OTREC

OREGON
TRANSPORTATION
RESEARCH AND
EDUCATION CONSORTIUM

FINAL REPORT

Exploratory Methods for Truck Re-Identification in a Statewide Network Based on Axle Weight and Axle Spacing Data to Enhance Freight Metrics: Phase II

OTREC-RR-12-04
May 2012

A National University Transportation Center sponsored by the U.S. Department of Transportation's Research and Innovative Technology Administration

EXPLORATORY METHODS FOR TRUCK RE-IDENTIFICATION IN A STATEWIDE NETWORK BASED ON AXLE-WEIGHT AND AXLE-SPACING DATA TO ENHANCE FREIGHT METRICS: PHASE II

Research Report

OTREC-RR-12-04

by

Mecit Cetin
Old Dominion University, Norfolk, VA

Christopher M. Monsere
Portland State University, Portland, OR

for
Oregon Transportation Research
and Education Consortium (OTREC)
P.O. Box 751
Portland, OR 97207

OTREC
OREGON TRANSPORTATION RESEARCH
AND EDUCATION CONSORTIUM

May 2012

Technical Report Documentation Page

1. Report No. OTREC-RR-12-04	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Exploratory Methods for Truck Re-Identification in a Statewide Network Based on Axle Weight and Axle Spacing Data to Enhance Freight Metrics: Phase II		5. Report Date May 2012 6. Performing Organization Code
7. Author(s) Mecit Cetin and Christopher M. Monsere		8. Performing Organization Report No.
9. Performing Organization Name and Address Portland State University PO Box 751 Portland, OR 97207 Old Dominion University 5115 Hampton Boulevard Norfolk, VA 23529		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 2010-327
12. Sponsoring Agency Name and Address Oregon Transportation Research and Education Consortium (OTREC) P.O. Box 751 Portland, Oregon 97207		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Vehicle re-identification methods can be used to anonymously match vehicles crossing two different locations based on vehicle attribute data. This research builds upon a previous study and investigates different methods for solving the re-identification problem and explores some of the factors that impact the accuracy of the results. To support this work, archived data from weigh-in-motion (WIM) stations in Oregon are used for developing, calibrating, and testing vehicle re-identification algorithms. In addition to the Bayesian approach developed by the researchers in the previous study, a neural network model is developed for solving the re-identification problem. The results from the testing datasets showed that both methods can be effective in solving the re-identification problem while the Bayesian method yields more accurate results. A comprehensive analysis is performed to investigate the key factors impacting the accuracy of the results. The analyses are performed by employing the Bayesian algorithm to match commercial vehicles that cross upstream and downstream pairs of WIM sites that are separated by long distances ranging from 70 to 214 miles. Data from 14 different pairs of WIM sites are used to evaluate how matching accuracy is impacted by various factors such as the distance between two sites, travel time variability, truck volumes, and sensor accuracy or consistency of measurements. After running the vehicle re-identification algorithm for each one of these 14 pairs of sites, the matching error rates are reported. The results from the testing datasets showed a large variation in terms of accuracy. It is found that sensor accuracy and volumes have the greatest impacts on matching accuracy whereas the distance alone does not have a significant impact. Overall, for estimating travel times and origin-destination flows between two WIM sites, the methods developed in this project can be used to effectively match commercial vehicles crossing two data collection sites that are separated by long distances.

17. Key Words vehicle re-identification, weigh-in-motion, freight performance measures	18. Distribution Statement No restrictions. Copies available from OTREC: www.otrec.us

19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 48	22. Price

ACKNOWLEDGEMENTS

The authors acknowledge the Oregon Transportation Research and Education Consortium (OTREC) for funding this research and the Oregon Department of Transportation (ODOT) Motor Carrier Division for providing the data. At ODOT, Dave Fifer and David McKane were particularly helpful. Christopher Higgins at Oregon State University contributed to the WIM archive by providing software tools and expertise. At Portland State University, Kristin Tufte and Michael Wolfe were instrumental in data management through the PORTAL umbrella. The National Science Foundation supported early development of PORTAL. The authors also acknowledge the help of graduate students Tanweer Rashid and Ilyas Ustun from Old Dominion University (ODU), who contributed to this work substantially by conducting data analyses.

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Most transportation agencies rely on point detectors (e.g., inductive loops, axle detectors) located at specific locations on highways to collect data on traffic volumes, vehicle classes and other relevant attributes of traffic. By utilizing these data collected from these point detectors, researchers have developed vehicle re-identification algorithms to match measurements at two sites that belong to the same vehicle. This enables tracking the movement of individual vehicles between different data collections sites, which in turn provides valuable information for the estimation of travel times, travel delays and origin-destination flows.

The main goal of this OTREC project is to investigate the factors that impact the accuracy of the re-identification algorithms and to implement such algorithms on larger datasets. By building upon a previous OTREC-funded study conducted by the authors, this research further contributes to the re-identification literature in two significant ways. First, the feasibility of using neural networks in solving the re-identification problem is investigated. Second, an extensive analysis is performed to understand the key factors affecting the matching accuracy of the re-identification algorithms.

Data from weigh-in-motion (WIM) stations provide a basis for the development and testing of these algorithms. The data supporting this research come from the WIM sites in Oregon, which are equipped with sensors that can measure axle weights, axle spacing, and gross vehicle weight estimates that are uniquely matched to each truck. Since some of the trucks (20-35%) are carrying radio-frequency identification (RFID) transponders, these measured attributes are also uniquely matched to transponder-equipped trucks. These particular trucks provide the needed data for model development, calibration and testing.

The neural network models are trained and tested on datasets from four different upstream and downstream pairs of WIM sites. While developing the neural network models, special attention is paid to ensure that the neural network design (e.g., number of neurons in the hidden layer) is optimal. The performance of neural network models is then compared to the Bayesian models developed by the authors in a previous study. Overall, with the exception of one case, the Bayesian models are found to outperform the neural networks based on the datasets considered in this study. The overall results show that both methods can be effective in solving the re-identification problem, while the Bayesian method yields more accurate results.

A comprehensive analysis is then performed to investigate the key factors impacting the accuracy of the results. The analyses are performed by employing the Bayesian algorithm to match trucks that cross upstream and downstream pairs of WIM sites that are separated by long distances ranging from 70 to 214 miles. Data from 14 different pairs of WIM sites are used to evaluate how matching accuracy is impacted by various factors such as the distance between two sites, travel-time variability, truck volumes, and sensor accuracy or consistency of measurements. After running the vehicle re-identification algorithm for each one of these 14 pairs of sites, the matching error rates are reported. The results from the testing datasets showed a large variation in terms of accuracy. It is found that sensor accuracy and volumes have the greatest impacts on matching accuracy, whereas the distance alone does not have a significant impact. Overall, for estimating travel times and origin-destination flows between two WIM sites, the methods developed in this project can be used to effectively match commercial vehicles crossing two data collection sites that are separated by long distances.

1.0 INTRODUCTION

Most transportation agencies rely on point detectors (e.g., inductive loops, axle detectors) located at specific locations on highways to collect data on traffic volumes, vehicle classes and other relevant attributes of traffic. By utilizing these data collected from these point detectors, researchers have developed vehicle re-identification algorithms to match measurements at two sites that belong to the same vehicle. This enables tracking the movement of individual vehicles between different data collections sites, which in turn provides valuable information for the estimation of travel times, travel delays and origin-destination flows.

Even though there are other technologies that can be utilized to track the movements of vehicles over transportation networks, most of these technologies (e.g., automatic vehicle identification (AVI) tags, license plate recognition) require installation of additional in-car and/or roadside devices and may have related privacy concerns. However, vehicle re-identification methods that are based on the vehicle attribute data collected by sensors already installed on roadways enable tracking vehicles anonymously and do not require substantial additional investment.

This OTREC project improves upon a previous study conducted by the authors on the same research theme. The main goal of the project is to investigate the factors that impact the accuracy of the re-identification algorithms and test such algorithms on additional datasets. Data from weigh-in-motion (WIM) stations provide a basis for the development and testing of these algorithms. The data supporting this research come from the WIM sites in Oregon, which are equipped with sensors that can measure axle weights, axle spacing, and gross vehicle weight estimates that are uniquely matched to each truck. Since some of the trucks (20-35%) are carrying radio frequency identification (RFID) transponders, these measured attributes are also uniquely matched to transponder-equipped trucks. These particular trucks provide the needed data for model development, calibration and testing.

This report describes the models used for re-identification, including a new algorithm developed based on neural networks. It also describes the analyses carried out to evaluate the impacts of different factors (e.g., distance between two sites, travel-time variability, truck volumes, and sensor accuracy or consistency of measurements) on the accuracy of the results produced by the re-identification algorithms.

Overall, the methods developed in this research can be used to support programs and applications for monitoring freight over the highways. One of the key aspects of monitoring freight has to do with determining the flow patterns (and travel times) of trucks, which can be achieved by uniquely identifying trucks at specific points along the roads or by tracking individual trucks using technology such as GPS. The re-identification method, in some circumstances, can be more advantageous as compared to other available options to track and re-identify trucks (e.g., GPS, automatic vehicle identification (AVI), license plate recognition) because of several reasons:

- Data from AVI transponders, such as the Oregon Green Light program, or from other types of electronic tracking systems might not be readily available to the public agencies involved in motor freight planning (e.g., MPOs, DOTs) due to privacy, jurisdictional, and institutional issues;

- Not all trucks are equipped with AVI transponders. However, with the re-identification methods all trucks can be potentially tracked since they all cross the WIM stations; and

- The proposed approach does not require installation of any new sensors since the input data are already collected at existing WIM and automatic vehicle classification (AVC) stations, whereas alternative technologies like license plate recognition requires additional investment.

1.1 OBJECTIVES

By building upon past and ongoing research by the principal investigators (PIs) and others in the areas of WIM data analysis, travel-time estimation for commercial trucks and vehicle re-identification methods, this research aims to contribute to the state-of-art and state-of-practice in freight movement by developing and testing novel vehicle re-identification methods to improve the ability to estimate truck movements in a transportation network. These methods capitalize on vehicle-attribute data, such as axle spacing and axle weights, which are already collected by numerous sensors installed on roadways.

The specific objectives of this project are:

- To investigate the factors that impact the accuracy of the re-identification algorithms that are developed for re-identifying commercial trucks based on vehicle-attribute data automatically collected by sensors installed at traffic data collection stations;

- To investigate alternative re-identification methods (e.g., neural networks) to solve the re-identification problem; and

- To implement such algorithms on larger datasets.

1.2 ORGANIZATION OF THE REPORT

This report is organized as follows: Chapter 2 provides an overview of some relevant studies on vehicle re-identification methods and applications. Chapter 3 describes the WIM data utilized for model development and testing in this project. Chapter 4 describes the problem of re-identification in detail and presents the Bayesian algorithm developed in the previous study conducted by the authors (Monsere, Cetin and Nichols, 2011). Chapter 5 describes the neural network model developed to solve the re-identification problem. Chapter 6 presents the results of the analyses carried out to evaluate the factors impacting the accuracy of the re-identification algorithms. Conclusions of the study are given in Chapter 7.

2.0 LITERATURE REVIEW

As explained in *A Concept for a National Freight Data Program: Special Report 276*, data on goods movements are needed to identify and evaluate options for mitigating congestion; improve regional and global economic competitiveness; inform investment and policy decisions about modal optimization; enhance transportation safety and security; identify transportation marketing opportunities; and reduce fuel consumption and improve air quality (TRB, 2003). This project contributes to a better understanding of freight movement by developing re-identification algorithms to estimate truck O-D (origin-destination) flows and travel times. Even though determining truck counts at particular locations on a transportation network is relatively easy to do, obtaining O-D data is, in general, more difficult since it requires uniquely re-identifying trucks at multiple points.

Since the mid-1990s, many research efforts have focused on methods to anonymously track vehicular movements by re-identifying individual vehicles at multiple locations utilizing existing sensors. The predominant objective has been to estimate travel times in order to characterize link performance. For this reason, the re-identification has focused primarily on passenger cars and light trucks, which typically make up the majority of traffic in urban areas where the link performance varies the most. Various techniques and technologies have been employed for the re-identification of vehicles including video/imaging (Shuldiner and Upchurch, 2001), and automatic vehicle identification (AVI) (Hellinga, 2001; Dion and Rakha, 2006). A more detailed explanation of these technologies and the associated techniques can be found in the *Travel Time Data Collection Handbook* (Turner, Eisele, Benz et al., 1998).

There have been several studies on re-identifying individual vehicles anonymously at multiple locations by utilizing data from existing inductive dual loop detectors (Sun, Ritchie, Tsai et al., 1999; Coifman and Cassidy, 2002; Coifman, 2003). While most of the previous studies are based on data from dual loops, some researchers also extended the application of the re-identification algorithms to data from single loops (Coifman and Krishnamurthy, 2007). Other than the traditional inductive loops that are embedded in the pavement, researchers have investigated new types of inductive loops, the so-called "blade sensors," to get more detailed characteristics of vehicles. These sensors are more sensitive than the typical inductive loops and are capable of capturing wheel locations (Oh, Ritchie and Jeng, 2007). In general, magnetic vehicle signatures from loops provide the raw data which is used to extract useful vehicle features or attributes to differentiate between different vehicles. The predominant application of vehicle re-identification has been to estimate travel times (Liu, Oh and Recker, 2002; Sun, Arr and Ramachandran, 2003; Oh, Tok and Ritchie, 2005). Application of new technologies and algorithms, such as those provided by Sensys Networks, are expanding the use of these re-identification approaches.

Less attention has been given to the techniques to re-identify commercial vehicles at multiple locations, even though such techniques can support numerous applications including estimating travel times for trucks, quantifying travel-time reliability, estimating truck flow patterns (i.e., origins-destinations), estimating empty-truck movements, trip-length estimation, pavement management, WIM-sensor accuracy, and weigh-station enforcement.

Recently, the authors of this report explored the use of axle-spacing and axle-weight data to re-identify commercial trucks at two WIM stations (Cetin and Nichols, 2009; Cetin, Nichols and Monsere, 2011). In the 2009 study by Cetin and Nichols, commercial trucks at two WIM stations in Indiana were matched based on re-identification techniques. They developed matching algorithms based on statistical mixture models and tested the performance of the algorithms on the data from these two WIM stations that are separated by one mile. The results showed that trucks were matched with 99% accuracy when both axle spacing and weights were used, and with 97% accuracy when only axle spacing was used. However, the WIM stations in this earlier study were only separated by one mile, and all trucks in the sample crossed both the upstream and downstream stations (Cetin and Nichols, 2009).

On the other hand, the datasets used in the 2011 study included WIM stations in Oregon that are separated by greater distances (more than 100 miles). This distance introduces additional complexities since travel times can vary significantly, and trucks can leave and enter the road in between the two stations (this was not the case in the Indiana dataset). Since not all trucks cross both stations, screening techniques were developed to identify vehicles that cross only one of the stations. It was observed that the algorithms can match trucks with approximately 90% accuracy, while the total number of trucks being matched at this accuracy level is about 95% of the actual common trucks that cross both upstream and downstream sites. These methods allow the user to trade off the accuracy vs. total vehicles being matched by adjusting a threshold parameter.

Other than the work by Cetin and Nichols in 2009 and Cetin, Nichols and Monsere in 2011, the only known previous application of WIM data for vehicle re-identification was conducted by the Norway Public Roads Administration for determining link travel times on the Oslo Toll Ring (Christiansen and Hauer, 1996). A prototype of the system was tested at the Winter Olympic Games in Lillehammer in 1994 and later refined with more advanced matching algorithms.

In general, vehicle re-identification methods rely on the variability within the vehicle population and the ability to accurately identify the pairs of measurements collected at upstream and downstream stations that are generated by the same vehicle. These measurements can either be the actual physical attributes of vehicles, such as length (Coifman and Cassidy, 2002) and axle spacing (Cetin and Nichols, 2009; Cetin, Nichols and Monsere, 2011) or some characteristics of the sensor waveform or inductive vehicle signature (Sun, Ritchie, Tsai et al., 1999). Researchers have developed various methods, such as lexicographic optimization (Sun, Ritchie, Tsai et al., 1999; Oh, Ritchie and Jeng, 2007) and decision trees (Tawfik, Abdulhai, Peng et al., 2004) to re-identify vehicles. In a typical implementation of these methods, a downstream vehicle is matched to the most “similar” upstream vehicle (or vice versa) based on some defined metric (e.g., Euclidian distance). The resulting accuracy of these methods depends on several factors, including the variation of the attribute data from vehicle to vehicle, number of attributes, the distance between data collection stations, variability of travel time, and type of the re-identification algorithm used. Given a particular set of factors, this accuracy may or may not be satisfactory for a given application. The impacts of different factors on the accuracy of the re-identification algorithms are investigated in this report. A summary of the findings presented in this report can be found in (Cetin, Monsere, Nichols et al., 2011).

3.0 WEIGH-IN-MOTION DATA

In this chapter, the assembly, processing and storage of the weigh-in-motion (WIM) and automatic vehicle classification (AVC) data is described. Oregon's prescreening/preclearance program for commercial motor vehicles at fixed weigh and inspection stations is called *Green Light*. There are 22 equipped stations on the Oregon state highway system. These locations are shown in Figure 3-1 with a corresponding list of stations shown in Table 3-1. The station number in the table is an internal number as part of PSU's archiving process. At each of the stations, approaching trucks are directed into the appropriate lane on the mainline highway. At a location upstream from the static weigh station, transponder-equipped trucks are identified by the reader. Participation in the *Green Light* program is high; on average about 40% of observed vehicles are equipped with transponders (though this varies from station to station). In addition to the transponder record, the vehicles are weighed in motion (by load cells). The observation consists of axle weights as well as the spacing of the axles. These data also include speed, timestamp, the lane of observation (some stations are multilane), length (calculated), gross vehicle weight (calculated), and a count of the number of axles (calculated). As part of the proprietary control program by the equipment vendor (International Road Dynamics), a sieved-based classification algorithm uses the axle-spacing information to classify vehicles. A more detailed description of the Oregon WIM system is provided by Elkins and Higgins (2008).

The unique aspect of Oregon's system is that this transponder and weight-related data are available together in one record. These transponder-equipped vehicles provide a large pool of data to develop, validate and test the vehicle re-identification techniques described within.

Table 3-1: List of stations

Number	Code	Name	Route	Direction	MP
1	FWB	Farewell Bend POE	I-84	WB	353.31
2	EMH	Emigrant Hill	I-84	WB	226.95
3	WYT	Wyeth	I-84	WB	54.3
4	CSL	Cascade Locks POE	I-84	EB	44.93
5	LGR	LaGrande	I-84	EB	258.52
6	ODF	Olds Ferry	I-84	EB	354.38
7	ASP	Ashland POE	I-5	NB	18.08
8	BOR	Booth Ranch	I-5	NB	111.07
9	WDN	Woodburn, NB	I-5	NB	274.15
10	WDS	Woodburn, SB	I-5	SB	274.18
11	BRE	Brightwood, EB	US-26	EB	36.51
12	BRW	Brightwood, WB	US-26	WB	36.31
13	JBS	Juniper Butte	US-97	SB	108.2
14	LWL	Lowell	US-58	WB	17.17
15	WLB	Wilbur	I-5	SB	130
16	ASH	Ashland, SB	I-5	SB	18.08

17	KFP	Klamath Falls POE	US-97	NB	271.73
18	BND	Bend	US-97	NB	145.5
19	JBN	Juniper Butte	US-97	NB	106.9
20	KFS	Klamath Falls, SB	US-97	SB	271.41
21	UMT	Umatilla POE	I-82	EB	183.8
22	RPT	Rocky Point	US-30	WB	16.53

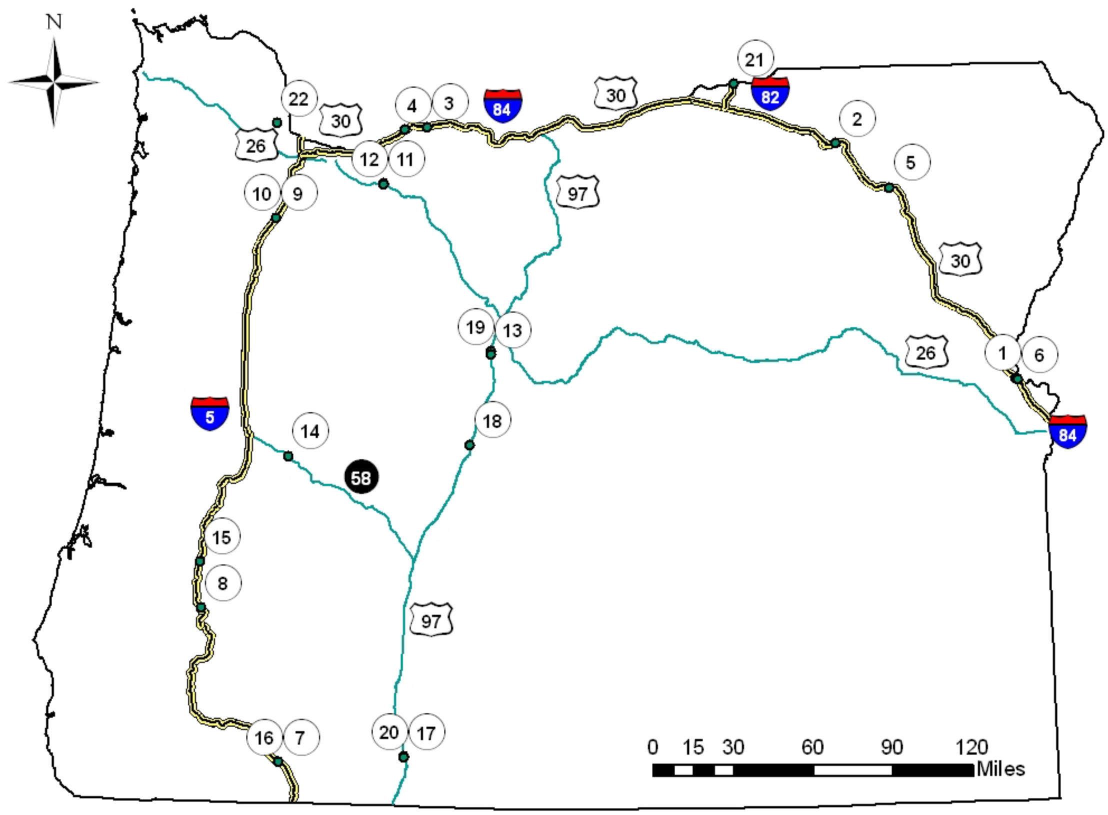

Figure 3-1 Map of WIM station locations in Oregon

3.1 DATA ARCHIVE

In support of this and other research, a WIM data archive was created (http://wim.its.pdx.edu/). This archive is housed under the Portland Transportation Archive Listing (PORTAL) umbrella at Portland State University's Intelligent Transportation Systems Lab. PORTAL is the official Archived Data User Service (ADUS) for the Portland metropolitan region as specified in the Regional ITS Architecture. PORTAL provides a centralized, electronic database that facilitates the collection, archiving and sharing of information/data for public agencies within the region. The creation of the PORTAL data archive was supported by a CAREER grant from the National Science Foundation (NSF). In addition, the Federal Highway Administration (through ODOT) has supported the purchase of hard disc storage, and the Portland metropolitan regional government (Metro) has invested in the ongoing support of the archive.

The archive stores data in a PostgreSQL-relational database management system (RDBMS). This archive implements a data warehousing strategy in that it retains large amounts of raw operational data for analysis and decision-making processes, and in that these data are stored independently of their operational sources, allowing the execution of time-consuming queries with no impact on critical operations uses. The database server is a Dell Server with two Quad Core Intel Xeon Processors running at 2.33 GHz with 8GB of memory. The database server runs Red Hat Linux. The RDBMS stores data physically on a 3.2 Terabyte redundant array of independent disks (RAID) providing both high-speed access and increased reliability through redundancy in the event of hardware failure. Offsite backups of the raw data are done once a week.

Monthly data are sent from ODOT via an FTP connection. These data are processed and then loaded in the WIM archive. A forthcoming OTREC report will describe the WIM data archive in detail (including data-quality efforts), but a short description follows. There are four primary tables in the WIM data. A schematic of the database is shown in Figure 3-2. The truck-level observations are loaded in a table called *wimdata*. A table *stations* includes the identifying information about each station. The table *stationmap* is a list of all possible routes (i.e., upstream-to-downstream station pairs), which defines the free-flow travel time, distance, and a parameter called upper time (currently time to travel between stations at 50 mph). Each possible link is given an identification that lists the upstream and downstream stations. An algorithm described in (Monsere, Wolfe, Alawakiel et al., 2009) produces a table *linktraveltime* of all trucks matched by transponder identification number between stations. The search algorithm matches a truck with a transponder at an upstream station with the same transponder at the downstream station. All matches within the time window of 0.75*free-flow time to 2*free-flow time are recorded. Free-flow time is defined as the time to traverse the route between stations at 55 mph (the posted speed limit for trucks on Oregon roadways).This table contains the upstream and downstream station numbers, tag number, timestamps of each observation, and whether the truck has been identified as a thru vehicle.

WIMDATA

timestamp	timestampwithtimezone
year	integer
month	integer
day	integer
hour	integer
minute	integer
seconds	integer
lane	integer
speed	integer
type	integer
length	integer
gvw	real
esal	real
sumlen	real
numaxles	integer
axl1	real
axl2	real
axl3	real
axl4	real
axl5	real
axl6	real
axl7	real
axl8	real
axl9	real
axl10	real
axl11	real
axl12	real
axl13	real
axl14	real
spc1	real
spc2	real
spc3	real
spc4	real
spc5	real
spc6	real
spc7	real
spc8	real
spc9	real
spc10	real
spc11	real
spc12	real
spc13	real
spc14	real
tag	text
stationnum	integer
gvw_zero	boolean
gvw_50	boolean
mph_10	boolean
mph_99	boolean
length_200	boolean
axle_sum_length	boolean
axle_sum_7	boolean
axle_first_5	boolean
num_axle_13	boolean
gvw_280	boolean
axle_spc_34	boolean
gvw_diff_7	boolean
truck_table	integer

STATIONS

stationnum	integer
station_code	character(3)
longname	text
name	text
route	character(5)
direction	character(2)
hwy_no	integer
roadbed	integer
mp	doubleprecision
lrs	charactervarying(15)
lat	doubleprecision
long	doubleprecision
filename	prefixtext

STATIONMAP

linkid	integer
up_station	integer
up_stationname	character(3)
dwn_station	integer
dwn_stationname	character(3)
freeflow	real
distance	real
uppertime	real

LINKTRAVELTIME

linkid	integer
up_station	integer
up_tag	text
up_timestamp	timestampwithtimezone
dwn_station	integer
dwn_tag	text
dwn_timestamp	timestampwithtimezone
thru_truck	boolean

Figure 3-2: Key table definitions for PSU PORTAL WIM Archive

4.0 THE BAYESIAN METHOD FOR RE-IDENTIFICATION

In general, the re-identification problem can be described as follows: Given two separate datasets that consist of vehicle attribute data (such as length, axle spacing, axle weights or some attributes of the magnetic signature), the re-identification algorithms attempt to match the pairs of measurements (one from each dataset) that belong to the same vehicle. These two datasets are collected at some upstream and downstream points in a transportation network. To simplify the discussion, an example is given in Figure 4-1which shows graphically two datasets for four vehicles that cross upstream and downstream stations. Each box represents a vehicle and the attribute data is indicated with horizontal bars. The actual matching is indicated with arrows in Figure 4-1. For both sites, vehicle number 3 only crosses one of the sites.

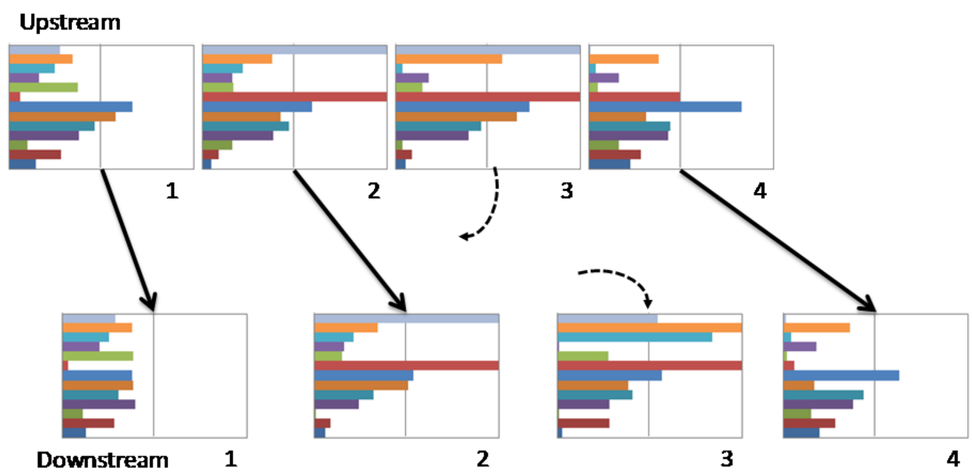

Figure 4-1: All vehicles are correctly matched while there is no match for one vehicle

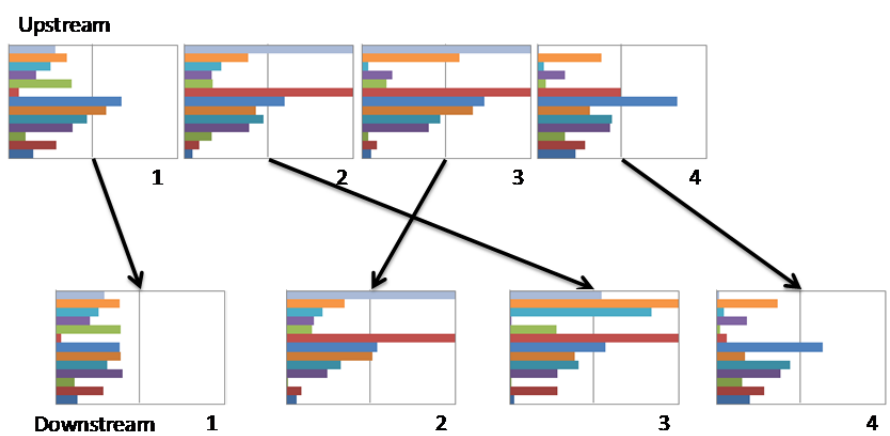

Figure 4-2: Vehicles 2 and 3 are mismatched

Vehicle re-identification algorithms attempt to match each vehicle in the downstream set to a vehicle in the upstream (or vice versa) based on some "similarity" measure that is a function of the attribute data between the two sites. These methods essentially capitalize on the variance in vehicle populations and the consistency or correlation of the measurements taken at the upstream and downstream stations. Figure 4-2 shows a potential outcome from a hypothetical algorithm for the same vehicles given in Figure 4-1. In this case, vehicles numbered 1 and 4 are matched accurately as the similarity measure is maximized for these pairs. On the other hand, vehicles 2 and 3 are mismatched. In reality, for downstream vehicle 3 there is no match at the upstream, but the matching algorithm identifies the upstream vehicle 2 as the best match among the four possibilities. Based on this simple illustration it can be observed that not only is a mechanism needed to identify the best match (in terms of the similarity in attribute data), but there also needs to be a method in place to screen out vehicles that cross one site but not the other.

The authors developed algorithms based on Bayesian statistics in Phase I of this project for solving the re-identification problem (Cetin, Nichols and Monsere, 2011; Monsere, Cetin and Nichols, 2011). The vehicle re-identification approach developed by the authors consists of two main stages. In the first stage, each vehicle from the downstream station is matched to the most "similar" upstream vehicle based on the posterior probabilities from a Bayesian model. For the second stage, several methods are developed to screen out vehicles that cross only one site. These methods increase the accuracy of matching, but may reduce the total number of vehicles matched. By setting a threshold value, these methods allow the user to trade off accuracy versus the total number of vehicles being matched. These methods involve calculating both the highest and the second-highest similarity measures for each vehicle being matched. Further details of these algorithmic steps can be found in Cetin, Nichols and Monsere (2011) and Monsere, Cetin and Nichols (2011).

The next two subsections provide a technical description of the problem and a brief overview of the Bayesian method for solving the re-identification problem.

4.1 NOTATION AND THE SEARCH SPACE

Let $\boldsymbol{U}$ and $\boldsymbol{D}$ be two non-empty sets that denote the vehicles crossing the upstream WIM station and downstream WIM station, respectively. Depending on various factors, including the station locations, WIM record validity (i.e., crossed sensors properly) and types of activity between the sensors, four general cases arise:

i) $U \subset D$ and $U \neq D$ (i.e., all vehicles crossing the upstream site also cross downstream site)

ii) $D \subset U$ and $U \neq D$ (i.e., all vehicles crossing the downstream site also cross upstream site)

iii) $U = D$ (i.e., all vehicles cross both sites)

iv) $D \not\subset U$, $U \not\subset D$, and $U \cap D \neq 0$ (i.e., not all vehicles in the upstream or downstream cross both sites)

Even though the fundamental re-identification problem is the same in all four cases, the search procedure in the third case is the simplest as all vehicles cross both sites. In this case, for any selected vehicle there is a match in the other set (i.e., there is a one-to-one mapping between the members of the two sets). One can apply not only statistical matching algorithms, but also assignment algorithms to assign all the members in one set to those in the other set while ensuring that each member is assigned only once. This is demonstrated in Cetin and Nichols (2009) and is shown to significantly improve the accuracy of matching vehicles.

The last case above (iv) is somewhat more difficult than the others since one needs to consider the possibility that a vehicle taken from one set might not have a match in the other set. In the first three cases, there is always a match for each vehicle in the smaller set (or in either set for case iii). The methods developed in this research can be used for any one of these four cases as the methods for screening can be applied to screen out vehicles that do not cross both sites. Without loss of generality, the methods (of the first stage) will be described for case ii where for each vehicle in D a match will be identified in U, which has more samples than set D. Then, in the second stage, the screening methods will be applied to the results of the first stage to determine which matched vehicles will be kept and which ones should be eliminated. In Chapter 6, the models are applied to datasets that fall into both case ii and case iv.

Let $\boldsymbol{X^U}$ and $\boldsymbol{X^D}$ be two matrices with the same number of columns that denote the data collected at an upstream station and a downstream station, respectively. $\boldsymbol{X^U}_i$ and $\boldsymbol{X^D}_j$ denote rows of these two matrices that correspond to the measurements (e.g., axle weights) taken for vehicle i at the upstream station and for vehicle j at the downstream station. Further, assume that the timestamps indicating arrival times of vehicles at each station are given and denoted by $\boldsymbol{t^U}_i$ for the upstream vehicles and $\boldsymbol{t^D}_j$ for the downstream vehicles. Given $\boldsymbol{X^U}$, $\boldsymbol{X^D}$, $\boldsymbol{t^U}_i$ and $\boldsymbol{t^D}_j$ the vehicle matching problem involves determining $\boldsymbol{X^U}_i$ and $\boldsymbol{X^D}_j$ that are generated by the same vehicle. Let δ_{ij} be a binary variable that equals 1 if $\boldsymbol{X^U}_i$ and $\boldsymbol{X^D}_j$ belong to the same vehicle and equal zero otherwise. The main objective of the matching algorithms is to estimate all δ_{ij}'s with minimum error.

As mentioned before, a two-stage approach is proposed in this research for the re-identification problem. In the first stage, for each vehicle in $\boldsymbol{D}$ a match is found in $\boldsymbol{U}$. This is accomplished by a Bayesian method as explained below. In the second stage, a new method is proposed to screen out mismatched vehicles to improve accuracy. These two stages are explained in detail in the subsequent sections.

For the first stage of re-identification, each vehicle in $\boldsymbol{D}$ needs to be matched to the most similar vehicle in $\boldsymbol{U}$. Since timestamp information is available for each vehicle, a reasonable "search space" from the upstream vehicle records ($\boldsymbol{U}$) can be identified based on travel times. Before the search starts to match a downstream vehicle j to an upstream vehicle i, a search space for vehicle j, denoted by S_j, is determined based on the timestamps at two stations ($\boldsymbol{t^U}_i$ and $\boldsymbol{t}^D{}_j$) and some defined time window. The variability in travel time can be captured by specifying minimum and maximum values for travel times. The minimum value (*minTime*) can be easily predicted based on an assumed maximum travel speed and the distance between the two stations. The maximum value can exhibit a large variation depending on the individual vehicle speeds, travel distance, and traffic flow interruptions between the two stations, and any pick-up, delivery or rest stops the driver may make. The maximum value (*maxTime*) can be taken as a multiple of the minimum time if no data exists or can be based on observations. The search space for a downstream vehicle j is then determined as follows:

$S_j = i \in \boldsymbol{U} \mid \boldsymbol{t}^{\boldsymbol{D}}{}_j - maxTime \leq \boldsymbol{t}^{U}{}_i \leq \boldsymbol{t}^{\boldsymbol{D}}{}_j - minTime \}$ (1)

Depending on the difference between *maxTime* and *minTime* or simply time window, the number of vehicles among which a match is to be found varies. Larger time windows will result in a larger number of vehicles in the search space, which can make the matching problem more difficult.

4.2 THE BAYESIAN METHOD

The Bayesian re-identification method relies on calculating the posterior probability of a match between two vehicles given two sets of data points collected for a vehicle pair (*i,j*) at the upstream and downstream stations. A vehicle *j* at the downstream station is matched to the upstream vehicle *i* that yields the largest probability of a match. The steps of the Bayesian method are formally explained below.

For each vehicle *j* in ***D***
  Identify a search space (see equation 1), $S_j \subset \boldsymbol{U}$
    For each $i \in S_j$
      Calculate $P(\delta_{ij} = 1|\text{data})$
    $m = \underset{i}{\text{argmax}}\ P(\delta_{ij} = 1|\text{data})$
  Match vehicle *j* to *m*, i.e., $\delta_{ij} = 1$ if $i = m$

Once a search space is identified, $P(\delta_{ij} = 1 \mid \boldsymbol{x}_{ij})$, the conditional probability that $\boldsymbol{X}^{\boldsymbol{U}}{}_i$ and $\boldsymbol{X}^{D}{}_j$ belong to the same vehicle given data (i.e., $\boldsymbol{x}_{ij} = \boldsymbol{x}^{U}{}_i \cup \boldsymbol{x}^{D}{}_j$), can be computed by the Bayes' theorem as follows:

$$P(\delta_{ij} = 1|x_{ij}) = \frac{f(x_{ij}|\delta_{ij}=1)P(\delta_{ij}=1)}{f(x_{ij}|\delta_{ij}=1)P(\delta_{ij}=1)+f(x_{ij}|\delta_{ij}=0)P(\delta_{ij}=0)} \quad (2)$$

In order to calculate this posterior probability, both the two conditional probability density functions (i.e., $f(x_{ij}|\delta_{ij}=1)$ and $f(x_{ij}|\delta_{ij}=0)$) and the prior probabilities (i.e., $P(\delta_{ij}=0)$ and $P(\delta_{ij}=1)$) are needed. The functions $f(x_{ij}|\delta_{ij}=1)$ and $f(x_{ij}|\delta_{ij}=0)$ are the density functions that characterize the collected data at two stations when it belongs to the same vehicle and different vehicles, respectively. As demonstrated in Cetin, Nichols and Monsere (2011) and Monsere, Cetin and Nichols (2011), when vehicles match (i.e., upstream and downstream measurements belong to the same vehicle) there is high correlation between the measurements, which is critical for re-identification to work effectively. On the other hand, when random data for upstream and downstream measurements are plotted the correlation disappears as expected and a roughly uniform distribution of points is observed. Since this amounts to an approximately uniform value for the density function, $f(x_{ij}|\delta_{ij}=0)$ in equation (2) can be replaced by some arbitrary constant (α). Furthermore, the travel-time information can be used to approximate the prior distribution $P(\delta_{ij}=1)$, as opposed to assigning a fixed value to the prior. If the probability density function for the travel time is denoted by, $f(t_{ij})$, then the posterior probability in equation (2) can be simplified to:

$$P(\delta_{ij} = 1|x_{ij}) \sim \frac{f(x_{ij}|\delta_{ij}=1)f(t_{ij})}{f(x_{ij}|\delta_{ij}=1)f(t_{ij})+\alpha} \quad (3)$$

where α is a positive arbitrary constant accounting for $f(x_{ij}|\delta_{ij}=0)$ and $f(\delta_{ij}=0)$. Since in matching vehicles only relative magnitude of this posterior probability is important, the selected value of α is not critical. In this research the simplified version (equation 3) is used and does not require the estimation of $f(x_{ij}|\delta_{ij}=0)$, which is an advantage in terms of model calibration and development.

In order to use equation 3, two probability distributions (i.e., $f(x_{ij}|\delta_{ij}=1)$ and $f(t_{ij})$) are needed to calculate the posterior probability. These probability density functions are found based on fitting finite mixture models to the training dataset as explained in Monsere, Cetin and Nichols (2011). Finite mixture modeling is a well-known semi-parametric technique for fitting a statistical distribution that is a weighted sum of multiple distributions. A mixture model is able to model quite complex distributions and can handle situations where a single parametric family cannot provide a satisfactory model (McLachlan and Peel, 2000). A sample mixture model is shown in Figure 4-3 for truck travel times between two stations.

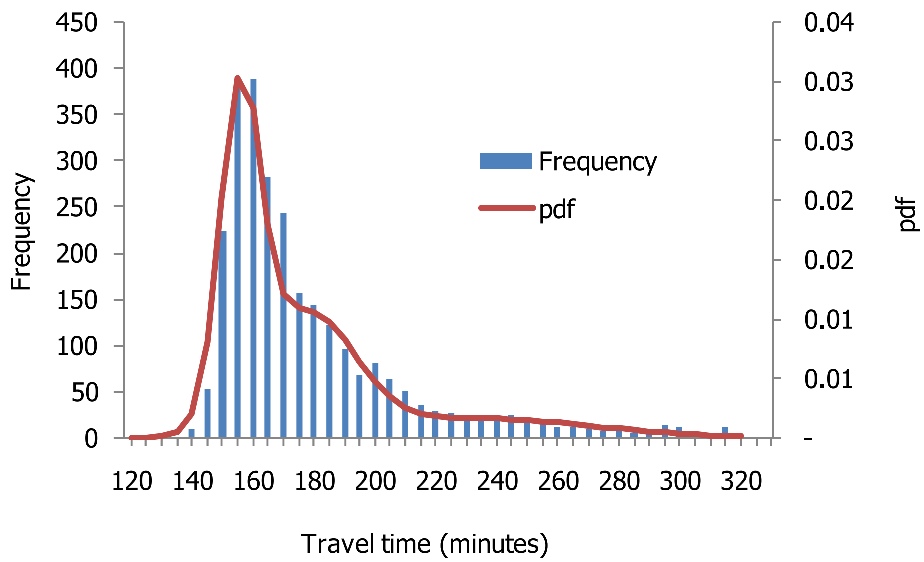

Figure 4-3 Travel time histogram for Link 234 and a probability density function (pdf) fit by mixture distributions

5.0 NEURAL NETWORKS FOR VEHICLE RE-IDENTIFICATION

In addition to the Bayesian method, the researchers also explored the applicability of neural networks in solving the vehicle re-identification problem. This section describes the steps taken to develop the neural network models.

5.1.1 Neural Networks

The neural network (NN) used for vehicle re-identification is a multilayer perceptron (MLP) as shown in Figure 5-1. In this figure, the left-most nodes represent the input layer, and the right-most nodes represent the output layer. The nodes in between represent the hidden layers. The number of neurons within the hidden layer is a variable that needs to be determined experimentally. The outputs from the previous-layer neurons become the input to all the neurons in the next layer. The connections between the neurons have weights, and these weights are the basis by which the MLP is able to learn from the given training data. Learning is achieved by adjusting the value of the weights, and this is done using the back-propagation algorithm. As shown in Figure 5-1, the neural network gives out an output of 1 or 0. For the re-identification problem, an output of 1 indicates a match, and an output of 0 indicates no match.

When a neural network receives an input pattern, the activation values of the neuron layers are propagated forward to the output neurons. The actual (known) output of the input pattern is compared with that of the output neuron, and an error value is computed. This error value is propagated backwards through the neural network to the hidden neurons to be used in weight adjustment. According to the universal approximation theorem, the back-propagation technique applied to a neural network of one hidden layer can approximate any function. The condition is that the neurons of the hidden layer must have nonlinear activation functions. In the majority of cases, the nonlinear function used is a sigmoid function.

Table 5-1 Pairs of WIM sites and the total number of trucks observed in neural network case

						Training Data (10/1 – 10/ 15)			Testing Data (Oct 16 - Oct 31)			
	Station ID			Time window (min)		Total Number of Trucks with AVI			Total Number of Trucks with AVI			
Link ID	Up	Dn	Distance (mi)	Min	Max	Up	Down	Common	Up	Down	Common	Avg Search Size
229	14	9	103	88	221	1,777	28,873	757	1,936	32,186	819	207
231	17	18	125	107	268	4,001	2,715	1,286	4,459	3,119	1,551	44
234	17	14	145	124	311	4,001	1,777	1,245	4,459	1,936	1,371	47
237	19	12	90	77	192	3,613	1,923	874	4,188	2,004	868	25

As mentioned previously, vehicular attributes consists of travel time, vehicle length, axle spacing and five axle weights. For the vehicle re-identification, data from two candidate vehicles - one from an upstream vehicle set and another from a downstream vehicle set - constitute an input pattern. The differences in the respective attributes of the two candidate vehicles are used as the input into the neural network. Since the attributes have different units (i.e., feet for distance and pound for axle weights) and measurement devices have different calibrations, it is important to normalize all input data (i.e., the differences of the vehicle attributes) before being used for neural network training and testing. Therefore, the data was normalized following formula:

$$\bar{\bar{X}}_i^{Attribute} = \frac{X_i^{Attribute} - \mu^{Attribute}}{\sigma^{Attribute}} \tag{4}$$

where *Attribute* specifies the travel time, vehicle length, axle weights or spacing, *i* represents the row, and $\bar{\bar{X}}_i^{Attribute}$ represents the normalized input for the indicated attribute and row. μ and σ are the mean and standard deviation, respectively, of each attribute. The mean and standard deviation of each attribute were calculated using data from only matching vehicles in the training dataset. These mean and variance were then used to normalize the testing dataset.

Table 5-1 provides a summary of the source data used for developing the neural network models. For each one of the four links analyzed, Table 5-1 gives the distance between the two sites, travel-time window, and total number of trucks observed in two time periods as indicated. The observations within the first 15 days of October 2007 are utilized for model training and the rest of the data for model testing. The minimum and maximum values of the time window are critical as they define the feasible time periods for the search of a matching vehicle.

The Java programming language was used in this research. The neural networks used in this research were built using a Java-based library that provides a neural network framework called JOONE (Java Object Oriented Neural Engine). The relevant documentation can be found in (sourceforge.net, 2011).

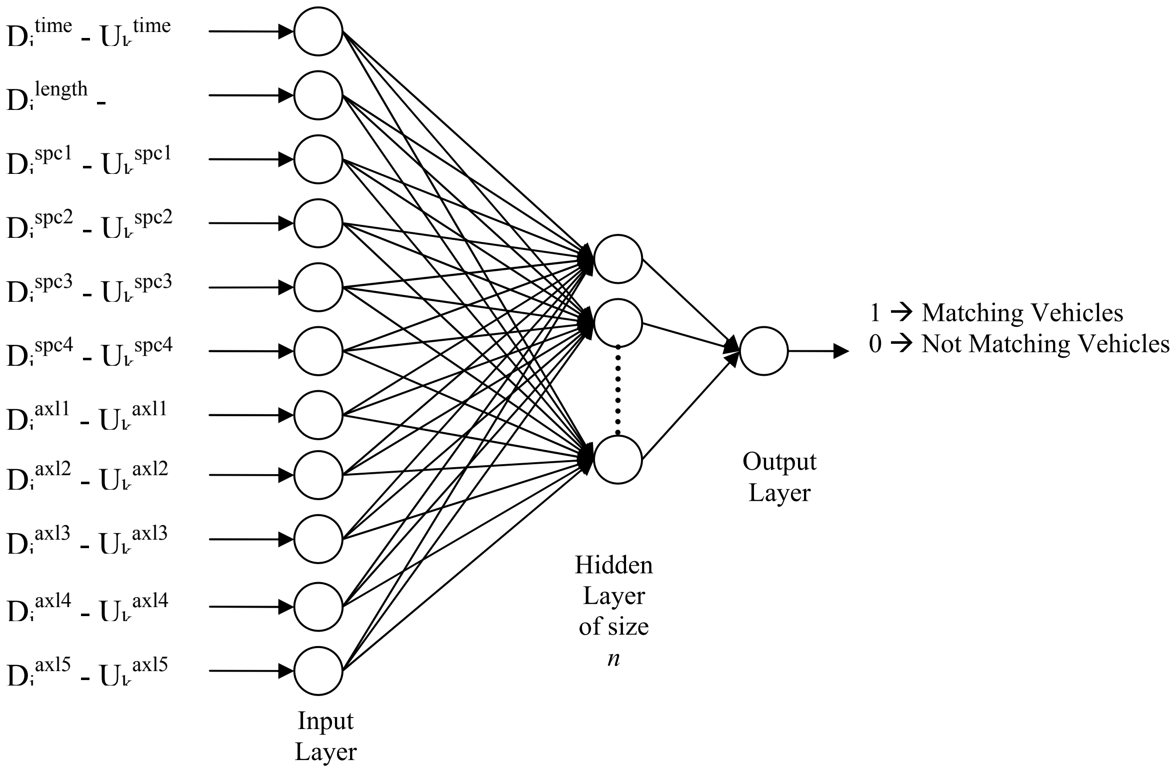

Figure 5-1 Diagram of the multilayer perceptron used for re-identification

Before discussing the experimental results, it is important to define the terms that are used to describe the output from the neural network. The vehicle re-identification problem can be categorized as a two-class problem. The neural network will identify vehicles that either match or do not match. Therefore, the output of the neural network can be something as simple as 1 and 0: 1 indicating a match and 0 indicating no match. However, in most cases, the neural network does not produce outputs as integers but as continuous values like 0.9523247 or 0.001258745. The easiest and most logical thing to do is to round up the output values to the nearest integer. Output values greater than or equal to 0.5 will be rounded to 1 and output values less than 0.5 will be rounded to 0. This threshold value of 0.5 will be used throughout the experiments to make a separation between Matching and No Matching vehicles.

Another important issue is the meaning of the output of the neural network. There are four possible output scenarios for the two neural network outputs (i.e., 0 or 1):

- Scenario 1: The system indicates a match when, in fact, it is a match.
- Scenario 2: The system does not indicate a match but, in fact, it is a match.
- Scenario 3: The system indicates a match but, in fact, there is no match.
- Scenario 4: The system does not indicate a match when, in fact, there is no match.

The terms used to describe the above four scenarios are borrowed from Signal Detection theory (Green and Swets, 1974). The terms Hit, Miss, False Alarm (FA) or Correct Rejection (CR) are used to describe Scenarios 1, 2, 3, and 4, respectively. These four terms will be used throughout the remainder of this section to describe the results of the neural networks.

While the output of the neural network model can be interpreted in four ways, Hits and FAs are of particular importance. In the ideal system, the number of Hits should be as high as possible while, at the same time, the number of FAs should be as low as possible. In the perfect scenario, the number of Hits should be the same as the number of matching vehicles and FA should be zero.

5.1.2 Determining the Size of the Hidden Layer

One of the biggest issues of using neural networks is the size of the hidden layer. Too few neurons in the hidden layer will result in poor performance, while too many neurons will result in over-learning and poor performance. It is therefore important to determine an "optimized" hidden-layer size to get the best result. This optimized hidden-layer size has to be determined experimentally. A manual search is first carried out by training NNs with various hidden-layer sizes for the training dataset of Link 234. Table 5-2 shows the results. For each hidden-layer size, 10 different NNs are trained to account for the randomness in the results (the back-propagation algorithm starts with a random set of weights). The results shown in Table 5-2 are the aggregated statistics of 10 data points. The final results show that having 25 neurons is better than other options since with this solution the average difference between Hit and FA is relatively high and the standard deviations are low.

Table 5-2 Summary of results as hidden-layer size is varied (Link 234 training data)

Hidden Layer Size	**Hits**		**False Alarm (FA)**		**Avg (Hit-FA)**
	Avg	**Stdev**	**Avg**	**Stdev**	
5	1054	22	299	21	755
10	1119	29	301	59	818
15	1110	27	259	42	851
20	1109	17	276	42	833
25	**1101**	**18**	**248**	**24**	**853**
30	1095	24	256	24	839
35	1109	18	265	35	844
40	1102	16	262	25	840
45	1105	27	265	50	840
50	1104	27	254	25	850
55	1109	23	254	26	855
60	1105	29	259	45	846

In addition to the manual method for searching the optimum hidden-layer size, simulated annealing (SA), a meta-heuristic optimization technique, is used to determine the best number of hidden-layer neurons. The details of the SA method developed for this problem can be found in Rashid (2011). The use of simulated annealing did not show any significant improvement in the accuracy of the neural network model. Therefore, the size of the hidden layer is kept at 25 neurons.

5.2 COMPARISION OF MODEL ACCURACIES

This subsection presents the results of both Bayesian and the neural network models when they are applied to the testing datasets. In order to solve the re-identification problem, a large number of comparisons need to be performed. For example, for link 229 there are 1,936 trucks in the upstream stations, for some of which (783) matching vehicles need to be identified from the downstream station dataset which contains 32,186 vehicles. Since on average there are 207 vehicles in the search space, 400,752 (1936*207) comparisons are expected to be made. The exact number of comparison is 392,751 since the size of the search space varies for each vehicle depending on its timestamp. This exact number is equal to the sum of Hit, Miss, FA and CR numbers listed

Table 5-3 for Link 229.

In

Table 5-3, the columns labeled 1-5 represent five separate trials performed for each of the four links. Each trial is based on a corresponding neural network model that is trained on the training datasets. Therefore, to generate the results shown in

Table 5-3, 20 (5*4) neural network models are created from the training datasets. The variation from one trial to another is due to the variation in the results of the back-propagation algorithm when applied each time to the same training dataset. The columns AVG and STDEV indicate the average and standard deviations of the results of five trials. The values in column "% Hit" are calculated by dividing the average Hits by the sum of Hits and Miss.

Based on the results in

Table 5-3, the performance of the neural network models varies from link to link; Link 231 having the best while Link 229 the worst performance. This may be expected as the size of the search space is much larger for Link 229 than that of Link 231 (i.e., 207 versus 47). Even though Link 237 has the lowest search space (25) its performance is not highest since there are other factors that contribute to the accuracy of matching, like the calibration and precision of the WIM sensors (Cetin, Monsere, Nichols et al., 2011)

The results in

Table 5-3 are generated when all output values greater than or equal to 0.5 are declared as a matching instance and output values less than 0.5 as not a match, as explained before. This threshold value of 0.5 can be increased to match fewer vehicles, and thereby, the error rate can be improved. For example, rather than trying to match all vehicles, a small but accurately matched sample may be more relevant for the particular application (e.g., travel-time estimation or WIM-sensor calibration). In that case, the error can be measured more appropriately by the equation below:

$$\frac{FA}{Hit + FA} \times 100 \tag{5}$$

Table 5-3 Summary of the results obtained from the neural network models for the four links

			Link 229					
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Hit**
Hit	532	521	488	498	541	516	22	66
Miss	251	262	295	285	242	267	22	
FA	308	282	238	249	289	273	29	
CR	391660	391686	391730	391719	391679	391695	29	
			Link 231					
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Hit**
Hit	1330	1321	1328	1317	1306	1320	10	86
Miss	200	209	202	213	224	210	10	
FA	404	311	312	283	273	317	52	
CR	134187	134280	134279	134308	134318	134274	52	
			Link 234					
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Hit**
Hit	1138	1129	1162	1099	1145	1135	23	83
Miss	223	232	199	262	216	226	23	
FA	280	271	335	242	331	292	40	
CR	89506	89515	89451	89544	89455	89494	40	
			Link 237					
Trials	**1**	**2**	**3**	**4**	**5**	**AVG**	**STD DEV**	**% Hit**
Hit	678	648	623	605	639	639	27	74
Miss	183	213	238	256	222	222	27	
FA	113	94	104	88	83	96	12	
CR	49552	49571	49561	49577	49582	49569	12	

The results of the Bayesian models and neural network models are compared to each other in terms of the error defined in equation 5. The threshold value is varied incrementally to generate results at different levels. These results are then plotted in Figure 5-2, which shows the error graphs of the Bayesian and the neural network models for Links 229, 231, 234 and 237. The vertical axis in each graph shows the error calculated by equation 5. The horizontal axis shows the total number of matched vehicles as a percentage of the common vehicles (or total actual matching vehicles) on that link, which is determined by,

$$\frac{Hit + FA}{Total\ Common\ Trucks} \times 100. \tag{6}$$

The number of common trucks for each link is listed in Table 5-1. These are 819, 1551, 1371 and 868 for Links 229, 231, 234 and 237, respectively.

The four charts in Figure 5-2 show the results of the Bayesian model and those of the five neural networks for each link. Overall, as fewer vehicles are being matched the error rate goes down. In the graphs for Link 229, when the system is matching 90% of vehicles, the error in the Bayesian model is around 25%. For the neural network models, the error at the same point is around 30%. This approximately 5% difference is generally maintained for nearly all percentage of matches for Link 229.

In the chart for Link 231, when the system is matching 90% of vehicles, the error in the Bayesian model is around 6%. For the neural network models, the error at the 90% match rate is somewhere around 12%. At 100% level, the error for the Bayesian model is about 10%, and it is 16% for the neural network models. In the figure for Link 234, when the system is matching 90% of vehicles, the error in the Bayesian model is around 11%. For the neural network models the error is around 16%. Finally, for Link 237, three of the neural network models seem to be outperforming the Bayesian model by about 3-4%, especially for values above 60% on the horizontal axis.

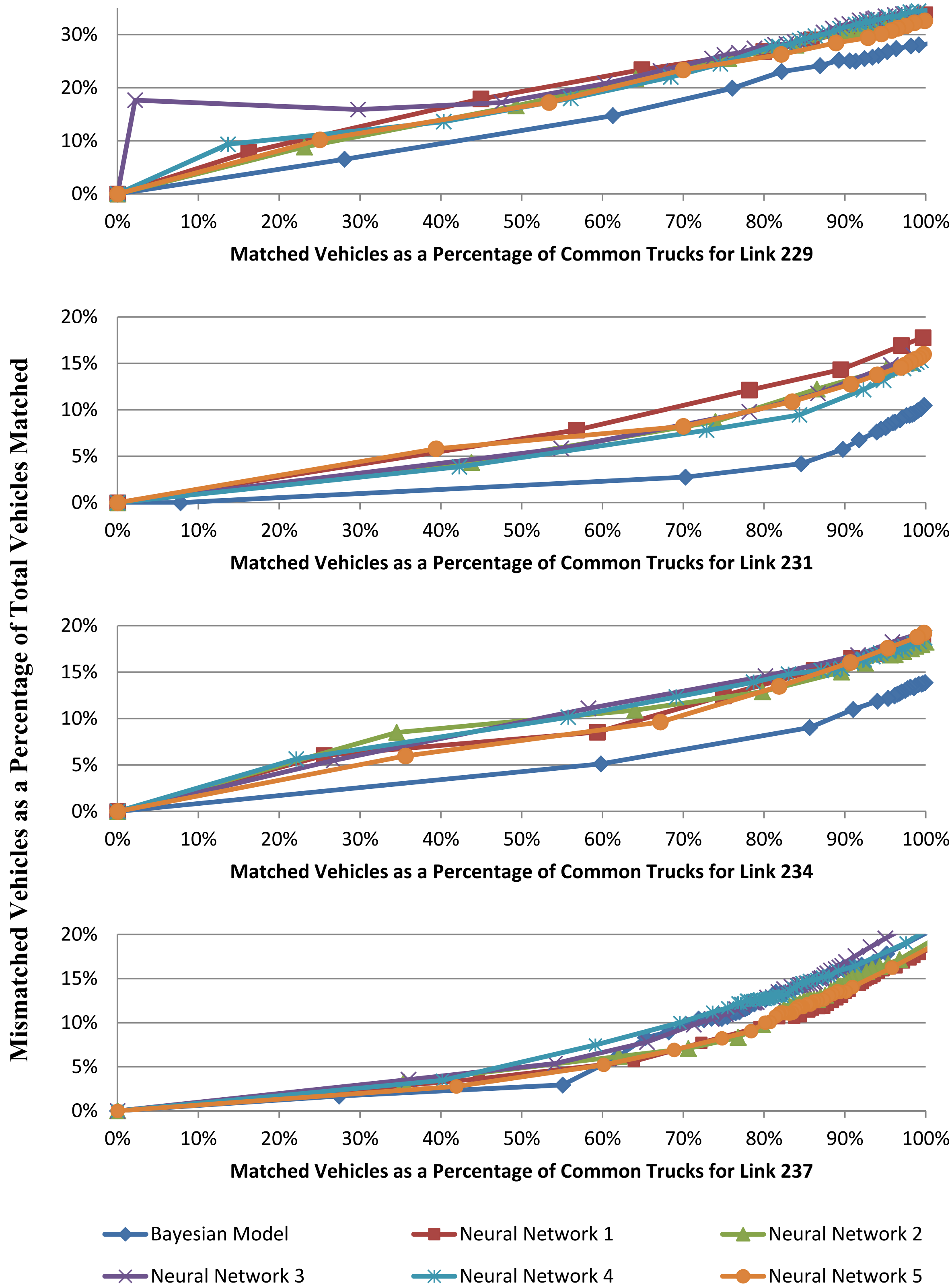

Figure 5-2 Comparison between the Bayesian and NN models for all four links

5.3 SUMMARY

To compare the performance of the Bayesian and neural network models in re-identifying vehicles, a large dataset from weigh-in-motion (WIM) stations in Oregon is utilized. The models are trained and tested on four different pairs of WIM sites (called links). For each one of these links, neural network models are developed and tested. Special attention is paid to ensure that the neural network design (e.g., number of neurons in the hidden layer) is optimal. The performance of neural network models is then compared to the Bayesian models. Overall, with the exception of one case, the Bayesian models are found to outperform the neural networks based on the datasets considered in this study.

The results from this study will be helpful to researchers in developing algorithms for solving the vehicle re-identification problem. In terms of future research, it will be desirable to investigate whether the Bayesian models and/or neural network models are transferable between two links as opposed to estimating a model for each link separately. These models can also be tested on different datasets to further evaluate their performance.

6.0 INVESTIGATING THE KEY FACTORS AFFECTING THE ACCURACY OF RE-IDENTIFICATION

In this section, the key factors that impact the accuracy of vehicle re-identification algorithms are investigated. The analyses are performed by employing the Bayesian re-identification algorithm to match vehicles that cross upstream and downstream pairs of weigh-in-motion (WIM) sites that are separated by long distances ranging from 70 to 214 miles. The data to support this research come from 17 fixed WIM sites in Oregon. Data from 14 different pairs of WIM sites are used to evaluate how matching accuracy is impacted by various factors including the distance between two sites, travel-time variability, truck volumes, and sensor accuracy or consistency of measurements. After running the vehicle re-identification algorithm for each one of these 14 pairs of sites, the matching error rates are reported. The results from the testing datasets showed a large variation in terms of accuracy. It is found that sensor accuracy and volumes have the greatest impacts on matching accuracy, whereas the distance alone does not have a significant impact.

6.1 WIM DATA

Table 6-1 provides a summary of the source data used in this study. For each one of the 14 links analyzed, Table 6-1 gives the distance between the two sites, travel-time window, and total number of trucks observed in two time periods as indicated. The observations within the first 15 days of October 2007 are utilized for model training and the rest of the data for model testing. The minimum and maximum values of the time window are critical as they define the feasible time periods for the search of a matching vehicle. The minimum time is found by dividing the distance by a 70 mph travel speed, whereas the maximum time is assumed to be 2.5 times the minimum time (approximately 28 mph). Even though not all trucks necessarily reach the downstream site within these time boundaries, these are deemed appropriate for the purpose of this research since most of the trucks travel within this time window. Table 6-1 also identifies the total number of trucks that had a transponder at the upstream station (Up), the downstream station (Down) and the number of trucks that were matched via transponder at both stations (Common). Finally, the last column in Table 6-1shows the average number of vehicles within the "search space" for each vehicle that is being matched in the testing dataset. These are the average number of vehicles among which a true match is identified. As the time window becomes larger or the vehicle volumes increase, the average search size increases which makes it more challenging to find the true match.

Table 6-1 Pairs of WIM sites and the total number of trucks observed

Link ID	Station ID Up	Station ID Dn	Distance (mi)	Time window (min) Min	Time window (min) Max	Training Data (Oct 1 - Oct 15) Total Number of Trucks with AVI Up	Down	Common	Testing Data (Oct 16 - Oct 31) Total Number of Trucks with AVI Up	Down	Common	Avg Search Size
201	1	2	126	108	271	11,762	12,874	8,549	11,294	5,353	2,968	87
202	2	3	173	148	370	12,874	13,716	3,570	5,353	15,419	1,684	181
205	3	10	79	68	169	13,716	32,645	2,284	15,419	35,873	2,546	191
208	4	5	214	183	459	15,986	13,516	4,110	17,460	15,247	4,579	236
210	5	6	96	82	206	13,516	7,388	5,594	15,247	8,836	6,900	94
211	7	8	93	80	200	16,355	16,589	8,710	18,217	18,718	9,814	118
214	8	9	165	141	354	16,589	28,873	8,126	18,718	32,186	9,518	309
217	9	4	70	60	150	28,873	15,986	1,931	32,186	17,460	2,098	147
223	11	13	90	77	193	2,068	3,951	1,151	2,152	4,559	1,179	28
227	13	20	163	140	350	3,951	3,881	1,174	4,559	4,217	1,375	49
229	14	9	103	88	221	1,777	28,873	757	1,936	32,186	819	207
231	17	18	125	107	268	4,001	2,715	1,286	4,459	3,119	1,551	44
234	17	14	145	124	311	4,001	1,777	1,245	4,459	1,936	1,371	47
237	19	12	90	77	192	3,613	1,923	874	4,188	2,004	868	25

6.2 APPLICATION OF THE BAYESIAN RE-IDENTIFICATION METHOD

The Bayesian method described in Section 4 is applied to the datasets for the 14 links discussed above. The overall methodology followed to match vehicles on each link has two main steps: model training and model testing. For each link dataset, mixture models are estimated based on the training dataset. These mixture models, $f(t_{ij})$ and $f(x_{ij}|\delta_{ij}=1)$, are then used as inputs in the testing phase on the new observations for the same link to re-identify vehicles. For screening out vehicles for which there is no match, a simple method (i.e., the naïve method described in Cetin, Nichols and Monsere (2011)) is employed to improve the accuracy. This is accomplished by not matching those vehicles for which the posterior probability in equation 3 (see Section 4) is less than a threshold which ranges from 0 to 1. The results are then analyzed by varying the threshold from 0 to 1for the screening method to evaluate how accuracy is changing. These are then used to obtain the final results which are shown as tradeoff graphs, as in Figure 6-1.

In Figure 6-1, the horizontal axis shows the total number of vehicles being matched which is presented as a percentage of the common trucks on each link since the number of common trucks is different on each link. The total numbers of common trucks on each link are shown in Table 6-1. This percentage on the x-axis can be larger than 100% since all vehicles observed at one site are being matched to the vehicles in the other site (when no screening is applied), and the number of those vehicles is always larger than the common trucks. The vertical axis shows the mismatched vehicles as a percentage of the total vehicles being matched. As the screening threshold is increased from zero to one, the total number of vehicles being matched decreases. The vehicles being matched are those for which the posterior probability in equation 2 is greater than the screening threshold. As it can be observed in Figure 6-1, there is a wide range of accuracy levels. The factors that can potentially explain this variation in accuracy are investigated in the next section.

6.3 EVALUATION OF FACTORS AFFECTING MATCHING ACCURACY

In this section, the potential factors that can explain the large range of error levels observed in Figure 6-1 are explored. For example, for link 231 the error level is 90% (i.e., when the total number of vehicles matched is 90% of the common vehicles on that link) is about 6%, whereas for link 205 it is 57%. Potential factors that can explain this large variation are listed in the columns of Table 6-2. It was hypothesized that distance between station pairs (*dist*), the travel-time variance (*ttvar*), sensor accuracy and the volume of trucks between each station-pair (link) would influence the accuracy of the matching algorithm. Variables characterizing these are summarized for each link in Table 6-2. The column heading is given with the variable name in (), followed by a description:

- Travel-time variance (*ttvar*): Variance of the travel times of the common vehicles that cross both sites in the training dataset.
- Set 1 (*set_1*): The total number of vehicles with transponders in the testing data from which a match is found for the vehicles in Set 2. Between the upstream and downstream sites, the site with the larger number of vehicles is represented as Set 1. Vehicles in Set 2 are being matched to the vehicles in Set 1.
- Set 2 (*set_2*): The total number of vehicles with transponders in the testing data for which a match is found in Set 1. Between the upstream and downstream sites, the site with the fewer number of vehicles is represented as Set 2.
- Common (*common*): The total number of vehicles with transponders in the testing dataset that cross both upstream and downstream sites.
- Average search size (*search*): The average number of vehicles considered in Set 1 for searching a match for each vehicle in Set 2. For a given vehicle, search size or search space is determined based on the time window as described previously.
- Mean and Standard Deviation of the Axle 1-2 spacing (*mu_sp1, sd_sp1*): For the vehicles matched by transponder ID, the average difference in measured distances between axles 1 and 2 at the upstream and downstream sites (normalized by the measured spacing at the upstream). If the WIM sites had identical calibration (not necessarily accurate) and vehicle speeds were constant over the sensors, the mean would be close to zero and the standard deviation would be low. A negative value indicates that the average spacing measured at the upstream site was less than the average spacing measured at the downstream site.
- Mean and Standard Deviation of the Axle 1 weight (*mu_ax1, sd_ax1*): For the vehicles matched by transponder ID, the average difference in measured axle 1 weights at the upstream and downstream sites (normalized by the measured weight at the upstream). Deviations of the

mean and standard deviation from zero are due to differences in sensor calibration and pavement smoothness at the two WIM sites.

- Percent of ODOT WIM Class 11 trucks (*pct_11*): The percentage of Oregon DOT class 11 trucks (FHWA class 9) among the common vehicles crossing both upstream and downstream sites. Generally, this is the most common truck type on roads in the United States.
- Error at 90% match (*error*): This is the dependent variable. This is the error shown on the y-axis of Figure 6-1 when x is 90% (i.e., when the total number of vehicles matched is 90% of the common vehicles on that link). Other values can also be selected to quantify the error but it will not impact the evaluation and results.

Table 6-2 Summary of link parameters

			Volume				Sensor accuracy					
			Total Number of Trucks with Tags (Testing Dataset)				Difference in Axle 1-2 Spacing			Difference in Axle 1 Weight		
Link ID	Dist. (mi)	Travel time variance	Set 1	Set 2	Common	Avg Search Size	Mean	St. Dev	% Class 11	Mean	St. Dev	Error at 90 % match
201	126	240	11,294	5,353	2,968	87	-5.1%	11.7%	76%	-1.4%	11.3%	0.380
202	173	584	15,419	5,353	1,684	181	9.5%	41.8%	72%	7.8%	39.5%	0.426
205	79	126	35,873	15,419	2,546	191	-0.1%	13.0%	51%	-15.0%	8.8%	0.571
208	214	825	17,460	15,247	4,579	236	15.1%	8.8%	71%	0.6%	20.1%	0.525
210	96	123	15,247	8,836	6,900	94	-9.2%	9.1%	73%	-2.1%	17.1%	0.246
211	93	156	18,718	18,217	9,814	118	2.9%	7.2%	85%	-3.5%	22.2%	0.499
214	165	713	32,186	18,718	9,518	309	-1.9%	8.6%	76%	14.2%	55.6%	0.492
217	70	92	32,186	17,460	2,098	147	-5.9%	11.4%	62%	6.9%	14.0%	0.479
223	90	129	4,559	2,152	1,179	28	1.1%	6.5%	39%	-9.2%	8.5%	0.123
227	163	536	4,559	4,217	1,375	49	2.9%	3.8%	92%	0.4%	9.0%	0.199
229	103	214	32,186	1,936	819	207	-1.8%	5.0%	78%	-4.6%	8.0%	0.251
231	125	191	4,459	3,119	1,551	44	-6.1%	3.3%	86%	4.0%	6.6%	0.059
234	145	315	4,459	1,936	1,371	47	-0.7%	5.3%	70%	1.1%	9.0%	0.106
237	90	38	4,188	2,004	868	25	-7.0%	3.6%	33%	10.1%	8.5%	0.162

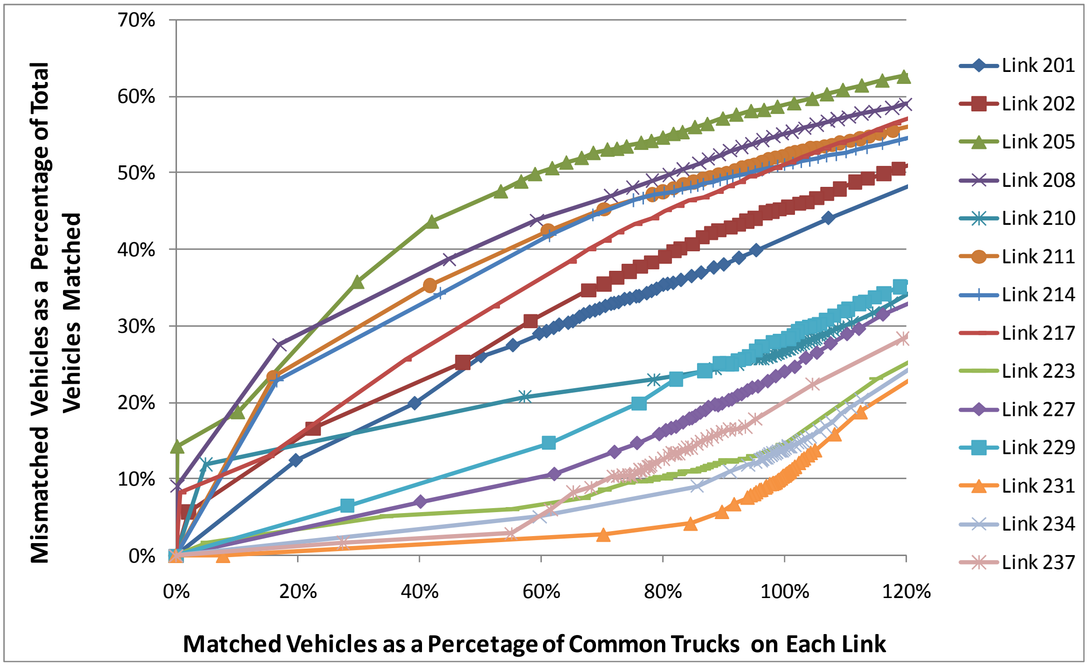

Figure 6-1 Percent error versus the total vehicles matched.

There are limited samples (14 points) to explore the variation of link-level error which may limit the robustness of conclusions. Nonetheless, in an effort to quantitatively analyze the limited data, a multivariate regression model was fit (error at 90% being the dependent variable). To begin, a scatterplot matrix of all candidate variables was created for exploratory data analysis. Because of the information density of these plots, a sample with fewer candidate variables is shown in Figure 6-2 with a subset of the candidate variables. In the plot, each cell with data points is scatterplot of two variables. The *x*-axis for each plot in a common column can be read along the top of the matrix; the labels for each *y*-axis for each common row can be read to the right of the matrix. The left-most plot in each row gives the variable name, a kernel density estimate, and a rug plot. In each plot pair, the dashed line is a simple linear regression model fit to the data in each cell. Scatterplots of all variables (not shown here) hinted at a strong correlation between the distance between stations (*dist*) and the travel-time variance (*ttvar*) and weaker correlations between the variables that measure the volume of trucks between station pairs (*set_1, set_2, common, search*). The correlation matrix of independent variables confirmed this observation (*dist* and *ttvar* correlation coefficient of 0.947). The exploratory plots also identified a possible outlier for the standard deviation of the axle 1-2 spacing error (*sd_sp1*) for link 202 (see rightmost data point in Column 5 plots in Figure 6-2).

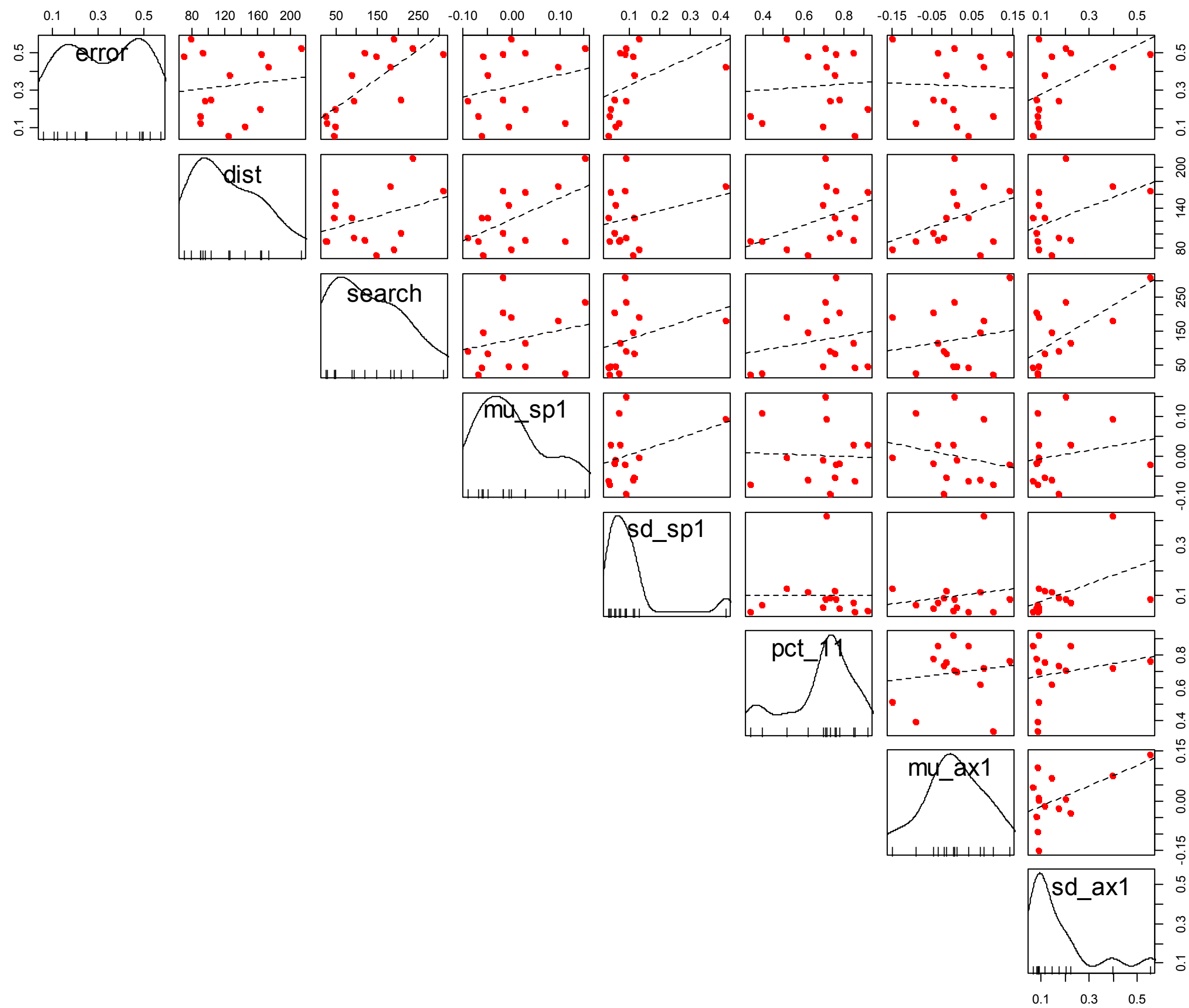

Figure 6-2 Scatterplot matrix

To fit the multivariate regression model, candidate variables were manually selected; the resulting model diagnostics (i.e., AIC, R^2 and parameter estimates) were compared. In most model combinations, *distance* and *ttvar* were not significant. *Search* and *sd_sp1* were significant, though *sd_ax1* was not. All diagnostic plots confirmed the outlier for the observation *sd_sp1* of link 202 (0.42). Though there is no reason to suspect that this data point is erroneous, it was shown to have high influence in the model diagnostic plots. The data point was removed and the regression model for *search* and *sd_sp1* recalibrated. Finally, recognizing that the intercept estimate was not significant (and that when search space is 0, error should be zero) the intercept was removed. The final selected model for *search* and *sd_sp1*, with the outlier removed, is given below:

```
           Coefficients:
          Estimate Std. Error t value Pr(>|t|)
search 0.0009897  0.0003097   3.195 0.008525 **
sd_sp1 2.6648429  0.5706331   4.670 0.000683 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
```

```
Residual standard error: 0.08609 on 11 degrees of freedom
Multiple R-squared: 0.9516,     Adjusted R-squared: 0.9428
F-statistic: 108.2 on 2 and 11 DF,  p-value: 5.82e-08
```

The adjusted R^2 is 0.94; the coefficients for the two variables are significant at the p=0.05 level. The results of the model quantitatively confirm the observations in the scatterplot. Reading the first row of independent variables plotted against *error* in Figure 6-2, both *search* and *sd_sp1* show strong linear relationships. It should be noted that removing the *sd_sp1* outlier does little to change the conclusions of the model (sign and relative magnitude of coefficients remain the same), though it does improve the overall fit.

The analysis hints that the error rate of the matching algorithm increases both with search size and decreased consistency of sensor measurements between sites (i.e., high standard deviation). The average search size is a function of both the distance (time window) and the volume of trucks between station pairs. More distance between stations and more vehicles in the search window increases the likelihood that the algorithm will incorrectly match vehicles. For sensor accuracy, the axle-spacing metric explained more of the error difference than the weight-accuracy metric (*explained in the next paragraph*). While the weight-sensors errors have much larger variance (see Figure 6-3), they are very similar for each link and are not able to explain much of the link-to-link error differences. The spacing-sensor error, however, is much more varied on a link basis.

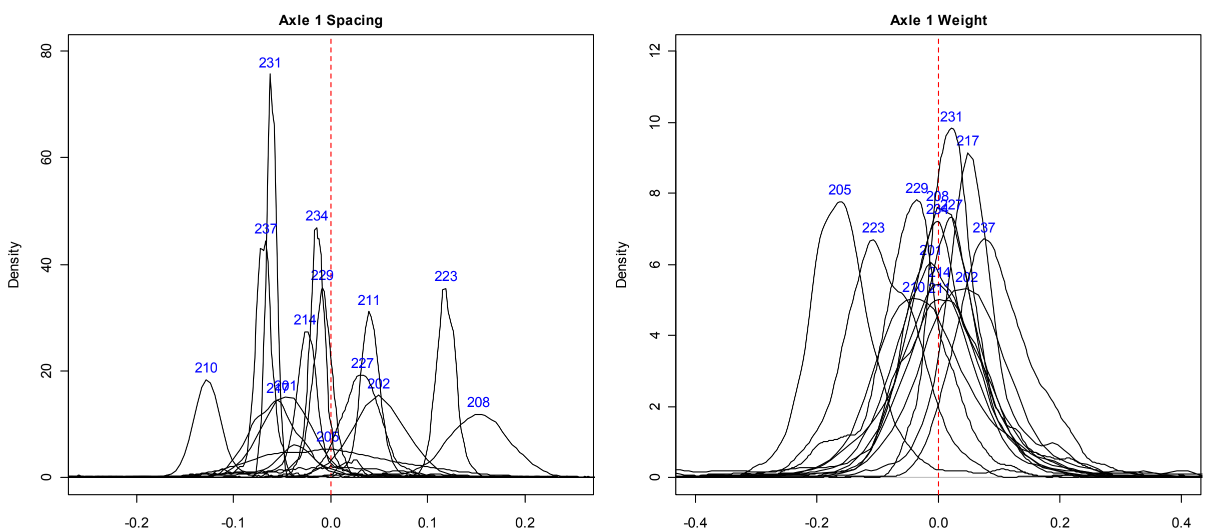

Figure 6-3 Kernel density estimates of normalized errors of the transponder-matched vehicles in training data set, axle 1-2 spacing and axle 1 weight

6.4 SENSOR ACCURACY

Some of the reasons for the upstream and downstream observations to not be the same for given vehicles include sensor calibration drift, failing sensors, vehicles not crossing the sensors properly, and the AVI tag not being properly matched to the WIM/AVC record. Vehicles not crossing the sensors properly, AVI tag mismatching, and some sensor failures are random events and would show up as outliers in the density plot. Sensor calibration problems and some sensor failures would be consistent across all observations, causing a shift in the density plot rather than outliers. More information on WIM-sensor errors can be found in the literature (Nichols, Bullock and Schneider, 2009)

In Figure 6-3, kernel density plots of the normalized errors of transponder-matched vehicles from the training data are shown. In the regression model, only the point estimates of the mean and standard deviation of these plots are used to explain link error. The means of the weight and spacing plots are most likely an indicator of the difference in sensor calibration parameters. For example, on link 205, the mean difference in weight is -15%, which indicates that the upstream sensors are measuring weights that are 15% lighter than the downstream sensors on that link. Assuming the calibration remains constant at a site, the difference in calibration precision (i.e., closeness to the correct value) between sites should not affect the error rate of the matching algorithm. This is qualitatively supported by the plots showing *error* vs. *mu_sp1* (row 1 column 4 in Figure 6-2) and *error* vs. *mu_ax1* (row 1 column 7 in Figure 6-2), where there is no linear relationship apparent. This is expected since the mixture model parameters that are estimated for a given pair of WIM sites can account for the observed differences in the measured parameters for matched vehicles.

The cause of increased standard deviations of the weight and spacing plots are more difficult to explain. The axle-spacing measurement reported at a WIM site is a function of the speed of the vehicle. If the speed of an individual vehicle is not constant as it crosses the sensor, the axle-spacing data for that vehicle will not be accurate. Therefore, locations with congested traffic conditions would likely report inaccurate axle-spacing data during congested periods. Another reason for increased standard deviations in the axle spacing could be failing axle-sensor detectors. Sometimes when a sensor fails, it will miss light-weight axles, causing axle-spacing data and classification data to be erroneous. Under ideal conditions, the standard deviation for this measurement should be very low, similar to link 231. The large variations illustrated by link 202, 205, 208, and others are unexpected and warrant further investigation.

Large standard deviations in the axle-weight error are most likely attributed to the pavement profile at the WIM sites, different sensor types, poor calibration or a failing sensor. If the pavement approaching the weight sensors is not smooth, the vehicle dynamics will increase and cause more variations on the measured weights compared to smooth pavement. However, the variations due to the pavement profile will vary from vehicle to vehicle, depending on the suspension type, axle spacing, vehicle speed, and other parameters that are not consistent.

An increase in standard deviation of the measurement error of the axle spacing or the axle weight between two sites is expected to increase the matching error rate. Consider a five-axle FHWA class 9 vehicle that crosses two WIM sites that have different calibration precision, but are consistent and repeatable in their measurements. Calibration factors for axle spacing, in general, are linear corrections. Therefore, the axle-spacing data are going to differ, but the difference should be constant by a percentage that is related to the difference in calibration factors, which is the mean that is estimated here. Calibration factors for axle weight can be linear or non-linear, depending on the type of sensor and its susceptibility to temperature variation. If the sensor measurements are not consistent and repeatable, those differences are not going to be consistent from vehicle to vehicle and result in high standard deviations. The matching algorithm is not capable of overcoming this deficiency in system performance. The sensors at the two WIM sites do not necessarily need to be correctly calibrated for successful re-identification, but the sensors need to perform consistent and repeatable measurements.

6.5 SUMMARY

The key factors that impact the accuracy in re-identifying vehicles are analyzed. Data from 17 WIM stations in Oregon is used to create 14 pairs of sites (called links). For each one of these 14 links, the Bayesian re-identification models are estimated and tested. The results from the testing dataset showed a large variation in terms of accuracy. Some links, such as links 234 and 231, have very low error rate, whereas others, such as links 205 and 208, have very large errors. Several explanatory variables, including distance between the sites, travel-time variability, truck volumes, and sensor accuracy or consistency of measurements, are analyzed to explain the variation in matching accuracy. It is found that the consistency of sensors measuring axle spacing and the number of vehicles in the search window have the greatest impact on matching accuracy. Error increases with search size and with decreased sensor consistency/repeatability. The distance alone between the two sites is not found to be a significant factor. These results imply that matching vehicles over long distances is feasible provided that sensors at two sites are accurate and provide consistent measurements and the volumes are not very high.

7.0 CONCLUSIONS

This project examined the use of vehicle attribute data that are typically obtained from WIM and AVC sensors for anonymously re-identifying commercial vehicles so that their movements can be tracked. Tracking the movement of individual vehicles between different data collections sites provides valuable information for the estimation of travel times, travel delays and origin-destination (OD) flows. Even though the data from transponder-equipped trucks can also be used for the estimation of travel times and OD flows, these trucks represent less than half of all trucks or a small fraction, depending on the selected sites. For example, in Oregon, on average, the rate of transponder-equipped trucks is about 40%. In addition, vehicle re-identification based on vehicle-attribute data does not raise any privacy concerns as is the case with other types of vehicle-tracking technologies (AVI, license plate recognition, etc.).

The authors developed both Bayesian and neural network models to solve the vehicle re-identification problem. To compare the performance of the Bayesian and neural network models in re-identifying vehicles, a large dataset from WIM stations in Oregon is utilized. The models are trained and tested on four different pairs of WIM sites (called links). For each one of these links, neural network models are developed and tested. Special attention is paid to ensure that the neural network design (e.g., number of neurons in the hidden layer) is optimal. The performance of neural network models is then compared to the Bayesian models. Overall, with the exception of one case, the Bayesian models are found to outperform the neural networks based on the datasets considered in this study.

The authors also investigated the factors that impact the accuracy of the re-identification algorithms. Data from 17 WIM stations in Oregon is used to create 14 pairs of sites (called links). For each one of these 14 links, the Bayesian re-identification models are estimated and tested. The results from the testing dataset showed a large variation in terms of accuracy. Some links, such as links 234 and 231, have very low error rate, whereas others, such as links 205 and 208, have very large errors. Several explanatory variables, including distance between the sites, travel-time variability, truck volumes, and sensor accuracy or consistency of measurements, are analyzed to explain the variation in matching accuracy. It is found that the consistency of sensors measuring axle spacing and the number of vehicles in the search window have the greatest impact on matching accuracy. Error increases with search size and with decreased sensor consistency/repeatability. The distance alone between the two sites is not found to be a significant factor. These results imply that matching vehicles over long distances is feasible provided that sensors at two sites are accurate and provide consistent measurements and the volumes are not very high.

Overall, for travel-time estimation purposes, the methods presented in this report can be used effectively to match commercial vehicles crossing two data collection sites that are separated by long distances.

8.0 REFERENCES

Cetin, M., C. M. Monsere, A. P. Nichols, et al. (2011). "Investigating the Key Factors Affecting the Accuracy of Re-identification of Trucks over Long Distances Based on Axle Measurement Data." Transportation Research Record **2243**. pp. 1-8.

Cetin, M. and A. P. Nichols (2009). "Improving the accuracy of vehicle re-Identification by solving the assignment problem." Transportation Research Record: Journal of the Transportation Research Board **2129**: 1-8.

Cetin, M., A. P. Nichols and C. M. Monsere (2011). "Bayesian Models for Re-identification of Trucks over Long Distances Based on Axle Measurement Data " Journal of Intelligent Transportation Systems **15**(1): 1-12.

Christiansen, I. and L. Hauer (1996). Probing for Travel Time: Norway Applies AVI and WIM Technologies for Section Probe Data. Traffic Technology International. **Aug/Sep 1996:** 41-44.

Coifman, B. (2003). "Identifying the onset of congestion rapidly with existing traffic detectors." Transportation Research Part a-Policy and Practice **37**(3): 277-291.

Coifman, B. and M. Cassidy (2002). "Vehicle reidentification and travel time measurement on congested freeways." Transportation Research, Part A (Policy and Practice) **36A**(10): 899-917.

Coifman, B. and S. Krishnamurthy (2007). "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure." Transportation Research Part C: Emerging Technologies **15**(3): 135-153.

Dion, F. and H. Rakha (2006). "Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data for Low Sampling Rates." Transportation Research Part B: Methodological **40**(9): 745-766.

Elkins, L. and C. Higgins (2008). Development of Truck Axle Spectra from Oregon Weigh-in-Motion Data for Use in Pavement Design and Analysis. Final Report FHWA-OR-RD-08-06, Oregon Department of Transportation.

Green, D. M. and J. A. Swets (1974). Signal detection theory and psychophysics. Huntington, N.Y., R.E. Krieger Pub. Co.

Hellinga, B. (2001). "Automated Vehicle Identification Tag-Matching Algorithms for Estimating Vehicle Travel Times: Comparative Assessment." Transportation Research Record **1774**: 106-114.

Liu, H. X., J.-S. Oh and W. Recker (2002). "Adaptive signal control system with online performance measure for a single intersection." Transportation Research Record(1811): 131-138.

McLachlan, G. and D. Peel (2000). Finite Mixture Models, John Wiley & Sons.

Monsere, C. M., M. Cetin and A. P. Nichols (2011). Exploratory Methods for Truck Re-identification in a Statewide Network Based on Axle Weight and Axle Spacing Data to Enhance Freight Metrics, Oregon Transportation Research and Education Consortium (OTREC), OTREC-RR-11-07.

Monsere, C. M., M. Wolfe, H. Alawakiel, et al. (2009). Developing Corridor-Level Truck Travel Time Estimates and Other Freight Performance Measures from Archived ITS Data. Final Report SPR 304-361, OTREC-RR-09-10.

Nichols, A. P., D. Bullock and W. Schneider (2009). "Detecting differential drift in weigh-in-motion wheel track sensors." Transportation Research Record(Compendex): 135-144.
Oh, C., S. G. Ritchie and S. T. Jeng (2007). "Anonymous vehicle reidentification using heterogeneous detection systems." IEEE Transactions on Intelligent Transportation Systems **8**(3): 460-469.
Oh, C., A. Tok and S. G. Ritchie (2005). "Real-time freeway level of service using inductive-signature-based vehicle reidentification system." IEEE Transactions on Intelligent Transportation Systems **6**(2): 138-146.
Rashid, T. (2011). Solving the Vehicle Re-Identification Problem by Using Neural Networks. MSVE, MS Thesis, Old Dominion University,Norfolk, VA.
Shuldiner, P. and J. Upchurch (2001). Automated Travel Time Data for a Regional Traveler Information System, Institute of Transportation Engineers.
sourceforge.net (2011). "An Object Oriented Neural Engine." from http://sourceforge.net/projects/joone/.
Sun, C., G. Arr and R. P. Ramachandran (2003). "Vehicle reidentification as method for deriving travel time and travel time distributions: Investigation." Transportation Research Record(1826): 25-31.
Sun, C., S. G. Ritchie, K. Tsai, et al. (1999). "Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways." Transportation Research Part C: Emerging Technologies **7**(4): 167-185.
Tawfik, A. Y., B. Abdulhai, A. Peng, et al. (2004). "Using decision trees to improve the accuracy of vehicle signature reidentification." Transportation Research Record(1886): 24-33.
TRB (2003). A Concept for a National Freight Data Program: Special Report 276, Transportation Research Board, The National Academies.
Turner, S., W. Eisele, R. Benz, et al. (1998). Travel Time Data Collection Handbook, Texas Transportation Institute, A&M University, College Station.

OTREC
OREGON TRANSPORTATION RESEARCH
AND EDUCATION CONSORTIUM

P.O. Box 751
Portland, OR 97207

OTREC is dedicated to stimulating and conducting collaborative multi-disciplinary research on multi-modal surface transportation issues, educating a diverse array of current practitioners and future leaders in the transportation field, and encouraging implementation of relevant research results.