UNIVERSITY OF CALIFORNIA,

IRVINE

Network-wide truck tracking using advanced point detector data

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Kyung Kate Hyun

Dissertation Committee:

Professor Stephen G. Ritchie, Chair

Professor R. Jayakrishnan

Professor Michael G. McNally

2016

DEDICATION

To my parents and sister

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWDEDGEMENTS

This has been a truly rewarding journey. I would like to thank a number of people who have supported me for these graduate years with patience and encouragement.

First, I would like to acknowledge my advisor Professor Stephen Ritchie. From the first day in ITS to the day of defense, he was always supportive and gave me various opportunities in research and teaching. I become more confident and passionate my research under his guidance and am very grateful to his advising environment and approach. I would like to thank my committee members, Professor Jayakrishnan and Professor McNally. From their comments and advices in my defense and classes, I could broaden my sights as a transportation engineer. I would like to express my gratitude for Professor Will Recker and Professor Houston for their critical comments and feedback for my dissertation proposal presentation.

I am truly grateful to my mentor, Dr. Andre Tok. With his insight and knowledge, I was able to be more motivated and enjoying my work, which could strengthen my dissertation work. I would like to thank his valuable research ideas, technical supports, writing tips, and encouragements and cheering wishes before my defense day.

It is fortunate to have supporting friends and fellow colleagues who share every moment together in this long journey. They make this challenging PhD life more pleasant and joyful. I would like to thank Dr. Sarah Hernandez for her cheers, writing supports, and valuable comments she gave me when I encounter obstacles in my research work. I truly enjoyed my graduate life with a lot of lunch discussion and weekend hiking with Kyungsoo Jeong, Ashley Lo, and Dr. Suman Mitra. For the five hours of dry run before my defense, I have to thank my friends Dr. Jared Sun and Youngeun Bae. I would like to thank my research team, Jun Park,

Ethan Sun and Yiqiao Li for their data collection and processing efforts. I would like to thank all my fellows at ITS who had gave me valuable insights and comments, Dr. Iris You, Dr. Jamie Kang, Dr. Jaeyoung Jung, Dr. Jinheon Choi and Dr. Jaehun Kim, Dr. Neda Mashoud, Dr. Vanessa Luong, Dr. Anupam Srivastava, Daisik Nam, Sunghi Ahn, and Kathrin Kuhne. I am thankful to my friends in Korea for their care packages, phone calls, and chats to cheer me up for this achievement.

I was fortunate to share my times with warm and supportive UC Irvine community; especially I am thankful to ITS staffs Myra Radlow, Kathy Riley, Anne Marie Defeo, and Ziggy Bates for their hard work. I am also grateful for the financial support from Caltrans, ARB, UCTC, WTS and KSEA.

I would like to show my gratitude to my dear friend, RG Kim. There is no word to express such a deep gratitude to him. He has been always by my side to give his hands for everything and helped me push through all difficulties during my whole graduate years. I truly thank his kindness and sacrifice, and I only wish I could do the same for him.

With utmost gratefulness, I would like to recognize my parents and sister Jin for their unwavering support and love. Their encouragement and belief is the most valuable power for me to keep moving forward for this achievement. They happily flew more than ten hours to visit me often so that we can share so much time together here in California. Their boundless sacrifice and love in California and Korea always makes me comfortable and manages my life emotionally balanced. I believe my family's support and love will be the most valuable asset in my future journeys.

CURRICULUM VITAE

EDUCATION

Ph.D. in Civil and Environmental Engineering Dec 2016
Institute of Transportation Studies, University of California Irvine, Irvine, CA
Dissertation: "Network-wide truck tracking using advanced point detector data"
Advisor: Dr. Stephen G. Ritchie
GPA: 3.98/4.0

M.S. in Urban Planning and Engineering June 2010
Yonsei University, Seoul, South Korea

B.S. in Urban Planning and Engineering June 2008
Yonsei University, Seoul, South Korea

RESEARCH EXPERIENCE

- **Graduate Student Researcher** Sep 2011 – Dec 2016
  Civil and Environmental Engineering Department, University of California, Irvine
  - Truck Count Study, California Department of Transportation (Caltrans)
  - Development of a New Methodology to Characterize Truck Body Types along California Freeways, California Air Resources Board
  - California Statewide Freight Forecasting Model Development, California Department of Transportation (Caltrans)
- **Graduate Student Researcher** July 2008 – June 2010
  Urban Planning and Engineering, Yonsei University, South Korea
  - An Evaluation of an Intelligent Transportation System and Urban Traffic Information System
  - Development of evaluation indices for public transit system with operators' views
  - Installation of an Intelligent Transportation System (ITS) in Seoul

PUBLICATIONS

- **Kyung (Kate) Hyun,** Andre Tok, and Stephen G. Ritchie (2016), Long Distance Truck Tracking from Advanced Point Detectors using a Selective Weighted Bayesian Model, *Journal of Transportation Research Part C, Under Review*
- **Kyung (Kate) Hyun**, Kyungsoo Jeong, and Stephen G. Ritchie (2016), Assessing crash risks considering vehicle interactions with trucks using point detector data, *Journal of Accident Analysis and Prevention, Under review*
- **Kyung (Kate) Hyun,** and Stephen G. Ritchie (2017), A sensor location decision model for truck flow measurement, *Accepted for the Journal of the Transportation Research Board*
- Andre Tok, **Kyung (Kate) Hyun,** Sarah Hernandez, Kyungsoo Jeong, Yue Sun, Craig Rindt, and Stephen G. Ritchie (2017), Truck activity monitoring system (TAMS) for freight transportation

analysis, *Accepted for Proceedings of Annual Meeting of the Transportation Research Board, Washington, D.C.*

- Sarah Hernandez and **Kyung (Kate) Hyun** (2017), Truck weight distributions at traffic count sites using WIM and GPS data, *Accepted for Proceedings of the Annual Meeting of the Transportation Research Board, Washington, D.C.*
- Kyungsoo Jeong, **Kyung (Kate) Hyun**, and Stephen G. Ritchie (2017), Influence of personal concerns about travel on travel behavior, *Accepted for Proceedings of the Annual Meeting of the Transportation Research Board, Washington, D.C.*
- Daisik Nam, **Kyung (Kate) Hyun**, Hyunmyung Kim, Kijung Ahn, R., Jayakrishnan (2016), Grid Cell Based Taxi Ridership Analysis Using Large Scale GPS Data, Model, *To be published in the Journal of the Transportation Research Board*
- **Kyung (Kate) Hyun**, Sarah Hernandez, Andre Tok, and Stephen G. Ritchie (2015), Truck body configuration volume and weight distribution: Estimation by using Weigh-in-Motion data, *Journal of Transportation Research Board* Issue 2478, pp.103-112
- Jun Lee, Hansun Jo, **Kyung (Kate) Hyun** and Jinhyuk Jung (2009) The effect analysis of one-side walking behavior using MDPM (Multi-directional Pedestrian Model), *Journal of the Korea Institute of Intelligent Transport Systems*, Vol.8, No.5, pp 151-159.

PRESENTATIONS

- **Kyung (Kate) Hyun,** Andre Tok, and Stephen G. Ritchie (2016), Long Distance Truck Tracking from Advanced Point Detectors using a Selective Weighted Bayesian Model, *95th Annual Meeting of the Transportation Research Board, Washington, D.C., and University of California Transportation Center Student (UCTC) Conference, UC Riverside*
- Kyungsoo Jeong, **Kyung (Kate) Hyun**, and Stephen G. Ritchie (2016), Relationship between travel concerns and activity using the 2009 National Household Travel Survey, *95th Annual Meeting of the Transportation Research Board, Washington, D.C.*
- **Kyung (Kate) Hyun,** Sarah Hernandez, Andre Tok, and Stephen G. Ritchie (2014), Estimating Truck Volumes by Body Configuration using Weigh-in-Motion Data, *2014 UCTC Student Conference, Cal Poly Pomona (Best Overall Poster Award)*
- **Kyung (Kate) Hyun**, Sarah Hernandez, Andre Tok, and Stephen G. Ritchie (2015), Truck body configuration volume and weight distribution: Estimation by using Weigh-in-Motion data, *94th Annual Meeting of the Transportation Research Board, Washington, D.C.*
- **Kyung (Kate) Hyun** and Bongsoo Son (2010). Development of bus service reliability index from the perspective of operators, *Journal of Korean Society of Transportation.*
- Jun Lee, Hansun Jo, **Kyung Hyun (Kate)** and Jinhyuk Jung (2008). The method of the signal control using time-sliced OD destination, *Journal of Korean Society of Transportation.*

TEACHING EXPERIENCE

- CEE 121: Highway Design and Analysis, UC Irvine — Fall 2015 & Fall 2016
- CEE 295: Student Seminar, UC Irvine — Fall 2012

- Graduate and Undergraduate Student Research Mentor, UC Irvine 2013-2016
- Transportation Planning, Environment and Transportation, and Urban Planning and Engineering, Yonsei University, Seoul, South Korea 2008-2010

HONORS

- Best poster award, UKC student research presentation 2016
- Best presentation award, KSEA-SWRC student research presentation 2016
- Graduate Scholarship, KSEA-KUSCO 2016
- First Place Overall Excellence in poster presentation, University of California Transportation Center (UCTC) Student Conference 2014
- Graduate Scholarship, Women's Transportation Seminar (WTS) 2013
- Best Paper Award, Journal of Korean Society of Transportation, South Korea 2008&2009
- Graduate and Undergraduate Scholarship, Yonsei University, Seoul, South Korea 2005-2010

PROFESSIONAL TRAINING

- TA Professional Development Training (TAPDP) Workshop, UC Irvine 2015
- Women's Transportation Seminar (WTS) Academy 2015

PROFESSIONAL SERVICE

- Serve as a reviewer: Transportation Research Board(TRB) 2015-2016
- Serve as a Secretary of General Affair, Korean Transportation Association in America (KOTAA) 2013-2016
- PATENT: Co-inventor of Korean Domestic Patent on system for estimating public transportation system
- LICENSE: Engineer Transportation, Korea

TECHNICAL SKILLS

- Programming: Python, C#
- Mathematical and statistical tools: R, MATLAB, Excel
- Transportation modeling : TransCAD, AutoCAD, GIS, Paramics, DYNASMART-P
- Database : PostgreSQL

ABSTRACT OF THE DISSERTATION

Network-wide truck tracking using advanced point detector data

By

Kyung Kate Hyun

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2016

Professor Stephen G. Ritchie, Chair

Trucks contribute disproportionally to traffic congestion, emissions, road safety issues, and infrastructure and maintenance costs. In addition, truck flow patterns are known to vary by season and time-of-day as trucks serve different industries and facilities. Therefore, truck flow data are critical for transportation planning, freight modeling, and highway infrastructure design and operations. However, the current data sources only provide partial truck flow or point observations. This dissertation developed a framework for estimating path flows of trucks by tracking individual vehicles as they traverse detector stations over long distances. Truck physical attributes and inductive waveform signatures were collected from advanced point detector systems and used to match vehicles between detector locations by a Selective Weighted Bayesian Model (SWBM). The key feature variables that were the most influential in distinguishing vehicles were identified and emphasized in the SWBM to efficiently and successfully track vehicles across road networks.

The initial results showed that the Bayesian approach with the full integration of two complementary detector data types – advanced inductive loop detectors and Weigh-in-Motion

(WIM) sensors – could successfully track trucks over long distances (i.e., 26 miles) by minimizing the impacts of measurement variations and errors from the detection systems. The network implementation of the model demonstrated high coverage and accuracy, which affirmed the capability of the tracking approach to provide comprehensive truck travel patterns in a complex network. Specifically, the model was able to successfully match 90 percent of multi-unit trucks where only 67 percent of trucks observed at a downstream site passed an upstream detection site.

A strategic plan to identify optimal sensor locations to maximize benefits from the truck tracking model was also proposed. A decision model that optimally locates sensors to capture the maximum truck OD and route flow was investigated using a goal programming approach. This approach suggested optimal locations for tracking implementation in a large truck network considering a limited budget. Results showed that sensor locations from a maximum-flow-capturing approach were more advantageous to observe truck flow than a conventional sensor location approach that focuses on OD and route identifiability.

1. Introduction

Trucks contribute disproportionally to traffic congestion, emissions, road safety issues, and infrastructure and maintenance costs. Therefore, timely and reliable truck flow data are of importance for transportation planning and investment analysis, traffic management, environmental and safety analyses, and operation and maintenance of infrastructure. In 2012, the most recent surface transportation authorization act, the Moving Ahead for Progress in the 21st Century Act (MAP-21) (FHWA, 2013), was enacted with the establishment of the national freight strategic plan (NFSP) (US DOT, 2016). The NFSP specifically focuses on increasing competitiveness and efficiency of freight movement. Main strategies of these programs are: i) reducing congestion to improve performance of the freight transportation system, ii) identifying and monitoring major freight corridors to support congestion mitigation and facilitate freight flows, iii) developing data collection and analytic tools for freight modeling that assists public and private sectors decision-making process. However, insufficient information on freight corridors, facilities, and movements creates a number of obstacles in freight planning and support. For example, incorrect information on truck activity on roadways often leads to inadequate restrictions or displacements of freight activities (US DOT, 2016). Due to the lack of data on freight movements, there have been difficulties in making accurate long-term forecasts of freight demands and facilities.

This dissertation develops a framework for estimating truck flow data utilizing advanced point detector systems. This study seeks solutions to identify primary truck routes and provide path flows of trucks along the truck corridors. An initial research was performed to estimate path flows along the same freeway corridor by tracking individual vehicles as they traverse

advanced point detector stations — Inductive loop detectors (ILD) and Weigh in motion (WIM). To capture dynamic truck activity in a more complex road network, the link-based tracking approach was further developed to enable network-wide tracking across multiple detector stations (ILDs) along different freeways. Using tracked vehicles, various applications were introduced including travel time estimation, detection calibration, and truck monitoring. To maximize benefits from the truck tracking model using point detector systems, a strategic plan that identifies the optimal sensor location was investigated and analyzed the impacts of sensor locations on OD and route flow measurements.

One of the main benefits of this study is leveraging existing point detection systems to anonymously track individual vehicles. Since ILDs collect temporally continuous real-time traffic data for the truck population, the ILD-based tracking framework facilitates the understanding of spatial and temporal truck flow patterns. Even though new technologies such as GPS, RFID tags and AVI systems have been used for traffic monitoring and performance measures, the proposed vehicle tracking approach has critical advantages in path flow estimations. While GPS and AVI systems only collect sample populations from the vehicles that are equipped with tracking devices, the proposed solutions can collect path flow data from the total truck population without any privacy concerns. Along with path information, detailed truck attributes such as axle configuration and truck type can be also obtained in the tracking process.

1.1 Key trends in truck movement

The vehicle miles traveled (VMT) of trucks have increased substantially in the US, with the VMT of multi-unit and single-unit trucks rising 365% and 282%, respectively, between 1970 and 2011 (Research and Innovative Technology Administration, 2011). Not surprisingly, increased truck movements have led to higher fuel and energy consumption despite technological advancements in drivetrain technology. It has been estimated that trucks are responsible for 25 percent of total fuel and energy consumption among highway transportation modes in the US (Research and Innovative Technology Administration, 2011).

According to NFSP (US DOT, 2016), the US population is expected to increase to 389 million by 2045 compared to 321 million in 2015, with economic growth doubled in size. Consequently, freight movements are expected to increase approximately by 42 percent by the year of 2040, which is equivalent to roughly 1.3 percent increase per year. Among various modes in freight transportation, trucks show the largest expected increase in flows by 2040 since they handle the most ton-miles in the US. Increasing freight demand will consequently yield substantial negative impacts on road networks. NFSP (US DOT, 2016) reported that if there is no capacity changes, truck and passenger vehicle traffic will increase peak-period congestion by 34 percent in 2040, compared to 10 percent in 2011. This increased traffic would slow vehicles nearly 30,000 miles and create stop-and-go conditions on an addition 46,000 miles in the US.

Trucks significantly contribute greenhouse gas emissions (GHG). As shown in Figure 1.1, trucks emit a much larger amount of carbon dioxide compared to other freight modes such as rail, waterborne transportation, and air cargo. Therefore, the increased truck flow will bring

considerable localized negative environmental impacts especially in neighborhoods adjacent to truck corridors and facilities.

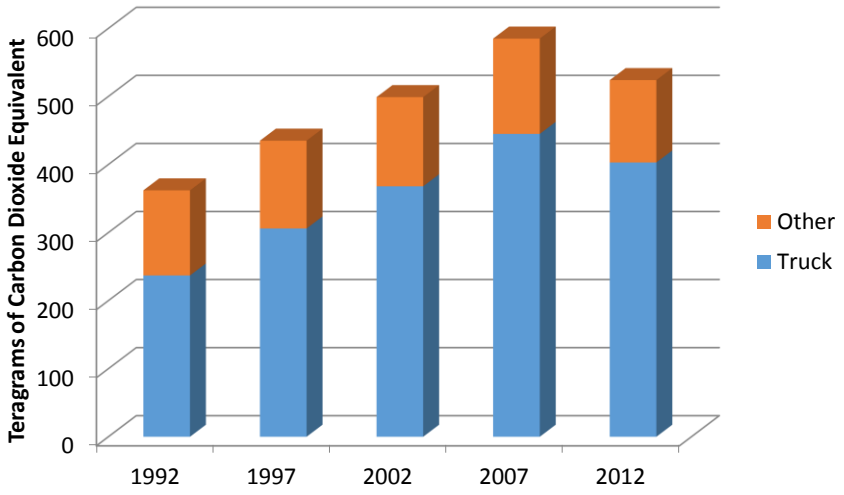


Figure 1.1 Freight transportation GHGs emissions from truck and other modes[1]

1.2 Limitation on current data sources

Truck traffic varies by location and time. Typically, passenger cars and local service trucks show heavy traffic during the morning and afternoon peak hours but very low during the night time. On the other hand, through trucks are reported to show constant traffic throughout the 24 hours, seven days a week (US DOT, 2016; Hallenback, 1997; Ogden, 1991).

Figure 1.2 presents traffic patterns of trucks and passenger vehicles by Day of Week (DOW) and Time of Day (TOD) at two different locations: (a) port of Long Beach area and (b) urban area on interstate highway I-5, in Los Angeles, California. As is known, passenger vehicle traffic shows two peaks at morning and evening hours at both locations during the weekdays. However, on the weekend, traffic volume is higher during the afternoon between 12PM to 6PM.

[1] Source : Beyond traffic (US DOT, National freight strategic plan)

Even though the traffic volumes were much higher at port area, the passenger vehicles showed similar travel patterns at two locations. However, trucks near port and urban areas have quite distinct travel patterns. In the port area, we observed three peaks that correspond to morning and afternoon peak during the day-time port operation hours, and night-time operation hours (PierPass program, 2016). Since the traffic was observed near the ports, truck travel patterns are expected to follow port operation hours. In addition, low weekend traffic could be because the port is not operated on the weekends. However, in the urban area, truck traffic has similar patterns for both weekdays and weekends where high volume was observed during the early morning (5AM-8AM), day time (10AM-2PM), and night time (8PM-10PM) although the weekends have lower volumes than the weekdays. This shows that the truck traffic has distinct spatial and temporal patterns which are closely related to their service industry and facility.

Trucks also show different monthly and seasonal trends compared to passenger vehicles. While some trucks show constant travel patterns throughout the year, others show higher volume at particular season, such as the harvest season for agricultural trucks and the holiday season for import and export containers. Directional variations in truck volume are also reported in several studies (Tok et al., 2016; US DOT, 2016).

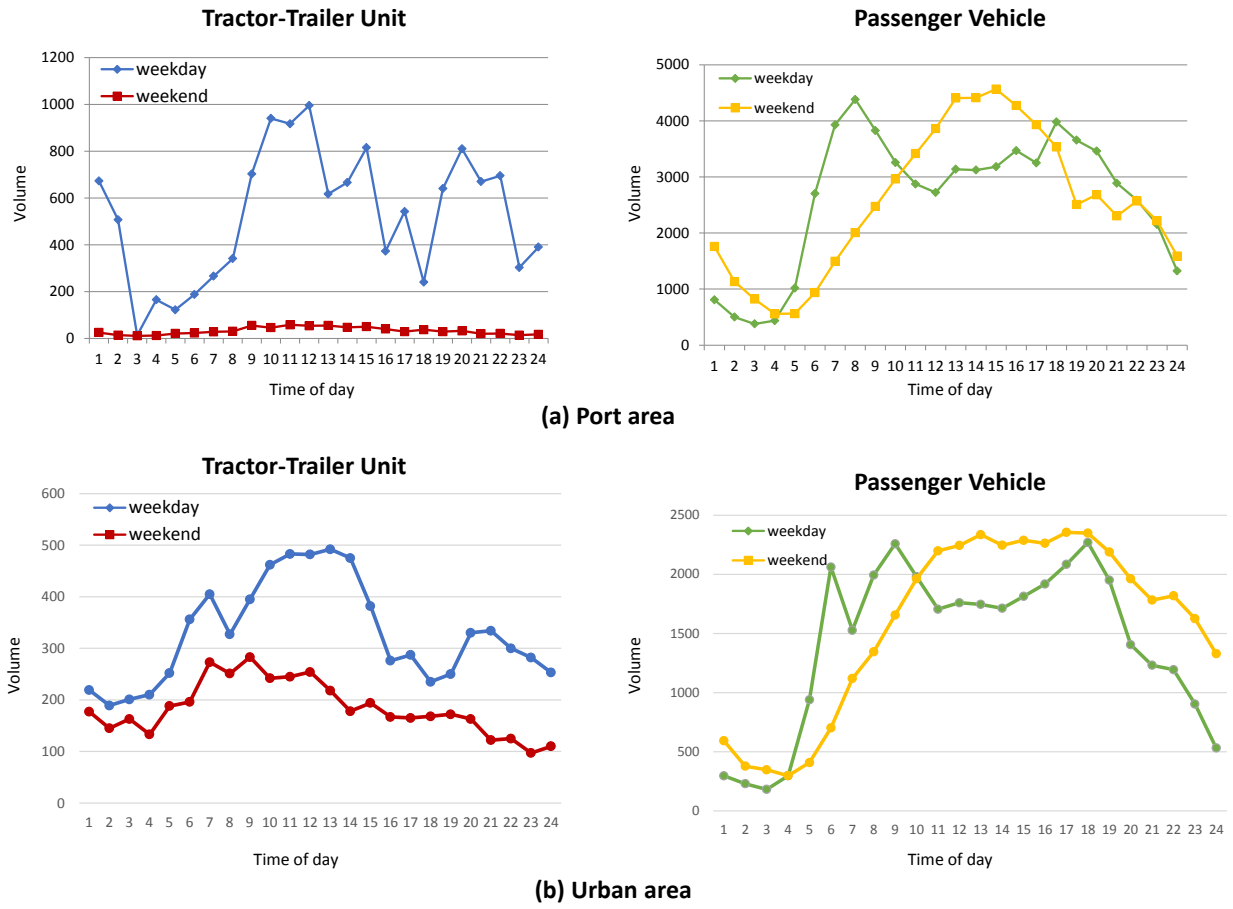


Figure 1.2 TOD and DOW traffic patterns of car and trucks at (a) port and (b) urban area

Hence, it is important to capture such variations in truck flow. However, limited availability of truck data makes it difficult to understand truck activity and movement. The main sources of truck data are truck surveys usually either from non-permanent limited duration studies, or truck counts at permanent facilities such as automated count stations and weigh-in-motion (WIM) sites, or temporary installed active sensor technology. The representative types and sources of the current truck data are summarized in Table 1.1.

Table 1.1 Truck data sources

Type	Source	Advantage	Limitation
Survey	• National level survey (Vehicle inventory and use survey (VIUS)) • State level intercept survey	• Detailed information on OD, truck category, weight, and VMT • Extensive data collection throughout the US or by State	• Obtain partial data from sampled population • Inaccurate responses and biased survey sampling • Limited spatial and temporal data collection span
Passive sensor technology	• Weigh in Motion (WIM) • Inductive Loop Detector (ILD)	• Obtain population data • Potential in obtaining truck flow and body type data with an additional modeling effort	• Obtain point observation • Measurement error from sensor calibration and sensitivity issues
Active sensor technology	• Automatic Vehicle Identification (AVI) • Electronic tolling • GPS and Bluetooth	• Flow data potentially with detailed truck information without additional modeling	• Typically short term observation for sampled population • Costly and privacy concerns

Surveys typically provide partial data obtained from sampled populations. The Vehicle Inventory and Use Survey (VIUS) (US Census Bureau, 2004) is the most common source used for truck activity data since it provides extensive information, including truck body configuration, average weight, and annual truck miles. However, these data are associated with significant limitations in their sample population. For one, VIUS cannot identify truck population statistics at the state level due to discrepancies in how the survey captures in-state and out of state trucks traveling in each state. In specific, trucks operating on California routes may be registered in California or in any other state. However, only trucks that have a home base and indicated their home base as California would have their vehicle miles traveled (VMT) reported for California. By definition, trucks with a home base perform at least 50 percent of their travel within their home base state. As a consequence, there is an inherent bias of under-sampling long-haul interstate truck activity at the state level. Furthermore, VIUS only captures trucks

owned or operated by carriers. Hence, intermodal containers and chassis typically owned by shippers are not included in the data. This is a concern for California, in particular, since it is home to three major US ports. The absence of intermodal container movement in the survey would introduce a significant bias in representing goods movement activity in the state.

Alternatives to VIUS include state level truck surveys, state or international truck registration records, state level truck intercept surveys (Lutsey, 2008), the International Registration Plan (IRP) (IRP, 2016), and the International Fuel Tax Agreements (IFTA) (IFTA, 2016). However, none of these commercial vehicle survey data provide temporal variations in truck path flow because of their limited temporal data collection spans.

Passive data collection through traffic sensors such as Weigh-in-Motion (WIM) and inductive loop detector (ILD) is able to provide permanent point observations of truck population. WIM systems provide truck volume, GVW, axle spacing and weight, axle count, vehicle length, and speed, from which axle based classification can be predicted (Lu et al., 2002). ILDs measure aggregate traffic volumes and occupancies. Through additional modeling efforts, conventional ILDs in double-loop configuration can be used to classify vehicles into length-based categories (Coifman et al., 2007); however, this coarse level of detail at point observation is insufficient to meet advanced freight modeling needs. Many researchers have utilized the point detection systems to obtain additional traffic information such as travel time and path flow. This approach is referred to as vehicle re-identification or vehicle tracking since individual or platoon of vehicles are tracked at point detector location. This approach obtains a greater level of details than that obtainable from simple extrapolations of point observations. The details and limitations of these studies will be discussed in the next chapter.

Along with increased implementation of Intelligent Transportation Systems (ITS) technologies, various types of path flow sensors have been applied to obtain truck flows (Srour, F., and Newton, D., 2006). Vehicle-identification sensors such as automatic vehicle identification (AVI) systems, electronic tolling technology, and license plate scanning are some of the widely utilized technologies that can measure traffic flows. The vehicle-identification sensors identify vehicle id and track vehicles between sensor locations. In the AVI system, for example, vehicles equipped with transponders communicate with AVI reader stations located in transportation networks. However, sampling bias is the main shortcoming of these sources, since travel information was only collected from a small fraction of vehicles that were equipped with electronic tags or captured by detection systems (Dion and Rakha, 2006).

Global Positioning System (GPS) is also capable of providing route choice, origin-destination and travel time. Even though the GPS monitors and collects vehicle path flow, the information can be obtained from only a small portion of vehicle, which would result in biased estimates in traffic flow or travel time. Bluetooth has similar capabilities and limitations in collecting truck information as the vehicles that connected to Bluetooth devices can be collected as samples.

In general, truck travel behavior is more complex than passenger vehicle travel since trucks serve different industries, facilities and commodity types. Hence, truck travel can vary significantly by season and time-of-day as mentioned earlier. Because of this complex behavior, path flow information from the total population would provide significant insight compared with point-based observations or sampled populations for freight modeling, highway infrastructure design and operation, capacity and level of service analysis, and energy and environmental impact analysis.

1.3 Objectives of proposed study

This study presents a novel framework that aims to estimate path flows of trucks across a region utilizing advanced point detector systems. Physical features of individual trucks were extracted from loop detectors and WIM systems. The truck features integrated in a Bayesian probabilistic model were used to match vehicles between detection locations. In addition, key features – ones that were the most influential in distinguishing matched vehicles – were further emphasized in the algorithm to improve matching accuracy over long distances. Figure 1.3 illustrates an overview of the tracking framework categorized into two phases: tracking model development and model implementation. Model development includes tracking feature processing and Bayesian modeling development; and model implementation describes detailed steps of vehicle tracking at detection locations.

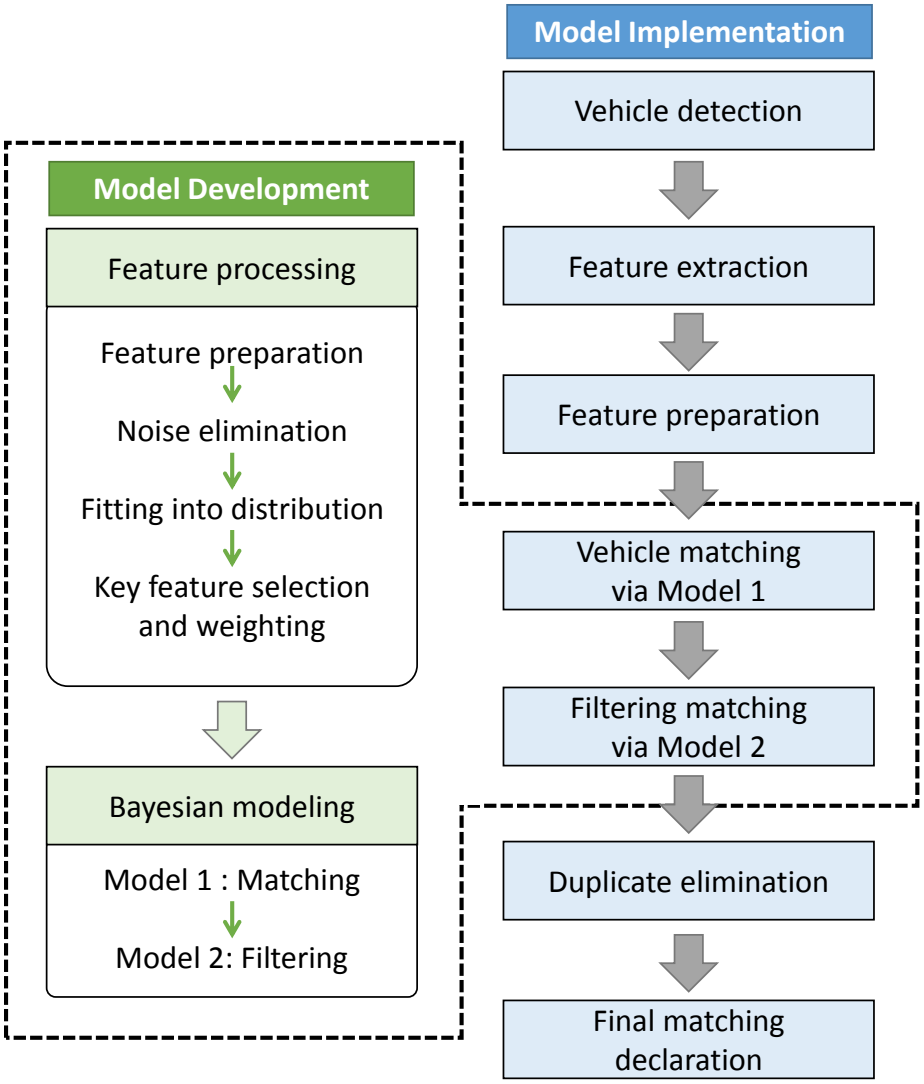


Figure 1.3 Flow chart of the tracking framework

Feature processing starts with a feature preparation. Vehicles' inductive signatures and physical attributes are extracted as vehicle matching features. Considering measurement errors from detection systems, feature noise elimination step is performed prior to fitting the features into probabilistic distributions. Separately, features with more discerning ability in distinguishing vehicles are selected and weighted in a Bayesian model. Since the tracking is

implemented over a long distance, some vehicles detected at a downstream location may not pass at an upstream location, which refers to missing vehicles in this study. To handle these missing vehicles, the tracking algorithm consists of two models, one for vehicle pair matching and another for missing pair filtering. The matching model finds potential matched vehicles at upstream for all detected trucks at the downstream location. The filtering model identifies potential missing vehicles among the matched pairs obtained from the first model. Any duplicate vehicles are eliminated in the last step of the tracking process.

This algorithm was initiated at a corridor level with one downstream and one upstream location. A full integration of the two advanced detection technologies – advanced inductive loop detectors and WIM sensors – at existing WIM sites were applied to the corridor level truck tracking. These two technologies collect complementary vehicle attributes, which allow vehicles to be more accurately and effectively identified when both systems are simultaneously utilized. A linear data fusion method was applied to the WIM and signature data to improve matching accuracy. This method allows both detection systems to complement each other by reducing impacts from errors in the measurements.

It should be noted that the challenge of vehicle tracking goes beyond just distinguishing different truck types, but further requires successful matching of trucks even within a candidate pool of ones that share similar physical configurations. Hence, the use of features with the selective weighting technique is ideal for capturing the salient differences in truck populations especially the trucks with the same body configurations.

The corridor level tracking was further developed to perform a network-wide tracking that captures dynamic truck activity in a more complex road network. This extension proposes

truck tracking across multiple detector stations along different routes utilizing inductive loop detector (ILD) infrastructure. To account for additional complexities in a larger spatial scope uch as varying distances and traffic states between tracking sites in the vehicle matching process, supplementary data sources are considered. These supplementary data sources include historical GPS trajectories which contain truck path flow data and body configuration estimates at all detection locations. Multiple scenarios are proposed with different combinations of the data sources. Bayesian updating approach was applied to integrate different data sources are t matching performances were compared by scenarios.

To accurately track trucks at point detector locations, it is important to determine the optimal sensor locations. Even though there have been considerable efforts devoted to determining optimal sensor locations to measure or estimate accurate traffic flow, general traffic OD or route identification was the main focus. However, considering the long travel distances of trucks, sensor locations that seek observability of ODs and routes may not acquire meaningful proportions of truck movements in a large-sized truck network. Therefore, this dissertation provides a decision model that optimally locates sensors to capture the maximum truck OD and route flow. This approach allows the tracking model to maximize its benefits with optimally locating sensors considering the primal interests of truck flows. Goal programming approaches with different weights that prioritize ODs or route flow were investigated. The proposed model was implemented in a real network in Los Angeles, California with actual truck flow data obtained from sampled truck GPS trajectories. The proposed approach provides optimal location solutions for not only point sensors with an integration of the vehicle tracking model but also alternative active sensors that already have vehicle tracking capability. Figure 1.4 describes an overall flow of the sensor location problem.

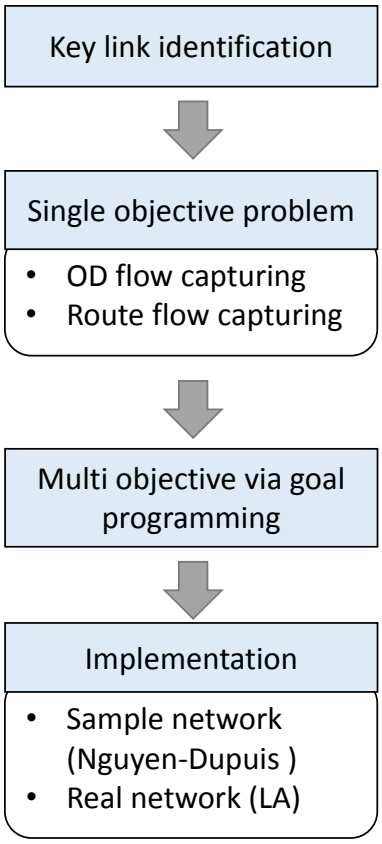


Figure 1.4 Flow chart of the sensor location problem

1.4 Organization

This dissertation consists of seven chapters.

Chapter 1 introduces key trends in truck movement and objectives of this dissertation.

Chapter 2 presents background of detection systems including ILD and WIM technologies and summarizes previous studies on vehicle tracking.

Chapter 3 demonstrates modeling approaches of the truck tracking framework. Bayesian classifier is first introduced and an extended form of selected and weighted Bayesian modeling (SWBM) is further discussed. Feature preparation such as noise elimination, key feature

selection and weighting methods are described. In addition, a step-by-step procedure of tracking algorithm is described to provide field implementation.

Chapter 4 introduces the corridor level tracking with the full integration of WIM and ILD systems using a linear data fusion method. Results from the Selective Weighted Bayesian Model (SWBM) are presented by truck categories (i.e., single unit and tractor-trailer units) with different linear fusion weights. The proposed modeling is compared to the previous approaches with the same tracking dataset. In addition, three applications on the corridor level tracking are introduced including travel time estimation, WIM calibration, and detailed truck monitoring with an integration of a body classification model.

Chapter 5 provides results of the network-wide tracking model using ILD technologies. Truck detection algorithm is introduced to identify vehicle types and validated with over 28,000 vehicles collected in California. In addition, three supplementary data sets, which correspond to GPS, truck body configuration estimates, and travel time, are integrated in tracking process using linear data fusion method. Results are compared with different linear fusion weights by truck types. As applications of the network-wide tracking, travel time estimation and truck monitoring results implemented over a larger network system are introduced.

Chapter 6 introduces the optimal sensor location problem. A decision model that optimally locates sensors to capture the maximum OD and route flow is investigated using a goal programming approach. The proposed model is implemented with a network of Los Angeles, California with actual truck flow data obtained from the sampled truck GPS trajectories.

Chapter 7 provides conclusion remarks. Contribution of this study and future studies are presented.

2. Background

Advanced point detection technologies of Weigh in Motion (WIM) and inductive loop detector (ILD) signature are the main data sources for truck tracking. Since WIM collects physical attributes of vehicles and loop records metallic inductance when vehicle passes over the sensor, different types of vehicle attributes are collected from these two detection systems. This chapter first introduces the WIM and ILD technologies and compares vehicle attributes collected from the detection systems.

Using the advanced point detection systems or other data sources such as active sensors, studies on vehicle tracking has been performed since 1990s. However, the main focus in previous studies was general traffic or passenger vehicles, and only recently more attention was given to commercial vehicles. This chapter introduces the general traffic tracking studies into two categories: (i) general vehicle tracking utilizing inductive signature technologies, (ii) general vehicle tracking utilizing other technologies. In addition, limitations on previous truck tracking studies are discussed with key literatures.

2.1 Weigh in Motion (WIM)

WIM devices have been used since the 1980s to collect data for truck routing, pavement management and design, weight enforcement, traffic safety, and transportation policy (Nichols, A., and D. M. Bullock, 2004). WIM systems initially introduced to improve operational efficiency of traditional static weigh stations (Lu et al., 2002). While trucks are required to stop to be weighted at the traditional weigh station, the WIM records instantaneous dynamic axle loads and spacing, number of axles, vehicle speed, lane, and time stamp with full a speed when a

truck traverses the sensors (FHWA, 2013). Detailed information recorded at WIM is described in Table 2.1.

Table 2.1 WIM records

Field	Data Type By Field	Field	Data Type By Field
1	Lane	16	Axle 2 Right Side weight (kips)
2	Month	17	Axle 2 Left Side weight (kips)
3	Day	18	Spacing between Axles 1 and 2 (feet)
4	Year	19	Axle 3 Right Side weight (kips)
5	Hour	20	Axle 3 Left Side weight (kips)
6	Minute	21	Spacing between Axles 2 and 3 (feet)
7	Second	21	Axle 4 Right Side weight (kips)
8	Vehicle Number	22	Axle 4 Left Side weight (kips)
9	Type	23	Spacing between Axles 3 and 4 (feet)
10	Gross Weight (kips)	24	Axle 5 Right Side weight (kips)
11	Overall Length (feet)	25	Axle 5 Left Side weight (kips)
12	Speed (mph)	26	Spacing between Axles 4 and 5 (feet)
13	Violation code	28 -39	Unused
14	Axle 1 Right Side weight (kips)	40	Direction (PAT System), unused (IRD System)
15	Axle 1 Left Side weight (kips)	41	Axle Number (PAT System only)

In California, there are approximately 106 operational WIM data collection sites along the major and minor truck corridors as shown on the map in Figure 2.1. The typical WIM system includes bending plate scales or pressure sensors that measure axle weight, and inductive loops sensors that detect the presence of the vehicle. A general WIM system in California classifies trucks into 14 FHWA axle-based classification categories using axle spacing and weight measurements and into 13 categories in FHWA schemes (FHWA, 2013) as shown in Figure 2.2. Based on the FHWA classification, a vehicle can be grouped as a passenger vehicle (FHWA class 2 and 3), bus (FHWA class 4), single-unit truck (FHWA class 5 to 7), or multi-unit truck (FHWA class 8 to 14).

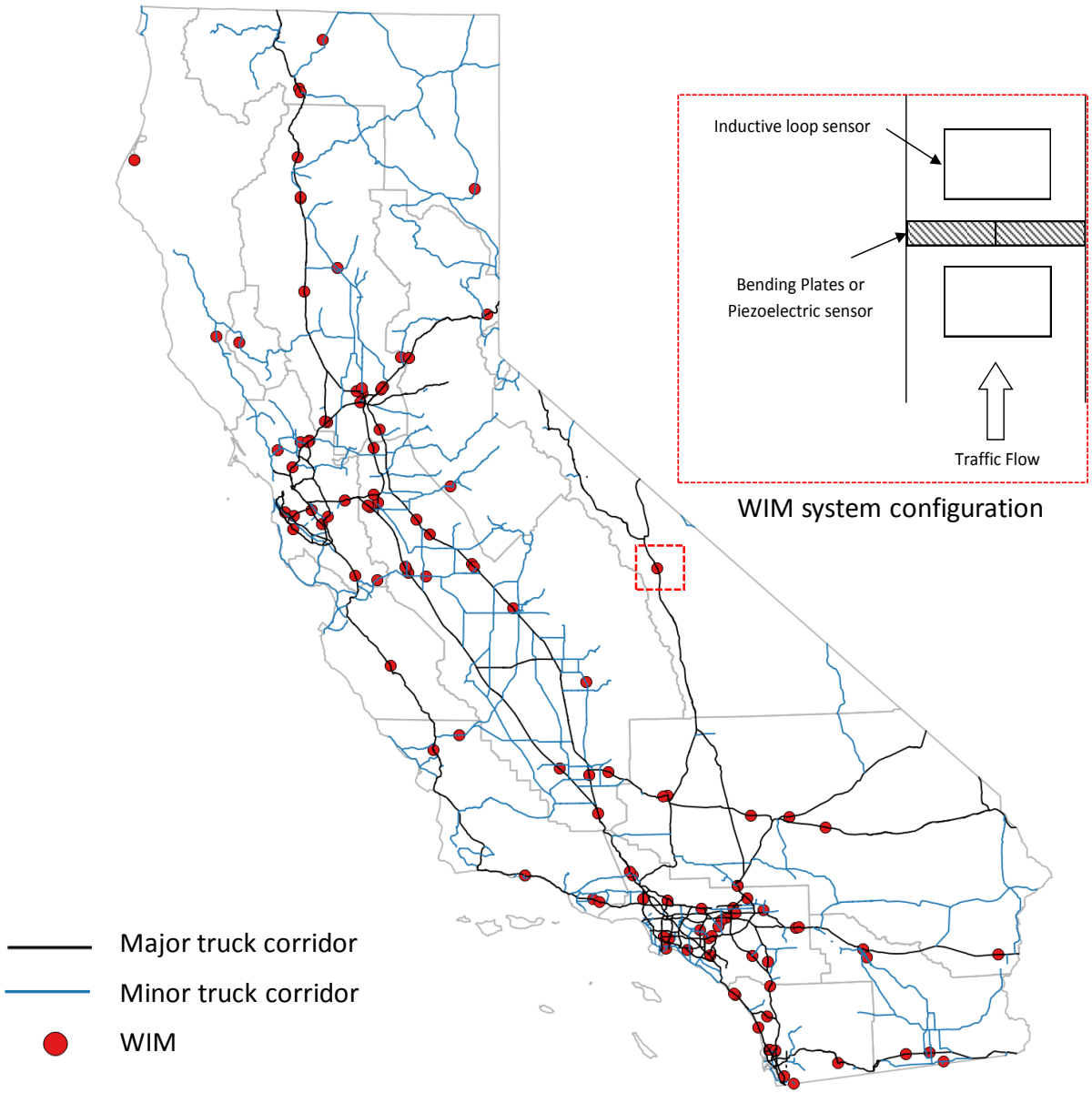


Figure 2.1 WIM systems in California

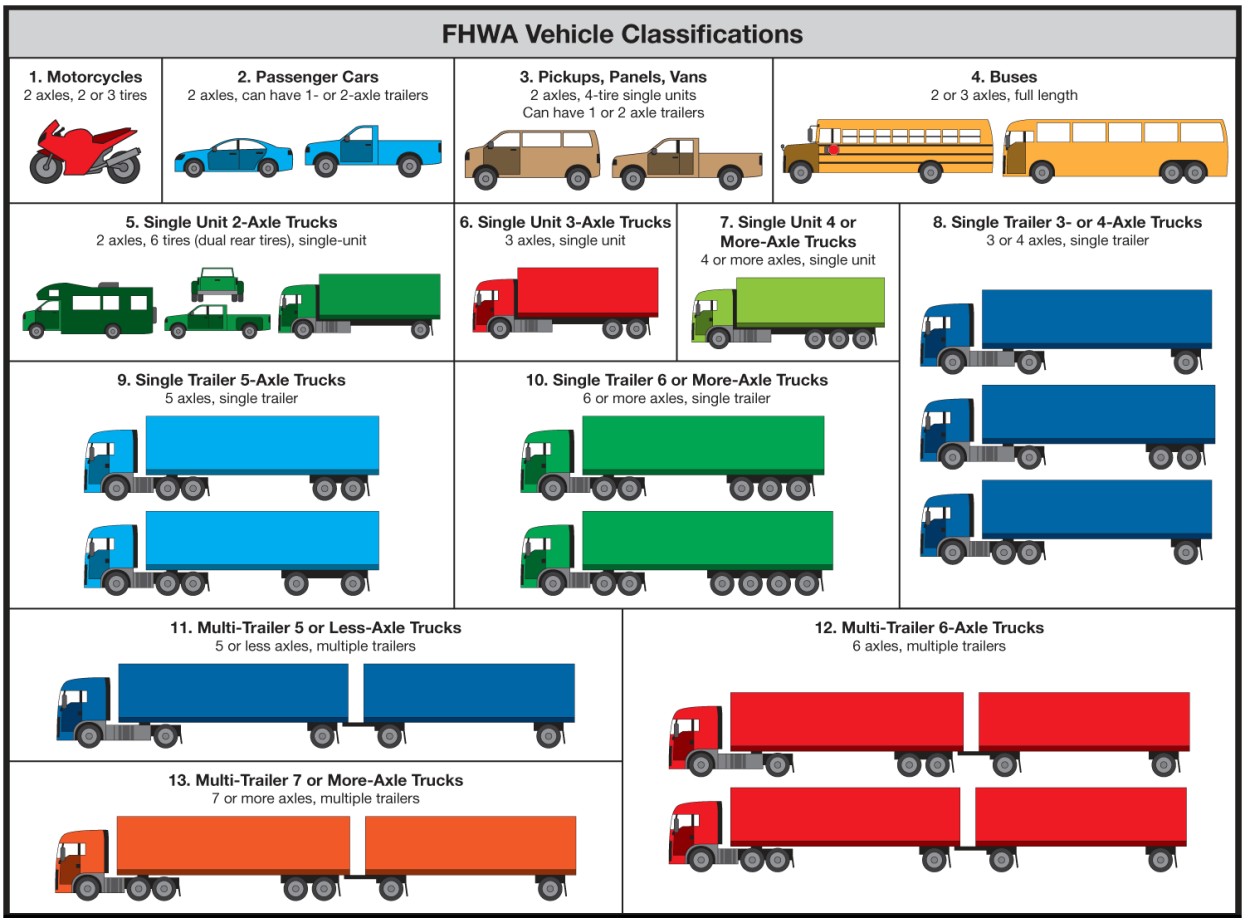


Figure 2.2 FHWA vehicle classification scheme of WIM systems

Agencies using WIM data are aware that WIM data may possess errors in speed, spacing, and weight measurements (FHWA, 2013). The inaccuracies are the result of several possible factors: (i) vehicle dynamics such as speed, acceleration, tire condition, load, and body configuration; (ii) site conditions such as pavement smoothness; (iii) environmental factors such as temperature and precipitation (Nichols, A., and D. M. Bullock, 2004, Papagiannakis et al., 2004). Prozzi and Hong (2007) modeled systematic and random load errors, where random error is the result of statistical fluctuations in estimation which can be over- or under- estimations of the true value. Systematic errors are persistent inaccuracies in which the true value is either

consistently over- or under-estimated. From the previous studies, systematic error are known to be minimized through proper calibration procedures but random disturbances in the data may persist after calibration (Papagiannakis et al., 2004, ASTM, 2009).

2.2 Inductive Loop Detector (ILD)

ILDs are the predominant detection system in the U.S., and one of the most common data sources in various applications including traffic performance measures, traffic operations such as ramp metering and signal control, and crash analysis (Golob et al., 2004; Oh et al, 2005; Zheng et al, 2010). Data from ILDs are typically aggregated in 30 second – or longer – intervals and produce measures of volume, occupancy and sometimes average speed. In California, these measures are publically available through the Performance Measurement System (PeMS, 2016) where over 25,000 ILDs provide traffic information on highway mainlines, highway ramps and, local arterials as shown in Figure 2.3.

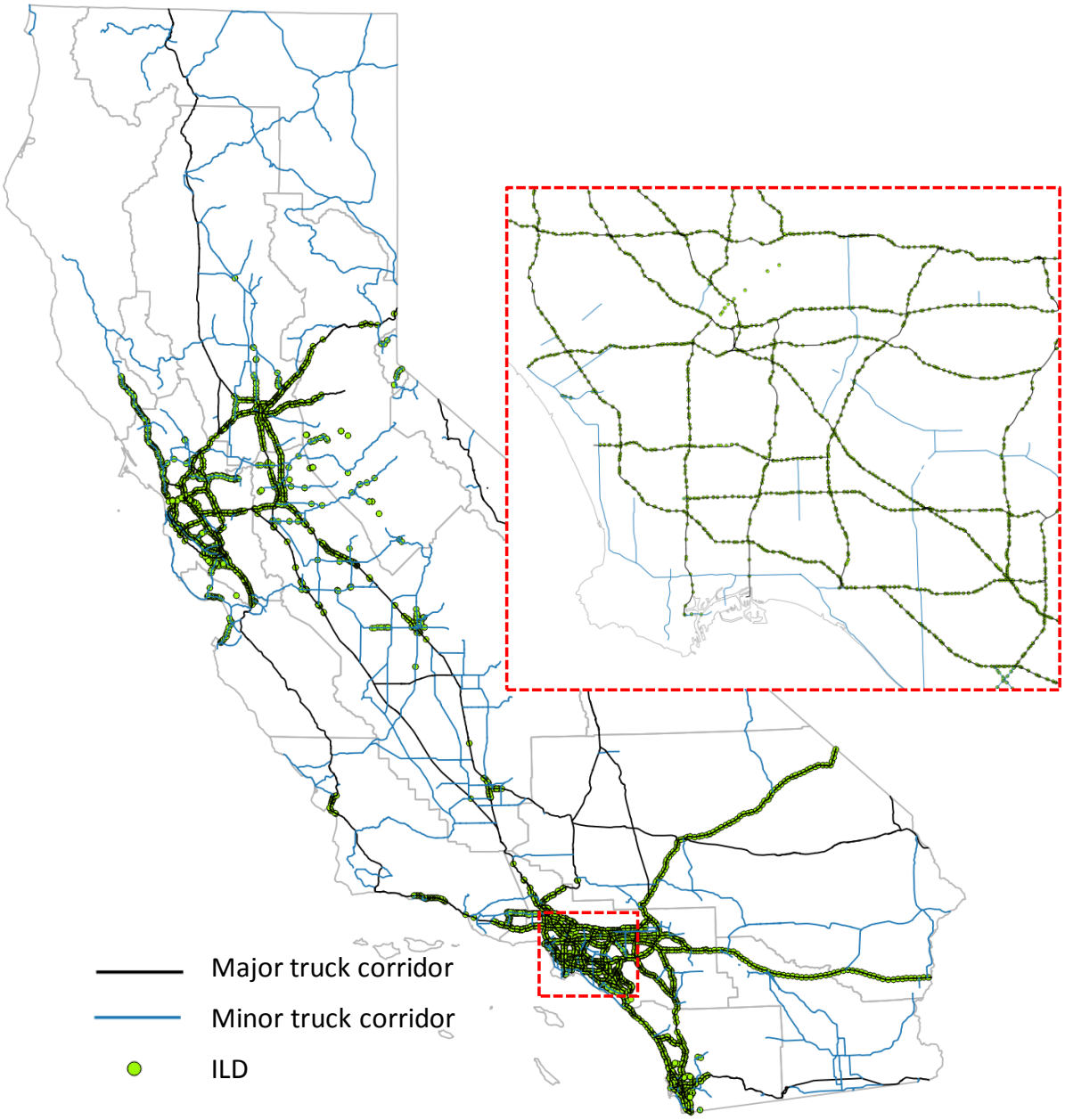


Figure 2.3 ILD systems in California

A conventional presence loop detector detects the presence of a vehicle in a bivalent mode. However, an advanced inductive loop technology that generates inductive signature waveforms of individual vehicles as shown in Figure 2.4. To convert a conventional ILD to an advanced ILD, the roadside hardware can be swapped out with no alterations to the in-pavement sensors. This allows the conversion to be relatively straightforward and cost effective since lane closures are not required.

The metallic composition of a vehicle affects the loop's inductance as it traverses the sensor. Advanced inductive signature technology produces a waveform signature for each vehicle at up to 1200 samples per second. With high sampling rates, the resulting inductive signature can be used to detect vehicle types or body bonfigurations while the conventional binary output can only measure vehicle presence.

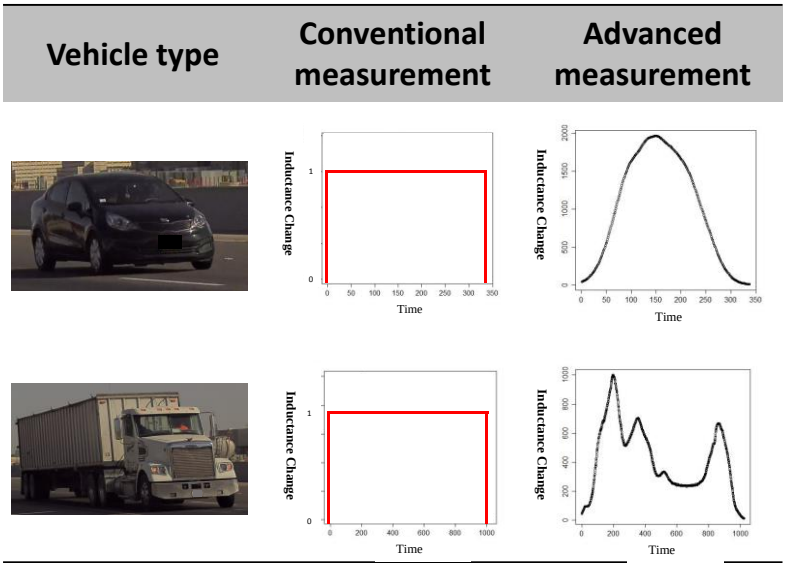


Figure 2.4 Inductive signature technology

Several examples of signatures by different truck body configurations are shown in Figure 2.5.

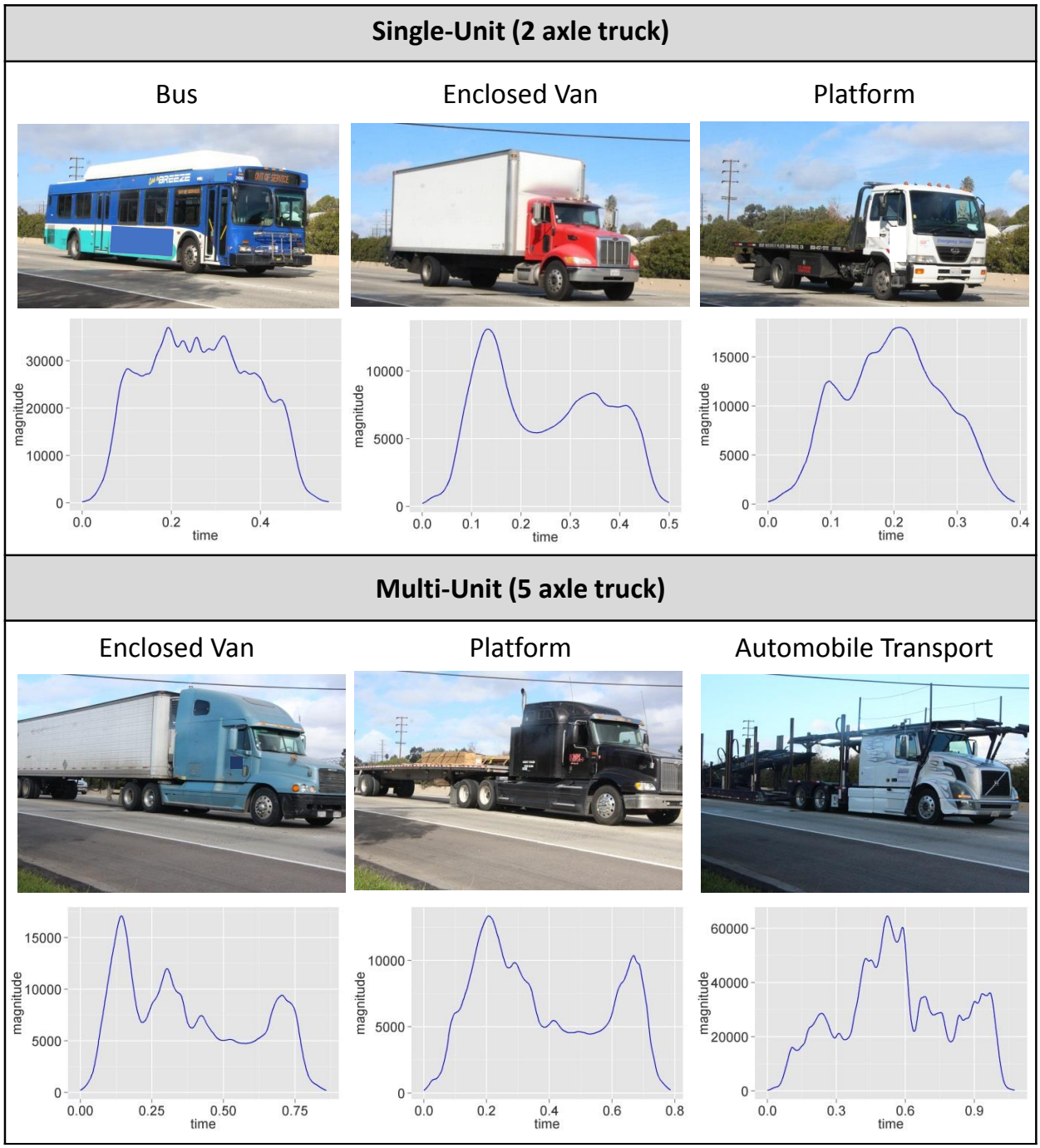


Figure 2.5 Signature from different truck body configuration

However, there are measurement abnormalities and variations in inductive signatures caused from several factors such as vehicle entrance angle, sensitivity of detector, and geometry of detection site. Figure 2.6 shows examples of abnormal signatures generated from such reasons. In all cases, two signatures, generated from the same vehicles and collected at two different detection sites (upstream and downstream) are compared. In case 1, the first part of signature shapes at two locations are significantly different because the vehicle changed the lane while passing the loop sensor at the upstream location. As a result, the upstream signature shows abnormal peaks around 0.25 second. Case 2 illustrates the signature variations from vehicle's lateral position on the loop. Considering that the loop was 6ft-long in its length, small changes in position of vehicle on top of the loop could significantly affect the shape of waveform. In this case, vehicle's different positions on the loop at two locations generate varying magnitude peak for the trailer part of the signature (circled area). Case 3 shows the results of sensor sensitivity on signature shapes. The truck type of case 3 was 20ft container with three rear axles. Due to the three rear axles, the latter part of signature has multiple peaks at upstream location. However, sensitivity and calibration issues of the downstream loop smooth the signature peak of the latter part and cannot capture the three axles accurately.

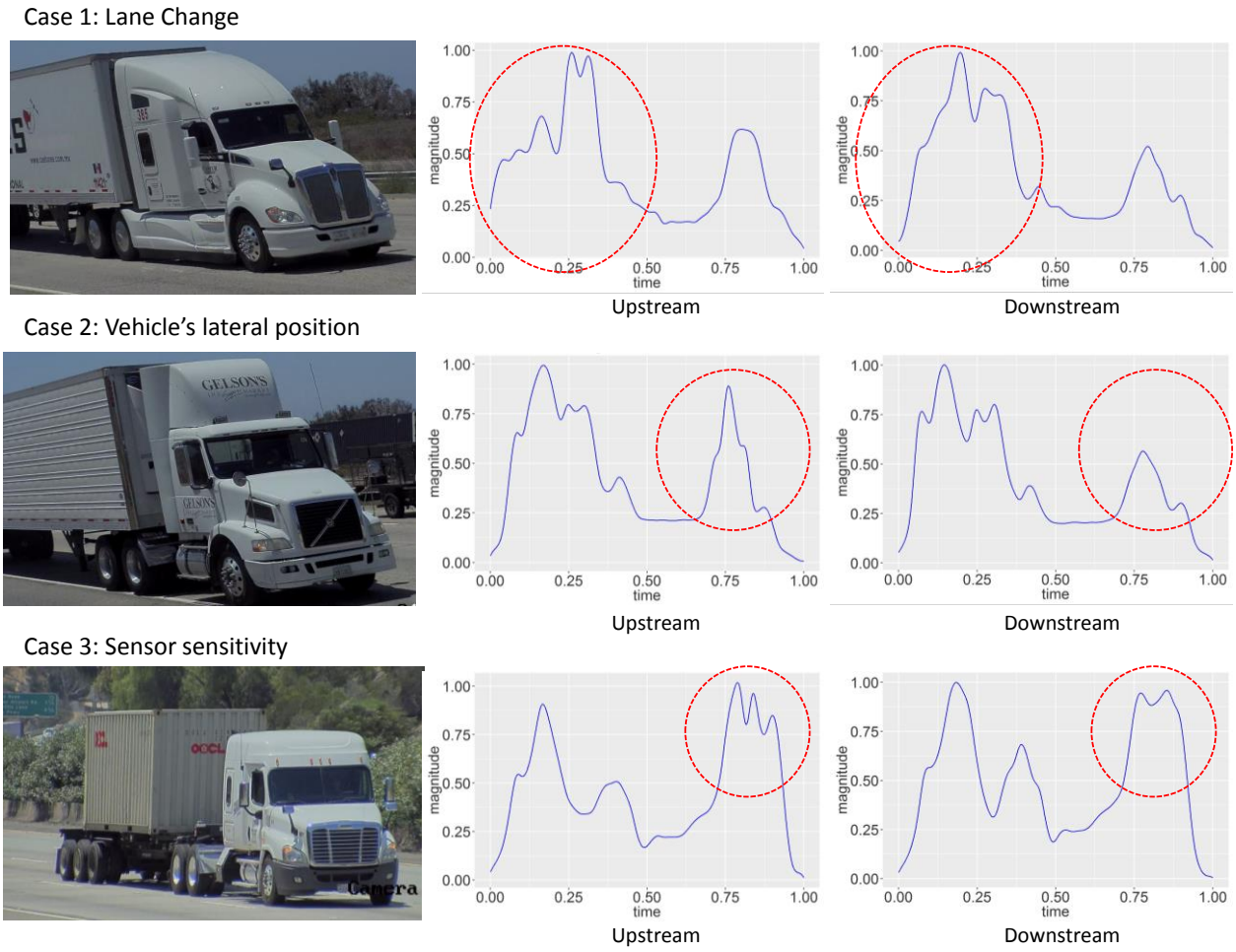


Figure 2.6 Measurement variations in signatures

2.3 Vehicle characteristics from WIM and inductive signatures

Even though signatures are capable of distinguishing vehicles by waveform shapes, if trucks have the same axle configurations and body types, similar signature shapes may be produced. Therefore it would be challenging to identify vehicles solely using the signatures. Figure 2.7 illustrates examples of the signatures from the same body configuration. First, three signatures from the most common body type, enclosed van, are compared and showed particular

common patterns. The second examples from the tank trailers are much more similar to each other since the tank trailer has the most uniform trailer type. As representative body configurations of dropped deck types, signatures from lowboy and automobile transport trucks are presented. Since the distance between dropped deck and loop sensor is close, the middle parts of the signature show high magnitudes in these two truck types.

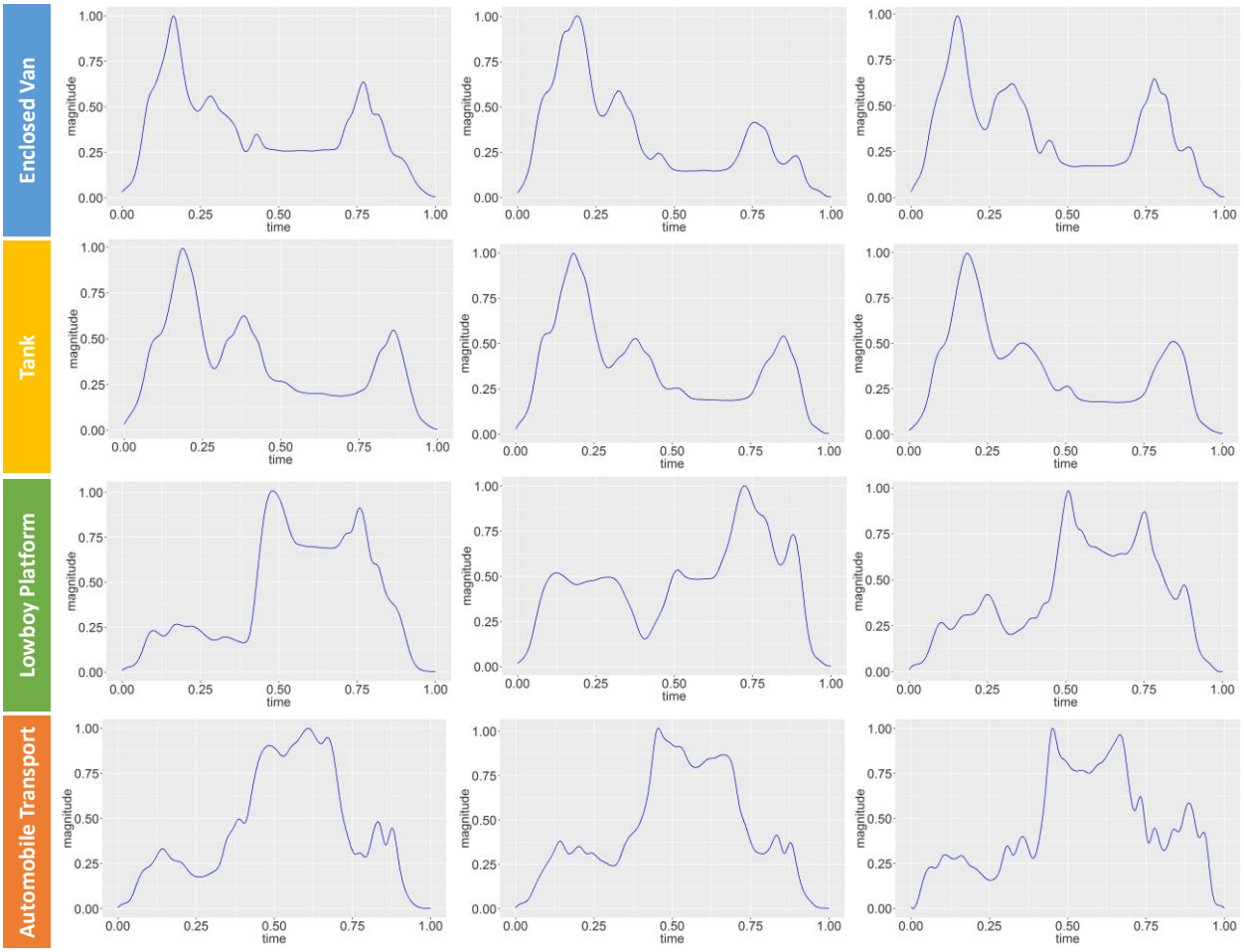


Figure 2.7 Comparisons of signatures in the same body configuration

Therefore, other physical attributes such as axle loads and spacing from the WIM can help distinguish trucks. To compare physical attributes among different truck types and also within the same truck category, five different trailer types from the five axle semi tractor-trailers (3S2 or FHWA Class 9 category) were analyzed. The 3S2s contain the largest variety of truck body configurations. To represent this diversity, this study used five trailer groups including vans, platforms, tanks, 40ft box containers, and 'other' type of trucks. Each group shares similar physical attributes and commodity types. An exploratory analysis was performed to show differences of a set of axle configuration variables among body configurations. A total of five attributes are used for this comparison as shown in Figure 2.8. Three attributes, spacing between the 3rd and 4th axle, overall vehicle length, and a derived measure called overhang, represent as length attributes. Spacing between the 3rd and 4th axles is the measured length in feet between the last tractor axle and the first trailer axle. Length refers to the distance in feet from the nose of the tractor to the tail of the trailer. Overhang represents the front and rear portions of the vehicle outside the axles and is obtained as the arithmetic difference between the overall length and the sum of all axle spacing measurements. Axle weight from the fourth left axle and gross vehicle weight (GVW) are also used as weight loading attributes. GVW is obtained by summing of all of the axle loadings. Since the axle loading is measured by axle and by side, 3S2s have a total of 10 axle loading measures from the first to fifth axles on left and right sides.

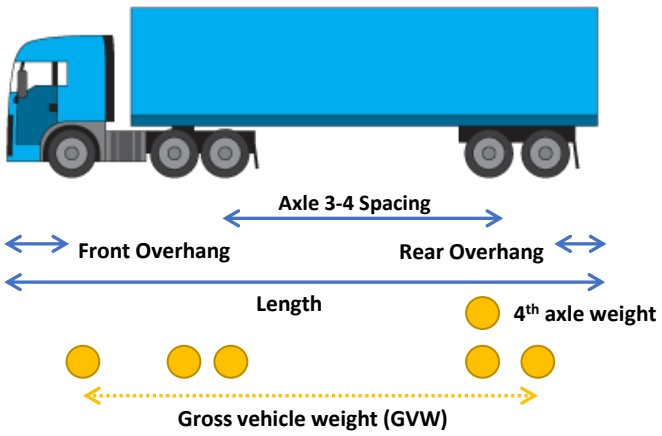


Figure 2.8 Axle configuration variables

Figure 2.9 shows boxplots depicting the descriptive statistics of these variables by body configuration. Van trailers have the longest length and overhang, and are distinctive from other body configurations. Generally, platforms have the second longest length. Tank type trailers have distinctive, shorter overhang. 40ft box containers have relatively longer overhang compared to their short length. A Kolmogorov-Smirnov (KS) hypothesis test (Washington et al., 2010) confirmed that the five body groups are indeed differentiable by length, axle spacing, and overhang.

Compared to the length-based measures, weight measures show more wide ranges in their distributions within the same body configuration. Therefore, it is shown that the length and weight measures together would better distinguish vehicles especially weights help identify ones that share similar length measurements.

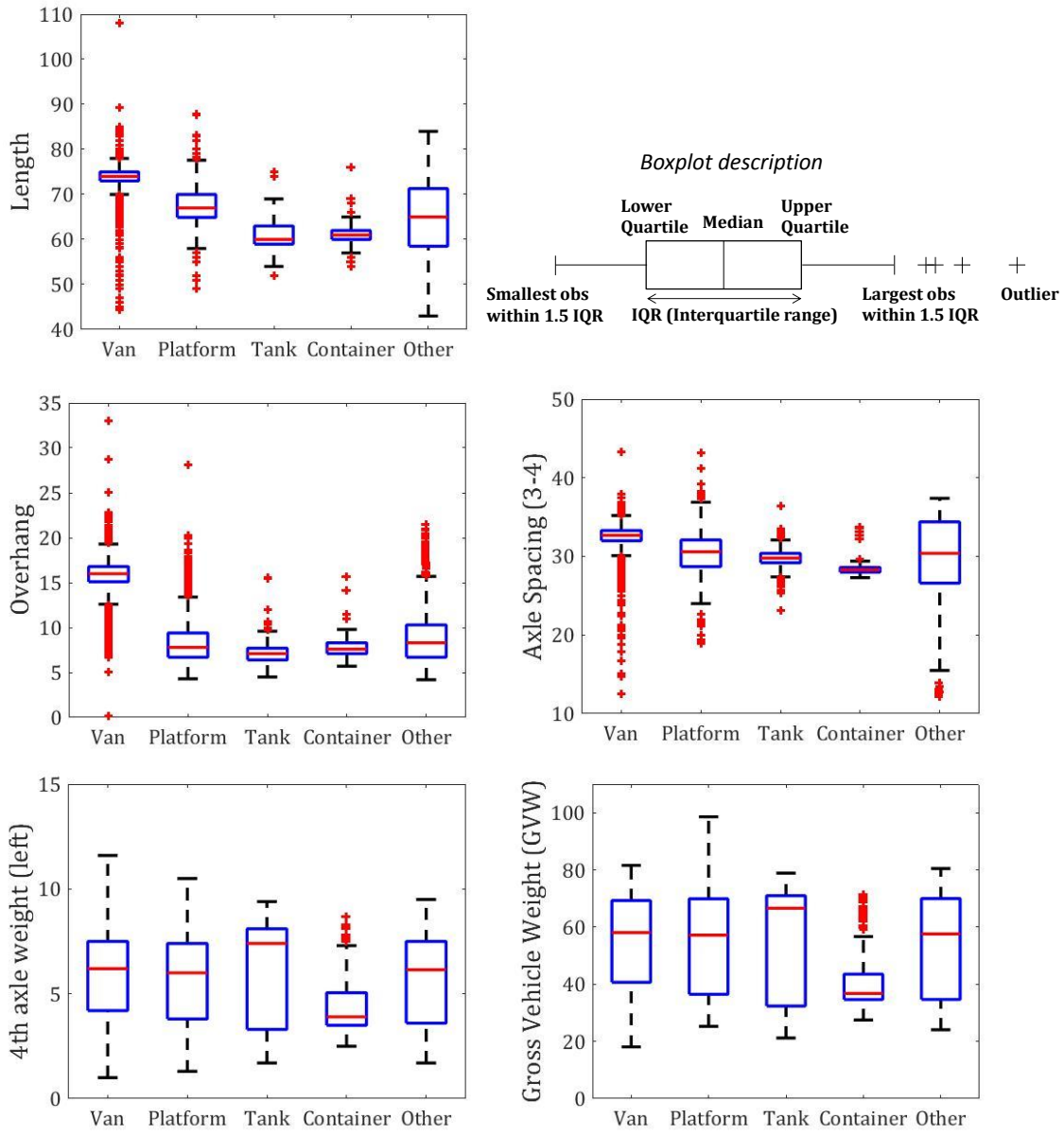


Figure 2.9 Descriptive statistics of WIM variables

2.4 Reviews on general traffic tracking study

In previous studies, vehicle tracking is also referred to as vehicle re-identification. The previous re-identification algorithms primarily sought to obtain accurate section-based travel

times, which aim to characterize link performance. For these reasons, general traffic, which comprises mostly passenger vehicles, has been the main focus. These studies have utilized various attributes for vehicle tracking such as inductive signature waveforms (Sun et al., 1999; Abdulhai and Tabib, 2003; Oh and Ritchie, 2003; Jeng et al., 2010), vehicle length from loop detectors (Coifman and Krishnamurthy, 2007), vehicle signature and color from multi detector fusion (Sun et al., 2004), video images (Sumalee et al., 2012), and magnetic sensors (Cheung et al, 2004; Kwong et al 2009; Tien et al, 2014). This section reviews the general traffic tracking studies categorized by utilized data sources.

2.4.1 Vehicle tracking utilizing inductive signature technologies

Bohnke and Pfannerstil (1986) first introduced the use of inductive waveforms to re-identify vehicle sequences. Kuhne (1991) followed by developing a freeway vehicle re-identification technique using dynamic traffic flow models. Extensive research using inductive vehicle signature has been performed since the late 1990's. Methodology, dataset, and accuracy ranges of the key literatures on general traffic tracking using ILD technology are summarized in Table 2.2.

Table 2.2 Summary of performance for key literatures on general traffic tracking

Author (year)	Dataset (Equipment)	Tracking distance	Accuracy ranges
Sun et al (1999)	Signature from ILDs	1.2 mile (freeway)	75% for passenger vehicles and 78% non-passenger vehicles
Abdulhai and Tabib (2003)	Signature from ILDs	1.2 mile (freeway)	56.3% on average (25% for passenger cars and 100% for heavy vehicles)
Oh et al (2003)	Synthetic signature from ILDs	Signalized intersection	73.6% for uncongested and 44.3% for congested condition
Jeng et al (2010)	Signature from ILDs	0.33 to 1.3 mile (freeway)	52.2% to 56.2% depends on the loop configuration
Cofiman and Krishnamurthy (2007)	Vehicle length from dual loop	0.91~0.97 mi (freeway)	36% ~ 41% for long vehicles
Sun et al (2004)	Signature from ILDs and video image	425 feet (arterial)	91.36% when using the best fusion weight

Sun et al (1999) proposed multi-objective optimization approach to formulate vehicle re-identification problem using inductive loop signatures. Arrival time, duration, and speed were collected from individual vehicles based on the signature attributes. Sun (1999) solved the re-identification problem by applying lexicographic method which was composed of a sequence of objective functions. Specifically, a total of five levels of objective functions were considered to match vehicles between detector stations, and hierarchical ordering of each objective function narrowed search space. Since matched vehicles in the upper level determined the search space for next level objective, computational burden was effectively reduced through this approach. In specific, from the first to third level objectives were goal programming that finds candidate vehicles at downstream detection location based on the time, percent signature magnitude, and length difference from the target vehicle. The fourth level objective identified the best matched vehicle using the signature magnitude, lane, and speed as matching features. Multiple distance measures such as Euclidean, correlation, similarity, Lebesque, and neural network modeling

were applied to estimate feature differences. At the last level optimization, the best distance measure was determined using posterior probability from the Bayesian approach. As a result, a correct matching rate (CMR) was 75% for passenger vehicles and 78% for non-passenger vehicles.

Oh (2003) developed an anonymous vehicle tracking algorithm focusing on passenger vehicles at signalized intersections. A probabilistic pattern recognizer based on the adaptive probabilistic neural network (APNN) was applied to match feature vectors from the inductive waveform. Feature vectors consisting of interpolated points at equally spaced increments along the waveform, and a spatial and temporal search space reduction based on the signal phase were implemented. Smoothing parameters from genetic algorithms (GAs) and self-organizing maps (SOMs) were proposed with the APNN. Based on the re-identification algorithm, Oh (2003) developed a tracking evaluation framework with a simulation program. Further implications including origin-destination estimation and real-time level of service analysis were investigated along multiple intersections. Algorithm performance was tested during congested and uncongested traffic conditions. The overall correct matching rate was 73.6% during uncongested conditions and 44.3% during congested conditions for the test data.

Tawfik et al (2004) adapted the lexicographic methods developed by Sun (1999) with a heuristic decision tree algorithm. In other words, the re-identification problem was solved as a classification problem via decision tree. The experimental evaluation was performed at a 1.2 mile freeway segment in Irvine, California. The first tree which was constructed with two features, lane change and length difference, showed 75% correct classification rate (CCR) on the test dataset. The second tree used travel time distribution to limit the search window. Time and speed difference were selected as feature vectors, and the tree showed 89% of correct matches

and 78% of correct mismatches. To eliminate false matches, an additional feature vector from the inductive loop signature was added. Multiple distance measures such as Euclidean, Lebesque, and first derivative estimated feature distances. In addition, signatures were transformed by the spatiotemporal approach so that spatial or temporal inconsistency on the loop detector can be reduced. Distance measures, speed difference, length difference and lane difference were used to develop more accurate decision tree and the results were compared to that of Sun (1999). The decision tree achieved 89% CCR, compared to 75% from the lexicographic optimization and 61% of the signature matching. The author additionally noted that the travel behavior such as persistence of lane and speed were the useful factors to match vehicles.

Jeng (2010) developed a real-time vehicle re-identification algorithm, called RTREID-2. To extract feature vectors, a cubic spline interpolation and piecewise slope rate (PSR) was introduced. Smooth curves through the data points were obtained by cubic spline interpolation, which also reduced all signatures to the same number of data points. A PSR technique transformed and compressed raw signatures to expedite processing time by reducing the size of input data in the algorithm. Once PSRs were estimated for each signature, differences in PSRs between the target vehicle at downstream and candidate vehicles at upstream were compared to find the pairs with the minimum difference. Different loops configurations were tested with the PSR approach and showed the matching accuracies ranging from 50.7 to 54.2 percent by the loop configurations.

Abdulhai (2003) identified a new distance measure to improve accuracy of re-identification algorithm. Conventional statistical measures (i.e., Euclidean, Correlation, Lebesgue, First derivative, MSE, and FFT), neural network (i.e., back propagation neural

network, time delay neural network and probabilistic neural network) and warping insensitive measures were introduced to match patterns in the vehicle inductance signature. The warping-insensitive matching measure transformed the waveform signature and supported the signature shape not to be affected by external spatio-temporal environment. A signature transformation horizontally shifted the signature by the function of time was employed. Through this process, the error from the matched vehicles' signatures was significantly reduced. The algorithm was applied to moderate and congested conditions with different candidate vehicle window sizes of 10, 20, and 30 vehicles using multiple distance measures. The result showed that the waveform transformation positively affected the performance with around 12% better performance for the congested condition and about 10% better performance for the moderate condition. Additionally, a neural network was selected as the best distance measure with the 48% accuracy for the before-transformed dataset and Euclidean distance measure was chosen for the after-transformed set with 56% accuracy. However, the experiment showed that the window size did not play an important role in re-identification performance.

Coifman (2007) focused on the development of vehicle re-identification algorithm for travel time measurement. Re-identification is conducted based on vehicle length obtained from conventional double and single loop detection systems. Individual occupancy measures were applied as the vehicle feature. Based on loop occupancy, long vehicles were identified as distinct vehicles and tracked by arrival sequence. Since the algorithm only stored the distinct vehicles detected at downstream location as target vehicles, vehicles were easily matched to the target vehicles by their longer lengths and lower frequency. As a result, possible matches were stored, and a travel time matrix (TTM) was built with candidate matching pairs. From the TTM, a maximum density matrix (MDM) was created to rule out false matches. About 40% of long

vehicles were re-identified, and travel time obtained by the proposed algorithm showed better capturing traffic flow rather than the point observation, especially during unstable traffic conditions.

2.4.2 Vehicle tracking utilizing other detector systems

A few studies have been conducted using other detector systems than ILD such as Bluetooth, video detector, and magnetic sensor. Sun et al (2004) presented a multi-detection fusion algorithm. In addition to the information from point detectors, color images from the video detectors were utilized for vehicle re-identification. To integrate data from multiple sources, a data fusion technique was investigated. The motivation was that the vehicle of the same model may yield high mismatching probability because the loop detector is only able to capture the metal composition of vehicle. Thus, when color information is added to signature features, vehicles would be better distinguished. To develop the algorithm, vehicle signature, speed, platoon traversal time, maximum inductive amplitude, and color were used as feature vectors. Each upstream platoon within the time window was compared to the downstream platoon with nearest neighbor classifier. A sensitivity analysis showed that the optimum platoon size for the re-identification accuracy maximized with three vehicles in one platoon. Weights for linear fusion of feature factors were determined through searching the best performance cases with the train dataset. The highest vehicle re-identification accuracy was shown at 92% when all the information was integrated. In comparison, the correct matching rate was 87% and 76% when the signature and color was solely used as feature vectors, respectively.

Sumalee et al (2012) proposed the probabilistic fusion algorithm for vehicle re-identification using video image data. Feature vectors including vehicle color, type and length constructed a probabilistic model based on Bayesian approach and Gaussian mixture model. Posterior probability density function was estimated by the feature vector's distance and prior probability density function was derived by the travel time. A data fusion rule for three feature vectors was performed with a logarithmic opinion pool approach. Furthermore, a bipartite matching method was introduced to remove 'overlapped matched' cases which are defined as the multiple vehicles matched to the same downstream vehicle. Testing was conducted along a 5km expressway in Bangkok, and 55% accuracy was shown for the total of 574 detected vehicles at the upstream location. The performance for color was superior to vehicle type and length, and the result indicated that the probabilistic fusion method outperformed than individual use of three variables.

Young et al (2012) applied Bluetooth technology for vehicle re-identification problem. Due to requirements for longer scanning periods, poor antenna performance, and small detection zones, utilizing Bluetooth technology for vehicle re-identification has been shown to be less feasible than using other detection technologies.

Another emerging technology for vehicle tracking is intrusive magnetic sensors. This technology is similar to the ILD in terms of capturing vehicle signatures using a device however it is known to be less susceptible than ILD for traffic conditions. Cheung et al (2004) reported that the magnetic sensor better captures the distinctive features than ILD since it records the changes in the fields caused by different parts of the vehicle while the loop records "the integration of the inductive signature over the traversal distance". Kwong et al (2009) developed

a statistical travel time estimation model using vehicle signatures from wireless magnetic sensors. Tian et al (2014) identified vehicles with a single node and multi nodes sensor deployments and showed 74.4 percent and 81.4 percent accuracies in truck tracking for single node and multi nodes, respectively.

2.5 Limitations on truck tracking studies

Researchers have only recently begun to investigate the problem of tracking commercial vehicles over longer distances using existing detector infrastructure. Cetin (2011) developed a vehicle re-identification algorithm using a Bayesian approach for five axle semi tractor-trailer trucks that correspond to FHWA class 9 trucks over a 100 mile corridor. Truck axle data and length from WIM stations were selected as feature vectors, and travel time obtained by the vehicle transponder was used for time window. In this study, a two-step algorithm was proposed. The first step consisted of developing the Bayesian re-identification algorithm to find a matched vehicle pair. The second step was to screen out false match using the posterior probability from the Bayesian model. Even though the vehicle pair with highest posterior probability was selected as a match at the first step, the second step decided if the selected match was indeed correct. The basic assumption was that if matching was incorrect, the posterior probability difference between matched and mismatched pair would be similar; otherwise truly matched pair has much larger posterior probability. Performance results were presented with various scenarios where the scenarios showed trade-off of the total matched vehicles and matching rate. Specifically, if 80 percent of common vehicles are correctly matched, 92 percent of total matches declared by algorithm were correct matches, however, if 88 percent of common vehicles are correctly matched, only 89 percent of total matches are shown to be correct. However, it should

be noted that a large proportion of the trucks (71 percent) of vehicles traverse both up- and downstream WIM stations in the study corridor, which would inherently lead to higher matching accuracy.

Jeng and Chu (2015) utilized inductive loop signatures and WIM to track trucks. The inductive loop signatures were the main data source to match vehicles and the WIM data were subsequently used to filter out mismatching vehicles after identifying potential matching vehicles. The algorithm was developed and tested for commercial trucks on 19 mile freeway in Orange County, California where 21 interchanges exist and traffic volume is heavy. Vehicles were matched based on proximity measures such that a vehicle pair with the minimum distance between signatures obtained at upstream and downstream stations was selected as the matching pair. WIM data filtering was applied with pre-determined threshold such as 0.1 for the axle parameter (e.g., axle 1 weight / gross vehicle weight), where a vehicle pair with less than the threshold was selected as a true match. However, since proximity measures were used to find potential matching pairs and a fixed threshold was applied in the filtering process, inaccurate calibration or different sensitivity in scales at different WIM sites may significantly affect matching results. In this regard, this fixed threshold approach with the separate analysis of WIM and signature features might eliminate many potentially correct matches, which makes it unsuitable for path flow estimation applications. Hence, the applications of this study mainly focused on measuring link performance and WIM calibration with a proportion of matched vehicle pairs.

3. Modelling Approach

A modeling approach to facilitate development of detailed truck flow data is described in this chapter. Based on Bayesian inference, a Selective Weighted Bayesian Model (SWBM) was developed to track individual vehicles between detector locations using physical attributes and inductive waveform signatures of individual trucks. Key feature variables were weighted in the SWBM to improve vehicle matching performance in a long distance tracking. Various statistical and machine learning algorithms were introduced for feature selection and weighting techniques.

While Weight-in-motion (WIM) and inductive loop point detectors (ILDs) were utilized for the corridor level tracking, the network-wide tracking used ILD and supplementary data sources such as historical GPS trajectories and body configuration model estimates to handle much more candidate vehicles detected at multiple upstream locations. Bayesian updating method and a linear data fusion were considered to integrate multiple sources in the network-wide tracking.

3.1 Modeling Background

3.1.1 Bayesian Inference

The SWBM is an extended form of a naïve Bayes classifier. As a family of probabilistic classifier in machine learning techniques, a naïve Bayes classifier is largely applied to various types of classification such as text categorization, medical diagnosis, and vehicle classification in transportation area.

Naïve Bayes is a type of a supervise learning that requires labeled training data. In other words, a pair of example should consist of input attributes (i.e., feature vectors) and output value

(i.e., class). In our cases, vehicle pairs (i.e., examples) of match and mismatch (i.e., classes) can be distinguished by vehicle attributes (i.e., feature vectors). Collected data were split into train and test dataset. The tracking model is developed by training dataset and validated with testing dataset. In general, the supervised learning algorithm analyzes the training data and produces a mapping function. In our cases, Bayesian classifier is the mapping function that matches classes and feature vectors. The mapping function determines classes of the unseen data in test dataset.

Bayes theorem can be expressed as follows:

$$p(x \cap \mu) = p(x) \cdot p(\mu|x) = p(\mu) \cdot p(x|\mu)$$

$$p(\mu|x) = p(\mu) \cdot \frac{p(x|\mu)}{p(x)}$$

$$\text{If, } \mu = (\mu_1, \mu_2, \ldots., \mu_j),\ p(\mu_j|x) = \frac{p(x|\mu_j)p(\mu_j)}{p(x)}$$

since $p(x)$ does not depend on μ,

$$p(\mu_j|x) \propto p(x|\mu_j) \cdot p(\mu_j)$$

where $p(x) = \sum_{j=1}^{J} p(x|\mu_j)p(\mu_j)$, $p(x|\mu_j)$ represents likelihood and $p(\mu_j)$ represents evidence (prior) where the class μ have j possible outcomes, x represents feature vectors.

If $x = (x_1, x_2, x_3, \ldots. x_n)$, $p(x|\mu_j)p(\mu_j)$ is equivalent to the joint probability model, $p(\mu_j, x_1, .. x_n)$. From the chain rule,

$$p(\mu_j, x_1, .. x_n) = p(x_1, .., x_n, \mu_j)$$

$$= p(x_1 | x_2, \ldots, x_n, \mu_j) \cdot p(x_2 | x_3, \ldots, x_n, \mu_j) \cdot p(x_3 | x_4, \ldots, x_n, \mu_j) \ \ldots$$

$$\cdot \ p(x_{n,-1} | x_n, \mu_j) \cdot p(x_n | \mu_j)$$

Conditional independence assumptions of the naïve Bayes (i.e., each feature is conditionally independent to other features given a category μ) gives us:

$$p(x_i | x_{i+1}, \ldots, x_n, \mu_j) = p(x_i | \mu_j)$$

Therefore,

$$p(\mu_j | x_1, x_2, \ldots x_n) \propto p(\mu_j, x_1, x_2, \ldots x_n)$$

$$\propto p(\mu_j) \cdot p(x_1 | \mu_j) \cdot p(x_2 | \mu_j) \cdot p(x_3 | \mu_j) \ \ldots \cdot p(x_n | \mu_j)$$

$$\propto p(\mu_j) \cdot \prod_{i=1}^{n} p(x_i | \mu_j)$$

3.1.2 Bayes classifier decision rule

In a naïve Bayes classifier, if we have two possible outcomes: $j = 0 \ or \ 1$, the final outcome can be decided by posterior probabilities.

$$\text{Decide} \begin{cases} \mu_1 \ if \ p(\mu_1|x) > p(\mu_2|x) = \frac{p(x|\mu_1)}{p(x|\mu_2)} > \frac{p(\mu_2)}{p(\mu_1)} \\ \mu_2 \ otherwise \end{cases}$$

If our action is a_i, our loss equals to $\lambda \ (\alpha_i|\mu_j)$. Then, the expected loss (r) with taking action of α_i can be obtained by the conditional risk as follows.

$$r(\alpha_i|x) = \sum_j \lambda\,(\alpha_i|\mu_j)\,p(\mu_j|x)$$

Bayes decision rule minimizes this error by selecting the action of α_i so that the total error, $R\,(\alpha_i|x)$, can be minimized.

For example, if $\alpha_1 = deciding\ \mu_1,\ and\ \alpha_2 = deciding\ \mu_2$, then loss = $\lambda_{ij} = \lambda\,(\alpha_i|\mu_j)$

$$R(\alpha_1|x\,) = \lambda_{11} \cdot p\,(\mu_1|x\,) + \lambda_{12} \cdot p\,(\mu_2|x\,)$$

$$R(\alpha_2|x\,) = \lambda_{21} \cdot p\,(\mu_1|x\,) + \lambda_{22} \cdot p\,(\mu_2|x\,)$$

Minimum risk decision rule should be:

$$\mu_1 \ \ if\ (\lambda_{21} - \lambda_{11})\,p\,(\mu_1|x\,) > \ (\lambda_{12} - \lambda_{22})\,p\,(\mu_2|x\,)$$

$$\mu_2\ otherwise$$

Decide $w_1\ if$

$$(\lambda_{21} - \lambda_{11}) \cdot \frac{p(x|\mu1)p(\mu_1)}{p(x)} > (\lambda_{12} - \lambda_{22}) \cdot \frac{p(x|\mu2)p(\mu_2)}{p(x)}$$

Thus,

$$\frac{p(x|\mu_1)}{p(x|\mu_2)} > \frac{(\lambda_{12} - \lambda_{22}\,)}{(\lambda_{21} - \lambda_{11}\,)} \cdot \frac{p(\mu_2)}{p(\mu_1)}$$

Since $\lambda_{11}\ and\ \lambda_{22}$ equal to zero,

$$\frac{p(x|\mu_1)}{p(x|\mu_2)} > \frac{(\lambda_{12})}{(\lambda_{21})} \cdot \frac{p(\mu_2)}{p(\mu_1)}$$

In the Bayes decision rule, there will be two types of errors as shown in Table 3.1. If the true class is μ_1 but μ_2 is assigned as a decision, we call this error as false negative or misdetection. On the other hand, assigned μ_1 is not matched to the true decision, μ_2, we call this error as false positive or false alarm.

Table 3.1 Type of errors in Bayes decision rule

		Assigned	
		μ_1	μ_2
True	μ_1	True	False negative (Type 1 error or mis-detection)
	μ_2	False positive (Type 2 error or false alarm)	True

3.1.3 Bayes classifier feature selection

A key assumption of the naïve Bayes model is the independence of attributes. This independence assumption is violated in the use of all attributes obtained from WIM. This is because the axle spacing and total length measurements of a vehicle are highly correlated with each other. Hence, a selective Bayesian model that uses subsets of attributes was applied to preserve the independence assumption and to improve matching performance.

For the feature selection method, this dissertation applies an information gain (IG). The IG feature selection is commonly used approach in a supervised learning algorithm, particularly for the decision tree model. IG uses entropy as a measure of purity in an arbitrary collection of examples. Let S be a set consisting of s data samples, and C_k be a class label of m case, $(k = 1, \ldots m)$. An expected information (H) for classifying a given sample is expressed as follows.

$$\mathrm{H}(\mathrm{s}_1, \ldots \mathrm{s}_\mathrm{m}) = -\sum_{k=1}^{m} p_k \log_2(p_k)$$

where $p_k$represents the probability that an arbitrary sample belongs to class C_k.

Information Gain (IG) entropy-based filter measures the amount of information about the class prediction in using subsets of attributes (Roobaert et al., 2006; Ratanamahatana and Gunopulos, 2002). Since the entropy measures the impurity or randomness, lower entropy indicates a feature with higher deterministic characteristics.

A set of S is partitioned into v subset $\{ \mathrm{s}_1, \ldots, \mathrm{s}_\mathrm{v}\}$. An attribute set of A for these subsets can be expressed as $(\mathrm{a}_1, \ldots \mathrm{a}_\mathrm{v})$. Therefore, s_j contains samples that have values of a_j in A. If class of subset is considered, S_kj represents a subset with an attribute value j that has k as a class. The entropy or expected information based on the partitioned subset is as follow.

$$\mathrm{E}(\mathrm{A}) = \sum_{j=1}^{v} \frac{(s_{1j} + \cdots + s_{mj})}{s} H(s_{1j}, \ldots, s_{mj})$$

Therefore, information gain (IG) can be as follows.

$$\mathrm{IG} = \mathrm{H}(\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_\mathrm{m}) - E(A)$$

An example in Figure 3.1 provides conceptual illustrations of Entropy. Two decision trees with different split variables compares entropy and corresponding IG.

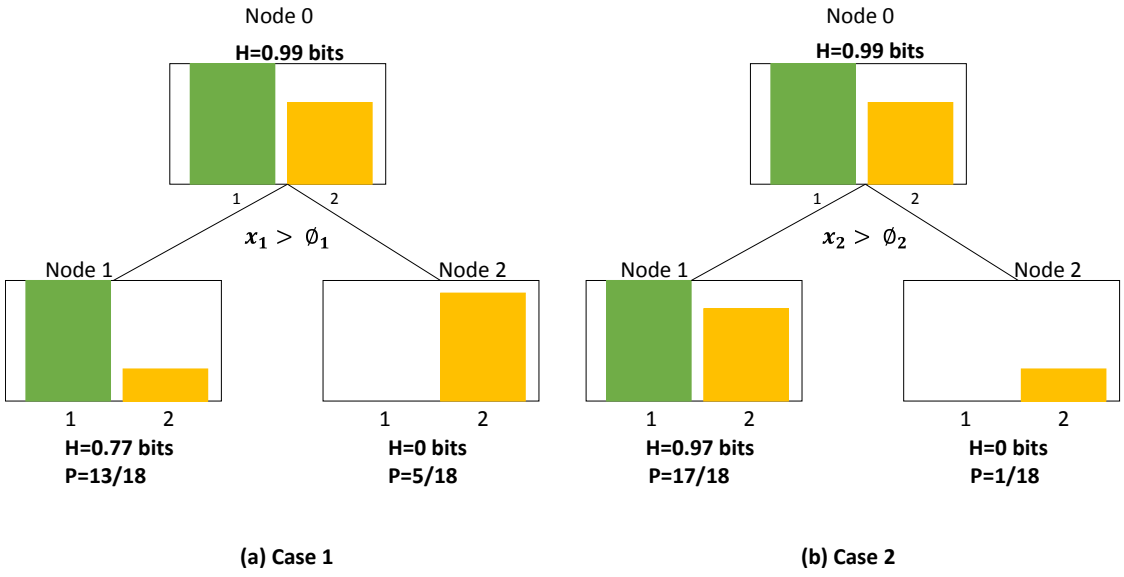


Figure 3.1 Entropy and Information Gain

IG is obtained as follows:

$$\text{IG} = prob(node\ 1) * H(node\ 0 - node\ 1) + prob(node\ 2) * H(node\ 0 - node\ 2)$$

$$= \text{H(node0)} - [\, prob(node1) * H(node1) + \big(1 - prob(node1)\big) * H(node2)]$$

Since the goal is to have less entropy after the node split, $H(node1)\ and\ H(node2)$ should be minimized. Consequently, larger IG will be obtained from less Hs. IGs for the case 1 and case 2 are 0.434 and 0.074, respectively, therefore the variable from case 1 (x_1) is selected as the first split variable.

In this dissertation, weights of attributes – entropy-based filters – were identified based on their correlations with class attributes of match and mismatch. Therefore, the subset with a higher IG was selected to improve the distinction between matched and mismatched pairs.

It should be noted that Entropy-based filters were not considered for features obtained from inductive signatures as they were assumed to be independent.

Hence, Bayes classifier considering feature selection from IG can be expressed as follows.

$$p(\mu_{ij} \mid x) \propto \frac{IG_m \cdot \Pi_{m=1}^{M} p(x_m|\mu_{ij})}{\sum_{i=1}^{I} IG_m \cdot \Pi_{m=1}^{M} p(x_m|\mu_{ij})}$$

3.1.4 Bayes prior

A central concept of Bayesian analysis is to update a prior to posterior distribution for parameter vectors based on a received dataset which are summarized through a likelihood function for the parameters (West, 1993).

$$posterior \propto prior \times likelihoood$$

$$\pi(\mu_j) = prior$$

$$\pi_t(\mu_j) = p(\mu_j|x_t) = current\ state\ at\ time\ t$$

$$\propto \pi(\mu_j) \times L\ (x_t|\mu_j)$$

$$= \frac{p(x_t|\mu_j)p(\mu_j)}{\sum_j p(x_t|\mu_j)p(\mu_j)}$$

where j represents match (μ_j=1) or mismatch (μ_j=0) in our case.

While posterior can be easily estimated based on a given dataset, prior represents one's belief about the probability before some evidence is taken into account. In Bayesian classification, one of the most important problems is to define a prior distribution since the prior significantly affects posterior probability. Figure 3.2 shows the final inferences from the prior distribution.

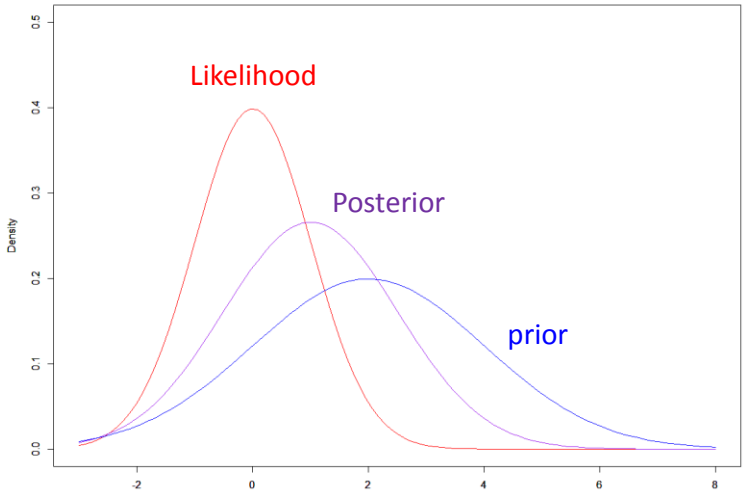


Figure 3.2 Influences from prior

A number of methods create prior distributions, largely categorized into two types: uninformative and informative priors. When a parameter θ is available for prior information, prior density can be estimated based on the parameter. However, if we have no prior information, uninformative prior is often used to minimally affect the final inferences. A representative uninformative prior is a uniform distribution. Informative priors give us numerical information on parameter distributions where the distribution can be estimated from direct observations using histogram and kernel density estimates or from chosen probabilistic models.

In our tracking models, two types of priors are assumed. In a corridor tracking model, a prior matching probability is assumed to be equal for all pairs. Since the tracking is implemented along the same corridor, without additional information, it would be reasonable to expect equal matching chances from all candidate vehicles. However, in a network-wide tracking, since candidate vehicles are collected from multiple upstream sites, prior information can provide extra matching probabilities if supplementary sources have information on detection locations or independent data on vehicle matching. For example, historic path flows from GPS trajectories can be used as a prior on potential upstream sites since GPS trajectories provide general travel flow between sites. In addition, direct comparisons of body configurations between target and candidate can give us an insight on matching probabilities for vehicle pairs. From these supplementary sources, discrete prior probabilities are estimated and updated for every matched pairs in vehicle matching and filtering processes. More details on data sources and prior probability will be discussed in Chapter 5.5.

3.1.5 Linear Data Fusion and the final SWBM

A linear combination method is known as an effective data fusion approach in integrating multiple data sources. In a general form, the linear fusion can be expressed as follows (Wu et al., 2011):

$$\mathrm{M(d)} = \sum_{i=1}^{n} wt_i * d_i$$

where $\mathrm{d_i}$ represents the normalized score of data sources, $\mathrm{wt_i}$is the weight assigned, and $\mathrm{M}(d)$is the calculated score of d.

In a vehicle tracking problem, Sun et al (2004) applied the linear data fusion approach with vehicle signature, velocity, color and platoon traversal time as follows.

$$\mathrm{M} = \mathrm{wt_s} \sum_j d\left(s_u^j, s_d^j\right) + \mathrm{wt_c} \sum_j d\left(c_u^j, c_d^j\right) + \mathrm{wt_v} \sum_j d\left(v_u^j, v_d^j\right) + \mathrm{wt_p} \sum_j d(p_u, p_d)$$

where $\mathrm{wt_s}$ is the fusion weight for vehicle signature, $\mathrm{wt_c}$ is the fusion weight for vehicle color, wt_v is the fusion weight for vehicle velocity, and $\mathrm{wt_p}$ is the fusion weight for platoon traversal time. In addition, $\mathrm{d}(\mathrm{x_u}, \mathrm{x_d})$ represents the feature distance between upstream and downstream site for feature vector j, in general.

In our study, the linear data fusion approach is applied to WIM and signature data with a range of [0, 5] for the fusion weight in the corridor level tracking. For the network-wide tracking, multiple data sources are linearly integrated with different ranges ([0, 5]) of fusion weights.

Final SWBM for corridor level and network-wide tracking model including prior, feature selection, and linear data fusion weights are summarized as follows.

Corridor model:

$$p\left(\mu_{ij} \left| W_{ij}^{1}, W_{ij}^{2}, .., W_{ij}^{m}, V_{ij}^{1}, V_{ij}^{2} \ldots V_{ij}^{l}\right.\right) = \frac{p\left(W_{ij}^{1}, W_{ij}^{2}, .., W_{ij}^{m}, V_{ij}^{1}, V_{ij}^{2} \ldots V_{ij}^{l} \left| \mu_{ij}\right.\right)}{\sum_{i=1}^{I} p\left(W_{ij}^{1}, W_{ij}^{2}, .., W_{ij}^{m}, V_{ij}^{1}, V_{ij}^{2} \ldots V_{ij}^{l} \left| \mu_{ij}\right.\right)}$$

$$= \frac{wt_W \cdot \Pi_{m=1}^{M} [\, \alpha_m \cdot IG_m \cdot p\left(W_{ij}^{m} \middle| \mu_{ij}\right)] \cdot wt_V \cdot \Pi_{l=1}^{L} [\alpha_l \cdot p\left(V_{ij}^{l} \middle| \mu_{ij}\right)]}{\sum_{i=1}^{I} wt_W \cdot \Pi_{m=1}^{M} [\, \alpha_m \cdot IG_m \cdot p\left(W_{ij}^{m} \middle| \mu_{ij}\right)] \cdot wt_V \cdot \Pi_{l=1}^{L} [\alpha_l \cdot p\left(V_{ij}^{l} \middle| \mu_{ij}\right)]}$$

where w_m and w_l represent linear fusion weights for WIM and signature, IG_m indicates information gain for WIM attributes and W_{ij} , V_{ij} denotes WIM and signature feature vectors of target vehicle j and candidate vehicle i, respectively

Feature label (α_m, α_l) will be explained in the next chapter of feature selection and weighting method (Chapter 3.4).

Network-wide model:

$$p\left(\mu_{ij} \left| V_{ij}^{1}, V_{ij}^{2} \ldots V_{ij}^{l}\right.\right) = \frac{\pi^{k}\left(\mu_{ij}\right) p\left(V_{ij}^{1}, V_{ij}^{2} \ldots V_{ij}^{l} \left| \mu_{ij}\right.\right)}{\sum_{i=1}^{I} \pi^{k}\left(\mu_{ij}\right) p\left(V_{ij}^{1}, V_{ij}^{2} \ldots V_{ij}^{l} \left| \mu_{ij}\right.\right)}$$

$$= \frac{wt_K \cdot \{ \Pi_{k=1}^{K} \pi^{k}\left(\mu_{ij}\right) \} \cdot wt_L \cdot \{ \Pi_{l=1}^{L} \alpha_l \cdot p\left(V_{ij}^{l} \middle| \mu_{ij}\right) \}}{\sum_{i=1}^{I} wt_K \cdot \{ \Pi_{k=1}^{K} \pi^{k}\left(\mu_{ij}\right) \} \cdot wt_L \cdot \{ \Pi_{l=1}^{L} \alpha_l \cdot p\left(V_{ij}^{l} \middle| \mu_{ij}\right) \}}$$

where π^k represents prior for data source k and its linear fusion weights is w_k

3.2 Feature Processing

3.2.1 Feature preparation

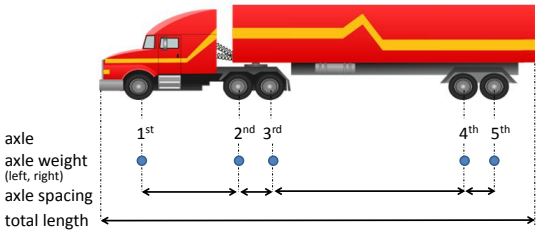
The truck tracking algorithm extracts vehicle attributes from WIM data and inductive loop signatures when the tracking is implemented between WIM stations and from inductive loop signatures only when the tracking is implemented between ILD sites. By comparing vehicle attributes, the target vehicle that detected at downstream site, is matched to candidate vehicles, which were detected at the upstream location at some time earlier than the target vehicle.

The WIM attributes include gross vehicle weight (GVW), total vehicle length, time duration a vehicle occupies a loop, axle spacing, and axle weight (see Figure 3.3(a)). Trucks have different numbers of axle spacing and weight attributes that vary by axle configuration, therefore up to 17 WIM attributes can be obtained from a truck. For trucks with six or more axles, only the first to fifth axle spacing and weight values were used as attributes. Using these attributes, WIM feature vectors are calculated with the difference between WIM attributes m of a target vehicle j and those for a corresponding candidate vehicle i. Simply, feature vectors are attribute differences between vehicle pairs.

Signature feature vectors represent the differences between the signature attributes of the target and candidate vehicles, where the attributes include 50 normalized magnitude measurements obtained at evenly distributed points along the temporal axis of the inductive signature as shown in Figure 3.3(b). Therefore, there are a total of 67 possible features from WIM and signature data.

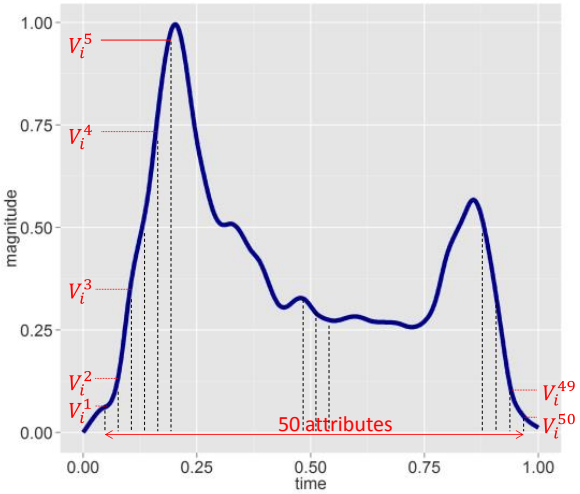
$$W_{ij}^{m} = W_{i}^{m} - W_{j}^{m}, m \in \{1, .., up\ to\ 17\}$$

- Duration ($m = 1$)
- Vehicle length ($m = 2$)
- Gross vehicle weight ($m = 3$)
- Axle spacing ($m = 4, \dots, 7$)
- Axle weight ($m = 8, \dots, 17$)



(a) WIM feature vectors (W_{ij}^{m})

$$V_{ij}^{l} = V_{i}^{l} - V_{j}^{l}, \quad l \in \{1, \dots 50\}$$



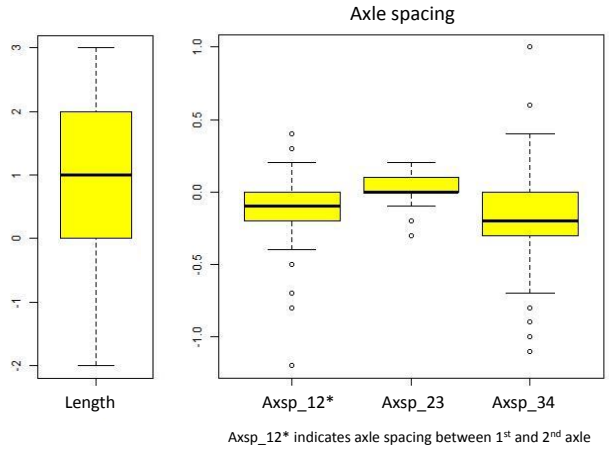
(b) Signature feature vectors (V_{ij}^{l})

Figure 3.3 Feature vectors: (a) WIM feature vectors (b) Signature feature vectors

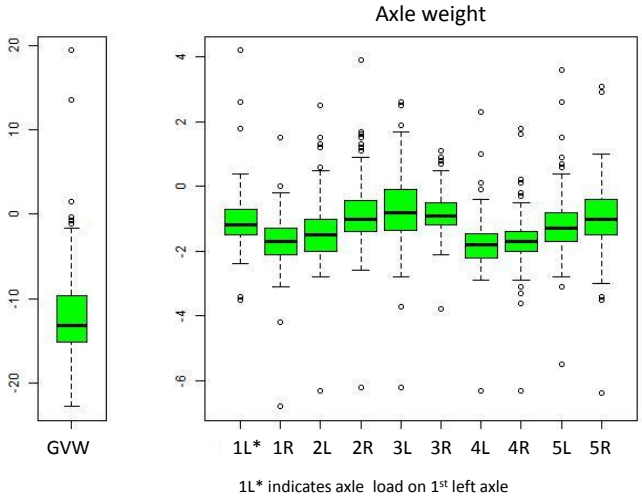
3.2.2 Feature Noise Elimination

Due to the sensor calibration and vehicle's lateral position, WIM and signature sensors can possess systematic or random noises. In order to minimize the impacts from these noises,

outlier features were initially identified and removed before estimating feature distribution. Figure 3.4 illustrates axle length and load differences between upstream and downstream detector measurements observed from the same vehicles.



(a) Axle length measurements



(b) Axle load measurements

Figure 3.4 Measurement errors in (a) axle length and (b) load attributes

From the visual inspection, outliers in the measurements are detected in both length and load attributes. This study used modified Z-scores method for outlier elimination (Iglewicz and Hoaglin, 2016). First, the Z-score of an observation is defined as:

$$\mathrm{Z_i} = \frac{Y_i - \bar{Y}}{s}$$

where $\bar{Y}$ represents a sample mean and s represents a sample standard deviation, respectively.

The modified Z-score is defined as

$$M_\mathrm{i} = \frac{0.6745\ (x_i - \hat{x})}{MAD}$$

where MAD denotes the median absolute deviation (= $\mathrm{median}\left(\mathrm{Y_i} - \hat{Y}\right)$) and $\hat{x}$ represents the median.

It is recommended that modified Z-scores with an absolute value of greater than 3.5 should be labeled as outliers. Therefore, in our study, the modified Z-scores of all 67 features were individually examined, and the attributes with greater than 3.5 modified Z-scores were identified as outliers and removed from the training feature sets.

Along with the outlier elimination, systematic measurement errors are corrected by WIM auto-calibration process. The constant differences in WIM data between upstream and downstream sites can be identified by matched vehicles. This constant difference is assumed as

a systematic error in the WIM measurement. Therefore, all the measures from one WIM site can be adjusted based on those from another WIM site from the same vehicles. Details will be discussed in Chapter 3.5.

3.3 Feature Distribution

After correcting systematic and random errors from the detection system, matched and mismatched feature distributions were estimated for Bayesian modelling. This study considered two possible approaches for feature distribution estimation.

First, non-parametric feature distribution is considered. Non-parametric probabilistic density function assumes that a suitably smooth density exists but the density does not represent a particular form of underlying distribution. The simplest example is a histogram. The histogram counts the number of observations falling into each bin so that the area of each bar of the histogram could be proportional to the number of observations falling into the corresponding interval. Although the histogram approach is very simple to implement, it has significant limitation from its dependence on the size of bin. In other words, if the bin size is excessively big or small, the histogram would be too smoothened or too fluctuated, and this could cause sensitivity issues on density estimations.

Another commonly used method in the non-parametric approximation of probability density function is Kernel density estimation. The Kernel density estimator is expressed as follows.

$$f_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

$$K(x) \geq 0, \int_{-\infty}^{\infty} K(x)dx = 1, K(x) = K(-x)$$

where $K(\cdot)$ is the kernel, h is a smoothing parameter called the bandwidth, x_i is samples, x is data center (average), and n is the number of samples

Similar to the histogram, the estimator highly depends on the bandwidth. For example, small bandwidth produces a more wiggly function and large bandwidth provides a more smoothed function. Kernels with different bandwidth are compared in Figure 3.5.

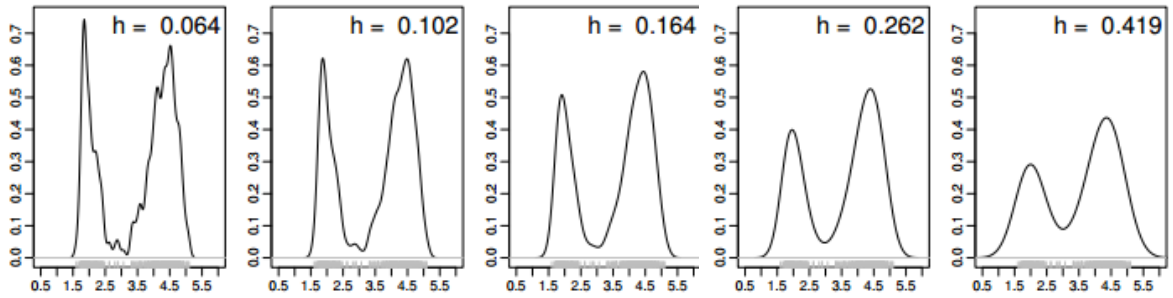


Figure 3.5 Kernel density estimation with different bandwidth[2]

Second method is a fitted distribution estimation approach. Features are assumed to be fitted to a particular distribution. Given the training data, parameters (θ) that determine distributions are chosen to maximize the likelihood function. In our study, features are shown

[2] Source: Computational statistics

normally distributed thus a probability density function for a continuous feature (x) was fitted to a Gaussian distribution with mean μ and standard deviation σ.

Again, the posterior probability ($\pi(\theta|D_n)$) of Bayes classifier can be expressed as follows.

$$\pi\ (\theta|D_n) = \frac{L(D|\theta)\ \pi(\theta)}{\int L(D|\theta)\ \pi(\theta) d\theta} \text{ where } \pi(\theta) \text{ is prior}$$

Let $D = \{x_1, x_2, x_3, \ldots x_n\}$ be the independent training samples and the likelihood function of $L(D|\theta)$ is expressed as follows.

$$L(D|\theta) = \prod_{k=1}^{n} p\ (x_k|\ \theta\ )$$

Since the distribution is assumed as Gaussian, $\theta = \mu$, the likelihood function is expressed as follows.

$$L(D|\theta) = \prod_{i=1}^{n} \exp \frac{\left\{ - \frac{(x_i - \mu)^2}{2\sigma^2} \right\}}{\sqrt{2\pi\sigma}}$$

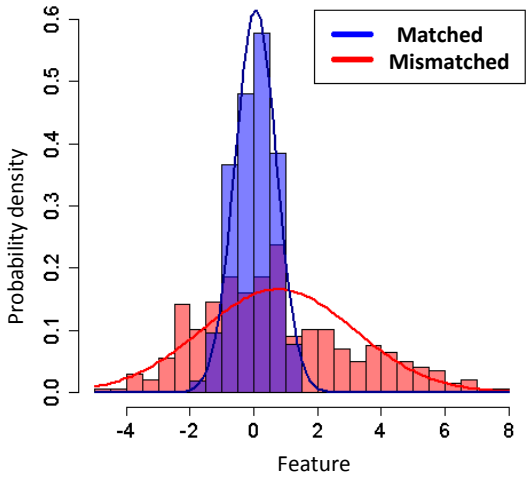
$$\propto \exp \frac{\left\{ - \sum_{i=1} \frac{(x - \mu)^2}{2\sigma^2} \right\}}{\sigma^n}$$

$$\propto \exp \frac{\left\{ -\frac{(\, n\mu^2 - 2n\bar{x}\mu + \sum_{i=1} x_i^2 \,)}{2\sigma^2} \right\}}{\sigma^n}$$

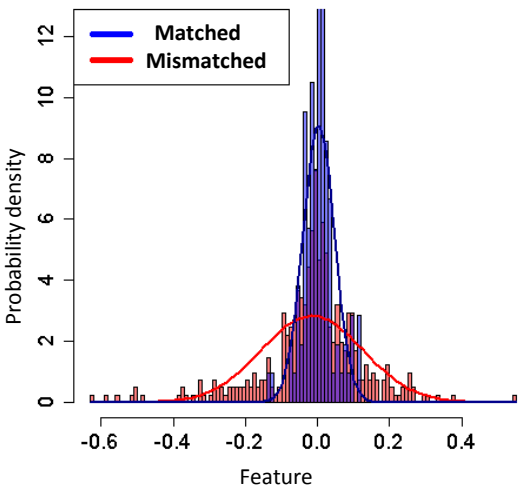
$$\propto \exp \frac{\left\{ -\frac{[n\,(\mu\, -\, \bar{x}\,)^2 + s^2]}{2\sigma^2} \right\}}{\sigma^n}$$

where $\bar{x}$ represents empirical mean and $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$

This study first identified matched pairs for all target vehicles in the training dataset. WIM and signature features' Euclidean distances between target and candidate vehicles were calculated for determining initial matched and mismatched pairs, and the vehicle pairs with minimum total feature distances were identified as matched pairs. The matched pairs were determined as either a true match or a mismatch using groundtruth data generated from visual matching of vehicles from collected side-fire images. Gaussian distributions were estimated for all 67 feature vectors from the pairs of true matches and mismatches. Figure 3.6 shows the estimated matches and mismatches Gaussian distributions overlaid with density histograms. Even though there are particular features that are more distinct between the match and mismatch distributions, in general, the matched pair distributions tend to have small variances centered at zero while the mismatched pair distributions typically show large variances.

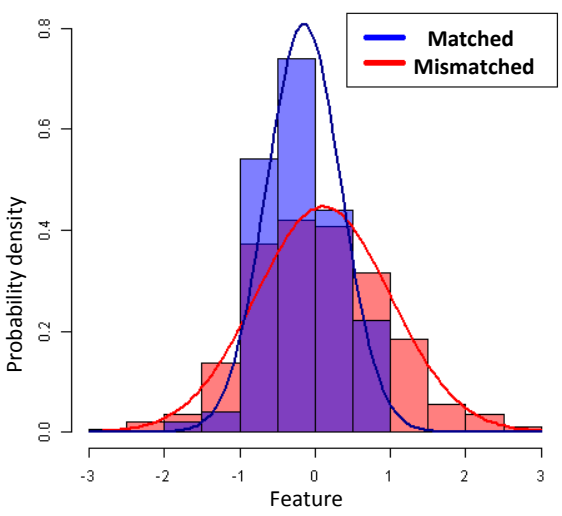


WIM feature Gaussian distribution
(fourth left axle weight)

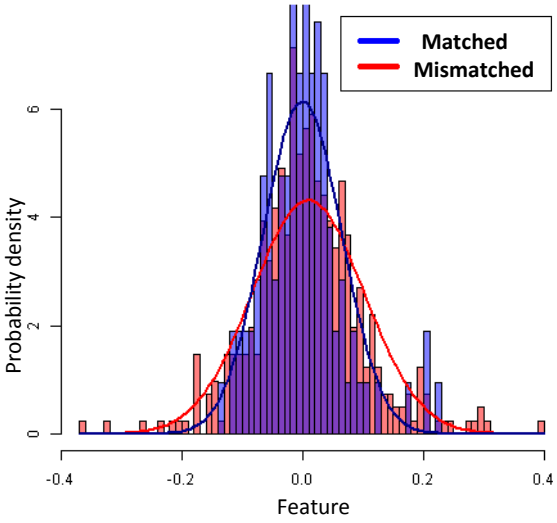


Signature feature Gaussian distribution
(13th feature)

(a) More distinct features



WIM feature Gaussian distribution
(first left axle weight)



Signature feature Gaussian distribution
(46th feature)

(b) Less distinct features

Figure 3.6 Gaussian distributions for matched and mismatched cases

3.4 Feature Selection and Weighting Method

Based on the Gaussian distributions of each feature, influence of the feature in distinguishing matched and mismatched pairs was further investigated using a statistical test and a clustering method. Features were categorized into four groups based on their variance and statistical differences in matched and mismatched distributions. For example, if the matched and mismatched distributions of a specific feature were not statistically different, the feature would not be expected to perform a significant role in determining vehicle matches. Conversely, if distributions were statistically different and the matched distribution had smaller variance than the mismatched distribution, the feature could play a significant role in distinguishing matches from mismatches.

3.4.1 Kolmogorov-Smirnov test

A Kolmogorov-Smirnov (KS) hypothesis test (Kim and Jennrich, 1973) was used to examine the statistical difference of the distributions. Two sample distributions are compared using the empirical distribution functions or cumulative fraction functions with a null hypothesis as follows.

H_o : Two data samples come from the same distribution

H_1 : Two data samples do not come from the same distribution

Since the KS test examines data only using the relative distance of distributions, it does not require any assumption on the distributions.

In Figure 3.7, KS statistics (D_{mn}) is $S_m(x) - S_n(x)$ where S_m and S_n are the empirical distribution of two samples. In our study, KS test is used to examine if the matched and mismatched distributions are statistically different at the 5% significance level.

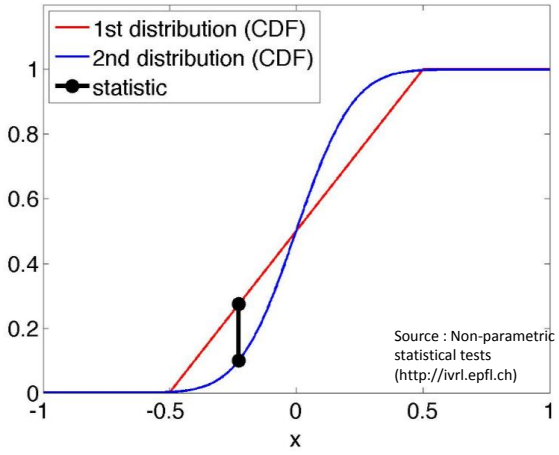


Figure 3.7 Kolmogorov-Smirnov (KS) hypothesis test

3.4.2 K-means Clustering

K-means clustering (Hartigan and Wong, 1979) was applied to categorize features based on their distribution characteristics. As an unsupervised classification algorithm, K-means solves clustering problems with k centroids where one centroid represents a cluster. The objective function is expressed as follows.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left|\left| x_i^j - c_j \right|\right|^2$$

where $\left|\left| x_i^j - c_j \right|\right|^2$ is a given distance measure between a data point x_i^j and the cluster center c_j, n represents total data point and k represents a given number of cluster. Figure 3.8 shows the illustration of five (K=5) clustering results.

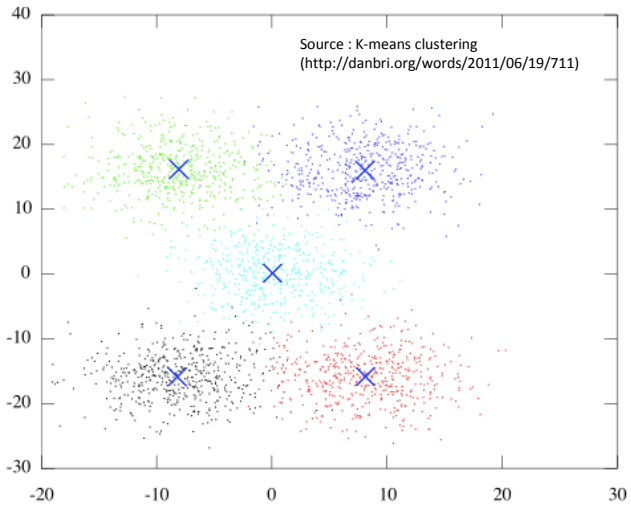


Figure 3.8 K-means clustering

In our study, k is defined as two, corresponding to matched and mismatched pairs.

3.4.3 Feature Labeling

Consequently, WIM and signature features were assigned one of four labels – critical, significant, insignificant, and inverse – in descending order of their ability to distinguish matches from mismatches. Figure 3.9 shows examples of feature distributions with four different labels. Different weights were assigned to these categories in the Bayesian tracking algorithm where critical feature distributions possessed the highest weight and inverse feature distributions had the lowest weight. These feature labels were separately analyzed for two different truck types corresponding to single-unit and multi-unit trucks.

1) Critical:

- matched and non-matched distributions are statistically different and
- matched distribution has small but non-matched distribution has large variance

2) Significant:

- matched and non-matched distributions are statistically different and
- matched and non-matched distributions have large variance

3) Insignificant:

- matched and non-matched distributions are not statistically different

4) Inverse:

- matched and non-matched distributions are statistically different and
- matched distribution has a larger variance than non-matched distribution

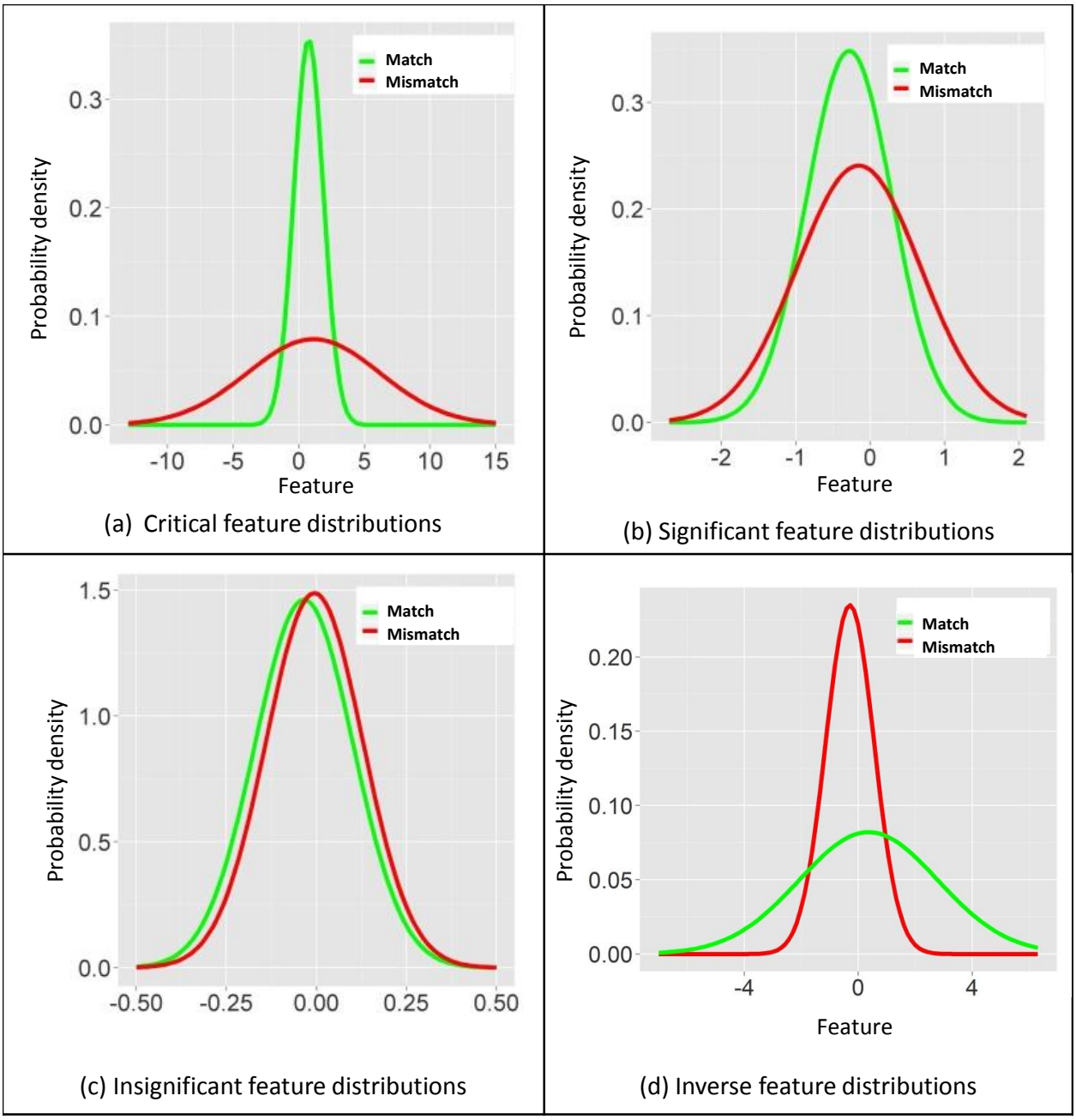


Figure 3.9 Comparisons of feature influence for multi-unit trucks

3.5 Implementation of Truck Tracking Algorithm

The tracking framework consists of two models. For every detected target truck detected at downstream, the first model determines the matched vehicle at upstream. However, since the tracking algorithm is implemented over a long-distance corridor or a large network that may contain multiple entries and exits, only a small fraction of trucks might be expected to traverse both stations. Therefore, after the first model assigns the match prediction for target trucks, there is a need in the second model to determine if the matched pair obtained is indeed a true match or a mismatch.

As shown in Figure 3.10, there are possible six cases in declaring match and mismatch to the target vehicle. The first four examples present the case where the target vehicle is observed at both upstream and downstream locations. In these cases, only if the matching model finds the correct match and the filtering model does not filter the vehicle, the matching can be successful (case A). Even though the matching model finds the correct match, this pair can be declared as a missing from the filtering model (case A-2). Or, if the first model finds the pair as a mismatch, incorrect outcomes will be obtained regardless of the filtering model outcomes (case B-1 and B-2). Case C-1 and C-2 shows the matching results when the target vehicle is observed at only downstream location. In this case, a declared matched pair from the first model should be filtered as a missing for the successful match (case C-1).

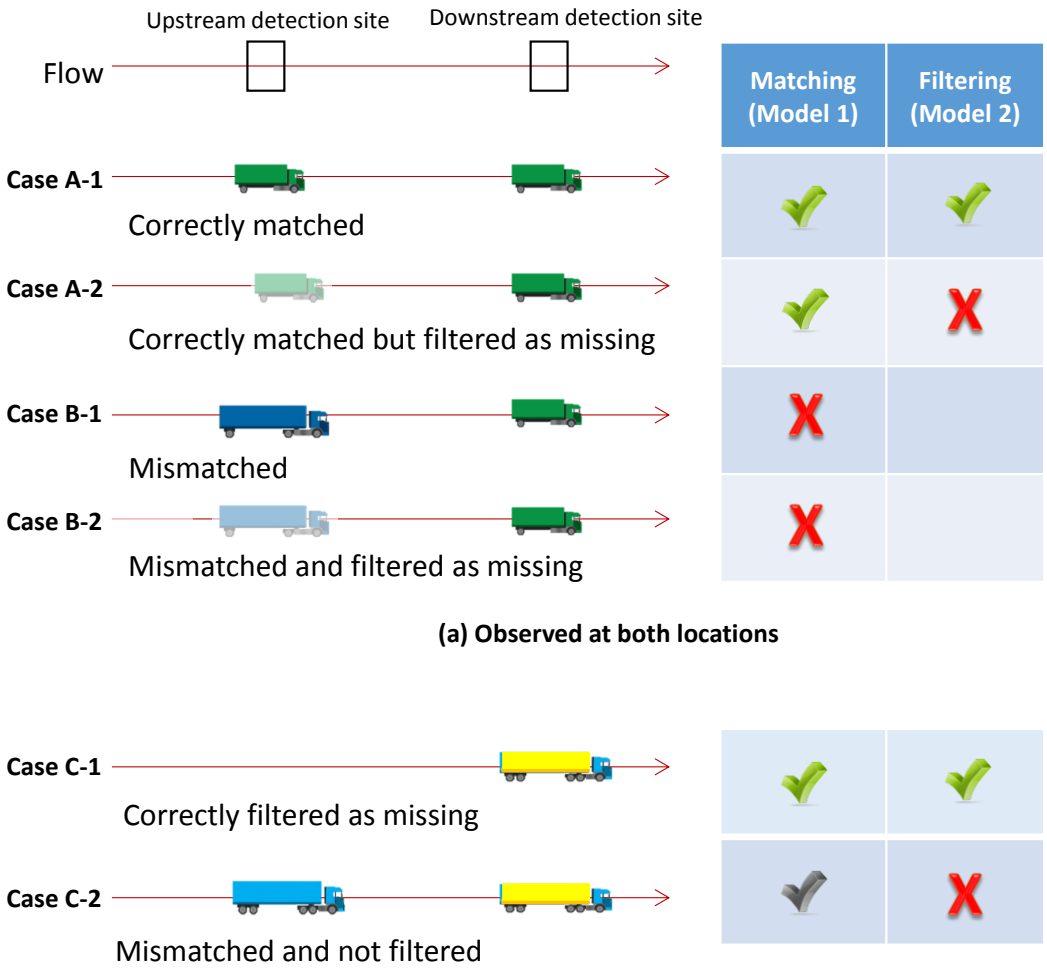


Figure 3.10 Cases of tracking results

In the field implementation of this tracking model, multiple sub-steps are performed prior to and within the matching and filtering models. A step-by-step implementation is described in this section.

Step 1. Candidate Vehicle Set Preparation

The first step searches candidate vehicle sets at upstream location based on their travel time and physical attributes, called search space reduction.

Step 1.1 Search space reductions

The search space reduction algorithm first designates a temporal window based on the travel time limits between upstream and downstream detection sites as follows:

$$\{minimum\ travel\ time\ (tt_{ff}) \leq Temporal\ window < maximum\ travel\ time\ (e.g., \alpha_{ts} * tt_{ff}\ )\}$$

where α_{ts} can be any value greater than 1

Trucks that pass the upstream location within the temporal window are selected as candidate matches for the target truck. Previous research studies in vehicle tracking have applied a relatively narrow travel time boundary to reduce computation costs (Abdulhai and Tabib, 2003; Oh and Ritchie, 2003; Dion and Rakha, 2006). However, in this case, the use of a narrow temporal window would introduce a significant negative impact on successful population matching because a narrow window eliminates many true matches. In this study, travel time at the speed limit between two locations was defined as the minimum travel time. The maximum travel time was determined as three times the minimum travel time for this study. The model is not particularly sensitive to the travel time boundary and this relaxed travel time window was designed to accommodate significant changes in traffic conditions between distant detector locations. To accommodate longer travel times of long-haul trucks with rest breaks at intermediate locations, an even more relaxed travel window may be considered. However, a

greatly increased search space could result in a high level of mismatches and not be practically feasible.

In a corridor level tracking, physical attributes of vehicles such as axle loading and spacing can be directly compared between target and candidate vehicles to further reduce the candidate vehicle set. Therefore, after the temporal window was applied to find candidate vehicles, WIM attributes of the candidate vehicles were compared with that of the target vehicle. Flexible upper and lower thresholds were applied to the WIM attribute window in consideration of systematic errors of the WIM measures (Prozzi and Hong, 2007). Through these steps, each target vehicle, j, was associated with its final candidate vehicles set, i ($i \in 1, \dots, I$).

$$\{W_j^m - \alpha_j^{lb} \leq W_j^m < W_j^m + \alpha_j^{ub}\}$$

where W_j^m indicates WIM attribute m for target vehicle j , and α_j^{lb} and α_j^{ub} are the lower and upper bounds of the window for W_j^m, respectively.

Step 1.2 Signature transformation

Loop sensitivity and a vehicle's lateral position over the sensor may affect the quality of the inductive signature waveform obtained from each vehicle. The purpose of signature transformation is to reduce these effects prior to extracting the vehicle signature tracking features. First, signature normalization and imputation were applied to individual signatures as shown in Figure 3.11. Normalization and imputation steps allow the signature features to be extracted over a common scale from zero to one.

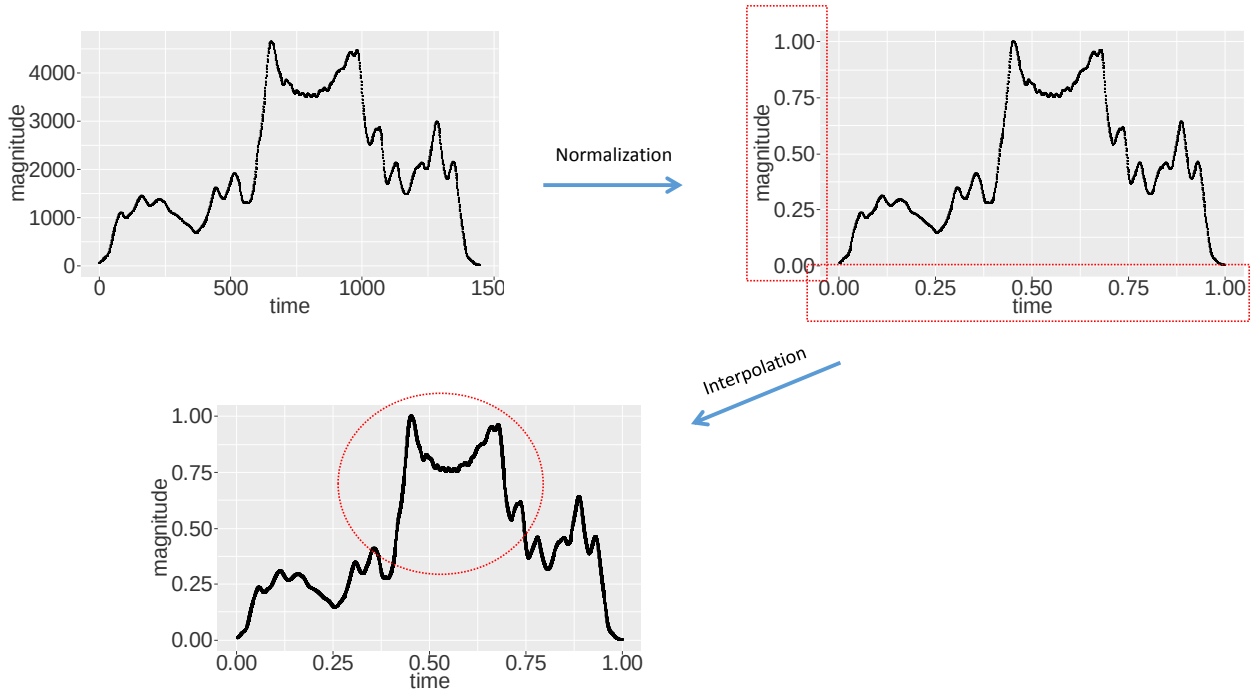


Figure 3.11 Signature normalization and imputation steps

Second, the candidate vehicle's signature, S_i, is horizontally and vertically transformed to fit the target vehicle's signature. The candidate signature is iteratively shifted and stretched until the minimum difference between the signature pairs is obtained. Consequently, this step minimizes differences of correctly matched pairs of signatures and further distinguishes matched pairs from mismatched pairs. Figure 3.12 a d Figure 3.13 illustrates procedures of the signature transformation. Figure 3.14 and Figure 3.15 show two comparison examples of correctly matched and mismatched signatures, respectively. Although the transformation step reduces the signature differences for both cases, the differences (i.e., distances) between the mismatched pairs are much higher compared with the matched pair after the transformation processes.

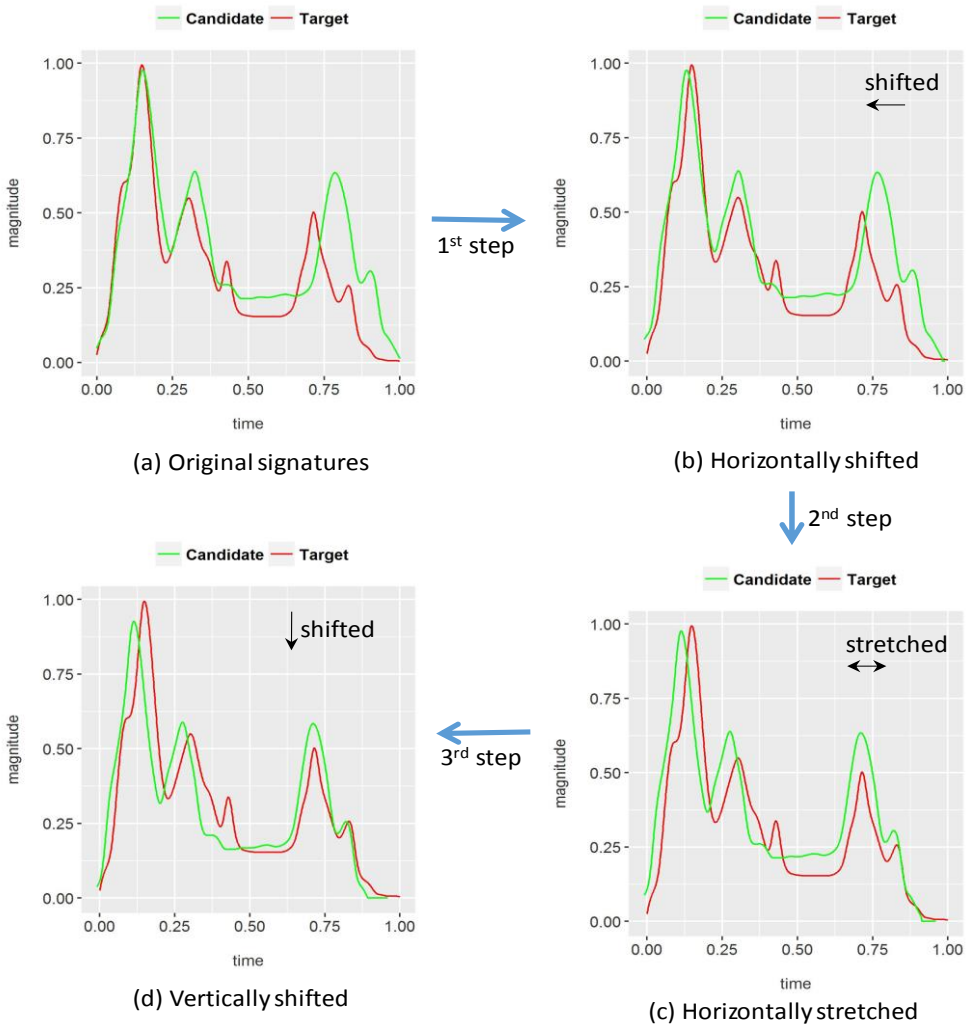


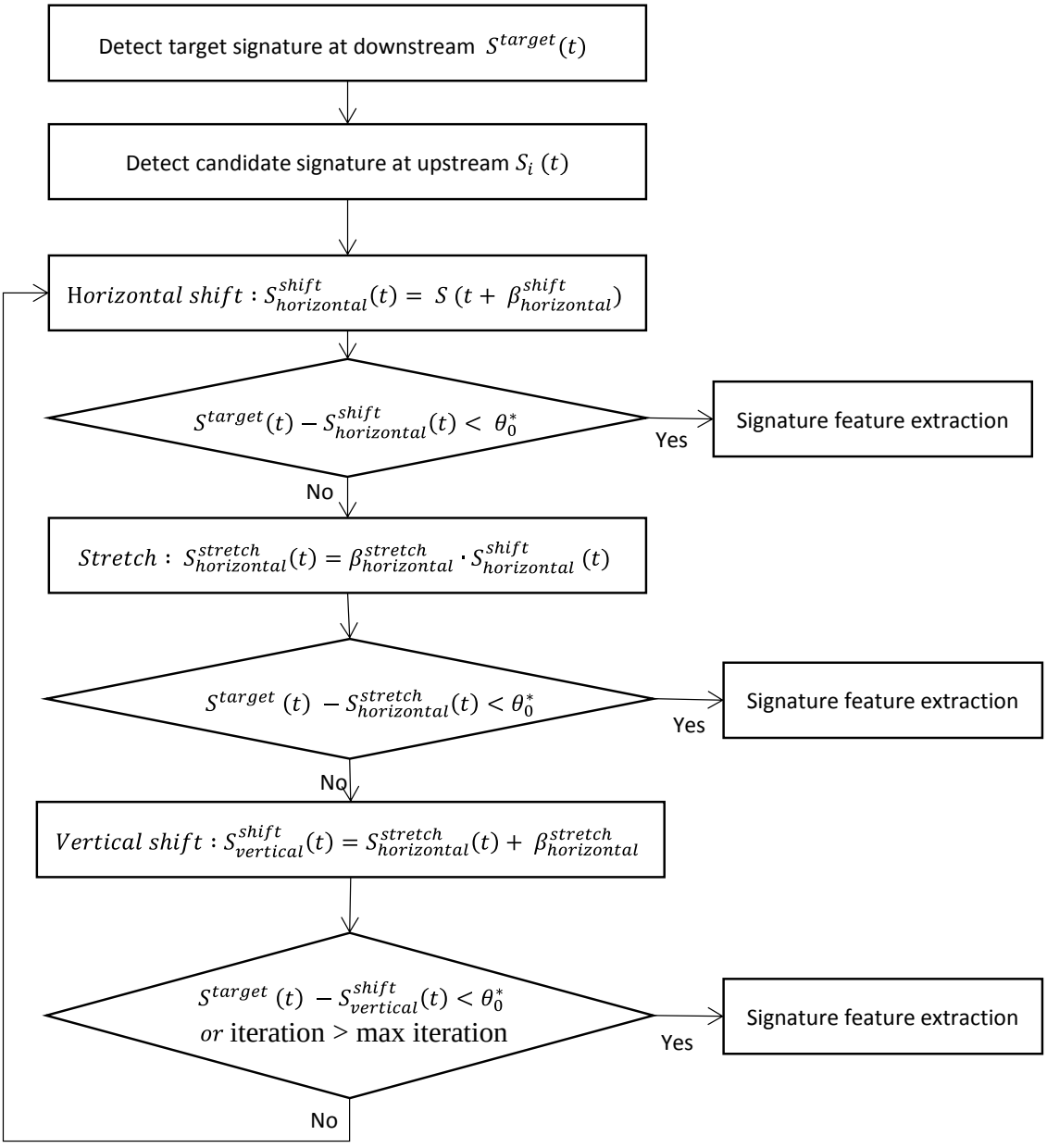
Figure 3.12 Horizontal and vertical signature transformation steps

Horizontal shift step: $S_i^{h.shift}(t) = S_i(t + \beta_{horizontal}^{shift})$ $[e.g., -0.20 \leq \beta_{horizontal}^{shift} \leq 0.20]$

Horizontal stretch step: $S_i^{stretch}(t) = S_i^{h.shift}(\beta^{stretch} \cdot t)$ $[e.g., 0.8 \leq \beta^{stretch} \leq 1.2]$

Vertical shift step: $S_i^{v.shift}(t) = S_i^{stretch} + \beta_{vertical}^{shift}$ $[e.g., -0.20 \leq \beta_{vertical}^{shift} \leq 0.20]$

where $\beta_{horizontal}^{shift}$, $\beta_{vertical}^{shift}$, and $\beta^{stretch}$ represent the horizontal shifting, vertical shifting, and horizontal stretching coefficient, respectively, t represents the time



θ^*_o represents the pre-determined threshold

Figure 3.13 Procedures for signature transformation

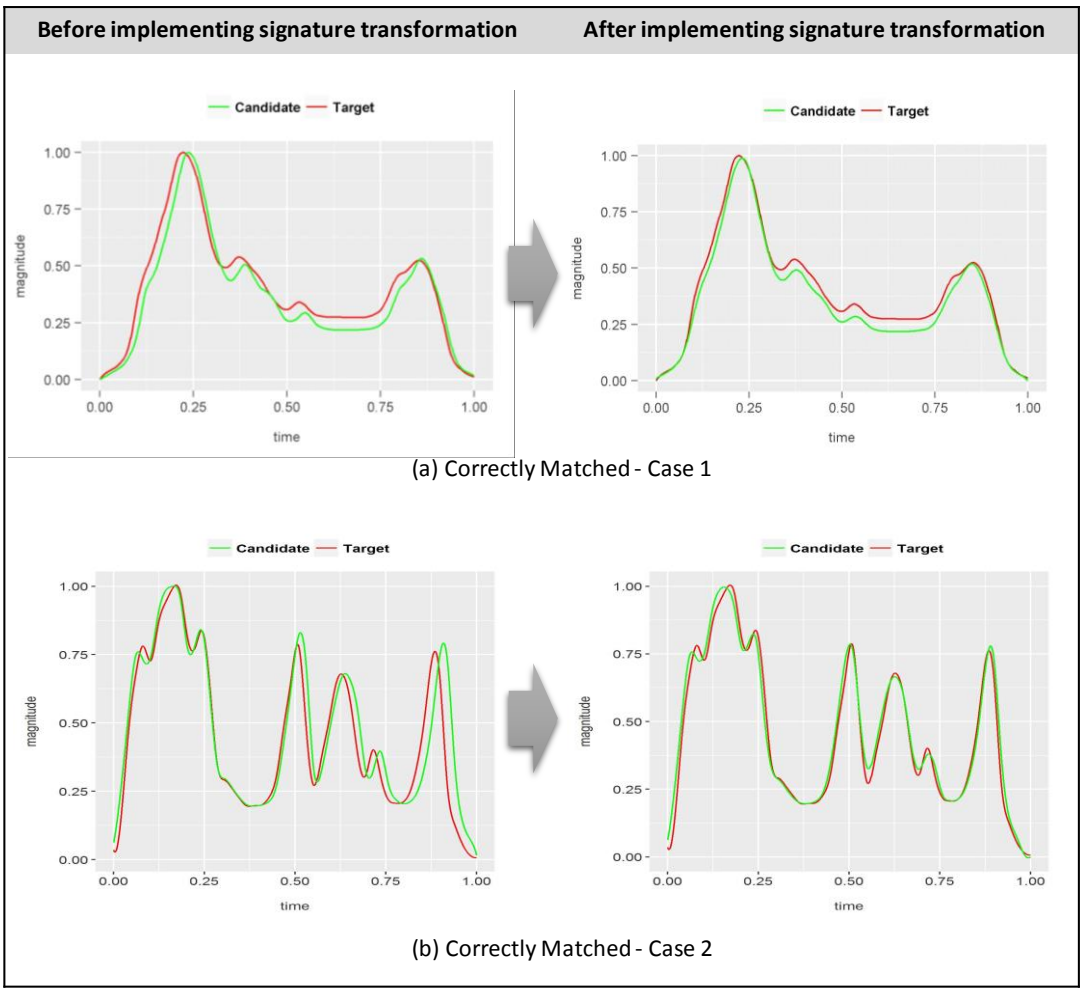


Figure 3.14 Signature transformation results for correctly matched pairs

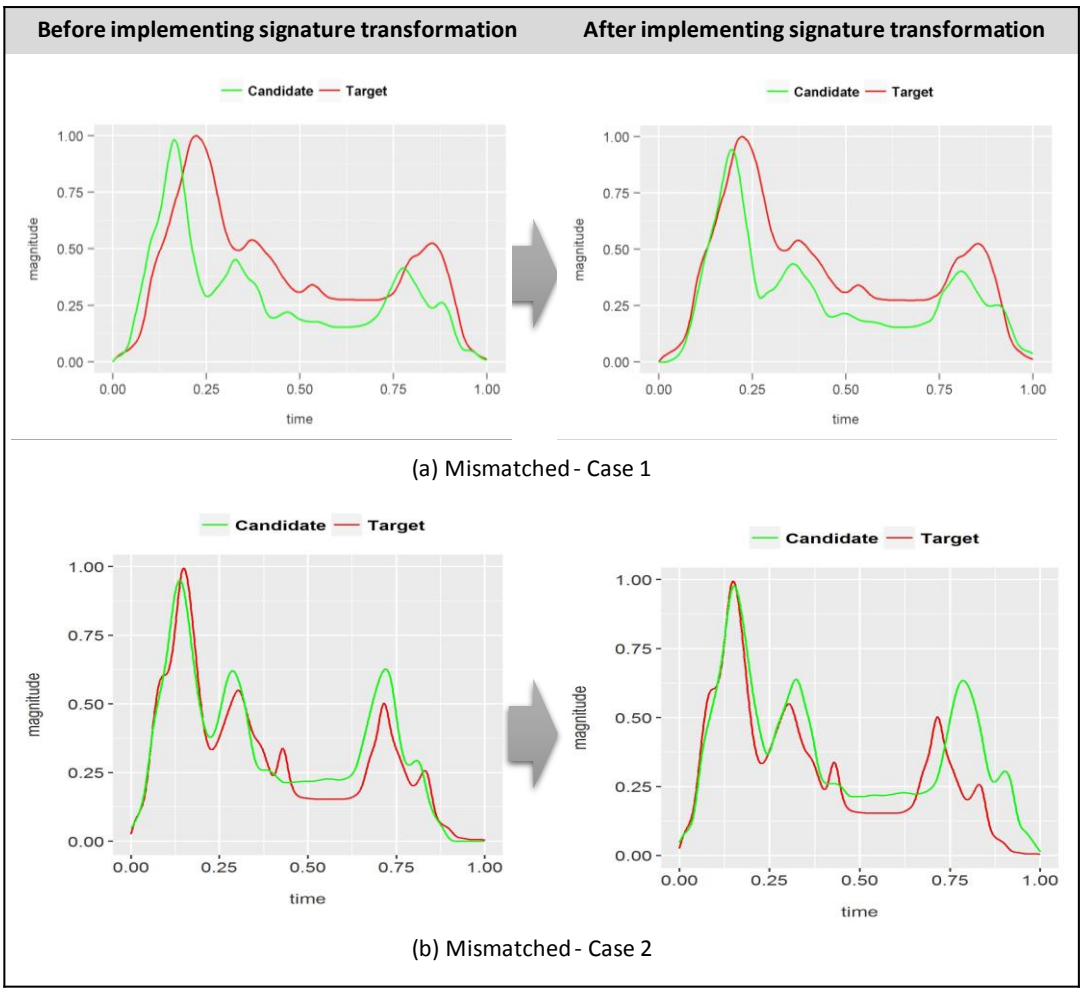


Figure 3.15 Signature transformation results for mismatched pairs

Step 1.3 WIM data auto-calibration

WIM data from the same vehicle may differ across sites due to systematic or random errors in WIM measures (Papagiannakis et al., 2008). These WIM errors, particularly systematic errors, may significantly affect the vehicle tracking process because the WIM data collected at up- and down-stream stations from the same vehicle can be very different. Therefore, a WIM data auto-calibration process was proposed to adjust for systematic errors and to minimize impacts from random noise.

Ideally, the average difference of each WIM attribute from the matched vehicles can be used to adjust calibration variances in WIM attributes. In field implementations where variances of WIM measures between sites are unknown, potential matched pairs can be defined by the signature distances. Only the signature pairs with smaller distances can be assigned as the potential match and used for the calibration.

Step 2. Vehicle Matching

Step 2.1 Best match search using Selective Weighted Bayesian Model (SWBM)

The best matched vehicle was identified among candidate vehicles obtained via search space reduction (step 1.1). Typically, a vehicle pair with the least feature distance can be chosen as a match. However, due to the similar physical attributes and signature shapes from the trucks with the same body configuration, the extended Bayes rule with selected weighted features was applied to distinguish matched pairs. This addresses the issue where a few erroneous measurements from noisy features may result in incorrect matches if the matches are determined using minimum feature distance alone as the criterion. However, these features would have less impact on matching decisions in the proposed method since the joint probabilities are considered using the Bayesian approach with multiple independent feature distributions.

For the corridor tracking where both signature and WIM data were used, vehicle features from these two detection systems were separately incorporated into the Bayesian model to find a match μ_{ij}. Assuming that the prior probability for every pair is held constant, a conditional

probability of a match was expanded to the joint probabilities of the WIM and signature attributes with the best linear data fusion weights

For the network-wide tracking, signature data are used as vehicle features. Prior probability is estimated based on supplementary data sources such as GPS trajectory and body configuration estimates. Linear data fusion approach is also applied to all data sources to better integrate multiple sources.

For both models, each candidate vehicle's matching probabilities were estimated based on matched feature distributions and prior probability, and the candidate vehicle with the highest matching probability was selected as the match.

Step 2.2 Matched pair filtering

Step 2.1 identified a match for every target vehicle. However, in an open system not every target vehicle would have arrived from the upstream detection site. Hence, some matched pairs need to be filtered as mismatches. In this step, two posterior probabilities, corresponding to true match, $p(\theta_{ij} = 1)$, and mismatch, $p(\theta_{ij} = 0)$, were estimated to the matched pairs and the result with the higher posterior probability was chosen as the final decision.

Probability of true match:

$$p(\theta_{ij} = 1 \,|X_{ij}) = \frac{p(\theta_{ij} = 1) \cdot \gamma_m \cdot \Pi_{m=1}^{M} p(X_{ij}^m | \theta_{ij} = 1)}{(\theta_{ij} = 1) \cdot \gamma_m \cdot \Pi_{m=1}^{M} p(X_{ij}^m | \theta_{ij} = 1) + (\theta_{ij} = 0) \cdot \gamma_m \cdot \Pi_{m=1}^{M} p(X_{ij}^m | \theta_{ij} = 0)}$$

Probability of mismatch:

$$p(\theta_{ij} = 0 \,|X_{ij}) = \frac{p(\theta_{ij} = 0) \cdot \gamma_m \cdot \Pi_{m=1}^{M} p(X_{ij}^{m} | \theta_{ij} = 1)}{(\theta_{ij} = 1) \cdot \gamma_m \cdot \Pi_{m=1}^{M} p(X_{ij}^{m} | \theta_{ij} = 1) + (\theta_{ij} = 0) \cdot \gamma_m \cdot \Pi_{m=1}^{M} p(X_{ij}^{m} | \theta_{ij} = 0)}$$

Where γ_m represents feature selection and weighting coefficient for attribute m, X represents matching features

A vehicle pair was confirmed as a match if its probability of a true match was greater than that of a mismatch. Conversely, a vehicle pair was confirmed as a non-match if its probability of a mismatch was greater than that of a true match.

Step 2.3 Eliminating duplicate matching pairs

From the Step 2.2, every target vehicle either declares its matching pair or identified as a missing. However, since the algorithm finds the matched pair based on the probability, the same candidate vehicles could be matched multiple times to different target vehicles. Therefore, as the last step, the multiply matched vehicles are eliminated. For example, if duplicates are found from two matched pairs, matching probability between pairs are compared and the pair with higher matching probability is declared as the final match. Then, the target vehicle from another pair goes back to the step 2.1 to select the second highest candidate as a match and proceeds to the filtering step. This duplicate removal step is implemented iteratively until no duplicates are found from matched pairs or all the best matching vehicles are exhausted in the candidate sets.

4. Corridor Tracking Model

This chapter introduces a corridor tracking model through a full integration of the two advanced detection technologies – advanced inductive loop detectors and WIM sensors – along the truck corridor. Since these two sensor technologies collect complementary vehicle attributes such as trucks' physical attributes from WIM sensors, and inductive signatures from inductive loop sensors, vehicles are able to be more accurately and effectively identified. Further advantages of the proposed tracking algorithm in its utilization of existing detection infrastructure are: (i) additional in-pavement retrofits are not required in vehicle tracking; (ii) private identifiable information is not collected when vehicles are tracked across sites; (iii) and the system provides collateral benefits to a recently developed advanced truck classification model using the fusion of the same detector technologies (Hernandez et al., 2016). The combination of these two systems has the potential to yield detailed tracking of commercial vehicles by their body configuration and industrial affiliation to yield a comprehensive data source of detailed truck activity.

4.1 Model overview

The truck tracking algorithm matches the vehicles between two detection stations using axle configuration and inductive loop signature attributes. After a truck is detected at a downstream location, the tracking algorithm subsequently searches for candidate trucks that have passed the upstream location at some earlier time. When a truck is detected at a downstream location, its axle configuration is determined by the WIM controller and categorized into either a

single-unit or multi-unit truck. Thus the algorithm only searches for candidate trucks within the same category. Figure 4.1 outlines the detailed steps of the vehicle tracking algorithm. Details in the tracking algorithm were discussed in Chapter 3.

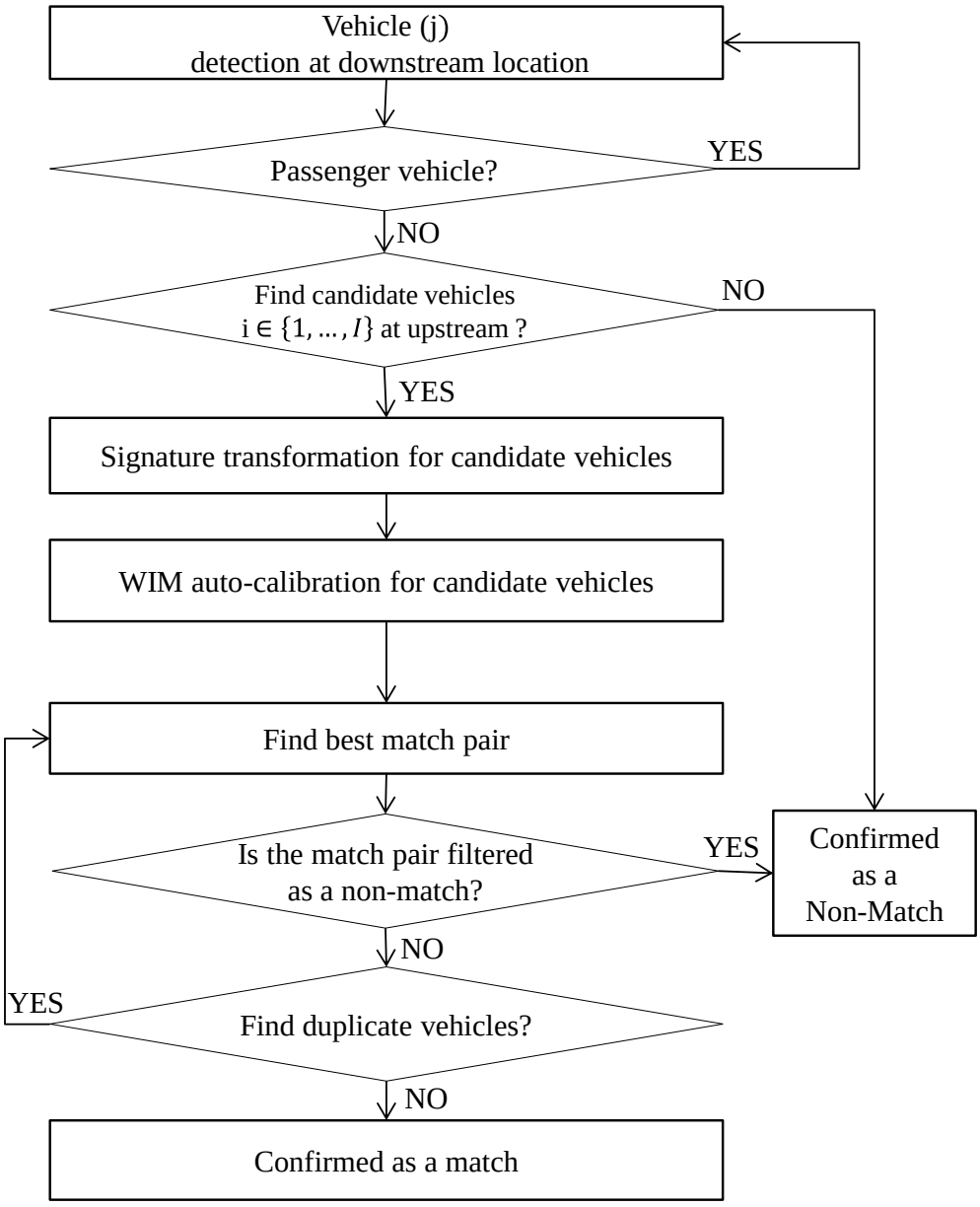


Figure 4.1 Steps of the corridor tracking implementation

4.2 Data

Truck axle and inductive signature data together with side-fire still images for each passing vehicle were collected at two WIM sites for model development (see Figure 4.2). WIM data, inductive signatures, and still images were stored in a database and manually integrated during an initial groundtruth data processing phase. Two WIM sites along the Interstate 5 freeway at San Onofre (upstream) and Leucadia (downstream) in California were selected for model development and validation (see Figure 4.3). The distance between these two locations is 26 miles, spanning two major freeway intersections and 17 entrance and exit ramps. WIM and signature data from the two outermost southbound (slow) lanes were collected. As shown in Table 4.1, a total of 471 trucks at the upstream station and 1,038 trucks at the downstream station were collected on January 9th and 10th, 2013. Most of the population comprised of multi-unit five-axle trucks. Only 14 percent of total trucks passed both upstream and downstream stations. These are referred to as common trucks as shown in Table 4.2.

Figure 4.2 Data collection site images

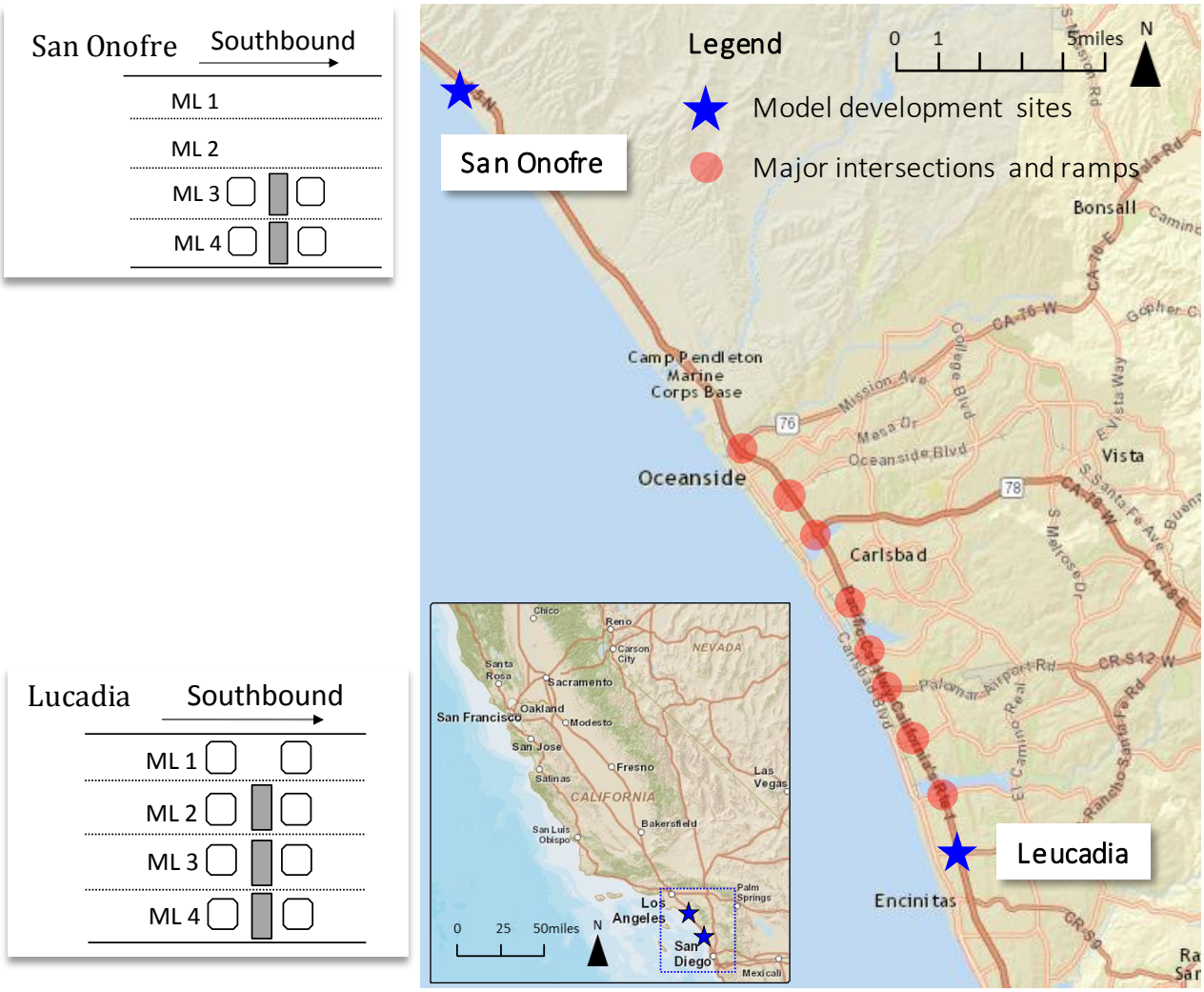


Figure 4.3 Data collection sites for corridor level tracking

Table 4.1 Data collection site description

Site Location	Distance	Collection Dates
San Onofre (SB I-5) to Leucadia (SB I-5)	26 miles	Jan. 09, 2013 (Testing) Jan. 10, 2013(Training)

Table 4.2 Corridor tracking testing and training data

Dataset	# of Trucks Collected at Downstream	% of Tractor-Trailer at Downstream	# of Trucks Collected at Upstream	% of Tractor-Trailer at Upstream	# Common Vehicles	% of Common Vehicle from the Total Vehicle Detected at Upstream
Testing	471	60%	50	86%	50	11%
Training	1038	65%	159	84%	159	15%
Total	1509	63%	209	84%	209	14%

4.3 Performance Measures

Five evaluation measures were used to analyze the performance of this model: Closed System Matching Rate – CSMR, two measurements of Open System Matching Rates – OSMR1 and OSMR2, and two measurements of False Matching Rates – FMR1 and FMR2. Open system refers to the tracking corridor that include multiple intersections and ramps, thus vehicles can enter or exit the tracking corridor during a tracking process. Therefore, every target vehicle would not be observed at upstream WIM site in open system. However, in close system, no exit and entry exists is assumed in the corridor, therefore all target vehicles are observed at upstream WIM site. The algorithm finds matches through two sequential processes – best match search and filtering. The filtering step is only needed in an open system deployment where there is a likelihood that a true match does not exist. Therefore, the performance measurements were designed to evaluate the model under both closed and open systems. To measure the model performance in a closed system, matching results from the best match step were compared only against vehicles that traversed both the upstream and downstream sites. Hence, CSMR is defined as the number of correctly matched pairs obtained from the best match search step divided by the total number of common vehicles, expressed as a percentage.

The OSMRs measure matching accuracy in an open system after the additional filtering step is implemented to predict if the best matches obtained are indeed true matches. OSMR1 represents the percent of correctly matched common vehicles that have passed both WIM sites in an open system, and OSMR2 is defined as the percent of correctly matched vehicles out of the total matches that are declared by the model. The FMRs indicate the matching inaccuracy of the model. Therefore, high CMRs and low FMRs are generally desirable. The FMR1 represents the percentage of error matches among the common vehicles, while FMR2 is defined as the number of error matches divided by the total number of vehicles observed at the downstream location. Care should be taken when evaluating different performance measures since the matching results are highly related to the geometric and traffic characteristics of the study corridor.

$$\text{CSMR} = \frac{\text{Initial Correct matches} *}{\text{Common vehicles}}$$

$$\text{OSMR1} = \frac{\text{Correct matches}^{**}}{\text{Common vehicles}}$$

$$\text{OSMR2} = \frac{\text{Correct matches}}{\text{Total matches}^{***}}$$

$$\text{FMR1} = \frac{\text{Error matches}}{\text{Common vehicles}} = \frac{\text{Total matches} - \text{Correct matches}}{\text{Common vehicles}}$$

$$\text{FMR2} = \frac{\text{Error matches}}{\text{Total vehicles}} = \frac{\text{Total matches} - \text{Correct matches}}{\text{Total vehicles}^{****}}$$

* Initial correct matches represent the number of vehicles that the algorithm matched correctly at the best match search step

**Correct matches represent the number of vehicles that the algorithm matched correctly

***Total matches represent the total number of vehicle matches declared by the algorithm

****Total vehicles represent the total number of vehicles detected at the downstream location

4.4 Results

Table 4.3 shows the matching results from the training and testing dataset with data fusion weights. Various combinations of weights were investigated to identify the best fusion weights from WIM and signatures. The matching rates where solely WIM or signatures were used are provided for comparison purposes.

The OSMRs were quite low when only WIM or signature features were used for tracking. However, the CSMR ranged from 71 percent to 100 percent for all cases regardless of the data fusion combination. This result indicates that the models using standalone WIM or signature data may be adequate in a closed corridor where the entire truck population is expected to traverse both the upstream and downstream detector sites, but are deficient in open system applications.

In the test data set, the OSMR1 obtained for multi-unit trucks was 77 percent. The OSMR2, which represents the matching accuracy of the multi-unit trucks that were declared as matches by the model, was 85 percent. These high values of OSMR1 and OSMR2 together show

that the model developed in this study was able to maintain a high accuracy of matches without sacrificing the proportion of vehicles tracked. The false matches of these vehicles were 14 percent out of the common vehicles (FMR1) and 4 percent in the total vehicles (FMR2).

Single-unit trucks in the test dataset showed OSMR1 and OSMR2 accuracies of 71 percent and 63 percent, respectively with FMR1 and FMR2 measurements of 43 percent and 3 percent, respectively. A likely reason for higher observed FMR1 from the single-unit trucks is that these vehicles possess fewer distinct physical attributes since they tend to be shorter in length and possess fewer axles. The single-unit trucks also tend to travel shorter distance, which may lower the matching accuracies. In addition, data for one of the lanes at the San Onofre site was incompletely captured, which may have contributed to a lower accuracy.

Table 4.3 Matching results of corridor level tracking

Dataset		Data Fusion Description	Fusion Weight		Number of trucks				Matching Rate				
			WIM	SIG*	Total	Com-mon	Total Match	Correct Match	CSMR	OSMR1	OSMR2	FMR1	FMR2
Test Data Set	All Trucks	Best combination	3	1	243	50	47	38	**88%**	**76%**	**81%**	**18%**	**4%**
		WIM only	1	0			42	10	76%	20%	24%	64%	13%
		SIG only	0	1			56	32	90%	64%	57%	48%	10%
	Multi-Unit Trucks	Best combination	3	1	148	43	39	33	**88%**	**77%**	**85%**	**14%**	**4%**
		WIM only	1	0			18	5	72%	12%	28%	30%	9%
		SIG only	0	1			47	28	91%	65%	60%	44%	13%
	Single-Unit Trucks	Best combination	3	1	95	7	8	5	**86%**	**71%**	**63%**	**43%**	**3%**
		WIM only	1	0			24	5	100%	71%	21%	271%	20%
		SIG only	0	1			9	4	86%	57%	44%	71%	5%
Train Data Set	All Trucks	Best combination	3	1	795	159	164	114	**87%**	**72%**	**70%**	**31%**	**6%**
		WIM only	1	0			140	23	72%	14%	16%	74%	15%
		SIG only	0	1			193	81	83%	51%	42%	70%	14%
	Multi-Unit Trucks	Best combination	3	1	526	134	124	93	**87%**	**69%**	**75%**	**23%**	**6%**
		WIM only	1	0			60	12	71%	9%	20%	36%	9%
		SIG only	0	1			136	62	82%	46%	46%	55%	14%
	Single-Unit Trucks	Best combination	3	1	269	25	40	21	**88%**	**84%**	**53%**	**76%**	**7%**
		WIM only	1	0			80	11	77%	44%	14%	276%	26%
		SIG only	0	1			57	19	88%	76%	33%	152%	14%

*SIG represents the inductive loop signature

4.5 Sensitivity analysis

During the tracking process, trucks could experience loading and unloading activity especially near port, intermodal rail facility, and warehouse area. Since the tracking model uses axle loading as a feature, this loading and unloading activity might significantly affect tracking performances. This section performs a sensitivity analysis on weight features to compare how the model performs when the axle loadings are included and not included in the feature set.

Figure 4.4 shows that 73 percent of mismatched pairs have the same body configurations from the corridor tracking. This high proportion of mismatch from the same body configuration indicates that similar waveform signatures would be generated from the same truck types, which results in lower performance in matching process.

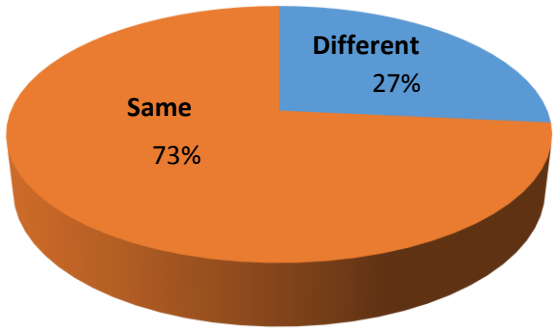


Figure 4.4 Identity of truck types from mismatched pairs

However, Figure 4.5 shows that these mismatched pairs showed quite different GVW compared to their vehicle length. The mismatched pairs from the test dataset of tractor-trailer units even showed 45 percent of weight differences between matched vehicles. On the contrary,

length differences between matched pairs are less than 10 percent. These results show that the GVW can play a significant role in distinguishing similar body types, which may be difficult to be differentiated by length or axle spacing measures.

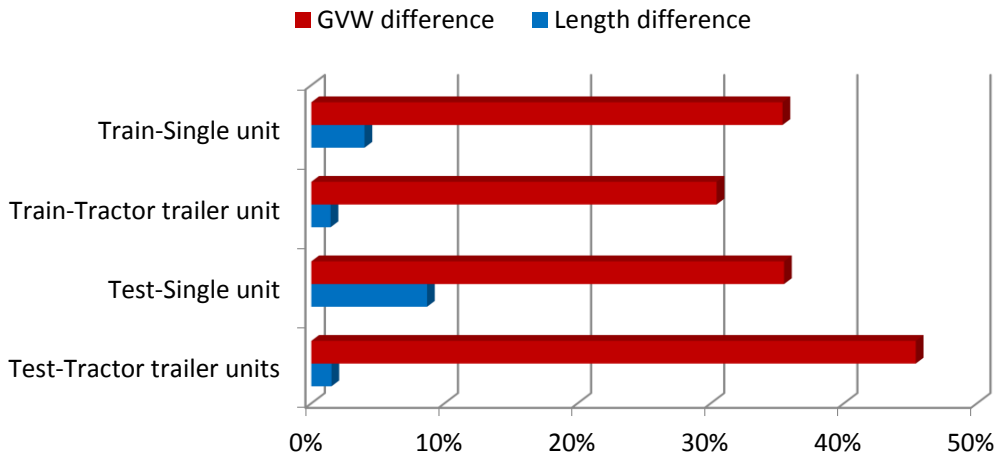
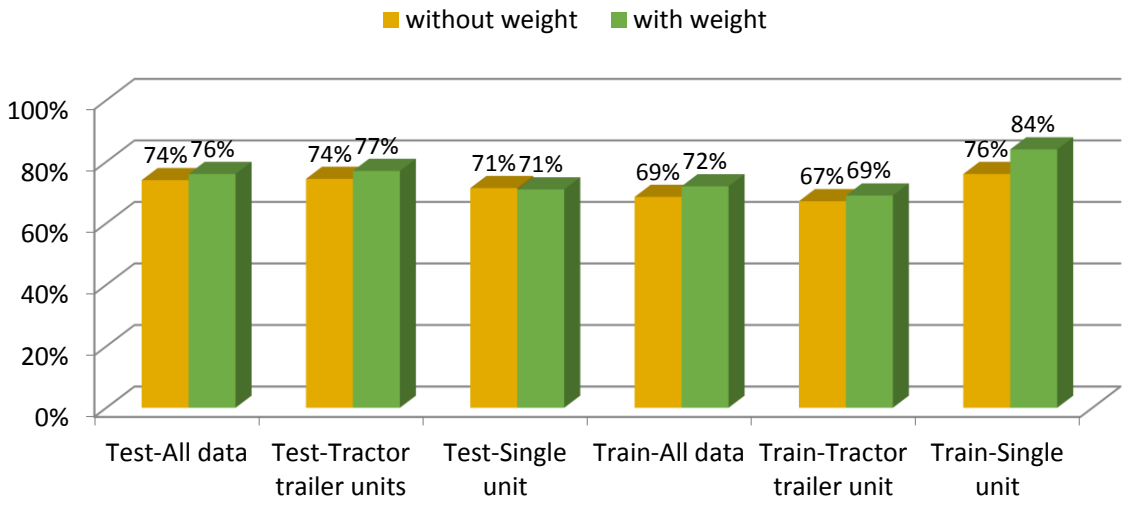
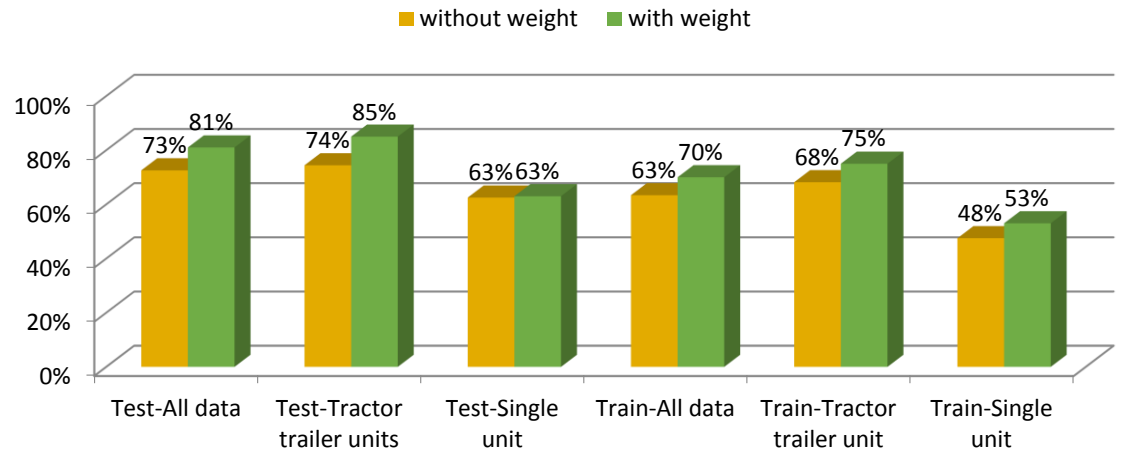


Figure 4.5 GVW and length differences from the mismatched pairs

Hence, the corridor tracking compared the tracking accuracy with and without the weight data in the feature sets. As shown in Figure 4.6, both OSMR1 and OSMR2 showed generally better performance when the GVW and axle loadings were used as features. The matching accuracies of OSMR2 rather than OSMR1 are improved when the weight measures were used in the tracking. These results indicate that the weight features effectively to filter missing pairs than other features.



(a) OSMR 1



(b) OSMR 2

Figure 4.6 Performance comparisons for weight features

4.6 Discussion

In this section, previous tracking approaches are compared to the proposed algorithm. The study by Jeng and Chu (2015) presented the most recent development in truck tracking using WIM and inductive signature data, with a focus on WIM calibration and travel time estimation

applications. The signature data were mainly utilized in a tracking process with proximity measures. WIM data were subsequently used in filtering matching vehicles as the last step of the tracking process. In addition, proximity measures were used to match vehicles so that a vehicle pair with the minimum distance between features was selected as a matching pair.

It should be noted that the performance of a tracking algorithm depends significantly on multiple factors such as vehicle type, tracking distance, freeway geometry (i.e., number of intersections or ramps), and total traffic flow. For example, if a tracking corridor contains fewer intersections, trucks will be more likely to traverse both upstream and downstream detection sites even though the tracking is performed over a very long distance, resulting in better tracking performance. In this regard, when performances of different tracking algorithms are compared, care should be taken especially if the tracking is implemented at different locations.

Hence, this section applied the tracking approaches from the previous studies to the same dataset that the proposed algorithm was used and compared matching performances. A total of three approaches were introduced: two approaches used proximity measures and one approach applied a Bayesian approach with equally weighted features as shown in Table 2. Since the proximity measures require a fixed parameter to declare matching pairs, two different fixed parameters were applied to each approach. The approaches of 1(a) and 1(b) only used signature features to find matching vehicles with different fixed parameters. The approaches of 2(a) and 2(b) were adopted from the study by Jeng and Chu (2015) therefore signature features were used to match vehicles and the WIM features were used as filtering purpose after determining potential matching pairs. For example, all the vehicles passing at downstream found their matching pairs using signature distances, and then the axle spacing and weight parameter were

used to filter mismatching pairs. Approach 2(a) specifically implemented the same fixed threshold method adopted from the study by Jeng and Chu (2015) for comparison purposes. After testing multiple values, thresholds from 2(b) were obtained as optimal threshold values for our dataset.

Table 4.4 Different tracking approaches

Approach	Features	Measures	Fixed parameter used?	Fixed parameter in matching
1(a)	Signature	Proximity measures	Yes	Sum of signature distances of matching pair < 2.0
1(b)	Signature	Proximity measures	Yes	Sum of signature distances of matching pair < 4.0
2(a)	Signature and WIM	Proximity measures	Yes	Axle spacing parameter* < 0.02 and Axle weight parameter** < 0.1
2(b)	Signature and WIM	Proximity measures	Yes	Axle spacing parameter* < 2.1 and Axle weight parameter** < 0.3
3	Signature and WIM	Bayesian approach with equally weighted features	No	-

* Axle spacing parameter differs by FHWA vehicle class. For example, axle spacing parameter for FHWA class 9 vehicles is defined as axle spacing between fourth and fifth axle divided by axle spacing between first and second axle. Details refer to Jeng and Chu (2015).

** Axle weight parameter is defined as first axle weight divided by GVW

Figure 4.7 illustrated correct matching rates of the five approaches and the proposed algorithm. Overall, the approaches with proximity measures showed lower correct matching rates than Bayesian approaches, and the performance significantly depends on the fixed thresholds. Although the approach 1(b) and 2(b) showed higher matching rate for OSMR1, OSMR 2 was lower than 20 percent. This is because too many vehicle pairs were declared as matching pairs, thus the proportion of correctly matching pairs among the total matching pairs

was very low. In contrast to the main applications of the previous studies, such as obtaining travel time estimation or detection site calibration, the most important aspect of a tracking process for path flow estimation is that total vehicles including not only true-matching pairs but also non-matching pairs should be accurately identified. In other words, for estimating path flow with tracked vehicles, every individual vehicle detected at a downstream station should be tracked and confirmed whether it passed at an upstream station or not. The proposed algorithm showed that the total vehicles were successfully tracked compared to the proximity measures.

Lower performance of proximity measures could be also caused by inaccurate calibration or different sensitivity in signature and WIM measures. The same vehicle could show varying vehicle length or signature shapes by detection locations if sensors have different sensitivity and calibration. However, when using the proximity measures, a few erroneous features could critically increase the total feature distance, which would result in lower matching accuracies. However, the proposed algorithm applied probabilistic approaches with selected features to ensure that the matching algorithm will be minimally affected by these exogenous factors. Further, signature transformation and WIM auto-calibration were added in the algorithm to reduce possible calibration errors or noise from the detection systems.

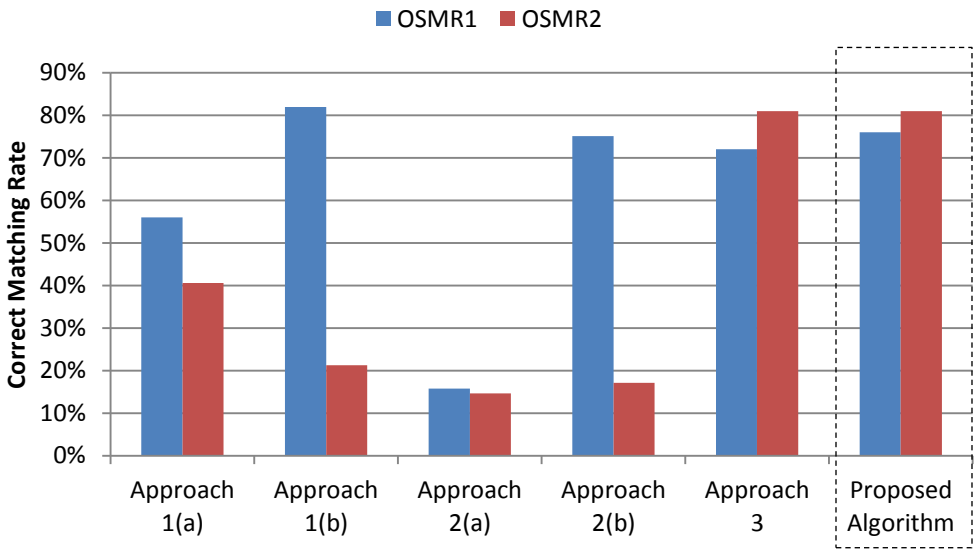


Figure 4.7 Comparison results

The proposed matching algorithm not only effectively captured the total vehicles tracked, but also achieved high matching rates (76% OSMR1 and 81% OSMR2). This is because the proposed algorithm simultaneously utilized both WIM and signature data through a Bayesian framework. The advantage of concurrently utilizing both detection systems is that the two systems can play complementary roles in distinguishing matching pairs. Therefore, trucks with the same trailer body configuration, which may have similar signature shapes, can be distinguished by their WIM data such as axle loads and gross vehicle weight. On the other hand, trucks with similar length and weight can be distinguished by their metallic composition obtained from the inductive signatures. Our results also showed that tracking methods solely using WIM or signature data only showed 24 percent and 57 percent matching accuracy (OSMR2) as opposed to 81 percent when WIM and signature data were both utilized (see Table 4.3). In addition, compared to results from the approach 3, the feature selection and weighting approaches applied to both the WIM and signature data enabled the Bayesian algorithm to

provide higher matching accuracy, which would be of great benefit for tracking in a corridor with heavy truck traffic.

4.7 Application

In this section, three applications of the corridor tracking results are presented, which includes truck monitoring along with the detailed truck body classification information, travel time estimation, and WIM calibration. With an integration of body configuration classification model, dynamic truck activity can be monitored with its detailed truck type and axle configurations. In travel time estimation and WIM calibration, more accurate travel time can be estimated between detection stations if a proportion of tracked vehicles that show higher matching probability is used.

4.7.1 Truck monitoring with detailed body classification

The tracking algorithm was implemented at a different location over a longer distance as a case study implementation (Figure 4.8). For the case study sites, a freeway corridor spanning two WIM sites separated by 65 miles and containing 6 major intersections in the California San Joaquin Valley was chosen. Truck axle and signature data were collected at two WIM sites at the Galt (downstream) and Keyes (upstream) stations and used for the case study implementation.

The data used in this analysis was obtained from June 17 to June 19, 2015 for 24 hours through the University of California, Irvine – Institute of Transportation Studies (UCI-ITS) Truck Activity Monitoring System (TAMS). The TAMS is an interactive web-based user interface that provides the spatial distribution of trucks by truck axle and body configurations. In

total, 7,003 multi-unit and 3,745 single-unit trucks were recorded at the downstream location (site Galt in Figure 4.8).

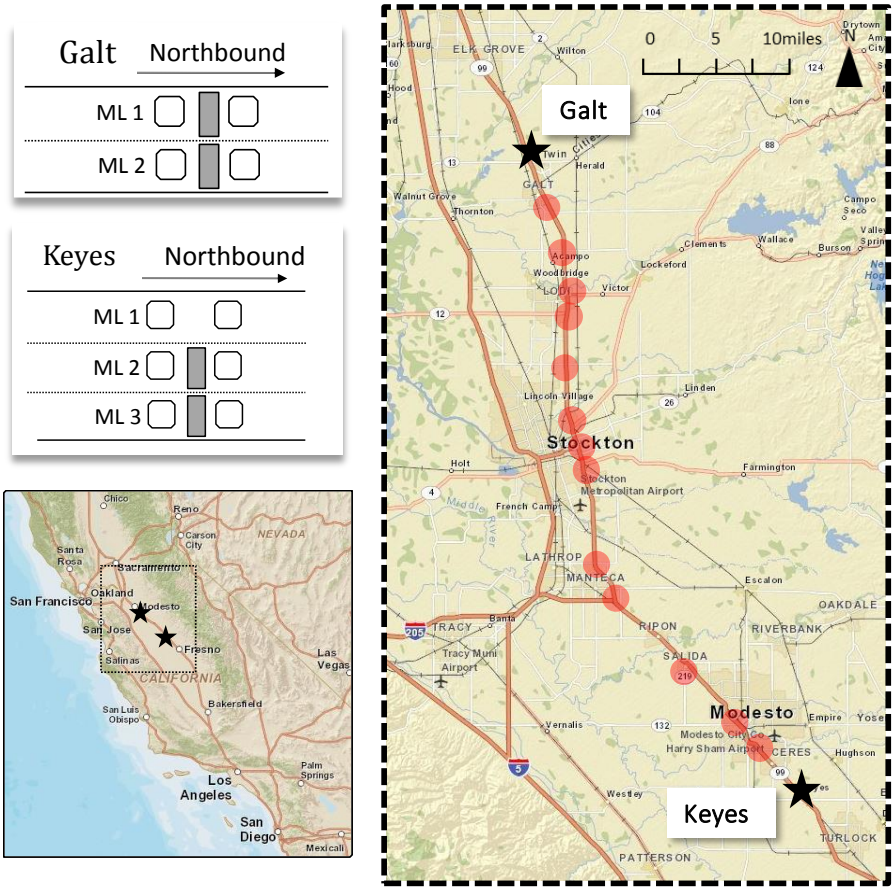


Figure 4.8 Case study site map for corridor level tracking

Table 4.5 Case study site description

Site Location	Distance	Site Description	Collection Dates	# of Trucks Collected at downstream WIM site	% of Tractor-Trailer at downstream WIM site
Keyes (NB SR-99) to Galt (NB SR-99)	65 miles	Case Study Implementation	June 17 -19, 2015	10,748	65%

Figure 4.9 shows the average detected volume and average matched rate by Time of Day (TOD). The TOD patterns for detected volume and matching rate show hourly variations in truck travel behavior. These TOD matching rates were estimated using the proposed tracking model, and the associated implications presented in this section are based on the model outcomes, assuming that the model provides accurate tracking results. The percentage match indicates the proportion of vehicles that traverse the entire corridor. To compare truck travel patterns by truck types, the truck classification model (Hernandez et al., 2016) was applied to estimate trailer body configurations. Two different trailer types, van and platform, for multi-unit 5 axle trucks (corresponding to FHWA class 9), were compared in hourly detected volume and matched proportions. For both trailer types, detected volumes were higher during the day-time than night-time. However, while the hourly matching rates of van types were constant for the whole day, platform type trucks showed higher matching rates during the night-time. Considering that the tracking rate does not depend on truck body configurations, these temporal variations in tracking rate between trucks with different trailer types can be inferred as attributable to dissimilar travel patterns, which can link to the industries and facilities they serve. Specifically, 10 to 20 percent of vans passed both WIM sites regardless of the time of day, which indicates 80 to 90 percent of van type trucks traveling over this freeway section presumably served local industries or used different routes. In addition, the platform-type truck higher matching rates during the night may imply that the platform trailers traveling during the night-time related highly to inter-regional rather than localized service compared to those traveling during the day. This application of the proposed algorithm demonstrates that it can provide valuable insights into truck travel patterns and industrial affiliations to yield a comprehensive truck activity data source.

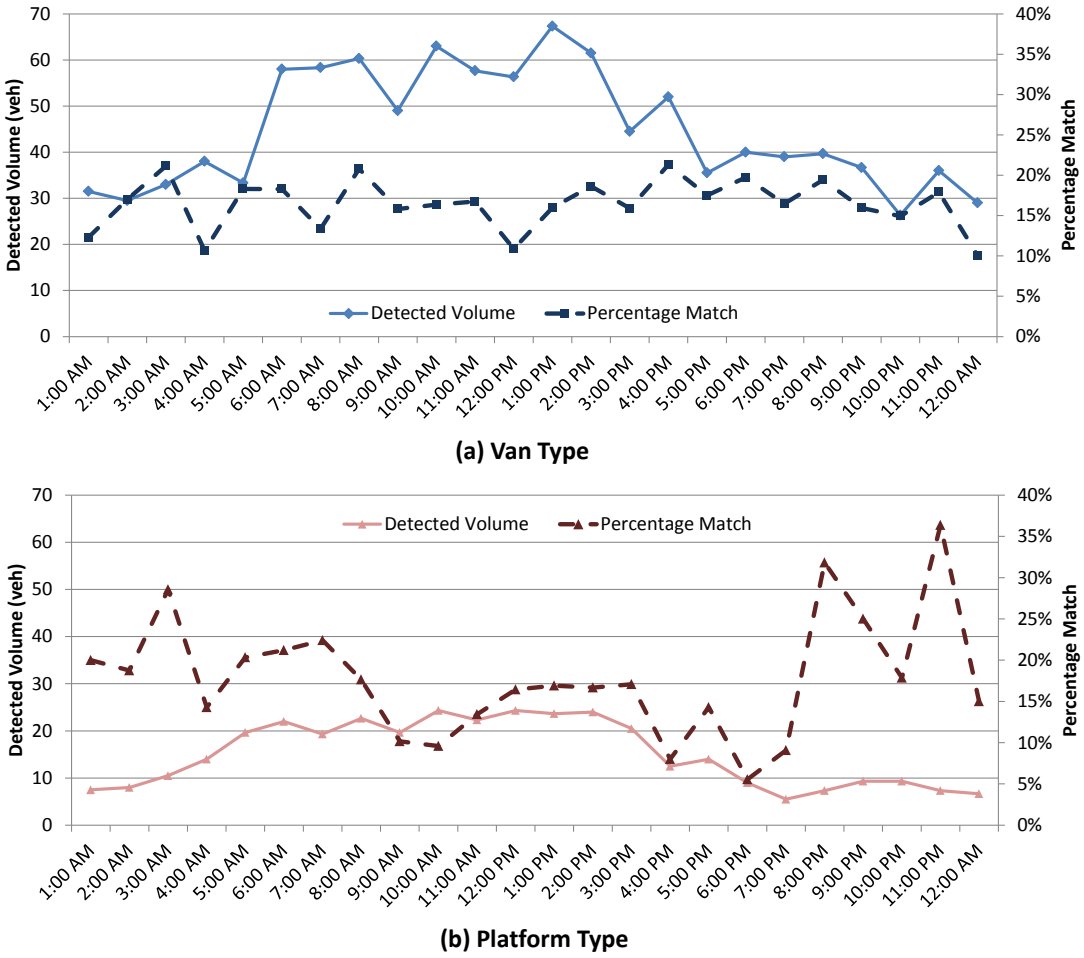


Figure 4.9 Comparison of detected and matched volume by truck type.

4.7.2 Travel time estimation

Using the matched vehicles from the tracking, travel time can be measured between two detection locations. However, different performances can be drawn by the applied samples since the timestamp between detection locations from the matched vehicles are used to estimate travel time. In other words, if the selected matched pairs which have higher probabilities in tracking model are used for travel time estimation, higher accuracy is expected than when using all the matched pairs. For travel time estimation, it is not important to maintain high proportion of

tracked population because a portion of samples with higher matching probability can provide more accurate travel time estimates.

To test travel time estimation performances, two different sets of samples are used, one with the full tracked population, and another with a portion of tracked vehicles with higher matching probabilities. Euclidean distance was used to select samples with higher matching probabilities where matched pairs with less than 2 of the total distance are considered as samples. Tracking results between San Onofre and Leucadia detection stations were used. Figure 4.10 and Figure 4.11 compare travel time estimations and MAPE (Mean Absolute Percent Error) from these two datasets. In the figures, average travel time for 5 minutes was illustrated as one time interval. MAPE represents average performance by comparing actual and estimated travel time in every time interval.

$$\text{MAPE} = \frac{\sum_n \frac{|Travel\ time\ (act)_n - Travel\ time\ (est)_n|}{Travel\ time\ (act)_n} * 100}{N}$$

where n represents time interval, act presents actual travel time from observations and est represents estimated travel time from tracking algorithm

The actual travel time ranges from 28 minutes to 34 minutes while estimated travel time was ranged from 25 minutes to 45 minutes when all the tracked vehicles were used. When the sampled tracked pairs were used, the range of estimated travel time was 24 minutes to 38 minutes. While the minimum and maximum MAPEs were 0% and 47.8 % with the median of

9.4% when all tracked pairs were used, the MAPEs were dropped to maximum of 32.9% and median of 6.9% when sampled tracked pairs were used (Table 4.6). This result confirmed that more accurate travel time can be estimated by sampled tracked pairs especially the maximum MAPE was significantly reduced.

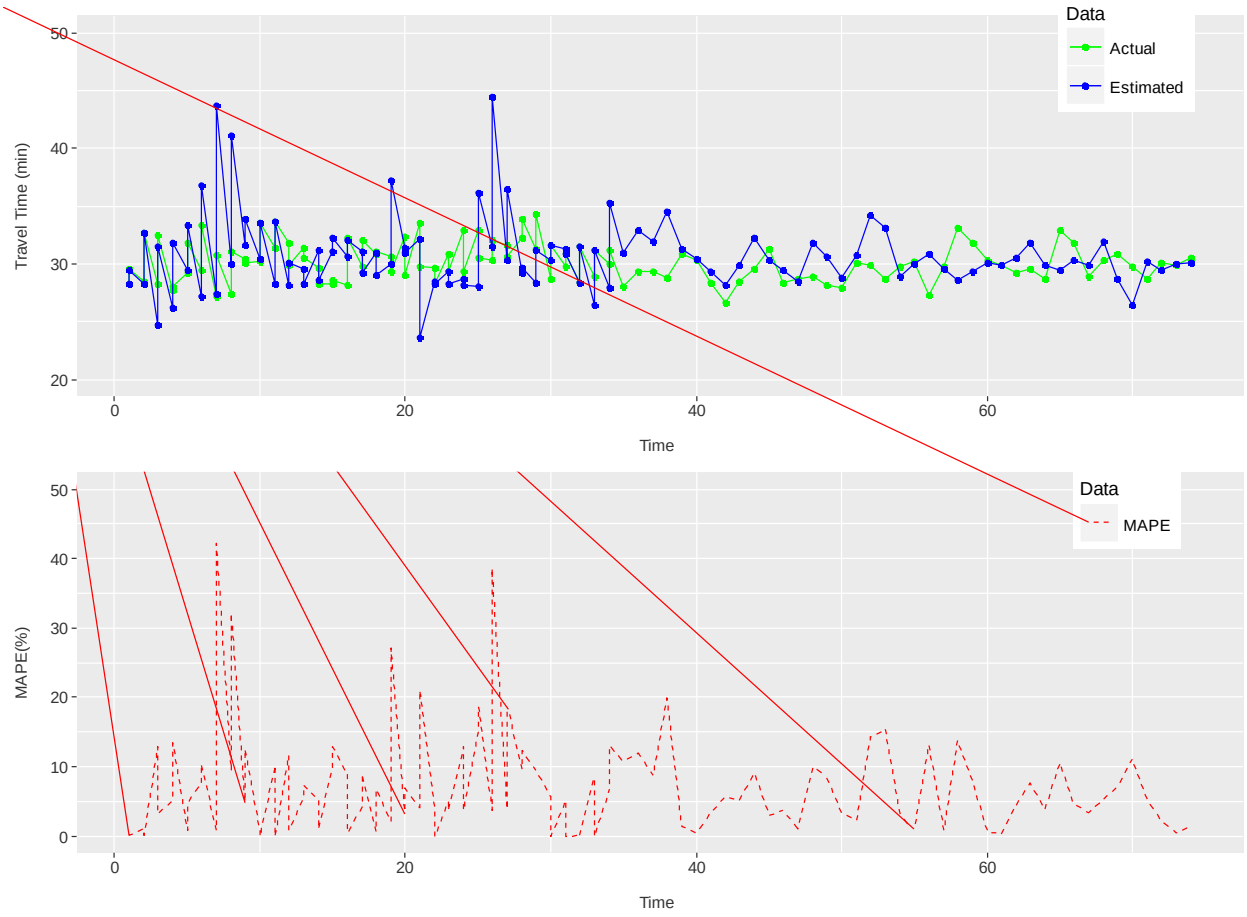


Figure 4.10 Actual and estimated travel time and MAPE with all tracked pairs

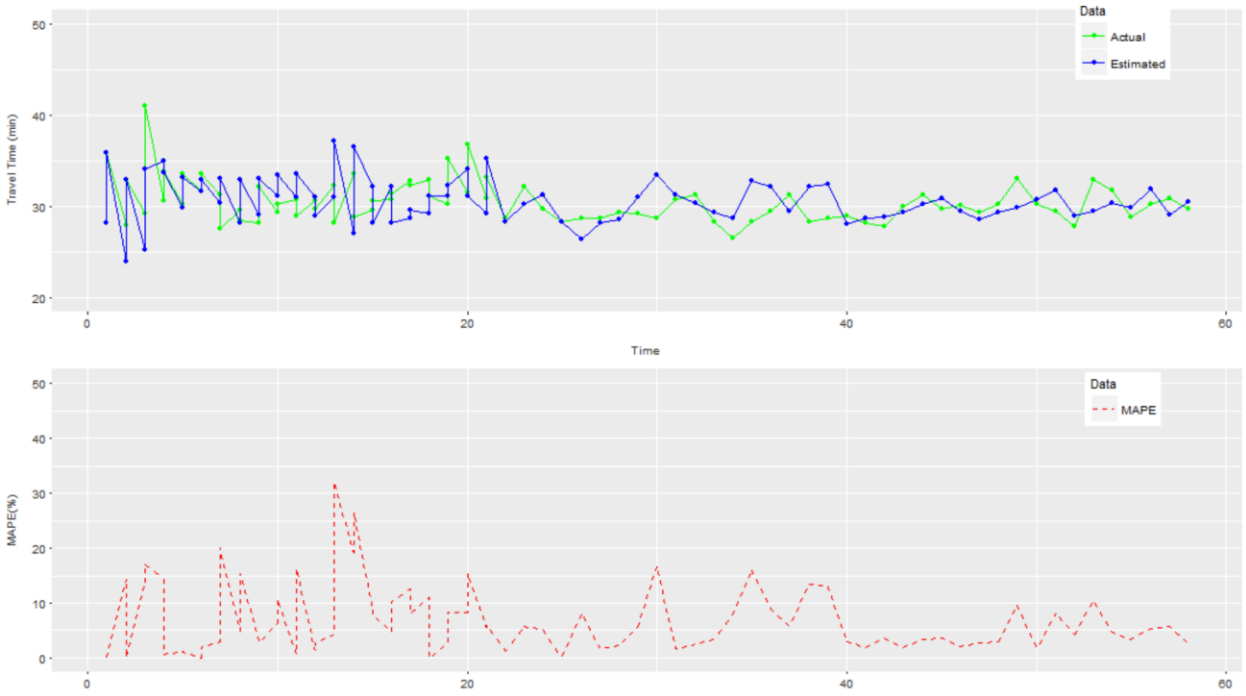


Figure 4.11 Actual and estimated travel time and MAPE with sampled tracked pairs

Table 4.6 Comparisons of MAPE

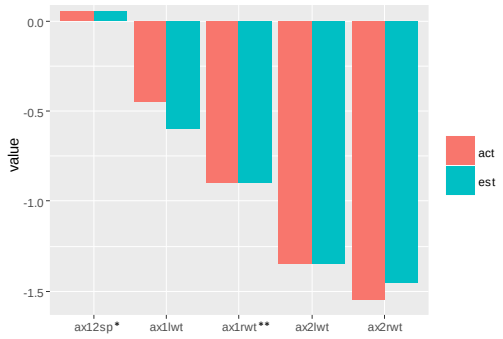
	All tracked pairs	Sampled tracked pairs
Min	0%	0%
1st Quantile	2.2%	2.5%
Median	6.2%	4.9%
Mean	9.4%	6.9%
3rd Quantile	13.3%	9.9%
Max	47.8%	32.9%

4.7.3 WIM calibration

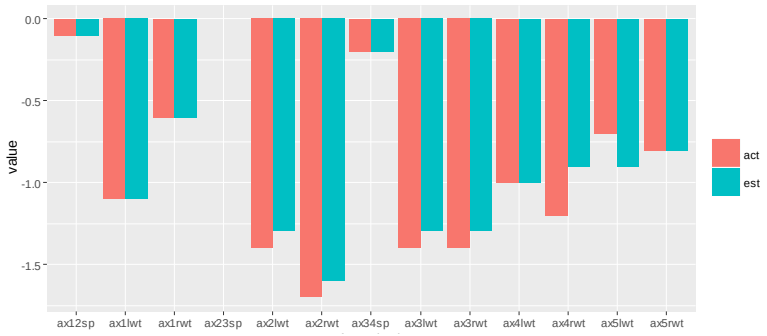
WIM calibration is another direct application from tracked vehicles. Differences of the WIM attributes between the WIM sites can be estimated from the matched vehicles. However, it is difficult to know which site has inaccurate measures since only relative differences can be obtained by the track vehicles. However, if the tracking is performed at multiple WIM sites, and the WIM data are compared at multiple sites, problematic sites could be identified. However, with a two WIM sites comparison, a degree of calibration issue can be only estimated by the relative differences in WIM estimates. The attribute differences in this chapter are estimated as follows.

$$\text{Value}_{\text{Attribute}} = Upstream_{Attribute} - Downstream_{Attribute}$$

Tracked vehicle between San Onofre and Leucadia WIM site are used for this application and the WIM data are separately examined by vehicle types. Figure 4.12 shows the actual and estimated differences in WIM measures. Overall, axle spacing and weight loads obtained at the downstream location were larger than those at the upstream location. Moreover, weight loads showed higher calibration errors than spacing measurements. However, it was shown that our matched vehicles would successfully correct the calibration errors since the pattern of off-calibrated values are pretty accurately tracked from the matched vehicles.



(a) Single units (Test dataset)



(b) Tractor-trailer units (Test dataset)

* axsp12 indicates axle spacing between 1st and 2nd axle

** ax1rwt indicates axle loading of 1st right axle

Figure 4.12 Differences in WIM measures

4.8 Conclusion

A Bayesian approach with selective weighted features was implemented to successfully distinguish matched truck pairs from mismatched ones in an open network for corridor level tracking. The results showed that a higher proportion of the truck population was successfully tracked with high matching accuracies at two WIM locations along a major freeway corridor spanning 26 miles and several interchanges. Matching accuracy for the trucks traversing both

WIM sites was 81 percent, where only 14 percent of trucks traversed both detection locations. The tracking algorithm, integrated with body classification modeling, was also implemented over a longer distance freeway section and showed temporally varying truck travel patterns by trailer types. This integrated tracking scheme shows significant potential in providing a new, comprehensive data activity source on truck movements and industry affiliations. Direct applications from the matched vehicles including travel time estimation and WIM calibration also provides valuable insights on the use-cases of tracking modeling.

5. Network-wide Tracking Model

The previous chapter investigated to estimate path flow by tracking individual vehicles along the same corridor. However, this tracking approach might be limited to capture dynamic truck activity in a complex road network since the tracking was focused on matching vehicles between two detector stations. Hence, in this chapter, the link-based tracking approach is extended to a network-wide tracking so that vehicles collected at multiple detector stations in different routes can be used in a tracking process.

5.1. Model overview

A network-wide tracking model adopts a general framework from the corridor level tracking. However, to handle significant amounts of matching candidate sets collected from multiple upstream locations, several steps were added to the original tracking model as shown in Figure 5.1. First, a vehicle classification step was considered. Since inductive loop signature has no capability in classifying vehicles, a truck detection algorithm was separately developed to categorize trucks into two groups (i.e., single-units and multi units) and to exclude passenger vehicles from the tracking process. In addition, a signature clustering step is introduced to more effectively utilize signature features to match vehicles as the ILD is the main data source in the network-wide tracking. To recognize varying distances and different traffic states between tracking sites, additional data sources were considered as supplementary information in the tracking model using Bayesian updating approach. Historical GPS, travel time, and truck body estimations are used as additional data sources. Consequently, tracking performances with four scenarios that consist of different combination of data sources are compared as follows.

Scenario 1: Only inductive signature data is available for vehicle tracking

Scenario 2: Signature and GPS data are available

Scenario 3: Signature and truck body classification estimates are available

Scenario 4: All of the above data sources are available

5.2 Truck detection algorithm

A binary truck detection algorithm was developed to distinguish passenger vehicles, single-unit trucks, and multi-unit trucks using vehicle duration from ILD data. The definition of multi-unit in this study refers to FHWA axle based class 8 to 13. Single-unit trucks represent FHWA class 4 through 7 (FHWA, 2013).

5.2.1 GM model for truck detection

Several previous studies have investigated the classification of vehicles based on ILD data from single loop sensors. Initial attempts provided percentage of long vehicles using aggregated flow and occupancy measures (Kwon et al, 2003; Wang and Nihan, 2003). Recently, Coifman et al (2009) developed an individual length-based classification scheme. Individual vehicle speed and length estimated by ILD data were utilized for vehicle classification. However, this approach requires several pre-determined traffic values to estimate individual vehicle length and speed such as defining speeds higher than 45 mph as free flow. These assumptions may result in inaccurate individual speed estimates under different traffic conditions, which eventually yield misclassification outcomes. Their results also showed that their model performance was not reliable during congested periods.

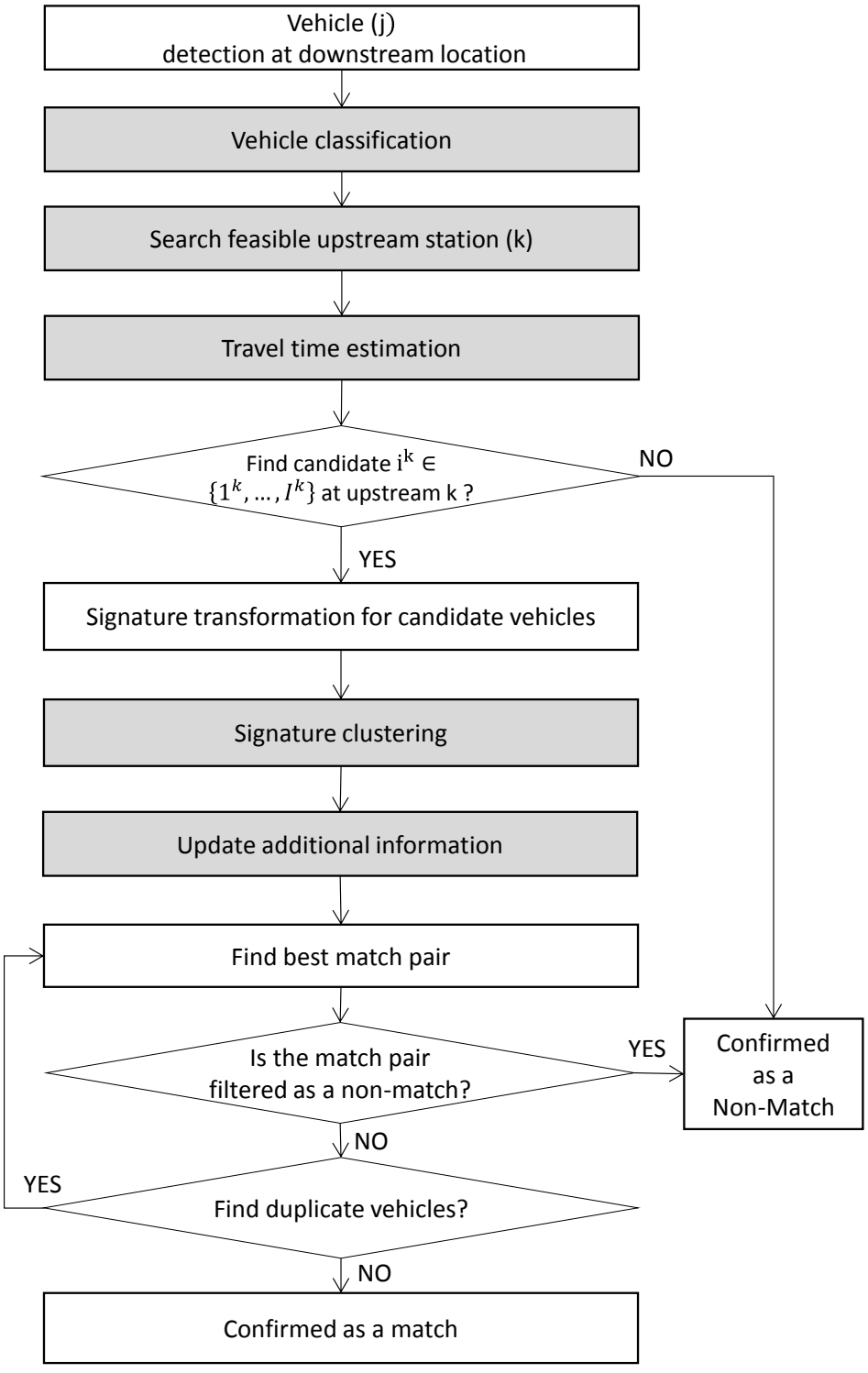


Figure 5.1 Flow chart of network-wide tracking

The proposed algorithm focuses on the difference in duration by vehicle type to identify trucks. Duration data of three different vehicle types which correspond to FHWA class 2-3, FHWA class 4-7, and FHWA class 8-13 were collected and analyzed as illustrated in Figure 5.2. Overall, duration ranges were distinct vehicle types; especially the duration of FHWA class 8-13 vehicles is noticeably longer than the others.

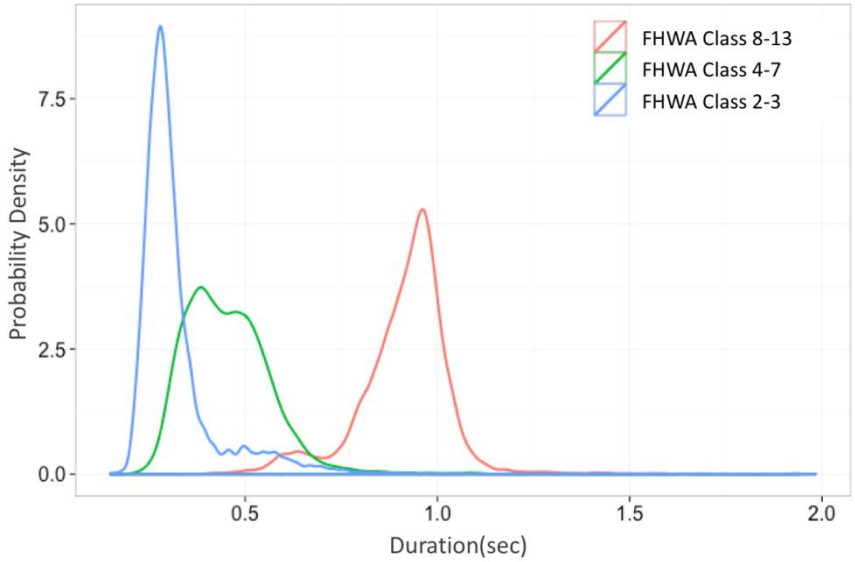


Figure 5.2 Duration distributions by vehicle type

However, the range of duration measures associated with vehicle types vary across traffic conditions. For example, the ranges are typically longer in congested traffic compared with uncongested traffic. Therefore, in order to use the duration range as an identifier of vehicle type, the range should be updated at short time intervals to effectively capture the changes in traffic state. This study developed an approach using the Gaussian Mixture (GM) model to determine and update the duration ranges by vehicle type over short time periods, deemed as 15 minutes to

obtain sufficient volumes to construct GM model. It should be noted that shorter or longer time period is easily adapted in the proposed GM model.

A GM model is a linear composition of Gaussian distributions, $\mathcal{N}(\mu_m, \Sigma_m)$ with a mixing proportion of p_m (Hastie et al., 2009). Since this study estimated three duration distributions – passenger vehicles, single-unit trucks and multi-unit trucks – tri-modal GM distributions were estimated by applying a mixing proportion of the three distributions.

$$f(x) = \sum_{m=1}^{M} p_m \cdot \mathcal{N}(x; \mu_m, \Sigma_m)$$

where m is number of mixture components, $\mathcal{N}(\mu_m, \Sigma_m)$ is a Gaussian distribution with mean μ and covariance matrix Σ, and p_m is the mixing proportion.

The proposed algorithm has several practical advantages. Implementation of the algorithm is not restricted by temporal or spatial conditions since the algorithm can reflect the change of traffic state in real-time. Additionally, any labor extensive data collection and process are not required to implement the algorithm at different ILD sites since the proposed algorithm does not need a training step or any assumptions for model development.

5.2.2 Validation of the truck detection algorithm

In our tracking dataset (see Table 5.2 and 5.3 for data description), the GM model shows 99 percent, 75 percent, and 95 percent of CCRs for passenger vehicle, single-unit truck, and multi-unit truck, respectively, as shown in Table 5.1. While passenger vehicles and multi-unit trucks showed high classification rates, approximately 19 percent of single-unit trucks were

classified as passenger vehicles. The body types of these misclassified trucks were mostly utility and service trucks which have relatively shorter durations.

Table 5.1 Truck detection algorithm results for tracking dataset

	Passenger Vehicle	Single-Unit	Multi-Unit	Total	CCR
Passenger Vehicle	206	1	2	209	99%
Single-Unit	96	382	32	510	75%
Multi-Unit	0	28	592	620	95%
Total	302	411	626	1339	88%

The proposed truck detecting GM algorithm was separately tested with 28,328 multi and single unit trucks collected at four ILD sites in California from the previous study (Hernandez et al., 2016). Since the passenger cars were not collected in this study, only single-unit and multi-units were used in GM model with a mixing proportion of two. Along with the individual vehicle duration and timestamp, side-fire images for each passing vehicle were stored together in a database to identify vehicle types. In the algorithm, every individual vehicle was classified into two types, multi-unit or single-unit trucks, based on its duration. In every 15 minutes, duration ranges for multi-unit trucks and single-unit trucks were updated in the GM model. Figure 5.3 shows results of two sample duration densities from an off-peak (a) and peak period (b), respectively. In these examples, the lower bound of duration for multi-unit trucks in an off-peak period was 0.42 seconds while a peak period showed a lower bound of 0.53 seconds, which was 0.11 seconds longer than that of the off-peak time period.

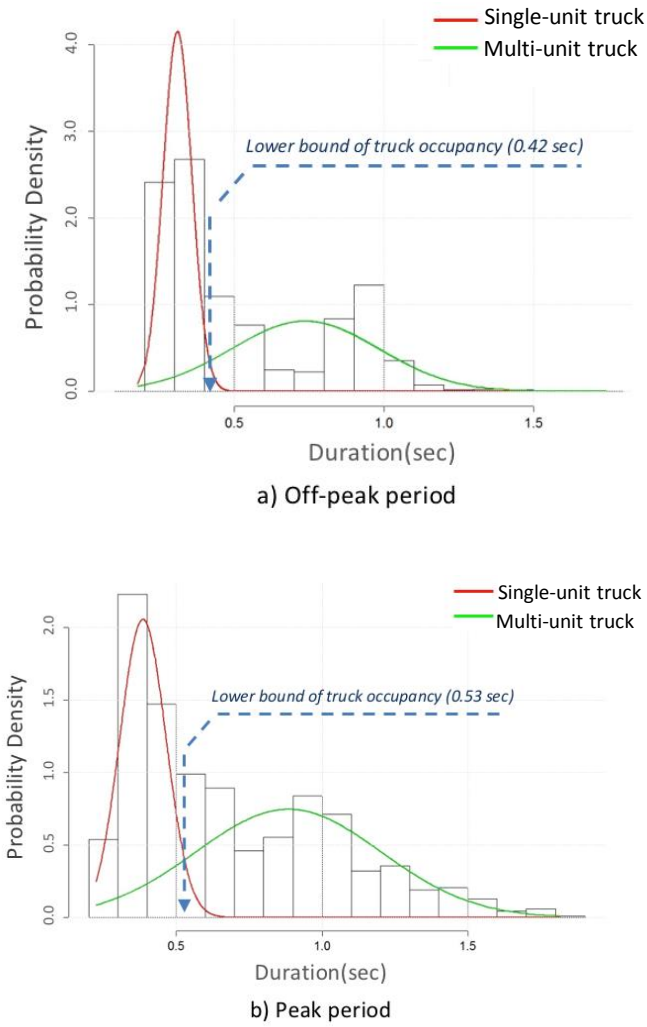


Figure 5.3 Duration distribution in off-peak (a) and peak (b) period

To validate the proposed GM model, the estimated vehicle types were compared to the actual vehicle types. A total of 323 15-minute periods were collected from the dataset. Since each time period provided classification results (i.e. probability of correct classification), summary statistics of mean, 25 percentiles, and 75 percentiles for correct classification cases are illustrated using a box-plot approach as shown in Figure 5.4. The average correct classification rates were 95% for single-unit trucks and 97% for multi-unit trucks. In other words, only 5% of

single-unit trucks were identified as multi-unit trucks, and 3% of multi-unit trucks were classified as single-unit trucks.

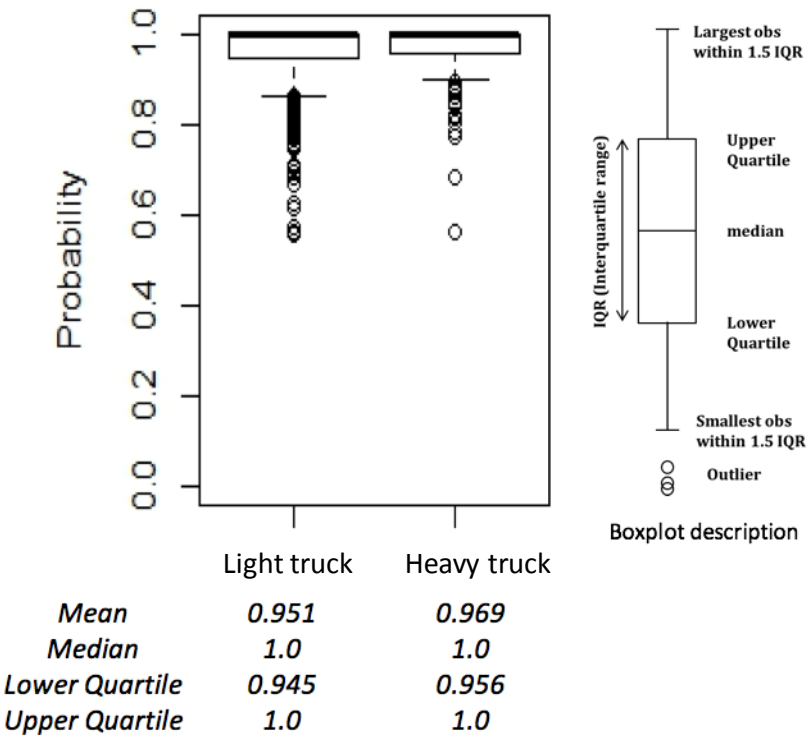


Figure 5.4 Truck detection algorithm results

The proposed algorithm was applied to a set of sample data obtained from 8 different ILD sites in California observed for 48 hours from March 29th (Tuesday) to March 30th (Wednesday), 2016. As shown in Figure 5.5, the proportion of multi-unit trucks varied temporally and spatially for the 48-hour period. Although repetitive time of day patterns were observed at some locations, the times with higher multi-unit truck traffic varied by site. For example, site 1 and 3 showed high proportions of heavy truck traffic during the night time while

high proportions of heavy truck traffic were observed during the daytime in site 2 and 4. These findings also confirmed that the proposed algorithm is capable of working as an independent truck detection platform.

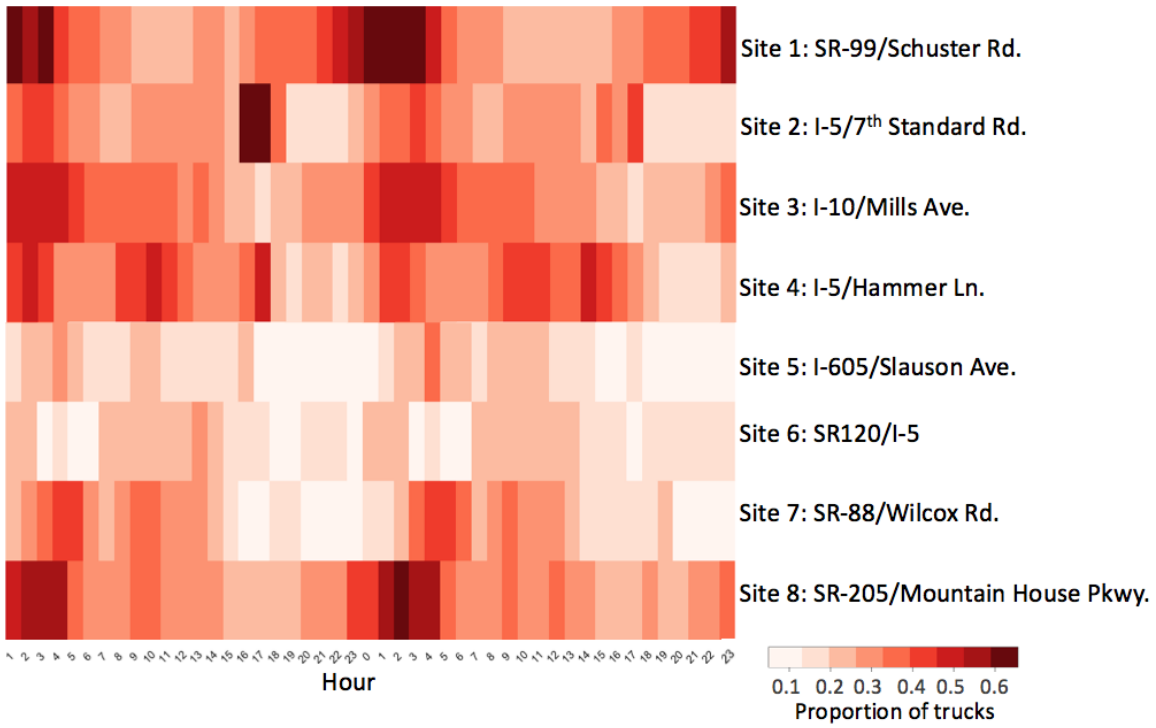
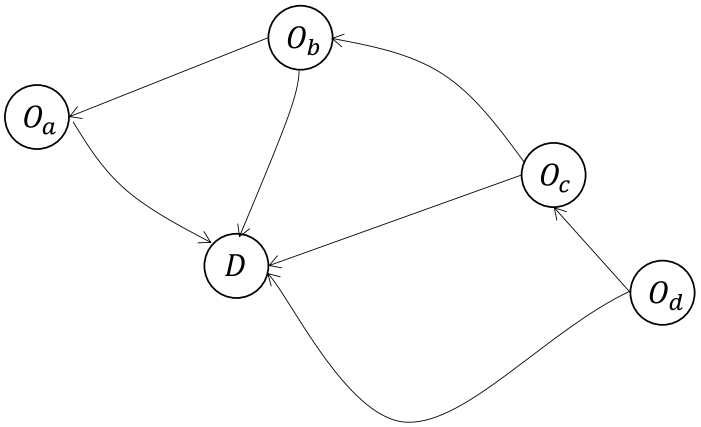


Figure 5.5 Proportion of heavy trucks in different segments of freeway in California

5.3 Search Space Identification

A network-wide tracking should match vehicles between one downstream location and multiple upstream locations. As a network becomes large, the number of possible upstream locations that reach to the downstream location can excessively increase. Therefore, it is important to identify feasible upstream locations in the network-wide tracking problem. An initial step is to find direct upstream sites that connect to the downstream site without passing

another site. Figure 5.6 illustrates a simple network with one downstream and four upstream locations.



OD	path
$[O_a, D]$	$O_a \rightarrow D$
$[O_b, D]$	$O_b \rightarrow D$ $O_b \rightarrow O_a \rightarrow D$
$[O_c, D]$	$O_c \rightarrow D$ $O_c \rightarrow O_b \rightarrow D$ $O_c \rightarrow O_b \rightarrow O_a \rightarrow D$
$[O_d, D]$	$O_d \rightarrow D$ $O_d \rightarrow O_c \rightarrow D$

Figure 5.6 Sample network for search space identification

In this case, four upstream locations $[O_a, O_b, O_c, O_d]$ can be directly reached to the downstream location [D]. To find feasible upstream and downstream location sets, only the shortest paths between the upstream and downstream locations are assumed to be selected by trucks. For example, if origin and destination is O_b and D, there are two possible paths: (i) $O_b \rightarrow D$ and (ii) $O_b \rightarrow O_a \rightarrow D$. Since the shortest travel time between these two locations can be obtained by the first path, only the first path is assumed to be used and considered in a tracking process. However, even though a direct path exists between two locations, for example O_d and D, if the travel time of the indirect path that passes through another location, such as $O_d \rightarrow O_c \rightarrow D$, is shorter than that of the direct path, $O_d \rightarrow D$, the indirect path would be chosen as the feasible tracking path. Therefore, location O_d is not chosen as the feasible upstream set of D. Consequently, the feasible sites of upstream set of this network are $[O_a, O_b, O_c]$.

5.4 Signature Clustering Approach

One challenge in vehicle tracking using ILD is to search a target vehicle among candidates that have the same truck or trailer body configurations because trucks with the same body configuration generate similar waveform signatures. The corridor tracking results (Chapter 4.5) confirmed that the most of the incorrectly matched target and candidate vehicle pairs had the same body trailer or truck types. In specific, 73 percent of the incorrectly matched pairs have the same truck types. This problem would make the network-wide tracking even more challenging because signatures are the main sources to distinguish vehicles.

Figure 5.7 depicts randomly chosen fifty signatures from the same trailer categories of livestock and tank. Even though the overall patterns of the same trailer type are very similar, we could visually found that there are discernible features that are more capable of identifying salient differences among vehicles. For example, all the signatures of the livestock trailers have high magnitudes from 0.5 to 1 second; however, the front part (up to 0.4 second) of the signatures showed more variations by trucks. Similarly, the middle parts of the tank signatures (between 0.3 to 0.8 second) are highly varied by vehicles. Therefore, these parts of signatures are seen to better distinguish vehicles. However, it would be difficult to find these features if all signatures with various waveform patterns are compared altogether since the distinct parts vary by overall signature pattern. However, if signatures that have similar patterns are grouped together, the distinguishable signatures features could be more salient.

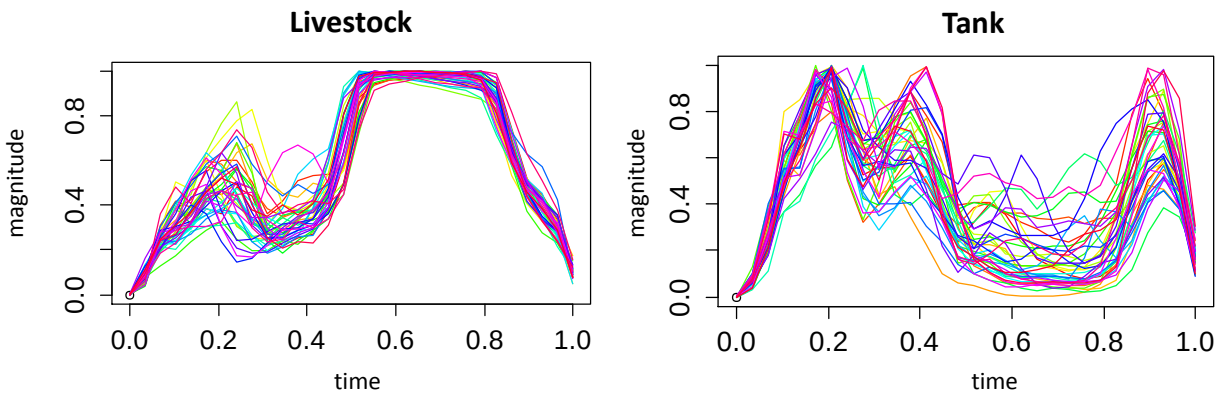


Figure 5.7 Comparison of signatures with different types of trailers

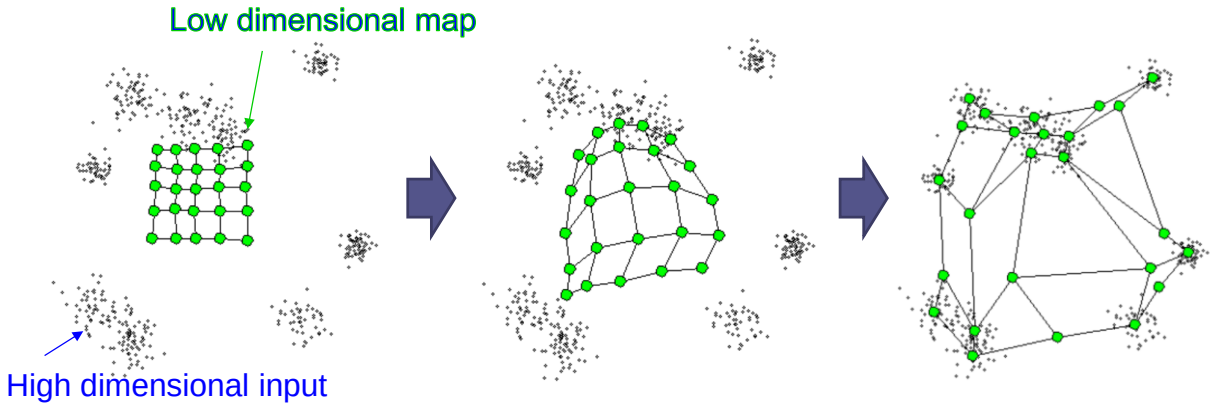
This study develops a signature clustering model to group trucks with similar signature shapes. Then, in each cluster, signature features that have larger variations are selected and weighted as more powerful features to distinguish vehicles. Parametric density functions are also separately estimated by signature clusters.

5.4.1 Self Organizing Map (SOM) for vehicle clustering

This study applies a SOM model for signature clustering. The SOM is an unsupervised clustering method using neural network algorithm. Since this method is unsupervised approach where data are not labeled for classification, vehicle signatures are clustered by their overall patterns, not classified by their body type labels. This feature allows this method to categorize high dimensional data solely depending on their inherent features.

It should be noted that signatures in the same trailer type may vary as observed in Figure 5.7. If there are common parts that have huge variations in several body configurations, the supervised classification algorithm may not work properly until the variations are removed and

the unique feature sets clearly indicate their corresponding vehicle type labels. However, it is impossible to entirely remove those overlaps; therefore, the SOM clustering would be more adequate approach in signature grouping since the waveform shapes determines clusters, not their body types.



Source : Eren Golge (https://www.quora.com/What-are-self-organizing-maps-How-do-they-work)

Figure 5.8 SOM model

SOM consists of two phase: algorithm training and mapping. The training step builds a map using given input data set. Then, a new input vector can be clustered using the map, which called mapping. The map consists of several nodes as shown in Figure 5.8. Every node is connected to the input, and no nodes are connected to each other. Each node has topological position and contains a vector of weights with the same number of input vectors. Initially, each node is assigned random weights between [0, 1]. A random input vector is chosen for training, and every node is examined to find the most similar node to the input vector, which refers to the best matching unit (BMU). The selection is implemented using Euclidean distance as a measure of similarity between the input vector and nodes. In other words, the distance between each

node's weight vector and input vector is calculated, and the node that has the closest weight vector to the input vector is tagged as the BMU. After identifying the BMU, a radius of neighborhood around the BMU is calculated to estimate the impact area (nodes) of the BMU. The radius decreases with an exponential decay function on each iteration.

The weights (w) of BMU and neighboring nodes are iteratively updated so that their weights can be more similar to the weight of the input vector.

$$\mathrm{w}(\mathrm{t}+1) = w(t) + \Theta(t)L(t)(V(t) - w(t))$$

$$\mathrm{L}(\mathrm{t}) = L_o \exp(-\frac{t}{\lambda})$$

$$\Theta(\mathrm{t}) = \exp(-\frac{dist^2}{2\delta^2(t)})$$

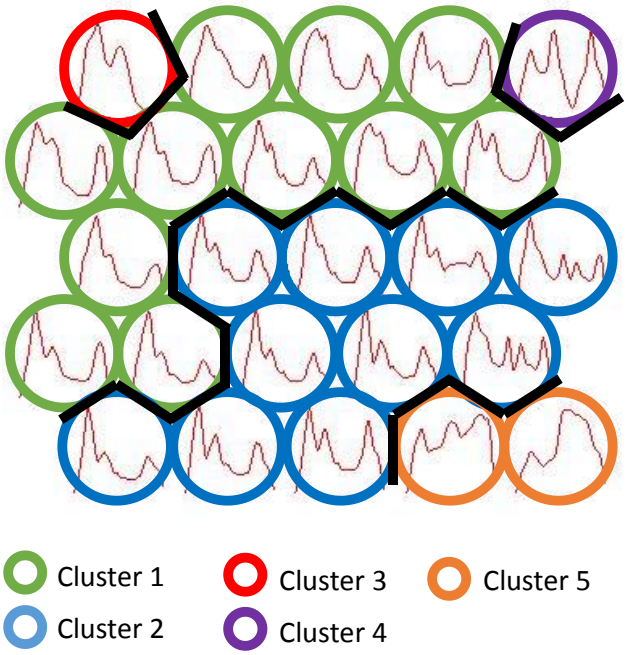
Where t is time step, $L(\cdot)$ is learning rate that decreases with time, L_o shows the previous learning rate of L, $\Theta(\cdot)$ is influence rate (neighborhood function) which shows amount of influence a node's distance from the BMU thus δ is an width of the lattice at the time t, V indicates input vectors, and w denotes the corresponding weight vector for each node, λ *and* $dist$ are constant.

While learning rate is monotonically decreasing over time, influence rate depends on the lattice distance. However, in general, initial stages where the neighborhood is broad have higher influence rates to affect more neighbors. At the last stage, however, weights are converged and highly affects to closer neighbors.

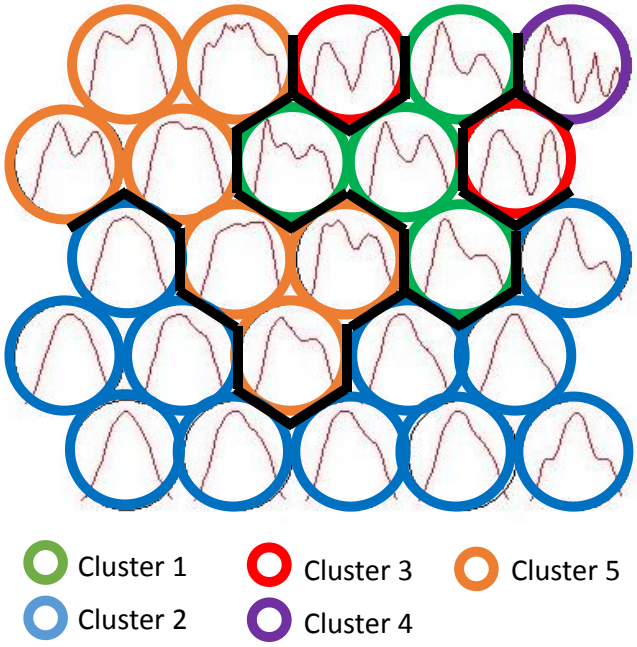
In SOM, nodes are placed onto a grid and each node has neighbors. In every step, nodes change their position on the grid. Therefore, the BMU adapts itself to the input vector, which further causes a change for its neighbors. Therefore, the neighbor vector is positioned closer to the BMU and increase difference with surrounding vectors.

5.4.2 Clustering results

This study trained a SOM with a large dataset that contains varieties of signature patterns from over 40 truck types. The data set includes 28,328 single and multi-unit trucks collected at four ILD sites in California, which is also used for validation of the truck detection algorithm in Chapter 5.2. This set was split into training and testing and the training data is used for cluster development. A total of 25 nodes are used in the map. After the training step, K-means clustering was used to find the optimal number of clusters. Figure 5.9 (a) and Figure 5.9 (b) show the optimal five clusters for multi-unit and single-unit, respectively, with a representative signature pattern in each node. Color in the figure represents clusters, for example, 11 green nodes in Figure 5.9 represent the ‘cluster 1’. Similarly, one node in red and purple color represents ‘cluster 3’ and ‘cluster 4’, respectively. It is visually confirmed that the nodes with the same cluster show similar signature shapes. For example, two nodes in ‘cluster 5’ both show signatures that have higher magnitude on the trailer part.



(a) Tractor-trailer unit Clusters



(b) Single unit Clusters

Figure 5.9 Clusters and signature patterns

The nodes in the same cluster contain similar signature patterns. Therefore, if each signature's body configuration is known, we could investigate associations between signature patterns and body configurations. In this study, trailer types of multi-unit were examined by clusters as shown in Figure 5.10. Most of vehicles in cluster 1 are enclosed van type while those in cluster 2 include the most variety of body configurations among five clusters. Cluster 3 however only includes one type of body type, logging trailers; and cluster 4 represents the units pulling multiple trailers. Cluster 5 represents lower deck trailers such as low boy platform, drop frame, and livestock.

Since each cluster has different waveform patterns, clusters would have different significant parts that have more abilities to differentiate vehicles. Therefore, parametric density functions along with feature selection and weighting are investigated by cluster. Figure 5.11 shows examples of feature distributions by cluster. Two features are selected as examples. Feature A (e.g., 13^{th} signature feature) shows that the cluster 1 has larger variances for both match and mismatch while the cluster 2 shows smaller variances for both match and mismatch. However, the cluster 3 has small variance in match but larger variance in mismatch. These differences indicate that features have varying ability in distinguishing match and mismatch, and further, the ability differs by clusters. Similarly, the feature B (e.g., 41^{st} signature feature) shows different variances of match and mismatch density distributions by clusters. As discussed in Chapter 3.5, feature weights are estimated and categorized into four labels, corresponding to critical, signature, insignificant, and inverse features based on their statistical differences and variances in distribution of match and mismatch pairs.

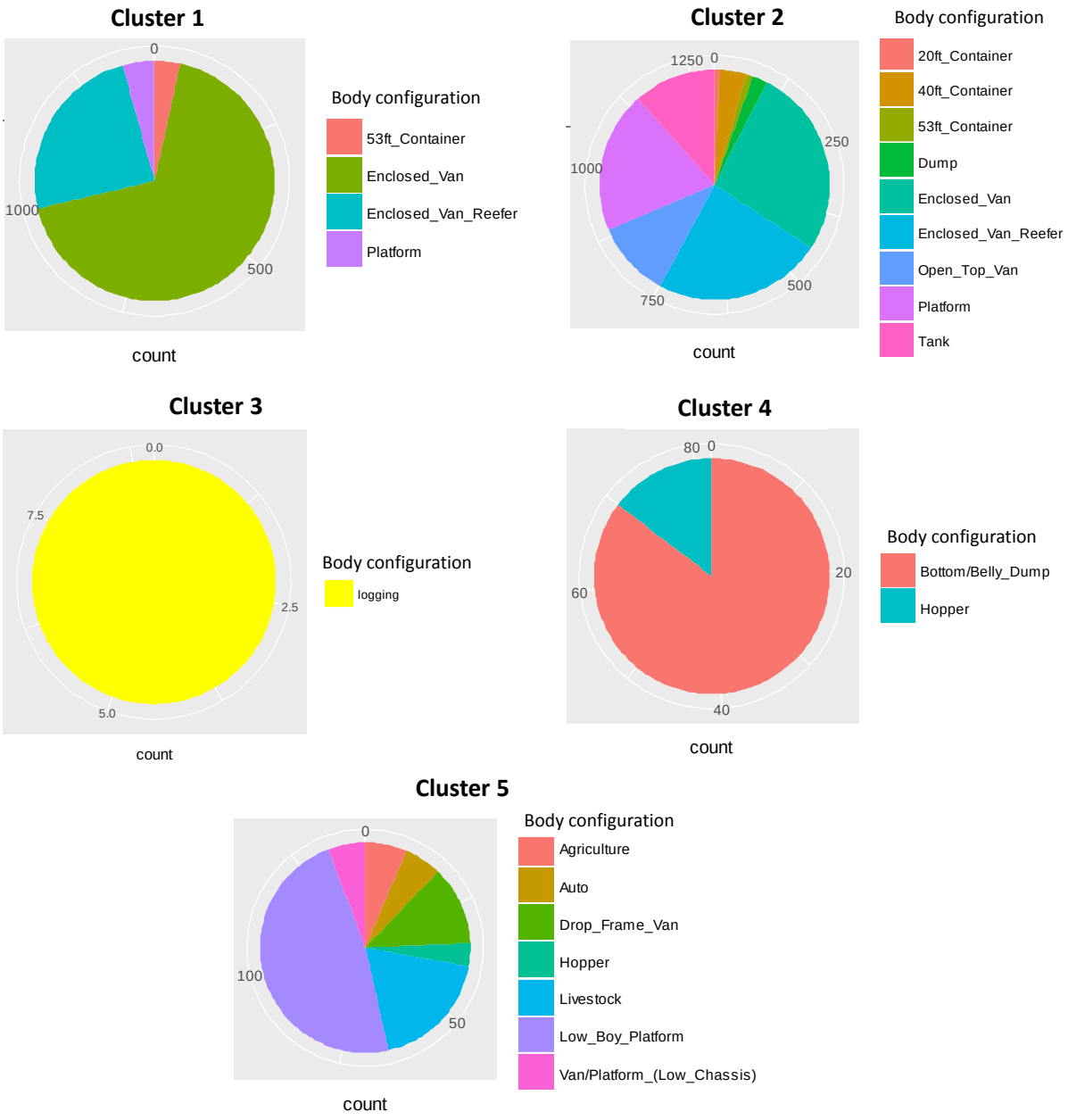
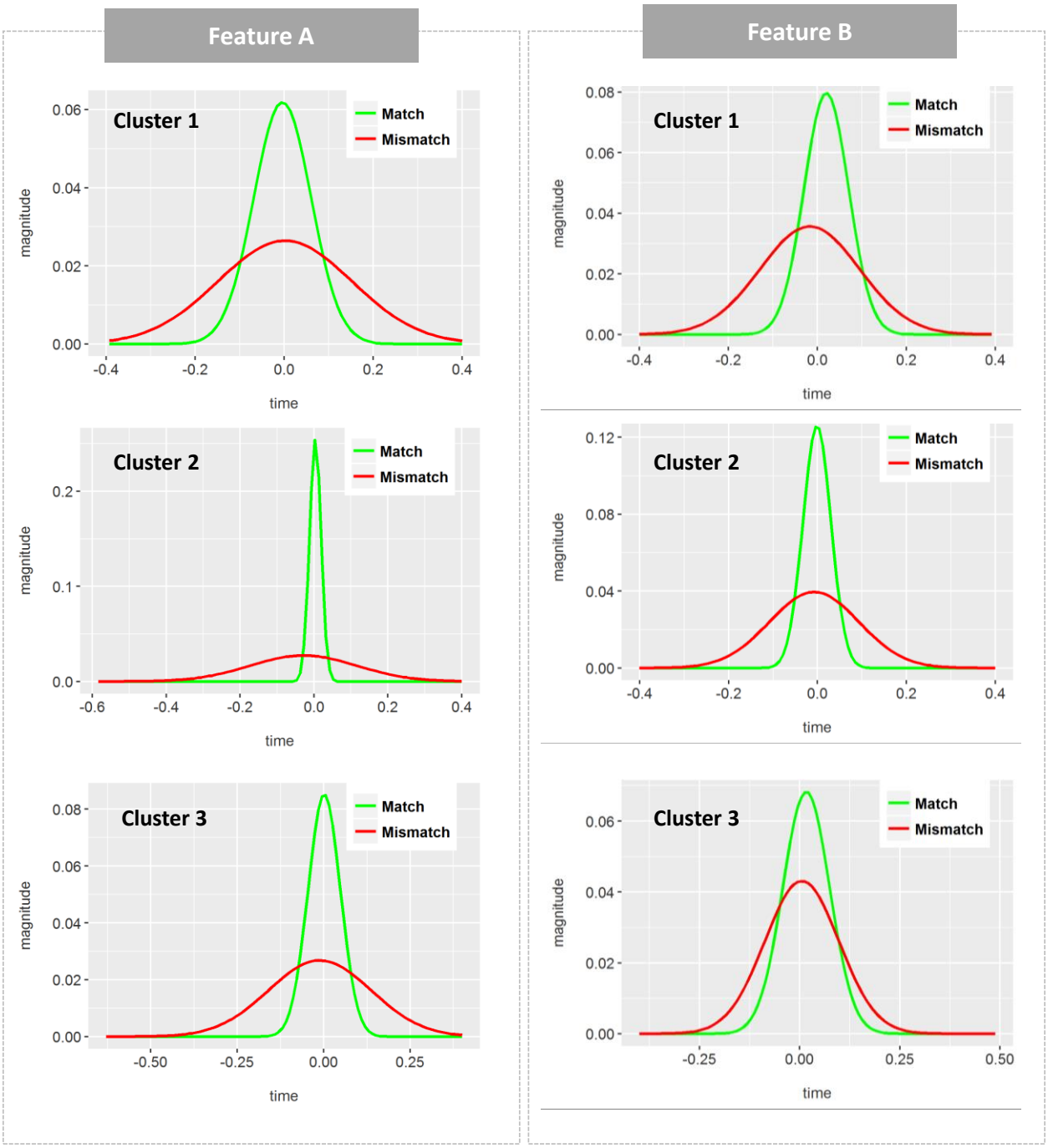


Figure 5.10 Body configurations by SOM cluster



*Feature A represents the 13th signature feature and Feature B represents the 41st signature feature.

Figure 5.11 Comparisons of parametric density functions by cluster

5.5 Data

To implement the network-wide tracking, we selected two upstream locations and one downstream location. The distance from each upstream to downstream location is 5.2 and 5.5 miles, spanning five major freeway intersections and entrance and exit ramps as shown in Figure 5.12 and Table 5.2. On July 7th, 2016, signature data from the two outermost northbound lanes were collected from the upstream #1 (U1) and the downstream sites (D), and two westbound lanes from the upstream #2 site (U2) were collected. A total of 424 vehicles were collected at the downstream locations where 58 percent of trucks are multi-unit trucks (Table 5.3). At U1, 222 trucks were collected and 74 percent of them were multi-unit units. At U2, 62 trucks were collected where 60 percent of the total population is multi-unit. There are 284 common trucks in this network, which is 67 percent of the total vehicle captured at downstream location.

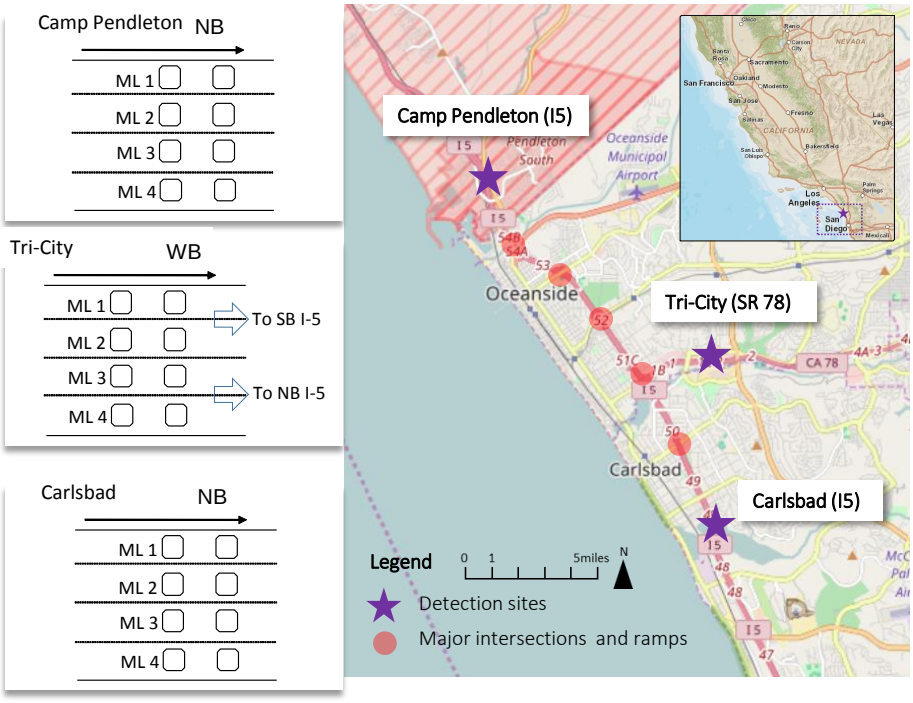


Figure 5.12 Data collection site map for network-wide tracking

Table 5.2 Data collection site description for network-wide tracking

Site Location	Distance	Site Description	Collection Dates
Camp Pendleton (I5) to Carlsbad (I5)	5.2 miles	D-U1 (Upstream Site #1)	July 7th 2016 11:20AM – 12:40PM
Camp Pendleton (I5) to Tri-City (SR 78)	5.5 miles	D-U2 (Upstream Site #2)	

Table 5.3 Data collected for network-wide tracking

Dataset	# of Trucks Collected at Downstream	% of Multi-unit at Downstream	# of Trucks Collected at Upstream	% of Multi-unit at Upstream	# Common Vehicles	% of Common Vehicle from the Total Vehicle Detected at Upstream
D-U1	424	58%	421	54%	222	67%
D-U2			118	49%	62	
Total			539	53%	284	

Inductive signature data and side-fire images for trucks were collected at three loop sites for validating network-wide tracking model (Figure 5.13). Similar to the corridor tracking process, inductive signatures and still images were stored in a database and manually linked through a groundtruth data processing.

Downstream Upstream 1 Upstream 2

Figure 5.13 Site Images for network-wide tracking

5.6 Additional Sources

5.6.1 GPS

GPS data were used to estimate spatial relationships between downstream and each upstream location based on route flows extracted by GPS trajectories. The underlying concept is that if vehicle flows from a particular upstream (U1) and downstream set is higher than the other upstream (U2) and downstream set, it can be assumed that vehicles from the upstream location (U1) have higher chance to be matched to the target vehicles on July 7th, 2016 downstream. This study utilized the GPS data collected from the American Transportation Research Institute (ATRI) for four weeks in each quarter in 2010. GPS vehicle trajectories were extracted, and path flows were estimated by each downstream and upstream pair. First, screen-lines for each detector stations (Downstream and two upstream sites) were created in Q-GIS as shown in Figure 5.14. Second, the GPS trajectories that pass each downstream-upstream set were queried at PostgreSQL database. Among a total of 2,307 trajectories that pass at the downstream location, 1,769 and 538 route flows were passed at U1 and U2, respectively. Therefore, prior matching probability for D-U1 and D-U2 pairs are 77% and 23%, respectively.

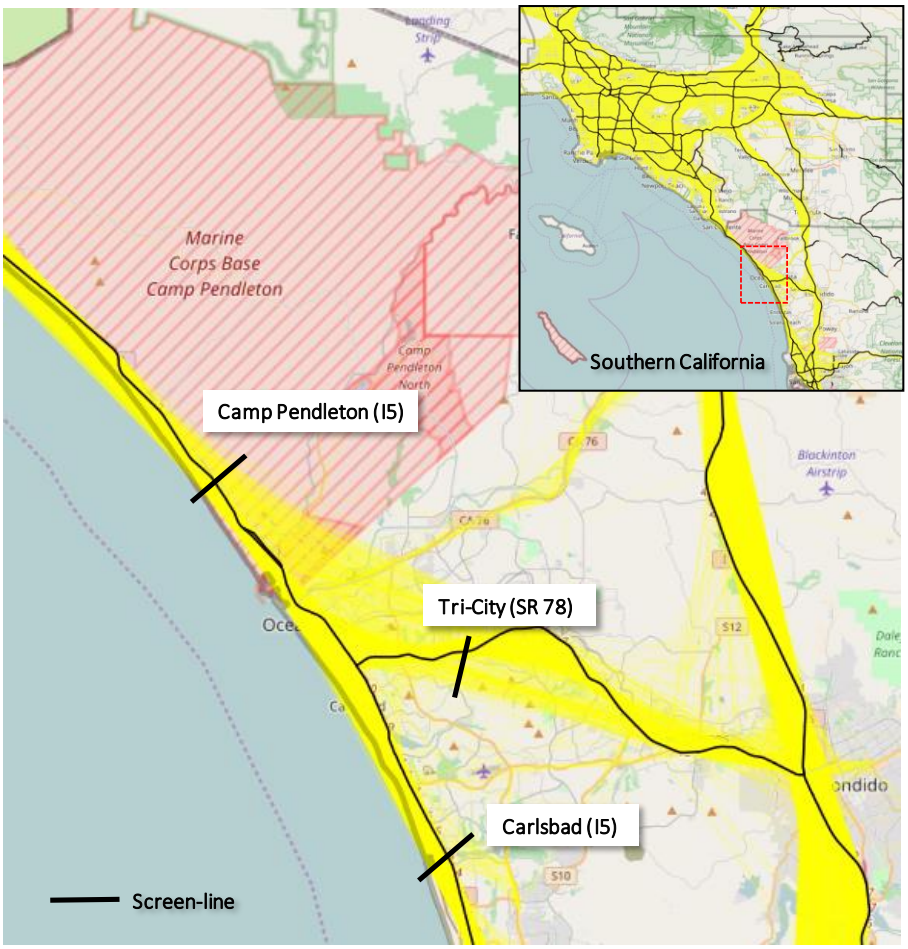


Figure 5.14 GPS trajectories among network-wide tracking sites

5.6.2 Truck body classification estimates

A detailed truck body classification model developed by Hernandez et al (2016) used inductive signatures to distinguish over 40 single-unit and multi-unit truck configurations from ILD sites as shown in Table 5.4.

Table 5.4 Truck classification scheme based on Inductive Signature Data

Units Type	Body Class	Units Type	Body Class
Single-Units without Trailer	Conventional Van/Platform	Multi-Units Tractor Trailers	Enc. Van
	Cab Over Van/Platform		53ft Container
	30ft Bus		40ft Container
	20ft Bus		40ft Container Reefer
	Multi Stop Van/RV		20ft Container
	Utility/Service		Platform
	Concrete		Tank
	Dumpster Transport		Open Top Van
	Garbage		Auto
	Bobtail		Low Boy Platform
	Dump Triple Rear		Drop Frame Van
	Street Sweeper		Dump
	Dump/Tank		Logging
Single-Units with Trailer	SU small trailer		Livestock
	Dump-Dump		Agriculture
	RV with Towed Vehicle		Beverage
	Concrete w/Lift Axle		Enclosed Van Reefer
	Tank-Tank		Platform/Tank
	Platform-Platform		Dump
	Tow Truck with vehicle	Multiple Semi Tractor Trailers	Pneumatic Tank
	Dump with Lift Axle		Hopper

The body classification model estimates individual body types for vehicles detected at every data collection sites. For the tracking process, two body type estimations were used, one is an individual estimation and another is an hourly estimation. An individual estimation compares body types estimated at downstream and upstream locations. If the same body types for target and candidate are estimated, the matching probability would be higher than the different body types are estimated. The matching probability is obtained from known correct classification rate (CCR) of body types from Hernandez et al (2015). An example of CCR for multi-unit is presented in Table 5.5. For example, if the downstream body type estimation is ‘enclosed van’ and the upstream body type is estimated as ‘enclosed van’, the matching probability is 83% based on the CCR of the ‘enclosed van’. Since the body configuration model includes the

classification error, even if the body types between target and candidate trucks are the same, the matching probability may not be 100 percent. If the body estimates are different at downstream and upstream locations, we take conservative approach to obtain a CCR. For example, if one location estimates 40ft container and another location estimates platform for one pair, we assumed that either body type could be a true value. Thus, we used maximum CCR between the case where 40ft container is misclassified as a platform and the case where platform is misclassified as a 40ft container. If the first case has 10% and the latter case has 20% of CCR, this study takes the bigger CCR as their matching probability.

Table 5.5 Body CCR Table for tractor-trailer units

Downstream Body type	Upstream Body type	CCR
40ft Container	40ft Container	64%
40ft Container	Platform	7%
40ft Container	Enclosed Van	7%
53ft Container	53ft Container	42%
53ft Container	Enclosed Van	53%
Enclosed Van	Enclosed Van	83%
Enclosed Van	Platform	3%
Enclosed Van	Enclosed Van Reefer	11%
Dump	Dump	57%
Dump	Tank	19%

While an individual estimation determines the matching probability between a target and a candidate vehicle, hourly estimation considers a potential upstream location based on the estimated body types. First, body type estimations from individual vehicles at each upstream are aggregated in an hour. Second, a target vehicle's body type is identified from the body classification model. Third, hourly aggregated volumes for the target's body type are compared to multiple upstream locations, and the probability of each upstream location is calculated based on volume proportions of the corresponding body type. An example of hourly estimation

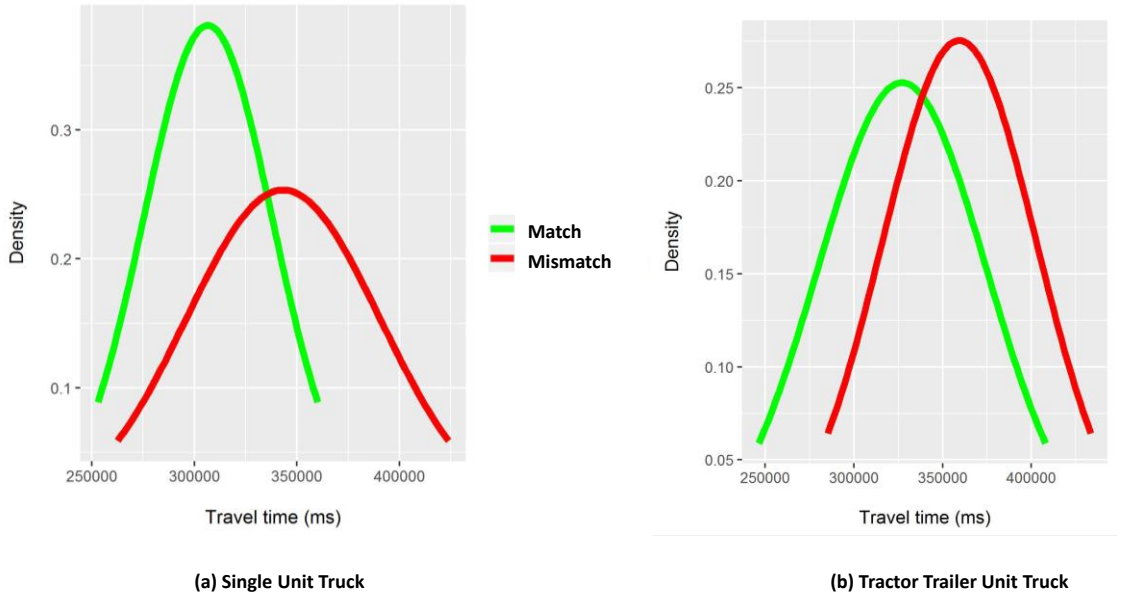
matching probability used in this study is presented in Table 5.6. For example, if the body type of target vehicle is estimated as '20ft container', the probabilities of two upstream locations are 21 percent (U1) and 79 percent (U2), respectively.

Table 5.6 Probabilities of upstream locations by truck body configuration

Body configuration	Probability for D-U1	Probability for D-U2
20ft Container	0.21	0.79
40ft Container	0.17	0.83
53ft Container	0.77	0.23
Platform	0.46	0.54
Drop Frame Van	0.35	0.65
Enclosed Van Reefer	0.42	0.58
Enclosed Van	0.63	0.37
Dump	0.62	0.38
Low Boy Platform	0.58	0.42
Open Top Van	0.58	0.42
Tank	0.72	0.28

5.6.3 Travel time

Travel time is an important indicator in search space reduction. However, since the network-wide tracking considers multiple upstream locations, travel time is utilized to find matched vehicles beyond the search space reduction. Similar to the signature feature vectors, travel time is also considered as a feature vector in the Bayesian model. Therefore, travel time distribution was estimated based on the potential match and mismatch pairs for each upstream. Figure 5.15 shows travel time distributions from U1 by truck types.



(a) Single Unit Truck

(b) Tractor Trailer Unit Truck

Figure 5.15 Travel time distributions for matched and mismatched pairs

5.7 Results

Prior to implementing the Bayesian model, a fixed parameter approach was applied. The fixed threshold was applied to the total distances of signature features. In other words, if the sum of signature features from a potential matching pair is less than the pre-determined threshold, the pair is declared as a match. The set of performance metrics developed for the corridor level tracking was also used for the network-wide tracking. As shown in Figure 5.16, the most balanced solution between OSMR 1 and OSMR 2 was found when three was used as the fixed threshold where 81 percent of OSMR 1 and 69 percent of OSMR 2 were obtained.

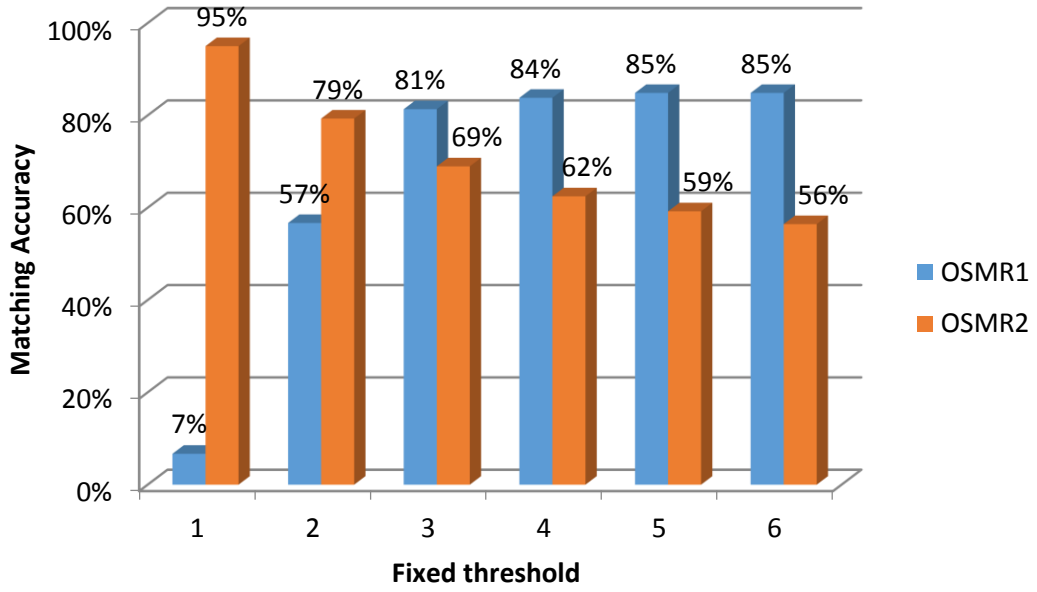


Figure 5.16 Fixed threshold approach results

Table 5.7 shows the results from the Bayesian model. The best fusion weights were found after applying varying fusion weights to the data sources. For comparison purposes, the matching rates from the scenarios that contain limited data sources are presented. The best fusion weights were found when more weight was applied to signature features than GPS data. Interestingly, body estimations are not seen as a helpful source based on the best fusion weight. This could be because the signatures are initially clustered based on their waveform shapes, thus body configuration information may not be additionally needed.

When signature and GPS were used for vehicle tracking with the best fusion weights, the matching accuracies for multi units was over 90 percent for all of CSMR, OSMR 1 and OSMR 2. When one or more data sources are missing, matching accuracies were dropped by 1 percent to 20 percent depending on the applied data. It should be noted that even though the distances

between tracking points are about 5 miles, only 68 percent of vehicles pass both downstream and upstream stations. However, the tracking algorithm was able to maintain total population with high matching accuracy based on balanced high OSMR1 and OSMR2 results. For multi-units, only two percent drop was observed after the filtering step was implemented, which indicates that the filtering step well performed to filter missing pairs for the multi-unit units.

However, overall performance for single-units was lower than multi-unit units. At the best fusion weights, 67 percent and 54 percent of OSMR 1 and OSMR 2 were obtained. Since less variety of truck types and axle configuration were observed for the single-units, signatures alone cannot capture the salient differences in vehicle features.

Table 5.7 Bayesian model results for network-wide tracking

Truck set	Data description	Fusion weight for available information					Number of trucks				Matching Rate				
		SIG	Travel time	Body type Ind. est.	Body type Hourly est.	GPS	Total	Common	Total Match	Correct Match	CSMR	OSMR1	OSMR2	FMR1	FMR2
All Trucks	Best Combination: SIG + GPS	3	0	0	0	1	424	284	303	238	**88%**	**84%**	**79%**	**23%**	**15%**
	All (equal weight)	1	1	1	1	1			265	205	77%	72%	77%	21%	14%
	SIG only	1	0	0	0	0			292	233	85%	82%	80%	21%	14%
	SIG + all Body est.	1	0	1	1	0			255	202	84%	71%	79%	19%	13%
	SIG + Body Ind est.	1	0	1	0	0			226	183	85%	64%	81%	15%	10%
	SIG + Body Hourly est.	1	0	0	1	0			281	218	85%	77%	78%	22%	15%
	SIG+Travel time	1	1	0	0	0			294	228	85%	80%	78%	23%	16%
	SIG + GPS (equal weight)	1	0	0	0	1			289	228	83%	80%	79%	21%	14%
Multi Unit Trucks	Best Combination: SIG + GPS	3	2	0	0	2	245	201	201	182	**93%**	**91%**	**91%**	**9%**	**8%**
	All (equal weight)	1	1	1	1	1			178	160	84%	80%	90%	9%	7%
	SIG only	1	0	0	0	0			194	178	91%	89%	92%	8%	7%
	SIG + all Body est.	1	0	1	1	0			171	156	91%	78%	91%	7%	6%
	SIG + Body Ind est.	1	0	1	0	0			155	142	92%	71%	92%	6%	5%
	SIG + Body Hourly est.	1	0	0	1	0			186	168	90%	84%	90%	9%	7%
	SIG+Travel time	1	1	0	0	0			196	175	90%	87%	89%	10%	9%
	SIG + GPS (equal weight)	1	0	0	0	1			197	180	89%	90%	91%	8%	7%
Single Unit Trucks	Best Combination: SIG + GPS	3	2	0	0	2	179	83	104	56	**76%**	**67%**	**54%**	**58%**	**27%**
	All (equal weight)	1	1	1	1	1			87	45	63%	54%	52%	51%	23%
	SIG only	1	0	0	0	0			98	55	70%	66%	56%	52%	24%
	SIG + all Body est.	1	0	1	1	0			84	46	66%	55%	55%	46%	21%
	SIG + Body Ind est.	1	0	1	0	0			71	41	69%	49%	58%	36%	17%
	SIG + Body Hourly est.	1	0	0	1	0			95	50	71%	60%	53%	54%	25%
	SIG+Travel time	1	1	0	0	0			98	53	71%	64%	54%	54%	25%
	SIG + GPS (equal weight)	1	0	0	0	1			92	48	70%	58%	52%	53%	25%

5.8 Discussion

This study used ATRI GPS trajectories as a supplementary dataset to provide additional information on truck movements. However, several limitations are found in GPS path flow. For one, GPS data collected from a sample of trucks that equipped with GPS devices represent the path flow of truck population. In addition, truck samples collected during shorter temporal span could represent the overall travel pattern. To address this problem, this study implemented a sensitivity analysis on GPS data and alternative data sources are investigated.

5.8.1 Sensitivity analysis on GPS data

GPS dataset used in this study consists of nine days from each of four months –February, May, August, and November – where each month represents different season. The data include five business (weekdays) and two sets of weekends. Among the path flows that passed at downstream site (D), flow proportion from upstream #1 site (U1) and upstream #2 site (U2) were presented by months, days and hours in Figure 5.17. For example, 72 percent and 78 percent of total flows were observed from U1-D set on May and August, respectively. It was shown that the path flows vary by season, day of week (DOW), and time of day (TOD), ranging from 64 percent to 81 percent on the basis of U1-D flows. Table 5.8 summarizes the path flow ranges.

Table 5.8 GPS sensitivity analysis cases

Data description*	Probability from U1	Probability from U2	Case description
Hypothetical lowest**	50%	50%	GPS-case1
Min	64%	36%	GPS-case2
25 percentile	74%	26%	GPS-case3
Median	77%	23%	GPS-case4***
75 percentile	78%	22%	GPS-case5
Max	81%	19%	GPS-case6
Hypothetical highest**	99%	1%	GPS-case7

* Data descriptions are on the basis of U1-D flow

**The lowest and the highest values are hypothesized for a comparison purpose even though these values do not exist in our GPS samples

*** GPS case-4 is the path flow proportion used for the proposed network-wide tracking model development and implementation

The sensitivity analysis was performed with seven path flow ranges. Five different ranges of path flow proportions were extracted from the GPS trajectories. For a comparison purpose, two cases that assume extremely biased path flow to one site are tested (i.e., GPS-case 1 and 7). Figure 5.18 illustrates the results of matching accuracies (OSMR1) by vehicle types. First, the tracking performances in two stations were varied by vehicle types. Multi-unit showed higher matching accuracy with a higher proportion of D-U1 flow as opposed to the single-units which has better performances with higher D-U2 flows. Since the GPS information does not provide path flow by vehicle type, aggregated flows from all trucks were applied to the model. However, higher accuracy of single-units when more weights were applied to D-U2 flows indicates that higher volumes of single-units may pass at U2 than U1 to reach downstream location. In other words, higher D-U2 flow proportion obtained from the historical GPS trajectories would make the tracking algorithm keep more single-units in the network, and which results in better performance of tracking for the single-unit. On the other hand, more multi-units

were remained as matches in the network when higher weight was applied to U1-D, and this improves multi-units' matching accuracy.

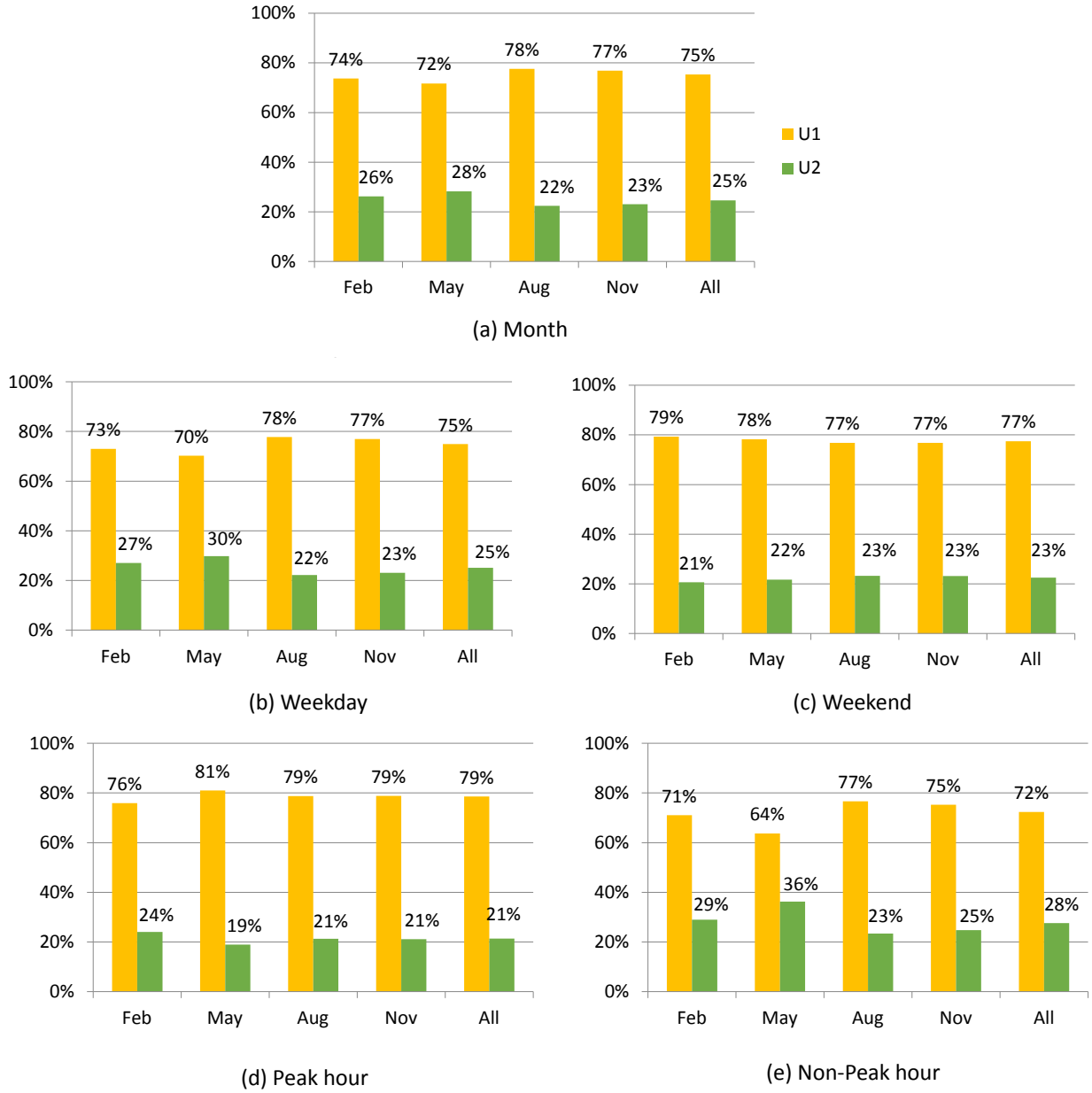


Figure 5.17 Proportion of GPS tracks by paths

However, variances in the GPS path flow rarely affected tracking performances. Only the extremely low or high path flow cases change matching accuracies up to approximately 10 percent.

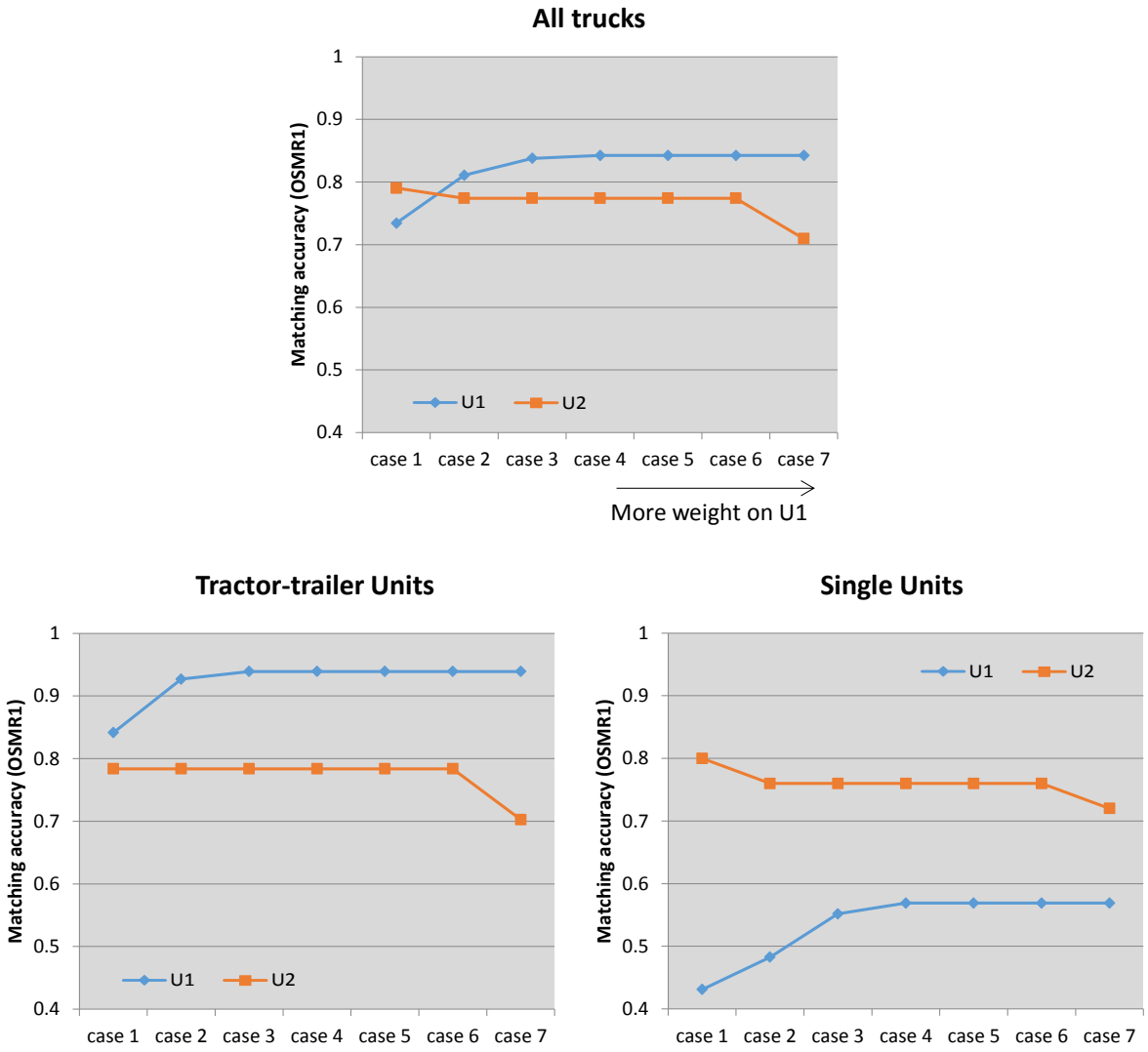


Figure 5.18 Results of GPS sensitivity analysis

5.8.2 Alternative data sources for truck flow information

Nonetheless, GPS data may possess more significant limitations due to the following reasons. First, sampled population may not represent the total population since the GPS trajectories are long-haul truck oriented. This problem raised another concern on single-units samples since the most of single-unit trucks serve local area. Second, temporal bias might be higher depending on the data aggregation interval. Even though the sensitivity analysis carried out in this chapter showed that the tracking performances were little influenced by GPS path flow variations, concerns still remain as the data quality would be significantly varying by location. For example, agricultural areas that observe high variances in flow by season, spatial and temporal variations in GPS data might be much higher than the proposed locations.

Hence, this study considers the use of alternative data source to obtain path flow. Considering that the aggregated volumes were extracted from GPS trajectories for each path, and further, the variations in this flow proportion little affect the matching accuracy, this study recommends a total detected volume from each upstream location as a proxy path flow. Even though the point volume does not represent the path flow, most of detected volume at upstream would pass downstream location in a short distance tracking, which makes this recommendation reasonable. In a long distance tracking, point volumes at multiple ILDs can be used to spatially interpolate and estimate path flow proxy between distant detection sites. In our study area, the proportions of the total detected volume at U1 and U2 are actually matched to the case 4 in the GPS sensitivity case, which is used for the model development and implementation. Since the path flow proxy from point volumes replaces the GPS flow estimates, the model proves its capability as a stand-alone tracking framework only using existing detection systems for path flow estimation in our study.

5.8.3 Model comparisons

One of the significant improvements in modeling between the network-wide tracking and the corridor level tracking is that the signature clustering is performed before the vehicle matching process. As a consequence, the tracking accuracies from the network-wide tracking are expected to be improved than the corridor level tracking. However, this better performance may not be the reason of clustering since the applied data set and tracking location is different for corridor and network-wide tracking. Hence, this section uses the same network-wide data set to compare tracking performances for two models, one with signature clustering and another without signature clustering, for parametric density estimation and feature selection and weighting steps.

Figure 5.19 presents the matching results of OSMR1 and OSMR2 by vehicle types. Overall, the tracking model with the signature clustering outperforms for all vehicle types in both measures. The performance improvements are more significant for single-units than multi-units as 13 percent and 9 percent higher matching accuracies were observed for single-units' OSMR1 and OSMR2, respectively. Due to their shorter length and simple metallic compositions, signatures are less distinguishable among single units. However, important signature features could be more distinguishable after the clustering is performed. We also found that the tracking performances were varied by dataset and detection locations. Multi-unit tracking accuracies were much higher even without the clustering in the network-wide tracking, compared to the corridor level tracking. This could be because the distance of the network-wide tracking is shorter than that of corridor level, which results in less variation in travel time especially for multi-unit trucks. Therefore, fewer matching candidates were dropped out in the search space

reduction step, which keeps more matched vehicles for the vehicle matching model in the network-wide tracking framework.

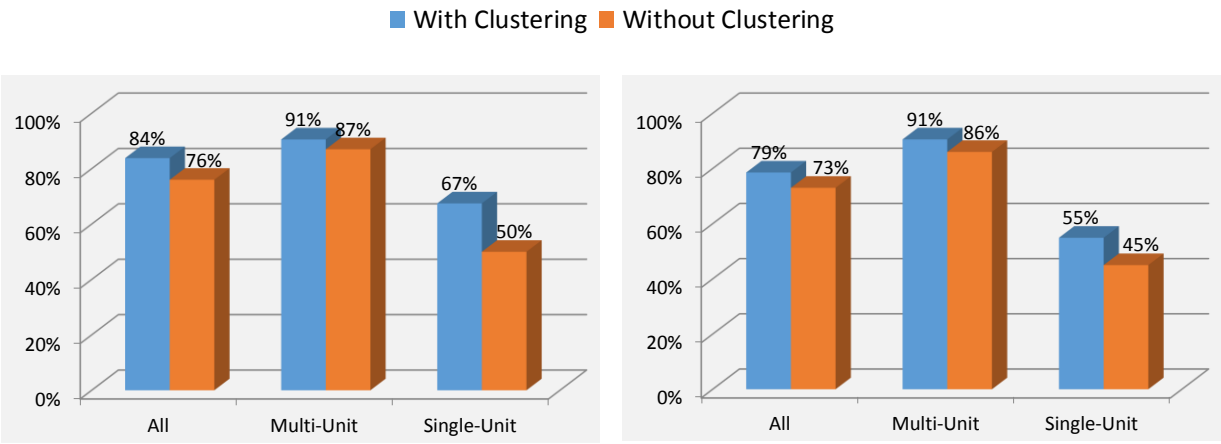


Figure 5.19 Signature clustering model comparison results

5.9 Application

In this section, two applications of the network-wide tracking are introduced. First travel time is estimated from tracked trucks. Second, truck monitoring results implemented in a larger network with an integration of truck body classification model are introduced.

5.9.1 Travel time estimation

Figure 5.20 shows the actual and estimated travel time from the tracking results where each time interval represents 1 minute in the figures. Overall pattern of travel time is depicted by tracked vehicles and MAPE shows error rate at each time interval. The minimum, median, and maximum APE is 0.01, 11.61, and 31.44 percent, respectively. It should be noted that the travel time estimation results can be differently reported with different aggregation time interval or data

filtering process. In this study, a portion of tracked vehicles with higher matching probability are used to estimate travel time, as discussed in Chapter 4.7.2.

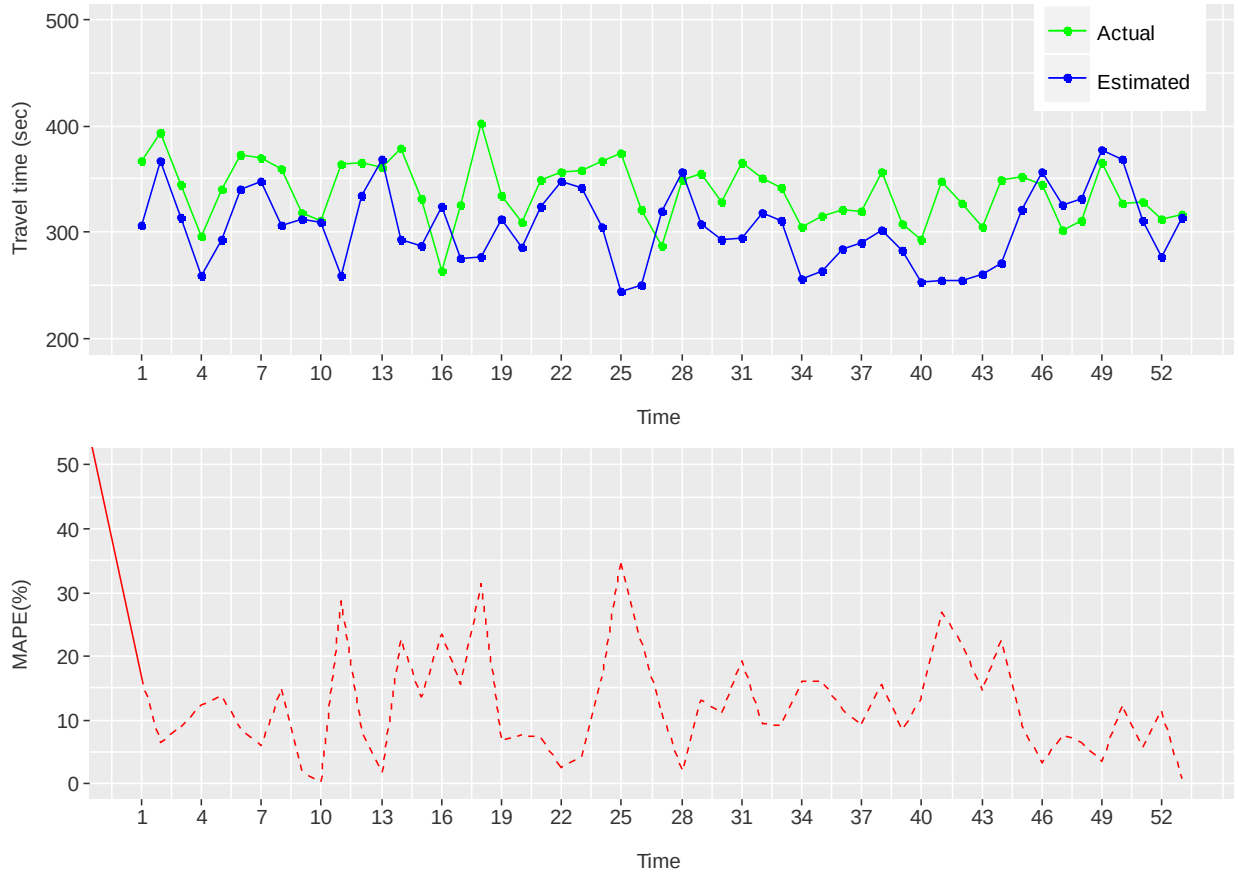


Figure 5.20 Actual and estimated travel time from network-wide tracking

5.9.2 Truck monitoring with detailed body classification

Since the SWBM tracks vehicles among multiple detection locations, trucks can be monitored with their route choice over a large network. This case study chose six ILD locations located on I-210, I-10 and SR-60 in Southern California as shown in Figure 5.21. These corridors are major routes connecting San Bernardino county and Los Angeles county. Since these three highways run parallel and serve as alternative routes for each other, truck tracking

can give us valuable insights on truck route choice by their industry/service types if truck vocation information can be integrated to the tracking model. Signature data and body configuration model estimates are collected on August 3rd, 2016 for 24 hours from UCI-TAMS. This application only considers heavy trucks in tracking process. The tracking model was implemented with total of 10,723 multi-unit trucks at the six detector sites. Table 5.9 demonstrates the total volume including single-units and passenger cars detected at monitoring sites.

Figure 5.21 Site map for network-wide truck monitoring

Table 5.9 Traffic monitoring site description

Site Location	Site Description	Vehicle collected	Multi-unit proportion	Collection Date
Azusa (I-210)	U1 (Upstream Site #1)	46,747	19%	Aug 3rd, 2016 00:00AM – 23:59PM
West Covina (I-10)	U2 (Upstream Site #2)	33,086	12%	
La Puente (SR-60)	U3 (Upstream Site #3)	48,714	23%	
Claremont (I-210)	D1 (Downstream Site #1)	29,696	14%	
Montclair (I-10)	D2 (Downstream Site #2)	54,342	17%	
Chino (SR-60)	D3 (Downstream Site #3)	53,970	16%	

Figure 5.22 shows the body configurations of multi-unit trucks at each detection location. There are seven representative body configurations including enclosed van, port containers (40ft container), domestic container (53ft container), dump, and platform types. The nearly located Los Angeles and Long Beach port complex is the busiest in the U.S., moving $180 billion in cargo between U.S. and Asian countries (White, 2012). Consequently, there are a substantial volume of trucks transporting imported and exported goods in intermodal containers from the Ports to adjacent cities where freight transfer facilities and distributions centers are located (Tok et al., 2016). In particular, 40ft intermodal containers, referred to as 'port trucks', are seen in heavy numbers along the corridors that serve the ports and inland cities. Commodities carried from the port in 40ft intermodal containers are commonly re-packaged into to 53ft containers at inland distribution centers or at near-dock rail yards before being shipped domestically to their final destination (Composition of the Global Fleet of Containers, 2008). Enclosed van is the most common body type for all stations, composed approximately 45 percent to 60 percent of total multi-unit trucks. The second common vehicles are either port container or enclosed van with a refrigeration unit. Site U2, U3, and D3 show the port container as the second common

type whereas site U1, D1 and D2 show the refrigerated enclosed van as the second common type. Notably, U3 and D3 show more domestic containers than other sites.

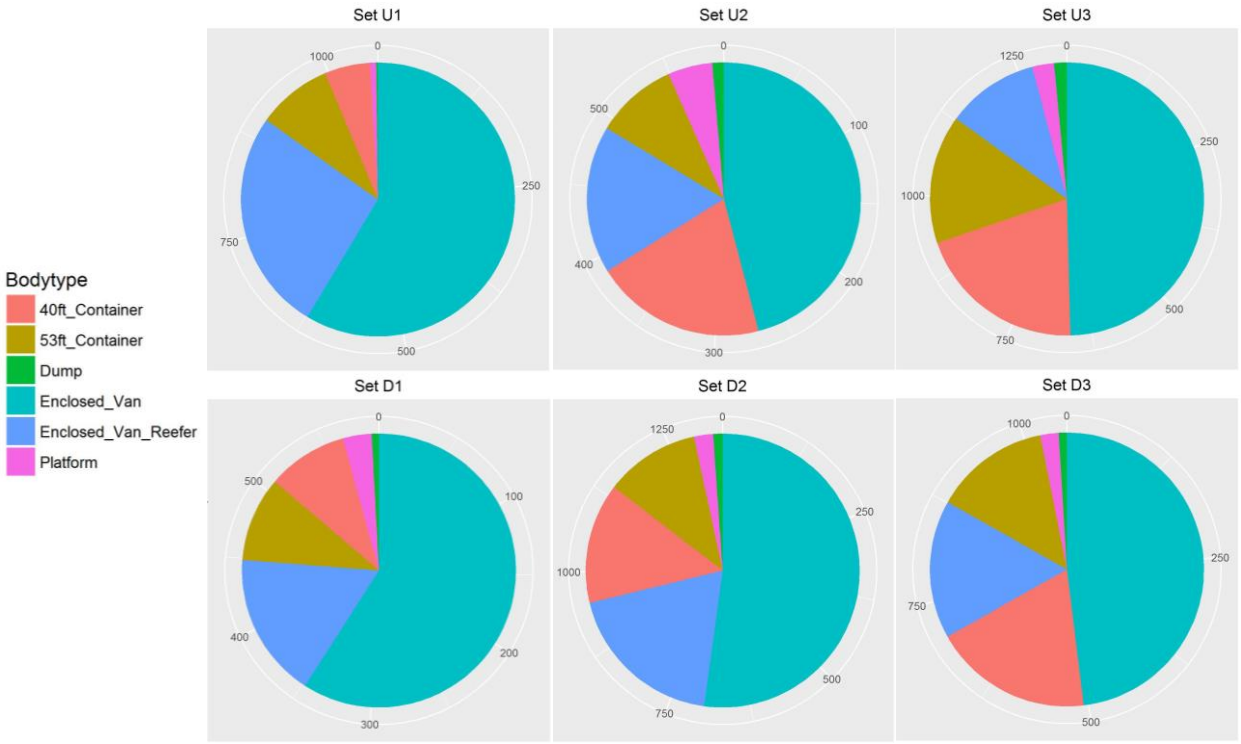


Figure 5.22 Body type at point detection point

Figure 5.23 compares the detected and matched volume for each monitoring site. Higher volume of trucks are observed at particular sites (U1, U3, D2, and D3), which can be explained by their close proximity to truck-related facilities such as warehouses, and intermodal rail facilities. Among the upstream sites, U3 shows the most volumes (11,183) while U2 has the least volumes (3,886). Among the downstream locations, D2 shows the most volumes of 9,394 and the least volume of 4,097 at D1. Interestingly, although U2 and D2 are located on the same corridor, D2 has almost 3 times of truck volumes than U2. Similarly, U1 shows 8,667 trucks; however, truck volume at D1 dropped to 4,097 even though the U1 and D1 are located along the

same corridor. Matched volumes are also illustrated in Figure 5.14. For example, matched volume for U1 and U2 is 1,086 and 632, respectively. The matched vehicle proportions at U1, U2, and U3 are 13%, 16%, and 13% of their total volumes, respectively. At downstream sites of D1, D2, and D3, matched vehicle proportions are 15%, 15%, and 13% of their total volumes. All the stations show comparable results for the matched volume proportions.

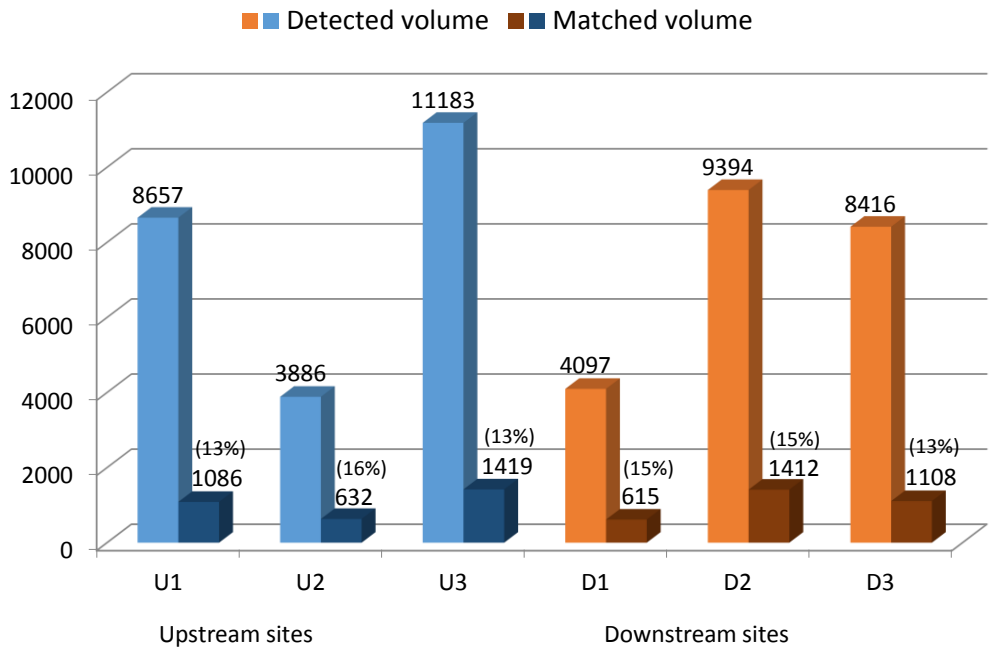


Figure 5.23 Detected and matched volume at monitoring sites

Figure 5.24 presents route flows by upstream location. For example, Figure 5.15 (a) shows the truck flow from U1 to downstream locations of D1, D2, and D3. Trucks passed at U1 more chose D3 than D2 or D1. Although U1 and D1 are located on the same highway, the least proportion of trucks was passed through both locations. Figure 5.15 (b) demonstrates truck flows between U2 and downstream locations. Notably, vehicle flow from U2 to D1 is significantly smaller than other downstream locations. Figure 5.15 (c) shows the truck route

choice for trucks detected at U3. Among 1,419 trucks that passed at U3, half of the trucks passed at D2. Overall, the route that is mostly used for trucks from U1 (I-210) is D3 (CA-60), from U3 (CA-60) and U2 (I-10) is D2 (I-10).

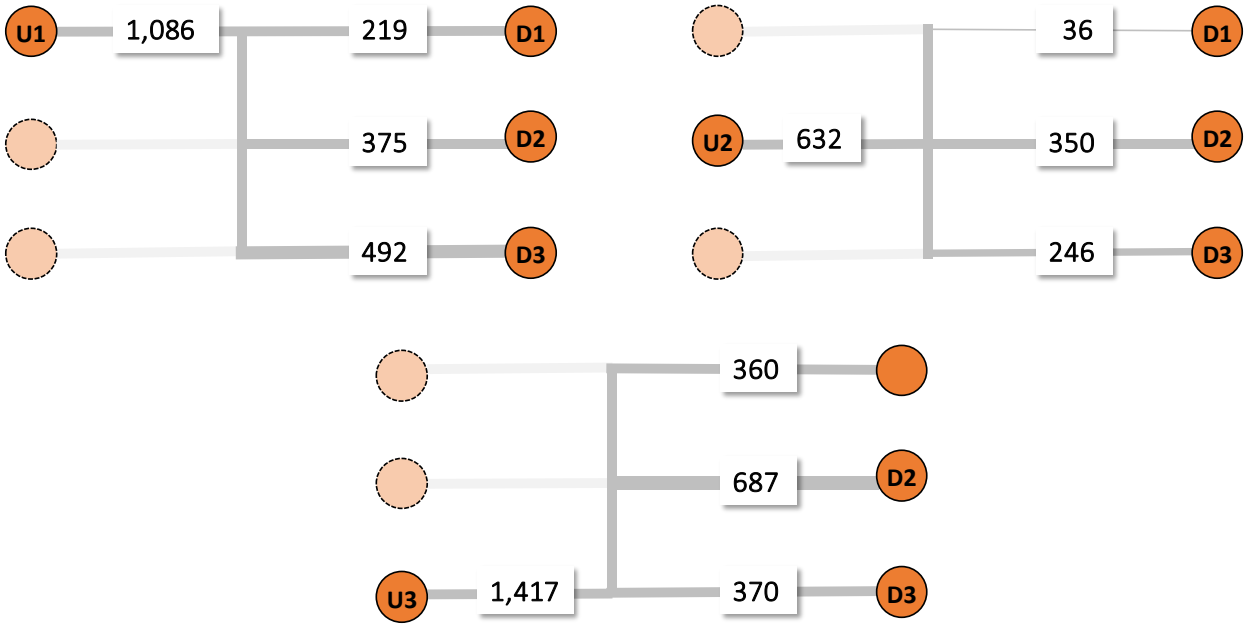


Figure 5.24 Truck flow between upstream and downstream sites

Figure 5.25 compares body configurations by routes. A total of 9 combinations of upstream-downstream sites are presented with six body configurations. As shown in Figure 5.24, the most common body type is enclosed van followed by either 40ft container or refrigerated enclosed van. Specifically, a large amount of container trucks were travelled to D3 (SR-60). This could be because D3 is closely located to two transcontinental rail lines and two international airports. These special geographic and land use zoning properties attract a great number port containers and domestic container movements.

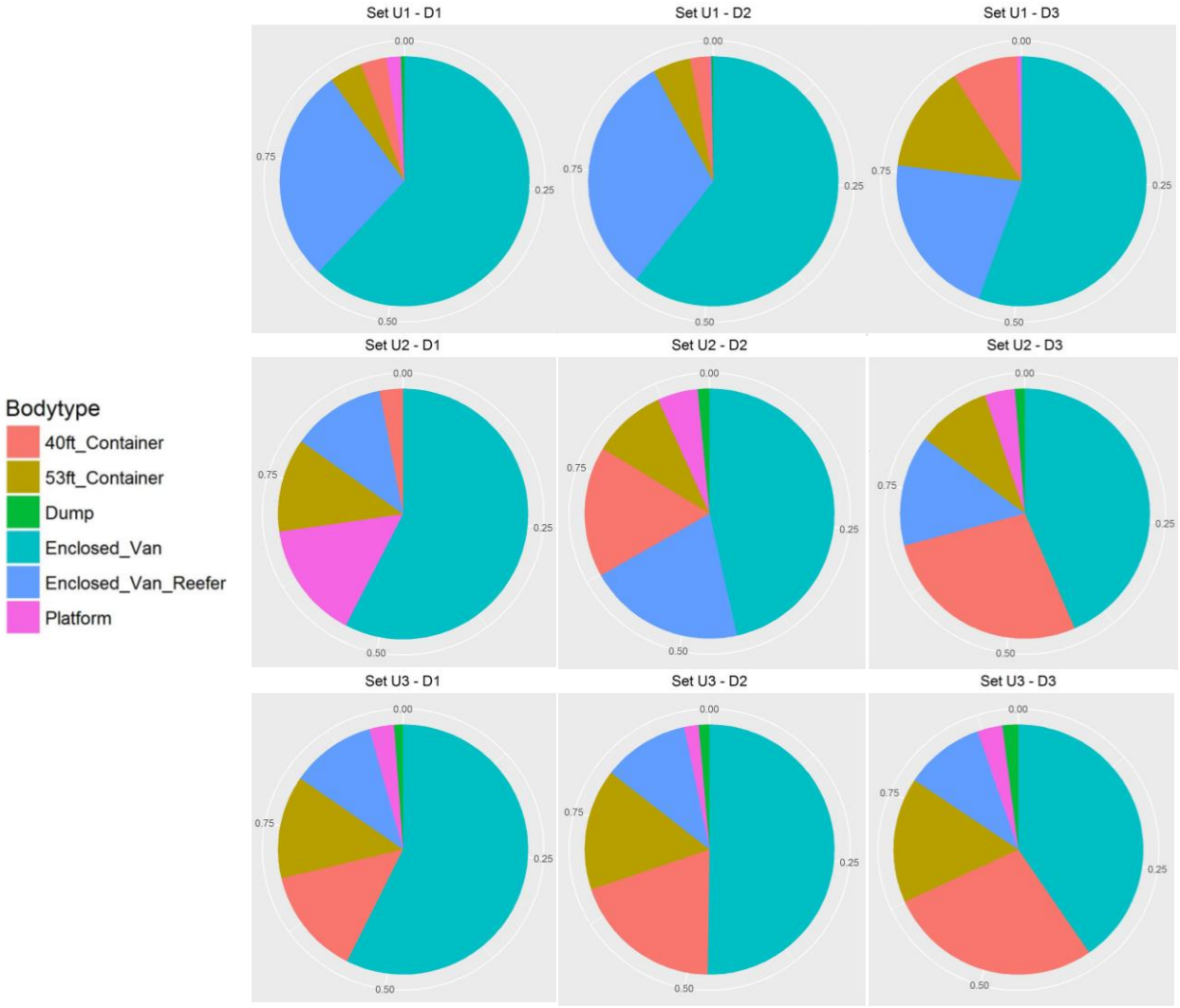


Figure 5.25 Truck body configurations by route

5.10 Conclusion

By extending a corridor level tracking model, a network wide tracking model, was developed in this chapter. Since the network-wide tracking approach considers multiple detector locations across a region, it facilitates to understand spatial and temporal truck flow pattern over a large network. The Bayesian approach (SWBM) considers multiple data sources in the tracking process. The SWBM updates matching prior probability utilizing multiple data sources

to improve matching accuracy. For example, GPS trajectories from ATRI were considered since it captures spatial relationships among detection locations. In addition, individual trucks' body classification estimates are used as additional supplementary data sources since direct comparisons in body types between matched vehicles are expected to affectively filter out potential mismatched and missing vehicles. Since truck route choice can be closely related to their service or affiliations types, hourly volume estimations on body configuration at upstream locations were estimated and used to choose potential upstream locations.

The developed model was tested along approximately 5.5 mile segments on I5 and CA-78 in San Diego, CA. Two upstream and one downstream location were selected where a total of 424 trucks were collected at downstream location. Sixty seven percent of the total trucks were common trucks. To find the best fusion weight, matching accuracies with different combinations of data and fusion weight for signature, body type estimation, travel time, and GPS were compared. As a result, 90 and 67 percent of correct matching rates are shown for multi-unit and single-unit trucks at the best data fusion weight combination. Due to less distinct signature features from less variety of signature shapes, lower performance was observed from the single-units. However, multi-unit units show 91 and 90 percent of correct matching rates for the common and total trucks, respectively. This study further performed sensitivity analysis for GPS data considering various limitations in GPS data including sampling bias and temporal variation. In addition, alternative path flow measures are considered to replace the GPS path flow and total detected volume was recommended as a path flow proxy in our study.

Two different applications of travel time estimation and truck monitoring with detailed body configuration information were implemented. Travel time estimation was performed with direct comparison of time-stamp collected at upstream and downstream locations from the

matched vehicles. Truck monitoring was implemented over a more complex network with six detector locations at port adjacent cities in southern California. Although six sites are closely located, distinct travel pattern was monitored by truck types, which shows the ability of tracking model to analyze temporal and seasonal variations of truck activities by affiliated industry with an integration of body classification modeling.

6. Sensor Location Problem

Truck activity data are of importance for transportation planning and investment analysis, traffic management, environmental and safety analyses, and operation and maintenance of infrastructure. As introduced in the earlier chapters, the developed truck tracking algorithms based on inductive signatures can be one of the most effective and economic method to obtain truck flow data. To maximize benefits from the tracking model, this chapter proposes a strategic plan to identify optimal sensor locations.

There have been considerable efforts devoted to determining optimal sensor locations to measure or estimate accurate traffic flow using sensors. Along with increased implementation of Intelligent Transportation Systems (ITS) technologies, State and regional agencies have utilized various types of sensors to obtain truck flows (Srour, 2006). Recently, vehicle-identification sensors have received attention since vehicle routes and ODs can be easily reconstructed with these sensors. Vehicle-identification sensors include automatic vehicle identification (AVI) systems, electronic tolling technology, and license plate scanning. The vehicle-identification sensors identify vehicle id and track vehicles between sensor locations. In the AVI system, for example, vehicles equipped with transponders communicate with AVI reader stations located in transportation networks.

However, the proposed tracking approach estimates path flow leveraging existing point detector sensors without additional vehicle identification detection systems. Therefore, this sensor location strategy can be applied to both sensor systems including point sensors with tracking model capability and vehicle identification sensors for truck path flow measurement.

In previous studies, identifiability of ODs and routes was the main focus in the sensor location problem, which is referred to as the observability problem. Gentili and Mirchandani (2005) identified the minimum number of sensors that provide all route observability (i.e., full observability) in the network. With a limited budget, their solutions tried to find the locations that observed the most ODs or routes. Castillo et al (2010) addressed problems for full route observability with the minimum number of sensors considering multiple uses of scanning devices. Cerrone (2015) focused on the full observability of routes and showed improved solutions with lower computational effort by examining the temporal order of license plate scans in the mathematical formulation.

The main focus of previous sensor location research was general traffic (i.e., passenger vehicles). However, the mathematical formulations for general traffic measurement can be easily adopted to observe truck flows. This is because the optimal sensor location problem entirely depends on the given vehicle flow in the model. In other words, if truck ODs and route information are used instead of general traffic information in a model, optimal sensor locations for full truck OD and route observability can be obtained. However, for practical reasons in large networks it is usually challenging to implement the full observability of truck ODs. Viti (2014) showed that generally 60 to 70 percent of the total links in the network should have sensor locations in order to fully observe OD pairs. Considering that trucks tend to travel long distances compared to passenger vehicles (Research and Innovative Technology Administration, 2011), a much larger transportation network would typically apply in a truck sensor location problem, which would require substantial investment in installing sensors in a real network application.

Therefore, this chapter focusses on development of a decision model to optimally locate sensors that capture truck flows. Truck travel tends to rely on the industry and commodity types

they serve (Roorda, 2011). Therefore, truck trip ends would be concentrated in particular locations or industrial and commercial areas such as warehouses, depots, and ports. In addition, due to their heavy weight and large physical dimensions, trucks prefer, or are only allowed to use, specific paths (Castillo et al., 2010, Roorda, 2011). In other words, trucks tend to travel between particular ODs using specific routes associated with their vehicle and service types. Hence, optimal locations of sensors for measuring truck movement can be strategically determined by considering their travel patterns. This study therefore focused on optimizing sensor locations that can obtain the maximum flow of truck movements. While observability in the sensor location problem identifies the locations that observe more ODs and routes, the maximum flow capturing problem prioritizes locations that are more utilized by trucks. The proposed approach would be advantageous under a limited budget because some sensor locations that seek observability may not capture a meaningful proportion of truck movements, especially in a large network.

However, relatively little attention has been given to flow capturing approaches even though it would be more desirable in a large network application. Hodgson (1990) initially investigated a flow capturing objective for a sensor location problem and focused on installed sensors capturing the least redundant information on the same OD while ensuring maximum OD flow was observed. Teodorovic et al., (2002) applied a bi-objective solution that maximized the total number of AVI readings and maximized the total number of ODs. Based on the work of Teodorovic et al (2002), Chen et al (2010a, 2010b) applied a multi-objective approach that considered maximizing OD observability and the number of AVI readings while minimizing installation cost. Mirchandani et al (2009) investigated optimal AVI sensor locations that maximized total vehicle miles monitored and minimized the variance of predicted travel times.

Minguez et al (2010) presented optimal traffic plate scanning locations that captured maximum relative route flow proportion for each OD pair under budget constraints.

While the previous studies focused on either OD or route flow capturing, this study investigated both OD and route flow as simultaneous goals in determining optimal sensor locations. In a practical implementation, a primal interest between OD and route flows can be adjusted depending on the specific goals to be achieved. In other words, in regional level planning, truck flows between particular origins and destinations would be the primal interest. However, truck route flow is more essential if a regional project focuses on infrastructure operation and pavement maintenance. While some may argue that both OD and route flows are important for traffic management and environmental impact studies, locating a sufficient amount of sensors that can observe all truck OD and route flows would be infeasible with limited resources and budget. Therefore, our approach introduces a multi-objective sensor location model that provides alternative solutions for both OD and route flow. A goal programming approach was applied to show the trade-offs between ODs and route flows by prioritizing goals, which can provide strategic plans for balancing OD and route flow-based solutions.

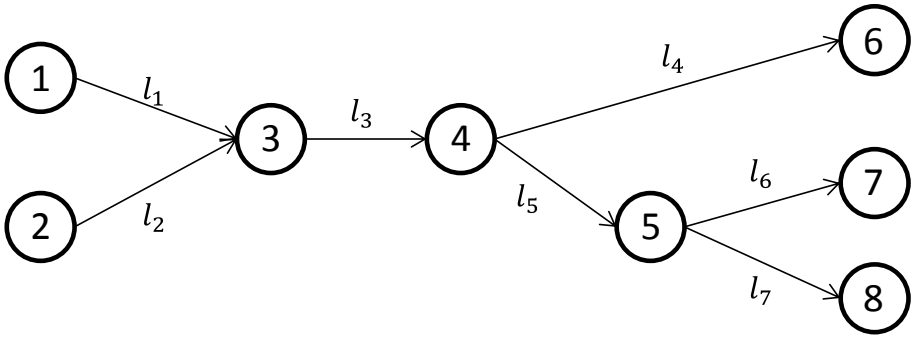
This study applied a proposed framework for truck corridors with actual truck flow data obtained from truck GPS trajectories, unlike previous studies that demonstrated applications of developed models to estimated OD and route flow data obtained from a traffic assignment model. This approach is able to present a more realistic truck travel pattern in a network. Again, it should be noted that the formulation developed in this study can be applied to any type of vehicle id sensors including AVI, license plate recognition, and point sensors of ILDs, for integration with vehicle tracking models.

6.1 Formulation

In this section, two single objective problems that identify sensor locations for maximum OD flow capturing and route flow capturing are initially solved. Based on the solutions from the single objective problems, a multi-objective problem was formulated to capture both OD and route flow using a goal programming approach.

6.1.1 Key link identification

Prior to solving the single objective problems, subsamples of links that predict exact traffic flow in the network, referred to as key links, were initially identified. To illustrate the concept of key links, a simple network composed with four ODs and four routes with seven links was considered as shown in Figure 6.1.



OD (y_a)	Route (z_r)
$\mathbf{y_1 = (1, 6)}$	$\mathbf{z_1 = (l_1, l_3, l_4)}$
$\mathbf{y_2 = (1, 7)}$	$\mathbf{z_2 = (l_1, l_3, l_5, l_6)}$
$\mathbf{y_3 = (2, 7)}$	$\mathbf{z_3 = (l_2, l_3, l_5, l_6)}$
$\mathbf{y_4 = (2, 8)}$	$\mathbf{z_1 = (l_2, l_3, l_5, l_7)}$

Figure 6.1 An elementary example network for key link identification.

In this network, if volume of link 1 (l_1) and link 3 (l_3) are known, the volume of link 2 (l_2) can be estimated by their linear relationship. In this regard, volumes of all the links in a network are not required to achieve full OD and route observability since volumes of some links can be identified using linear relationships with other links. The previous study by Castillo et al (2008) introduced a binary programming problem that solved the optimal subset of links for distinguishing all routes in the network as follows.

$$\text{minimize} \quad L^{\text{route}} = \sum_j x_j \tag{1}$$

$$\text{subject to} \quad x_j \cdot w_{rj}^{route} \geq 1 \tag{2}$$

$$x_j \cdot \xi_{rj}^{route} \geq 1 \tag{3}$$

where $w_{rj}^{route} = 1$ if link j is contained in route r but not in r′ ($r \neq r', r \text{ and } r' \in R$, R indicates all routes considered in the network), ξ_{rj}^{route} indicates a route/link incidence matrix, $\xi_{rj} = 1$ if route r contains link j

The model identifies a minimum subset of links that provide full observability of routes. Constraint 2 guarantees that selected links are able to distinguish route r from all other routes r′ considered in the network. Constraint 3 ensures that all routes contain at least one selected link.

Similar to this formulation, key links for full OD observability can be identified as follows.

$$\text{minimize} \quad L^{OD} = \sum_j x_j \tag{4}$$

$$\text{subject to} \quad x_j \cdot w_{vj}^{OD} \geq 1 \tag{5}$$

$$x_j \cdot \xi_{vj}^{OD} \geq 1 \tag{6}$$

where $w_{vj}^{OD} = 1$ if link j is contained in OD v but not in OD v′ ($v \neq v', r \text{ and } v' \in V$, V indicates all OD pairs considered in the network), ξ_{vj} represents an OD/link incidence matrix, $\xi_{vj} = 1$ if OD v contains link j

6.1.2 Single objective problem

Two single objectives problems, which correspond to maximum flow capturing for ODs and routes, are formulated in this section. With budgetary limitations, sensors cannot be located at all the key links identified from the model (1) – (6) in a large network. Therefore, in the first model, sensor locations that capture ODs with large flows were prioritized to be chosen. An OD pair is captured only if sensors are located at the most upstream and the most downstream links of the corresponding OD. OD flows are assumed to be known from an existing OD matrix.

Model 1. OD flow capturing:

maximize $\sum_v t_v \cdot y_v$ (7)

subject to $y_v \geq \sum_j \delta_{vj} \cdot x_j - 1 \quad \forall v, j \in L^{OD}$ (8)

$y_v - \delta_{vj} \cdot x_j \leq 0 \quad \forall v, \ j \in L^{OD}$ (9)

$\sum_j x_j \leq N \quad j \in L^{OD}$ (10)

$x_j, y_v \ binary \ (0 \ or \ 1)$ (11)

where y_v is a binary variable representing an OD capturing such that $y_v = 1$ if OD v is captured by located sensors, t_v represents the OD flow for OD pair v, δ_{vj} indicates the link/OD incidence matrix where $\delta_{vj} = 1$ if OD v contains the most upstream or the most downstream link j, N indicates the maximum number of sensors considering a budget constraint, and L^{OD} represents the key link for full OD observability

The objective of this problem is to capture the maximum total volume of OD flows with located sensors. Constraint 8 guarantees that if at least two links are selected for sensor locations, a unique OD v should be captured by forcing y_v to be 1. If only one link is selected from constraint 8 to capture an OD, constraint 9 rejects to choose the link as the optimal sensor location. Therefore, both constraint 8 and 9 ensure that the model utilizes sensors most effectively by locating two sensors at the most upstream and downstream links to capture a unique OD. Constraint 10 indicates budget limitations. Constraint 11 requires binary variables

for OD and link selection. All of the selected links should be the key links identified from the previous model (j $\in L^{OD}$).

For comparison purposes, an observability model that seeks maximum OD observability is also presented. The maximum observability can be obtained with the objective function of maximizing $\sum_v y_v$ with the constraint set (8) – (11), referred to as *Model 2 (OD observability model).*

Model 2. OD observability model:

maximize $\sum_v y_v$ (12)

subject to $y_v \geq \sum_j \delta_{vj} \cdot x_j - 1 \quad \forall v, j \in L^{OD}$ (13)

$y_v - \delta_{vj} \cdot x_j \leq 0 \quad \forall v, j \in L^{OD}$ (14)

$\sum_j x_j \leq N \quad j \in L^{OD}$ (15)

x_j, y_v *binary* (0 *or* 1) (16)

Another single objective model for maximum route flow capturing was also investigated. Similar to the OD flow capturing model, sensor locations that capture routes utilized by higher volumes of trucks are prioritized to be selected in this model.

Model 3. Route flow capturing:

maximize $\sum_r p_r \cdot z_r$ (17)

subject to $z_r \geq \sum_j \xi_{rj} \cdot x_j - (n^r - 1) \quad \forall r,\ j \in L^{route}$ (18)

$z_r - \xi_{rj} \cdot x_j \leq 0 \quad \forall r,\ j \in L^{route}$ (19)

$\sum_j x_j \leq N \qquad j \in L^{route}$ (20)

$x_j, y_a\ binary\ (0\ or\ 1)$ (21)

where z_r is a binary variable representing a route flow such that $z_r = 1$ if route r is captured from located sensors, p_r is out-of-date route flow, ξ_{rj} indicates route/link incidence matrix where $\xi_{rj} = 1$ if route r contains link j, n^r is a number of key links in route r, L^{route} represents the key link for full route observability

An objective function 17 identifies sensor locations that observe higher route flow. Constraint 18 and 19 together ensure that routes are distinguished with the minimum number of links. If links cannot identify any unique route, they are not chosen as the optimal locations. In other words, if a link j cannot capture unique route r, z_r is forced to be a zero in constraint 19. On the contrary, if all key links in route r are selected, z_r forced to be one in constraint 18.

For comparison purposes, the route observability problem is presented with the objective function of maximizing $\sum_r z_r$ with the constraint set (18) – (21), referred to as *Model 4 (Route observability model).*

Model 4. Route observability model:

maximize maximizing $\sum_r z_r$ (22)

subject to $z_r \geq \sum_j \xi_{rj} \cdot x_j - (n^r - 1) \quad \forall r,\ j \in L^{route}$ (23)

$z_r - \xi_{rj} \cdot x_j \leq 0 \quad \forall r,\ j \in L^{route}$ (24)

$\sum_j x_j \leq N \qquad j \in L^{route}$ (25)

x_j, y_a $binary$ $(0\ or\ 1)$

6.1.3 Goal programming approach

Through sections 6.1.1 and 6.1.2, key links and optimal sensor locations that capture maximum OD and route flow, were formulated by single objective models. In this section, both OD and route flows are considered together as multi-objectives to be captured by located sensors.

Instead of producing an optimal solution as shown in the single objective problem, multi-objective problems introduce Pareto-optimality that shows non-dominant solutions (Hwang and Masud, 2012). In Pareto-optimality, solutions cannot be improved without worsening other solutions. Therefore in a multi-objective problem, there can be a solution that reaches optimality for one objective, however the solution may not be the optimal in another objective.

Consequently, Pareto-optimality finds optimality in the most efficient manner where no other alternatives produce better solutions (Hwang and Masud, 2012). A goal programming approach is one of the widely used techniques to solve multi-objective problems (Charnes and Cooper 1977; Tamiz et al., 1998). Goals are implemented to be met as closely as possible. This goal programming model introduces a deviation term in the objective function and the deviation should be minimized to reach the goals. Often weights are considered in the objective function so that more important goals can be met more closely (Charnes and Cooper 1977). The mathematical formulations of the goal programming (*Model 5*) for maximizing OD and route flow capturing are introduced as follows.

Model 5: multi-objective model for maximum OD and route flow capturing

$$\text{minimize} \quad w_1 \cdot d_1^+ + w_2 \cdot d_2^+ \tag{26}$$

$$\text{subject to} \quad \sum_v t_v y_v + d_1^+ - d_1^- = g_1 \tag{27}$$

$$\sum_r p_r \cdot z_r + d_2^+ - d_2^- = g_2 \tag{28}$$

$$y_v \geq \sum_j \delta_{vj} \cdot x_j - 1 \quad \forall v, j \in L^{OD} \tag{29}$$

$$y_v - \delta_{vj} \cdot x_j \leq 0 \quad \forall v, \ j \in L^{OD} \tag{30}$$

$$z_r \geq \sum_j \xi_{rj} \cdot x_j - (n_j^r - 1) \quad \forall r, \ j \in L^{route} \tag{31}$$

$$z_r - \xi_{rj} \cdot x_j \leq 0 \quad \forall j \in L^{route} \tag{32}$$

$$\sum_j x_j \leq N \qquad j \in [L^{OD} \cup L^{route}] \tag{33}$$

$$d_1^+, d_1^-, d_2^+, d_2^-, w_1, w_2 \geq 0 \quad (34)$$

$$x_j,\ \mathrm{y_v}\ binary\ (0\ or\ 1) \quad (35)$$

where d_1^+, d_1^-, d_2^+, d_2^- are the deviations from the goals, w_1 and w_2 are the weights for goal 1 (OD flow capturing) and goal 2 (route flow capturing), respectively

The objective function of equation 26 solves the minimum deviation from the goals. Constraints 27 and 28 demonstrate the goal for two objectives which correspond to OD and route flow capturing. The goals $\mathrm{g_1}$and $\mathrm{g_2}$ can be obtained from the optimal solution of the single objective model 1 and model 3, respectively. Since the objective is to find the maximum OD flow, d_1^- in constraint 27 should be determined as zero. Similarly, d_2^- becomes zero in constraint 29. Therefore only d_1^+ and d_2^+ remain as deviation terms to be minimized in the objective function 26. Constraints 29 to 33 and 35 follow the same formulations of the previous single objective problems, and constraint 34 indicates the non-negativity to the deviation and weight terms.

6.2 Sample network example – Nyuyen-Dupuis network

The proposed single and multi-objective models were implemented with a sample network example. A widely used network example, Nguyen-Dupuis network, was utilized in this study. The network is composed of a total of 13 nodes and 38 links as shown in Figure 6.2. Corresponding OD and route information were obtained from the study by Castillo (17) as presented in Table 6.1.

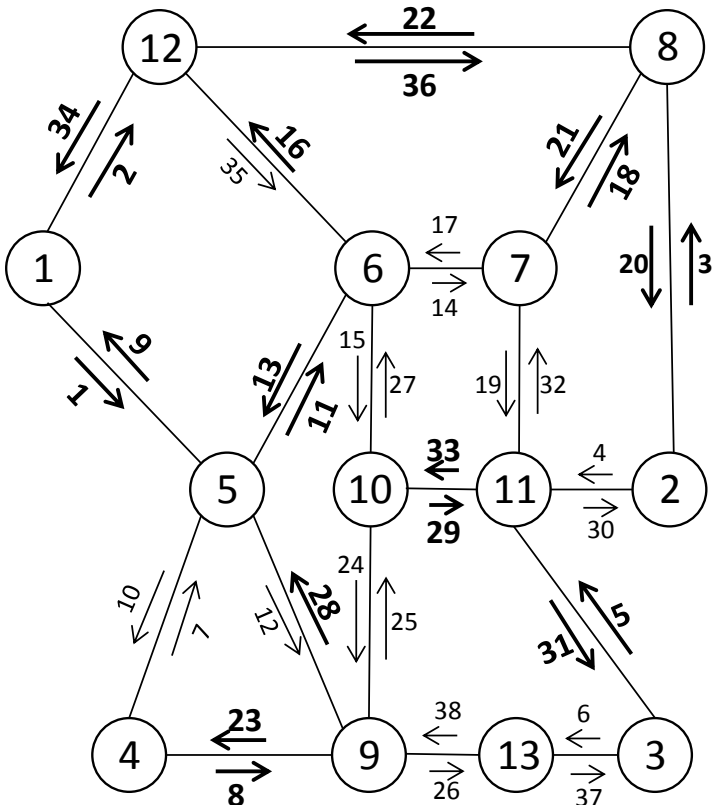


Figure 6.2 Nguyen-Dupuis network

Table 6.1 OD Pairs and Routes in the Nguyen-Dupuis Example

OD & Route		Links	OD flow	Route flow
OD pair 1 (1-2)				
	route 1	1-11-14-18-20		0.1203
	route 2	2-35-14-18-20	210	0.0579
	route 3	2-36-20		0.8218
OD pair 2 (1-3)				
	route 4	1-11-14-19-31		0.1574
	route 5	1-11-15-29-31		0.1894
	route 6	1-12-25-29-31	430	0.0618
	route 7	1-12-26-37		0.4311
	route 8	2-35-14-19-31		0.0758
	route 9	2-35-15-29-31		0.0891
OD pair 3 (1-8)				
	route 10	1-11-14-18	320	0.1203
	route 11	2-35-14-18		0.0579
	route 12	2-36		0.8218
OD pair 4 (2-1)				
	route 13	3-21-17-13-9	210	0.1268

	route 14	3-21-17-16-34		0.0655
	route 15	3-22-34		0.8077
OD pair 5 (2-4)				
	route 16	3-21-17-13-10	320	0.1171
	route 17	3-21-19-33-28-23		0.0376
	route 18	4-33-28-23		0.8453
OD pair 6 (2-12)				
	route 19	3-21-17-16	50	0.076
	route 20	3-22		0.925
OD pair 7 (3-1)				
	route 21	5-32-17-13-9		0.1497
	route 22	5-32-17-16-34		0.0773
	route 23	5-33-27-13-9	430	0.172
	route 24	5-33-27-16-34		0.0888
	route 25	5-33-28-24-9		0.0781
	route 26	6-38-24-9		0.434
OD pair 8 (3-4)				
	route 27	5-33-28-23	110	0.1526
	route 28	6-28-23		0.8474
OD pair 9 (3-12)				
	route 29	5-32-17-16	40	0.4653
	route 30	5-33-27-16		0.5374
OD pair 10 (4-2)				
	route 31	7-11-14-18-20	320	0.1225
	route 32	8-25-29-30		0.8416
	route 33	8-25-29-32-18-20		0.0359
OD pair 11 (4-3)				
	route 34	8-25-29-31	110	0.1253
	route 35	8-26-37		0.8747
OD pair 12 (4-8)				
	route 36	7-11-14-18	210	0.7735
	route 37	8-25-29-32-18		0.2265
OD pair 13 (8-1)				
	route 38	21-17-13-9	320	0.1268
	route 39	21-17-16-34		0.0655
	route 40	22-34		0.8077
OD pair 14 (8-4)				
	route 41	22-17-13-10	210	0.7567
	route 42	21-19-33-28-23		0.2433
OD pair 15 (8-12)				
	route 43	21-17-16	60	0.075
	route 44	22		0.925
OD pair 16 (12-2)				
	route 45	35-14-18-20	50	0.0659
	route 46	36-20		0.9341
OD pair 17 (12-3)				
	route 47	35-14-19-31	40	0.4598
	route 48	35-15-29-31		0.5402

OD pair 18 (12-8)				
	route 49	35-14-18	60	0.0659
	route 50	36		0.9341

This study used the solver from the optimization software Gurobi (2012) using a Python API to solve the proposed models for the network examples. As shown in Table 6.2, a total of 18 and 10 key links were identified for the full route and OD observability models (1) - (3) and (4) - (6), respectively. Therefore, only 47 percent (=18/38) and 26 percent (=10/38) of links should be monitored with sensors for full route and OD observability in the Nguyen-Dupuis network. The attributes of these key links are illustrated in bold in Figure 6.2.

Table 6.2 Key links in the Nguyen-Dupuis Example

Model	Key link
Full route observability model	Link 1, 2, 3, 5, 8, 9, 11, 13, 18, 20, 21, 22, 28, 29, 31, 33, 34, 36
Full OD observability model	Link 1, 3, 5, 8, 16, 18, 20, 21, 23, 31

The single objective solutions from Model 1 to Model 4 are compared in Figure 6.3. Figure 6.3(a) compares solutions for OD flow capturing and number of observed ODs from Model 1 (OD flow capturing model) and Model 2 (OD observability model) by increasing number of sensors from 1 to 18. Figure 6.3(b) compares solutions for route flow capturing and number of observed routes from Model 4 and Model 5. As shown in Figure 6.3(a), the observability model and the flow capturing model observed the same number of ODs when the same number of sensors was located. However, significantly higher OD flow was captured when

the flow capturing model was applied. This result indicates that the flow-capturing model outperformed the observability model on obtaining the OD information. Similarly, the flow capturing model observed generally higher route flows than the observability model. However, the number of captured routes was higher with the route observability objective.

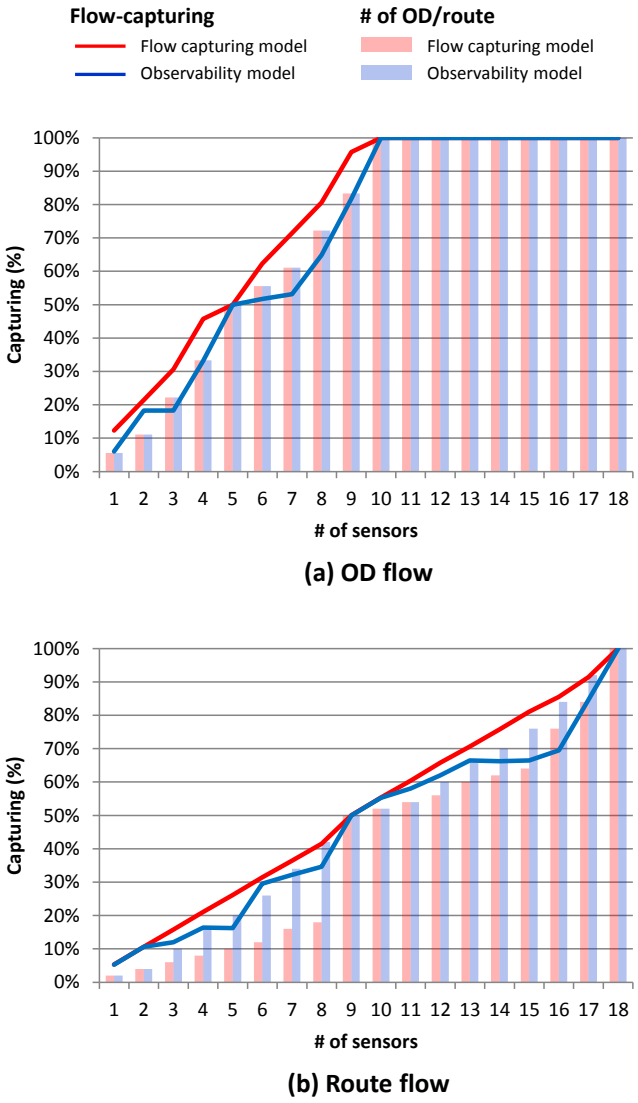


Figure 6.3 Results from the single objective problems for (a) OD and (b) route flow with Nguyen-Dupuis network.

Table 6.3 shows results of the goal programming approach by different weight combinations for the OD and route flow-capturing objectives. The optimal OD and route flows from the single objective problem were used as goals in the goal programming model. Results showed that the optimal locations differed according to the weights applied to the two objectives. When the weight of OD and route was equally applied (weight [OD:Route]=1:1), captured OD and route flow was 62% and 20%, respectively, with seven located sensors. However, if the OD flow capturing objective was prioritized over the route flow capturing objective (weight [OD:Route]=2:1), captured OD flow was increased to 71% while captured route flow was reduced to 9%. On the contrary, when the objective of route flow capturing was prioritized over the OD flow capturing objective (weight [OD:Route]=1:2), captured route flow was increased to 32% while captured OD flow was decreased to 50%. From these results, it can be concluded that similar proportions of ODs and route flow can be captured when more weight is applied to the route flow objectives because much higher OD flows than route flow can be measured with a small number of located sensors.

Table 6.3 Goal Programming Model Results with Different Weight Factors.

Weight factor[OD : Route] = 1:1					
# sensors	Selected links	# OD (%)	OD flow (%)	# Route (%)	Route flow (%)
3	1, 5, 31	3 (17%)	900 (26%)	3 (6%)	222 (6%)
7	1, 5, 8, 18, 20, 29, 31	10 (56%)	2180 (62%)	12 (24%)	716 (20%)
10	1, 3, 5, 8, 18, 20, 21, 23, 29, 31	15 (83%)	3350 (96%)	14 (28%)	724 (21%)
14	1, 2, 3, 5, 8, 11, 13, 18, 20, 21, 23, 29, 31, 36	15 (83%)	3350 (96%)	30 (60%)	1975 (56%)
16	1, 2, 3, 5, 8, 11, 13, 18, 20, 21, 22, 23, 29, 31, 34, 36	15 (83%)	3350 (96%)	37 (74%)	2573 (74%)
Weight factor [OD : Route] = 2:1					
# sensors	Selected links	# OD (%)	OD flow (%)	# Route (%)	Route flow (%)
3	5, 21, 23	3 (17%)	1070 (31%)	3 (6%)	23 (1%)
7	1, 5, 8, 18, 20, 21, 31	10 (56%)	2500 (71%)	12 (24%)	330 (9%)
10	1, 3, 5, 8, 18, 20, 21, 23, 29, 31	15 (83%)	3350 (96%)	14 (28%)	724 (21%)
14	1, 2, 3, 5, 8, 11, 16, 18, 20, 21, 23, 29, 31, 36	15 (83%)	3500 (100%)	30 (60%)	1779 (51%)
16	1, 2, 3, 5, 8, 11, 16, 18, 20, 21, 22, 23, 29, 31, 34, 36	15 (83%)	3500 (100%)	37 (74%)	2377 (68%)
Weight factor [OD : Route] = 1:2					
# sensors	Selected links	# OD (%)	OD flow (%)	# Route (%)	Route flow (%)
3	1, 5, 31	3 (17%)	900 (26%)	3 (6%)	222 (6%)
7	1, 8, 11, 18, 20, 29, 31	9 (50%)	1750 (50%)	17 (34%)	1112 (32%)
10	1, 2, 5, 8, 11, 18, 20, 29, 31, 36	10 (56%)	2180 (62%)	26 (52%)	1771 (51%)
14	1, 2, 3, 5, 8, 11, 18, 20, 22, 23, 29, 31, 34, 36	13 (72%)	2820 (81%)	31 (62%)	2334 (67%)
16	1, 2, 3, 5, 8, 11, 13, 18, 20, 21, 22, 23, 29, 31, 34, 36	15 (83%)	3350 (96%)	37 (74%)	2573 (74%)

A sensitivity analysis on weight factors was performed is illustrated in Figure 6.4. When equal or more weight was given to the OD flow capturing than route flow capturing objectives, over 90% of OD flows were captured with only nine sensors while only 50% of route flows were measured. A higher weight on OD flow, such as [OD:route] of [1:1], [2:1], and [3:1], did not demonstrate noticeable differences in OD flow capturing outcomes. However route flow

capturing showed more sensitive results to the applied weights. These findings provide insight on the optimal weights for multiple goals under a limited budget. For example, if the goal is observing at least 50% of ODs and route flows under a budget of maximum of 10 sensors, the weight on the route flow capturing objective should be twice that of the OD flow capturing objective. However, if the objective is to observe at least 70% of ODs and route flows with a budget of maximum of 15 sensors, much higher weight on route flow capturing would be recommended (e.g., [OD:route] =[1:3]).

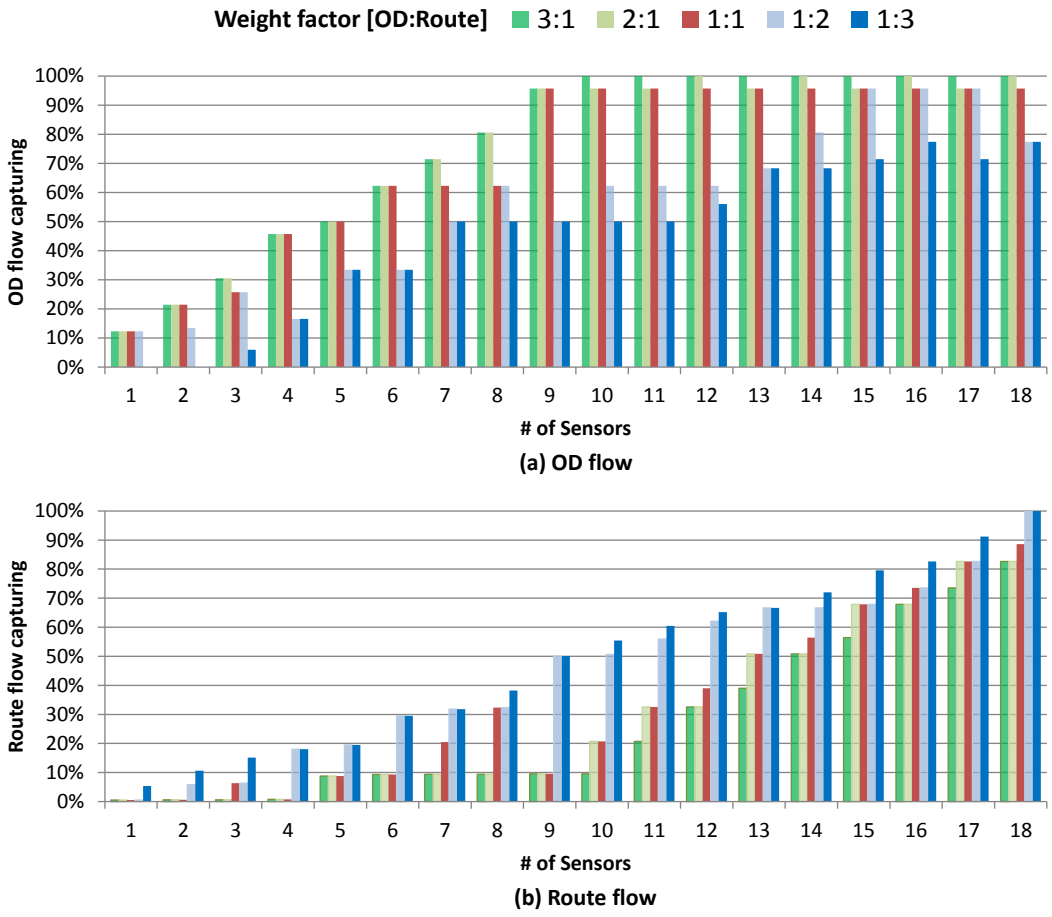


Figure 6.4 Weight factor sensitivity analysis for (a) OD flow capturing and (b) route flow capturing

6.3 Real network example – Los Angeles network

The proposed models were implemented in a real network example in Los Angeles, California. While previous studies typically used OD and route flows obtained from a traffic assignment model, this study applied actual truck flows sampled from GPS tracks. The GPS data were obtained from the American Transportation Research Institute (ATRI) and collected from four weeks in each quarter in 2010. Among a total of 144 OD pairs, OD pairs that contained at least two sensors between OD locations and had at least two routes were extracted for this study, which included 15,323 GPS trajectories as illustrated in Figure 6.5.

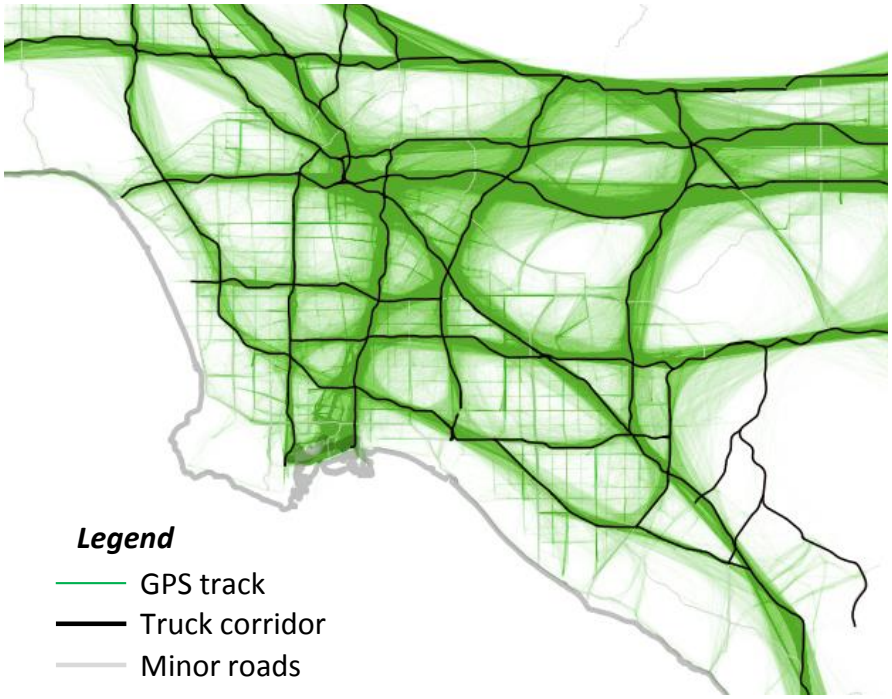


Figure 6.5 GPS trajectory in Los Angeles area.

The study area included 144 links consisting of 65 OD pairs and 144 routes. From the previous formulations (1)-(6), key links were initially determined. A total of 74 and 49 key links, corresponding to 51% and 34% of network links, were identified as the key links for full route and OD observability, respectively. Results from the single objective models are compared in Figure 6.6. Both OD and route flow capturing were significantly higher with the flow-capturing model than with the observability model, while the captured number of ODs and routes was higher in the observability models. For example, if 20 sensors were located, only 40 percent of route flow could be captured in the observability model (model 4) while 75 percent of flow could be captured in the flow capturing model (model 3). However, differences in the number of observed routes between the observability and flow capturing models were at most 10 percent for both OD and route observability.

Figure 6.7 presents flow capturing results from the goal programming approach for different weight factors. Overall, more balanced solutions for OD and route flow capturing were achieved when the same weights were applied to the two objectives when more than 30 sensors were located in the network. However, if fewer than 30 sensors were present, flow capturing results significantly depended on the weights applied to the different objectives. This is because, as shown in the Figure 6.6, both OD and route flows captured by additional sensors were significantly increased by including up to 30 sensors. Therefore, flow capturing results were highly sensitive to the changes in applied weights with less than 30 located sensors.

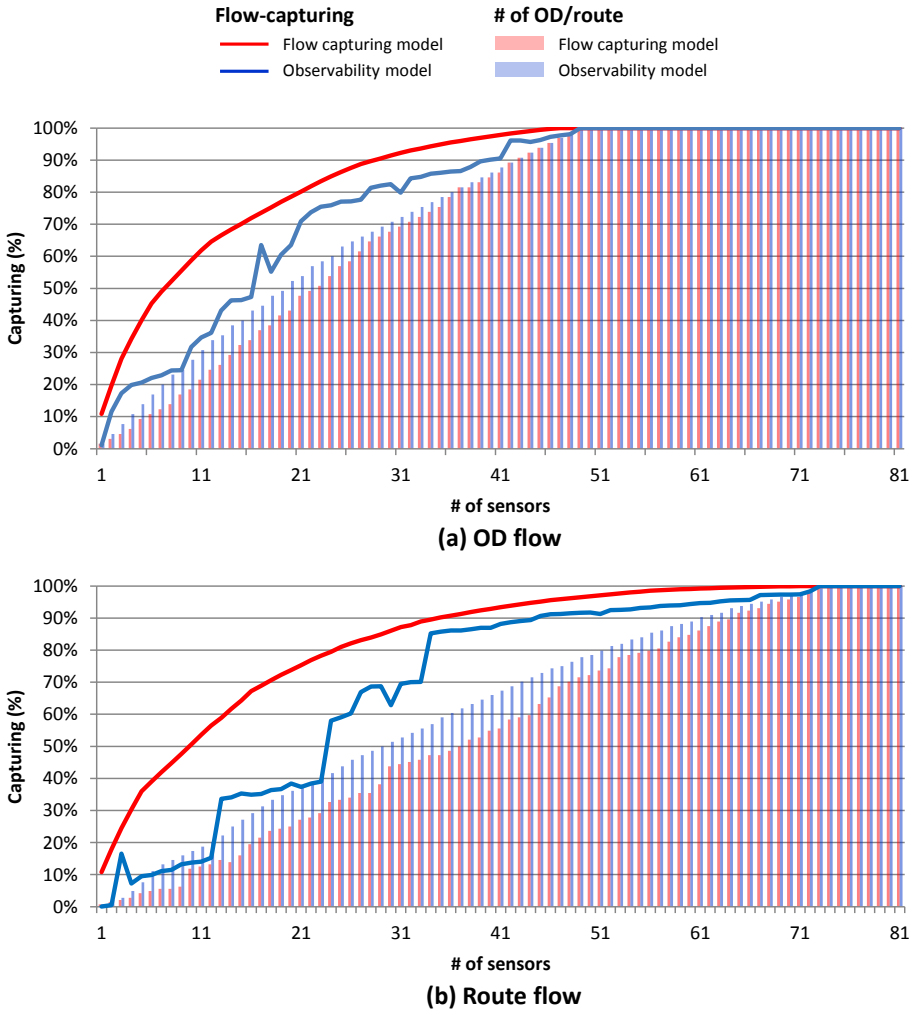


Figure 6.6 Results from the single objective problems for (a) OD and (b) route flow with Los Angeles network.

Interestingly, compared to the Nguyen-Dupuis network, the Los Angeles example showed different sensitivity for the weight factors on flow capturing results. While the Nguyen-Dupuis network showed that more weights on the route flow objective yielded balanced OD and route flow capturing results, the Los Angeles example produced similar OD and route flow capturing when the same weights were applied to two objectives. Considering that the Nguyen-

Dupuis network captured general traffic patterns from the traffic assignment models, and the Los Angeles network is based on truck travel trajectories, this result may be caused by different travel patterns of passenger vehicles and trucks. In other words, distinct truck travel patterns such as route-specific trips for particular ODs may yield such results since more weight on ODs produced excessively biased results in OD flow-capturing, and higher weights on routes provided inferior results than when equal weights were applied in the Los Angeles network.

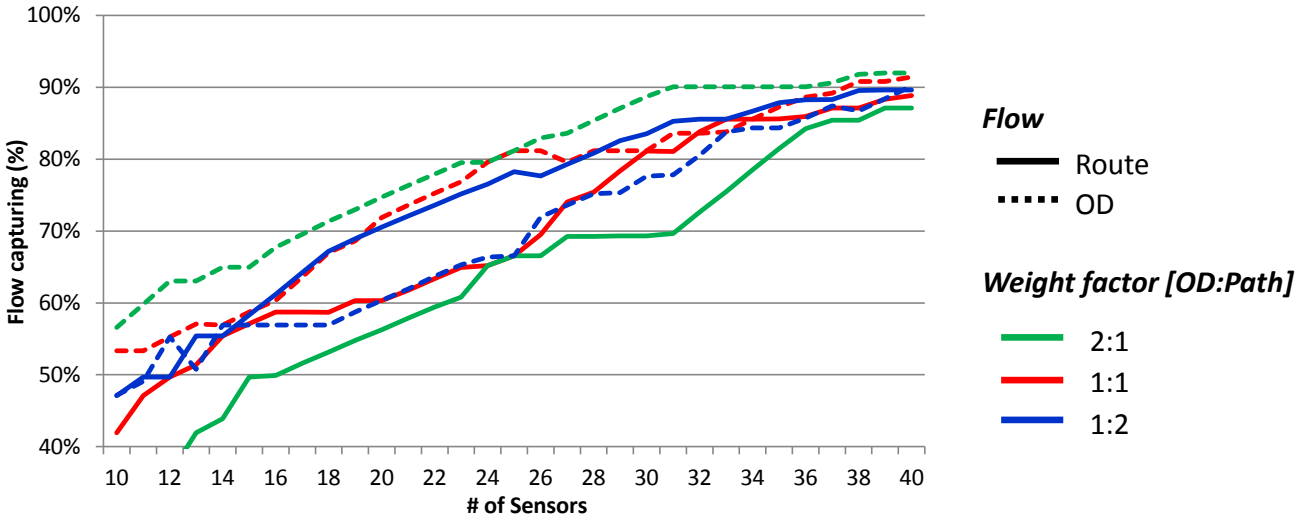


Figure 6.7 Flow capturing results from the goal programming approach.

6.4 Conclusion

This chapter investigated optimal sensor locations for monitoring truck movements. In this paper, practical implications were given highest consideration on identifying optimal sensor locations. First, this study focused on maximum flow capturing as an optimization problem while previous studies focused on identifying observability of ODs and routes. Since trucks travel long distances, the conventional full observability problem cannot be implemented in a

large-sized truck network. However, considering trucks' distinct travel patterns that include frequent trips to specific ODs with particular routes, sensor locations that observe more flow would provide a better understanding of truck flows under budgetary limitations than the sensor locations that observe more ODs and routes. Second, the proposed model suggested a multi-objective approach to achieve both OD and flow capturing in a sensor location model. Our approach considers maximum OD and path flow capturing as goals to be achieved; however, these goals were differently prioritized in the model depending on a prior objective. Several alternative solutions with different weight factors that prioritize those objectives were compared, and trade-off relationships between OD and route flow capturing were identified under a limited budget. Application of the proposed model to a sample Nguyen-Dupuis network showed that significantly larger OD and route flows were captured by the proposed flow-capturing model than the conventional maximum observability approach. The results from the Los Angeles network with actual truck flow data showed that more balanced solutions for OD and route flow can be obtained by equally weighting objectives of OD and route flow capturing when sufficient numbers of sensors are located in the network. However, the weight factor should be adjusted by considering budget and the objective OD and route flow-capturing.

These findings indicated that a sensor location model should be determined by analyzing the various objectives. The proposed goal programming approach allows decision makers to investigate alternative sensor locations by considering trade-offs between OD and path flow capturing and to determine the appropriate priority of goals by considering systematic truck flow coverage and the budget. Therefore, the empirical results of this study provide insights to practitioners on sensor location problems as a decision support tool for strategic planning.

7. Concluding Remarks

In California, the recently published sustainable freight action plan specified multiple goals to improve freight efficiency and competitiveness throughout the state. Two pilot projects in the action plan seek specifically solutions to reduce freight congestions and encourage zero emissions in critical truck corridors using advanced technologies. This dissertation proposes a framework for tracking and monitoring truck flow in a highway network utilizing existing point detection systems. Compared to the current data sources, the proposed approach is more advantageous to assist aforementioned pilot projects since spatially and temporally varying truck flows from total truck population can be continuously tracked and monitored in a large network by leveraging existing detection systems. In addition, the proposed modeling contributes to provide not only truck path information but also detailed truck attributes such as axle configuration and type with an integration of previously developed truck body and axle configuration models. These are of critical importance in freight modeling, emissions estimation, pavement maintenance, and vehicle surveillance.

7.1 Summary of contribution

In this dissertation, truck tracking was implemented at two spatial scopes of corridor and network-wide levels. While the corridor level tracking utilizes the complementary two detection systems, ILD and WIM, the network-wide model uses ILD systems with supplementary data sources. For a modelling, Bayesian approach was applied with extensive feature selection and weighting techniques.

The proposed tracking model composes with two sub-models. In the first model, every truck detected at downstream location (i.e., target) finds their matched vehicle (i.e., candidate) at upstream location. Specifically, matching probabilities of every target and its corresponding candidate vehicles were estimated based on their features represented by physical attributes and waveform signatures. The vehicle pair with the highest matching probability is declared as a match at the first model. Since the tracking is implemented over a long distant corridor that contains multiple highway interchanges and ramps, not every vehicle detected downstream site would not pass the upstream sites. Therefore, the second filtering model handles those vehicles (i.e., missing) that enter or leave the tracking corridor during the tracking process.

The tracking model includes novel feature selection and weighting methods. The features that had greater discerning ability to identify vehicles were prioritized over the weaker ones, resulting in higher matching accuracy for truck population tracking across longer distances. Considering that trucks have similar physical attributes or signature shapes by their axle configurations and trailer types, this feature selection and weighting techniques help identifying vehicles and improving matching accuracy even in a complex road network. For the network-wide tracking, signature clustering was performed to group signatures by waveform shapes prior to the feature selection. This approach enhances vehicle matching and filtering through separate feature density estimation and weighting by signature clusters since discernible features that are more capable of identifying salient differences among vehicles were better captured in clusters.

In addition, the proposed tracking algorithm features a relaxed temporal search space to increase the likelihood of including the true matched vehicle even over long distances. Several data pre-processing methods including signature transformation and WIM auto-calibration steps

were performed to minimize the impacts of measurement variations from the detection systems employed in the tracking process.

In the network-wide tracking adopts Bayesian updating approach on the framework of the corridor model. The network-wide tracking is implemented with multiple detection locations, therefore different travel time and traffic state between tracking locations could challenge tracking performance. Moreover, substantial amount of vehicles from multiple upstream locations should be handled in a matching process. Hence, additional data sources such as GPS vehicle trajectory, travel time, and body classification model estimates were used as supplementary data of ILD signatures. Four scenarios composed with different combinations of data sources were compared tracking performances and the best combination and fusion weight of data sources was identified.

In addition, a separate modeling for vehicle classification was investigated using vehicle duration from ILD data for the network-wide tracking. The tracking model would provide better matching accuracies when targets find their matched vehicles in the same vehicle category such as single units and multi units. However, since ILD has no capabilities in distinguishing vehicle types, a truck detection model was developed to classify trucks into single-units and multi-units and exclude passenger vehicles in the tracking process.

Both corridor level and network-wide tracking models were able to successfully match total population of trucks while maintaining high matching accuracy. Specifically, the corridor level tracking showed 81 percent of correct matching rate even though only 14 percent of trucks passed upstream and downstream locations. A network wide tracking showed 91 percent and 67

percent of correct matching rates for multi-unit trucks and single-units trucks, respectively. The best fusion weights differ by tracking level and data sources. In the corridor tracking, more weights on WIM than signature features result in better performance while more weights on signature than additional data sources showed higher matching rate in the network-wide tracking.

Tracked trucks from a corridor and network were used for various applications such as travel time estimation, WIM calibration, and truck monitoring with integration of truck body configuration model. These applications show potential use-cases of vehicle tracking outcomes. Travel time estimation and WIM calibration results showed overall good performance when sampled populations with higher matching probabilities were used rather than applying the whole matched population for the applications. Truck monitoring results provide valuable insights on truck route choice in a complex network. Since truck body configurations are closely related to service type and affiliated industry, temporal and spatial distribution of trucks obtained by the proposed vehicle tracking and monitoring frameworks give us detailed truck movement than what is observable at existing point detections.

7.2 Future study

While this dissertation proposes a new and comprehensive framework for truck path flow estimation leveraging existing detection systems, there are several improvements to be considered in a future work. In a network-wide tracking, due to sparse WIM deployment, signatures would be the main data sources to be utilized. Therefore, the quality of signatures from ILD system along with feature selection and treatment methods would highly affect tracking performance. Therefore, future work could investigate sensitivity in loop configuration

and geometry for robust tracking framework on varying ILD systems. In addition, the proposed feature selection from fifty magnitude attributes in one signature could be expanded with additional features such as magnitude differences especially for single-unit trucks. Second, the tracking model will be implemented at different locations for spatial transferability test. The tracking algorithm might be affected by complexity of network system and truck travel patterns of particular locations as trucks travels are highly related to service industry. Hence, a more complex network system that contains multiple downstream and upstream detection sites can be considered with manual data groudtruthing for further model validation. Third, parametric density estimations for travel time could consider varying traffic states between detection stations. Due to the data availability, this work considers stable traffic state. However, in field implementation, unexpected congestion or crash could significantly increase travel time in tracking corridors, thus dynamic density estimation could be considered. For example, as shown in truck detection GM model, travel time estimation could be updated in shorter time interval.

Besides the methodological improvement works, several practical consideration are presented as a future study. The outcomes of proposed individual vehicles tracking are expected to assist many government and state agencies through various applications. For one, primary truck routes with detailed truck types could be identified and managed. In addition, performance measures such as truck travel time and average delays along the truck corridor could provide valuable inputs for highway monitoring and operations. Moreover, this dissertation can generate high value impacts for Greenhouse Gas (GHG) estimations and in guiding policies towards the reduction of GHGs. Specifically, emissions estimations could be improved through the availability of a finer spatial resolution of heavy-heavy duty trucks (HHDT) spatial distribution and their route choice. Further, the dissertation work helps improve and support current freight

modeling efforts for regional and statewide agencies. Statistics relating to truck movement in various types of trucks such as freight and non-freight, special purpose trucks (i.e., logging, agricultural, and livestock), can be provided as inputs to support freight modeling efforts.

Therefore, through the tracking framework, agencies in California such as Caltrans, CARB, and CEC can gain further insights in understanding truck movements to guide freight planning, sustainable goods movement, and infrastructure investment while mitigating its negative impacts such as greenhouse gas, traffic congestion, safety concerns, and infrastructure wear. Potential immediate and future benefits are categorized into seven topics and further discussed with related previous studies as follows.

1) Truck activity monitoring: The proposed network truck tracking study is expected to provide a better understanding of truck travel characteristics and travel behavior through truck route choice information by body configuration. Temporal and spatial variations on truck travel patterns and weight distribution will be investigated by truck routes and by body configuration, which would be a great benefit on a decision making for sustainable goods movement. An investigation will also involve the enhancement of truck weight distribution model as one of the main applications of the tracking model. The spatial relationships between WIM and ILD can be updated by tracking results. Weight distributions from WIM can be spatially interpolated to ILD sites by the estimate spatial relationships, which provide much finer resolutions of weight distributions over a large network. The weight distributions will be further disaggregated by body types or axle configurations to identify key corridors for weight enforcement.

2) Freight model validation: If vehicles can be tracked along the consecutive points with high matching accuracy, point-to-point OD between the tracking stations can be measured, instead of estimated. The OD information will be further detailed in truck class or weight category based on the truck detection algorithm or WIM records. In addition, truck weight distribution such as loaded and unloaded weight distribution by body configuration helps validating the recently developed California Statewide Freight Forecasting Model (CSFFM). For example, state-level truck characteristics such as average payload and loaded weights can be updated with the representative samples of weight estimations by proposed study. In addition, flow of short/long haul trucks and empty/fully loaded trucks will be able to assist validation of freight forecasting models.

3) Truck travel performance measure: Oh (2003) suggested new criteria for real-time level of service (LOS) using average delay of the tracked vehicles. Kwon (2010) also utilized WIM data to estimate truck travel times using matching results. As shown in this dissertation, truck travel time can be directly obtained by matched vehicles across multiple detector stations. Estimated travel time can be further analyzed across temporal scales such as time-of-day, day-of-week, and seasons, which can assist highway operation, capacity, and level of service.

4) Sensor calibration: The proposed algorithm directly assists to the WIM examination and calibration. Kwon (2010) showed that weight can be used as automatic mis-calibration identifier by comparing vehicle weights between two stations from the same vehicle. Weight comparison gives us a clue of weight bias at one station, and subtracting (or adding) biased weight to that station automatically shift weight, which correct for the bias at 'bad' station. However, tracking should be implemented across multiple WIM sites to use a tracked truck as an

investigator for WIM accuracy to find 'bad' stations. The future study expands the tracking framework to multiple WIM sites to develop a WIM calibration model that capture and adjust errors in WIM measurements.

5) Optimal sensor location identification: Truck path flow data can assist determining sensor location such as advanced inductive loop and WIM. Considering varying travel patterns by truck type and service type, sensor location should be decided with a consideration of prior interest of flows to capture. In this regards, this dissertation showed that different sensor location should be selected for route identifiability and flow capturing. The future study will further investigate vehicle tracking in a more complex and longer highway stretches and compare various sensor location strategies. Since the proposed modeling requires an advanced inductive signature card replaced by a conventional card to obtain waveform signatures, utilizing both conventional bivalent and advanced waveform ILD systems could provide more economical strategy for path flow estimation. Hence, sensor optimal location model utilizing both conventional and advanced ILD can be investigated considering a prior path flow measurement.

6) Emissions estimations: Liu (2011) developed a system to estimate and monitor on-road emissions in real time using tracking algorithm. Liu (2011) aggregated vehicle types for emissions modeling and applied multi-layer perceptron neural network for vehicle re-identification on the aggregated vehicle types. Since vehicle mix and vehicle activity directly affect emission rates, average truck speed and truck class from tracking algorithm will obtain more accurate emission rates. Typically, vehicle registration databases and truck surveys are used to supply input data for the emissions models therefore estimated weight distribution at

finer spatial resolution along with VMT from the tracking model would facilitate improvements in emissions modeling.

7) Safety: Despite the large impact of trucks on road crashes, studies on truck safety impacts have not been well investigated due to the lack of available data. Oh (2003) proposed real-time safety index (RSI) based on the safety distance in car-following situation by vehicle tracking results. This dissertation could further investigate individual truck travel patterns using from tracking model combined with weight and body configuration information to facilitate the development of more comprehensive and reliable measures of truck exposure and crash factors.

8. References

Abdulhai, Baher, and Seyed M. Tabib. "Spatio-temporal inductance-pattern recognition for vehicle re-identification." *Transportation Research Part C: Emerging Technologies* 11.3 (2003): 223-239.

Andre Tok, Kyung (Kate) Hyun, Sarah Hernandez, Kyungsoo Jeong, Yue Sun, Craig Rindt, and Stephen G. Ritchie. "Truck activity monitoring system (TAMS) for freight transportation analysis.", accepted for *Proceedings of Annual Meeting of the Transportation Research Board*, Washington, D.C. (2016)

Board of Equalization, "International Fuel Tax Agreement (IFTA)." www.boe.ca.gov/sptaxprog/mciftamain.htm.

Böhnke, P., and Elmar Pfannerstill. "System for the automatic surveillance of traffic situations." *ITE journal* 56.1 (1986): 41-45.

Bureau of Transportation Statistics, "National Transportation Statistics." US DOT, www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/index.html.

CA DMV, "International Registration Plan IRP." www.dmv.ca.gov/portal/dmv/?1dmy&urile=wcm:path:/dmv_content_en/dmv/vehindustry/irp/irpinfo.

Caltrans PeMS, "Caltrans Performance Measurement System (PeMS)." US DOT, pems.dot.ca.gov/.

Castillo, Enrique, et al. "Optimal use of plate-scanning resources for route flow estimation in traffic networks." *IEEE Transactions on Intelligent Transportation Systems* 11.2 (2010): 380-391.

Castillo, Enrique, José María Menéndez, and Pilar Jiménez. "Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations." *Transportation Research Part B: Methodological* 42.5 (2008): 455-481.

Cerrone, C., R. Cerulli, and M. Gentili. "Vehicle-ID sensor location for route flow recognition: Models and algorithms." *European Journal of Operational Research* 247.2 (2015): 618-629.

Cetin, Mecit, Christopher M. Monsere, and Andrew P. Nichols. "Bayesian models for reidentification of trucks over long distances on the basis of axle measurement data." *Journal of Intelligent Transportation Systems* 15.1 (2011): 1-12.

Charnes, Abraham, and William Wager Cooper. "Goal programming and multiple objective optimizations: Part 1." *European Journal of Operational Research* 1.1 (1977): 39-54.

Chen, Anthony, Piya Chootinan, and Surachet Pravinvongvuth. "Multiobjective model for locating automatic vehicle identification readers." *Transportation Research Record: Journal of the Transportation Research Board* 1886 (2004): 49-58.

Chen, Anthony, Surachet Pravinvongvuth, and Piya Chootinan. "Scenario-based multi-objective AVI reader location models under different travel demand patterns." *Transportmetrica* 6.1 (2010): 53-78.

Cheung, Sing, et al. "Traffic measurement and vehicle classification with single magnetic sensor." *Transportation research record: journal of the transportation research board* 1917 (2005): 173-181.

Coifman, Benjamin, and Sivaraman Krishnamurthy. "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure." *Transportation Research Part C: Emerging Technologies* 15.3 (2007): 135-153.

Dion, Francois, and Hesham Rakha. "Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates." *Transportation Research Part B: Methodological* 40.9 (2006): 745-766.

Federal Highway Administration Office of Policy and Governmental Affairs, "A Summary of Highway Provisions." *Moving Ahead for Progress in the 21st Century Act (MAP-21),* 17 July 2012, www.fhwa.dot.gov/map21/summaryinfo.cfm.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1. Springer, Berlin: Springer series in statistics, 2001.

Gelman, Andrew, et al. *Bayesian data analysis*. Vol. 2. Boca Raton, FL, USA: Chapman & Hall/CRC, 2014.

Gentili, Monica, and Pitu B. Mirchandani. "Locating active sensors on traffic networks." *Annals of Operations Research* 136.1 (2005): 229-257.

Golob, Thomas F., Wilfred W. Recker, and Veronica M. Alvarez. "Freeway safety as a function of traffic flow." *Accident Analysis & Prevention* 36.6 (2004): 933-946.

Hallenbeck, Mark E., and Soon-Gwam Kim. *Final Technical Report for Task A: Truck Loads and Flows*. No. WA-RD 320.3. 1993.

Hernandez, Sarah, Andre Tok, and Stephen G. Ritchie. "Multiple-Classifier Systems for Truck Body Classification at WIM Sites with Inductive Signature Data." *Transportation Research Board 94th Annual Meeting*. No. 15-1377. 2015.

Hodgson, M. John. "A Flow- Capturing Location- Allocation Model." *Geographical Analysis* 22.3 (1990): 270-279.

Hwang, C-L., and Abu Syed Md Masud. *Multiple objective decision making—methods and applications: a state-of-the-art survey*. Vol. 164. Springer Science & Business Media, 2012.

Iglewicz, and Hoaglin. "Detection of Outliers." US DOC, www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm.

Jeng, Shin-Ting, and Lianyu Chu. "Tracking heavy vehicles based on weigh-in-motion and inductive loop signature technologies." *IEEE Transactions on Intelligent Transportation Systems* 16.2 (2015): 632-641.

Jeng, Shin-Ting, Yeow Chern Andre Tok, and Stephen G. Ritchie. "Freeway corridor performance measurement based on vehicle reidentification." *IEEE Transactions on Intelligent Transportation Systems* 11.3 (2010): 639-646.

Kim, P. J., and Robert I. Jennrich. "Tables of the Exact Sampling Distribution of the Two-sample Kolmogorov-Smirnov Criterion Dm; n; m< n." *Selected tables in mathematical statistics* 1 (1973): 80-129.

Kuhne, Reinhart D. *Freeway control using a dynamic traffic flow model and vehicle reidentification techniques*. No. 1320. 1991.

Kwon, Jaimyoung, and Karl Petty. "Vehicle re-identification using weigh-in-motion data for truck travel time measurement and sensor calibration." *17th ITS World Congress*. 2010.

Kwon, Jaimyoung, Pravin Varaiya, and Alexander Skabardonis. "Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation." *Transportation Research Record: Journal of the Transportation Research Board* 1856 (2003): 106-117.

Kwon, Jaimyoung, Pravin Varaiya, and Alexander Skabardonis. "Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation." *Transportation Research Record: Journal of the Transportation Research Board* 1856 (2003): 106-117.

Kwong, Karric, et al. "A practical scheme for arterial travel time estimation based on vehicle re-identification using wireless sensors." *University of California, Berkeley* 94720 (1700).

Lee, C. E. "Standards for highway Weigh-in-motion (WIM) systems." *ASTM Standardization News* 19.2 (1991).

Liu, Hang, Yeow Chern Andre Tok, and Stephen G. Ritchie. "Development of a real-time on-road emissions estimation and monitoring system." *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2011.

Lu, Q., et al. "Truck traffic analysis using weigh-in-motion (WIM) data in California." *University of California, Berkeley, Institute of Transportation Studies, Pavement Research Center, Berkeley, California, USA* (2002).

Lutsey, Nicholas P. "Assessment of Out-of-State Heavy-Duty Truck Activity Trends In California." *Institute of Transportation Studies* (2008).

Martin Maechler, and Peter Bühlmann. "Chapter 2, Nonparametric Density Estimation." stat.ethz.ch/education/semesters/ss2011/compstat/.

Mínguez, R., et al. "Optimal traffic plate scanning location for OD trip matrix and route estimation in road networks." *Transportation Research Part B: Methodological* 44.2 (2010): 282-298.

Mirchandani, Pitu B., M. Gentili, and Yang He. "Location of vehicle identification sensors to monitor travel-time performance." *IET Intelligent Transport Systems* 3.3 (2009): 289-303.

Nichols, Andrew P., and Darcy M. Bullock. "Quality control procedures for weigh-in-motion data." (2004).

Ogden, K. W. "Truck movement and access in urban areas." *Journal of Transportation Engineering* 117.1 (1991): 71-90.

Oh, Cheol, and Stephen Ritchie. "Anonymous vehicle tracking for real-time traffic surveillance and performance on signalized arterials." *Transportation Research Record: Journal of the Transportation Research Board* 1826 (2003): 37-44.

Oh, Jun-Seok, et al. "Real-time estimation of accident likelihood for safety enhancement." *Journal of transportation engineering* 131.5 (2005): 358-363.

Optimization, Gurobi. "Gurobi optimizer reference manual." *URL: http://www. gurobi. com* 2 (2012): 1-3.

Papagiannakis, A. Thomas, and Rich Quinley. *High speed weigh-in-motion system calibration practices*. Vol. 386. Transportation Research Board, 2008.

PierPass program, http://www.pierpass.org/monthly-newsletters/pierpass-july-2016-newsletter/ accessed by July 2016

Prozzi, Jorge A., and Feng Hong. "Effect of weigh-in-motion system measurement errors on load-pavement impact estimation." *Journal of Transportation Engineering* 133.1 (2007): 1-10.

Ratanamahatana, C.A. and Gunopulos, D., Scaling up the naive Bayesian classifier: Using decision trees for feature selection (2002.).

Rodrigue, Jean-Paul et al. *The Geography of Transport Systems*. people.hofstra.edu/geotrans/about.html.

Roobaert, Danny, Grigoris Karakoulas, and Nitesh V. Chawla. "Information gain, correlation and support vector machines." *Feature Extraction*. Springer Berlin Heidelberg, 2006. 463-470.

Roorda, Matthew. "Data collection strategies for benchmarking urban goods movement across Canada." *Transportation Letters* 3.3 (2011): 175-199.

Srour, F., and Diane Newton. "Freight-specific data derived from intelligent transportation systems: Potential uses in planning freight improvement projects." *Transportation Research Record: Journal of the Transportation Research Board* 1957 (2006): 66-74.

Sumalee, Agachai, et al. "Probabilistic fusion of vehicle features for reidentification and travel time estimation using video image data." *Transportation Research Record: Journal of the Transportation Research Board* 2308 (2012): 73-82.

Sun, Carlos C., et al. "Vehicle reidentification using multidetector fusion." *IEEE Transactions on Intelligent Transportation Systems* 5.3 (2004): 155-164.

Sun, Carlos, et al. "Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways." *Transportation Research Part C: Emerging Technologies* 7.4 (1999): 167-185.

Tamiz, Mehrdad, Dylan Jones, and Carlos Romero. "Goal programming for decision making: An overview of the current state-of-the-art." *European Journal of operational research* 111.3 (1998): 569-581.

Tawfik, Ahmed, et al. "Using decision trees to improve the accuracy of vehicle signature reidentification." *Transportation Research Record: Journal of the Transportation Research Board* 1886 (2004): 24-33.

Teodorovic, Dusan, et al. "Genetic algorithms approach to the problem of the automated vehicle identification equipment locations." *Journal of Advanced Transportation* 36.1 (2002): 1-21.

Tian, Yin, et al. "A vehicle re-identification algorithm based on multi-sensor correlation." *Journal of Zhejiang University SCIENCE C* 15.5 (2014): 372-382.

U.S. Census Bureau, "Vehicle Inventory and Use Survey." 2004, www.census.gov/svsd/+www/vius/products.html.

UC Irvine, "Truck Activity Monitoring System." *Truck Activity Monitoring System*, freight.its.uci.edu/tams/.

US DOT, "Travel Monitoring and Traffic Volume." *Traffic Monitoring Guide*, 2013, www.fhwa.dot.gov/policyinformation/tmguide/.

USDOT, "National Freight Strategic Plan." *Department of Transportation*, US DOT, www.transportation.gov/freight/nfsp.

Viti, Francesco, et al. "Assessing partial observability in network sensor location problems." *Transportation research part B: methodological* 70 (2014): 65-89.

Washington, Simon P., Matthew G. Karlaftis, and Fred Mannering. *Statistical and econometric methods for transportation data analysis*. CRC press, 2010.

West, Mike. "Mixture models, Monte Carlo, Bayesian updating, and dynamic models." *Computing Science and Statistics* (1993): 325-325.

Wu, Shengli, Yaxin Bi, and Xiaoqin Zeng. "The linear combination data fusion method in information retrieval." *International Conference on Database and Expert Systems Applications*. Springer Berlin Heidelberg, 2011.

Young, Stanley Ernest, et al. "Detection Probability Models for Bluetooth Re-identification Technology."

Zheng, Zuduo, Soyoung Ahn, and Christopher M. Monsere. "Impact of traffic oscillations on freeway crash occurrences." *Accident Analysis & Prevention*42.2 (2010): 626-636.