# A STUDY ON USE OF WIDE-AREA PERSISTENT VIDEO DATA FOR MODELING TRAFFIC CHARACTERISTICS

Md Rauful Islam

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfilment of the requirements for the degree of

Master of Science
In
Civil Engineering

Kathleen L. Hancock, Chair
Pamela Murray-Tuite
Hesham Rakha

October 19, 2018
Blacksburg, Virginia

Keywords: Persistent Video Data, Vehicle Trajectories, Map Matching, Traffic Data, Car-following Model, Model Calibration.

A Study on Use of Wide-Area Persistent Video Data for Modeling Traffic Characteristics

Md Rauful Islam

## ABSTRACT

This study explores the potential of vehicle trajectory data obtained from Wide Area Motion Imagery for modeling and analyzing traffic characteristics. The data in question is collected by PV Labs and also known as persistent wide-area video. This video, in combination with PVLab's integrated Tactical Content Management System's spatiotemporal 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. This thesis explores the use of data collected by the PVLab's system for an approximate area of 4 square kilometers area in the CBD area of Hamilton, Canada for use in understanding traffic characteristics. The data was collected for two three-hour continuous periods, one in the morning and one in the evening of the same day. Like any other computer vision algorithm, this data suffers from false detection, no detection, and other inaccuracies caused by faulty image registration. Data filtering requirements to remove noisy trajectories and reduce error is developed and presented. A methodology for extracting microscopic traffic data (gap, relative velocity, acceleration, speed) from the vehicle trajectories is presented in details.

This study includes the development of a data model for storing this type of large scale spatiotemporal data. The proposed data model is a combination of two efficient trajectory data storing techniques, the 3-D schema and the network schema and was developed to store trajectory information along with associated microscopic traffic information. The data model is designed to run fast queries on trajectory information. A 15-minute sample of tracks was validated using manual extraction from imagery frames from the video. Microscopic traffic data is extracted from this trajectory data using customized GIS analysis. Resulting tracks were map-matched to roads and individual lanes to support macro and microscopic traffic characteristic extraction. The final processed dataset includes vehicles and their

trajectories for an area of approximately 4-square miles that includes a dense and complex urban network of roads over two continuous three-hour periods.

Two subsets of the data were extracted, cleaned, and processed for use in calibrating car-following sub-models used in microscopic simulations. The car-following model is one of the cornerstones of any simulation based traffic analysis. Calibrating and validating these models is essential for enhancing the ability of the model's capability of representing local traffic. Calibration efforts have previously been limited by the availability and accuracy of microscopic traffic data. Even datasets like the NGSIM data, are restricted in either time or space. Trajectory data of all vehicles over a wide area during an extended period of time can provide new insight into microscopic models. Persistent wide-area imagery provides a source for this data. This study explores data smoothing required to handle measurement error and to prepare model input for calibration. Three car-following models – the GHR model, the linear Helly model, and the Intelligent Driver model – are calibrated using this new data source. Two approaches were taken for calibrating model parameters. First, a least square method is used to estimate the best fit value for the model parameter that minimizes the global error between the observed and predicted values. The calibration results outline the limitation of both the WAMI data source and the models themselves. Existing model structures impose limitations on the parameter values. Models become unstable beyond these parameter values and these values may not be near global optima. Most of the car-following models were developed based upon some kinematic relation between driver reaction and expected stimuli of that response. For this reason, models in their current form are ill suited for calibration with noisy microscopic data. On the other hand, the limitation of the WAMI data is the inability of obtaining an estimate of the measurement errors. With unknown measurement errors, any model development or calibration becomes questionable irrespective of the data smoothing or filtering technique undertaken. These findings indicate requirements for development of a new generation of car following model that can accommodate noisy trajectory data for calibration of its parameters.

A Study on Use of Wide-Area Persistent Video Data for Modeling Traffic Characteristics

Md Rauful Islam

## GENERAL AUDIENCE ABSTRACT

The decision making process undertaken by transportation agencies for planning, evaluating, and operating transportation facilities relies on analyzing traffic and driver behavior in both aggregated and disaggregated manner. Different computational tools relying on representative models of aggregate traffic flow measures and/or driver behavior are used in the decision support system tools. Field data is used not only as an input for the computational tools but also to develop, calibrate, and validate the models representing a particular aspect of traffic and driver behavior. Different approaches have been undertaken to collect the data required for analyzing traffic and driver behavior. One of the applied approach is to collect trajectory (i.e. position, speed, acceleration) information of vehicles in the analysis zone. However, this data collection approach is often limited to relatively small stretch of a roadway for short duration due to high cost of collection and limitation of technology. As a result, the models developed and calibrated using these data often lack generalization power for different situation. This study explores the potential of a new data source to address the aforementioned limitations. The data used in this study collects the trajectory information for the whole population of vehicles in the study area by collecting wide-area (WAMI) video data. The data is collected by Canada based imaging solution company PV Labs. The collection area is relatively large to cover wide range of roadway types and traffic operation system. A framework has been developed to extract traffic flow measures from the trajectory data. The extracted traffic flow measures are then applied to calibrate the car-following model. The car-following model attempts to mimic the longitudinal movement of real-world drivers following another vehicle in front of them. The calibration results outline the limitations of the WAMI data. Although, this dataset is capable of capturing traffic measures for different driving condition, the lack of information about measurement error imposes limits on the direct application of the data for model calibration. Findings of this study can be applied for refinement of the video data capture technology and subsequent application in modelling traffic characteristics as well as development of new and calibration of existing driver behavior model.

## ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere appreciation to Dr. Kathleen Hancock for providing me the opportunity to conduct this research and for being an eager helping hand in any problem I faced during the research process. I have learned an insurmountable amount of knowledge from her both academic and otherwise. She is a great friend, mentor, and a motivational guide. I would also like convey my appreciation to the members of my committee, Dr. Pamela Murray-Tuite, and Dr. Hesham Rakha, who have provided valuable input to improve the scope and the basis of the research, and for being a guiding hand throughout my Graduate Studies. Additionally, I would like to express my gratitude to the instructors from whom I have taken courses here at Virginia Tech, namely Dr. Alireza Haghighat, Dr. Kevin Heaslip, Dr. Samuel Tignor, Dr. Antoine Hobeika, and Dr. Gerardo Flintsch. I am especially in debt to Dr. Alireza Haghighat for providing valuable input into some of the complex analytical process adopted in this research.

Thanks to the CEE, and other NCR facility stuff especially Jeny Beausoliel, Neil Bradish, Erik Finlayson, and James Murphy for all their support. I want to thank all my colleges in the Systems Urban Mobility Laboratory (SUMLab) for their support, advice and for enriching my experience at Virginia Tech in various ways. Special thanks to my roommates Arifur Rahman, and Mahmudul Hasan for their friendship, support, and care. I cherish their friendship.

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

Finally, I would like to dedicate this thesis to my loving parents for reasons all too familiar, which I am incapable of expressing in written words.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ACRONYMS

| | |
|---|---|
| AP | Action Point model |
| CA | Collision Avoidance model |
| CBD | Central Business District |
| DBMS | Database management systems |
| DGPS | Differential Global Positioning Systems |
| DoT | Department of Transportation |
| FHWA | Federal Highway Administration |
| GF | Generalized Force model |
| GHR | Gazis-Herman_Rothery car-following model |
| GM | General Motor's model |
| GOF | Goodness of Fit |
| IDM | Intelligent Driver Model |
| IMSE | Integrated Mean Square Error |
| MoP | Measure of Performance |
| MPO | Metropolitan Planning Organization |
| NAD | North American Datum |
| NDS | Naturalistic Driving Study |
| NGSIM | Next Generation SIMulation |
| OD | Origin-destination |
| OV | Optimal Velocity model |
| SHRP | Strategic Highway Research Program |
| UTM | Univeral Transverse Mercator coordinate system |
| WAMI | Wide Area Motion Imagery |
| WGS84 | World Geodetic System 1984 |

# 1. INTRODUCTION

State Department of Transportation (DoT), Metropolitan Planning Organizations (MPO), and other public agencies in charge of management and development of transportation infrastructure face increased demands for limited resources. To address this demand more reliance is placed on advanced technologies and computational tools to support decision making system. These tools in turns help facilitate in planning, evaluating and operating transportation facilities. Most of these tools rely on representative models of aggregate traffic flow measurements and/or microscopic measures of vehicle or driver behaviors. In most cases, these models were developed, calibrated and validated using sample data which were generally fixed in time or space or both. However, there is a stark contrast between the planning models with that of operations models. The planning and evaluation models mostly leverage large-scale socio-economic and demographic data. On the other hand, the operations models are largely based on small-scale vehicle characteristics, traffic flow parameters. and driver behavior data. New generation of data sources such as mobile sensors and naturalistic driver data are becoming available to address the gap in microscopic traffic data sources. However, capturing movement information about every individual vehicle trajectory over a large area for an extended period of time has not previously been available to transportation researchers or practitioners. Vehicle trajectory data collected by means of Wide-Area Motion Imagery (WAMI) can be helpful in addressing this data need.

The capability to collect airborne persistent wide-area motion imagery (WAMI) data was originally developed to support military, and surveillance activities for use in intelligence collection and reconnaissance. Enhanced camera platform stabilization techniques, advanced image capturing 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, Turcotte et al. 2012). Now, 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, Turcotte et al. 2012). Current examples of public sector traffic related information derived from

WAMI include intersection vehicle movement counts, vehicle classification, and origin-destination counts for complicated interchange or critical area. Most of these measures consist of counts of vehicles at fixed locations within the view frame. A powerful potential, which this study begins to explore, is the capture and use of vehicle trajectory data extracted from video, which provides paths of individual vehicles linked to corresponding spatiotemporal derived characteristics. The comprehensive nature of WAMI data and its ability to capture information about both microscopic vehicle behavior and macroscopic traffic flow information means it has the potential for reevaluating the microscopic/macroscopic models used thus far as well as a cornerstone of developing new models. In the following sections, the motivation behind exploring new data sources as well as their potential uses in both macroscopic and microscopic traffic data extraction is explained briefly.

### 1.1. MOTIVATION FOR USING NEW DATA SOURCE

Traditional traffic data collection efforts can be broadly classified into two groups, macroscopic and microscopic data. Macroscopic data includes traffic flow, origin-destination, average speed, and travel time. Traditionally, these data have been collected by means of on-road sensors, surveys, license plate recognition or simply by roadside interview. Some information like origin-destination (OD) is very difficult and costly to get. Usually, OD data has been collected in an aggregate manner by means of different types of surveys. Loop detectors have been used to collect traffic flow and travel speed. Recently, many novel data collection methods (i.e. cellular phone, car blue-tooth tracking) have been proposed for the purpose of estimation of OD, flow, and speed. Several prediction models have been developed to predict congestion level on the roadway based on these data. Although technologically superior to traditional methods, even these novel technologies lack the extensive trajectory information that is available from WAMI data.

Microscopic traffic data collection has been more challenging in the past. Two methods most popularly used previously are instrumented vehicles and roadside video. Vehicles are usually instrumented with LIDER, DGPS, and even with a video camera to collect inter-vehicle spacing, relative velocity, speed, and acceleration data (Hoogendoorn, Van Zuylen et al. 2003, Ranjitkar, Nakatsuji et al. 2005, Marczak and

Buisson 2012). These data collection processes usually suffer from significant measurement error (Punzo, Formisano et al. 2005, Ossen and Hoogendoorn 2009). Data sampling interval also plays important role in the accuracy of the data (Treiber and Kesting 2013). Using the instrumented vehicle method, data from a small sample of driver population can be obtained and the data may not be representative of actual driver behavior as this method may suffer from influenced drivers. Roadside video data is an improvement in this aspect. Roadside video data is usually processed using computer vision algorithm to obtain vehicle trajectory (Alexiadis, Colyar et al. 2004). They are usually restricted spatially to a stretch of the road and suffers from traditional perks of image processing data like false detection, no detection, vehicle overlapping and discontinuity (Punzo, Borzacchiello et al. 2011). So, there is a justifiable case for exploring new microscopic data source. Wide area motion imagery data opens a new door in this aspect as it comes in the form of vehicle trajectory data that can be used to extract microscopic traffic data for a larger spatiotemporal extent.

WAMI data is usually collected at an interval from an airborne platform attached to either fixed winged of rotary aircraft. PV Labs, the provider of the WAMI data presented in this thesis used fixed-wing aircraft to collect the data at 1-second interval. Images are then processed using PV Labs' own proprietary computer vision algorithm to locate the vehicles in the image. One of the most important obstacles in using any computer vision algorithm is the case of false detection and/or no detection. The WAMI data is similar to other computer vision algorithm in this regard. However, as the PV Labs' vehicle tracking algorithm store vehicles with individual tracking ID, speed, and heading, several data filters can be designed using several trajectory properties to extract the most likely trajectory that represents actual vehicles on the network.

As trajectory data has both two spatial dimension along with additional time dimension attached to it handling and storing data can become cumbersome, hence the need for a data model suited for this particular task. Another challenge of using a large-scale trajectory dataset is to develop data models that can facilitate efficient handling of a different query run on the database.

Thus, there is ample motivation for using WAMI data to explore large spatiotemporal scale microscopic/macroscopic traffic behavior. To accomplish this task, data filtering is also required to validate the vehicle trajectory as well as to develop efficient data model to store this type of large-scale data.

## 1.2. MOTIVATION FOR IMPROVED CALIBRATION AND VALIDATION OF SIMULATION MODELS

Microscopic traffic simulation is a frequently 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 the facility and/or traffic characteristics. However, use of simulation must be done in a cautious manner as the underlying theory and model developed for ideal conditions may not match the ground truth an analyst trying to assess. One way to avoid this shortfall of "black box" simulation is to calibrate the models describing traffic flow in a network using ground truth data.

Because of increasing use of simulation, improving calibration and validation processes is gaining attention. Difficulty in obtaining microscopic behavioral data has been regarded as a considerable drawback in understanding real phenomena and has affected the development and calibration of traffic simulation model (Punzo, Formisano et al. 2005). Traditionally calibration and validation effort have concentrated on matching macroscopic measures (flow, speed, capacity) generated by the micro-simulation model with representative field data. This is due to the fact that most of the models used in micro-simulation are based on sub-models that are estimated using conceptual bases supported by empirical data (Brackstone and McDonald 1999).

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. 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 and McDonald 1999, Punzo, Formisano 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 and McDonald 1999). 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. However, in the current state, the PVLabs data does not provide any information on the type or the size of the vehicle. Vehicle type and size is an important input for almost all discretionary lane-changing model. As a result, the PVLabs trajectory data more suitable for calibration of car-following model.

There continues to be a lack of established methods for calibration of microscopic model parameters. Traditionally, parameter estimation has been indirectly determined by setting up an appropriate optimization framework that attempts to reproduce macroscopic performance measure of 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 and Hoogendoorn 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 and Hoogendoorn 2009). Optimization methods used for parameter estimation have also been shown to produce significantly different results (Punzo, Ciuffo et al. 2012).

Irrespective of differences in outcome, these calibration studies are important because they provide the opportunity to assess the performance of existing car-following models. The outcomes of these studies are used to obtain empirical insights into car-following behavior. Insights can include the extent of heterogeneity between drivers (Ossen and Hoogendoorn 2005) and the degree of multi-anticipative or spatiotemporal variability of driver anticipation (Hoogendoorn, Ossen et al. 2006). Each methodology has

its own strengths and weakness, but a common weakness is that none cover a large geographical area consisting of different types of transportation facilities. So there is a need for calibration of the car-following model using microscopic dataset collected in a relatively large spatial area and covering longer periods of time.

## 1.3. OBJECTIVES OF THE RESEARCH

WAMI data in form of vehicle trajectory open a new horizon for transportation data collection and analysis. The objective of this study was to explore the use of WAMI data for use in modeling traffic behavior. To achieve this objective, the research was twofold. First, the vehicle trajectory data obtained by means of WAMI was explored for its quality, attributes, consistency, and coverage. Before applying this new data source, its quality required measuring and appropriate filtering to reduce data noise expected from this kind of data. Data noise can come from false detection, no detection, image registration, and vehicle tracking method used by PVLabs. Several geographic analyses were also required for extraction of macroscopic and microscopic traffic data from the trajectory data. This filtered trajectory data was then used to extract useful traffic related data. The second objective of the research ws to assess the quality of the extracted microscopic data and use it for calibration of car-following models. Like any other microscopic data, extracted microscopic data may need data smoothing to reduce data noise. Any data smoothing method should be limited by individual vehicle trajectory consistency as well as platoon consistency.

To achieve the objectives of the study the following task were accomplished:

- Develop filtering method to extract true vehicle trajectory from noisy dataset
- Develop map-matching algorithm to associate 2-dimensional trajectory data to appropriate roadway segment
- Understand Storage and Processing of large-scale vehicle trajectory data
- Extract macroscopic traffic data (flow, density, OD) from the trajectory data
- Extract microscopic traffic characteristics to support calibration work on car-following models from the tracking data
- Apply data smoothing technic on microscopic data to reduce data noise

- Calibrate car-following model by both deterministic and stochastic approach using smoothed microscopic data.
- Validate calibration result by comparing with the previous studies

## 1.4. THESIS OUTLINE

This thesis is organized into five individual chapters. Chapter 3 provides an overview of the data used in this study. Data overview also includes a brief description of data collection system used by PVLabs and attributes of the data and data structuring requirement for this particular type of data. Chapter 2 focuses on an in-depth literature review on the use of time series tracking data in development and calibration and/or validation of microscopic traffic simulation model. This chapter also provides an overview of different data models proposed in the literature. Chapter 4 discusses parameter estimation of car-following models using vehicle trajectory data from wide area motion imagery. Chapter 5 highlights key findings of this study and provides direction for future research.

# 2. LITERATURE REVIEW

A literature review was performed to identify current uses of trajectory data for obtaining traffic characteristics as well as developing, calibrating and validating simulation models. Recently, several companies have used wide-area motion captures to obtain traffic data. However, the dataset in this thesis is different in content and larger in size. A thorough study of current practice on how trajectory data are currently represented in both transportation and other fields that use spatiotemporal data.

After a review of data uses and storage models, common sub-models that calculate the longitudinal and lateral movement of the simulated vehicles used in microscopic simulation were reviewed. These models include car-following, lane-changing, and gap acceptance model. As discussed in Chapter 1, this study is focused on car-following. Although, lane-changing and gap acceptance models are beyond the scope of this research, a brief review of lane-change and gap acceptance models used in commercially available simulation packages was included to understand possible interactions with car following in microscopic simulation.

## 2.1. USE OF VIDEO AND TRAJECTORY DATA IN TRAFFIC ANALYSIS

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 uses the similar data collection technology used by PVLabs. Miovision uses similar data collection method to measure several macroscopic traffic performance matrices including traffic volume, speed, delay time, signal time split, signal phase occupancy ratio, and traffic monitoring in a construction zone. Data From Sky uses Traffic Drone to collect similar aerial imagery and extract similar traffic information as Miovision. TrafficVision uses static cameras to collect volume, movement type, incident monitoring, and construction management related data.

## 2.2. USE OF VIDEO AND TRAJECTORY DATA IN MICROSCOPIC MODEL DEVELOPMENT

The methodology of estimating vehicle trajectory information can be broadly classified into two categories; i) post-processing of vehicle positional data collected from onboard sensors, usually GPS devices (instrumented vehicle) and ii) post-processing of video data (remote sensing). Processing of vehicle positional information obtained from onboard sensors is relatively straightforward due to the availability of GPS devices and geospatial tools. Onboard 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 (Ossen and Hoogendoorn 2008, Ossen and Hoogendoorn 2009, Punzo, Ciuffo et al. 2012). Although video and other remote sensing data are 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 questionable. 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 is discussed in more detail in the following paragraphs.

Trajectory data for use in model development, calibration, and/or validation using onboard sensors are mostly collected under controlled test conditions using a limited number of vehicles. Gurusinghe *et al.* collected GPS data from a platoon of ten vehicles on two parallel sections (Gurusinghe, Nakatsuji et al. 2002) with a sampling rate of 0.1 seconds, the authors determined that this high-resolution data can facilitate accurate car-following analysis. The same dataset was then used by Ranjitkar *et al*. to examine the stability of stimulus-response based car-following models with increased platoon size and found large intra and inter-vehicle variation in reaction time (Ranjitkar, Nakatsuji et al. 2003). The authors concluded that the stimulus-response based car-following model might be too simple to explain car-following behavior. Brockfeld *et al.* collected GPS track data from nine vehicles on a test track in Japan for calibration and validation of 10 different car-following models (Brockfeld, Kühne et al. 2004). The authors concluded

that the differences between individual drivers are larger than the differences between different models and because these models were developed for specific conditions, they are not capable of generalizing to other situations. Ranjitkar *et al.* 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 which is contradictory to the findings of Brockfeld *et al*. (Brockfeld, Kühne et al. 2004, Ranjitkar, Nakatsuji et al. 2005). These findings have importance in relation to this study. 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, Nakatsuji 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 and Simonelli 2005). This dataset was used to calibrate four microscopic car-following models 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.* and Brockfeld *et al.* where intra-personal and inter-personal variability was greater than the 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 example of terrestrial video data collection was done in the Next Generation Simulation (NGSIM) program. The 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 were 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 seconds. Some analysis of NGSIM data has shown processing issues such as vehicle overlaps and unusual change in acceleration. 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, Ramanujam et al. 2009, Wan, Jin et al. 2014), arterial lane selection model for urban intersections (Choudhury and Ben-Akiva 2008), lane position and lane-changing execution model for different vehicle classes (Moridpour, Sarvi et al. 2010, Aghabayk, Moridpour et al. 2011). This dataset has also been used in calibrating existing models (Monteil, Billot et al. 2014), as well as determining the effect of correlation between model parameters (Kim and Mahmassani 2011).

Ozaki used video data collected from the 32$^{nd}$ floor of a city office building and found that driver's reaction times were not stable over time, space, or driving condition (Ozaki 1993). Ahn *et al.* 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 and Hoogendoorn used a similar approach to collect data from Freeway A2 in Utrecht, in the Netherlands to estimate the parameters for the stimulus-response based (GHR) car-following model and found that car-following behavior varies widely within the observed driver population (Ossen and Hoogendoorn 2005). The same dataset was used in calibration and parameter estimation of other car-following models,(Hoogendoorn and Hoogendoorn 2010, Hoogendoorn, Hoogendoorn et al. 2011) and for determining heterogeneity in car-following characteristics (Ossen and Hoogendoorn 2011).

A similar approach was used for the development of the multi-anticipated car-following model by Hoogendroon *et al.* (Hoogendoorn, Ossen 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, Olson et al. 2006, Board 2013). Disadvantages of this method include the possibility that drivers may be influenced by the cameras and 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.

## 2.3. DATA MODEL FOR STORING TRACK 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 are 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 (Brakatsoulas, Pfoser et al. 2004). 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 important property 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 (Brakatsoulas, Pfoser et al. 2004). 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 database size necessary for this project.

Developing trajectories dataset includes the following task: (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 (Brakatsoulas, Pfoser et al. 2004). When defining the data model, additional considerations include establishing the relationship between a trajectory and its environment, and interactions between individual trajectories. Examples of the former include determining the position of the vehicle on the road such as distance from upstream roadway node, distance to downstream roadway node, lane position etc. at each discrete time step. Examples of interactions between trajectories include headway between vehicles and relative speeds and accelerations.

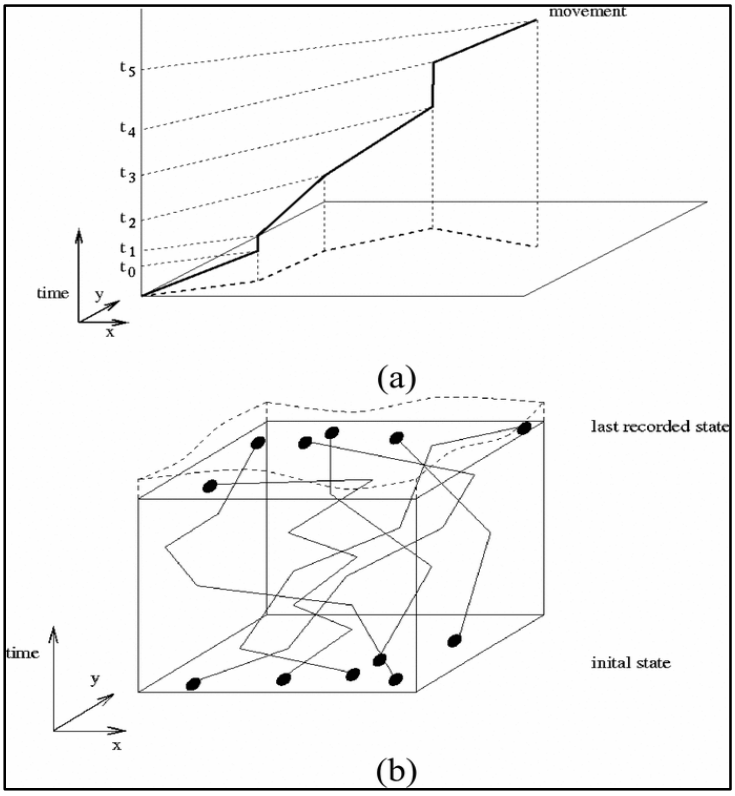



**Figure 1: 3-D Representation of Trajectory Data (Source: (Brakatsoulas, Pfoser et al. 2004))**

Brakatsoulas *et al.* proposed two different schemas for preprocessing and storing large trajectory databases: the 3-D

schema and the network schema. In the 3-D schema, only the vehicle position is recorded at a discrete interval of time. The data can be represented as a 3-dimensional plot of X and Y in space with time as the third dimension as shown in Figure 1. 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. PVLabs provides the data in this format. However, this format cannot be used to determine the interaction of a vehicle with its environment. Out of the two data storing method, the network schema was determined to be more effective.

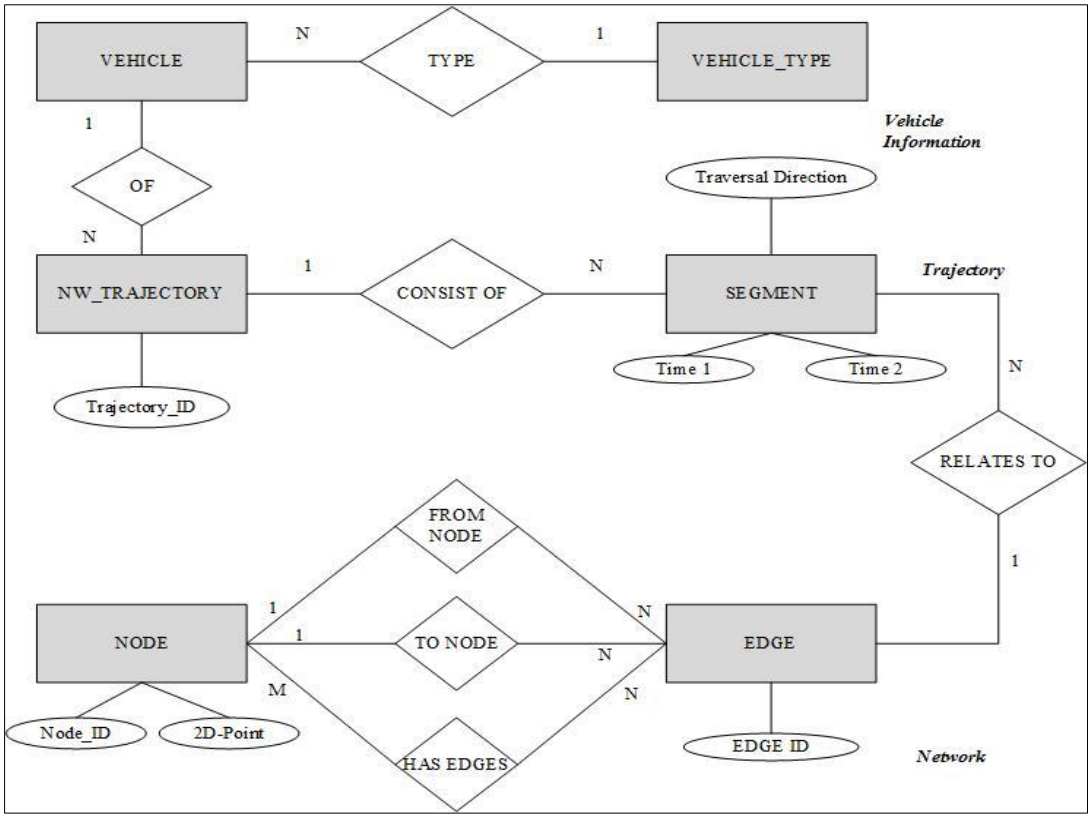



**Figure 2: Network Data Schema of Trajectory Data (Source: (Brakatsoulas, Pfoser et al. 2004))**

The Network Schema uses a map-matching algorithm and is illustrated in Figure 2. The method of associating positional data to a particular position on a roadway is known as map-matching. In this method, the trajectory data is stored in six tables as indicated by shaded boxes in Figure 2. Each trajectory is stored as a 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 (Brakatsoulas, Pfoser et al. 2004). 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.

To make a better use of geo-processing and visualization capabilities of GIS software, the data need to reside in a single database which is not achievable by the network schema. Another shortcoming of the network schema is the mismatch in geographical resolution between the tracking layer and the roadway layer. Brakatsoulas *et al.* used GIS tracking data which was collected at a sample rate of 30 seconds. During a 30 second, 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.* will require modification as discussed in Chapter 3.

## 2.4. MICROSCOPIC TRAFFIC FLOW 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. This section presents the theoretical background and parameters of the car-following and lane-changing models. As the parameters are the only element of the model which can be manipulated to obtain an output that matches reality, calibration effort is focused on finding the approximate value of these parameters using optimization technique.

### 2.4.1. CAR-FOLLOWING MODEL

The mathematical formulation describing the behavior of vehicles following each other in the traffic stream is commonly known as a 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). Detail understanding of the car-following and lane-changing models is required to set up suitable calibration and validation method. These models are presented below along with their advantages and disadvantages.

**Gazis-Herman-Rothery (GHR) model (GM Model):**

The GHR model and related GM model are one of the oldest and well-studied of the car-following models. The first generation was put forward in the late 1950's by Chandler *et al.* (Chandler, Herman et al. 1958) and was based on the hypothesis that driver's reaction (acceleration/deceleration) to any number of traffic conditions is proportional to the difference between the speed of the leader and follower vehicle. The structure of the model puts it into broad classification of the stimulus-response model. The basic structure of the model can be shown in equation 1.

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

Where the *response* is represented as vehicle acceleration/deceleration by which driver control the vehicle, the *stimulus* is the relative speed between the leader and follower vehicle. *Sensitivity* is the only term needs to be estimated and varies with different model and different drivers. This model also includes driver's perception-reaction time T as the delay in driver's response. The delay in response is due to the time lapses before the driver notices a change in the driving situation and can react to the situation. 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 not 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).

$$a_n(t) = cv_n^m(t)\left[\frac{\Delta v(t-T)}{\Delta x^l(t-T)}\right] \quad (2)$$

Where,

an (t) = Acceleration/Deceleration of the driver at time t
C, m, n = Constant to be determined
$v_n\ (t)$ = Velocity of the follower vehicle
$\Delta v(t-T)$ = Relative velocity at time t-T
$\Delta x\ \ (t-T)$ = Relative Gap between leader and follower
T = Perception Reaction time

Many studies were undertaken to calibrate this model using different combinations of *m* and *l* value (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 contradictorily resulting in a wide combination of *m* and *l* value (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.

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

The concept of safety distance was first proposed by one of the pioneers of 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 3.

$$\Delta x(t-T)_{min} = L_n + \frac{L_n * v_n\,(t)}{10} \quad (3)$$

Where,

| | |
|---|---|
| $\Delta x(t-T)_{min}$ | = Minimum Headway Distance |
| $L_n$ | = Length of the Vehicle |
| $v_n\,(t)$ | = Speed of the vehicle in Mile per hour |

The length of the vehicle is also considered to be the jam-density spacing which is defined as the clearance distance between cars at a 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 4.

$$\Delta x(t-T)_{min} = 1.34 v_n(t) + 6 \quad (4)$$

The Pipes model assumes a linear relationship between vehicle's desired speed and spacing as well as vehicle acceleration and the relative speed between the leader and follower vehicle. This relationship can be expressed as Equation 5 and 6. 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 a larger gap.

$$v_n(t) = \min\big(\lambda[x_{n-1}(t-T) - x_n(t-T) - L_n], v_f\big) \quad (5)$$

$$a_n(t) = \lambda[v_{n-1}(t-T) - v_n(t-T)] \quad (6)$$

Where,

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

Field studies conducted by Rakha and Crowther on the Pipes model have suggested that even with simplistic assumptions, speeds prediction by 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 was proposed by Van Aerde and Rakha which is an amalgamation of Pipes and Greenshield's macroscopic traffic flow model (Van Aerde 1995, Van Aerde and Rakha 1995). This model introduces an additional non-linear term to address the insensitiveness of Pipes model to traffic density at higher speeds. The mathematical formulation is shown in Equation 7.

$$\Delta x_n(t-T) = C_1 + C_3 v_n(t) + \frac{C_2}{v_f - v_n(t)} \quad (7)$$

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 |
| C1, C2 & C3 | = Model Parameter |

The first two terms of the Van Aerde model provide linear increases of headway with speed while the third term ensures that spacing increases when speed approaches free flow speed of the facility. This model also ensures 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, the speed at capacity, capacity and jam density spacing. This is an advantage since calibration can be done using easily collectible macroscopic data (Rakha and Arafeh 2010). This model 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 8.

$$\Delta x(t-T)_{min} = \alpha v_{n-1}^2(t-T) + \beta_1 v_n^2(t) + \beta v_n(t) + b_0 \qquad (8)$$

Where,

| | |
|---|---|
| $\Delta x(t-T)_{min}$ | = Minimum Headway |
| $v_{n-1}(t-T)$ | = Speed of the leader |
| $v_n\ (t-T)$ | = Speed of the follower |
| α, β1, β, b0 | = Constant to be determined |

Model parameters α 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. β 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 and proposed 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). The overall consensus is that improved model 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. Helly proposed a car-following a 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 Equation 9 and 10.

$$a_n(t) = C_1 \Delta v(t-T) + C_2 [\Delta x(t-T) - D_n(t)] \quad (9)$$

$$D_n(t) = \alpha + \beta v_n(t-T) + \gamma a_n(t-T) \quad (10)$$

Where,

| | |
|---|---|
| $a_n(t)$ | = Acceleration of the follower vehicles at time t |
| T | = Driver Reaction time |
| $\Delta v(t-T)$ | = Relative velocity between vehicles at time t-T |
| $\Delta x(t-T)$ | = Relative distance between vehicles at time t-T |
| $v_n(t-T)$ | = Velocity of the follower at time t-T |
| α, β, γ, C1, C2 | = Calibration constant |

Subsequent calibration of the model resulted in wide-ranging values of T, $C_1$, and $C_2$. Calibration of this model requires estimation of 6 parameter T, $C_1$, $C_2$, α, β and γ, which makes it more difficult than most (Panwai and Dia 2005). In later years several variations of this model and their calibration result have been proposed. Although this model 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 a 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 the 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). The first threshold is of relative velocity. It is described as $d/dt\ (\sim \Delta v/\Delta x^2) \sim 6 \times 10^{-4}$. If this threshold is exceeded, the driver will decelerate until a relative velocity is no longer perceived. If the perception of relative velocity remains under this threshold then the 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 the change in visual angle. Typically 10% change in visual angle is required for the driver to perceive that he/she is closed 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 a 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 to simulate traffic.

**Intelligent Driver Model (IDM):**

The Intelligent Driver Model (IDM) is collision frees continuous function model proposed by Treiber *et al* (Treiber, Hennecke et al. 2000). This model is a multi-regime model where the vehicle acceleration is defined as a continuous function of dynamics of the lead-follower pair of vehicles. The main difference is that the IDM doesn't use the time lag used in the previous model in term of the reaction time. Instead, the IDM tries to incorporate the desired time headway parameter value to incorporate the similar effect of the reaction time. The mathematical formulation of the model is shown in 25 and 26.

$$\dot{v_\alpha} = a\left[1 - \left(\frac{v_\alpha}{v_0}\right)^\delta - \left(\frac{S^*(v_\alpha, \Delta v_\alpha)}{S_\alpha}\right)^2\right] \tag{11}$$

$$where, S^*(v_\alpha, \Delta v_\alpha) = S_0 + S_1\sqrt{\frac{v_\alpha}{v_0}} + T^\alpha v_\alpha + \frac{v_\alpha \Delta v_\alpha}{2\sqrt{ab}} \tag{12}$$

Where,

$\dot{v_\alpha}$ = Acceleration of vehicle α at any time
$a$ = Desired maximum vehicle acceleration
$b$ = Desired maximum vehicle deceleration
$T^\alpha$ = Desired time headway
$v_0$ = Desired free flow speed
$S_0$ = Minimum desired headway/Jam distance
$v_\alpha$ = Current velocity of vehicle α
$\Delta v_\alpha$ = Relative velocity between vehicle αand its leader
$S_\alpha$ = Current spacing of vehicle α

According to Treiber *et al* (Treiber, Hennecke et al. 2000) the parameter $S_1$ can be considered as zero and the value of $\delta = 4$. So the there are five parameters ($a, b, T^{\alpha}, v_0, and\ S_0$ ) that needs calibrating.

**Other Car-Following Models:**

Other than the car-following models already discussed many more car-following models have been proposed using many different approaches. 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 13 and 14.

$$a_t(t) = k[V * \Delta x(t) - v_n(t)] \quad (13)$$

$$V(\Delta x) = V_1 + V_2 \tanh[C_1(\Delta x - l_c) - C_2] \quad (14)$$

Where,

| | |
|---|---|
| $a_n(t)$ | = Acceleration of the follower vehicles at time t |
| $\Delta x(t)$ | = Relative distance between vehicles at time t |
| V | = Optimum velocity function |
| $V_1$ | = 6.75 m/sec |
| $V_2$ | = 7.91 m/sec |
| C1 | = 0.13 m-1, |
| C2 | = 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 condition (Gong, Liu et al. 2008). However, unrealistic acceleration and deceleration have 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 break, 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. The potential use of these models in a comprehensive simulation environment is yet to be explored.

### 2.4.2. LANE-CHANGING AND GAP-ACCEPTANCE MODELS

Modeling lane-changing behavior is a complicated issue. The complication arises due to the 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 also consist of complex processes. Lane-Changing 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 3 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 3, have many different forms. These models are typically associated with specific micro-simulation software packages and are explained as they are implemented in a multistep process instead of as an independent formulation. In current state the PVLabs trajectory data does not have adequate information like vehicle size and type that is required for analyzing lane changing behavior or even the fidelity

(preferably 0.1 sec) required for gap acceptance model. As a result, any further investigation on this issue is not pursed in this study.

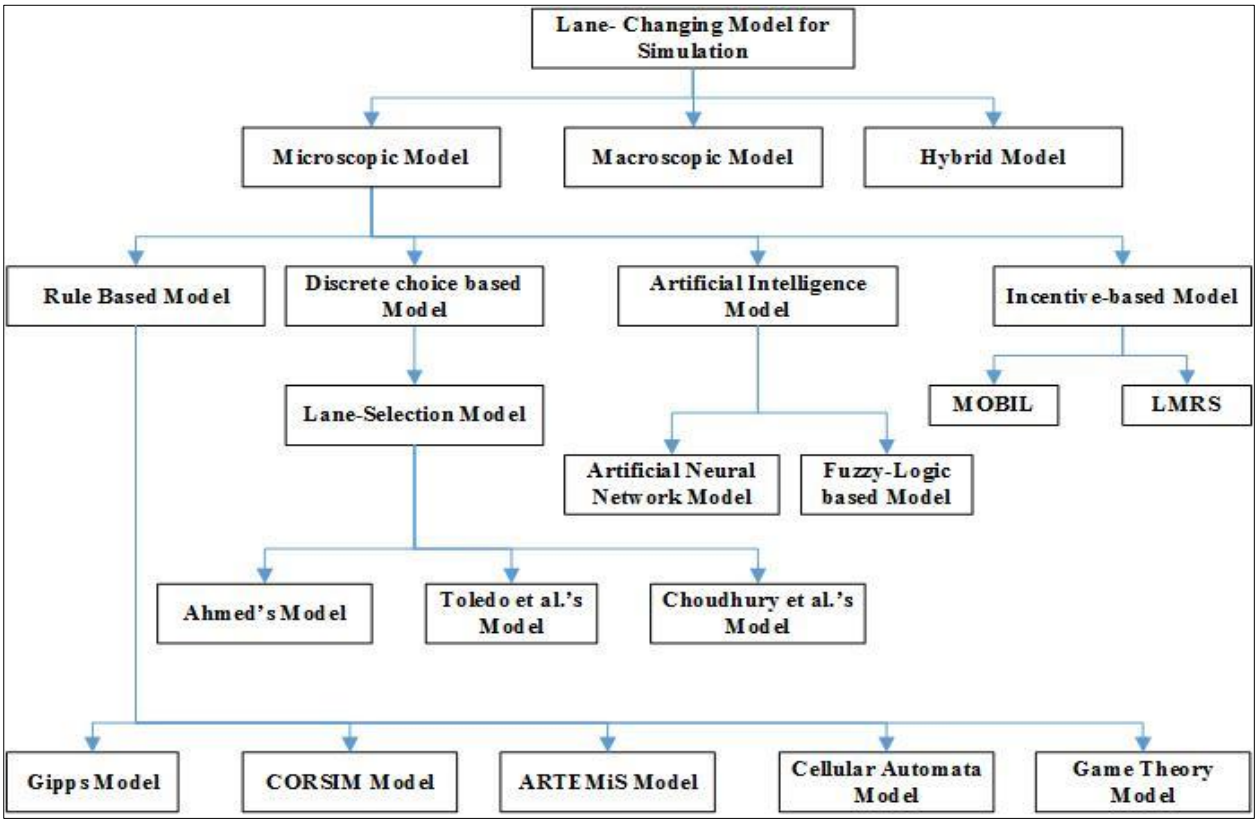



**Figure 3: Classification of lane-changing models (Source: (Rahman, Chowdhury et al. 2013))**

## 2.5. SUMMARY

Several car-following models were reviewed in detail, as was calibration and validation of these models. Data structure requirements for large-scale trajectory data were also studied. Results of this review provide the necessary background to begin using the WAMI data to improve and calibrate existing models. The models reviewed vary in their structure and conceptual background. Any model has a certain realm of validity. In traffic flow theory, the question of validity has been answered by what has been called the visual Turing test: it looks realistic (Brockfeld and Wagner 2006). Several studies have been undertaken to examine the strength and weakness of these models. None of the models stands out as best or worst (Brockfeld and Wagner 2006). However, studies suggest that microscopic data is better suited for evaluating microscopic parameter of traffic flow models especially the car-following model.

# 3. USE OF VEHICLE TRAJECTORY FROM WIDE-AREA MOTION IMAGERY FOR TRAFFIC DATA MINING

## 3.1. INTRODUCTION

This chapter provides details description of the WAMI trajectory data used in this study. Previous attempts of collecting vehicle trajectory data by different methods is discussed along with their strengths and weaknesses. WAMI data is then compared with previous data collection attempts. A brief discussion is provided on the data collection method as well as attributes included with trajectory points. Because the extraction process of vehicle trajectory from aerial images is proprietary of PVLabs, the focus of this study is on the resulting trajectory data and not on data extraction from the images. Vehicle trajectory comes in the form of a sequence of point locations of the vehicle in the view frame. Data was collected for two 3-hour peak period on the same day in Hamilton, Ontario, Canada.

Initial studies of the trajectory data revealed a significant amount of data noise due to false detection. A filtering method was then developed to reduce data noise using three attributes of individual trajectories namely its duration in the view frame, its location relative to the roadway, and sinuosity of the vehicle path constructed from trajectory points. The filtering process is a simple weighted attribute filtering process, which is then validated using vehicle count from 15 minutes of image data.

Filtered vehicle trajectories are then attached to roadway segment by means of map-matching. This helps to transform vehicle movement on two-dimensional space on to linear road section. This matching helps to obtain traffic-related information from a sequence of trajectory points. The position of trajectory points and the direction of travel is used to develop the map-matching algorithm. The accuracy of the map-matching algorithm is also estimated. Trajectory data is then stored in a data model developed for storing large-scale three-dimensional data along with related traffic and route information.

Finally, different macroscopic and microscopic data extraction processes are discussed. Macroscopic data includes OD of individual trajectory, traffic flow, travel speed/time and their distribution.

The microscopic data includes instantaneous speed, acceleration, the gap between leader-follower pair, and lane assignment. Finally, a discussion on the strength and weaknesses of WAMI data is presented followed by recommendations for overcoming these weaknesses.

## 3.2. BACKGROUND

The range of dataset traditionally been used in transportation planning and operation is out of the scope of this research. Rather, the focus of this study is on the dataset that closely resembles the WAMI data. This kind of image data can be seen as a conglomeration of two kinds of microscopic data collection effort; the instrumented vehicle method and the roadside video data collection method. Although the WAMI data can provide OD data for the area under coverage, the geographical extent required for planning purpose cannot be covered by current technology. However, this data can be a provider of OD of a specific location in an urban area. Estimation of OD and other macroscopic performance measure requires simple query in concurrent with a map matching algorithm. Details are discussed in section 3.7. In-depth literature review on the macroscopic measure, estimation is avoided in this study. Rather, the microscopic measures that can be derived from the data is focused in this research. In the following paragraphs, microscopic data collection methods using the instrumented vehicle, roadside video and their usages in modeling traffic characteristics are discussed.

Instrumented vehicles are usually equipped with positional sensing equipment like Differential GPS, LIDER or even camera which tracks the vehicle during the data collection period. These onboard sensor data used mainly for model development, calibration, and/or validation are mostly collected under controlled test conditions using a limited number of vehicles. Gurusinghe *et al.* collected GPS data from a platoon of ten vehicles on two parallel sections (Gurusinghe, Nakatsuji et al. 2002) with a sampling rate of 0.1 seconds, the authors determined that this high-resolution data can facilitate accurate car-following analysis. The same dataset was then used by Ranjitkar *et al.* to examine the stability of stimulus-response based car-following models with increased platoon size and found large intra and inter-vehicle variation in reaction time (Ranjitkar, Nakatsuji et al. 2003). Brockfeld *et al.* collected GPS track data from nine vehicles

on a test track in Japan for calibration and validation of 10 different car-following models (Brockfeld, Kühne et al. 2004). Ranjitkar *et al.* used a similar data collection method to measure the performance of six different car-following models (Brockfeld, Kühne et al. 2004, Ranjitkar, Nakatsuji et al. 2005). Because these data were collected on a test track, they may 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 and Simonelli 2005). This dataset was used for calibration and validation of four microscopic car-following models.

While the instrumented vehicle data have been used for calibration and validation of the microscopic models, the vehicle trajectory data obtained from video analysis has primarily been used for model development. Methods 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 example of terrestrial video data collection was done in the Next Generation Simulation (NGSIM) program. The 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 (FHWA 2008). The length of each roadway segment was between 500 and 640 meters and the data was collected for different durations and time of the day. A customized software application, NG-VIDEO, was used to track the position of every vehicle at a frequency of 0.1 seconds. Some analysis of NGSIM data has shown processing issues such as vehicle overlaps and unusual change in acceleration (Punzo, Borzacchiello et al. 2011). The NGSIM data have been used for development, calibration, and validation of many microscopic traffic flow models including the Target lane-changing model (Choudhury 2005), cooperative/forced freeway merging behavior (Choudhury, Ramanujam et al. 2009, Wan, Jin et al. 2014), arterial lane selection model for urban intersections (Choudhury and Ben-Akiva 2008), lane position and lane changing execution model for different vehicle classes (Moridpour, Sarvi et al. 2010, Aghabayk, Moridpour et al. 2011). This dataset has also been used in calibrating existing models (Monteil, Billot et al.

2014), as well as determining the effect of correlation between model parameters (Kim and Mahmassani 2011).

Previously, Ozaki 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 (Ozaki 1993). Ahn *et al.* used video data collected from a building near two signalized intersections in Oakland, California, for verification of Newell's simplified car-following theory (Ahn, Cassidy et al. 2004). 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 and Hoogendoorn used a similar approach to collect data from Freeway A2 in Utrecht, in the Netherlands to estimate the parameters for the stimulus-response based (GHR) car-following model and found that car-following behavior varies widely within the observed driver population (Ossen and Hoogendoorn 2005). The same dataset was used in calibration and parameter estimation of other car-following models, (Hoogendoorn and Hoogendoorn 2010, Hoogendoorn, Hoogendoorn et al. 2011) and for determining heterogeneity in car-following characteristics (Ossen and Hoogendoorn 2011). A similar approach was used for the development of the multi-anticipated car-following model by Hoogendroon *et al.* (Hoogendoorn, Ossen et al. 2006).

Onboard sensor data suffers from processing challenges and 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 (Ossen and Hoogendoorn 2008, Ossen and Hoogendoorn 2009, Punzo, Ciuffo et al. 2012). Although video and other remote sensing data are 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 questionable. 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.

A primary advantage of terrestrial and airborne video data collection is the ability to observe vehicles without influencing drivers. 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). Skycomp uses the same data collection technology used by PVLabs. Miovision uses similar data collection method to measure several macroscopic traffic performance matrices including traffic volume, speed, delay time, signal time split, signal phase occupancy ratio, and traffic monitoring in a construction zone. Data From Sky uses Traffic Drone to collect similar aerial imagery and extract similar traffic information as Miovision. TrafficVision uses static cameras to collect volume, movement type, incident monitoring, and construction management related data.

From the literature review, there is no one-size-fits-all approach for collection of microscopic data for the purpose of developing, calibrating and validating different driver behavior and other performance measurement models. There are opportunities for new data collection methods that can combine both the macroscopic and microscopic data. The WAMI data provides such an opportunity.

## 3.3. WAMI DATA OVERVIEW

Wide area motion imagery for the collection of airborne intelligence and reconnaissance was first developed as a tool for military intelligence. High-resolution imagery from aircraft in visible and infrared bands provides valuable detection capabilities based on target shapes and temperatures respectively. Different stabilization techniques and camera technology have been used for successful detection and tracking of stationary and moving objects. This kind of imaging spectrometers has the capability of yielding a large amount of rich information about the area under investigation (Puckrin, Turcotte et al. 2012). This

section provides a brief summary of the WAMI data collection system including the attributes of the data as processed and provided by PVLabs.

### 3.3.1. WAMI DATA COLLECTION

PVLabs is an advanced imaging solutions company specializing in the design and development of turnkey aerial imaging systems. The basic principle behind PVLabs wide area motion imagery data collection system is to detect moving objects from high-resolution images of an area captured continuously from an airborne platform, either fixed-wing or rotor. PVLabs uses Gen-V Technology for steering and stabilization of a camera fixed to a Gyrostabilized Gimbal Airborne Platform mounted on the aircraft. The WAMI tracking data used in this study was collected on 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 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 railroad and a shipping yard along Lake Ontario as shown in Figure 4. The collection occurred over two three-hour periods, one during the morning peak and one during the afternoon peak. The interval of image capture is 1 second. Figure 4 shows individual time-stamped vehicle points for a 15-minute interval during the evening peak period.

Once the video is collected, the PVLabs WAMI data processing system geospatially references the images, detect the vehicles, assign unique track ID and calculates other attribute listed in section 3.3.2. This image registration and the detection algorithms are proprietary to PVLabs and are beyond the scope of this study. It can be clearly seen from the unfiltered data in Figure 4, that there are a lot of data points that are unlikely to be an actual vehicle on the road. Some vehicle trajectory is completely outside of the roadway network and some have a very unrealistic path. Because of these limitations, special attention is required in evaluating data accuracy and data smoothing which is one of the major motivation of this study. These shortcomings are explored in section 3.5.

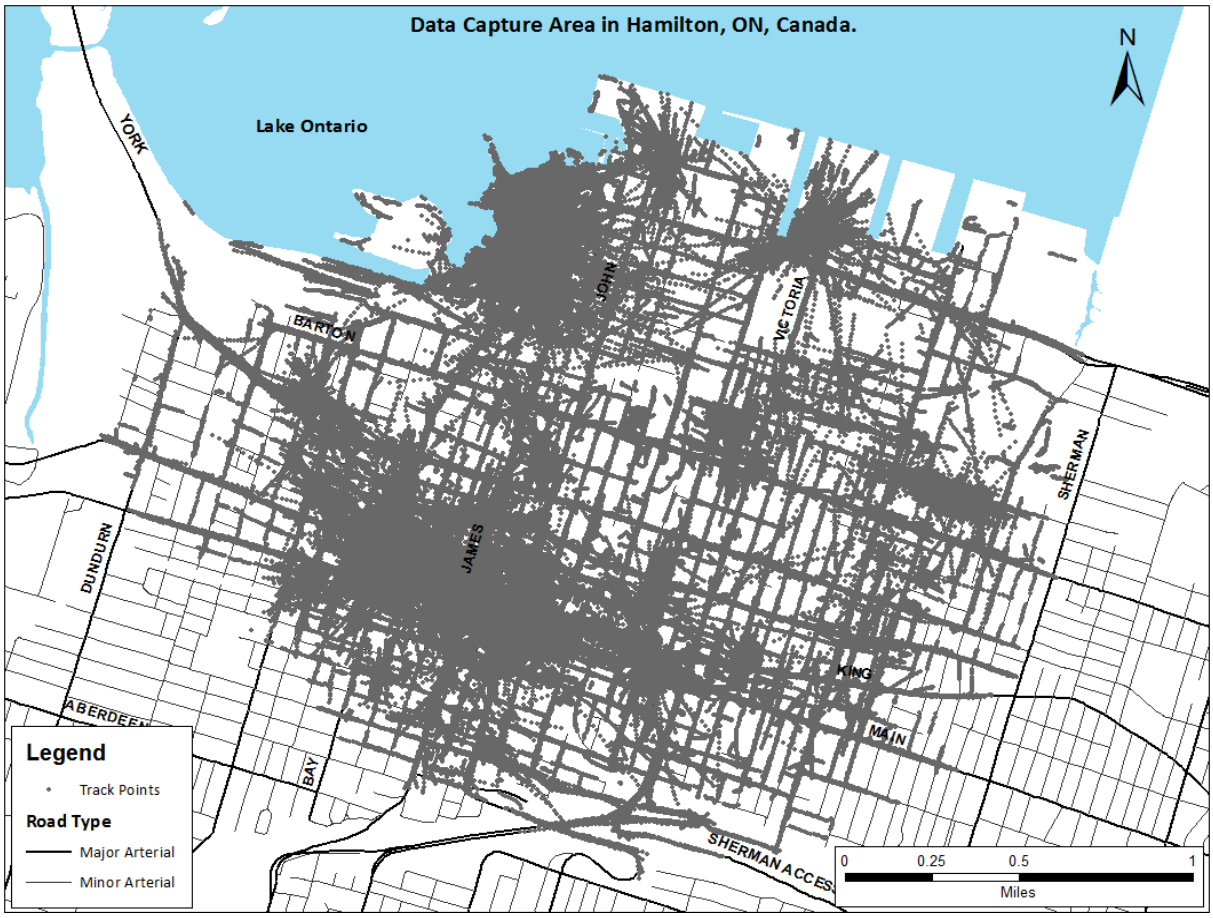



**Figure 4: Coverage Area and Tracking Points**

### 3.3.2. DATA ATTRIBUTES

The processed WAMI tracking data consists of one record per vehicle per second which resulted in 6.58 gigabytes for the morning peak and 6.92 gigabytes for the evening peaks. The attributes of the vehicle trajectory data extracted by PVLabs WAMI content management system is provided below. The name of each data attribute is reported as provided by PVLabs.

- **Frame No:** Unique identification number for every instance the image is taken
- **Frame Time:** Time of the image taken. This is also the time of detection of a vehicle at the particular location.
- **Track ID:** Unique identification number for every vehicle identified.
- **UUID:** Unique identification number for every vehicle in every frame.
- **Vehicle Size in Pixel:** Number of Pixel the vehicle occupies in the image.
- **Speed in Pixel:** Speed of the vehicle in pixel size. It is calculated by comparing the position of the vehicle at two consecutive frames.
- **Target Centroid Latitude:** Latitude of the centroid of the vehicle. The centroid is the geometric centroid of the pixels identified as a vehicle.

- **Target Centroid Longitude:** Longitude of the centroid of the vehicle. The centroid is the geometric centroid of the pixels identified as a vehicle.
- **Target Centroid Speed:** Speed in Kilometer per Hour of the centroid of the vehicle.
- **Target Centroid Heading:** Bearing of the vehicle's heading.

To extract both microscopic and macroscopic traffic information considerable effort is needed in data enrichment, refinement, and data mining. One drawback of this data is that it doesn't provide a classification of the vehicle. Any effort to classify vehicles using the pixel size of the vehicles has been unsuccessful due to high variation in pixel size even for individual tracks. Variations in pixel size occur due to changing vehicle or camera position at each image capture. The tracking point in the data is representing the centroid of the tracked vehicle.

## 3.4. TRAJECTORY DATA FILTERING

The term data filtering has been used with different meanings in the literature. Sometimes it has been used as a mean to separate good data from the bad and sometimes it has been used to describe data smoothing. According to Rayner (1971), filtering is just a general name given to the process of systematically modifying data which are arranged in a sequence or array. In case of the WAMI data presented in this study the term 'data filtering' is applied to the process of selecting falsely detected vehicle from the images and remove them from the dataset. In this kind of trajectory data, filtering can be based upon the whole or part of vehicle trajectory. For example, the individual trajectory can lie inside or outside (e.g. parking space, driveway) of the roadway network. The filtering technique must consider the objective of trajectory data analysis. If the study of driver behavior is the main objective of the study, then the portion of trajectory inside the network is of paramount importance. Then again, for the purpose of OD estimation, the whole vehicle trajectory becomes significant. In this study, a simple filtering technique is proposed which uses both spatiotemporal and trajectory characteristics of individual vehicles. The filtering technique is a constraint-based, multidimensional query driven process which is easy to understand and to apply.

### 3.4.1. TEMPORAL FILTERING

Temporal Filtering has been used in this thesis to refer to filtering based on the temporal aspect of the data. Vehicle tracks of different duration are shown in Figure 5. From this figure along with Figure 4, it is clear that some tracks in the data do not represent any vehicle. Some vehicle tracks can be seen on top of buildings, on parks or even on the Lake Ontario. This false detection might be caused by moving pixels in the images caused by registration error or other moving objects like flying birds or even pedestrian. When segmented by track duration a trend between false detection of vehicles and their duration in the view frame can be seen. Most of the false detection is concentrated within track duration less than 30 seconds. There is much less noise in track duration between 30-60 seconds. However, numerous tracks can be seen with duration more than one minute, which are circular in shape. These tracks emerge from stabilization and image registration error. Stabilization and image registration error causes adjacent pixels to move around from frame to frame thereby causing false detection.

To filter data based on their trajectory duration, different time span was chosen to observe their spatial relation with the underlying network, as shown in Figure 5. It was found that most data noises were associated with trajectories less than 30 seconds. To develop filtering algorithm, track duration is converted to an ordinal attribute. The ordinal attributes are trajectory ranking based on their duration. Trajectory duration of 1-10 second is assigned a ranking of 1, 11-30 second is assigned 2, 31-60 second is assigned 3 and others are ranked 4. The similar ranking is done for spatial attributes and sinuosity attribute. The trajectories with lower rank have a higher probability of false detection. Using this method spatial and temporal aspect of the trajectory is converted to simple attribute which can be used to mine the data. Three attributes are combined into a single attribute by means of the weighted sum. Final data mining is done based on this combined attribute. A histogram of trajectory duration is shown in The reduction of trajectory count is clearly visible throughout the range of duration but the most of the reduction is in the short trajectory duration. This outcome validate the use of a temporal aspect of the trajectory data as trajectory with lower duration has a lower rank and highly likely to be a false detection.

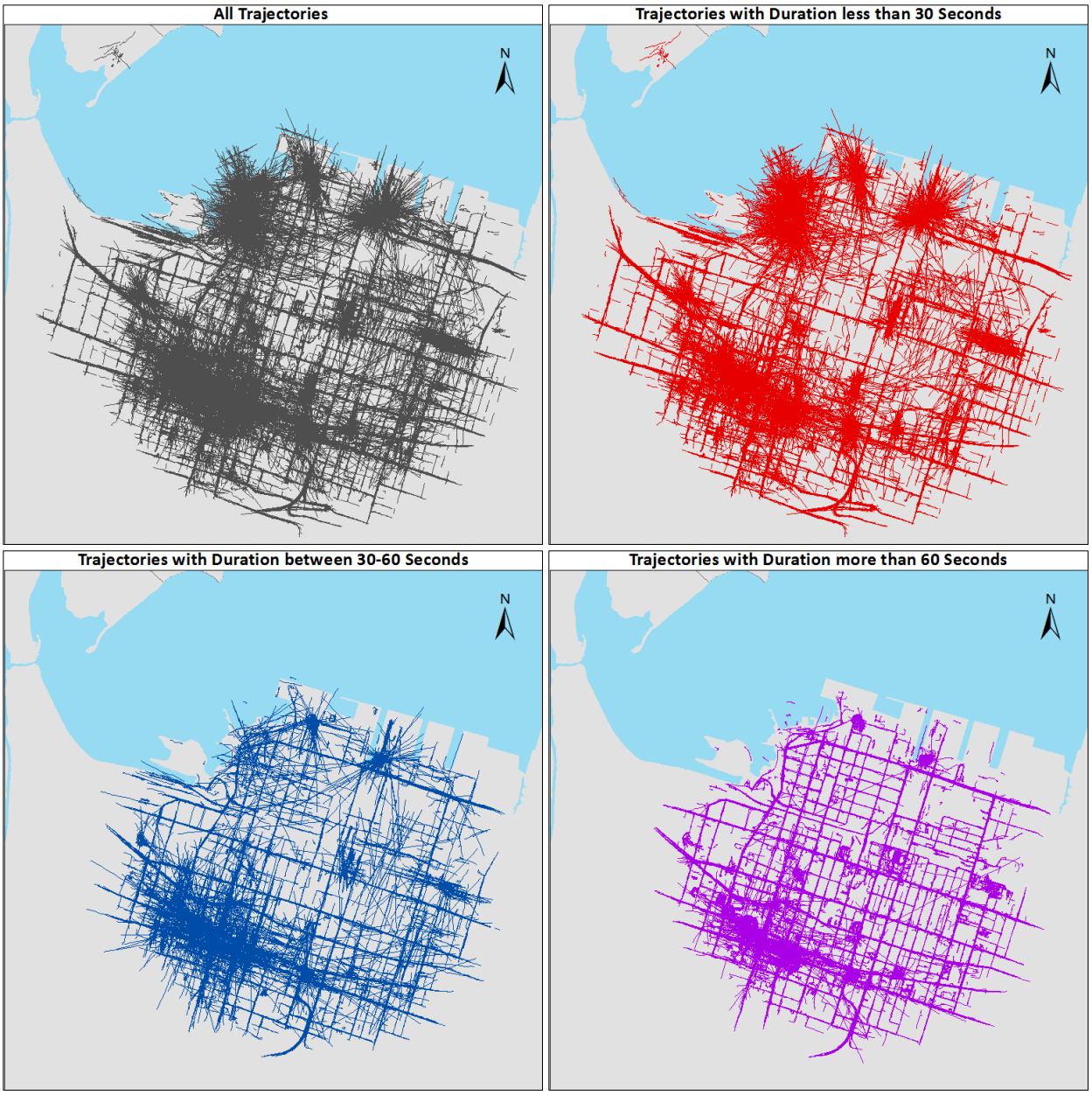



**Figure 5: Trajectory with Different Duration**

### 3.4.2. SPATIAL FILTERING

Spatial filtering allows query based upon the position of the trajectory point. This method is highly dependent on the positional accuracy of the trajectory point. The positional accuracy of tracked vehicles is reliant on the image registration system and errors in either the registration or the detection algorithm will result in errors in the trajectory and resulting microscopic parameters (Ossen and Hoogendoorn 2008). Spatial filtering is built upon a detail polygon layer of the roadway network. The polygon layer contains

the whole length and breadth of the network under investigation. The location query is conducted on every tracking point to locate whether they lie within or outside the polygon layer. If 25% of tracking point of each trajectory is within the network that trajectory is assigned a rank of 1, 26-50% is assigned ranking 2, 51-75% is assigned ranking 3 and the rest is assigned a ranking of 4. Similar to temporal ranking, lower spatial ranking also represents a higher likelihood of false detection.

### 3.4.3. FILTERING BASED ON TRACK SINUOSITY

Sinuosity of a curve having at least one inflection point is the ratio of the curvilinear length and the Euclidean distance between the endpoints of the curve. The value of this dimensionless property of the curve ranges from 1 (e.g. straight line) to infinity (e.g. a closed loop, where the shortest path length is zero) or for an infinitely-long actual path. Although the arcs under consideration are piecewise linear, the sinuosity index provides a good indication of the actual shape of the curve. This property of the trajectory is selected as an attribute to filter the data due to the fact that most of the long duration ambiguous trajectory is concentrated near high rise building and the Lake Ontario. These trajectories are circular in nature. High sinuosity value is an indicator of the stationary object picked up by the image processing algorithm as vehicle trajectory. The trajectories are ranked from 1-4 based on their sinuosity value similar to temporal and spatial filtering attribute. The Euclidean distance between the beginning and end point of the trajectory is calculated by extracting the first and last point of every trajectory and calculating the straight line distance between them. The curve length is calculated by summing the straight line distance between successive tracking points of the trajectory. Trajectory with sinuosity value of 1-1.25 is assigned rank 4, 1.26-1.5 is assigned rank 3, 1.51-2.0 is assigned rank 3 and others are assigned rank 1. Similar to previous filtering attributes lower ranks have a high likelihood of false detection.

### 3.4.4. ATTRIBUTE WEIGHTING

Attribute weighting is a critical part of combining the three different filtering attribute into one attribute which can be used to filter the data using simple query. Weighting can be absolute or relative. In this study relative weight has been used. Higher weight has been assigned to spatial and sinuosity attributes

(40% each) and lower weight is assigned to temporal attribute (20%). The temporal attribute is given less weight as the trajectory of the same vehicle can be spliced into multiple trajectories as its trace goes missing in an urban canyon. The combined weight is calculated using equation 15.

$$A_t = w_{temp} * R_{temp} + w_{spt} * R_{spt} + w_{sin} * R_{sin} \quad (15)$$

Where,

$w_{temp}$ = Weight of temporal attribute
$R_{temp}$ = Ranking of temporal attribute
$w_{spt}$ = Weight of spatial attribute
$R_{spt}$ = Ranking of spatial attribute
$w_{sin}$ = Weight of sinuosity attribute
$R_{sin}$ = Ranking of sinuosity attribute

The combined attributed rank of any trajectory can also vary from 1-4 as the individual ranks. A higher value of rank represents a high likelihood of positively identified tracks. The cut off value of the combined attributes is selected by means of data validation. This process is described in section 3.4.5. Validation of this filtering algorithm is conducted on by means of visually matching the filtered data with the actual images taken by PV Labs from the airborne platform.

### 3.4.5. VALIDATION OF DATA FILTERING

Validation of filtered data is necessary to ensure the efficiency of the filtering algorithm and accuracy of the filtered data. Removal of actual trajectory from the dataset will significantly impact the estimation of macroscopic and microscopic data extracted from the tracking data. One method of validation of filtering algorithm can be by means of matching the vehicles in the images taken by PV Labs and the vehicles in the filtered data. This method can be used to validate both the PV Labs computer vision algorithm as well filtering algorithm proposed in this study. Each image taken by the PV Labs is off the size of 200 MB in JPEG format. For each three-hour peak period, the total number of images taken is 10,800. This high number of images with their large size makes the total size of the image data collected more than 2 TB. Manually matching the images with the filtered data is almost impossible and somewhat

unnecessary. In this study, validation effort conducted on the two main arterials of Hamilton namely King Street and Main Street is presented. Many high rise buildings are located along these roads and as seen in Figure 4, a lot of data noise exists along these routes. Validation of filtering is done on the first ten minutes of afternoon peak period. Ten minutes of data consist of 600 frames. Vehicles on the two arterials are counted at every 5-seconds interval. A total of 120 data point is collected by means visually counting the vehicles in the images. The manual count is matched with the vehicle count for different combined parameter cutoff value. The plot of vehicle count from the trajectory data against the vehicle count from the images is shown Figure 6. For ideal condition, the scatter plot in Appendix A would lie close to the line with a slope of 1 as drawn in the figure. At lower combined parameter most of the data points lie over the line indicating a higher number of false detection. The point gets shifted lower and lower with the increased cutoff value of the combined parameter. Beyond the combined cutoff value of 2.2, most of the point lies below the line indicating a shift from removing false detection to the removal of actual vehicle trajectory. A cutoff value of 2.2 is used for the farther analysis presented in this study.

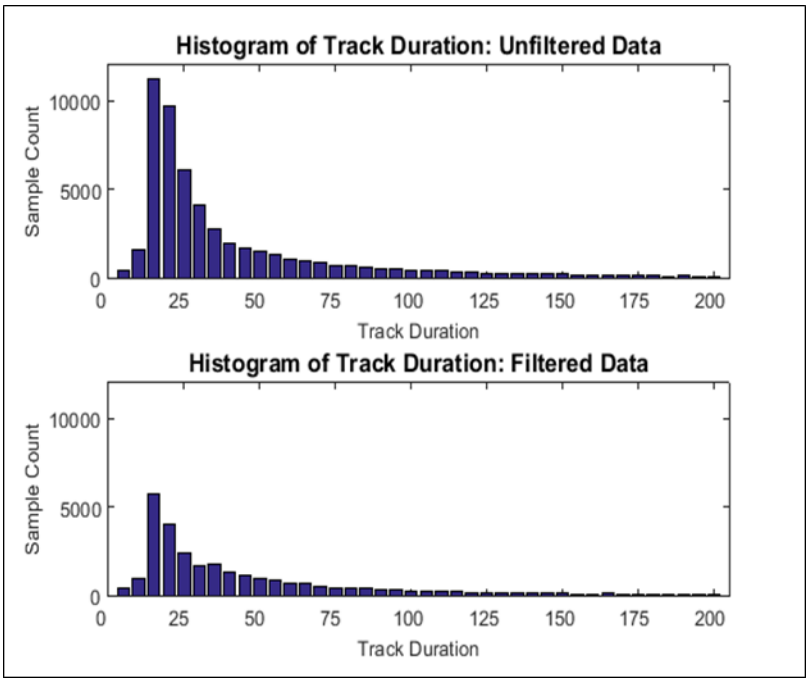



**Figure 6: Histogram of Track Duration**

### 3.5. MAP MATCHING

The purpose of map matching is to match the tracked point to correct link on the network. The process is straightforward for roadway section with relatively long lengths. However, it becomes more problematic with decreasing sampling rate and link length. The high sampling rate of the WAMI data reduces the risk of missing vehicle trace over a link. Nonetheless, intersection density is more important for this dataset as the data is collected in a dense urban network. A strong argument can be made for this kind of big dataset, whether to filter the data first and then do the map matching or do the filtering after map matching. Filtering the data first would reduce the number of records and reduce the processing time required for map matching algorithm. However, map matching the data first would give an additional attribute on which the

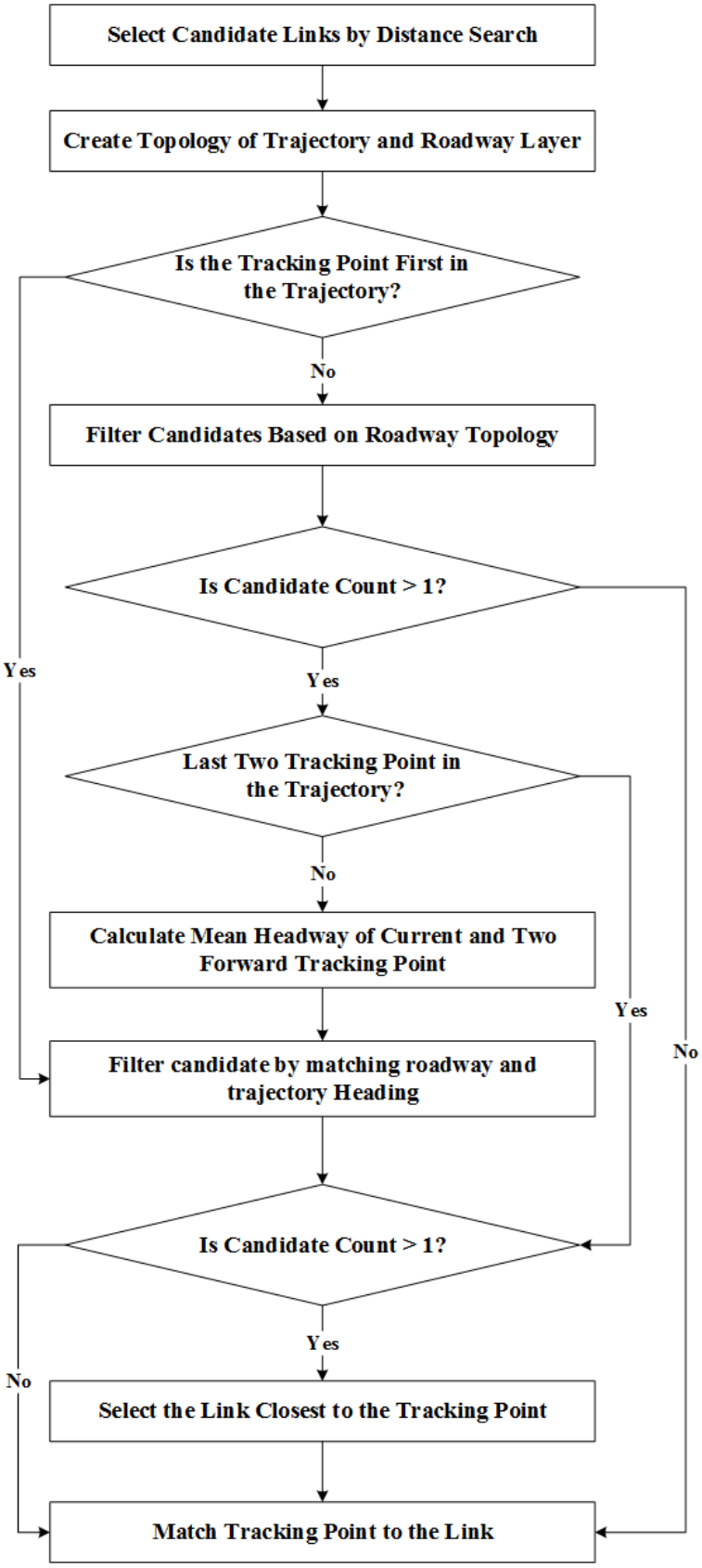



**Figure 7: Map Matching Algorithm**

filtering algorithm can be improved upon.

### 3.5.1. MAP MATCHING ALGORITHM USED

Geography and topology of the underlying network on which the tracking points are being matched to, are important inputs of map matching. A roadway network consists of a set of curves, *A*. Each of them is called an arc. Each arc consists of segments which are piecewise linear and can be defined by a set of points. The beginning and the end point of the arc are called nodes and the interim points are called vertices or shape points. A vehicle trajectory also consists of a set of point, $P^t$. In map matching algorithm, each point of the trajectory is assigned to an arc of the roadway segment. Each data point can be associated with a set of candidate roadway network, $C^t$ which is a subset of *A*. For simplicity, a one-to-one correspondence between tracking point and the matched roadway arc is assumed.

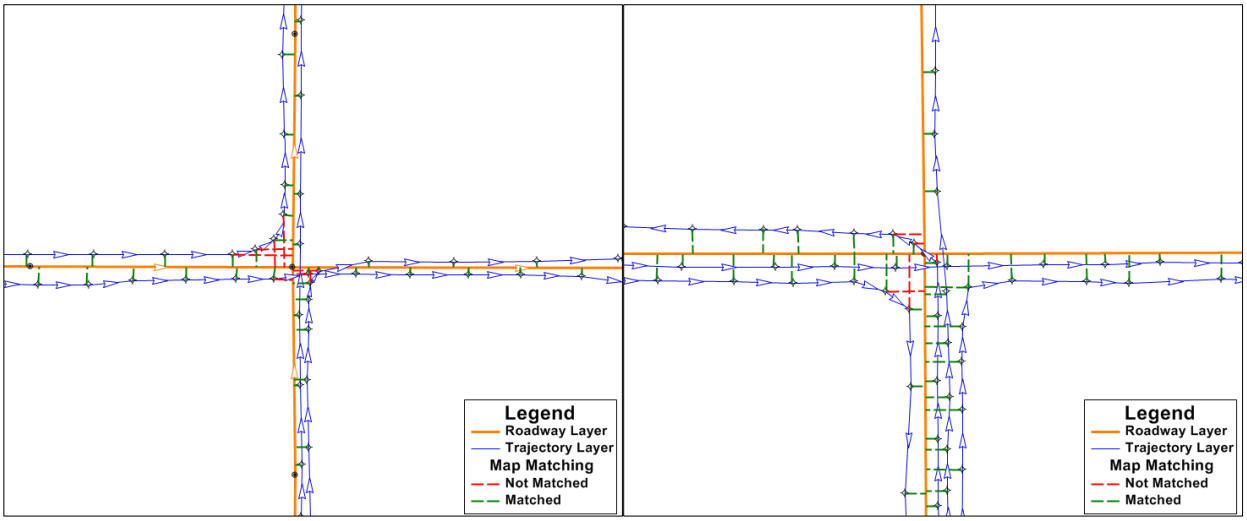



**Figure 8: Illustration of Map Matching Algorithm**

Several approaches exist in the literature about map matching (White, Bernstein et al. 2000, Greenfeld 2002, El Najjar and Bonnifait 2005, Quddus 2006, Quddus, Ochieng et al. 2007, Velaga, Quddus et al. 2009, Miwa, Kiuchi et al. 2012). In this study, both geometric and topological analysis to match the point to map is used. Geometric map matching is a simple search algorithm. Traditionally, two distinct methods have been used for geometric map matching; the point-to-point match and the point-to-curve match. In point-to-point matching approach, each of the track points is matched to the closest 'node' or

'shape point' of a road segment way. This method is very fast to implement but it is very sensitive to the way the underlying layer is digitized. In the point-to-curve matches the point at matched to an arc that is the closest. This method is better than a point-to-point match but suffers from error in dense urban network. In algorithm used in this study, the point-to-curve method to select a set of candidate arcs for each trajectory point is used. An arc is selected as a candidate if it is within a maximum search distant from the trajectory point. This set of candidate the forms a primary set, $C^{t1}$.

Topology, as defined in geography, is the relationship between adjacent objects. In case of roadway line layer, the topology is the connectivity between arcs. In the algorithm used, topologies of the roadway network and the heading of the links and trajectories is used. The heading of the links and trajectories are calculated from true north. The roadway network used in map matching is geographically matched with the direction of traffic and the topology at the intersection is built based on allowed turn movement. This allows reduction of the number of candidate links if the vehicle trajectory is in the transition from one link to another.

At each timestamp or point, only those links are selected as the new candidate set, $C^{t2}$, which can be reached from the previously matched link. If the point is the first of last of a unique vehicle trajectory, then this step is skipped and it is matched with the candidate link who's heading closely matches with the heading of the vehicle trajectory. After topological matching, track and link headway matching is conducted to select only one link from the new candidate set. The link whose heading is closest to the heading of the point is selected as the matched link. The algorithm is shown in Figure 7. Matching result is shown in Figure 8 graphically.

### 3.5.2. PERFORMANCE OF THE ALGORITHM

Performance of the map matching algorithm is calculated after the data has been filtered. The filtering technique is discussed in the previous section. The accuracy of the algorithm is estimated of for two main arterials, King, and Main Street. The validation is done by comparing the map matched point with

the images. The accuracy is evaluated using the first fifteen minutes of afternoon peak. The accuracy of the algorithm was tested for 1,000 vehicle trajectories by visually matching the vehicle location on the roads with the images. For the two one-way arterial under consideration 97.9% trajectory point was found to be positively matched. Although the accuracy of image registration and thereby positional accuracy is not assessed, which have a high impact on microscopic measurements, the map matching algorithm works well for this particular type of data.

## 3.6. DATA MODEL OF STORING TRACK 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 are 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 (Brakatsoulas, Pfoser et al. 2004). One important property of vehicle trajectories is that vehicles cannot be at any discrete position over space. Their movements are bounded by their surrounding environment such as roads, buildings, obstructions, etc. (Tiakas, Papadopoulos et al. 2009). Another important property is that two vehicles cannot occupy the same space at the same 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 (Brakatsoulas, Pfoser et al. 2004). 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 to the author's knowledge, no currently available commercial data management products include a structure that effectively manages this type of data. Although ESRI's GIS software, ArcGIS©, can display and perform limited analysis on trajectory data, it operates on individual points that define the trajectory, not the trajectory itself, and the processing time becomes impractical for the database size of this magnitude.

Developing trajectories include: (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 (Brakatsoulas, Pfoser et al. 2004). When defining the data model, additional considerations may include establishing the relationship between a trajectory and its environment, and interactions between individual trajectories. Examples of the former include determining the position of the vehicle on the road such as distance from upstream link node, distance to downstream link 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 schemas for preprocessing and storing large trajectory databases: the 3-D schema and the network schema. In the 3-D schema, only the vehicle position is recorded at a discrete interval of time. 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. PVLabs 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 uses a map-matching algorithm. In this method, each trajectory is stored as a 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 (Brakatsoulas, Pfoser et al. 2004). 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. This schema also needs significantly smaller data storing space making it more computationally efficient. Details about these data schemas can be found in the original paper (Brakatsoulas, Pfoser et al. 2004).

To make a better use of geo-processing and visualization capabilities of any GIS software, the data need to reside in a single database which is not achievable by the network schema. Another shortcoming of the network schema is the mismatch in geographical resolution between the tracking layer and the roadway layer. Brakatsoulas *et al.* used GIS tracking data which was collected at a sample rate of 30 seconds. During

a 30 second, 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.* requires modification. A modified Brakatsoulas *et al.* data model is shown in Figure 9. At each observation, a unique ID for individual tracking point as well as individual vehicle tracks is preserved.

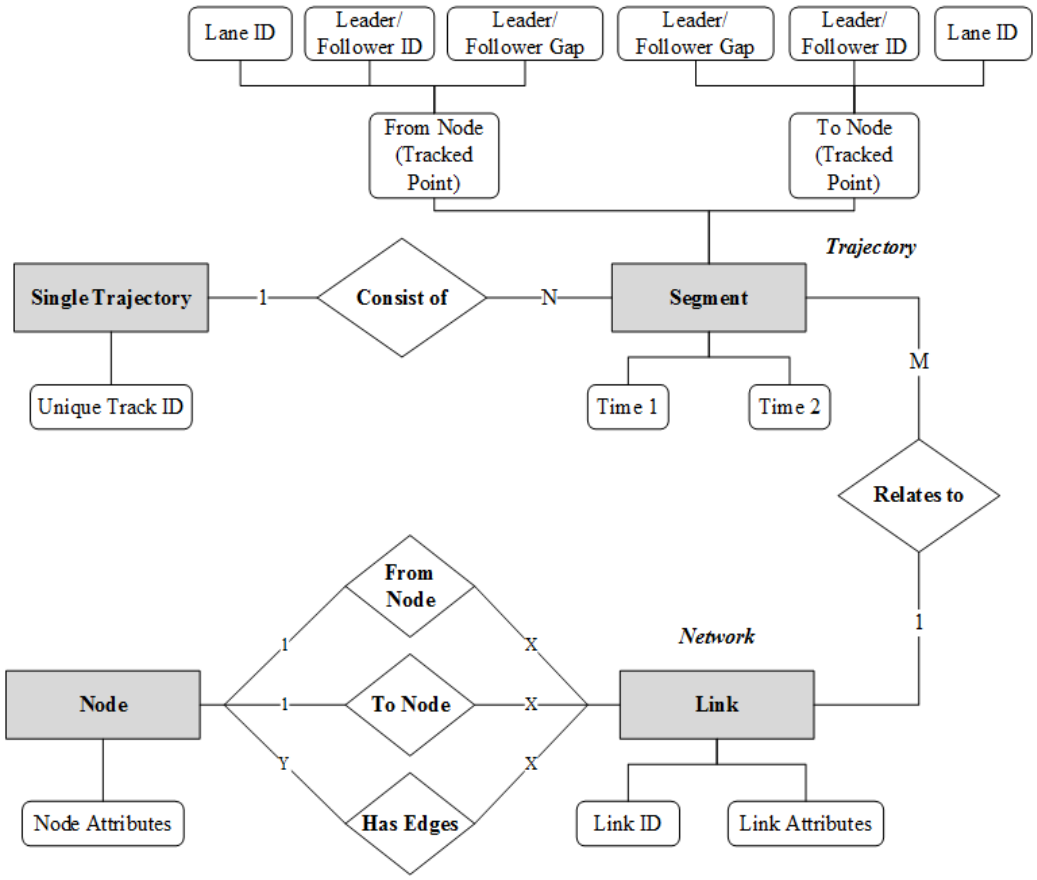



**Figure 9: Data Schema for Storing tracking data taken from Brakatsoulas *et al.***

## 3.7. TRAFFIC DATA EXTRACTION

ArcGIS version 10.2.2 has been used to calculate flow, density, headway, and lane position of the vehicle. ArcGIS has limited capability in storing and analyzing of large datasets. To overcome this obstacle and to use the geo-processing capability of ArcGIS, its Python extension has also been used. The use of this extension is limited to headway calculation only in which distance headway of vehicle is calculated for

tracking point. The geographic projection system of the data provided by PVLabs is WGS84, which is a latitude-longitude based coordinate system. Any calculation of distance based on this projection system is prone to error in measurement. To somewhat reduce this error, the data is projected to a planer projection system. The planer projection system used in this study is 'NAD_1983_UTM_ZONE_17N', which is the Transverse Mercator projection system for the region.

### 3.7.1. MACROSCOPIC DATA

Link-wise flow, density and travel time/speed are calculated based on the same link-node relationship. Matching the tracking data with roadway network is done by overlaying the tracking layer with the roadway line layer. ArcGIS tracking analyst tool is used to convert time series of position (point) data to a line layer. This method calculates the distance from the difference between successively ordered features in a track. Length of each link is the instantaneous speed of vehicle. Flow and density of each link is calculated by aggregating individual tracks for each fifteen-minute time interval. Two different methods are used to calculate average speed/travel time on the network. In one method, average speed/travel time is calculated for each individual link of the road. Average speed is also calculated for every individual track. The data extracted for the tracking data is as the following:

#### 3.7.1.1. OD OF EACH TRAJECTORY ON THE VIEW AREA

Origin and destination of individual vehicle can be extracted from the WAMI data. The extracted origin and destination for individual vehicle during peak period from 3:00 PM to 3:15 PM is shown in Figure 10. First point of appearance of every track is identified as its origin and the last one is identified as the destination. Most of the data OD points in the surveyed area are located in parking lots. This kind of OD estimation becomes problematic of urban canyons. If a vehicle disappears behind a large building, its trajectory gets terminated and when it reemerges it become a new trajectory. As a result, estimation of actual number of trips made becomes uncertain. This effect can clearly be seen in Figure 10. There is a cluster of OD point near the center of each image. Most of the tallest buildings in Hamilton are located in this particular area. As vehicle disappears behind the tall buildings in this area, trajectory get terminated

and high number of destination point is located near this area. The high density of origin can be explained by reemergence of vehicle as a part of completely new trajectory.

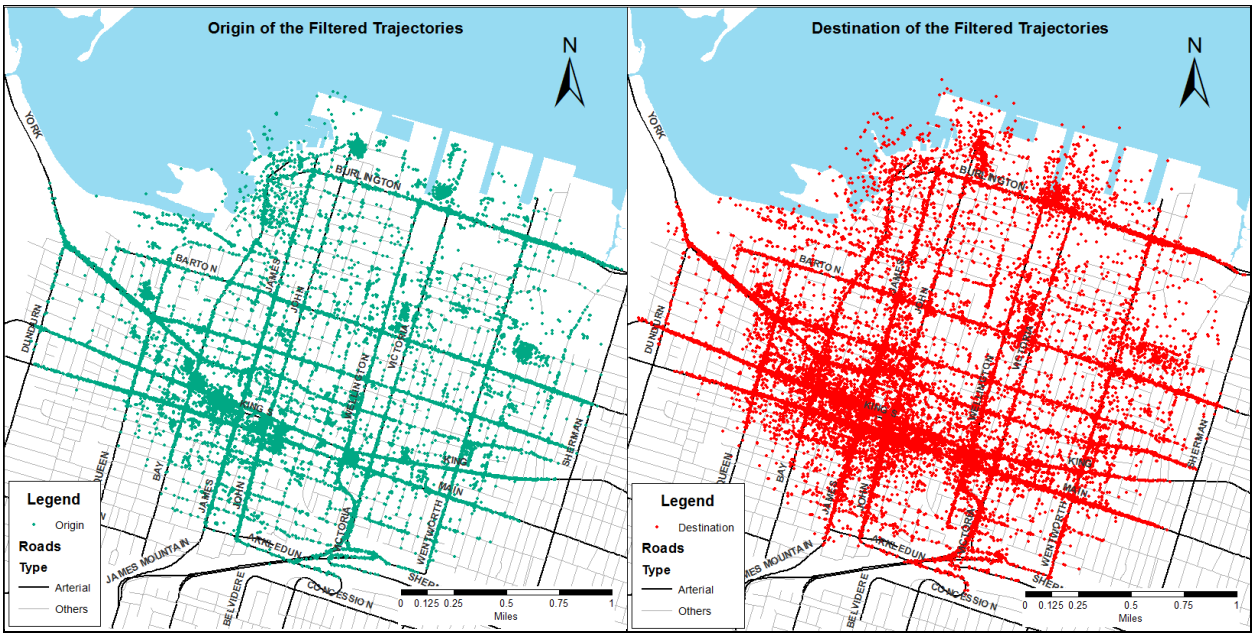



**Figure 10: Origin and Destination of Individual Tracks**

### 3.7.1.2. AVERAGE TRAVEL TIME/SPEED ON EACH LINK

Speed and travel time of roadway segment can only be estimated after map matching the trajectories with the underlying roadway line layer. The mean speed that is easily estimated from the trajectory data is the arithmetic mean instantaneous speed over a roadway section. This mean speed does not fit the traditional meaning of time mean speed or space means speed but as this average instantaneous speed gives an average speed of over the roadway segment, it closely resembles space mean speed. The average mean instantaneous speed is calculated after the trajectory is matched to a roadway segment. The instantaneous speed at all the trajectory point in the segment is averaged to estimate the mean speed of that particular link. The estimated mean speed for only the arterial roadway network in the surveyed area is shown in Figure 11 along with its standard deviation

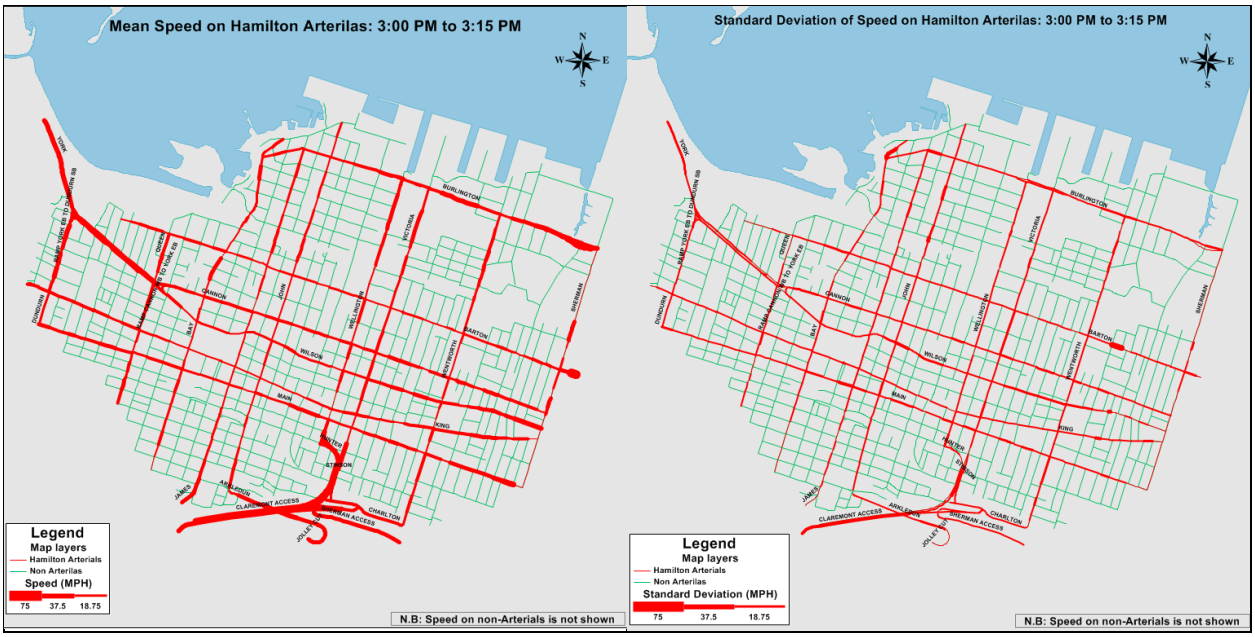



**Figure 11: Mean Speed and their Standard Deviation**

### 3.7.1.3. TRAFFIC FLOW ON NETWORK

Similar to speed and travel time estimation, traffic flow also can only be estimated after map matching the trajectories with the network. Flow is estimated by counting individual trajectories matched to a particular link. Counting whole trajectory instead of every track point reduced the chance of double counting. Traffic flow on the arterial network of Hamilton is shown in Figure 12.

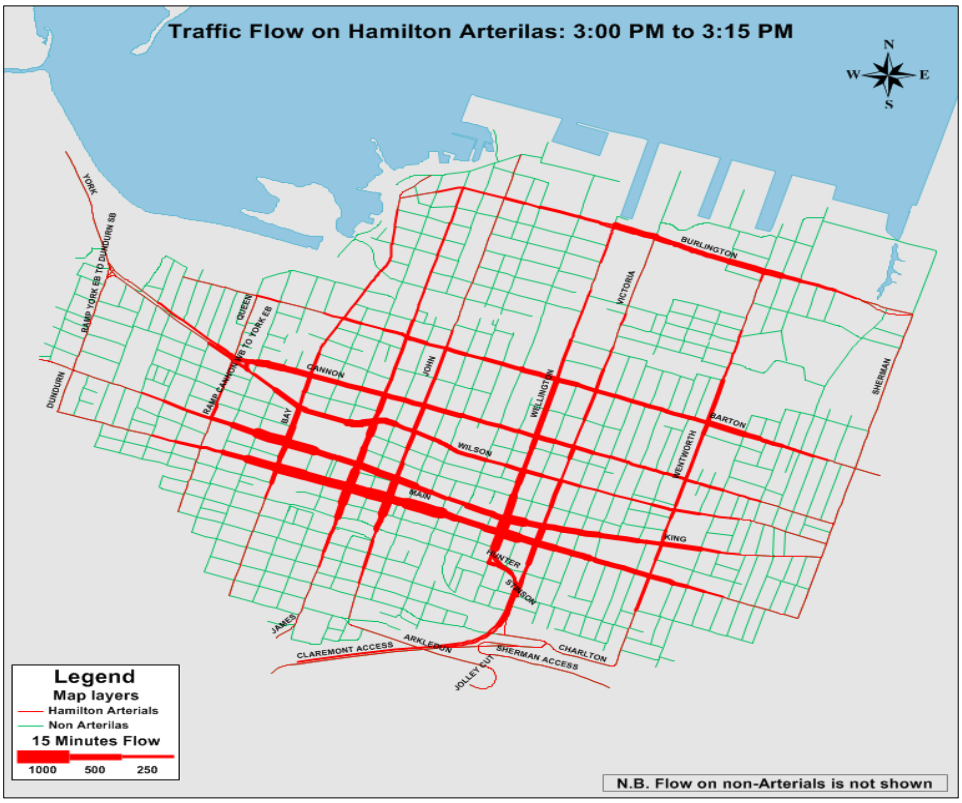



**Figure 12: Flow in Arterials of Hamilton**

#### 3.7.1.4. OTHER MACROSCOPIC MEASURES

In addition to OD, speed, and flow, macroscopic traffic measure that can be extracted from the WAMI data are vehicle density on roadway segment, lane-wise distribution, turning movement, route choice and so on. However, the focus of this research is on the modeling of microscopic car-following behaviour which is the focus of the next section.

### 3.7.2. MICROSCOPIC DATA

The microscopic traffic measures that can be mined from the WAMI data are instantaneous speed, instantaneous acceleration, headway and lane changing the behavior of the driver. The length of the vehicle is required to estimate the bumper-bumper gap between leader-follower pair. As vehicles cannot be classified from the WAMI data a value of 6 meters is deducted from the estimated pixel center to center distance to estimate the gap between vehicles. The gap between every vehicle and three other vehicles on the same lane in front and back is calculated to study platoon behavior. The headway calculation is terminated beyond the 500-meter range both in front and back of every vehicle. The microscopic traffic measures are stored in data schema shown in Figure 9. Examples of some microscopic traffic measures are shown in Figure 13. For simplicity, a single pair of leader-follower is selected and their speed, acceleration, and the gap is shown as time series in Figure 13. These microscopic traffic measures are suitable for calibration and validation of microscopic traffic flow models like car-following model and lane changing model. However, calibration of gap acceptance model is problematic as higher sampling rate (0.1 sec) is preferable for this model. Methodology and result of calibration and validation result are discussed in a Chapter 4.

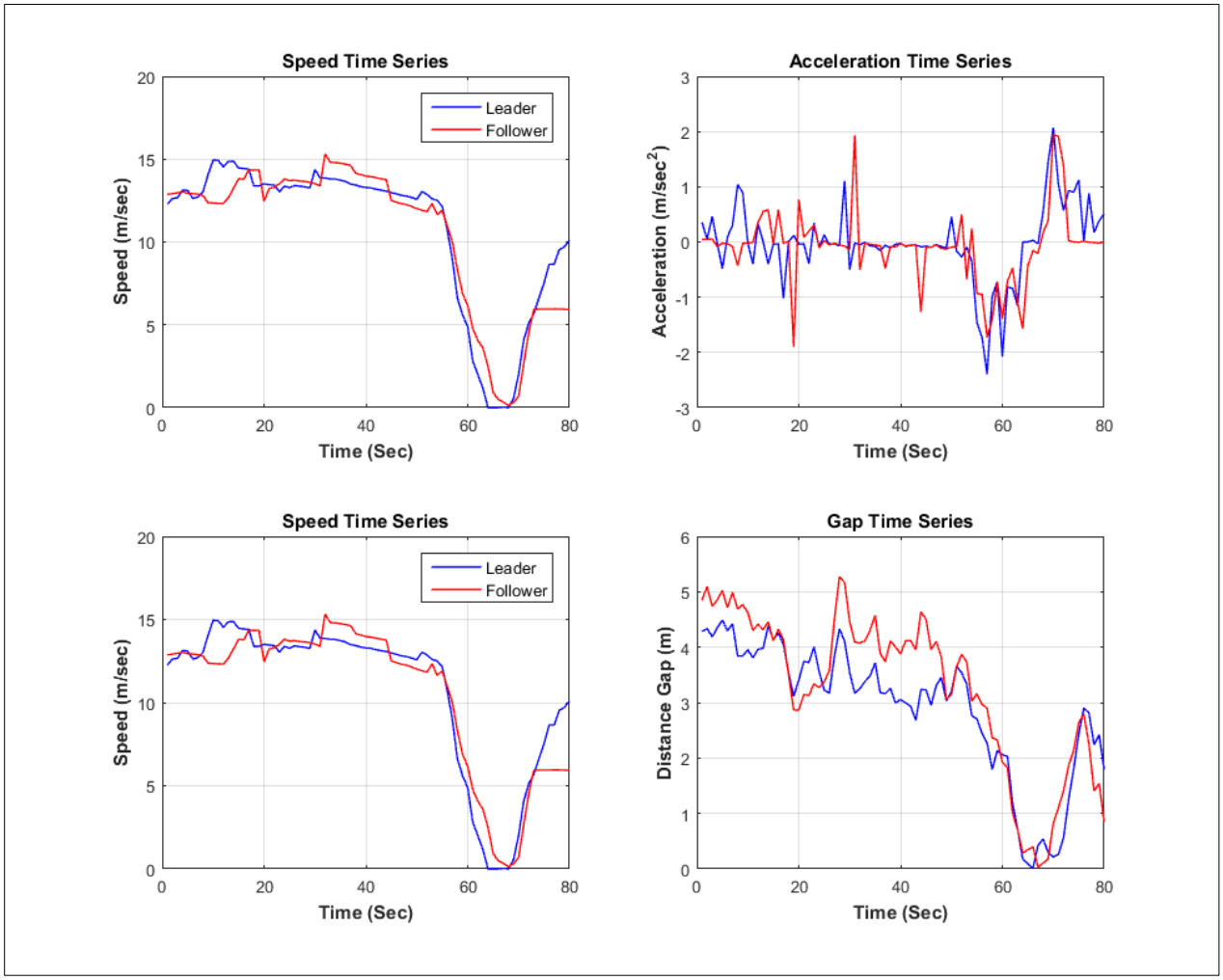



**Figure 13: Microscopic Traffic Flow Measures**

## 3.8. SUMMARY

In this chapter, the potential of vehicle trajectory data obtained from Wide Area Motion Imagery for obtaining microscopic and macroscopic traffic measurements is explored. The WAMI data provides a new way of a collection of traffic data. It is a move away from traditional data collection effort which is mostly bound by temporal or spatial or both restrictions. Although still constrained by limited data collection area, WAMI data provides the opportunity of collecting both microscopic and macroscopic data for a long duration. Nevertheless, the WAMI data suffers from noises caused by both false identifications of the vehicle and measurement error. Handling of this kind of large spatiotemporal data also possesses serious challenges. Some of the issues are discussed in this study along with some solution technique.

One of the most important issues is filtering the data to remove false detections. A limited amount of available information makes filtering difficult. Three different aspects of the data, spatial location related to roadway network, track duration and track sinuosity, are used to develop a filtering algorithm. The algorithm is better suited for filtering the whole trajectory rather than an individual point in the trajectory. Validation of the filtering algorithm reveals ample opportunity for development of new filtering algorithms as the results obtained from this research are limited.

Another important requirement of the WAMI data to be useful in identifying traffic characteristics is matching the trajectory to the underlying roadway network. Current methodologies of map matching with some modification are capable of handling this problem with reasonable success. However, to use the algorithm presented in this study, the underlying road network needs geographical and topological enrichment. Storing of this kind of trajectory data and subsequent derived macroscopic and microscopic traffic data was identified as one of the critical issues. A modification of existing data storing model is proposed in this research that is capable of addressing simplified query about the whole trajectory as well as individual trajectory point.

Derivation of microscopic and macroscopic traffic measures revealed additional challenges associated with this data such as trajectory splitting in urban canyons and measurement error in vehicle position and subsequent estimation of speed and acceleration. However, this error can be reduced by data smoothing and other techniques that are explored in the literature. This makes false and no identification of vehicles the most important aspect of future development. The long duration of the data makes it suitable for observing behavioral changes of drivers. The findings of this research can be used for refinement of computer vision algorithm used for trajectory extraction from images.

# 4. PARAMETER ESTIMATION OF CAR-FOLLOWING MODELS USING VEHICLE TRAJECTORY DATA FROM WIDE-AREA MOTION IMAGERY

## 4.1. INTRODUCTION

The mathematical formulation describing the behavior of vehicles following each other in the traffic stream is commonly known as car-following model. Car-following models represent a molecular approach of modeling driver behavior, which models how a driver follow the preceding vehicle in a traffic stream (Brackstone and McDonald 1999, Rakha, Pecker et al. 2007). The purpose of these models was to assess the effect of any fluctuation of driving condition on traffic flow. Currently, these models are the basis of different microsimulation models, which have wide-ranging application in network design, alternatives evaluation, planning, intelligent transportation systems analysis, and assessment of autonomous and connected vehicle's impact on traffic flow.

The first generation of these models was developed in the 1950s and 1960s. They were validated and refined based on varying traffic measurements collected in the field (Rahman 2013). These models were built upon some fundamental relationship between different driving stimulus, vehicle dynamics, and driver response. A later generation of car-following models tried to incorporate some complex parameter related to driver's decision making and reaction process. Driver's decision making factors like route choice, response to information from V2V/V2I or any other communication are being incorporated into the latest generation of models. As a result, use of these models has moved on from traffic analysis to dynamic vehicle control and safety system as well as autonomous cruise control. However, depending on the structure of the model, multiple parameters requires estimation and calibration.

Rigorous calibration of car-following and other models used in the microscopic simulation is required for obtaining reliable results for local traffic behavior. Calibration of the car-following model has relatively higher importance than others, as key traffic flow parameter like speed, flow, and density is highly related to the car-following model. 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). The models calibrated in this study are the GHR model, the linear Helly model (variation of CA model), and the Intelligent Driver model.

The procedure used in calibrating car-following models is widely varying in the literature. Initial efforts on the calibration of the car-following model were concentrated on converting the microscopic model into macroscopic traffic stream model and calibrating the traffic stream model with macroscopic data like flow, density and speed (Punzo and Tripodi 2007, Rakha, Pecker et al. 2007). One drawback of this method is that the traffic stream model usually represents a steady state condition of the roadway. Steady state condition is defined as a state where acceleration and deceleration of vehicles are in a range of ±0.05g or simply a short time or distance where traffic state remain constant (Rakha and Crowther 2002). However, even within this steady state definition, traffic can be in any of the three following states; emergency, car-following, and free flow depending on the relative time headway between the vehicle and its leader. In these three states, the car-following model only applies to the region when drivers accelerate/decelerate to maintain his/her desirable gap with the leading vehicle. Using macroscopic data with traffic stream model to calibrate car-following model can skew the parameter estimation to traffic where car-following regime may not apply. So, there is ample opportunity to calibrate the car-following model for suitable driving conditions for different roadway characteristics.

The recent development of sensing, image processing, and data capture technology has enabled researchers to develop high fidelity vehicle trajectory data. High fidelity trajectory data like the Next Generation Simulation (NGSIM) data has been used to calibrate car-following model (Treiber and Kesting 2013, Rahman, Chowdhury et al. 2015). However, the NGSIM and other high fidelity data are restricted either spatially or temporally or both. Although the WAMI is of lower fidelity (1Hz) than the NGSIM data (10Hz) it provides more spatial and temporal coverage. Interpolation between two consecutive time-stamped locations can be used to obtain vehicle trajectory information at any intermediate time. This way

the low fidelity of the WAMI data can be addressed in case of an application for calibration of the car-following model.

The framework of parameter estimation has received attention from the research community. Most of the studies in the literature focus on a deterministic framework of calibrating the car-following model. The deterministic approach estimates a near optimal solution of parameters to minimize the difference between the observed and simulated data. However, recent studies have revealed a significant difference in behavior among individual drivers as well as the same driver for different conditions (Kesting and Treiber 2008). A distribution of parameters may be a better representation of the driver behavior resulting in more reliable analysis output (Kesting and Treiber 2008, Rahman, Chowdhury et al. 2015). In this chapter, calibration results conducted in both deterministic and stochastic frameworks is presented. However, the stochastic method used in this study in conducted to obtain parameter in a maximum likelihood framework.

The rest of the chapter is organized into six sections. In section 4.2 a brief literature review is provided on the different type of traffic data collection method and their uses in calibration and validation of the car-following model. Details about the WAMI data is presented in section 4.3 along with the data smoothing and sampling requirement used to prepare model input. In section 0, the models calibrated in this study is discussed in details. Calibration methodology used in this research is presented in section 4.5. Calibration results are presented in section 4.6. Finally, in section 4.7, discussion about some significant outcome of the study and some recommendation of future researchers is presented.

## 4.2. BACKGROUND

Previously, due to unavailability of comprehensive microscopic data, a different set of macroscopic data have been used in the literature. For example, Toledo *et al*. used sensor and loop detector data (flow and speed) from Stockholm, Sweden to calibrate the driver behavior model of the MITSIM software package (Ben-Akiva, Davol et al. 2000, Ben-Akiva, Toledo et al. 2003). The authors then validated the calibration result using macroscopic traffic measures like traffic flow, travel time, and queue length (Ben-

Akiva, Toledo et al. 2003). Punzo *et al*. used flow, speed and vehicle classification data collected at Napoli-Salerno (E-45) two-lane freeway in Italy to calibrate the Gipps car-following model (Punzo and Tripodi 2007). To calibrate the model parameters, the authors converted the microscopic car-following model into a multi-regime steady state traffic stream model whose output matches the collected real-world data. A similar approach was taken by Rakha *et al*. who used loop detector data to fit a derived steady state traffic stream model from the Gipps car-following model (Rakha, Pecker et al. 2007). Rakha *et al*. used a similar approach to calibrate the Van Aerde traffic stream model (Rakha and Arafeh 2010). Although this method provides a fast and dirty option for calibrating car-following model by using relatively cheaper macroscopic traffic data, the drawback lies in the steady-state assumption as they might not be representative of the car-following region of driving states. Additionally, this kind of calibration provides an aggregate estimate of the whole population. Previous studies have shown a large variability in behavior among drivers as well as large intra driver variation in different condition (Kesting and Treiber 2008). To, account for the variability in driver population microscopic vehicle trajectory data is proposed to the suitable choice (Punzo, Borzacchiello et al. 2011).

Vehicle trajectory data can be collected using different methods. However, the accuracy of the trajectory data varies based on the collection method. The measurement accuracy has a crucial impact on the reliability of the calibration results (Ossen and Hoogendoorn 2008). As classified by Ossen and Hoogendoorn there are two main methods of trajectory data collection: 1) data collection through field experiments and 2) data collection using driving simulators (Ossen and Hoogendoorn 2008). The authors further classified field data collection into two groups namely: 1) instrumented vehicles or floating car method and, 2) aerial observation or remote sensing method.

Instrumented vehicles are usually equipped with position sensing equipment like Differential GPS along with other sensors like LIDER or even camera which tracks the equipped vehicle and follower/leader vehicles. For example, Gurusinghe *et al.* collected GPS data from a platoon of ten vehicles on a test track in Japan using DGPS with a sampling rate of 0.1 second (Gurusinghe, Nakatsuji et al. 2002). Upon

analyzing the data quality, the authors determined that, this kind high resolution data can facilitate accurate car-following analysis. The authors also estimated driver reaction time graphically by time series plot of relative speed and acceleration of the follower vehicle. This dataset was then used by Ranjitkar *et al.* to examine the stability of stimulus-response based car-following models with different platoon size and found the stimulus-response based car-following model is unstable both locally and asymptotically. The authors also found large intra and inter-vehicle variation in reaction time (Ranjitkar, Nakatsuji et al. 2003). Ranjitkar *et al.* also concluded that the stimulus-response based car-following model might be too simple to explain driver's behavior in different conditions. Brockfeld *et al.* used same dataset to calibrate ten different car-following models with varying number of parameters(Brockfeld, Kühne et al. 2004). Most important findings of this study was that, no model was better than others with statistically significance in explaining the car-following behavior. The authors concluded that parameters estimated at one condition ends up causing high error in others (Brockfeld, Kühne et al. 2004). Ranjitkar *et al.* used the test track data again to evaluate performance of six different car-following models using speed and headway as objective variable (Ranjitkar, Nakatsuji et al. 2004, Ranjitkar, Nakatsuji et al. 2005). The authors found calibration result with speed as objective variable more accurate than headway, which was caused by nosier headway data compared to less noisy speed data (Ranjitkar, Nakatsuji et al. 2004). One important finding is large variation in reaction time of driver from model to model. This finding lays down the ground for the argument that perception-reaction time should be independent of the other model parameters and should only be dependent on the driver population. Upon these findings a strong argument can be made to remove reaction time from calibration parameter and microscopic data is better as ground truth data rather than macroscopic data.

An important drawback of test track data is that they were collected on a test condition and may 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 and Simonelli 2005). This dataset was used for calibration and validation of four microscopic car-

following models. The validation of the calibrated model using the real world traffic data resulted in higher errors than the test track data indicating the need of using real-world traffic data for more robust and accurate calibration.

The second method of trajectory data collection is by means of video. Methods 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 example of terrestrial video data collection has been done in the NGSIM program. Video data were 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 (Halkias and Colyar 2006, Halkias and Colyar 2007, Halkias and Colyar 2007). The length of each roadway segment was between 500 and 640 meters and the data was collected for different durations and time of the day (Halkias and Colyar 2006, Halkias and Colyar 2007, Halkias and Colyar 2007). A customized software application, NG-VIDEO, was used to track the position of every vehicle at a frequency of 0.1 seconds and subsequently calculate traffic data like speed, acceleration, and gap (Halkias and Colyar 2006, Halkias and Colyar 2007, Halkias and Colyar 2007). Analysis of NGSIM data has shown processing issues such as vehicle overlaps, platoon inconsistency, vehicle misidentification, unexpected merging and overtaking, lane identification error and unusual lane changing, and unusually high fluctuation of acceleration (Punzo, Borzacchiello et al. 2011). Regardless of inherent drawbacks, the NGSIM data have been used for development, calibration, and validation of many microscopic traffic flow models including the Target lane-changing model (Choudhury 2005), cooperative/forced freeway merging behavior (Choudhury, Ramanujam et al. 2009, Wan, Jin et al. 2014), arterial lane selection model for urban intersections (Choudhury and Ben-Akiva 2008), lane position and lane changing execution model for different vehicle classes (Moridpour, Sarvi et al. 2010, Aghabayk, Moridpour et al. 2011). This dataset has also been used in calibrating existing models (Monteil, Billot et al. 2014), as well as determining the effect of correlation between model parameters (Kim and Mahmassani 2011). To overcome the drawbacks of the NGSIM data Montanino and Punzo

proposed a multistep data filtering method (Montanino and Punzo 2013). A key aspect of this data filtering method are i) eliminating outliers that causes improbable values for speed and acceleration by local reconstruction of the vehicle trajectory and ii) reduce the residual random disturbances from the time series output while still preserving the driving dynamics (Montanino and Punzo 2013).

Another type of video data used in calibration of the car-following model was done by Ossen and Hoogendoorn (Ossen and Hoogendoorn 2005). The authors used an aerial camera mounted on a helicopter to collect data from Freeway A2 in Utrecht, in the Netherlands to estimate the parameters for the stimulus-response based (GHR) car-following model and found that car-following behavior varies widely within the observed driver population (Ossen and Hoogendoorn 2005). The same dataset was used in calibration and parameter estimation of other car-following models,(Hoogendoorn and Hoogendoorn 2010, Hoogendoorn, Hoogendoorn et al. 2011) and for determining heterogeneity in car-following characteristics (Ossen and Hoogendoorn 2011). A similar approach was used for the development of the multi-anticipated car-following model by Hoogendroon *et al.* (Hoogendoorn, Ossen et al. 2006).

In summary, a wide range of data with varying accuracy and cost have already been used in calibration of the car-following model. Onboard sensor data suffers from measurement errors caused by limitations of the sensors and requires complex data filtering and data smoothing (Ossen and Hoogendoorn 2008, Ossen and Hoogendoorn 2009, Punzo, Ciuffo et al. 2012). Data collected in this manner is not representative of the driver population but only the test vehicles. On the other hand, video and other remote sensing data are capable of providing more comprehensive information about driver behavior for a population. Nevertheless, video data processing is challenging, spatially restricted and of questionable quality (Punzo, Borzacchiello et al. 2011). So far video data has not been collected in enough location and enough duration to represent behavior for different types of road that constitute a dense urban area network as well as to observe driver behavioral change in the different level of congestion. WAMI data provides the opportunity to overcome the shortcomings of previous data collection effort. This microscopic data also provides an opportunity for obtaining a distribution of every parameter of any model due to a large number

of sample size on varying road type and traffic control measures. It is expected that use of distribution of parameter is likely to reduce error in the simulated output. This chapter pursues the validity of these assumptions.

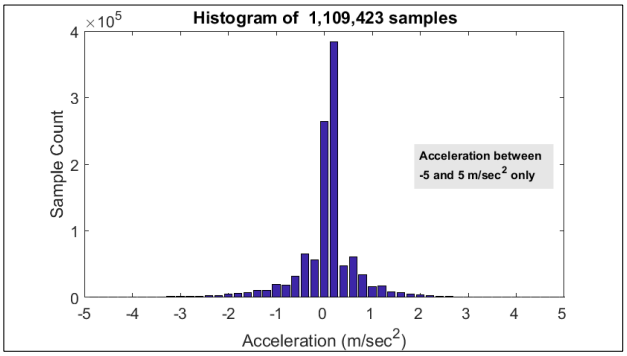
## 4.3. TRAJECTORY DATA PREPARATION

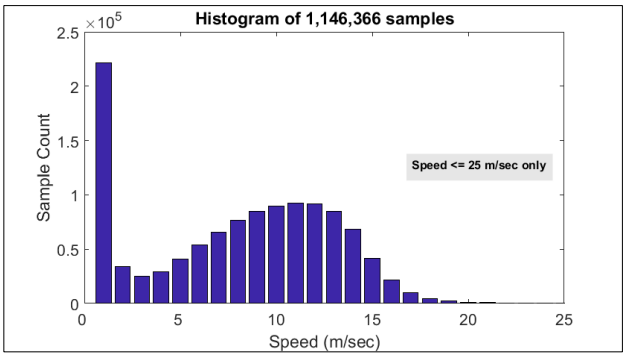
As discussed in section 3.6 the processed microscopic trajectory data is stored in a network schema containing its trajectory, route, and inter-vehicle interaction (platoon) related information. However, quality of the processed trajectory data is essential to obtain any degree of confidence in the validity of the data set. Quality control issues like negative spacing, sudden and high magnitude change in acceleration, and frequent lane change due to faulty image registration must be addressed before applying the data in model calibration. Another important issue for using trajectory data for calibration purpose is to select the dataset that is applicable for driving regime car-following models operate in. As a result, the entire trajectory dataset could not be applied in calibration. Sampling the dataset that most likely to represent car following regime was required for obtaining acceptable parameter estimation. These issues are covered in this section.

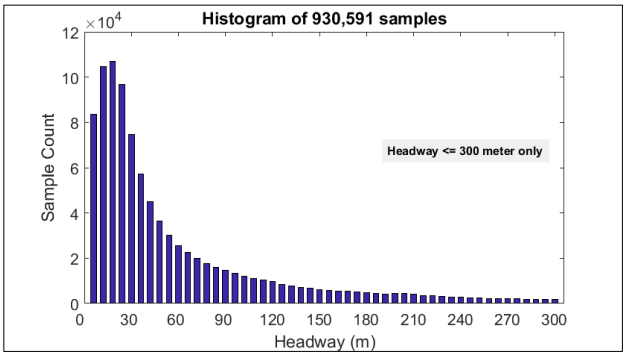
### 4.3.1. TRAJECTORY DATA ANALYSIS

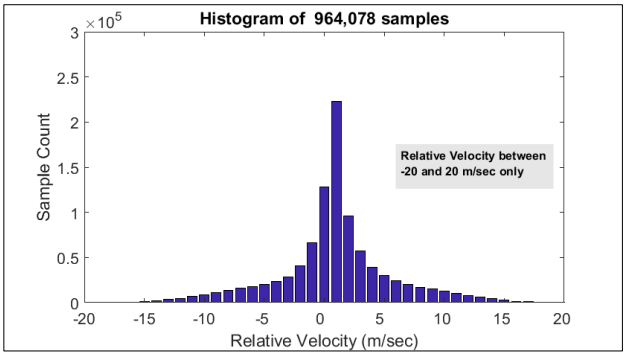
Quality of vehicle trajectory data can be assessed in multiple ways. One option is to assess fluctuation in individual trajectories and another option is to use their aggregate measures along with their associated distribution. The distribution of microscopic measures like speed, acceleration, and the inter-vehicle gap can give important information on the quality of the data. Generally, the change in the microscopic measures mentioned above is smooth in nature. Too many erratic changes in microscopic measurement are indicative of measurement error. In the following figure, the distribution and descriptive statistics of vehicle speed, acceleration, and the gap are provided for two main arterials of the city of Hamilton namely the Main street and the king street.

## King Street

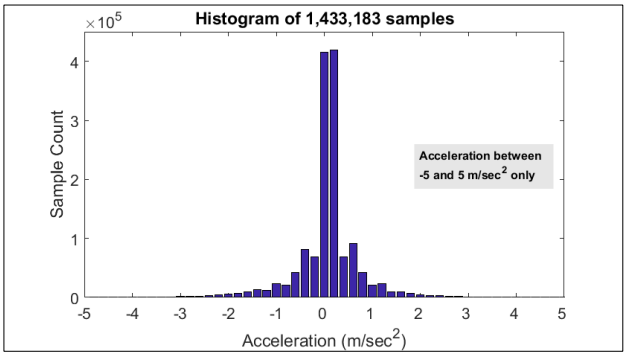


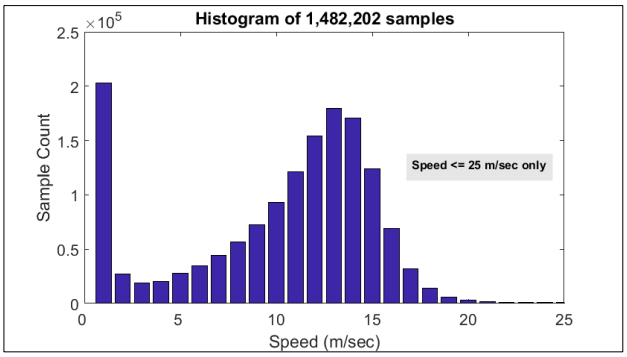



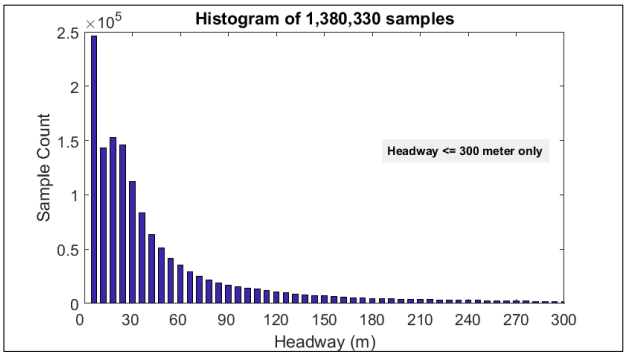



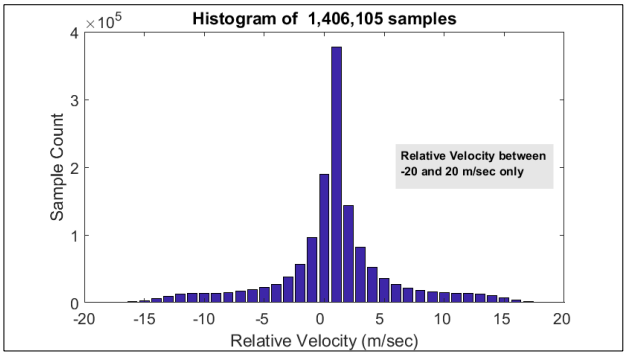




## Main Street



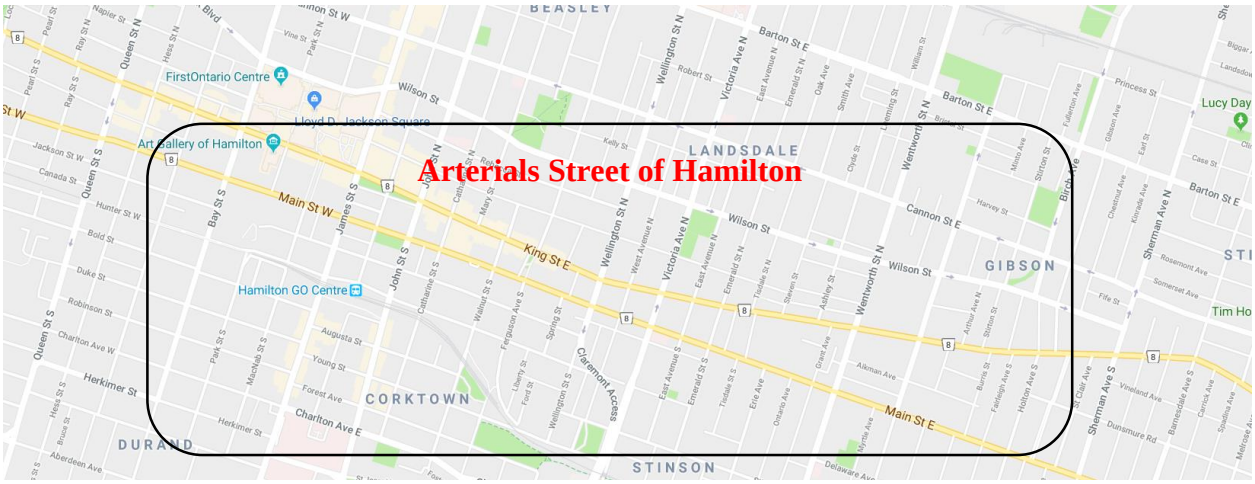



**Figure 14: Distribution of Microscopic Data before Smoothing.**

*source: Maps-Google Map data 2018*

### 4.3.2. DATA SMOOTHING

All the microscopic traffic measures are estimated from the position of the vehicle during the consecutive discrete time interval. Any error in vehicle location and image registration is likely to cause the higher magnitude of error in any other derived measure. In this case, speed and acceleration are estimated from the positional data. An error of 1 meter in the position is going to cause 1 meter/sec error in the speed estimation and 1 meter/sec$^2$ error in the acceleration estimation. 1 meter/sec$^2$ error is acceleration is more consequential in estimating model parameter than 1-meter error in vehicle position. For details about the error propagation in vehicle trajectory data and the requirements of data smoothing please refer to (Montanino and Punzo 2013). Measurement error can lead to considerable bias the parameter prediction of car-following models. Although Treiber, and Kesting (Treiber and Kesting 2013) did not find any significant impact of data smoothing on the estimated parameters, measurement error is an important issue that needs addressing. One important aspect of smoothing time series trajectory data is that a vehicle in a traffic stream is on a dynamic system which should be considered while applying any smoothing application. Smoothing acceleration, speed, and inter-vehicle spacing must be done maintaining individual trajectory and platoon consistency. To smooth individual trajectory, the following rules must apply:

$$x_n(t + \Delta t) = x_n(t) + v_n(t) * \Delta t \quad (16)$$

$$v_n(t + \Delta t) = v_n(t) + a_n(t) * \Delta t \quad (17)$$

From the trajectory data, speed is estimated by calculating the distance traveled by the vehicle in two consecutive times spent and the acceleration is estimated by calculating the difference in speed the said time step. So, the trajectory consistency rule is already applied in the dataset. As a result, any measurement error in the data comes from faulty positional information. However, it is easier to smooth the acceleration and speed than the vehicle positional data as their time series data has a causal relationship between each time steps. In other words, vehicle acceleration or speed can be expressed as a function of time. Vehicle speed can be recalculated from the smoothed acceleration data. To avoid negative spacing, vehicle overlapping, and inconsistence car-following behavior inter-vehicle spacing also needs to be recalculated based upon the smoothed acceleration data. For the purpose of maintaining platoon consistency the following rule must apply:

$$x_n(t) = x_{n+1}(t) + l_{n,n+1}(t) \quad (18)$$

Another method of data smoothing is to treat vehicles as a sole dynamic system and one consistent estimation problem which can be solved using a Kalman Filter. However, designing a Kalman Filter requires variance of estimation error in sensor measurement which is yet to be established for WAMI data. As a result, a whole trajectory reconstruction based method starting from smoothing of acceleration is suitable for this data. A three step process is used in this study for data smoothing and trajectory reconstruction. The process is as followed:

**Step 1: Smooth acceleration data using suitable method**

For a large scale implementation like this study the selected smoothing methods needs to be fast and robust from outliers. Six different smoothing method were applied on a sample of data to observe their

performance. These methods are, moving average, Local regression using weighted linear least squares and a 1st degree polynomial (lowess), Local regression using weighted linear least squares and a 2nd degree polynomial (loess), Savitzky-Golay filter, and robust version of lowess and loess. For all method a smoothing span of 5 was used, which for this data indicated acceleration of each data point is estimated based on the neighboring 5 seconds of data. The plots of raw and smoothed data for different methods is shown in Appendix B. It is clear from the figure that lowess and robust version of lowess performs much worse than other method. In loess and robust version of lowess the smoothed value is more influenced by the preceding values than the moving average and Savitzky-Golay filter. However, the performance of the moving average method and the Savitzky-Golay method were very similar. Due to ease of implementation and lower processing time, moving average method is used to smooth the acceleration time series of this trajectory data.

**Step 2: Recalculate Speed from Smoothed Acceleration Data**

In this step of trajectory reconstruction, the vehicle speed is recalculated using the smoothed acceleration data. Vehicle acceleration at current time step is added with the speed to obtain the speed of the next time step. The process can be written in the following mathematical formulation:

$$v_n(t+1) = v_n(t) + a_n^S(t) * \Delta t \tag{19}$$

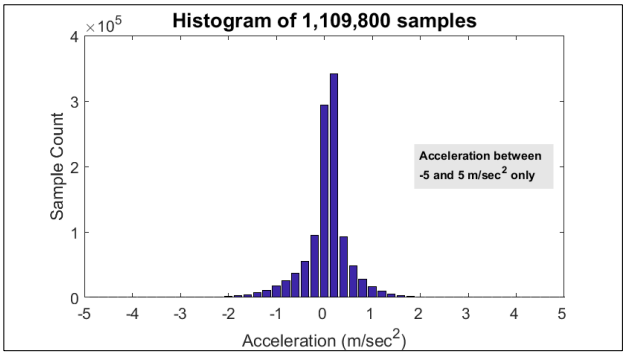
**Step 3: Recalculate Inter-Vehicle Spacing**

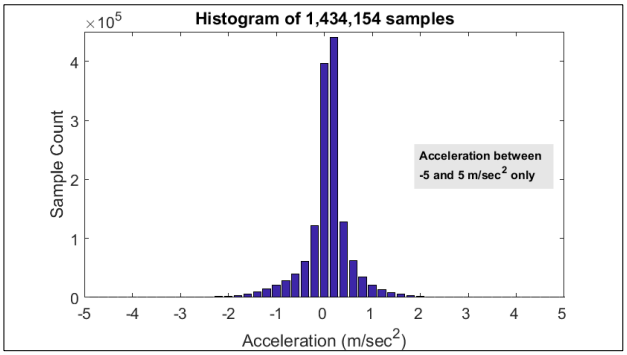
The concluding part of trajectory reconstruction is to recalculate the vehicle spacing from the smoother speed. To accomplish this task, the difference in speed between the smoothed data and raw data is calculated at every trajectory points. When multiplied by the time interval (1 sec) this speed difference indicates a change in vehicle spacing due to the smoothing of vehicle acceleration and speed. The inter-vehicle spacing between every leader-follower pair is then calculated using the following mathematical expression.

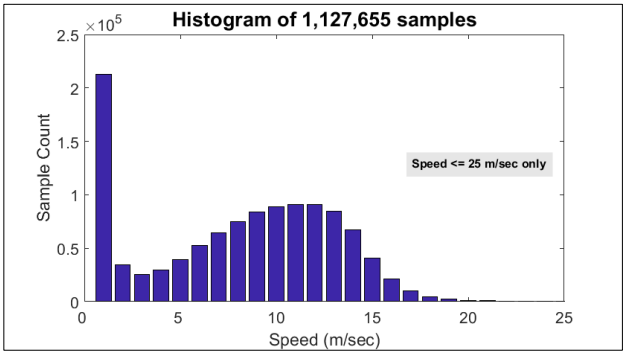
$$l^{s}_{n,n+1}(t) = l_{n,n+1}(t) + \Delta v_n(t) * \Delta t - \Delta v_{n+1}(t) * \Delta t \quad (20)$$

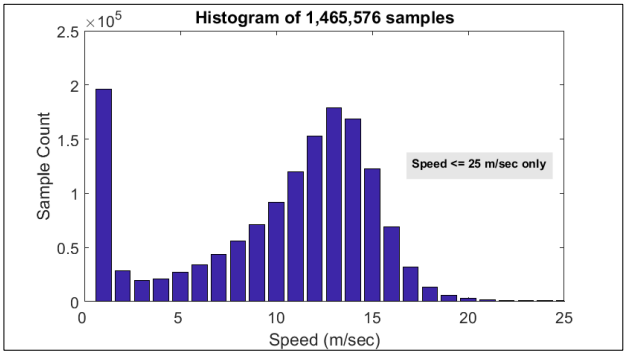
The trajectory reconstruction is stopped at the calculation of vehicle spacing. Recalculation of vehicle position on the 2-D surface is not conducted because this is not required to calibrate the car-following model and is out of the scope of this study.

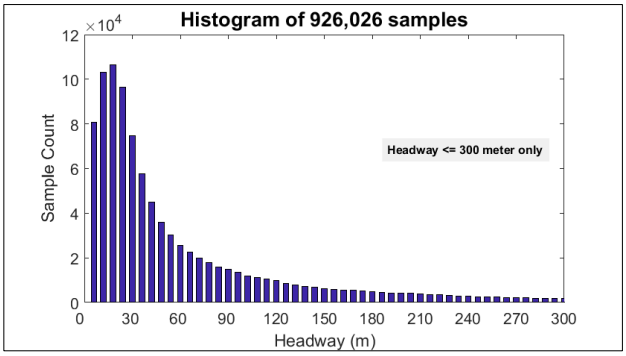
**King Street** | **Main Street**

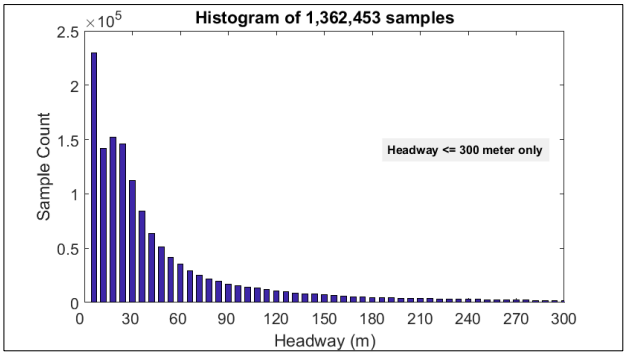


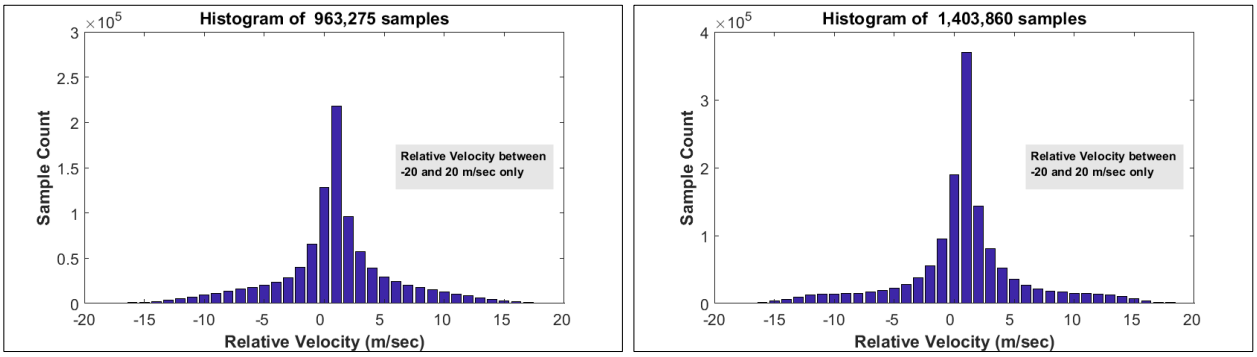



**Figure 15: Distribution of Microscopic Data after Smoothing.**

### 4.3.3. DATA SAMPLING FOR CALIBRATION

Sampling the trajectory data to calibrate car-following model is important for two reasons. First, sampling should remove extreme data points caused by measurement error. These values are often outside the realm of physical possibility. As discussed in section 4.2, any global least square error optimization algorithm used to estimate parameters is susceptible to noise in the input data. Although, there are mathematical ways to reduce the impact of an outlier on the estimated parameter, removing them completely is a more effective way to obtain good parameter estimate. Additionally, as discussed before, the car-following models are developed to predict diver behavior with respect to the trajectory information of the leading vehicle. WAMI data is collected on the roadway where all three regimes of traffic are observable. The emergency regime can sometimes be observed at approach or departure from an intersection or near-crash situation. These traffic regimes can be observed on the speed distribution plot shown in Figure 14 and Figure 15. Selection of trajectory data points in car following regime can be obtained by filtering the trajectories by providing restriction on the microscopic measure. As car following model require an offset time (Sangster, Rakha et al. 2013), another way to filter data is to select only the leader-follower pair with longer event duration. An event is defined as the consecutive time period a particular follower vehicle is following a particular leader vehicle. it. Sangster, Rakha et al. used event duration of more than 5 seconds in the calibration of car-following model (Sangster, Rakha et al. 2013). If noise reduction is undertaken for trajectories with longer event duration (acceleration ±0.05g), then the resulting data points is expected to represents steady state car

following regime. In this study both approaches have been used to select microscopic data for the calibration analysis.

#### 4.3.3.1. SAMPLING ON THE ENTIRE TRAJECTORIES

A large number of observations have zero speed values. These values are indicative of vehicles stopped at traffic light. Excluding the zero values, the speed distribution shows a normal distribution skewed towards the lower end. Car-following regime is observable in the region between intersections where the drivers are responding to their prospective leader. As the data is collected during peak congested period it would be acceptable to assume that most of the observation would be in the car-following/emergency regime. Free flow regime is observable near the end of the green signal cycle of a traffic controlled intersection. On the speed distribution plot, this regime is represented along the right tail. To address these data sampling issues, a multistep data mining approach is undertaken to select set of microscopic data most likely to represent car-following regime.

**Step 1: Sampling on Time Headway**

Most traffic simulation tools trigger a transition between the three traffic regimes based on the time headway between the leader-follower pair. A higher time headway threshold is used to separate traffic between the free flow regime and car-following regime. Similarly, lower time headway threshold is used to separate vehicles from a car-following regime and emergency regime. Often some distribution of threshold is used to represent heterogeneity among drivers. However, for simplicity, it is preferable to use a single value of upper and lower time headway. For this study, the upper limit of time headway is chosen as 3s and the lower limit is chosen as 0.25s. Time headway is calculated by dividing space headway at any instance by the instantaneous speed. If the vehicle speed goes near to zero, the time headway might be relatively high. But this condition is the only representation of stop-go condition (emergency regime) rather than a car-following condition. Due to this reason filtering based on space headway alone is undertaken in next step.

**Step 2: Sampling on Space Headway**

Filtering based on space headway is done to compliment the time headway filtering. Most of the longer headway is filtered away by using the time headway filtering. Filtering out data point very close to each other in terms of space headway is likely to remove the stop-go condition of traffic. Although vehicles can follow each other in close proximity their driving characteristic is likely to be governed by the emergency regime. For the purpose of this study, any space headway lower than 2 times the vehicle body length is discarded from calibration data sample. Vehicle length used to calculate gap was 6m. Using this method any gap lower than 12m is discarded from the calibration analysis.

**Step 3: Sampling on Speed**

Filtering on speed is done to reduce the number of outliers from the calibration data. As the distribution of speed data has some central tendency and the car-following regime is more likely to be around the median a variance based approach is taken to filter data based on speed. Data is only selected if they within a 1 standard deviation span around the mean. The mean and standard deviation used in this method is estimated prior to any filtering based on time or space headway.

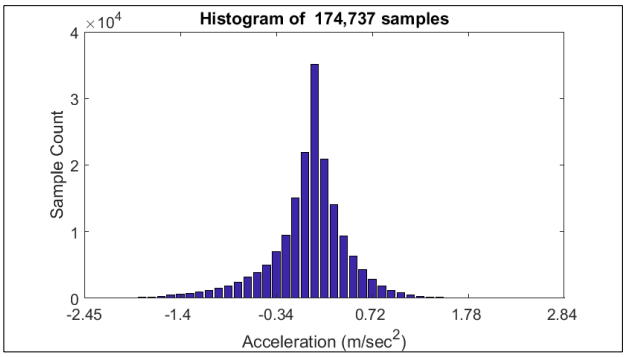
**Step 4: Sampling on Acceleration**

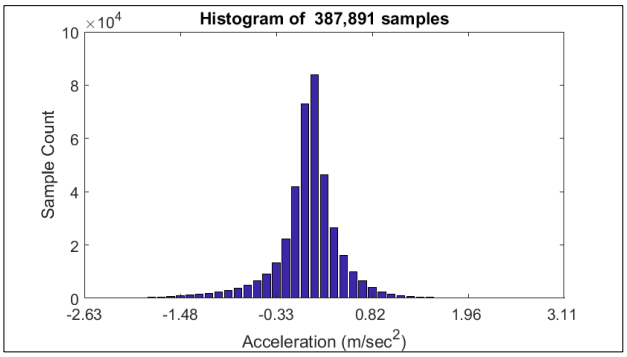
The purpose of filtering based on acceleration is similar to filtering based on acceleration. Acceleration, in fact, shows a better central tendency than speed. In both cases, distribution of acceleration shows normal distribution around mean of 0. Similar to speed based filtering, data is only selected if they within a 1 standard deviation span around the mean. The mean and standard deviation used in this method is estimated from the whole set of smoothed microscopic data.

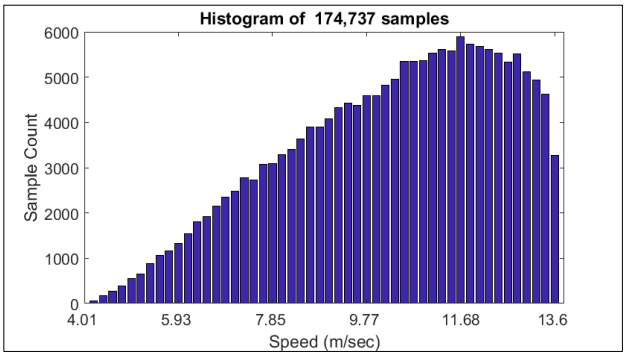
**Step 5: Sampling on Relative Velocity**

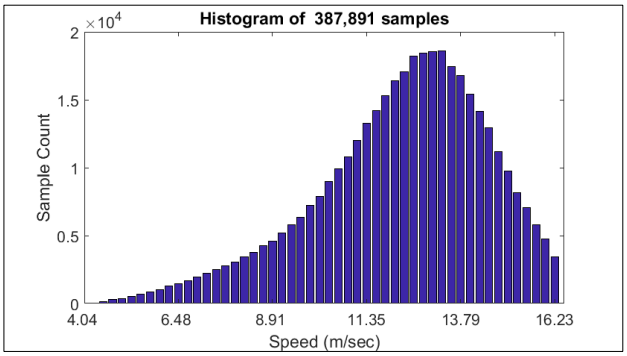
This filtering step is undertaken on the dataset after the first four steps are completed. This step removes noisy data points in terms of relative velocity. Like acceleration, relative velocity also shows a central tendency around zero. Similar to acceleration based on filtering relative velocity based filtering is also based on its distribution. Data is only selected if they are within 2 standard deviation span around the mean. The distribution of acceleration, speed, and headway after data filtering is shown below.

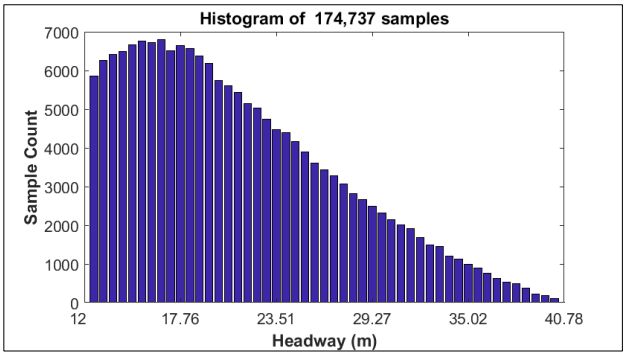
**King Street** | **Main Street**

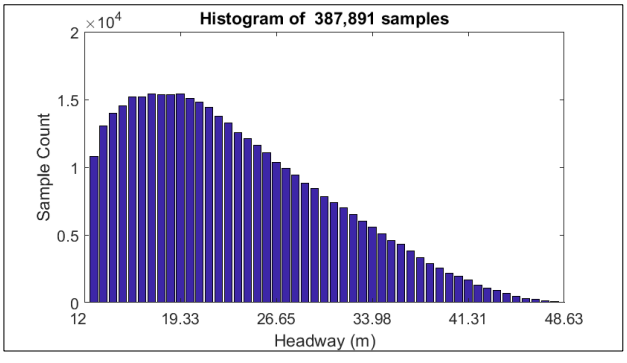


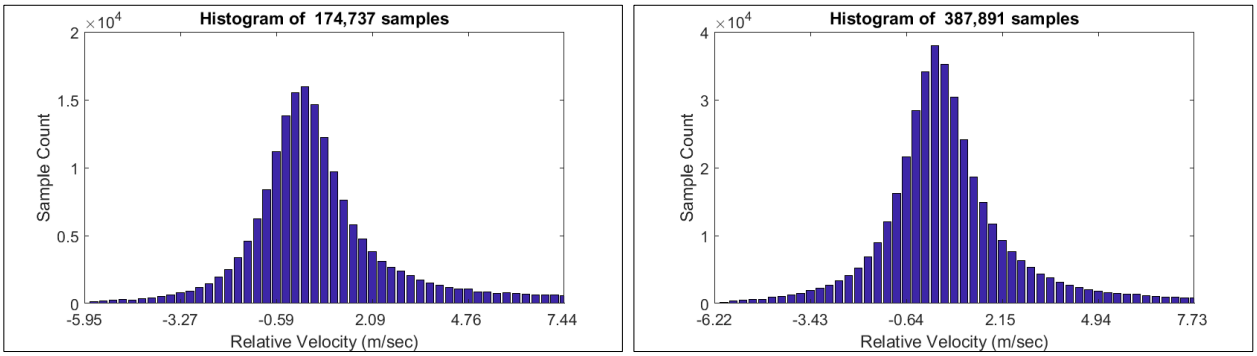



**Figure 16: Distribution of Microscopic Data after Sampling.**

In case of King Street, only 15.23% of data points meet the filtering requirement, while for Main Street the amount proportion was 26.14%. This method of data sampling is designed to find data points that are most likely in the car-following regime in a congested urban network. This data reduction technique does not consider the interaction between individual leader-follower pair but considers their position in the distribution of microscopic traffic parameters. In this kind of dense urban network, the duration of a leader-follower event is expected to be shorter and using a limiting criterion on how long a follower follows any particular leader in a traffic stream might reduce the amount of sampled data points further. The speed-spacing plots for the total and the sampled data is shown in the following.

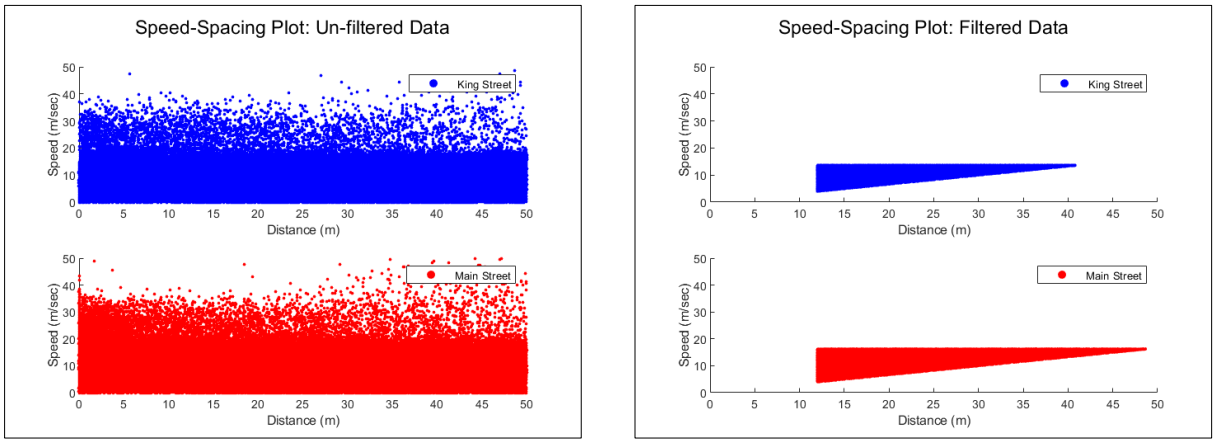



**Figure 17: Speed vs Spacing plot for the entire and sampled dataset**

Form Figure 17 it is clear that filtering technique based on constraint on microscopic measurement just samples the data form a narrow region of the entire space.

#### 4.3.3.2. SAMPLING BASED ON EVENT DURATION

In this method the duration of every leader-follower event is calculated. The event duration helps in identifying car following behavior without putting any restriction on the microscopic measures. This method helps in data reduction while maintaining trajectory consistency. As data is selected based on individual events not individual data points, this method maintains consistency for every event. For this study event duration of equal to or more than 30 seconds was selected to reduce sample size for calibration. Noise reduction technique as mentioned in the previous section is also undertaken. Only steady-state condition defined by Ozaki (Ozaki 1993) (acceleration ±0.05g) is used for filtering the data. Low values of speeds are also removed. The fundamental diagram of the events selected for the analysis before and after noise reduction is shown in the following figure.

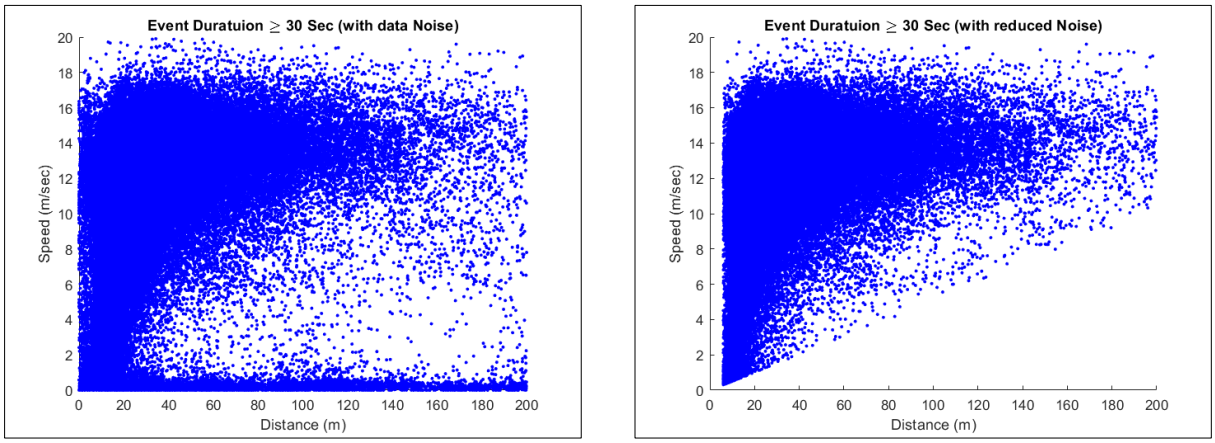



**Figure 18: Speed vs Spacing Plot for Event Based Sampling**

## 4.4. CAR-FOLLOWING MODELS AND THEIR INPUT PREPARATION

Three different car-following models are calibrated in this study. These models not only vary on model structure but also model input. For example, the GHR model is non-linear and the Helly model is linear. Both GHR and Helly model use perception-reaction time to model the time lag in applying control mechanism. On the other hand, the IDM model predicts the instantaneous value of acceleration. The time lag between a stimulus and driver's reaction is modeled by using a term called desired time headway. The structure of the models under consideration and data input requirement is discussed in the following.

### 4.4.1. STIMULUS-RESPONSE BASED MODEL

The GHR model and related GM model are one of the oldest and most-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 a function of the difference between the speed of the leader and follower vehicles. The model structure puts it into the broad classification of stimulus-response models. The basic structure of the model is presented in Equation 1.

$$\text{Response (t + T)} = function\ (sensitivity, stimulus\ (t)) \tag{21}$$

Where the *response* is representative of vehicle control (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 which signifies the delay in response to the stimulus. The delay in response is a measure of the time that elapses before the driver notices a change in stimulus and starts responding to it. Early in its development, the proposed model was linear. Rapid development led to a more generalized non-linear approach and several validation studies suggested that driver sensitivity was not constant for every condition. Later studies suggested that sensitivity increased with vehicle speed and decreased with the increasing 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).

$$a_n(t) = \alpha v_n^{\beta}(t)\left[\frac{\Delta v(t-T)}{\Delta x^{\gamma}(t-T)}\right] \tag{22}$$

Where,

| | |
|---|---|
| an (t) | = Acceleration/deceleration of driver n at time t |
| α, β, γ | = Constants to be determined (Needs to be Calibrated) |
| $v_n(t)$ | = Velocity of the follower vehicle n at time t |
| $\Delta v(t-T)$ | = Relative velocity at time t-T |
| $\Delta x(t-T)$ | = Relative gap between leader and follower at time t-T |
| T | = Perception-reaction time |

The input data for this models is the relative velocity between leader and follower, gap, and speed of the follower car itself. The output of the model is the acceleration of the follower. Model input data is already available from the WAMI data itself while the output data require some interpolation between consecutive time points. This is due to the difference between the frequencies of WAMI data and that required for the car-following model. From WAMI data response/output data is only obtained at 1 sec interval where in case of simulation a 0.1 sec time is preferable. For the purpose of this study, perception-reaction time is modeled independently from the model parameter as discussed in section 4.2. Perception-reaction time is obtained from a log-normal distribution with a mean value of 0.73 sec and standard deviation of 0.18 which is taken from values listed in the literature (Schweitzer, Apter et al. 1995, Green 2000). It is assumed that for the time horizon a vehicle is captured in the WAMI data, the perception-reaction is constant for an individual vehicle. Perception-reaction time is estimated with the precision of 0.1 sec and the response variable for that reaction time is estimated by linear interpolation between the corresponding time steps for that vehicle. For example, if the perception-reaction time is 1.8 sec then the acceleration at 1.8 sec is estimated by linear interpolation between acceleration at a time interval of 1 sec and 2 sec after the time stamp of the data point. Sample calibration input for GHR model is shown in Appendix C.

### 4.4.2. LINEAR (HELLY) MODEL

Although linear, this class of model differs from the first generation GHR (GM) model in that Helly proposed a car-following a model where the relationship between acceleration and the difference between the current and desired gap, and relative velocity is linear (Helly 1959). The mathematical formulation is shown in Equations 3 and 4.

$$a_n(t) = C_1 \Delta v(t - T) + C_2 [\Delta x(t - T) - D_n(t)] \quad (23)$$

$$D_n(t) = \alpha + \beta v_n(t-T) + \gamma a_n(t-T) \quad (24)$$

Where,

$a_n(t)$ = Acceleration of the follower vehicles at time t
T = Driver reaction time
$\Delta v(t-T)$ = Relative velocity between vehicles at time t-T
$\Delta x(t-T)$ = Relative distance between vehicles at time t-T
$v_n(t-T)$ = Velocity of the follower at time t-T
α, β, γ, C1, C2 = Calibration Constants

Subsequent calibration of the model resulted in wide-ranging values of T, $C_1$, and $C_2$. Calibration of this model requires estimation of 6 parameter T, $C_1$, $C_2$, α, β and γ, which makes it more difficult than most (Panwai and Dia 2005). As in this study the driver reaction time is modeled before calibration the number of parameter to be estimated reduced to 5. The 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). The input data of this model is similar to the stimulus-response based model with the addition of acceleration of the follower vehicle. Output/response variable is estimated using a similar methodology as discussed in section 4.4.1. Sample calibration input is shown in Appendix C.

### 4.4.3. INTELLIGENT DRIVER MODEL

The Intelligent Driver Model (IDM) is collision free continuous function model proposed by Treiber *et al* (Treiber, Hennecke et al. 2000). This model is a multi-regime model where the vehicle acceleration is defined as a continuous function of dynamics of the lead-follower pair of vehicles. The main difference is that the IDM doesn't use the time lag used in the previous model in term of the reaction time. Instead, the IDM tries to incorporate the desired time headway parameter value to incorporate the similar effect of the reaction time. The mathematical formulation of the model is shown in 25 and 26.

$$\dot{v_\alpha} = a\left[1 - \left(\frac{v_\alpha}{v_0}\right)^{\delta} - \left(\frac{S^*(v_\alpha, \Delta v_\alpha)}{S_\alpha}\right)^2\right] \quad (25)$$

$$where, S^*(v_\alpha, \Delta v_\alpha) = S_0 + S_1\sqrt{\frac{v_\alpha}{v_0}} + T^\alpha v_\alpha + \frac{v_\alpha \Delta v_\alpha}{2\sqrt{ab}} \quad (26)$$

Where,

$\dot{v_\alpha}$ = Acceleration of vehicle α at any time
$a$ = Desired maximum vehicle acceleration
$b$ = Desired maximum vehicle deceleration
$T^\alpha$ = Desired time headway
$v_0$ = Desired free flow speed
$S_0$ = Minimum desired headway/Jam distance
$v_\alpha$ = Current velocity of vehicle α
$\Delta v_\alpha$ = Relative velocity between vehicle α and its leader
$S_\alpha$ = Current spacing between vehicle α and its leader

According to Treiber *et al* (Treiber, Hennecke et al. 2000) the parameter $S_1$ can be considered as zero and the value of. So the there are five parameters ($a, b, T^\alpha, v_0, and\ S_0$ ) that needs calibrating. The parameter b is only active is case of non-steady state of traffic where the following vehicle is approaching a slower or stationary vehicle in front. Input variables are follower velocity, spacing, and approaching speed of the follower vehicle. All of these input and output is available from the processed WAMI data. Sample calibration input is shown in Appendix D.

## 4.5. CALIBRATION METHOD

Calibration method used to estimate parameter has a profound impact on the result. Two different approaches for calibration is applied. In the global approach, a complete data trajectory is compared with a simulated trajectory and the objective function is defined in terms of a sum of squared errors (SSE) of gaps or speeds or acceleration. In local approach, the dependent model variables are compared with observed value by maximum likelihood technique. This is done separately for each data point. The global approach is often termed as simulation-based and the local approach is termed as model-based (Ciuffo, Punzo et al. 2008).

### 4.5.1. GLOBAL LEAST-SQUARED ERRORS CALIBRATION

In the global deterministic approach, trajectories are extracted from the simulation and compared with the observed trajectories by formulating objective functions in terms of sums of squares error of the chosen variables. 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, Ciuffo 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 an absolute and relative measure of the gap are provided in Equations 6 and 7, respectively.

$$Min,\ \ S_s^{abs} = \sum_{i=1}^{n} \left(S_i^{sim} - S_i^{obs}\right)^2 \tag{27}$$

$$Min,\ \ S_s^{abs} = \sum_{i=1}^{n} \left(\ln S_i^{sim} - \ln S_i^{obs}\right)^2 \tag{28}$$

The objective function in 28 is less sensitive to outliers and measurement error than 27. This is because the objective function of Equation 27 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 28 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, Ciuffo 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, Ciuffo et al. 2012). In this study, the IMSE approach

is not applied as the primary focus of the study is assessing the quality of the WAMI data and more examples is available to compare for regular LSE approach.

### 4.5.2. LOCAL MAXIMUM-LIKELIHOOD CALIBRATION

The maximum likelihood parameter calibration stems from the popular Maximum Likelihood estimation. It's a statistical approach which attempts to answer the question: Which value of model parameters are most likely to describe a given set of data? Primarily this method is used to estimate the parameters for a probability density function (PDF) of a particular distribution. If the distribution of the residual between the endogenous model parameter and the observed value can be fitted with any statistical distribution, then the MLE approach can be leveraged to estimate the parameters of the model. A likelihood function on residual is defined and MLE method is used to maximize this likelihood function, which along with appropriate optimization algorithm finds the parameter value that maximizes the agreement between the model and the data. This approach can be applied to models of arbitrary complexity. The MLE approach is same as LSE approach in a Gaussian model.

In all of the selected model for this study, the dependent variable is vehicles acceleration. Acceleration is calculated as a function of relative velocity, spacing, speed, desired speed, and desired headway. If independent and identically distributed (IID) zero mean Gaussian noise with unknown variance is assume then the prediction of acceleration can be written as,

$$a_{obs} = a_{sim} + \epsilon, \qquad \epsilon \sim N(0, \sigma^2) \tag{29}$$

$$a_{obs} = f(v_n, \Delta v, \Delta x, \dots) + \epsilon, \qquad \epsilon \sim N(0, \sigma^2) \tag{30}$$

As the error term of the model has well defined statistical properties the probability of observing an outcome given the parameters can be well defined. Due to the IID assumption the joint probability distribution of observing the data for a given set of parameters, $\beta$ can be written as a product of probabilities of individual data point:

$$L(\beta) = \prod_{i=1}^{n} p(a_i^{obs}|\beta) \tag{31}$$

As the individual probabilities are a very small fractional number, estimation of the joint probability often become infeasible even with high precision computing. Using log-likelihood is a good work around this problem. As the log is strictly monotonously increasing function it doesn't change the maximization criterion. The likelihood function can be transformed into:

$$\tilde{L}(\beta) = \sum_{i=1}^{n} \ln(p(a_i^{obs}|\beta)) \tag{32}$$

By using the log-likelihood function the parameter estimation problem turns into the calibration problem of the following format:

$$\hat{\beta} = \arg\max(\tilde{L}(\beta)) \tag{33}$$

With this objective function, the parameter estimation problem turns into a non-linear optimization problem. The main criticism of this approach is the IID assumption of the error term which is often serially correlated in case of the car-following data (Treiber and Kesting 2013). In this study, the "STAT4" and "BBMLE" package of statistical software R is used. A negative log-likelihood function is defined over the residual of the dataset. A zero-mean Gaussian distribution with unknown variance is used as a starting point of the calibration work. The "SANN" and "L-BFGS-B" method is used to optimize for a model parameter. The BFGS is a quasi-Newton method, which maintains and updates an approximation of the Hessian matrix based on the first-order derivative. This method cannot work with box restriction on the parameter values for which the L-BFGS-B method is used. The "mle" function of "STAT4" package often fails to invert the Hessian matrix due to poor starting point estimate. In those cases, the "mle2" function of "BBMLE" package is used.

## 4.6. RESULTS

Calibration is conducted in four different way using the two dataset sampled. A maximum likelihood calibration is conducted using with and without constraints on the parameter value. The same procedure is also followed for the Least-Squares method. The unconstrained parameter estimation gives somewhat of a global optimum solution (limited by the optimization algorithm) but the results can lead to unacceptable parameter values due to non-linear model structure and unknown distribution of the residual. Constraints on parameter values is one way to avoid this effect. For unconstrained calibration the “SANN” optimization method of “mle2” function of R is used. "SANN" is a variant of simulated annealing proposed by Belisle (1992). It is a stochastic global optimization method that works for non-differentiable function. This method is good for obtaining a good value on a rough surface. For constrained parameter estimation "L-BFGS-B" method is used. It’s a quasi-Newton method that allows box constraint on the parameter value. The two different dataset selected by the two approach used in data sampling is The calibration results of the three models are presented below.

### 4.6.1. PARAMETER ESTIMATES FOR GHR MODEL

The estimated parameter of the GHR obtained by the four different estimation process is listed in Table 1. The initial values of the parameters are selected from the default parameter values used in TransModeler. The driving regime is divided into acceleration and deceleration defined by positive and negative relative velocity with the leader respectively. The unconstrained estimate of the parameters varies greatly from the constrained estimation especially in the case of parameter $\alpha$ and $\gamma$. The parameter $\alpha$ represent model constant its variation is more acceptable. The parameter $\gamma$ is the exponent of the gap between a leader follower pair.

**Table 1: Parameter Estimates for GHR model**

| Dataset Used | Regime | Parameter | Starting Point for Estimation | Estimation with parameter Constraints | | | | Unconstrained Parameter Estimation | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Min | Max | Maximum-Likelihood Estimate | Least-Squared Estimate | Maximum-Likelihood Estimate | Least-Squared Estimate |
| Sampled Trajectory | Accel-eration | $\alpha_a$ | 2.15 | 1.5 | 2.5 | 1.5000 | 1.5000 | 3.6601 | 0.1694 |
| | | $\beta_a$ | -1.67 | -2.0 | -1.5 | -1.7316 | -1.6091 | -1.4406 | -0.0942 |
| | | $\gamma_a$ | -0.89 | -1.2 | -0.7 | -0.7000 | -0.7000 | 0.0959 | 0.0960 |
| | Decel-eration | $\alpha_d$ | -2.328 | -2.5 | -2.0 | -2.5000 | -2.0000 | -1.5230 | -0.8399 |
| | | $\beta_d$ | 0.961 | 0.8 | 1.2 | 1.2000 | 0.8000 | -1.0387 | -0.3795 |
| | | $\gamma_d$ | 1.116 | 0.9 | 1.4 | 0.9000 | 1.3676 | -0.0350 | 0.1388 |
| Event Based Dataset | Accel-eration | $\alpha_a$ | 2.15 | 1.5 | 2.5 | 1.5000 | 1.5000 | 0.0011 | 0.2362 |
| | | $\beta_a$ | -1.67 | -2.0 | -1.5 | -2.0000 | -1.6781 | -1.3740 | -0.1846 |
| | | $\gamma_a$ | -0.89 | -1.2 | -0.7 | -0.7000 | -0.7000 | 1.0929 | 0.2346 |
| | Decel-eration | $\alpha_d$ | -2.328 | -2.5 | -2.0 | -2.0000 | -2.0000 | 0.0834 | -0.8399 |
| | | $\beta_d$ | 0.961 | 0.8 | 1.2 | 0.8000 | 0.8000 | -0.0336 | -0.3795 |
| | | $\gamma_d$ | 1.116 | 0.9 | 1.4 | 1.4000 | 0.9000 | 0.0525 | 0.1388 |

In all cases the estimate4 of $\gamma$ is lower than expected. This indicates a lower sensitivity of vehicle response variable with respect to the gap in front. This can be explained by the fact that the data location is a dense urban network operating under traffic light where vehicles can anticipate the movement of its leader with much greater certainty. In both constrained maximum likelihood and least squared estimate the parameter $\alpha$and $\gamma$are at or near the boundary condition imposed upon them. However, the variation of parameter estimate indicates that the global optimization algorithms may be getting stuck at local optima. Other parameter estimation method like Bayesian estimation method can be helpful to get out the local optima.

### 4.6.2. PARAMETER ESTIMATES FOR LINEAR (HELLY) MODEL

The Helly model has 5 parameters that need calibration, thus making it more complex than the GHR model. Similar to the GHR model 4 different calibration approach was applied for the Helly model and the results are listed in Table 2. As this model is linear the distribution of the residual is more livelily

to follow a normal distribution than the non-linear model like the GHR. As a result, the constrained and unconstrained maximum likelihood estimate are close to each other than that of GHR model. The initial values of the parameter are taken from Rahman (Rahman 2013). The parameter α is a model constant so its variability among the different estimates is acceptable. The only other parameter that varies among the estimates is the parameter C2. The parameter C2 is associated with gap between leader follower pair. This behavior is similar to the GHR model where the model exhibited less sensitivity to gap between a leader follower pair. However, the results can't not be compared for this two cases as the GHR model were optimized for two different driving response where the Helly model is calibrated for both those responses together. Other three parameters are mostly within the expected values predicted in previous studies.

**Table 2: Parameter Estimates of Linear (Helly) model**

| **Dataset Used** | **Parameter** | **Initial Value** | **Estimation with parameter Constraints** | | | | **Unconstrained Parameter Estimation** | |
|---|---|---|---|---|---|---|---|---|
| | | | Min. | Max. | Maximum-Likelihood Estimate | Least-Squared Estimate | Maximum-Likelihood Estimate | Least-Squared Estimate |
| Sampled Trajectory | α | -0.9424 | -1.1 | -0.7 | -0.9420 | -0.7000 | -1.0495 | -0.5077 |
| | β | 1.3930 | 1 | 2 | 1.0000 | 1.7593 | 2.0144 | 1.7520 |
| | γ | -1.2969 | -1.6 | -1 | -1.6000 | -1.6000 | -1.4131 | -1.5017 |
| | C1 | -0.1052 | -0.2 | -0.05 | -0.0510 | -0.0515 | -0.0502 | -0.0508 |
| | C2 | -0.0015 | -0.003 | 0.003 | 0.003 | 0.0030 | 0.0085 | 0.0063 |
| Event Based Dataset | α | -0.9424 | -1.1 | -0.7 | -1.0263 | -0.7000 | -1.3382 | -0.4625 |
| | β | 1.3930 | 1 | 2 | 1.0000 | 1.6502 | 0.8982 | 1.5587 |
| | γ | -1.2969 | -1.6 | -1 | -1.6000 | -1.6000 | -1.3719 | -1.4867 |
| | C1 | -0.1052 | -0.2 | -0.05 | 0.0500 | -0.0601 | -0.0402 | -0.0684 |
| | C2 | -0.0015 | -0.003 | 0.003 | 0.0016 | 0.0030 | 0.0011 | 0.0058 |

### 4.6.3. PARAMETER ESTIMATES FOR INTELLIGENT DRIVER MODEL

IDM model is different from the previous two models as all of its parameters are intuitive and each of them relate to different driving regimes. For example, the parameter b is not used in case of free flow

state or when the leader is faster than the follower. In case of free flow condition vehicle acceleration is governed by the desired speed parameter. In case of vehicle approaching a slower or stopped vehicle the acceleration is governed by the desired gap parameters. In case of stop go condition the acceleration is governed by the desired minimum gap parameter. As all of the parameter of this model are non-negative the only way to calibrate the parameters is by using constraint on the parameter. Due to the intuitive nature of the model parameters calibrating them using noisy trajectory data becomes complicated as it is hard to separate the multiple driving regime from the microscopic data alone. The parameter estimates for the IDM model are presented in Table 3. All five parameters are estimated using data points where the approaching speed is positive and vehicles are in deceleration regime. All parameters other than the desired speed are on the boundary imposed upon them. The results indicate that using the residual of acceleration as the minimization function may not be suitable for calibration of the IDM model. Velocity or gap between leader follower have been proven to be more suitable in the literature. The distribution of the residual for this non-linear model is also a contributing factor for the ill poised solution obtained by using a normal distribution.

**Table 3: Parameter Estimates of Intelligent Driver Model**

| **Dataset Used** | **Parameter** | **Initial Value** | **Estimation with parameter Constraints** | | | |
|---|---|---|---|---|---|---|
| | | | Min. | Max. | Maximum-Likelihood Estimate | Least-Squared Estimate |
| Sampled Trajectory | a | 0.5 m/sec$^2$ | 0.5 | 2 | 0.5 | 0.5 |
| | b | 1.5 m/sec$^2$ | 0.5 | 3.5 | 3.5 | 3.5 |
| | $v_0$ | 15 m/sec | 8 | 20 | 18.2387 | 16.0508 |
| | $S_0$ | 2 m | 2 | 10 | 2.0 | 2 |
| | $T^{\alpha}$ | 1.1 Sec | 0.5 | 2 | 0.5 | 0.5 |
| Event Based Dataset | a | 0.5 m/sec$^2$ | 0.5 | 2 | 0.5 | 0.5 |
| | b | 1.5 m/sec$^2$ | 0.5 | 3.5 | 3.5 | 3.5 |
| | $v_0$ | 15 m/sec | 8 | 20 | 17.4518 | 16.6087 |
| | $S_0$ | 2 m | 2 | 10 | 2.0 | 2 |
| | $T^{\alpha}$ | 1.1 Sec | 0.5 | 2 | 0.5 | 0.5 |

## 4.7. SUMMARY

The potential of microscopic traffic data derived from WAMI data for calibration of several car following models is explored. Three separate car following models were chosen representing three different approaches of modeling car following behavior. The GHR model is basically a kinematic model that relates vehicle speed, gap, and relative velocity in a non-linear relationship without imposing any constrain that removes collision in simulated trajectory. The Helly model is also kinematic in nature but its linear and has restriction on minimum gap between a leader follower pair to remove collision in the simulation. The IDM model takes this concept one step further by incorporating intuitive parameters in the model that represent real world driving behavior of drivers. Calibrating these three models using the new dataset provides insight into different modeling approaches as well as the quality of the microscopic data itself. Application of both global least squared method and local maximum likelihood method is incorporated to remove some uncertainty with the calibration approach itself.

Before applying the new WAMI data to calibration, a data quality analysis was undertaken. The data is assessed based on the distribution of the four microscopic traffic flow parameter. These distributions give an overall picture of the diving condition on the roadway under consideration as well as amount of outliers in the data. The data analysis revealed large number of zero velocity points and a small number of outliers. As the data is from a traffic controlled urban arterial the zero velocity may represent both stopped traffic or outliers. As car following models is only applicable for steady state driving condition, some data smoothing and sampling. Data smoothing is conducted to remove some of the outliers and data sampling is conducted to select dataset that most likely to represent car-following condition on the road. A trajectory reconstruction method is applied along with data smoothing to reduce error propagation.

In the calibration work, the driver reaction time is modeled separately and not included as a calibrated parameter. This method of model parameter estimation is consistent with literature (Rahman, Chowdhury et al. 2015). Driver's reaction time is a physiological phenomenon that is not model specific

and including them in model calibration would lead to overfitting of the model. Finding optical reaction time specific to a particular model obfuscate the true descriptive power of the models themselves.

Out of the three models calibrated the parameters of the Helly model are mostly within the expected range compared to the GHR and the IDM model in case of both the dataset. This is due to the fact that this model is linear in nature making it easier to guess the distribution of the residual. This is critical for a maximum-likelihood estimation. In case of least-squared method the linear model provides a smoother surface than a non-linear model. Any future research should include the application of Baysian inverse approach to calibrate the parameters of non-linear car following models.

However, it is clear from this study as well as many calibration studies in the literature that most of the car following models are too restricted to be calibrated using real world noisy trajectory data. An appropriate range of parameter values is still up for debate, solving a calibration problem is not enough as stability of the model using the optimal solution also requires further investigations. Vehicle trajectory data like the WAMI is expected to become more accurate with the progression of imaging and computer vision technology. To apply this kind of dataset for car following analysis, a new generation of models is warranted that are more robust than the current ones.

# 5. CONCLUSIONS AND RECOMMENDATIONS

This exploratory study assessed the use of vehicle trajectory extracted from persistent wide area motion imagery (WAMI) data for traffic data generation and its application in modeling traffic characteristics. As the data represents vehicles with their corresponding tracks, it has the potential to support the calculation of macro and microscopic traffic characteristics for (1) individual vehicles, (2) individual tracked position, (3) individual 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 actual vehicle trajectory was used for further analysis, because the image processing data suffers from false identification, and positional inaccuracy. This required filtering the point and trajectory data that was provided by PVLabs to select only those points that most likely represented vehicles driving on the transportation system. This included developing a multi-dimensional trajectory 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. The filtered tracks were then map-matched to road centerlines 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 1,000 vehicle trajectory 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 stores microscopic characteristics along with vehicle trajectory information, and their relative position on the roadway. The schema developed by Brakatsoulas *et al* (2004) was modified to include the distance between vehicle centroids (as defined by pixels representing the vehicle in the image) of the target vehicle and each of three 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 sub model like the car-following model.

Several car-following were reviewed in detail in terms of their structure, as well as calibration and validation of their parameters. 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 calibration process were applied to three continuous hour period for the portions of tracks that occurred on the two principle one-way arterials, King and Main St, which span the entire view frame and go through the middle of the CBD. The review of the car following models indicated that the kinematic relation that form the base of those model might make them too restrictive for calibration using noisy trajectory data. The most important factor for obtaining good calibration result is to select the appropriate driving regime from microscopic data that facilitate the car following behavior. This is due to the fact that this particular trajectory data represents all driving condition like, stop and go, free flow, and car-following. As the car following models are particular to one regime, the input data for calibration should also represent that regime. In this study, this task is accomplished by means of sampling form the distribution of the microscopic parameters. As both stop-go, and free flow regime is governed by the time and space headway between a leader follower pair, it is used as the first step for sampling. Other microscopic parameters are used to remove some data noise and uncertain driving condition. However, the calibration results were found to be unsatisfactory due to restricted nature of the models when applied with noisy data.

This research illustrated the methods required for using WAMI data to obtain both macroscopic and microscopic traffic data for a long duration of time, wider location, and can cover multiple class of streets and roadway. However, the data is limited by absence of any vehicle classifications information which is necessary for lane changing and gap acceptance behavior analysis. Another issue is the lack of information of any measurement errors. If these issues are addressed, then WAMI data for a complete population of vehicle has the potential for developing new generation of traffic flow models as well as improving the existing ones with robust calibration and validation.

# REFERENCES


Aghabayk, K., S. Moridpour, W. Young, M. Sarvi and Y.-B. Wang (2011). "Comparing Heavy Vehicle and Passenger Car Lane-Changing Maneuvers on Arterial Roads and Freeways." Transportation Research Record: Journal of the Transportation Research Board **2260**(-1): 94-101.

Ahn, S., M. J. Cassidy and J. Laval (2004). "Verification of a simplified car-following theory." Transportation Research Part B: Methodological **38**(5): 431-440.

Alexiadis, V., J. Colyar, J. Halkias, R. Hranac and G. McHale (2004). "The next generation simulation program." Institute of Transportation Engineers. ITE Journal **74**(8): 22.

Aron, M. (1988). "Car following in an urban network: simulation and experiments." PLANNING AND TRANSPORT RESEARCH AND COMPUTATION.

Bando, M., K. Hasebe, A. Nakayama, A. Shibata and Y. Sugiyama (1995). "Dynamical model of traffic congestion and numerical simulation." Physical Review E **51**(2): 1035-1042.

Ben-Akiva, M., A. Davol, T. Toledo, H. N. Koutsopoulos, W. Burghout, I. Andréasson, T. Johansson and C. Lundin (2000). "MITSIMLab for Stockholm Enhancements, Calibration and Validation."

Ben-Akiva, M. E., T. Toledo, H. N. Koutsopoulos and A. Davol (2003). "Calibration and validation of microscopic traffic simulation tools: Stockholm case study." Transportation research record(1831): 65.

Board, T. R. (2013). "Strategic Highway Research Program: Naturalistic Driving Study." Naturalistic Driving Study, Nov 2014.

Brackstone, M. and M. McDonald (1999). "Car-following: a historical review." Transportation Research Part F: Traffic Psychology and Behaviour **2**(4): 181-196.

Brakatsoulas, S., D. Pfoser and N. Tryfona (2004). Modeling, storing and mining moving object databases. Database Engineering and Applications Symposium, 2004. IDEAS '04. Proceedings. International.

Brockfeld, E., R. Kühne and P. Wagner (2004). "Calibration and Validation of Microscopic Traffic Flow Models." Transportation Research Record: Journal of the Transportation Research Board **1876**(-1): 62-70.

Brockfeld, E. and P. Wagner (2006). Validating microscopic traffic flow models. Intelligent Transportation Systems Conference, 2006. ITSC '06. IEEE.

Chandler, R. E., R. Herman and E. W. Montroll (1958). "Traffic dynamics: studies in car following." Operations research **6**(2): 165-184.

Choudhury, C. and M. Ben-Akiva (2008). "Lane Selection Model for Urban Intersections." Transportation Research Record: Journal of the Transportation Research Board **2088**(-1): 167-176.

Choudhury, C., V. Ramanujam and M. Ben-Akiva (2009). "Modeling Acceleration Decisions for Freeway Merges." Transportation Research Record: Journal of the Transportation Research Board **2124**(-1): 45-57.

Choudhury, C. F. (2005). Modeling lane-changing behavior in pesence of exclusive lane M.S. Thesis.

Ciuffo, B., V. Punzo and V. Torrieri (2008). "Comparison of Simulation-Based and Model-Based Calibrations of Traffic-Flow Microsimulation Models." Transportation Research Record: Journal of the Transportation Research Board **2088**(-1): 36-44.

El Najjar, M. E. and P. Bonnifait (2005). "A road-matching method for precise vehicle localization using belief theory and Kalman filtering." Autonomous Robots **19**(2): 173-191.

FHWA. (2008). " NGSIM – Next Generation SIMulation." Retrieved 7/1, 2018, from https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

Fritzsche, H.-T. (1994). "Amodel for traffic simulation."

Gazis, D. C., R. Herman and R. W. Rothery (1961). "Nonlinear follow-the-leader models of traffic flow." Operations Research **9**(4): 545-567.

Gong, H., H. Liu and B.-H. Wang (2008). "An asymmetric full velocity difference car-following model." Physica A: Statistical Mechanics and its Applications **387**(11): 2595-2602.

Green, M. (2000). "" How long does it take to stop?" Methodological analysis of driver perception-brake times." Transportation human factors **2**(3): 195-216.

Greenfeld, J. S. (2002). Matching GPS observations to locations on a digital map. Transportation Research Board 81st Annual Meeting.

Gurusinghe, G., T. Nakatsuji, Y. Azuta, P. Ranjitkar and Y. Tanaboriboon (2002). "Multiple Car-Following Data with Real-Time Kinematic Global Positioning System." Transportation Research Record: Journal of the Transportation Research Board **1802**(-1): 166-180.

Halkias, J. and J. Colyar (2006). Interstate 80 Freeway Dataset. Washington, DC, U.S. Department of Transportation.

Halkias, J. and J. Colyar (2007). Lankershim Boulevard Dataset. Washington, DC, U.S. Department of Transportation.

Halkias, J. and J. Colyar (2007). US Highway 101 Dataset. Washington, DC, U.S. Department of Transportation.

Hanowski, R. J., R. L. Olson, J. S. Hickman and T. A. Dingus (2006). The 100-Car Naturalistic Driving Study: A Descriptive Analysis of Light Vehicle-Heavy Vehicle Interactions from the Light Vehicle Driver's Perspective, Data Analysis Results.

Helbing, D. and B. Tilch (1998). "Generalized force model of traffic dynamics." Physical Review E **58**(1): 133.

Helly, W. (1959). "Simulation of bottlenecks in single-lane traffic flow."

Hoogendoorn, S. and R. Hoogendoorn (2010). "Generic Calibration Framework for Joint Estimation of Car-Following Models by Using Microscopic Data." Transportation Research Record: Journal of the Transportation Research Board **2188**(-1): 37-45.

Hoogendoorn, S., R. Hoogendoorn and W. Daamen (2011). "Wiedemann Revisited." Transportation Research Record: Journal of the Transportation Research Board **2260**(-1): 152-162.

Hoogendoorn, S., S. Ossen and M. Schreuder (2006). "Empirics of Multianticipative Car-Following Behavior." Transportation Research Record: Journal of the Transportation Research Board **1965**(-1): 112-120.

Hoogendoorn, S., H. Van Zuylen, M. Schreuder, B. Gorte and G. Vosselman (2003). "Microscopic traffic data collection by remote sensing." Transportation Research Record: Journal of the Transportation Research Board(1855): 121-128.

Jiang, R., Q. Wu and Z. Zhu (2001). "Full velocity difference model for a car-following theory." Physical Review E **64**(1): 017101.

Kesting, A. and M. Treiber (2008). "Calibrating Car-Following Models by Using Trajectory Data: Methodological Study." Transportation Research Record: Journal of the Transportation Research Board **2088**(-1): 148-156.

Kikuchi, S. and P. Chakroborty (1992). Car-following model based on fuzzy inference system.

Kim, J. and H. Mahmassani (2011). "Correlated Parameters in Driving Behavior Models." Transportation Research Record: Journal of the Transportation Research Board **2249**(-1): 62-77.

Kometani, E. and T. Sasaki (1959). "Dynamic behaviour of traffic with a non-linear spacing-speed relationship."

Marczak, F. and C. Buisson (2012). "New Filtering Method for Trajectory Measurement Errors and Its Comparison with Existing Methods." Transportation Research Record: Journal of the Transportation Research Board **2315**(-1): 35-46.

Michaels, R. (1963). Perceptual factors in car following. Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow (London, England), OECD.

Miwa, T., D. Kiuchi, T. Yamamoto and T. Morikawa (2012). "Development of map matching algorithm for low frequency probe data." Transportation Research Part C: Emerging Technologies **22**: 132-145.

Montanino, M., B. Ciuffo and V. Punzo (2012). Calibration of microscopic traffic flow models against time-series data. Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on.

Montanino, M. and V. Punzo (2013). "Making NGSIM data usable for studies on traffic flow theory: Multistep method for vehicle trajectory reconstruction." Transportation Research Record: Journal of the Transportation Research Board(2390): 99-111.

Monteil, J., R. Billot, J. Sau, C. Buisson and N.-E. El Faouzi (2014). "Calibration, Estimation, and Sampling Issues of Car-Following Parameters." Transportation Research Record: Journal of the Transportation Research Board **2422**(-1): 131-140.

Moridpour, S., M. Sarvi and G. Rose (2010). "Modeling the Lane-Changing Execution of Multiclass Vehicles Under Heavy Traffic Conditions." Transportation Research Record: Journal of the Transportation Research Board **2161**(-1): 11-19.

Ossen, S. and S. Hoogendoorn (2005). "Car-Following Behavior Analysis from Microscopic Trajectory Data." Transportation Research Record: Journal of the Transportation Research Board **1934**(-1): 13-21.

Ossen, S. and S. Hoogendoorn (2008). "Validity of Trajectory-Based Calibration Approach of Car-Following Models in Presence of Measurement Errors." Transportation Research Record: Journal of the Transportation Research Board **2088**(-1): 117-125.

Ossen, S. and S. Hoogendoorn (2009). "Reliability of Parameter Values Estimated Using Trajectory Observations." Transportation Research Record: Journal of the Transportation Research Board **2124**(-1): 36-44.

Ossen, S. and S. P. Hoogendoorn (2011). "Heterogeneity in car-following behavior: Theory and empirics." Transportation Research Part C: Emerging Technologies **19**(2): 182-195.

Ozaki, H. (1993). Reaction and anticipation in the car-following behavior. Proc. of 12th International Symposium on Theory of Traffic Flow and Transportation.

Ozaki, H. (1993). Reaction and Anticipation in the Car Following Behavior. 12th International Symposium on the Theory of Traffic Flow and Transportation, Berkeley, California Elsevier Science Publishers B.V., Amsterdam,.

Panwai, S. and H. Dia (2005). "Comparative evaluation of microscopic car-following behavior." Intelligent Transportation Systems, IEEE Transactions on **6**(3): 314-325.

Peng, G. H. and D. H. Sun (2010). "A dynamical model of car-following with the consideration of the multiple information of preceding cars." Physics Letters A **374**(15–16): 1694-1698.

Pipes, L. A. (1953). "An operational analysis of traffic dynamics." Journal of applied physics **24**(3): 274-281.

Puckrin, E., C. S. Turcotte, M.-A. Gagnon, J. Bastedo, V. Farley and M. Chamberland (2012). Airborne infrared hyperspectral imager for intelligence, surveillance, and reconnaissance applications.

Punzo, V., M. T. Borzacchiello and B. Ciuffo (2011). "On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data." Transportation Research Part C: Emerging Technologies **19**(6): 1243-1262.

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**(-1): 11-24.

Punzo, V., D. Formisano and V. Torrieri (2005). "Nonstationary Kalman Filter for Estimation of Accurate and Consistent Car-Following Data." Transportation Research Record: Journal of the Transportation Research Board **1934**(-1): 1-12.

Punzo, V. and F. Simonelli (2005). "Analysis and Comparison of Microscopic Traffic Flow Models with Real Traffic Microscopic Data." Transportation Research Record: Journal of the Transportation Research Board **1934**(-1): 53-63.

Punzo, V. and A. Tripodi (2007). "Steady-State Solutions and Multiclass Calibration of Gipps Microscopic Traffic Flow Model." Transportation Research Record: Journal of the Transportation Research Board **1999**(-1): 104-114.

Quddus, M. A. (2006). High integrity map matching algorithms for advanced transport telematics applications, Imperial College London, United Kingdom.

Quddus, M. A., W. Y. Ochieng and R. B. Noland (2007). "Current map-matching algorithms for transport applications: State-of-the art and future research directions." Transportation Research Part C: Emerging Technologies **15**(5): 312-328.

Rahman, M. (2013). "Application of parameter estimation and calibration method for car-following models."

Rahman, M., M. Chowdhury, T. Khan and P. Bhavsar (2015). "Improving the Efficacy of Car-Following Models With a New Stochastic Parameter Estimation and Calibration Method." Intelligent Transportation Systems, IEEE Transactions on **16**(5): 2687-2699.

Rahman, M., M. Chowdhury, T. Khan and P. Bhavsar (2015). "Improving the Efficacy of Car-Following Models With a New Stochastic Parameter Estimation and Calibration Method." IEEE Transactions on Intelligent Transportation Systems **16**(5): 2687-2699.

Rahman, M., M. Chowdhury, X. Yuanchang and H. Yiming (2013). "Review of Microscopic Lane-Changing Models and Future Research Opportunities." Intelligent Transportation Systems, IEEE Transactions on **14**(4): 1942-1956.

Rakha, H. and M. Arafeh (2010). "Calibrating Steady-State Traffic Stream and Car-Following Models Using Loop Detector Data." Transportation Science **44**(2): 151-168.

Rakha, H. and B. Crowther (2002). "Comparison of Greenshields, Pipes, and Van Aerde Car-Following and Traffic Stream Models." Transportation Research Record: Journal of the Transportation Research Board **1802**(-1): 248-262.

Rakha, H., C. Pecker and H. Cybis (2007). "Calibration Procedure for Gipps Car-Following Model." Transportation Research Record: Journal of the Transportation Research Board **1999**(-1): 115-127.

Ranjitkar, P., T. Nakatsuji and M. Asano (2004). "Performance Evaluation of Microscopic Traffic Flow Models with Test Track Data." Transportation Research Record: Journal of the Transportation Research Board **1876**(-1): 90-100.

Ranjitkar, P., T. Nakatsuji, Y. Azuta and G. Gurusinghe (2003). "Stability Analysis Based on Instantaneous Driving Behavior Using Car-Following Data." Transportation Research Record: Journal of the Transportation Research Board **1852**(-1): 140-151.

Ranjitkar, P., T. Nakatsuji and A. Kawamura (2005). "Experimental Analysis of Car-Following Dynamics and Traffic Stability." Transportation Research Record: Journal of the Transportation Research Board **1934**(-1): 22-32.

Sangster, J., H. Rakha and J. Du (2013). "Application of Naturalistic Driving Data to Modeling of Driver Car-Following Behavior." Transportation Research Record: Journal of the Transportation Research Board **2390**(-1): 20-33.

Schweitzer, N., Y. Apter, G. Ben-David, D. Liebermann and A. Parush (1995). "A field study on braking responses during driving. II. Minimum driver braking times." Ergonomics **38**(9): 1903-1910.

Skycomp. (2014). "Skycomp Traffic Studies and Data Collection. Available at: http://www.skycomp.com/ " Retrieved Nov 6, 2014.

Tiakas, E., A. N. Papadopoulos, A. Nanopoulos, Y. Manolopoulos, D. Stojanovic and S. Djordjevic-Kajan (2009). "Searching for similar trajectories in spatial networks." Journal of Systems and Software **82**(5): 772-788.

Treiber, M., A. Hennecke and D. Helbing (2000). "Congested traffic states in empirical observations and microscopic simulations." Physical Review E **62**(2): 1805-1824.

Treiber, M. and A. Kesting (2013). "Microscopic Calibration and Validation of Car-Following Models – A Systematic Approach." Procedia - Social and Behavioral Sciences **80**(0): 922-939.

Treiterer, J., and J. A. Myers (1974). The Hysteresis Phenomenon in Traffic Flow. Sixth Symposium on Transportation and Traffic Flow Theory, Berkeley, California Elsevier Science Publishers B.V., Amsterdam,.

Van Aerde, M. (1995). Single regime speed-flow-density relationship for congested and uncongested highways. 74th Annual Meeting of the Transportation Research Board, Washington, DC.

Van Aerde, M. and H. Rakha (1995). Multivariate calibration of single regime speed-flow-density relationships. Proceedings of the 6th 1995 Vehicle Navigation and Information Systems Conference.

Velaga, N. R., M. A. Quddus and A. L. Bristow (2009). "Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems." Transportation Research Part C: Emerging Technologies **17**(6): 672-683.

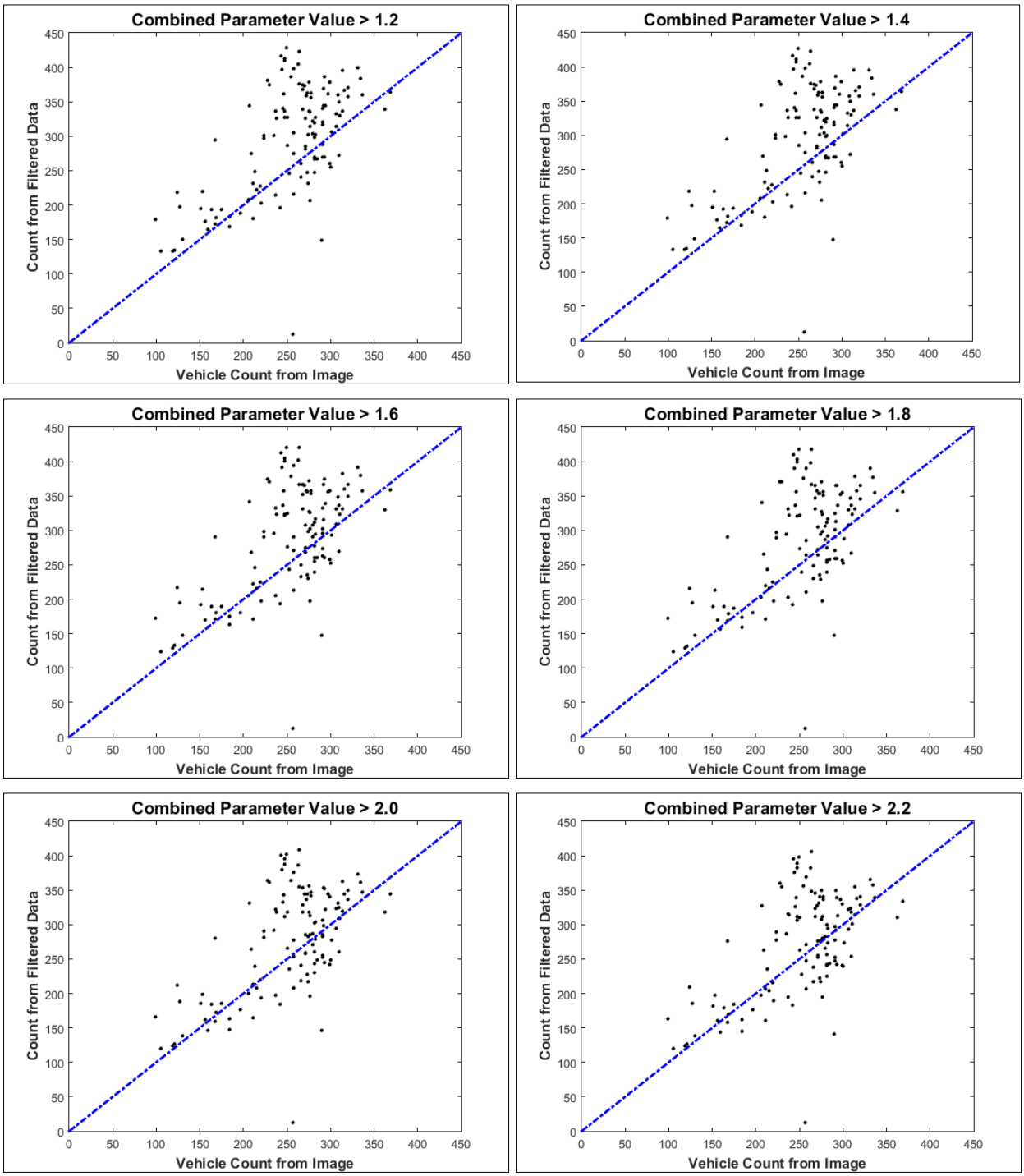
Wan, X., P. Jin, F. Yang, J. Zhang and B. Ran (2014). "Modeling Vehicle Interactions During Merge in Congested Weaving Section of Freeway Ramp." Transportation Research Record: Journal of the Transportation Research Board **2421**(-1): 82-92.

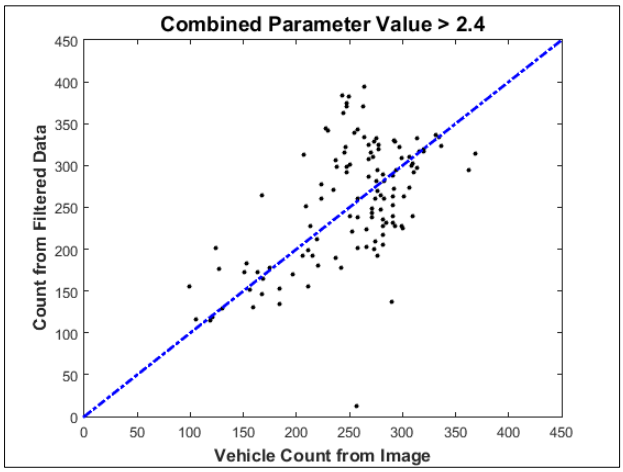
White, C. E., D. Bernstein and A. L. Kornhauser (2000). "Some map matching algorithms for personal navigation assistants." Transportation Research Part C: Emerging Technologies **8**(1–6): 91-108.

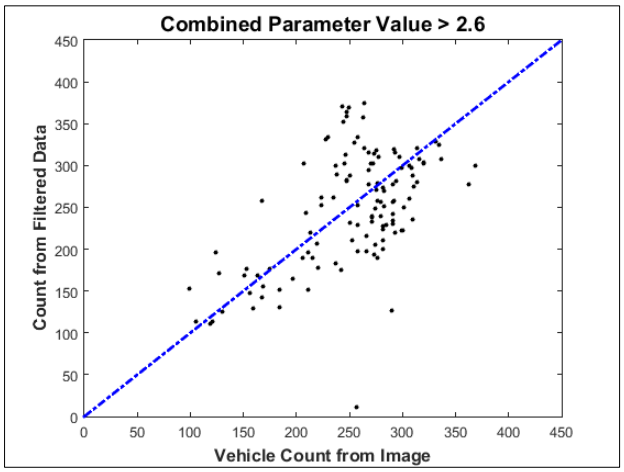
# APPENDICES

# APPENDIX A

Vehicle count plot for different filtering criterion

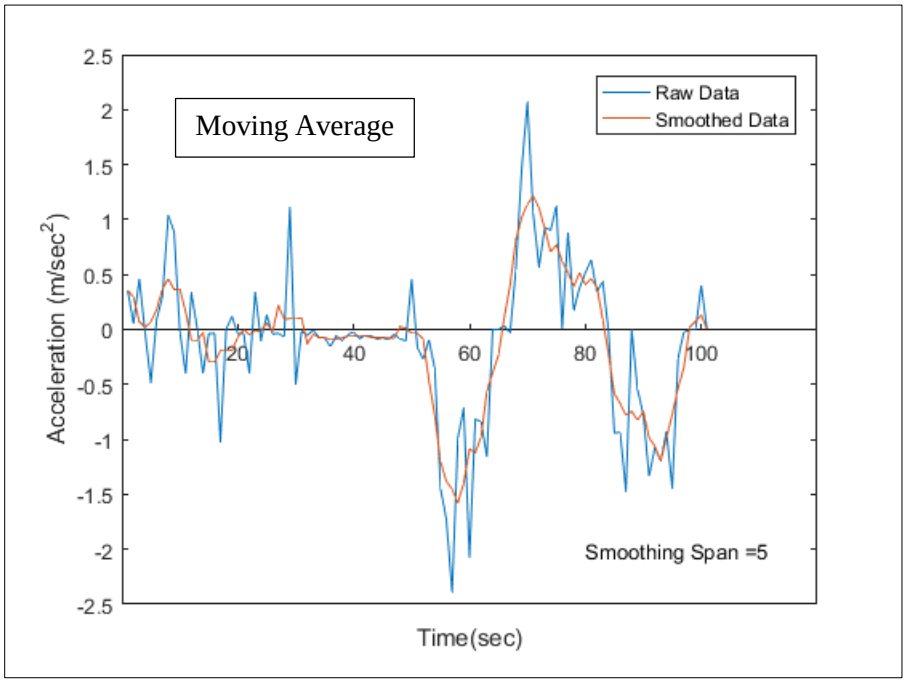


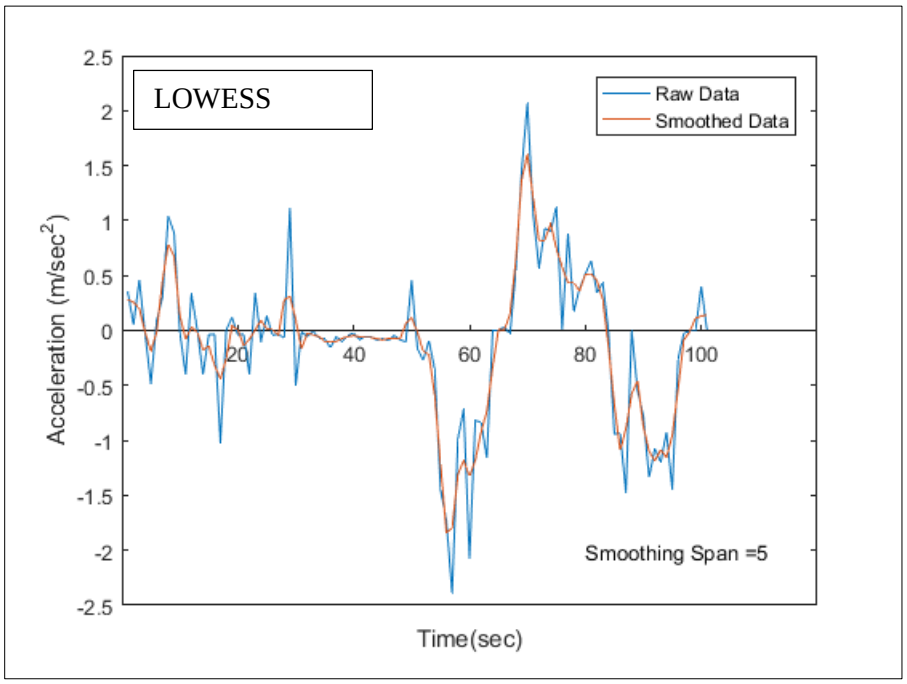

Combined Parameter Value > 2.4
Count from Filtered Data
Vehicle Count from Image
0
50
100
150
200
250
300
350
400
450

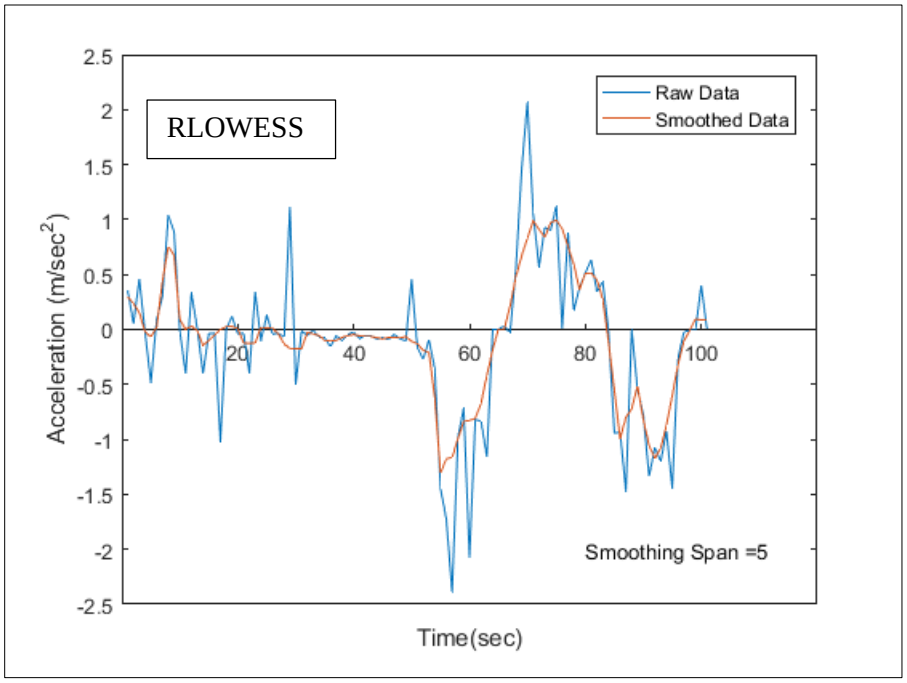

Combined Parameter Value > 2.6
Count from Filtered Data
Vehicle Count from Image
0
50
100
150
200
250
300
350
400
450

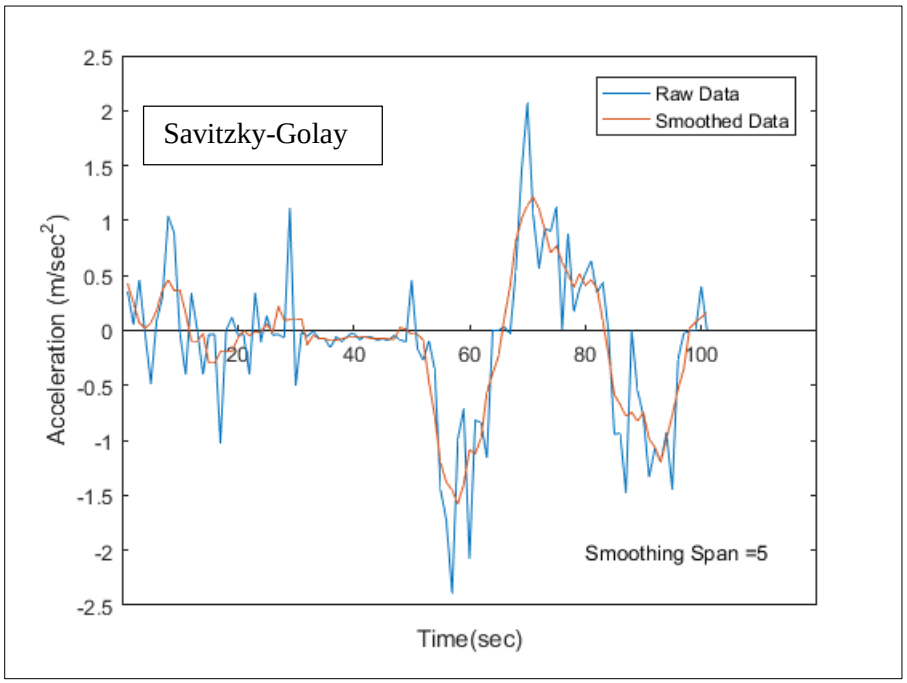
## APPENDIX B

Time series plot of Acceleration for different smoothing technique




RLOWESS
Raw Data
Smoothed Data
Acceleration (m/sec$^2$)
Time(sec)
Smoothing Span =5
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
20
40
60
80
100


Savitzky-Golay
Raw Data
Smoothed Data
Acceleration (m/sec$^2$)
Time(sec)
Smoothing Span =5
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
20
40
60
80
100

## APPENDIX C

Sample data for model calibration for GHR and Helly Model

| Current Acceleration ($m/s^2$) | Current Velocity (m/s) | Current Gap with Leader (m) | Current relative Velocity (m/s) | Acceleration after PRT (m/s2) | Time Headway with the Leader (s) |
|---|---|---|---|---|---|
| 0.253 | 9.8652 | 20.488 | 1.8078 | 0.19596 | 2.0768 |
| -0.3726 | 8.143 | 12.042 | -0.6748 | -0.36819 | 1.4788 |
| 0.0312 | 6.5596 | 18.82 | -1.8552 | 0.026502 | 2.8691 |
| -0.0335 | 12.253 | 23.139 | 0.2023 | 0.060594 | 1.8884 |
| -0.1597 | 10.51 | 18.232 | -1.0612 | -0.10079 | 1.7348 |
| -0.1652 | 10.592 | 13.159 | 0.1259 | -0.16022 | 1.2423 |
| 0.0343 | 10.361 | 13.052 | -0.2353 | 0.13104 | 1.2598 |
| -0.8712 | 12.184 | 19.337 | -1.2096 | -0.91165 | 1.587 |
| 0.0662 | 11.656 | 30.698 | 2.6885 | 0.03679 | 2.6336 |
| 0.25 | 9.2016 | 17.188 | 0.0852 | 0.49545 | 1.8679 |
| 0.2131 | 6.2067 | 18.546 | -1.7067 | 0.12922 | 2.988 |
| 0.0704 | 10.541 | 27.173 | 0.1659 | 0.067879 | 2.5778 |
| -0.4378 | 9.6153 | 17.564 | 0.3063 | -0.23395 | 1.8267 |
| -1.4602 | 10.481 | 14.379 | 3.2968 | -0.90706 | 1.372 |
| -0.6612 | 11.072 | 13.505 | 1.1401 | -0.56191 | 1.2198 |
| 0.2289 | 11.428 | 20.443 | -0.8259 | 0.22708 | 1.7888 |
| 0.0213 | 6.6145 | 15.35 | -0.0385 | -0.00991 | 2.3206 |
| 0.0147 | 8.7174 | 22.357 | -0.4169 | 0.019132 | 2.5646 |
| -0.2141 | 10.188 | 12.543 | 1.262 | -0.06878 | 1.2312 |
| 0.4555 | 6.4181 | 12.015 | 1.254 | 0.46192 | 1.8721 |
| 0 | 11.9 | 22.728 | 0.3241 | 0.017464 | 1.9099 |
| -0.4083 | 12.065 | 24.871 | 1.6967 | -0.35189 | 2.0615 |
| 0.3738 | 9.6005 | 17.285 | 1.5319 | 1.675 | 1.8005 |
| -0.092 | 9.6916 | 15.821 | -0.3125 | -0.22322 | 1.6324 |
| 0.0395 | 10.74 | 19.352 | -5.3773 | -0.06191 | 1.8019 |
| -0.4793 | 12.719 | 14.06 | 1.3952 | -0.46992 | 1.1054 |
| -0.7629 | 5.6174 | 14.89 | 2.4259 | -0.41119 | 2.6507 |
| 0.8837 | 8.2948 | 16.718 | -2.7984 | 1.0027 | 2.0155 |
| 0.0863 | 7.4042 | 14.841 | 0.0859 | -0.25004 | 2.0044 |

## APPENDIX D

Sample data for model calibration for IDM Model

| Current Acceleration ($m/s^2$) | Current Velocity (m/s) | Current Gap with Leader (m) | Current relative Velocity (m/s) | Time Headway with the Leader (s) |
|---|---|---|---|---|
| 0.0455 | 9.8198 | 23.541 | 1.9153 | 2.3973 |
| 0.5647 | 9.2155 | 18.531 | -0.8287 | 2.0108 |
| 0.214 | 10.089 | 17.547 | 0.748 | 1.7392 |
| -0.0052 | 11.243 | 14.371 | 0.8994 | 1.2781 |
| 0.6927 | 10.174 | 30.169 | -1.1338 | 2.9652 |
| -0.6354 | 8.4214 | 21.768 | -0.6596 | 2.5849 |
| -0.0295 | 10.999 | 16.2 | 3.9058 | 1.4729 |
| -0.5694 | 11.066 | 17.688 | 0.4299 | 1.5985 |
| 0.1272 | 11.265 | 13.436 | -0.0556 | 1.1928 |
| -0.3632 | 13.218 | 17.21 | 0.2832 | 1.302 |
| 0.2182 | 5.9067 | 13.807 | 0.8207 | 2.3375 |
| -0.0846 | 8.0564 | 15.588 | 2.1606 | 1.9348 |
| -0.0931 | 10.255 | 12.571 | -0.3405 | 1.2259 |
| -0.3127 | 6.4135 | 17.216 | -1.0156 | 2.6844 |
| 0.0939 | 7.9185 | 12.591 | 0.6002 | 1.59 |
| -1.0188 | 10.645 | 15.014 | -0.3563 | 1.4104 |
| 0.3849 | 11.115 | 26.58 | -0.8491 | 2.3914 |
| 0.168 | 11.35 | 30.202 | -0.947 | 2.6611 |
| -1.0827 | 12.064 | 24.299 | 4.5329 | 2.0141 |
| -1.3732 | 13.24 | 28.688 | 5.3383 | 2.1667 |
| 0.0323 | 10.844 | 20.687 | 0.1316 | 1.9078 |
| 0.322 | 10.712 | 23.708 | -0.4177 | 2.2131 |
| -0.1513 | 12.169 | 20.661 | 0.5555 | 1.6978 |
| 0.3397 | 7.7345 | 19.187 | -1.573 | 2.4807 |
| 0.0142 | 6.2859 | 12.935 | 3.0925 | 2.0577 |
| 0.2507 | 10.49 | 25.489 | -0.43 | 2.4298 |
| -0.0299 | 11.019 | 23.898 | 1.1865 | 2.1687 |
| 0.1801 | 12.993 | 33.486 | -1.007 | 2.5773 |
| 0.4537 | 12.711 | 17.436 | 0.0579 | 1.3717 |