a summary of the proposed model to support the project analysis and the process used to populate the model.

Background of Existing Data Models for Storing Trajectory Data

The ability to effectively use WAMI data requires a non-traditional data model, particularly with respect to transportation. In geometry, a path and associated information about a moving object is referred to as a trajectory. For vehicle movements, a trajectory is a trace of the vehicle over time. Data about moving objects consists of three different types of information (i) geospatial data, (ii) non-geospatial attribute data and (iii) trajectories (Chandler, Herman et al. 1958). One important property of vehicle trajectories is that vehicles cannot be at any discrete position over space. Their movement is bounded by their surrounding environment such as roads, buildings, obstructions, etc. (Tiakas, Papadopoulos et al. 2009) Another constraint is that two vehicles cannot occupy the same space at a single instance in time. Non-spatial data consists of other thematic information such as vehicle type, vehicle size etc.

Both geospatial and attribute data structures are mature research subjects and most commercial DBMS products allow for their efficient storage and manipulation (Chandler, Herman et al. 1958). Common data mining tasks including classification, regression and geospatial clustering are also well documented. Trajectory data, on the other hand, is a relatively new field of research and no currently available commercial products include a structure that effectively manages this type of data. Although ESRI’s GIS software, ArcGIS©, can display and perform limited analysis of trajectory data, it operates on individual points that define the trajectory, not the trajectory itself, and the processing time becomes impractical for the size database necessary for this project.

Developing tracks includes: (i) pre-processing the data to correct errors in positional measurements, (ii) defining the conceptual data model that meets system requirements, and (iii) storing the data (Chandler, Herman et al. 1958). When defining the data model, additional considerations include establishing the relationship between a track and its environment, and establishing and modeling interactions between tracks. Examples of the former includes determining the position of the vehicle on the road such as distance from upstream node, distance to downstream node, lane position etc. at each discrete time step. Examples of interactions between trajectories include headway between vehicles and relative speeds and accelerations.
Brakatsoulas et al. proposed two different schema for preprocessing and storing large track databases: the 3-D schema and the network schema. In the 3-D schema only the vehicle position is recorded at discrete interval of time. The data can be represented as a 3-dimentional plot of X and Y in space with time as the third dimension as shown in Figure 9. This trajectory representation is adequate to derive properties of a vehicle’s movement such as lane position and distance headway calculations between leader and follower vehicle. PV Labs provides the data in this format. However, this format cannot be used to determine the interaction of a vehicle with its environment. The network schema was determined to be more effective.

The method of associating positional data to a particular position on a roadway is known as map-matching. The Network Schema proposed by Brakatsoulas et al uses a map-matching algorithm and is illustrated in Figure 10. In the network schema, the trajectory data is stored in six tables as indicated by shaded boxes in Error! Reference source not found.. Each track is stored as collection of segments where each segment is associated with a particular edge/link of the roadway network. Two timestamps are recorded with each segment to store the temporal aspect of trajectory data (Chandler, Herman et al. 1958). Association between the nodes and edges of the roadway network is also captured in the schema. The network schema is better equipped for data mining tasks like classification, regression and clustering, but the 3-D schema

![Figure 9. 3-D Representation of Trajectory Data (source: Chandler et al, 1958).](image)
is also required for extracting vehicle interaction characteristics necessary for traffic modeling and calibration purposes.

To make use of geoprocessing and visualization capabilities of GIS software, the data need to reside in a single database. Brakatsoulas et al. (2004) used GIS tracking data which was collected at a sample rate of 30 seconds. During a 30-second trajectory, a vehicle can traverse more than one link, especially in urban areas which have a dense road network, resulting in shorter links. Because the WAMI data includes all trajectories and multiple trajectories can be associated with any given link/edge at any given time step, the data models proposed by Brakatsoulas et al. (2004) require modification. This research uses the modified model shown in Figure 11 where the microscopic characteristics are shown in the upper right. Although not represented in the figure, these characteristics are captured for five leader-follower pairs before and after each individual vehicle for future research into platooning behavior and queuing analysis.
Traffic Data Extraction

To extract and calculate both microscopic and macroscopic traffic characteristics, such as flow, density, headway, and lane position of each vehicle, a combination of geospatial processing and network analysis capabilities are required. This section provides an overview of the data extraction necessary for both populating the data model with the microscopic characteristics and for obtaining macro-level data for validation purposes. A combination of SQL Server, ArcGIS 10.2.2, Python scripts, MatLab, and TransCad were used to mine, visualize, extract and process the data.

In each data extraction analysis, a unique ID for individual tracking point (vehicle location) as well as each individual track is preserved. This helps to manage the dataset efficiently. A limitation to the WAMI data is that vehicles cannot be classified without manual review of the video. Although the data includes pixel size (number of pixels used to identify the vehicle), efforts to classify vehicles using this value have been unsuccessful due to high variability, even for a single track. Because the tracking point represents the centroid of the vehicle, the length of the vehicle is required to estimate bumper-bumper gaps between leader-follower pair. The value stored in the database is thus the centroid-centroid distance, not the physical gap.
Data Preprocessing. The geographic projection of the data provided by PVLabs is WGS84, which uses latitude and longitude measures referenced from the center of the earth and typically reported in decimal degrees. This can be problematic for measuring small distances both in terms of accuracy and precision. Because the study area is relatively small, a planer projection is more appropriate and computationally more efficient for calculating the required traffic characteristics. Therefore, the data were re-projected to NAD83 UTM ZONE17N which is the Transverse Mercator projection for the region of Canada where the data were collected.

The next step was to create a track feature set. Using ArcGIS tracking functions, each track’s time series of points were converted to a set of linear features resulting in the track dataset. Points were then related to tracks according to the data schema in Figure 11.

Macroscopic Data

The fundamental macroscopic traffic characteristics consist of flow, density, speed and by extension, travel time. In traditional approaches, these are either measured using a sample of vehicles over the area being analyzed or are extracted as an aggregate output or measure from a simulation. With WAMI data, these characteristics can be directly measured from the total population at any level of aggregation either spatially or temporally for the data collect.

Origin-destination (OD) of each track. For the WAMI dataset, the origin of a vehicle track is defined as the location where an identified vehicle enters the view frame or where it is first identified within the view frame. Likewise, the destination of the track is identified where the vehicle leaves the view frame or the last location where the vehicle was identified. Figure 12 shows the Origin and Destination points, respectively, for each track in the 15-minute analysis.

![Figure 12: Origin and Destination of Individual Tracks](source: Islam et al, 2016)
period. It should be noted that the points that appear off the roadway centerlines in the figure occurred in parking lots or on private roads. Also, the geospatial representation of the water is at a lower resolution than the point locations resulting in some points appearing in the water. However, based on manual review, these are valid locations within the port and along the coast.

As indicated, the data was collected for downtown Hamilton which includes several high-rises creating an urban canyon with heavy shadows at street level. This resulted in a splitting of vehicle tracks where a track ends as it enters the shaded area and starts as a new track when it leaves the shadow. The results of this can be seen in Figure 12 in the downtown areas where the large clusters of points exist. Linking these tracks is beyond the scope of this research. However, the impact of these splits does not have an effect except for the links that are shaded.

**Average speed and travel time.** Because points (vehicle locations) occur at regular one-second intervals, the distance between points represents the near instantaneous speed of each vehicle at that location. This characteristic provides several options for calculating speeds at a single location, over any length of a single or coincident tracks, over any length of a roadway link or corridor, over any set of roads or links of similar characteristics, or over the entire road network. It also allows for the calculation of both space mean and time mean speeds as well as the arithmetic mean of instantaneous speeds along a length of roadway. Speeds associated with tracks are readily calculated directly from the track time series. Speeds associated with specific locations, a roadway segment or the network are calculated after the track is map matched to the segment. Figure 13 shows the latter for the 15-minute period of the validation data for each link of the arterials in the view area. Also included in the figure on the right is the corresponding

![Figure 13: Mean Speeds and Standard Deviations on Arterials (source: Islam et al, 2016)](image-url)
standard deviation of these speeds. It should be noted that these values include speeds at every point including when the vehicle is stationary.

Once speed has been determined, travel time is readily calculated as the distance over speed for whatever segment(s) have been identified.

**Traffic Flow.** Traffic flow on links can also be estimated after map matching tracks to the network. Flow is estimated by counting tracks matched to each link. The result is the number of vehicles per the time period that the tracks were selected for. Figure 14 provides the vehicles per the same 15-minute time period, again for arterials in Hamilton.

**Other Measures.** Additional macroscopic measures that can be extracted include:

- vehicle density by road segment – number of tracks on the segment at a given time step
- lane-wise distribution – number of tracks by distance from centerline on a given segment
- turn movements – number of tracks that change direction at specified locations (intersections or into parking facilities)
- route choice – number and variation of tracks by O-D

**Microscopic Data**

Microscopic measures that can be mined from the WAMI data include instantaneous speed, instantaneous acceleration, and headway or gap of each individual vehicle. As indicated, these measures are with respect to the vehicle centroid. Instantaneous speed is included with the dataset and can be readily validated as the distance between points (divided by the 1-second time between frames/points). From this, instantaneous acceleration can be derived.

Headway is defined as the distance from front bumper of the lead vehicle to front bumper of the following vehicle. For this research, gap, which is the distance from rear bumper of the lead car to front bumper of the following car, is calculated. Because vehicle classification cannot be determined, an average vehicle length of 6 meters is assumed and subtracted from the centroid
to centroid distance to estimate the gap between vehicles. The gap between each vehicle and the five vehicles leading and the five vehicles following it are included in the database to allow for studying platooning and queueing behavior in the future. This calculation is terminated if the gap exceeds 500 meters. An example of a single leader-follower pair time series for each of the microscopic measures is provided in Figure 15. These are the measures required to calibrate and validate car following sub-models.

Another microscopic measure is lane changing behavior for individual vehicles. This characteristic is not included in the schema since it requires more advanced processing which includes roadway characteristics. Although originally part of the proposed scope, it was replaced by the more involved data processing and validation requirements.

![Figure 15: Microscopic Time Series Traffic Flow Measures for a Sample Leader-Follower Vehicle Pair.](source: Islam et al, 2016)
MICROSCOPIC TRAFFIC SUB-MODELS

Typically, microscopic traffic sub-models have been developed independently. For example, no causal or interactive relationships exist between car-following models and lane-changing models. However, some lane-changing models have been developed in conjunction with other models like gap acceptance and acceleration models. As discussed in the introduction, a lack of consensus exists among traffic engineers and researchers about the validity of each model, particularly for representing car following behaviors.

Background

This section presents the theoretical background and parameters of the car-following and lane-changing models.

Car-Following Models

The mathematical formulation describing the behavior of vehicles following each other in the traffic stream is commonly known as Car Following theory which is based on the molecular approach of modeling behavior where drivers follow the preceding vehicle in a traffic stream (Brackstone and McDonald 1999). These theories were first developed in the 1950s and 1960s and are validated and refined based on traffic measurements collected in the field (Rahman 2013). The most familiar car-following models can be classified into five categories: the Gazis-Herman-Rothery (GHR) model, the Collision Avoidance (CA) model, the Linear Model, the Fuzzy-logic-based model, and the Optimal Velocity (OV) model and its variations (Brackstone and McDonald 1999, Panwai and Dia 2005). These models are presented below along with their advantages and disadvantages.

Gazis-Herman-Rothery (GHR) model

The GHR model and related GM model are the oldest and most well-studied of the car-following models. The first generation was put forward in the late 1950s by Chandler et al (Chandler, Herman et al. 1958) and was based on the hypothesis that a driver’s acceleration/deceleration reaction is proportional to the difference between the speed of the leader and follower vehicles. The structure of the model puts it into the broad classification of stimulus response models. The basic structure of the model is presented in Equation 1.

$$\text{Response} (t + T) = \text{function} (\text{sensitivity}, \text{stimulus} (t))$$  \hspace{1cm} (1)

Where, response is represented as vehicle acceleration/deceleration and stimulus is the relative speed between the leader and follower vehicle. Sensitivity is the only term that needs to be estimated and varies with different models and different drivers. This model also includes
driver perception-reaction time, T. The delay in response is a measure of the time that elapses before the driver notices a change. Early in its development, the proposed model was linear. Rapid development led to a more generalized approach and several validation studies suggested that driver sensitivity was constant for every condition. Later studies suggested that sensitivity increased with vehicle speed and decreased with the distance between the follower and leader vehicles. Gazis, Herman and Rothery proposed a more generalized version of the stimulus response based model as shown in equation 2 (Gazis, Herman et al. 1961).

\[ a_n(t) = c v_n^m(t) \left( \frac{\Delta v(t - T)}{\Delta x^l(t - T)} \right) \]  

(2)

Where,
- \( a_n(t) \) = Acceleration/deceleration of driver n at time t
- \( c, l, m \) = Constants to be determined
- \( v_n^m(t) \) = Velocity of the follower vehicle n at time t
- \( \Delta v(t - T) \) = Relative velocity at time t-T
- \( \Delta x^l(t - T) \) = Relative gap between leader and follower at time t-T
- T = Perception-reaction time

Many studies were undertaken to calibrate this model using different combinations of \( m \) and \( l \) values (Brackstone and McDonald 1999). After 1972, most calibration studies were performed for specific traffic or driving conditions once it was determined that these parameters varied based on conditions. However, findings of these studies were contradictory resulting in a wide combination of \( m \) and \( l \) values (Brackstone and McDonald 1999). This variability is also one of the reasons for the lack of further investigation of the GHR model, particularly after 2000.

Bando et al proposed a car following model named Optimum Velocity Model which used optimal velocity as a function of headway (Bando, Hasebe et al. 1995). This model is similar to the stimulus-response class model except that the stimulus is a function of the relative gap rather than speed while sensitivity is constant. One advantage of this mode is that it does not require the use of perception reaction time. Mathematical formulation of this model is shown in Equation 3 and 4. This model is rarely used in commercial simulation software.

\[ a_e(t) = k [V \times \Delta x(t) - v_n(t)] \]  

(3)

\[ V(\Delta x) = V_1 + V_2 \tanh \left( C_1(\Delta x - l_c) - C_2 \right) \]  

(4)
Where,

\[ a_n(t) = \text{Acceleration of the follower vehicles at time } t \]
\[ \Delta x(t) = \text{Relative distance between vehicles at time } t \]
\[ V = \text{Optimum velocity function} \]
\[ V_1 = 6.75 \text{ m/sec} \]
\[ V_2 = 7.91 \text{ m/sec} \]
\[ C_1 = 0.13 \text{ m}^{-1} \]
\[ C_2 = 1.57 \]

The mathematical formulation and the calibration parameters of the OV function were proposed by Helbing and Tilch (Helbing and Tilch 1998). One important aspect of this model is that it can represent different traffic flow (Gong, Liu et al. 2008). However, unrealistic acceleration and deceleration has been reported when compared with field data (Peng and Sun 2010).

In 1998 Helbing and Tilch proposed a successor to the OV model known as the Generalized Force (GF) Model (Helbing and Tilch 1998). An additional negative relative speed term is included to limit estimation of unrealistic acceleration or deceleration. It is a common notion that if leading cars are traveling much faster, then the following vehicle will not brake, even if its headway is smaller than the safe distance. This behavior cannot be explained by either the OV or GF models (Jiang, Wu et al. 2001). To address this, a similar model named Full Velocity Model using both positive and negative relative speed was proposed by Jiang et al (Jiang, Wu et al. 2001). This model is based on non-linear fluid dynamics, phase transition, and stochastic processes. Potential use of these models in a comprehensive simulation environment is yet to be explored.

**Safety Distance or Collision Avoidance (CA) model**

The concept of safety distance was first proposed by one of the pioneers of the car following theory, Louis A Pipes (Pipes 1953). The basis of the model evolved from the safe driving rule coined in California Motor Vehicle Code which states:

“A good rule for following another vehicle at a safe distance is to allow yourself at least the length of a car between your vehicle and the vehicle ahead of you for every ten miles per hour of speed at which you are traveling.”

The mathematical formulation of this model is shown in Equation 5.

\[
\Delta x(t - T)_{\text{min}} = L_n + \frac{L_n \cdot v_n(t)}{10}
\]  

(5)
Where,
\[ \Delta x(t - T)_{min} = \text{Minimum headway distance} \]
\[ L_n = \text{Length of vehicle} \]
\[ v_n(t) = \text{Speed of vehicle in mph} \]

The length of the vehicle is also considered to be the jam-density spacing which is defined as the clearance distance between cars at maximum density or the distance between vehicles at zero speed. Assuming a vehicle length of 6 meters and expressing all measurements in metric, the Pipes model reduces to equation 6.

\[ \Delta x(t - T)_{min} = 1.34v_n(t) + 6 \] (6)

The Pipes model assumes a linear relationship between vehicle’s desired speed and spacing as well as vehicle acceleration and relative speed between the leader and follower vehicle. These relationships can be expressed as Equations 7 and 8. This formulation is insensitive to traffic density which is not the case in reality. As the speed of the vehicle goes up, drivers tend to leave larger gaps.

\[ v_n(t) = \min(\lambda[x_{n-1}(t - T) - x_n(t - T) - L_n], v_f) \] (7)
\[ a_n(t) = \lambda[v_{n-1}(t - T) - v_n(t - T)] \] (8)

Where,
\[ x_{n-1}(t - T) = \text{Distance of the leader from Upstream Datum} \]
\[ x_n(t - T) = \text{Distance of the follower from Upstream Datum} \]
\[ L_n = \text{Length of the Vehicle / Jam Density Spacing} \]
\[ v_f = \text{Free Flow Speed} \]
\[ \lambda = \text{Driver Sensitivity Factor} \]

Field studies conducted by Rakha and Crowther on the Pipes model have suggested that even with simplistic assumptions, speeds predicted using the Pipes model are consistent with field data from 20-70 km/h (Rakha and Crowther 2002). Beyond this range the Pipes model overestimates speed (Rakha and Crowther 2002). This model requires calibration of three parameters; free flow speed, jam density spacing, and the driver sensitivity factor. The Pipes model is embedded in CORSIM, a microscopic simulation software developed by the FHWA. A variation known as Pitt’s model is used in FRESIM and the Weidemann 74 variation is used in VISSIM (Rakha and Crowther 2002). Although the mathematical formulations of Pitt’s and
Wiedemann 74 differ from Pipes, the primary building block is safety distance as proposed by Pipes.

One important variation of Pipes model which is an amalgamation of Pipes and Greenshield’s macroscopic traffic flow model was proposed by Van Aerde and Rakha (Van Aerde 1995, Van Aerde and Rakha 1995). This model introduces an additional non-linear term to address the insensitiveness to traffic density at higher speeds. The mathematical formulation of is shown in Equation 9.

$$\Delta x_n (t-T) = C_1 + C_3 v_n(t) + \frac{C_2}{v_f - v_n(t)} \quad (9)$$

Where,
- $\Delta x_n (t-T)$ = Gap of the n’th vehicle from its follower
- $v_f$ = Free flow speed of the facility
- $v_n(t)$ = Speed of n’th vehicle at time t
- $C_1, C_2 & C_3$ = Model parameters

The first two parameters of the Van Aerde model provide linear increases of headway with speed while the third parameter insures that spacing increases when speed approaches free flow speed of the facility. This model also insures that vehicles will never exceed the free flow speed of the facility. All parameters are facility dependent. Calibration of the model requires estimation of four macroscopic parameters; free-flow speed, speed at capacity, capacity and jam density spacing which is an advantage since calibration is done using easily collectable macroscopic data (Rakha and Arafeh 2010). This is incorporated in INTEGRATION.

Another method of modeling car following behavior using minimum distance approach is the Collision Avoidance (CA) model. The mathematical formulation of CA was proposed by Kometani and Sasaki (Kometani and Sasaki 1959). This model specifies a safe following distance based on the Newtonian equation of motion which assumes that a follower maintains a safe distance from the leader such that if the leader slams on the brakes, the follower will have enough space to react to the sudden change and stop before colliding with the leader. The mathematical formulation is shown in Equation 10.

$$\Delta x(t-T)_{min} = \alpha v_{n-1}^2 (t-T) + \beta_1 v_n^2 (t) + \beta v_n (t) + b_0 \quad (10)$$

Where,
- $\Delta x(t-T)_{min}$ = Minimum Headway
\[ v_{n-1}^2(t - T) = \text{Speed of the leader} \]
\[ v_n^2(t - T) = \text{Speed of the follower} \]
\[ \alpha, \beta, \beta_1, \beta_0 = \text{Constant to be determined} \]

Model parameters \( \alpha \) and \( \beta_1 \) are measures of deceleration (\( \alpha = -1/2b_n \) and \( \beta_1 = 1/b_{n-1} \)) used by the leader and follower, respectively. \( \beta \) is the perception reaction time of the follower and \( b_0 \) is the jam density spacing. Early calibration of this model led to unrealistic results where drivers were braking at over 1700 m/s\(^2\).

In later years several variations were developed including the Gipps model which is one of the most widely used (Gipps 1981, Brackstone and McDonald 1999, Panwai and Dia 2005). Gipps included several mitigating factors to the original CA model proposing to use a more realistic maximum deceleration value (3 m/s\(^2\)) and increasing the reaction time by \( T/2 \). This model is appealing to traffic engineers because it is derived from a straightforward premise: a driver adapts his speed to smoothly reach the desired speed or to safely proceed behind the leader (Ciuffo, Punzo et al. 2012). As with the Van Aerde model, calibration of the Gipps model requires estimation of four traffic stream parameters, free-flow speed, speed-at-capacity, capacity, and jam density. Details about calibration of Gipps model can be found in (Rakha and Wang 2009) and (Ciuffo, Punzo et al. 2012). Overall consensus is that improved performance can be achieved by careful calibration of the model.

**Linear (Helly) Model**

Although linear, this class of model differs from the first generation GHR (GM) model in that Helly proposed a car following model where the relationship between acceleration and the difference between the current and desired gap is linear (Helly 1959). The mathematical formulation is shown in Equations 11 and 12.

\[
a_n(t) = C_1 \Delta v(t - T) + C_2 [\Delta x(t - T) - D_n(t)] \quad (11)
\]

\[
D_n(t) = \alpha + \beta v_n(t - T) + \gamma a_n(t - T) \quad (3)
\]

Where,
\[
\begin{align*}
 a_n(t) & = \text{Acceleration of the follower vehicles at time } t \\
 T & = \text{Driver reaction time} \\
 \Delta v(t - T) & = \text{Relative velocity between vehicles at time } t-T \\
 \Delta x(t - T) & = \text{Relative distance between vehicles at time } t-T \\
 v_n(t - T) & = \text{Velocity of the follower at time } t-T \\
 \alpha, \beta, \gamma, C_1, C_2 & = \text{Calibration constants}
\end{align*}
\]
Subsequent calibration of the model resulted in wide ranging values of T, C1 and C2. Calibration of this model requires estimation of 5 parameter C1, C2, α, β and γ, which makes it more difficult than most (Panwai and Dia 2005). In later years several variation of this model and their calibration result has been proposed. General consensus is that although this has some advantages over GHR the same criticisms apply (Brackstone and McDonald 1999, Panwai and Dia 2005). Use of this model has been very limited with the notable exception of SITRAS-B (Aron 1988).

*Psychophysical or action point model (AP)*

The basis of these models was first proposed by Michaels in an attempt to model car following based on human perception where driver reaction is dependent on the driver’s ability to perceive change in the apparent size of the leading vehicle (Michaels 1963). This is premised on the perception of relative velocity through visual change as subtended by vehicle ahead (Brackstone and McDonald 1999). Driver reaction is limited by the threshold of this perception.

Three different thresholds of perception are implemented in the model proposed by Michaels. (Michaels 1963). First threshold is given as \( \frac{d}{dt}(\Delta v/\Delta x^2) \sim 6 \times 10^{-4} \). If this threshold is exceeded, the driver will decelerate until a relative velocity is no longer perceived. If perception of threshold of relative velocity remains under this threshold then behavior is governed by another threshold based on spacing. This space-based threshold is relevant at close headways. Thus, for any change to be noticeable, \( \Delta x \) must vary by a “just noticeable distance” (JND) which depends on change in visual angle. Typically 10% change in visual angle is required for the driver to perceive that he/she is closing to the leader. Upon crossing this second threshold, the driver sets a determined acceleration/deceleration until the third threshold is crossed. This threshold is obtained from a series of perception-based experiments that require passengers in test vehicles to observe a target vehicle and make a decision whether the car-following gaps are widening or shortening. Calibration and validation of individual elements and threshold of this kind of model have had limited success (Brackstone and McDonald 1999). A derivative of this model developed by Fritzsche is used in the English simulation software PARAMICS (Fritzsche 1994).

*Fuzzy-Logic-Based Model*

A relatively new development in car following theory is the use of fuzzy-set theory, which describes how adequately a variable fits the description of a term. This approach was first applied to the GHR model by Kikuchi and Chakroborty (Kikuchi and Chakroborty 1992). This
approach is unique because the human driver is considered a fuzzy system rather than a precise machine, and thus, is more likely to represent real human driving behavior.

The model divides the selected inputs into a number of fuzzy sets. Logical operators are then used to produce fuzzy output sets or rule-based car-following behaviors. For example, two principal inputs to the decision-making process can be relative speed and the separation divergence (or the ratio of vehicle separation to the driver’s desired following distance). A typical fuzzy rule for the car-following model would then have the form: If Distance Divergence is “Too Far” and relative speed is “closing,” then the driver’s response is “No Action”. However, it is difficult to calibrate the membership function, which is the most important part of the model (Brackstone and McDonald 1999, Panwai and Dia 2005). Due to the complex nature of this model and the difficulty in calibration and validation, this approach is mostly used in understanding properties of other models rather than in use in simulation software.

**Lane Changing Model**

Although the scope of this research was modified to perform the validation and calibration only on the car following model, the initial research included an assessment of lane changing models. The result of that review is included here as a useful reference for future research using WAMI data or other comprehensive data sources.

Predicting lane changing behavior is a complicated problem. The complication arises due to asymmetric behavior of drivers in determining requirements for changing lanes and variability in gap acceptance behavior. Lane-changing models have been developed independently as well as in conjunction with other models like gap acceptance models and acceleration models. Since the 1980s, many lane-changing models have been developed for micro-simulation and vehicle automation like adaptive cruise control. Because these models consist of multiple parameters and interactions, calibration and validation are also consist of complex processes. Models that have been developed to date can be categorized into four groups: rule-based models, discrete choice based models, artificial intelligence models, and incentive-based models (Rahman, Chowdhury et al. 2013). Figure 8 provides an overview of these categories along with their offspring.

Unlike car following models which consist of relatively straight-forward and easily encapsulated formulations, lane changing models typically consist of a complex set of processes and, as shown in Figure 16, have many different forms. These models are typically associated with specific micro-simulation software packages and are explained as they are implemented.
instead of as an independent formulation. From a review of existing lane-changing models, rule-based and discrete choice-based models appear to be the most commonly used. These models have been widely implemented in microscopic traffic simulation software. Among them, rule-based lane-changing models are based on the perspective of drivers. For rule-based models, typically the subject vehicle’s lane-changing reasons are evaluated first. If these reasons warrant a lane change, a target lane from the adjacent lane(s) is selected. A gap acceptance model fitted based on field data/simulation data is then used to determine whether the available gaps should be accepted.

Most discrete choice-based lane-changing models are based on logit or probit models. For discrete choice-based models, the lane-changing maneuver is usually modeled as either MLC or DLC following three steps: 1) checking lane change necessity, 2) choice of target lane, and 3) gap acceptance. Each of these steps can be formulated as a probit or logit model. Depending on which step and the number of lanes, the subject driver may face a binary or multi-choice decision. Similar to rule-based models, discrete choice model parameters and utility functions need to be calibrated using field data. In existing discrete choice-based lane-changing models, the heterogeneities in drivers and vehicles (i.e., driver aggressiveness, driving skill level, vehicle

Figure 16: Classification of lane-changing models (source: Rahman et al, 2013)
acceleration performance) have not been given adequate consideration. A major reason is that existing traffic data and data collection technologies cannot provide information that is detailed enough for developing and testing such models. Nevertheless, these characteristics are important for accurately describing real-world lane-changing behaviors and relevant explanatory variables should be incorporated into the utility functions of future discrete choice-based lane-changing models.

Three commercial micro-simulation packages which incorporate different lane changing processes were reviewed in detail. TransModeler© is Caliper Corporation’s versatile traffic simulation package that models a wide array of traffic planning and modeling tasks which integrates complex simulation applications with a GIS environment. PTV Group’s VISSIM© is a widely used microscopic, behavior-based multi-purpose simulation package. INTEGRATION, developed by Van Aerde/Rakha, is a traffic simulation model which integrates vehicle dynamics with traffic flow theory. The accuracy of microscopic traffic simulation is largely dependent on the lane changing model which requires careful calibration to effectively represent real life driver behavior.

*TransModeler* (Caliper 2007)

The lane changing procedures in TransModeler is classified as a discrete choice model from Figure 8. TransModeler models lane changing behavior in three steps which examine the feasibility, desirability, and safety of a lane change maneuver, respectively:

- **Step 1: Selection of eligible lanes**
- **Step 2: Lane change decision-making and target lane selection**
- **Step 3: Gap evaluation and lane change execution**

**Step 1: Lane Eligibility Selection.** The first step in the lane changing maneuver is to explore the current and neighboring lanes to determine whether a lane changing maneuver is feasible or not. The reason for ineligibility, as described in TransModeler, includes barrier, solid lane stripe, lane blockage or lane closed, toll booth closed, vehicle type restriction, a vehicle’s distance to freeway exit or intersection, and connectivity of the prospective lane with vehicle’s path. If more than one lane is eligible, then the selection of target lane depends on the lane changing regime. MLC takes priority over DLC. FLC is a special case of DLC where an acceptable gap has not found or the location before which a lane change must be executed in very near, or both

**Step 2: Lane Changing Regime Selection.** After the selection of eligible lanes comes the process of lane selection and lane changing decision making. If Step 1 includes more than one alternative, the selection of target lanes depends on the lane changing regime:
• Discretionary Lane Change (DLC)
• Mandatory Lane Change (MLC)
• Forced Lane Change (FLC)

The distinction between lane change regimes was first used in CORSIM (Halati, Lieu et al. 1997) which was based on the lane change model proposed by Gipps (Gipps 1986). Similar to the model used in TransModeler the Gipps model considered the necessity, desirability, and safety of performing the lane change. In this rule based model, MLC is performed when the driver must leave the current lane (e.g. to stay on its path or avoid a lane blockage). DLC is made due to the driver perceiving an improved driving condition in the target lane which may include an increase in speed or better line of sight. In an attempt to improve the rule based decision making process used in CORESIM, Ahmed et al proposed a discrete choice framework that captures both MLC and DLC decision processes (Ahmed, Ben-Akiva et al. 1996, Ahmed 1999). Ahmed proposed a third category, FLC, which occurs when a gap is not sufficient for safe maneuvering but is created by the driver to execute a lane-changing maneuver in heavily congested traffic conditions (Ahmed 1999). The probability of selecting MLC, DLC or FLC is calculated using a vector of explanatory variables affecting the decision making process. Although this model accurately describe driver’s MLC, DLC and FLC decision, it fails to capture trade-offs between the different regimes and also fails to define conditions that trigger MLC (Toledo, Koutsopoulos et al. 2003). Built on the work of Ahmed et al., a probabilistic lane change model using a discrete choice framework was proposed by Toledo et al (Toledo, Choudhury et al. 2005). TransModeler uses a modified version of this model. Details of the models and lane changing decision making process used in TransModeler is described in the following.

Selection of the lane changing regime depends on several parameters. During the simulation, TransModeler maintains a general “picture” of every vehicle’s path immediately downstream of that vehicle. The downstream distance is determined by look-ahead distance which is limited by three parameters:

• Time headway (default: 90 sec)
• Maximum number of links (default: 4)
• Maximum distance (default: 2414.02 m)

Time headway takes priority. Each driver is assumed to be aware of the path approximately 90 seconds into the future. Free flow speed is used to calculate the distance downstream for this time headway. The next two parameters, maximum number of links and maximum distance, are used to improve the computational efficiency and are based on the
assumption that the driver is not concerned about turn movements beyond a certain distance. As part of looking ahead, TransModeler considers lane drops as well as turns or exits. If a diversion is necessary, TransModeler identifies a critical point along its path within the look ahead distance which is used to define the lane changing regime. Any lane change that is required beyond a critical point is mandatory even if the current lane is favorable and connected to its path. Instead of using fixed values for determining the critical point, TransModeler uses a distribution across the driver population for different facility types to capture the difference in driver aggressiveness as shown in Table 2.

<table>
<thead>
<tr>
<th>Percent of Drivers (%)</th>
<th>On Street (m)</th>
<th>On Freeways (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>213</td>
<td>381</td>
</tr>
<tr>
<td>5.0</td>
<td>229</td>
<td>457</td>
</tr>
<tr>
<td>5.0</td>
<td>244</td>
<td>533</td>
</tr>
<tr>
<td>15.0</td>
<td>259</td>
<td>610</td>
</tr>
<tr>
<td>30.0</td>
<td>274</td>
<td>686</td>
</tr>
<tr>
<td>15.0</td>
<td>290</td>
<td>762</td>
</tr>
<tr>
<td>10.0</td>
<td>305</td>
<td>838</td>
</tr>
<tr>
<td>5.0</td>
<td>320</td>
<td>914</td>
</tr>
<tr>
<td>5.0</td>
<td>335</td>
<td>991</td>
</tr>
<tr>
<td>5.0</td>
<td>351</td>
<td>1067</td>
</tr>
</tbody>
</table>

In TransModeler, lanes between different segments of the roadway are connected using a lane connectivity bias. In case of lane drops, a response distance from the lane drop location is used to prevent vehicle from changing lanes at the last moment. Beyond this point connectivity bias begins to affect mandatory lane change behavior. Any lane change upstream of the critical point is considered to be DLC which depends on the driver’s perception of improved driving quality in the target lane.

TransModeler provides two different discretionary lane change (DLC) models to compute the utility of each lane in the choice set. Both models evaluate the desirability of different lanes but differ in the selection strategy of eligible lanes as well as the lane changing decision making process. Both use a multinomial logit formulation and vary in the number of lanes considered at each evaluation step. A short description of the models and their input variables are provided below.

**DLC Model 1 - Neighborhood Lane Model:** In the neighborhood lane model, only eligible adjacent lanes, left or right, are part of a driver’s choice set. This reflects a more myopic
view of the drivers. Perceived gain in speed is the principal variable of the model. A multinomial logit model is used to calculate the utility of a lane. The utility function is a linear combination of the variables listed in Table and is formulated per Equation 4.

\[
P_t(NL|V_s) = \frac{1}{1 - e^{-x_n^{NL}(t)\beta^{NL}}} \tag{4}
\]

Where:

\[
P_t(NL|V_s) = \text{The probability of selecting a current or neighboring lane}
\]

\[
x_n^{NL} = \text{Vector of explanatory variables affecting lane change decisions}
\]

\[
\beta^{NL} = \text{Estimated Parameter corresponding to vector of parameter}
\]

<table>
<thead>
<tr>
<th>Table 3. Lane Choice Utility Function (Source: Caliper 2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable (X)</strong></td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Path influence factor</td>
</tr>
<tr>
<td>Minimum speed gain (fps)</td>
</tr>
<tr>
<td>Average Speed gain (fps)</td>
</tr>
<tr>
<td>Heavy Vehicle Ahead</td>
</tr>
<tr>
<td>On Ramp Ahead</td>
</tr>
<tr>
<td>Off Ramp Ahead</td>
</tr>
<tr>
<td>Slow Vehicle in Passing Lane</td>
</tr>
<tr>
<td>Lane not Connected</td>
</tr>
<tr>
<td>Heavy Vehicle (Subject)</td>
</tr>
<tr>
<td>Transit Vehicle (Subject)</td>
</tr>
<tr>
<td>Tailgated (Subject)</td>
</tr>
<tr>
<td>Same Direction as Previous DLC</td>
</tr>
<tr>
<td>HOV towards HOV lane</td>
</tr>
<tr>
<td>HOV away from HOV lanes</td>
</tr>
<tr>
<td>Open Space ahead at toll booth</td>
</tr>
</tbody>
</table>

Lanes may be connected or disconnected to the vehicle path. If a lane brings the subject vehicle into or closer to its path, it is perceived as beneficial to the driver. The path influence factor is a function of the distance from the nearest downstream critical point before which the vehicle must be in the connected lane and the required number of lane changes. The path influence factor has a functional form following a logistic curve defined by the distances at which the path influence factor has values of 75% and 50% (Table 3). The curve tappers to zero at large distances and approaches 100 as the vehicle approaches the critical point. Lower values of distance at path influence factor of 75% and 50% indicates higher aggressiveness of the driver. As more aggressive driver tends to change lanes at the last moment. This parameter, when appropriately calibrated, insures that drivers tend to choose connected lanes and over disconnected lanes upstream of the nearest critical point, after which, the lane changing regime.
changes to MLC. This model prevents vehicles from making sudden lane changes at the last moment. Minimum and average speed gains is calculated using equation 5 and 6 respectively. The number of vehicles used in calculating speed gain depends on the number of vehicles downstream of the subject vehicle within the ‘search’ distance which is not specified. The default value of the path influence factor is provided in Table.

Minimum Speed Gains:

\[ V_{i}^{MinGain} = \min\{ V_{i}^{Min} , V^{Desired} \} - V^{Current} \quad i=L,R \quad (5) \]

Where:
\( i \) = Left lane or Right lane
\( V_{i}^{MinGain} \) = Minimum speed gain in prospective lane \( i \);
\( V_{i}^{Min} \) = Minimum speed gain in prospective lane \( i \);
\( V^{Desired} \) = Subject vehicle’s desired speed;
\( V^{Current} \) = Subject vehicle’s current speed.

Average Speed Gain:

\[ V_{i}^{AvgGain} = \min\{ V_{i}^{Avg} , V^{Desired} \} - V^{Current} \quad i=L,R \quad (6) \]

Where:
\( i \) = Left lane or Right lane
\( V_{i}^{AvgGain} \) = Average speed gain in prospective lane \( i \);
\( V_{i}^{Avg} \) = Average speed observed in prospective lane \( i \);
\( V^{Desired} \) = Subject vehicle’s desired speed;
\( V^{Current} \) = Subject vehicle’s current speed.

<p>| Table 4: Path Influence Factor (unit distance per number of lane changes required) |
|--------------------------------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Factor</th>
<th>On Street (ft)</th>
<th>Off Street (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>400.0</td>
<td>2000.0</td>
</tr>
<tr>
<td>50</td>
<td>800.0</td>
<td>5000.0</td>
</tr>
</tbody>
</table>

**DLC Model 1 - Target Lane Model:** While the neighborhood lane model reflects a more myopic view of the driver, the target lane model accounts for a broader view. In the target lane model, drivers choose the lane which has the highest perceived utility from all lanes on a segment, which models a driver’s ability to choose a lane based on its specific utility such as
HOV or tolled lanes. The parameters and coefficients of the utility function for those parameters are listed in Table.

The path plan factor acts similarly to the path influence factor of the Neighboring Lane Model. The parameter this factor decreases with the increased number of lane changes required for the subject vehicle to return to its path. In other words, the utility of a lane decreases as the number of lane changes from that lane to a lane connected to the vehicles path increases.

<table>
<thead>
<tr>
<th>Variable</th>
<th>On Street</th>
<th>On Freeways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lane 1 constant (First from right)</td>
<td>-1.696</td>
<td>-1.696</td>
</tr>
<tr>
<td>Lane 2 constant (Second from right)</td>
<td>-0.571</td>
<td>-0.571</td>
</tr>
<tr>
<td>Lane 3 constant (Third from right)</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>Lane Density (veh/km)</td>
<td>-0.013</td>
<td>-0.013</td>
</tr>
<tr>
<td>Average Speed in Lane (m/sec)</td>
<td>0.176</td>
<td>0.176</td>
</tr>
<tr>
<td>Front Vehicle Spacing (m)</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>Relative Front Vehicle Speed</td>
<td>0.115</td>
<td>0.115</td>
</tr>
<tr>
<td>Tailgated Dummy</td>
<td>-4.395</td>
<td>-4.395</td>
</tr>
<tr>
<td>Current Lane Dummy</td>
<td>2.686</td>
<td>2.686</td>
</tr>
<tr>
<td>One lane Change from Current Lane</td>
<td>-0.845</td>
<td>-0.845</td>
</tr>
<tr>
<td>Number of additional Lane changes</td>
<td>-3.338</td>
<td>-3.338</td>
</tr>
<tr>
<td>Path Plan Impact - 1 Lane Change Required</td>
<td>-2.549</td>
<td>-2.549</td>
</tr>
<tr>
<td>Path Plan Impact - 2 Lane Change Required</td>
<td>-4.953</td>
<td>-4.953</td>
</tr>
<tr>
<td>Path Plan Impact - 3 Lane Change Required</td>
<td>-6.953</td>
<td>-6.953</td>
</tr>
<tr>
<td>Next Exit Dummy – Lane Changes Required</td>
<td>-0.872</td>
<td>-0.872</td>
</tr>
<tr>
<td>$\theta^{MLC}$</td>
<td>-0.417</td>
<td>-0.417</td>
</tr>
<tr>
<td>$\pi_1$</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>$\pi_2$</td>
<td>0.086</td>
<td>0.086</td>
</tr>
<tr>
<td>Alpha Lane 1 ($\alpha^{lan1}$)</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Alpha Lane 2 ($\alpha^{lan2}$)</td>
<td>0.086</td>
<td>0.086</td>
</tr>
<tr>
<td>Alpha Lane 3 ($\alpha^{lan3}$)</td>
<td>-1.412</td>
<td>-1.412</td>
</tr>
<tr>
<td>Alpha Lane 4 ($\alpha^{lan4}$)</td>
<td>-1.072</td>
<td>-1.072</td>
</tr>
</tbody>
</table>

The formulation of this factor, shown in Equation 16, insures that the number of lane changes does not impact the discretionary lane change if the vehicle is a long distance from a downstream exit.

\[
PPI = \beta_i,\ path (d^{exit})^{\theta^{MLC}}
\]  
\[\text{(7)}\]

Where:

- $\beta_i, \ path$ = Path plan impact coefficient for target lane $i$;
- $d^{exit}$ = Distance from downstream turn or exit;
\[ \theta_{MLC} = \text{Path plan impact power.} \]

The values of \( \theta_{MLC} \) and \( \beta_{i, \text{path}} \) are negative which insures that if \( d^{exit} \) decreases, the impact of PPI increases in the decision making process. When \( d^{exit} \) decreases, vehicles are more likely to stay on the connected path. The target lane model also attempts to capture the unobservable (latent) characteristics of individual drivers. For each individual driver, a latent variable \( (V_n) \) is drawn from a distribution assumed to represent the population. The alpha parameter listed in Table 5 is multiplied by the latent variable \( (V_n) \) when calculating the utility of different lanes. While using density and average speed of different lanes, TransModeler only uses data from a certain length downstream or upstream of the vicinity of the subject vehicle. The default downstream length is 200 m and upstream length is 0 m. Although this model was developed for modeling lane change in the presence of an exclusive lane, its flexible structure of makes it suitable for more general purposes.

The target lane model was developed by Choudhury et al for the NGSIM program sponsored by the Federal Highway Administration (FHWA) (Choudhury 2005). Two different sets of data, one aggregated and one disaggregated, were used to calibrate and validate the model. The disaggregated data consisted of detailed behavior data such as vehicle trajectories (Choudhury 2005). The explanatory variables and parameters of the model, such as the subject vehicle speed and relation between subject and other vehicles, were derived from this disaggregated data. This work is done outside of simulation environment. The data were collected by the FHWA in 1983 for a 997-meter long four-lane section of southbound I-395 in Arlington, VA, which included one on-ramp and two off-ramps. Attributes included position, lane and dimension of every vehicle on the section at 1 second intervals. The aggregated data was collected from 1.5 miles section of I-80 in Emeryville and Berkley California which included 4 on- and 3 off-ramps with one continuous access HOV lane. Attributes included lane specific traffic counts, occupancies and speed. The data was collected for two weeks at 30 seconds intervals and was divided into two sets, one for each week. The MITSIM simulation software was used to validate the model developed from the disaggregate data. One set of the aggregated data was used to calibrate the MITSIM model and the second set was used to validate the calibrated model. Calibration and validation were focused on the evening peak period. (Choudhury 2005).

**Gap evaluation and lane change execution.** TransCad includes three different methods for evaluation gaps; non-linear, linear, and NGSIM. Tables 6 through 8 provide the factors for each model, respectively.
**Non Linear Gap Acceptance (Caliper 2007):**

Table 6: Non Linear Gap acceptance Parameter used in TransModeler

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DLC-Lead</th>
<th>DLC-Lag</th>
<th>MLC-Lead</th>
<th>MLC-Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (m)</td>
<td>1.0</td>
<td>1.5</td>
<td>1.0</td>
<td>1.50</td>
</tr>
<tr>
<td>Constant ($\beta_0^g$)</td>
<td>0.508</td>
<td>2.020</td>
<td>0.384</td>
<td>0.587</td>
</tr>
<tr>
<td>Follower Slower ($\beta_1^g$)</td>
<td>0.153</td>
<td></td>
<td></td>
<td>0.048</td>
</tr>
<tr>
<td>Follower Faster ($\beta_2^g$)</td>
<td>0.420</td>
<td>0.188</td>
<td></td>
<td>0.356</td>
</tr>
<tr>
<td>Follower Speed ($\beta_3^g$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remaining Distance ($d$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance of Gap Error ($\sigma^g$)</td>
<td>0.488</td>
<td>0.526</td>
<td>0.859</td>
<td>1.070</td>
</tr>
</tbody>
</table>

**Linear Gap Acceptance (Caliper 2007):**

Table 7: Linear Gap acceptance Parameter used in TransModeler

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DLC-Lead</th>
<th>DLC-Lag</th>
<th>MLC-Lead</th>
<th>MLC-Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (m)</td>
<td>1.0</td>
<td>1.5</td>
<td>1.0</td>
<td>1.50</td>
</tr>
<tr>
<td>Constant ($\beta_0^g$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follower Slower ($\beta_1^g$)</td>
<td>0.200</td>
<td>0.150</td>
<td>0.150</td>
<td>0.100</td>
</tr>
<tr>
<td>Follower Faster ($\beta_2^g$)</td>
<td>0.350</td>
<td>0.450</td>
<td>0.300</td>
<td>0.350</td>
</tr>
<tr>
<td>Follower Speed ($\beta_3^g$)</td>
<td>0.450</td>
<td>0.500</td>
<td>0.400</td>
<td>0.450</td>
</tr>
<tr>
<td>Remaining Distance ($d$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance of Gap Error ($\sigma^g$)</td>
<td>1.000</td>
<td>1.500</td>
<td>1.000</td>
<td>1.500</td>
</tr>
</tbody>
</table>

**NGSIM Gap Acceptance (Caliper 2007):**

Table 8: NGSIM Gap acceptance Parameter used in TransModeler

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DLC-Lead</th>
<th>DLC-Lag</th>
<th>MLC-Lead</th>
<th>MLC-Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (m)</td>
<td>1.0</td>
<td>1.5</td>
<td>1.0</td>
<td>1.50</td>
</tr>
<tr>
<td>Constant ($\beta_0^g$)</td>
<td>1.541</td>
<td>1.426</td>
<td>1.541</td>
<td>1.426</td>
</tr>
<tr>
<td>Follower Slower ($\beta_1^g$)</td>
<td>6.210</td>
<td>0.640</td>
<td>6.210</td>
<td>0.640</td>
</tr>
<tr>
<td>Follower Faster ($\beta_2^g$)</td>
<td>0.130</td>
<td>0.130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Follower Speed ($\beta_3^g$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha ($\alpha^g$)</td>
<td>-0.008</td>
<td>-0.205</td>
<td>-0.008</td>
<td>-0.205</td>
</tr>
<tr>
<td>Variance of Gap Error ($\sigma^g$)</td>
<td>0.854</td>
<td>0.954</td>
<td>0.854</td>
<td>0.954</td>
</tr>
</tbody>
</table>

**VISSIM (Vision 2007)**

The lane change model used in VISSIM falls under the general category of rule based models. This model divides lane changing into two regimes: Necessary Lane Change and Free Lane Change. Similar to TransModeler, the necessary lane change is mandated only to remain on the path of the vehicle. The free lane change is executed to achieve better driving conditions. This model does not account for variations of aggressiveness in the driver population but is fixed
based on subject vehicle and the trailing vehicle in the target lane. It is assumed that the subject vehicle is more aggressive in breaking than the trailing vehicle. This model also does not account for the relation between lane change decision and gap and speed difference between the subject vehicle and the lead vehicle in the target lane. Only a minimum headway between vehicles on the target lane is used as part of gap acceptance. If the lead vehicle is slower than the subject vehicle then the subject driver may be reluctant to change lanes. The selection criterion of eligible lane however is up to the user. Two different lane selection methods are available but are explicitly specified by the modeler:

- Free Lane Selection
- Slow Lane Rule

In Free lane selection, vehicles are allowed to overtake in any lane. The Slow Lane Rule allows overtaking in the fast lane (usually the adjacent left lane) only if the speed in the fast lane is above 60 km/h. For slower speeds, “undertaking” can occur in the slow lane (usually the adjacent right lane) with a maximum speed difference of 20 km/h between subject and lead/lag vehicle in the target lane and collision time with a lead vehicle in the slower lane is more than a threshold value (default: 11 sec). The rule associated with lane change is myopic in nature and is unlikely to account for lane changes associated with exclusive lanes.

The parameters associated with lane changing decision follow:

- Maximum deceleration of the lane changing vehicle and the trailing vehicle on the target lane
- Distance required for unit reduction of acceleration
- Accepted acceleration of two vehicles

The primary driver behaviors associated with lane change is maximum and accepted deceleration for subject and trailing vehicle in the target lane. The primary assumption is that the subject vehicle will only execute a lane change if the distance between the subject and the trailing vehicle is high enough such that, if the subject vehicle applies its maximum breaking due to some emergency, the trailing vehicle will have the time to stop without crashing into it. The maximum acceptable deceleration is only applicable if the distance to emergency stop location (i.e. stopping line at intersection) is zero. The maximum deceleration parameter decreases linearly with the increase in distance to the emergency stop location. The default value for this decrease in maximum deceleration is -1 ms\(^2\) per 300 meter.

The gap acceptance parameters associated with lane change execution are simplistic. A diffusion time threshold is used which removes a vehicle from the simulation if the vehicle is
unable to execute the lane change within that threshold. Shorter diffusion times can result in incomplete vehicle trips in standstill conditions. The Gap Acceptance parameters consist of:

- Minimum Headway (front/rear): the clear distance between the subject and front or rear vehicle at standstill condition in the current lane. If the distance is higher than this value than lane changing action is considered
- Safety distance reduction factor: the measure of gap between the subject vehicle and the leading or trailing vehicle in the target lane. It is a function of safety distance calculated from the maximum deceleration parameter of the lane change decision process described earlier.
- Additional parameter which corresponds to cooperative lane change that is similar to forced lane change in TransModeler. Maximum relative speed, collision and deceleration are fixed in this model to initiate cooperative lane change. These parameters are fixed for all drivers

Calibration of the lane changing model in VISSIM requires the estimation of mean values of the parameters for the location under consideration. The model structure is unlikely to capture complex lane change decision of drivers.

**INTEGRATION** (Rakha and Zhang 2004)

The lane change model used in INTEGRATION also falls under broad category of rule based models. The lane change reason is evaluated first, which is then followed by the application of the gap acceptance model to determine whether an available gap should be accepted. Similar to VISSIM the lane change decision in INTEGRATION is divided into two categories; mandatory and discretionary lane changes. Mandatory lane change takes place when the current lane is no longer a feasible option and the driver must change lanes to remain connected to its path. Discretionary lane change occurs when the adjacent lane is perceived to provide for better driving conditions.

**Mandatory Lane Change:** Instead of using a rigid boundary between discretionary and mandatory lane change, INTEGRATION uses a stochastic method to smoothly transition from one regime to the other. This is done by defining two boundaries upstream of the point before which the vehicle must be in a connected lane. The first boundary, located farther upstream, is denoted as the “softwall”; while the second boundary is denoted as the “hardwall” and is located closer to the physical diverge point or intersection. These boundaries can vary significantly. The mean location of “hardwall” is 10n times the jam density spacing and for the “softwall” is 100n times the jam density spacing. A coefficient of variation of 0.5 is used to incorporate variation of the boundaries from vehicle to vehicle. This lane changing logic also incorporates a gap acceptance process, where the size of an acceptable gap in the adjacent lane is a function of the
vehicle’s speed, distance to the point where the lane change should be completed, and the time spent by an individual vehicle to search for a gap. Each lane change action requires two seconds which can be modified by the user. This time does not vary with the increase of number of lane changes required for the vehicle to be in connected path.

**Discretionary Lane Change** (Rakha and Zhang 2004): Discretionary lane change is considered when an MLC is not required and the driver perception of driving conditions on other lanes is better than the current lane. Driving condition is a measure of perceived speeds in the current and adjacent lane. Speeds in the current lane and adjacent lanes to the left and right are compared at every second. This process is similar to the Neighborhood lane model of TransModeler but differs in estimation of utility/desirability of different lanes. The perceived speed of current and adjacent lanes are calculated using Equation 8 and 9 respectively.

\[
\hat{u}_i = u_c + \min(u_c f_3 + f_4, f_5) \sqrt{pcu}
\]  
(8)

\[
\hat{u}_i = u_i + \left(\frac{n}{2} - \text{ABS}\left(\frac{n-1}{2} + 1 - i\right)\right)(u_i f_7 + f_8) + \left(\frac{7 - i}{n}\right)^2 (pcu - 1)
\]  
(9)

Where:

\[ i = \text{lane number from right side of the road} \]
\[ u_c = \text{Vehicle speed in current lane (km/h)} \]
\[ \hat{u}_i = \text{Perceived speed in lane “i” (km/h)} \]
\[ u_i = \text{Actual speed in lane “i” (km/h)} \]
\[ f_3 = \text{Relative inertia factor (unitless), (Default: 0.2)} \]
\[ f_4 = \text{Absolute inertia factor (km/h), (Default: 5 km/h)} \]
\[ f_5 = \text{Absolute Maximum inertia factor (km/h), (Default: 10 km/h)} \]
\[ f_7 = \text{Relative speed factor (unitless), (Default: 0.01)} \]
\[ f_8 = \text{Absolute speed factor (km/h), (Default: 1 km/h)} \]
\[ pcu = \text{Vehicle length equivalency factor and} \]
\[ n = \text{Number of lanes on roadway section} \]

Speeds on current and adjacent lanes are calculated using the steady state car following model as discussed in that section. The INTEGRATION model considers a pre-specified bias for a vehicle to remain in the current lane by adding an inertia factor to the vehicle’s desired speed when computing its perceived speed. This value is estimated as minimum of \( u_c f_3 + f_4 \) or \( f_5 \) multiplied by the square root of the \( pcu \) of the vehicle. The maximum inertia that can be added to the desired speed is \( f_5 \sqrt{pcu} \). The use of this factor reduces the number of unnecessary lane changes by increasing the attractiveness of the current lane. It also insures that larger vehicles are more reluctant to make lane changes due to their high \( pcu \) value which results in increased
perceived speed/desirability of the current lane. In addition, a bias is applied towards travel in specific lanes depending on the total number of lanes when vehicles travel outside the influence area of merge and diverge sections. Specifically, the model biases passenger cars to travel towards the middle lanes for roadways with three or more lanes and trucks or other larger vehicle towards the right lanes. This bias is achieved by altering the perceived speed in a specific lane using the estimation formula of Equation 9. The formulation insures additional utility (speed) is added for trucks or larger vehicle \((pcu>1)\) in right lanes which decreases with increase in lane number. This additional utility has no effect on passenger cars. A different lane bias factors is prescribed for lanes near merge section to bias through movement toward middle or left lane rather than shoulder lane. Users can override the default lane bias parameters. For example, if the user specifies a factor of 3.0 the perceived speed on all lanes other than the biased lane is computed as 1/3 the perceived speed.

**Gap Acceptance:** After the lane change decision is made, the available gap in the target lane is evaluated with respect to the driver-specific stochastic critical gap which is modeled as either a normal or lognormal distribution of gap acceptance behavior. Gaps that exceed the driver-specific critical gap are accepted while those that are less than the critical gap are rejected. The model also considers the impact of wait time on the critical gap by considering a linear decay function in the critical gap over time. The calibration parameters of gap acceptance model include:

a) The mean critical gap  
b) The coefficient of variation for stochastic gap acceptance  
c) Decay time for decay function

The critical gap varies by type of movement, number of lane-changes required, and type of intersection. For example the critical gap for a left turn movement is 0.5 s longer than the base through movement. On the other hand, the critical gap for a right turn movement is 0.5 s shorter than the through movement critical gap. The critical gap is increased by 0.5 s for each additional lane of travel. These parameters are hard coded in the software and are consistent with HCM and AASHTO Green Book procedures.

**Car Following Model Calibration and Validation**

This section presents an overview of the calibration processes to be used on the WAMI data along with preliminary results for calibration of the linear Helly car following model. The full results will be available in the final thesis for M. R. Islam and in the paper under development by Islam et at. (2016b). Car following is the cornerstone of microsimulation analysis. Several models have been developed as discussed above. Calibrating and validating
these models is essential for determining the validity of the simulation to accurately represent traffic behavior and predict the resulting performance measures.

Calibration of uncertain input parameters of traffic flow models against field data is the accepted approach to addressing the epistemic uncertainty related to un-modeled details not predicted by the model (Montanino et al. 2012). Because of this, model parameters have been indirectly derived by means of an optimization framework. The capability of a model to reproduce real tracks is directly related to the setup of the optimization framework itself (Punzo et al. 2012). In case of time series data, a different goodness of fit measure has been proposed by Montanio et al (Montanino et al. 2012). In a general framework, the specification of the estimation problem relies on several factors, including:

i. Choice of the model parameters to calibrate,
ii. Choice of input data
iii. Influence of data sampling rate and data smoothing
iv. Choice of the estimation method
v. Choice of the Measure of Performance (MoP) to describe the status of the system,
vii. Choice of the optimization algorithm to solve the problem.

Different estimation techniques have been used over time to solve the problem of indirect estimation of model parameters against limited trajectory data. Recent developments in sensing, tracking, image processing, and data capture technology have enabled the capture of vehicle trajectory data over larger areas. However, even high fidelity trajectory data like the NGSIM data, which has been used to calibrate car following models as well as for development of new models, is restricted either spatially or temporally or both. Because WAMI data consists of the entire population of vehicles within a relatively large area for an extended period, a more comprehensive dataset can be used to calibrate and validate car following models for multiple road types, geometries and traffic flow situations.

**Data Sampling and Smoothing.**

One major factor is the sampling interval of the data. In most simulation software, a vehicle trajectory is estimated every 0.1 seconds. WAMI data is collected at 1 second intervals. Treiber and Kesting (2013) explored the impact of sampling rate on parameter estimation of the IDM (Intelligent Driver Model). A major finding was that the difference in the results is insignificant if the sampling interval does not exceed one second and the authors concluded that a sample interval of 1 second is sufficient for calibrating car following models.
Another important aspect of calibrating microscopic models is the effect of data smoothing on parameter estimation. For “noisy” trajectory data, data smoothing is important, particularly since noise is expected to increase with an increase in data sampling rate. Several data smoothing techniques have been proposed in the literature including locally weighted regression (Toledo, Koutsopoulos et al. 2007) and non-stationary Kalman filtering (Punzo et al. 2005). Preliminary evaluation of the WAMI data has indicated that it is less noisy than the data used by Punzo, et al (2005).

The actual impact of data smoothing on parameter estimation of flow models is still under debate. Treiber and Kesting (2013) explored the impact of data smoothing on parameter estimation and found that data smoothing is not necessary in global calibration. For local calibration, a small bias with the acceleration parameter was found but was determined to not be of great importance.

**Model Calibration**

This research explored two calibration options, the global deterministic approach and stochastic calibration. To most effectively take advantage of the WAMA data, the stochastic approach was structured using five components; Markov Chain Monte Carlo (MCMC) simulation using the Bayesian estimation theory, Statistical Inverse Problem, Bayesian Inference: from Prior to Posterior, Bayesian Estimate, and MCMC. Figure 17 provides a general overview of the calibration process. Preliminary results from both calibrations are provided after a summary of each of the techniques.

**Deterministic Calibration**

In the global deterministic approach, tracks are extracted from the simulation and compared with the observed tracks by formulating objective functions in terms of sums of squares of the variables to be calibrated. The choice of variables impacts which sum of squares is minimized and has an impact on parameter estimates. Speed-based measures were found to be insensitive to parameters controlling gaps, whereas acceleration based measures are insensitive to the desired speed of the vehicle (Punzo et al, 2012). Absolute and relative gap based objective functions have been found to be the most effective in parameter estimation using a minimum least square error method. Mathematical formulation of the objective functions of absolute and relative measure of the gap are provided in Equations 19 and 20, respectively.
\[
\begin{align*}
\text{Min, } S_{s}^{abs} &= \sum_{i=1}^{n}(S_{i}^{sim} - S_{i}^{obs})^2 \\
\text{Min, } S_{s}^{abs} &= \sum_{i=1}^{n}(\ln S_{i}^{sim} - \ln S_{i}^{obs})^2
\end{align*}
\]  

(19) \hspace{2cm} (20)

The objective function in equation 20 is less sensitive to outliers and measurement error than 19. This is because the objective function of Equation 19 is focused on larger gaps corresponding to acceleration or free flow regime. This objective function will provide a better fit to trajectory data on uncongested roadway sections. On the other hand, Equation 20 is a relative measure and can provide a better fit for all traffic conditions. Recent studies on this approach have revealed that the use of error measures as well as statistical GOF functions, can lead to ill-posed problems (Punzo et al, 2012). This is caused by the integral nature of the objective functions which locally cumulate the errors, but are unaware of dynamics of consecutive observations. Time series vehicle tracking data observations or data points are auto-correlated (time-wise correlated). To overcome this shortcoming, an Integrated Mean Square Error (IMSE) based optimization framework has been used (Montanino et al, 2012).

\textit{Stochastic Calibration}

In this approach, endogenous model variables (acceleration or gap) are compared with the observed data separately for each data point. This approach assumes explicit noise of a given distribution either in the model, the data, or both. For each time step, the data contain the vector of all exogenous variables needed for the model and a vector of all endogenous variables. A log likelihood function is defined as the joint probability that the model predicts all data points given a certain data parameter vector. The formulation of likelihood functions varies from model to model. Overall this approach provides global optimum solutions rather than local optima. This approach also provides variations of each parameter throughout the population.

\textbf{Step 1: Markov Chain Monte Carlo (MCMC) simulation using the Bayesian estimation theory.} The calibration process using the deterministic minimizing approach requires estimating the parameters to find the average value. In contrast, if the distribution of each of the car-following model parameters can be estimated, the aggregate behavior of all vehicles can be better simulated. A Bayesian framework can then be applied to the selected car-following model to provide a comparison with the deterministic approach.

This Bayesian framework uses prior distributions and vehicle trajectory data to estimate the statistical distribution of the parameters for the selected car-following model. The general framework is illustrated in Figure 18. As indicated, the prior probabilities are transformed into posterior probabilities for each parameter that uses the Bayes rule. The Metropolis Hasting algorithm then calculates the Bayes estimate of the model parameters. Finally, the model is validated against another data set. An overview of each of the five components is provided in the following sections.
Step 2: Statistical Inverse Problem. Bayesian statistics allows the inference of a relationship between the results of observation with theoretical predictions. Figure 18 shows the definition of an inverse problem. Given a parameter vector, v (q), the result of the observations is represented as an observation vector, m (q). If d is the actual observations of m, or more broadly, m + ε (noise), then P (d | q) is the conditional probability of the observation given the cause. The inverse is then P (q | d) which is the conditional probability of possible causes, given that some effect has been observed. This inverse probability represents the given knowledge of v after measuring m. For the WAMI data in the context of inverse problem theory, v (q) is the image and m (q) is the data.

Step 3: Bayesian Inference: From Prior to Posterior. In a Bayesian framework, the prior distribution of the parameter sets of the selected model are used to estimate the posterior distribution of the parameter using Bayesian inference. If q is the vector of model parameters (q1,
$q_2, q_3, ..., q_k^T$ with $k$ elements of the selected model and $v(q)$ is a generic model, then $m(q) = C[v(q)]$ is a model of observations, where $C$ transforms the generic model (i.e., velocity) to an observable quantity such as acceleration, headway, etc. The next step is to estimate the distribution of $P(d|q)$, called the Bayes estimate of the parameter.

**Step 4: Bayesian Estimate.** In this step, the prior distribution of the parameter of the selected model is assumed to be multivariate normal, which means that the mean square error is minimized:

$$P(d|q) \propto e^{-\frac{1}{2}(d-m(q))^T \Sigma^{-1}(d-m(q))}$$

(21)

Where, $\Sigma$ is the corresponding covariance matrix. Thus, the Bayes estimate, or expected value, is obtained from $q$ given by the following definition:

$$E(q|d) := \int_{\mathbb{R}^k} q P(d|q) \, dq$$

(22)

Because this integral is difficult to solve directly, the Markov chain Monte Carlo (MCMC) method is used to estimate the solution.

**Step 5: Markov Chain Monte Carlo (MCMC) Method.** A large number of random samples are generated from the posterior distribution of $q$ for the Bayes estimate. The Gibbs sampler and the metropolis Hasting algorithm are typical algorithms used to generate very large numbers of random samples. In the preliminary work, a special type of the Metropolis Hasting algorithm, shown in Figure , is used. For a more general form, see Gelman et al (2014).

If a large enough number of random samples, $(q^{(t)})_{t=1}^{r}$, are generated with $r$ random samples after an appropriate burn-in time (defined as the length of the time for the algorithm to run before collecting actual samples of the parameter) from the posterior distribution

---

**Figure 20:** Bayesian estimation process using Markov Chain Monte Carlo Method (source: Islam et al, 2016b)
of \(q\), the Bayesian estimate can be approximated by its sample mean:

\[
E(q|d) \approx \frac{\sum_{i=1}^{r} q^i}{r}
\]

(23)

With a given a current sample, \(q^i\), a new random sample, \(q^{i+1}\), is generated using the following algorithm:

Step 1: Generate \(\bar{q} \sim N(q^i, \Psi)\), where \(\Psi\) is a covariance matrix.

Step 2: Calculate the acceptance ratio:

\[
a := \frac{P(d|\bar{q})P(\bar{q})}{c} = \frac{P(d|q^i)P(q^i)}{c} = \frac{P(d|q^i)P(q^i)}{P(d|q^i)P(q^i)}
\]

(24)

Step 3: If, \(a \geq 1\), set \(q^{t+1} = \bar{q}\). Else set \(q^{t+1} = q^i\) with probability \(a\), and \(q^{t+1} = q^i\) with probability \(1-a\).

Step 4: Stop if \(r + b\) samples are produced, otherwise set \(t = t + 1\) and go to step 1.

In general, any parameter set can be used as the starting element \(q^0\). However, \(q^0\) may be selected from calibration results from previous studies, allowing the burn-in time to be minimized. This is important as running this algorithm can take substantial computational resources and time.

**Preliminary Results**

To perform the preliminary calibration, five minutes of microscopic data from two one-way major arterials, King and Main Streets are used. This consisted of 62,503 records of vehicle interactions. First, reaction time for every vehicle is estimated from a normal distribution with a mean of 1.5 seconds and standard deviation of 0.44 seconds as indicated by Triggs and Harris (1982). For subsequent points, one fifth of the standard deviation is used. This results in reaction times varying for every driver and also at every instance. Speed, acceleration, gap and relative speed are calculated for each reaction time by linearly interpolating between subsequent time stamps.

The preliminary calibrations are performed for the five model parameters of the linear Helly car following model: \(a, \beta, \gamma, C1\) and \(C2\) where \(a\) is the linear offset, \(\beta\) relates to the velocity of the follower, \(\gamma\) relates to the acceleration of the follower, \(C1\) relates to the relative velocity between vehicles and \(C2\) relates to the relative distance between vehicles.
**Deterministic Calibration**

Using the global deterministic approach and an objective function similar to Equation 20, the optimal parameter values were calculated and are shown in Table 9. The error in simulated acceleration is estimated for each individual track and varies from 12% to -28% which is similar to previous studies.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Calibrated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-0.9424</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.3930</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-1.2969</td>
</tr>
<tr>
<td>$C_1$</td>
<td>-0.1052</td>
</tr>
<tr>
<td>$C_2$</td>
<td>-0.0015</td>
</tr>
</tbody>
</table>

**Stochastic Calibration**

The MCMC algorithm is used to generate 200,000 samples of the parameters. Histograms and convergences of each parameter is shown in Figure . The calibrated parameters from the deterministic calibration in Table 9 is used as the starting point. As any distribution and covariance parameter is unknown, a large value of standard deviation is used for initial generation of parameters. The initial standard deviation of $\alpha$, $\beta$, $\gamma$, $C_1$ and $C_2$ was assumed to be 5, 3, 4, 1 & 1 respectively. Using the methodology outlined above, errors for every vehicle were estimated. These errors ranged from 9%-24% showing a slight improvement from the deterministic approach. Some resulting statistics from the 200,000 samples are shown in Error!

Reference source not found..

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Calibrated Value</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-0.2358</td>
<td>2.0525</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.0237</td>
<td>0.2236</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.8756</td>
<td>0.5734</td>
</tr>
<tr>
<td>$C_1$</td>
<td>-0.0518</td>
<td>0.4687</td>
</tr>
<tr>
<td>$C_2$</td>
<td>-0.0267</td>
<td>0.0982</td>
</tr>
</tbody>
</table>
Figure 21: Statistics Associated with Coefficients Resulting from Stochastic Calibration Process
A significant reduction in the variance from the initially assumed value is observed in the sample. This result indicates that the parameters are converging to an expected distribution. Key findings from the preliminary stochastic calibration include:

- All the parameters follow previously assumed distributions but their variance has significantly reduced when compared to the assumed values.
- All the parameters accept alpha converge very quickly.
- A wide range of results should be expected when considering the entire dataset.
- Data with longer durations will provide parameter distributions which represent driving conditions that have higher variability.

As indicated, these are preliminary results and a more detailed analysis and evaluation will be available in the M. R. Islam’s thesis.
CONCLUSIONS

This exploratory study assessed the use of persistent wide area motion imagery (WAMI) data for use in transportation planning and operations. Because the data include every vehicle with corresponding tracks, it has the potential to support the calculation of macro- and microscopic traffic characteristics for (1) individual vehicles, (2) individual tracks, (3) road segments, (4) classes of road segments, and (5) the global population of vehicles, tracks, and roads within the data collection area which can then be used to calibrate and validate models to estimate or predict these characteristics in comparable locations and situations.

The first phase of the research explored the data and ensured that the appropriate information was used for further analysis. This required filtering the point and trajectory data that was provided by PVLabs to select only those points that represented vehicles driving on the transportation system. This included developing a multi-dimensional filtering process that evaluated the length of trajectories in seconds, how much of a track occurred on the traveled way (road polygons), and how circuitous the track was. Each category was ranked from 1 to 4 and then weighted. Based on a comparison of vehicle points from the filtered tracks for a 15-minute period to vehicles manually located in a series of 5-second interval images, a cut-off of 2.2 was determined to effectively select actual vehicle tracks. The tracks were then map-matched to road centerlines within the frame to allow calculation of macroscopic traffic characteristics and to establish which lanes vehicles were occupying to ensure that the microscopic car-following measures were calculated for the correct leader-follower pairs in the traffic flow. The effectiveness of the map-matching algorithm was validated against a sample of 100 tracks and had a match rate of 97.9% on one-way arterials and 96.5% for two-way arterials and local streets.

A new data model was developed which efficiently stores microscopic characteristics with trajectory data. The schema developed by Brakatsoulas et al (2004) was modified to include the distance between vehicle centroids of the target vehicle and each of five vehicles in front and behind. Because the data is captured and stored at 1-second intervals which translates to vehicle speed, this structure supports the validation and calibration of the basic microscopic traffic submodels.

Several car-following and lane-changing models were reviewed in detail as was calibration and validation of these models. Results of this review provided the necessary background to formulate the process for use of WAMI data to calibrate car-following models. Both a global deterministic and stochastic process were applied to a 15-minute period for the portions of tracks that occurred on the two principle one-way arterials, King and Main Sts, which span the entire view frame and go through the middle of the CBD, for the Linear Helly model. The stochastic process was structured using five components; Markov Chain Monte Carlo
(MCMC) simulation using the Bayesian estimation theory, Statistical Inverse Problem, Bayesian Inference: from Prior to Posterior, Bayesian Estimate, and MCMC. The preliminary results indicated that the global deterministic calibration produced results similar to previous work. Results of the stochastic calibration showed a significant reduction in the variance from the initially assumed value indicating that the parameters are converging to an expected distribution. Other findings include:

- All the parameters follow previously assumed distributions but their variance is significantly reduced when compared to the assumed values.
- All the parameters accept alpha converge very quickly.
- Data with longer durations will provide parameter distributions which represent driving conditions that have higher variability.

This research showed the efficacy of using WAMI data to calculate macro- and microscopic traffic characteristics for any combination of time, location and class of roadway. One limitation that was identified was the inability to accurately classify vehicle type without manually reviewing imagery which could result in introducing error in some gap calculations. The research also supported the premise that having a complete population of vehicle tracks has the potential to improve calibration of current microscopic submodels for use in microsimulations.

RECOMMENDATIONS

This report presents results of trajectory data extraction, processing, and filtering; the development of a data schema to efficiently store tracks with microscopic traffic characteristics, and the preliminary results of a car following model calibration. Recommendations include the following:

- Complete the calibration and validation of select car following submodels for one-way arterials in an urban environment.
- Calibrate and validate car following submodels for other types of roadways including 2-lane local and collector roads, both way arterials in an urban environment.
- Assess the effectiveness of models in accurately representing micro- and macroscopic vehicle behaviors on dense urban networks.
- Assess the variability of micro- and macroscopic vehicle behaviors for different time increments and for morning and afternoon peak periods.
- Expand the evaluation of traffic submodels to lane changing and gap acceptance.
- Evaluate the transferability of the results to other locations and times.
These are only a few of the possible extensions to the exploratory research performed in this study. Because the processed dataset resulting from this research includes all vehicles and their trajectories for an area of approximately 4-square miles that includes a dense and complex urban network of roads over a three-hour period, the opportunities for mining and analyzing it are nearly inexhaustible.
REFERENCES


Islam, M. R. and K. Hancock (2016), "Vehicle Trajectory Extraction and Validation from Wide Area Motion Imagery Data for Use in Transportation Modeling", submitted to Elsevier Part XX.

Islam, M. R., K. Hancock, H. Rakha (2016b) "Parameter Estimation of Car Following Models Using Vehicle Trajectory Data from Wide Area Motion Imagery", under development.


Punzo, V., B. Ciuffo and M. Montanino (2012). "Can Results of Car-Following Model Calibration Based on Trajectory Data Be Trusted?" Transportation Research Record: Journal of the Transportation Research Board 2315: 11-24.


Triggs, T. J. and W. G. Harris, (1982). "Reaction time of drivers to road stimuli".

