AN INVESTIGATION OF THE DETERMINANTS OF TRUCK DRIVERS' ROUTING BEHAVIOR

A Thesis

by

XIAOQIANG KONG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Yunlong Zhang
Committee Members,	Daren B. H. Cline
	William (Bill) L. Eisele
Head of Department,	Robin Autenrieth

August 2017

Major Subject: Civil Engineering

ABSTRACT

Understanding truck drivers' routing selection behavior according to congestion level, travel time reliability, and other factors can not only help transportation agencies improve the efficiency of traffic management but also increase the accuracy of travel time predictions. However, most of the existing studies on this subject used non-empirical methods such as stated preference, experimental, and theoretical modeling and simulations because real field data were not available.

This research analyzes 17,024 observed trips on I-495 crossing through Maryland, Virginia, and Washington, D.C., to explore how truck drivers make routing decisions based on real-time congestion information, travel time reliability, and other factors such as rush hour and day of the week. The east loop of I-495 is defined as the east route, and the north loop is defined as the north route. The results show that the odds of selecting the north route significantly decrease if the travel time index ratio between the north route and east route increases. The research also demonstrates that the planning time index ratio has a significant impact on routing selection. Also, freight drivers' routing decisions are influenced differently by factors such as morning rush hour, afternoon rush hour, and whether it is a weekday or weekend. A similarly detailed aggregate freight dataset from the Dallas–Fort Worth area validates the results from the Maryland dataset.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Zhang, and my committee members, Dr. Cline and Dr. Eisele, for their guidance, patience, and engagement in my research study.

I am very grateful to Dr. Schrank for his help throughout this research. Special thanks go to the Maryland Department of Transportation's State Highway Administration and the Texas Department of Transportation's Transportation Planning and Programming Division for giving permission to use these invaluable freight datasets.

Thanks to Dr. Lingtao Wu for his helpful comments and suggestions in the research methods, and for sharing his expertise in R coding process.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis committee consisting of Dr. Zhang and Dr. Cline from the Department of Statistics and Dr. Eisele of the Texas A&M Transportation Institute. The data used in this research are from Maryland Department of Transportation's State Highway Administration and the Texas Department of Transportation's Transportation Planning and Programming Division. Dr. Eisele helped to get the permission of using the data from Maryland. Dr. Schrank from the mobility group of Texas A&M Transportation Institute helped to get the permission of using the data from Texas. Dr. Cline provided supports in the statistical analysis part. All other work conducted for the thesis was completed by the student independently.

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

NOMENCLATURE

DFW	Dallas–Fort Worth
EE	External to external geospatial type
EI	External to internal geospatial type
GPS	Global positioning system
IE	Internal to external geospatial type
II	Internal to internal geospatial type
OD	Origin-destination
PTI	Planning time index
PTIRATIO	Ratio of PTI on different routes
SP	Stated preference
TTI	Travel time index
TTIRATIO	Ratio of TTI on different routes
Waypoints	Geospatial points recorded by global positioning system devices

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

Today, more and more researchers have started to notice the major role trucks play in people's lives. Especially in the United States, many goods are transported via trucks, which make a significant contribution to congestion and environmental issues. Better managing the transportation system and addressing congestion issues require a comprehensive understanding of traffic participants' characteristics, including the routing behavior of freight vehicles.

Studies related to the routing behavior of passenger cars are extensive. Most of the existing routing behavior studies applied stated-preference (SP) methods, which mainly collected data through experiments. It is rare to find routing behavior studies that use empirical data, and studies of truck drivers' routing behavior are even rarer. The reason is that traffic data from freight companies are usually kept confidential.

Existing literature has revealed a wide range of influential factors that could have impacts on the routing decision-making process. The majority of researchers put their efforts into using SP or simulation methods to approach the issue. With the development of advanced navigation technologies, more affordable navigation units could generate a greater impact on drivers' routing decisions. However, the drivers' rate of compliance with routing recommendations provided by these navigation tools is unknown. This psychological self-selection process becomes an insurmountable obstacle for SP and other experimental methods because critics could question if these imaginary responses from participants in the laboratory reflect their real selections under actual conditions.

With empirical data, the compliance rate will not be an impediment for researchers to study travel behavior because what the data show is the real decisions made by those drivers under real conditions. Detailed freight traffic information is extremely hard to access due to collection difficulties arising from freight companies' labeling of the data as highly classified. This study may be the first research, to the best knowledge of the author, in academic and practical studies on detailed disaggregate freight data with geospatial information.

This study analyzes empirical traffic data from Maryland and Dallas–Fort Worth (DFW), Texas. A significant portion of the data were collected through embedded global positioning system (GPS) devices in fleet vehicles. Therefore, these two datasets could fulfill the purpose of studying the routing behavior of fleet drivers.

This research poses five research questions:

1. How do different levels of congestion on two similar routes affect truck drivers' routing decision-making process?
2. How does travel time reliability impact routing decisions?
3. How do different geospatial-type trips influence routing selection?
4. How do morning and afternoon rush hours affect routing decisions?
5. How do weekdays versus weekends influence the routing preferences of truck drivers?

2. LITERATURE REVIEW

Understanding routing behavior is critical for transportation planning and operation, and numerous studies have been done in this area. Researchers show wide agreement that the influential factors that could affect the routing decision-making process include the availability of alternate routes, travel time, travel time reliability, the level of congestion, and concern for safety (*1*-*3*).

The previous studies of truck drivers' routing behavior mostly applied research methods such as SP method and simulations because of the limited accessibility of real freight routing data (*4*). Researchers have had difficulty obtaining detailed disaggregate freight data because freight companies do not want to disclose this information to their competitors (*5*). This barrier results in a gap in the analysis of the relationship between truck drivers' routing decision and congestion, travel time reliability, rush hours, and other factors.

In 1993, Bitzios and Ferreira noticed that passenger travel is markedly different from freight travel. However, they were not able to prove this finding through a convincingly statistical approach (*6*). The first paper regarding the analysis of detailed freight data and proving this theory was published in 2001. The author, for the first time, revealed that the travel time distribution of commercial vehicles is different from that of passenger vehicles. The analysis of variance results demonstrated that the coefficient of variation of the travel time distribution of passenger cars is relatively constant. However, the distribution of commercial vehicles could vary significantly as a result of the time of day

(*7*). The particular features of goods transportation make the routing behavior of commercial vehicles different from that of passenger cars. For instance, the drivers of light trucks are more concerned about congestion than heavy-truck drivers are, and heavy-truck drivers are more sensitive to the hierarchy of the road than light-truck drivers are (*8*).

Figliozzi et al. summarize the characteristics of the freight routes and discuss the relationship between travel speed and distance. The authors also show that the morning rush hour period exhibits the highest level of congestion (*5*). In 2004, Hunt and Abraham found the probability of delay has a significant impact on the routing behavior of commercial vehicle drivers. This research proved commercial vehicle drivers are more sensitive to the length of delay rather than total driving time. In other words, the study shows the travel time reliability affects the routing decision-making process (*2*).

Routing decision-making processes have been investigated over the years. The popular methods used in research are SP, experimental, and theoretical modeling and simulation (*9*). Theoretical modeling and simulation methods will not be the focus of this literature review.

The SP survey is one of the most frequently used methods. Most of the literature mentioned in this review uses this method. This method is easy to control because the researcher can control all the conditions and make all the assumptions before the participants are even involved.

With this method, researchers demonstrated that the travel time reliability and variability of the travel time could influence the routing selection behavior. Receiving traffic information could shape the routing decision-making process. However, the effect of travel time variability on the path choice varies across individuals (*10*). Eisele et al. used a nonparametric local regression (LOESS) statistical procedure to prove that the travel time distribution characteristics between passenger vehicles and commercial vehicles are different, and that commercial vehicles show more travel time variability than passenger vehicles do (*7*).

With the continual advance of technological development, real-time traffic information is becoming more and more accessible to drivers. In contrast to previous routing decision-making processes, which were mainly based on experience, the locus of current routing decision-making is a combination of experience and real-time traffic information. The compliance rate of navigation systems can differ significantly between drivers and different times of day. For example, the compliance rate is higher when the recommended route choice serves the driver's interests rather when it is arbitrary (*11*). With increasing accessibility of real-time traffic information and navigation software, routing decision-making behaviors are changing. Golob and Regan found the level of local congestion is positively related to the purchase of routing software (*12*). Researchers have realized the importance of knowing the driver's compliance rate with real-time traffic information. However, the main method used by researchers is still SP. This approach, however, is questionable. The main criticism of this method is that the participants might not act the same way in reality as they claim in the experiment (*13*).

The lack of support from the empirical data makes such research results questionable from the start.

3. METHODOLOGY

3.1 Raw Data Description and Site Selection

The Maryland Department of Transportation's State Highway Administration provided the Maryland INRIX origin-destination (OD) datasets. INRIX collected 400 gigabytes of data during February, June, July, and October 2015, which include 19,690,402 trips and 1,376,720,203 waypoints in Maryland, Virginia, and Washington, D.C. From the whole dataset, 60 percent of trips came from fleet resources, 31 percent came from consumer resources (auto clients), and 9 percent came from mobile resources (mobile devices). About 77 percent of trips happened internally within Maryland.

These detailed disaggregate traffic data are composed of trip information and recorded waypoints with geospatial information. Trip information includes data provider type, device ID, geospatial type, provider driving profile, vehicle weight class, etc. Waypoint data provide geospatial information with longitude and latitude and the capture time of each waypoint.

Two types of data collection approaches were used: an embedded GPS and a mobile application. Three data provider types are listed in the datasets, including consumer, fleet, and mobile. The consumer is mainly auto clients such as BMW, which collects data through an embedded GPS in its connected vehicles. Data gathered via mobile applications are in the consumer category as well. However, exact information about whether these mobile users are passenger cars or trucks is not available. A strong assumption is that the majority of the mobile map application users are passenger vehicles because most fleet vehicles have their particular navigation systems. All fleet data are gathered through freight providers. Thus, fleet data are collected from reliable data sources, which could produce reliable analysis results.

3.1.1 Research Area Selection

The purpose of this research is to find out if real-time traffic information, travel time reliability, and other factors could affect the routing selection. To test these factors' impacts, it is important first to eliminate other factors that could largely influence routing selection such as tolls, different travel distances, and various hierarchies of routes. Two identical or similar routes are necessary for this study. The length and function class of the research routes should be similar, and both should be toll-free. The study tries to evaluate the impact of the real-time traffic situation on routing behavior. Using a proper length of routes is critical because obtaining traffic information for a long length could be problematic (*14*). Therefore, the length of the selected routes is a constraint.

The I-495 east loop and I-495 north loop were selected to fulfill these requirements. They provide a sufficient number of observations. In Figure 1, from area A to B, the distance is 34.7 miles for the north loop and 38.3 miles for the east loop. The estimated driving time during free-flow speeds is 36 minutes for the north loop and 40 minutes for the east loop. The coordinates of the selected areas are −76.941, 39.043, −76.932,39.036 for area A and −77.183,38.788, −77.175,38.782 for area B. The trips passing areas A and B are filtered from the original dataset for research purposes. The exported data are divided into two datasets based on the directional information: northbound trips and southbound trips. Trips that passed area A first and then passed area B are in the southbound dataset, and trips that passed area B first and then passed area A are in the northbound dataset.

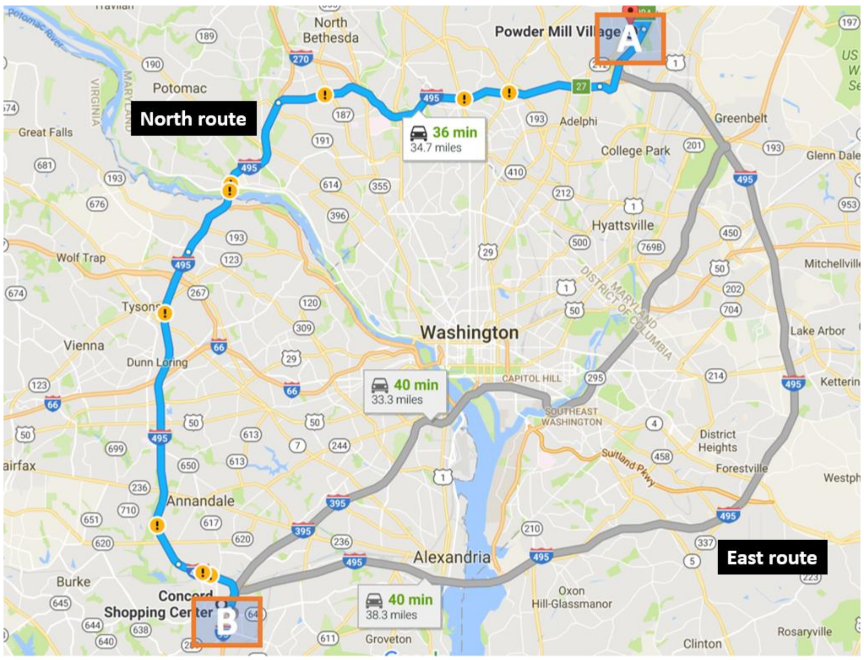

Figure 1 Selected research area in Maryland

An important step in the data cleaning process is eliminating trips passing areas A and B but going through downtown Washington, D.C., instead of using I-495. These data are not considered because they do not contribute to the study of behaviors on the two similar routes. Only a very small portion of trips recorded in this dataset pass downtown Washington, D.C. Another important step is excluding delivery trips. Delivery trips could have a strong influence on the research results because their routing selection is based only on the locations of their delivery stops. After exclusion of trips that did not travel entirely on the I-495 loop and trips that involved local delivery, the dataset is ready for further data analysis.

3.2 Two Datasets and Descriptive Data Summary

In the southbound and northbound datasets, the shares of trip information categories (such as provider type and geospatial type) are dramatically distinctive from their shares in the initial dataset.

After data cleaning, the southbound dataset consists of 2,157,488 waypoint records and 9,276 trips in the four months. The northbound dataset has 1,550,912 waypoint records and 7,775 trips in the four months.

Figure 2 shows that about 83 percent of trips are fleet trips; 13 percent of those trips are from mobile resources, and only 4 percent are from consumer resources.

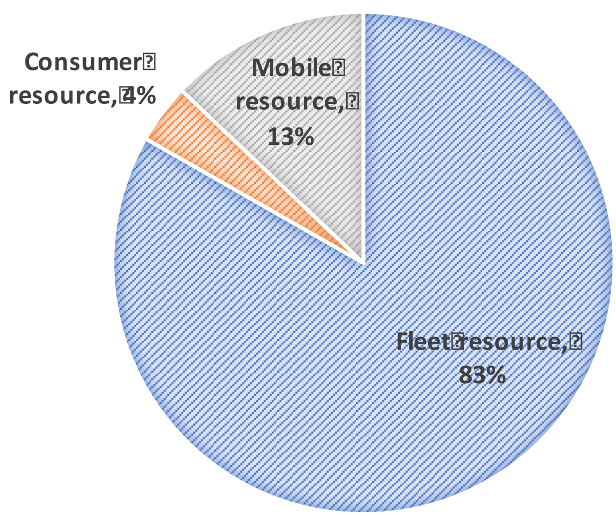

Figure 2 Proportion of the data resources in the selected research area in the Maryland dataset

3.2.1 Geospatial Type

Trips were classified by where they began and ended:

- Trips that began internal to Maryland and ended internal to Maryland (II).
- Trips that began external to Maryland and ended external to Maryland (EE).
- Trips that began internal to Maryland and ended external to Maryland (IE).
- Trips that began external to Maryland and ended internal to Maryland (EI).

Table 1 shows that in the original data about 77 percent of trips are II, and only 3 percent are EE. In the southbound dataset, 67 percent of trips are IE, about 30 percent are EE, and almost 0 percent are EI or II. In the northbound dataset, 70 percent of trips are EI, 30 percent are EE, and almost 0 percent are IE or II.

Table 1 also shows that the percentages of geospatial types in the two sub-datasets (northbound and southbound) are different from those in the original data. The reason for this is that the boundary of Maryland cuts halfway through the I-495 loop from northwest to southeast. According to the definitions of the geospatial types, the trips in the southbound dataset represent trips that pass area A first and then pass area B. Area A is inside Maryland, and area B is outside Maryland. This situation leads to all trips in the southbound dataset ending outside Maryland, which is IE. For the same reason, all trips in the northbound dataset start outside Maryland and end inside Maryland, which is EI.

Table 1 Statistic summary of the geospatial characteristics of trips in the selected Maryland dataset

Geospatial Type	Original Data	Southbound	Northbound
EE	722,678	3,067	2,239
	3%	33%	30%
EI	1,906,726	3	5,205
	10%	0%	70%
IE	1,923,633	6203	0
	10%	67%	0%
II	15,137,365	3	2
	77%	0%	0%
Sum	19,690,402	9,276	7,446

Note: EE = external to external, EI = external to internal, IE = internal to external, II = internal to internal.

3.2.2 Vehicle Type

A summary of the trip by provider driving profile shows 46 percent of trips are field service/local delivery fleets, 31 percent of total trips are consumer vehicles, and 22 percent are for-hire/private truck fleets. As seen in Table 2, in the two sub-datasets, a majority of those trips passing area A and area B are for-hire or private truck fleets. This is because heavy-duty transport relies on trucking fleets.

Table 2 Statistic summary of the provider type profile characteristics of trips in the selected Maryland area

Provider Driving Profile	Original Data	Southbound	Northbound
Consumer vehicle	6,155,314	1,354	1,136
	31%	15%	15%
Taxi/shuttle/town car service fleets	145,053	30	16
	1%	0%	0%
Field service/local delivery fleets	9,075,413	1,037	907
	46%	11%	12%
For-hire/private trucking fleets	4,314,622	6,855	5,387
	22%	74%	72%
Sum	19,690,402	9,276	7,446

3.2.3 Trip Density

The peak of the trip density distribution of the northbound dataset is lower than that of the southbound dataset. Figure 3 shows that the trip density peaks around 2 p.m. for both directions and then diminishes gradually. The density reaches its lowest point at midnight. The two curves in Figure 3 illustrate that truck drivers are inclined to avoid rush hours.

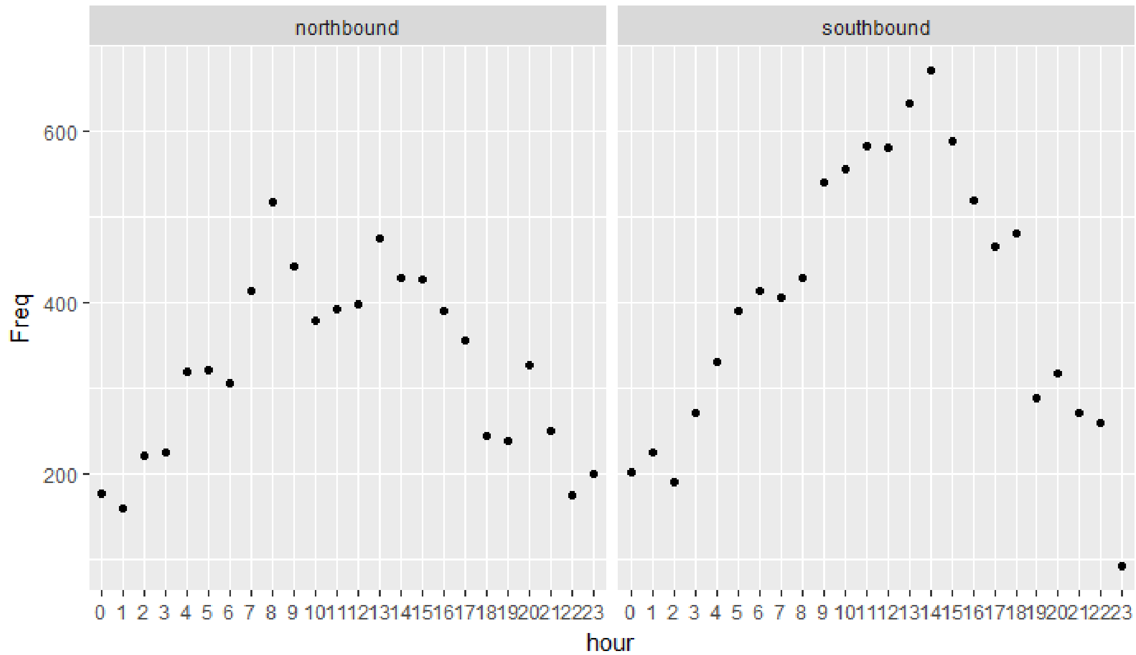

Figure 3 Hourly fleet trip distribution based on direction in the Maryland dataset

3.3 Methodology

For the purpose of this research, fleet trips are the main concern for two reasons: 83 percent of the data are fleet trip data that could provide a substantial number of observations, and the fleet data resources are more reliable than consumer and mobile data resources.

This research focuses on two critical factors, the ratio of the congestion level on each route (travel time index ratio [TTIRATIO]) and the ratio of travel time reliability on each route (planning time index ratio [PTIRATIO]). These two factors could have significant contributions to the routing decision process. Based on previous studies, the geospatial type and morning and afternoon rush hours are expected to have significant impacts on routing decisions. The routing behavior could act differently during weekdays and weekends in response to different traffic patterns.

The routing decision for a fleet driver to decide to travel on the north loop or the east loop is the binary discrete response. Thus, the binary response logistic regression is suitable to identify a relationship between routing choices and independent variables.

3.3.1 TTIRATIO

Transportation researchers and engineers consider the travel time index (TTI) an indicator of the level of congestion. The common approach to calculating the TTI is to use the average travel time divided by the free-flow travel time.

To compare the congestion level on the two loops, this research uses the TTIRATIO to represent the different congestion conditions on both routes. For research consistency, the TTIRATIO is always defined as the TTI of the north loop over the TTI of the east loop:

$$\text{TTI} = \frac{mean\ of\ travel\ time\ of\ all\ observations\ during\ past\ one\ hour\ on\ selected\ route}{free\ flow\ travel\ time\ on\ this\ route}$$

$$\text{TTIRATIO} = \frac{TTI\ of\ north\ loop}{TTI\ of\ east\ loop} =$$

$$\frac{mean\ of\ travel\ time\ on\ north\ loop \big/ free\ flow\ travel\ time\ on\ north\ loop}{mean\ of\ travel\ time\ on\ east\ loop \big/ free\ flow\ travel\ time\ on\ east\ loop}$$

$$= \frac{mean\ of\ travel\ time\ on\ north\ loop * free\ flow\ travel\ time\ on\ east\ loop}{means\ of\ travel\ time\ on\ east\ loop * free\ flow\ travel\ time\ on\ north\ loop}$$

For both directional datasets, the boxplot in Figure 4 states that the majority of the TTIRATIO is concentrated between 1.0 and 1.4. Some extreme values do exist in the dataset.

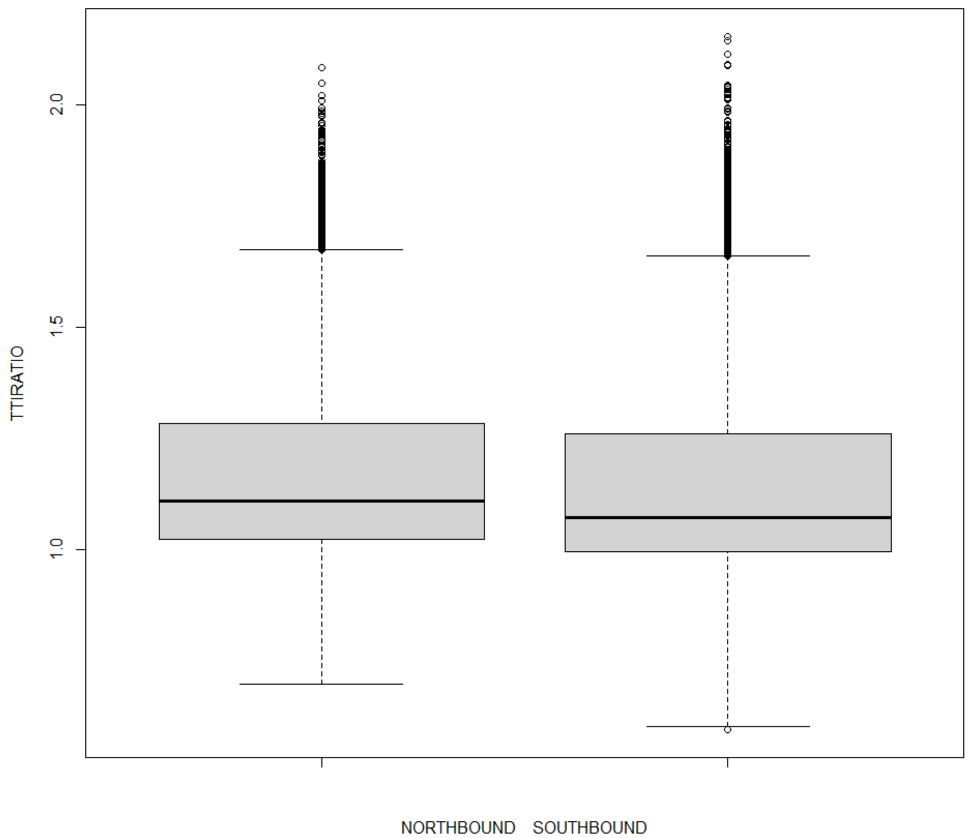

Figure 4 Boxplot of the TTIRATIO in the Maryland dataset

3.3.2 PTIRATIO

The planning time index (PTI) is a tool that measures travel time reliability. From Federal Highway Administration operations, the simplest method to measure this travel time reliability is to use the 90th or 95th percentile peak-period travel time divided by the free-flow travel time.

To compare the travel time reliability of the two routes, the PTIRATIO measures the difference in travel time reliability between the north loop and east loop. The PTIRATIO is defined as the PTI of the north loop divided by the PTI of the east loop:

PTI =

$$\frac{95\text{th}\ percentile\ of\ travel\ time\ of\ all\ observations\ during\ past\ one\ hour\ on\ selected\ route}{free\ flow\ travel\ time\ on\ this\ route}$$

PTIRATIO =

$$\frac{PTI\ of\ north\ loop}{PTI\ of\ south\ loop} =$$

$$\frac{95\ percentile\ of\ peak\ period\ travel\ time\ of\ north\ loop \Big/ free\ flow\ travel\ time\ of\ north\ loop}{95\ percentile\ of\ peak\ period\ travel\ time\ of\ east\ loop \Big/ free\ flow\ travel\ time\ of\ east\ loop}$$

$$= \frac{95\text{th}\ percentile\ speed\ of\ north\ loop * free\ flow\ travel\ time\ of\ east\ loop}{95\ percentile\ of\ peak\ period\ travel\ time\ of\ east\ loop * free\ flow\ travel\ time\ of\ north\ loop}$$

The PTIRATIO boxplot in Figure 5 shows that most of the values are between 1.0 and 1.4. The southbound dataset has more variation than the northbound dataset.

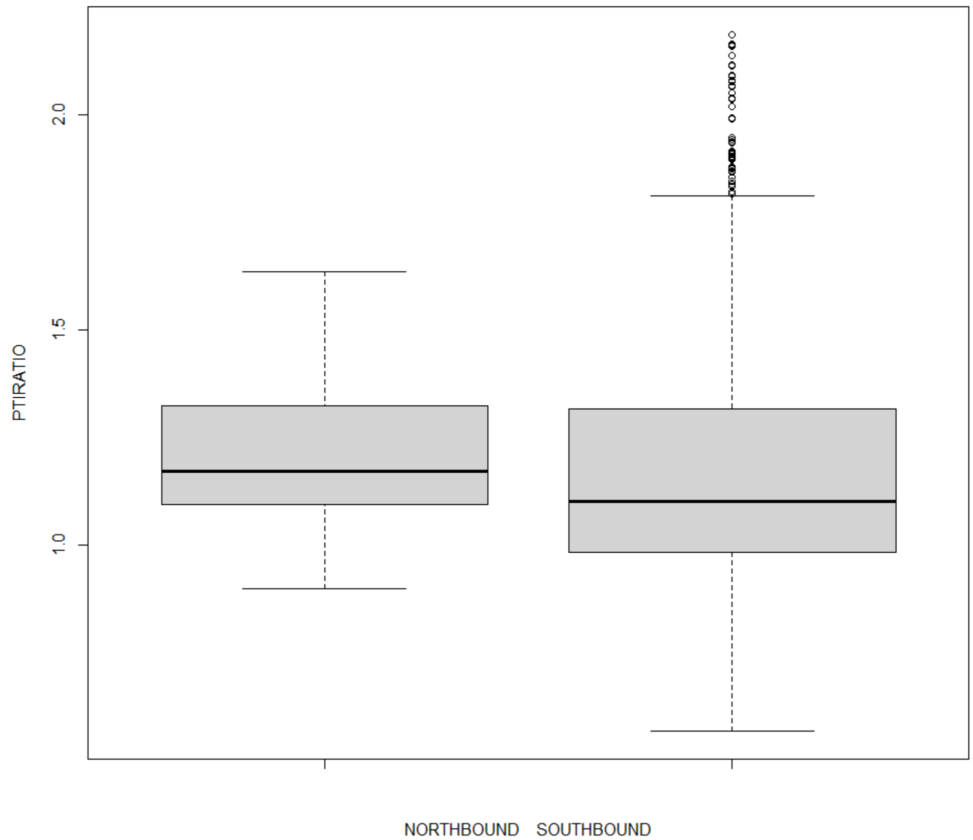

Figure 5 Boxplot of the PTIRATIO in the Maryland dataset

3.3.3 Logistic Model

Logistic regression is a regression model that can establish a relationship between discrete results with given various independent variables (*15; 16*). In this case, the binary response (0 and 1) represents the drivers' routing choices (1 = north loop, 0 = east loop). Table 3 illustrated all variable that will be applied in this logistic model. In the logistic regression model, the log odds of the response are modeled as a linear function of these explanatory variables:

$$Logit(p_i) = \log\left[\frac{p_i}{1 - p_i}\right] = \alpha + \beta_i X_i$$

p_i = probability of the response with given explanatory variables

α = intercept parameter

$\frac{p_i}{1 - p_i} = \text{odds}$

β_i = coefficient parameter of the vector of the explanatory variable X_i, which represents how much the explanatory variable can impact the log odds

X_i = explanatory variables (*17*)

Table 3 Variables considered in the logit model

<table>
<tr><th colspan="3">Variable</th><th>Type</th><th>Explanation</th></tr>
<tr><td>Dependent variable</td><td>Y</td><td>Route selected</td><td>Dichotomous</td><td>Y = 1 if observed vehicle selected north loop;
Y = 0 if observed vehicle selected east loop</td></tr>
<tr><td rowspan="6">Independen t variable</td><td>X1</td><td>TTIRATIO</td><td>Continuous</td><td>Calculated by definition</td></tr>
<tr><td>X2</td><td>PTIRATIO</td><td>Continuous</td><td>Calculated by definition</td></tr>
<tr><td>X3</td><td>Geospatial type</td><td>Dichotomous</td><td>X4 = 1 if the geospatial type of the trip is EE;
X4 = 0 if the geospatial type of the trip is IE in the southbound dataset;
X4 = 0 if the geospatial type of the trip is EI in the northbound dataset</td></tr>
<tr><td>X4</td><td>Off-rush hour</td><td rowspan="2">Nominal</td><td rowspan="2">X4 = 0 if the observation happened during off-rush hour;
X5 = 1 if the observation happened during morning rush hour (6:00 a.m. to 9:30 a.m.);
Set the afternoon rush hours = 2 as the reference category (3:30 p.m. to 6:30 p.m.)</td></tr>
<tr><td>X5</td><td>Morning rush hour</td></tr>
<tr><td>X6</td><td>Weekday</td><td>Dichotomous</td><td>X8 = 1 if the day is from Monday to Friday;
Set the weekend (Saturday or Sunday) = 0 as the reference category</td></tr>
</table>

4. RESULTS

Binary response logistic regression has been applied to both datasets. Table 4 and Table 5 show the estimated coefficients of those predictors with their corresponding odds ratios.

Each directional dataset has been subdivided into three sub-datasets according to their provider types: consumer, fleet, and mobile. Therefore, for each travel direction (southbound or northbound), there are four datasets including the one with all trips. Model 2, which is the regression with consumer trip dataset, is not discussed here because it only accounts for less than a 4 percent share of the total trips in each directional dataset. All models have the same predictors. All independent variables used in this regression model are presented in the tables, including those variables with a significant level lower than 95 percent. The odds ratio is a good indication of the expected change in the log odds of the response variable if the value of the independent variable increases by one unit.

Table 4 Logistic regression results for the Maryland southbound dataset

	Dependent variable:						
	total trips	consumer trips	fleet trips	mobile trips	consumer trips	fleet trips	mobile trips
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TTIRATIO	-0.821 p=0.000***	-1.271 p=0.006***	-0.829 p=0.000***	-0.780 p=0.007***			
PTIRATIO					-0.289 p=0.270	-0.296 p=0.003***	-0.368 p=0.072*
GEOSPATIAL(IE)	0.106 p=0.0125***	-0.219 p=0.191	0.117 p=0.0135**	0.063 p=0.300	-0.047 p=0.423	0.120 p=0.0115**	0.086 p=0.236
OFF RUSH HOUR	-0.079 p=0.141	-0.494 p=0.092*	-0.162 p=0.0225**	0.436 p=0.026**			
MORNING RUSH HOUR	-0.303 p=0.001***	-1.086 p=0.0085***	-0.250 p=0.0065***	-0.453 p=0.069*			
WEEKDAY	-0.123 p=0.0185**	-0.028 p=0.462	-0.119 p=0.044**	0.302 p=0.0095**			
CONSTANT	0.245 p=0.065*	1.342 p=0.0475**	0.188 p=0.152	0.111 p=0.400	-0.374 p=0.254	-0.656 p=0.00000***	0.193 p=0.258
Observations	9,696	370	8,073	1,253	370	8,073	1,253
Log Likelihood	-5,881.343	-226.876	-4,728.133	-846.178	-233.660	-4,756.404	-860.963
Akaike Inf. Crit.	11,774.690	465.751	9,468.267	1,704.355	473.321	9,518.808	1,727.927
Note:							*p<0.1; **p<0.05; ***p<0.01

Maryland Southbound Logistic regression results
===
all

Maryland Southbound Logistic regression results
===================================
significent test are one-side tests

Table 5 Logistic regression results for the Maryland northbound dataset

	Dependent variable:						
	total trips (1)	consumer trips (2)	fleet trips (3)	mobile trips (4)	consumer trips (5)	fleet trips (6)	mobile trips (7)
TTIRATIO	-0.482	-0.749	-0.399	-1.016			
	p=0.0001***	p=0.137	p=0.0035***	p=0.001***			
PTIRATIO					-0.684	-0.091	-0.179
					p=0.247	p=0.329	p=0.343
GEOSPATIAL(EI)	0.191	0.372	0.293	-0.231	0.347	0.305	-0.235
	p=0.001***	p=0.116	p=0.00001***	p=0.074*	p=0.131	p=0.000005***	p=0.068*
OFF RUSH HOUR	0.026	0.026	-0.087	0.606			
	p=0.409	p=0.483	p=0.243	p=0.0245**			
MORNING RUSH HOUR	-0.175	-0.284	-0.223	0.061			
	p=0.0835*	p=0.348	p=0.05**	p=0.437			
WEEKDAY	-0.325	-0.615	-0.245	-0.097			
	p=0.000005***	p=0.0375*	p=0.002***	p=0.258			
CONSTANT	-0.586	-0.318	-0.768	-0.028	-0.817	-1.426	-0.533
	p=0.00175***	p=0.392	p=0.001***	p=0.480	p=0.251	p=0.00000***	p=0.164
Observations	7,775	305	6,465	1,005	305	6,465	1,005
Log Likelihood	-3,925.983	-140.229	-3,139.843	-606.902	-142.988	-3,151.193	-618.945
Akaike Inf. Crit.	7,863.967	292.457	6,291.685	1,225.803	291.977	6,308.386	1,243.890
Note:							*p<0.1; **p<0.05; ***p<0.01

Maryland Northbound Logistic regression results

significant test are one-side tests

4.1 Influential Factors

4.1.1 Constant

The estimated coefficient of the intercept is the log odds of a driver choosing to travel on the north loop with all continuous variables held at the hypothetical value of zero, and all categorical variables are at the reference level.

4.1.2 TTIRATIO

The TTIRATIO, which measures the congestion condition on both routes, has a negative influence on the decision-making process. The result shows that the odds of selecting the north route significantly decrease if the TTIRATIO between the north route and east route increases. It is reasonable to switch to the alternative route when drivers find another one more congested. The coefficients of every model shown in Table 4 and Table 5 are highly significant. This research focuses on explaining the results from the fleet trip dataset.

As mentioned in the variable description, Figure 4 shows that most of the TTIRATIO values are between 1.0 and 1.4. The traditional approach to explain the coefficient of the continuous variable is that each estimated coefficient represents the expected change in the log odds ratio of driving on the north loop with one unit increase in the TTIRATIO. However, in this case, increasing one unit in the TTIRATIO is not meaningful. In this research, the expected change in the odds ratio in using the north loop is calculated using a 0.1-unit increment in the corresponding value.

In the fleet trip dataset:

The odds ratio $OR(TTIRATIO_S) = \exp(-0.0829) = 0.92 (p = 0.000^{***})$

$OR(TTIRATIO_N) = \exp(-0.0399) = 0.96 (p = 0.0035^{***})$

A 0.1-unit increase of the TTIRATIO means the average travel time on the north loop is 10 percent greater than the average travel time on the east loop. There is an 8 percent decrease in the odds of choosing the north loop in the southbound direction and more than a 4 percent decrease in the odds of using the north loop in the northbound direction. Drivers are making a rational decision on routing choice because switching to the alternative route uses less travel time.

4.1.3 Geospatial Type

This research sets the EE geospatial type as the reference group. The regression results show the geospatial type is a significant factor that affects routing selection. The following odds ratio shows that the odds of IE or EI trips using the north route are 12 to 34 percent greater than that of EE trips. Trips starting and ending outside Maryland prefer the north loop.

In the fleet trip dataset:

The odds ratio $OR(GEOSPATIAL_{IE\_vs\_EE\_S}) = \exp(0.117) = 1.12 (p = 0.0135^{**})$

$OR(GEOSPATIAL_{EI\_vs\_EE\_N}) = \exp(0.293) = 1.34 (p = 0.00001^{***})$

4.1.4 Rush Hour

Rush hour is a nominal variable with three unordered levels. The prediction model has two dummy variables: off-rush hour and morning rush hour. Afternoon rush hour is defined as the reference group. The coefficient of the off-rush hour is the log of the odds ratio between the off-rush hour and the afternoon rush hour. The coefficient of morning rush hour is the log of the odds ratio between the morning rush hour and afternoon rush hour, with other variables maintaining certain values.

In the fleet trip dataset:

The odds ratio $OR(\text{RUSH HOUR}_{\text{off-rush hour\_vs\_afternoon rush hour\_S}})$

$= \exp(-0.162) = 0.85(p = 0.0225^{**})$

$OR(\text{RUSH HOUR}_{\text{off-rush hour\_vs\_afternoon rush hour\_N}})$

$= \exp(-0.087) = 0.92(p = 0.243)$

$OR(\text{RUSH HOUR}_{\text{morning rush hour\_vs\_afternoon rush hour\_S}})$

$= \exp(-0.250) = 0.78(p = 0.0065^{***})$

$OR(\text{RUSH HOUR}_{\text{morning rush hour\_vs\_afternoon rush hour\_N}})$

$= \exp(-0.223) = 0.80(p = 0.05^{**})$

The odds ratios of the off-rush hour versus afternoon rush hour in both directions vary slightly from each other. The odds ratios of the morning rush hour versus afternoon rush hour in both directions do not change much as well. The four odds ratios are all less than 1. This pattern indicates the odds (the percentage of fleet drivers driving on the north loop over the percentage of fleet drivers driving on the east loop) of the off-rush hour and the odds of the morning rush hour are less than the odds of the afternoon rush hour by about 10 and 25 percent, respectively. First, the two dummy variables have significant effects on the routing decision except for the off-rush hour in the northbound direction. Second, compared to the afternoon rush hour, truck drivers prefer the east loop in the off-rush hour and morning rush hour.

However, the off-rush hour is only significant in the southbound direction and not significant in the northbound direction. The possible explanation could be glare effects or road construction during the data collection months.

In the mobile trip dataset:

The odds ratio $OR(\text{RUSH HOUR}_{\text{off-rush hour\_vs\_afternoon rush hour\_S}})$

$= \exp(0.436) = 1.55 (p = 0.026^{**})$

$OR(\text{RUSH HOUR}_{\text{off-rush hour\_vs\_afternoon rush hour\_N}})$

$= \exp(0.606) = 1.83 (p = 0.0245^{**})$

$OR(\text{RUSH HOUR}_{\text{morning rush hour\_vs\_afternoon rush hour\_S}})$

$= \exp(-0.453) = 0.64 (p = 0.069^{*})$

$OR(\text{RUSH HOUR}_{\text{morning rush hour\_vs\_afternoon rush hour\_N}})$

$$= \exp(0.061) = 1.06 (p = 0.437)$$

The odds ratio pattern shows that in the fleet trip dataset, the odds ratio of those rush hours in the mobile trip dataset varies significantly. In the southbound direction, the odds of the off-rush hour are 55 percent greater than the odds of the afternoon rush hour. In the other direction, the odds of the morning rush hour are greater than the odds of the afternoon rush hour by more than 70 percent. Moreover, they are statistically significant. In other words, drivers are more likely to choose the north loop than the east loop in both directions during the off-rush hour than in the afternoon rush hour. Interestingly, the preference of routing selection does not show significant distinction between the two rush periods. The mobile trips have the same routing behavior between the morning rush hour and afternoon rush hour.

By comparing the results from the two different datasets, the difference could have been caused by the various navigation systems used by drivers. Even with the same traffic situation, algorithms of travel time prediction could be different with different navigation tools, even their predicted travel time may very close. This could lead to different routing suggestions. Another explanation is the various drivers' rate of compliance with the navigation system recommendations.

4.1.5 Weekday

The binary variable weekday has a significant impact on the preference of route selected by fleet drivers. The weekend is defined as the reference group (weekend = 0).

In the fleet trip dataset:

The odds ratio $OR(WEEKDAY_{weekday\_vs\_weekend\_S}) = \exp(-0.119)$

$= 0.88(p = 0.044^{**})$

$OR(WEEKDAY_{weekday\_vs\_weekend\_N}) = \exp(-0.245) = 0.78(p = 0.002^{***})$

The estimated coefficients indicate that fleet drivers have less preference for the north loop on weekdays compared to on weekends. The odds on weekdays are 10 to 20 percent less than the odds on weekends.

4.1.6 PTIRATIO

The PTIRATIO measures the travel time reliability. The reason for excluding this variable from previous models is multicollinearity. Previous research found that the estimated coefficient of this variable is positive, which means that an increase in the PTIRATIO attracts more drivers to use the north loop. In other words, more drivers prefer to use the north loop when it shows less reliability. This is because of the collinearity problem between the PTIRATIO and other variables. Figure 6 illustrates the curve of the PTIRATIO. This multimodal curve arrives at its peaks during rush hours. This could lead to collinearity between the PTIRATIO and rush hour.

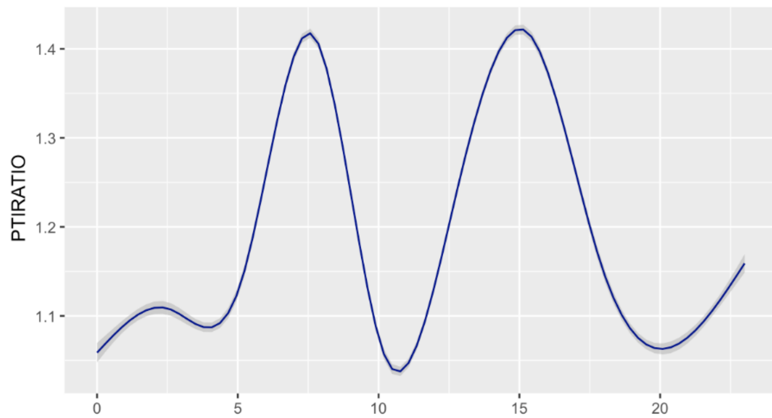

Figure 6 PTIRATIO distribution in the Maryland dataset

However, the coefficients in model 6 are negative values when the regression model does not include those highly correlated variables. A 0.1-unit increase in the PTIRATIO is used to explain its corresponding odds ratio. The reason is the same as why a 0.1-unit increase is used in the TTIRATIO. Figure 5 shows the majority of the PTIRATIO is between 1.0 and 1.4.

In the fleet trip dataset:

The odds ratio $OR(PTIRATIO_S) = \exp(-0.0296) = 0.97(p = 0.003^{***})$

$OR(PTIRATIO_N) = \exp(-0.0091) = 0.99(p = 0.329)$

In the mobile trip dataset:

The odds ratio $OR(PTIRATIO_S) = \exp(-0.0368) = 0.96(p = 0.072^{*})$

$OR(PTIRATIO_N) = \exp(-0.0179) = 0.98(p = 0.343)$

Every 0.1-unit increase of the PTIRATIO brings a 3 and 1 percent decrease in the odds of using the north loop in the southbound and northbound directions. Based on the statistical results, southbound drivers seem to have slightly more concern about travel time reliability compared to northbound drivers. The results from mobile trips have no significant impact on routing selection as well. Comparing the results by data provider type, travel time reliability does have a negative impact on the routing selection process. However, those impacts are not a big concern for passenger vehicle drivers. Some fleet drivers consider travel time reliability a critical evaluator during routing selection. Having a sense of travel time reliability on particular routes requires experience. Normally, only frequent drivers on the particular routes have anticipation of travel time reliability. This prerequisite could be the reason why the regression does not have sufficient data to reject the null hypothesis.

5. DFW OD DATASET

5.1 Data Description and Research Area Selection

The Texas Department of Transportation's Transportation Planning and Programming Division provided the DFW INRIX OD datasets.

INRIX collected 340 gigabytes of data from January to April in 2016, which includes 29,138,492 trips and 1,608,100,000 waypoints in the DFW area. Unlike the Maryland dataset, this dataset has no data collected from mobile resources. The fleet resources are the main portion of the dataset.

As shown in the Figure 7, a part of I-30 and I-35E and a part of I-30 were selected as the research area. The north point is defined as the conjunction of I-30 and I-20 east of Dallas, and the south point is defined as the conjunction of I-35E and I-20 south of Dallas. The north route is a part of I-30 and I-35E, and goes through downtown Dallas. The east route is a part of I-20. Their lengths are about 22.5 miles and 23.9 miles, respectively. For data filtering purposes, the coordinates of the north point are −96.812, 32.934, −96.774,32.911, and the coordinates for the south point is −96.726,32.645, −96.659,32.602.

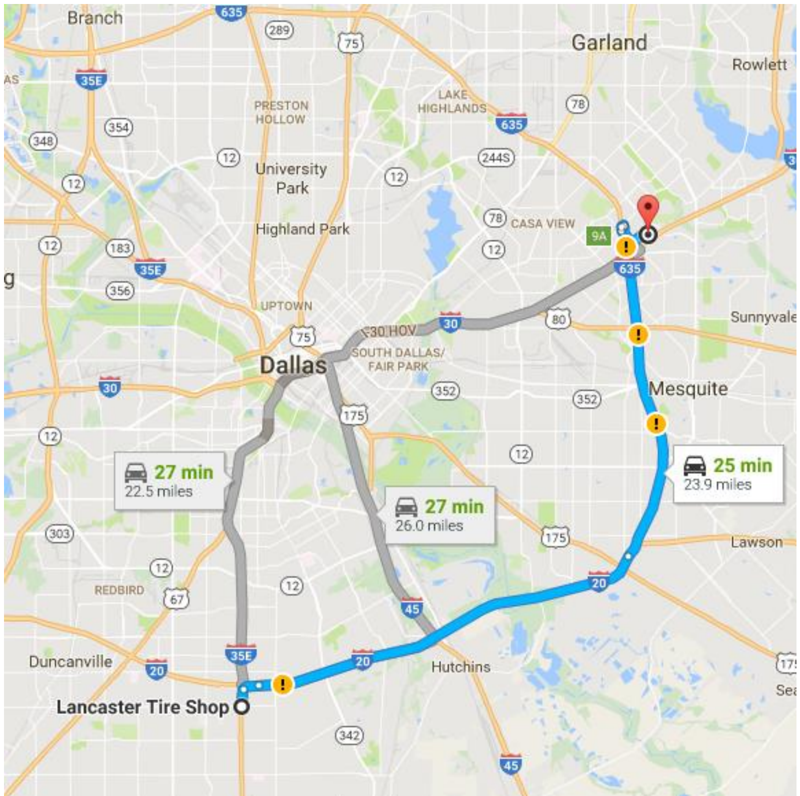

Figure 7 Selected research area in Dallas

5.2 Two Datasets and Descriptive Data Summary

The data cleaning process for the DFW dataset is similar to that for the Maryland dataset. Two directional datasets were obtained from the original dataset.

The southbound dataset has 94,829 waypoints and 1,212 trips. The northbound dataset has 105,384 waypoints and 1,324 trips. Figure 8 shows that, in both datasets, 85 percent of the data were collected by fleet resources, and 15 percent of the data were collected by consumer resources.

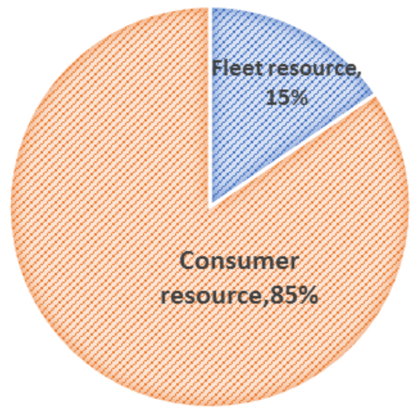

Figure 8 Proportion of the data resources in the selected research area in the DFW dataset

5.2.1 Geospatial Type

In Table 6, both datasets show that almost half of the trips are II trips. Ten percent of those trips start and end outside Texas. In the southbound dataset, 10 percent of the trips start from outside Texas and end inside Texas. More than 30 percent of trips are IE. In the northbound dataset, more than 30 percent of trips are EI, and only 10 percent of trips are IE.

Table 6 Statistic summary of the geospatial characteristics of trips in the selected DFW dataset

Geospatial Type	Southbound	Northbound
EE	144	148
	11.88%	11.18%
EI	124	429
	10.23%	32.40%
IE	383	163
	31.60%	12.31%
II	561	584
	46.29%	44.11%
Sum	1212	1324

Note: EE = external to external, EI = external to internal, IE = internal to external, II = internal to internal.

5.2.2 Vehicle Type

Shown in Table 7, nearly 70 percent of trips are in the for-hire/private trucking fleets category for both directions. The ratios of other two provider driving profile types in two datasets are about 15 percent. The distributions in both datasets are similar.

Table 7 Statistic summary of the provider type profile characteristics of trips in the selected Maryland area

Provider Driving Profile	Southbound	Northbound
Consumer vehicle	165	205
	13.61%	15.48%
Field service/local delivery fleets	200	198
	16.50%	14.95%
For-hire/private trucking fleets	847	921
	69.88%	69.56%

5.2.3 Trip Density

Figure 9 shows that the fleet trip density in the selected dataset is a unimodal distribution. The trip density reaches its peak at 11 a.m. In Figure 3, the unimodal distribution of Maryland fleet trips has its mode at 2 p.m. Both graphs explain that fleet drivers prefer avoiding travel at morning and afternoon rush hours.

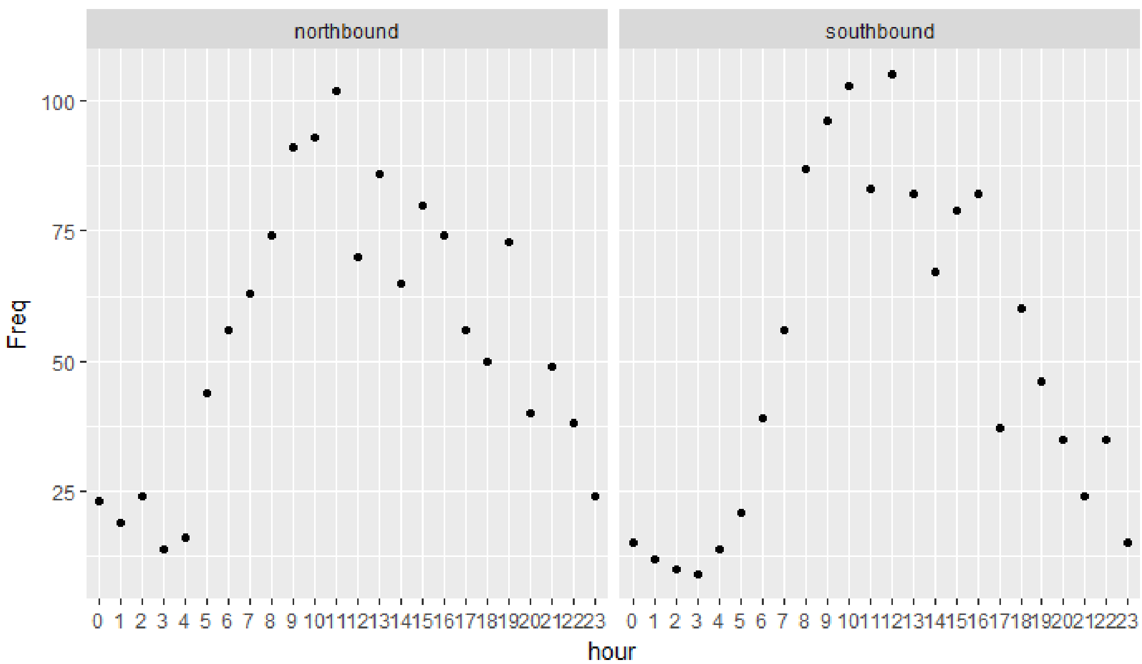

Figure 9 Hourly fleet trip distribution based on direction in the DFW dataset

5.3 Results

This case study applied to the DFW dataset the same methodology as the one used with the Maryland dataset. The two cases have the same data cleaning process, the same independent and dependent variables, and the same regression methods.

Although there are some discrepancies between the two datasets, the DFW dataset has only two data resources: consumer and fleet. The two datasets have different geospatial types. As mentioned in the data description, the research area in the Maryland dataset covers parts of two states (Maryland and Virginia). Thus, most of the trips are EI or IE, and there are no II trips. However, in the DFW dataset, most of the trips are II, and there are EI, IE, and II trips as well.

Binary logistic regression has been applied to the two directional datasets. In the DFW dataset, 85 percent of trips are fleet trips. The explanation of results focuses on the results generated from fleet trips. As in the Maryland dataset, the PTIRATIO has been excluded from models to avoid the multicollinearity effect with the TTIRATIO and the rush hour factors. The results are presented in following Table 8 and Table 9.

Table 8 DFW northbound logistic regression results

	Dependent variable:				
	total trips (1)	consumer trips (2)	**fleet trips** (3)	consumer trips (4)	fleet trips (5)
TTIRATIO	-1.211 P=0.00001***	-1.360 p=0.056*	-1.281 p=0.00015***		
PTIRATIO				-0.997 p=0.005***	-0.120 p=0.182
GEOSPATIAL(IE)	0.578 p=0.006***	1.436 p=0.080*	0.484 p=0.035***	1.631 p=0.005**	0.421 p=0.055*
GEOSPATIAL(EI)	1.342 p=0.000***	3.429 p=0.000005***	0.928 p=0.00005***	3.119 p=0.000005***	0.760 p=0.0005***
GEOSPATIAL(II)	1.735 p=0.000***	2.824 p=0.0001***	1.567 p=0.000***	2.800 p=0.000025***	1.430 p=0.000***
OFF RUSH HOUR	0.694 p=0.00005***	1.416 p=0.0005***	0.675 p=0.00025***		
MORNING RUSH HOUR	0.468 p=0.03**	1.444 p=0.031**	0.483 p=0.02**		
WEEKDAY	-0.149 p=0.184	-0.681 p=0.072*	0.133 p=0.241		
CONSTANT	-0.341 p=0.393	-1.156 p=0.150	-0.469 p=0.144	-0.682 p=0.180	-0.909 p=0.00025***
Observations	1,324	205	1,119	205	1,119
Log Likelihood	-862.648	-109.717	-730.921	-114.359	-739.962
Akaike Inf. Crit.	1,741.296	235.435	1,477.843	238.718	1,489.923
Note:				*p<0.1; **p<0.05; ***p<0.01	

DFW Northbound Logistic regression results
===
all

Table 9 DFW southbound logistic regression results

	Dependent variable:				
	total trips (1)	consumer trips (2)	**fleet trips** (3)	consumer trips (4)	fleet trips (5)
TTIRATIO	-0.944 p=0.003***	-3.204 p=0.007***	-0.748 p=0.019**		
PTIRATIO				-0.623 p=0.235	-0.163 p=0.264
GEOSPATIAL(IE)	1.645 p=0.000***	0.313 p=0.393	1.511 p=0.00000***	0.067 p=0.477	1.460 p=0.00000***
GEOSPATIAL(EI)	0.237 p=0.218	-1.373 p=0.246	0.435 p=0.088*	-1.739 p=0.165	0.397 p=0.106
GEOSPATIAL(II)	1.600 p=0.000***	-0.045 p=0.484	1.630 p=0.000***	-0.306 p=0.393	1.522 p=0.000***
OFF RUSH HOUR	0.346 p=0.092*	0.240 p=0.365	0.387 p = 0.042**		
MORNING RUSH HOUR	-0.345 p=0.072*	-0.753 p=0.184	-0.209 p = 0.205		
WEEKDAY	-0.207 p=0.127	-0.514 p=0.192	0.115 p = 0.293		
CONSTANT	-0.557 p=0.146	5.098 p=0.012**	-1.296 p = 0.0135**	2.401 p=0.040**	-1.481 p=0.00002***
Observations	1,212	165	1,047	165	1,047
Log Likelihood	-778.820	-65.001	-663.165	-73.922	-668.327
Akaike Inf. Crit.	1,573.640	146.002	1,342.330	157.843	1,346.654
Note:					*p<0.1; **p<0.05; ***p<0.01

DFW Southbound Logistic regression results

significent test are one-side tests

5.3.1 TTIRATIO

Figure 10 shows the range of the TTIRATIO in the DFW dataset is from 0.9 to 1.2. This small range makes using a 0.1-unit increment in the TTIRATIO to explain the estimated coefficient reasonably.

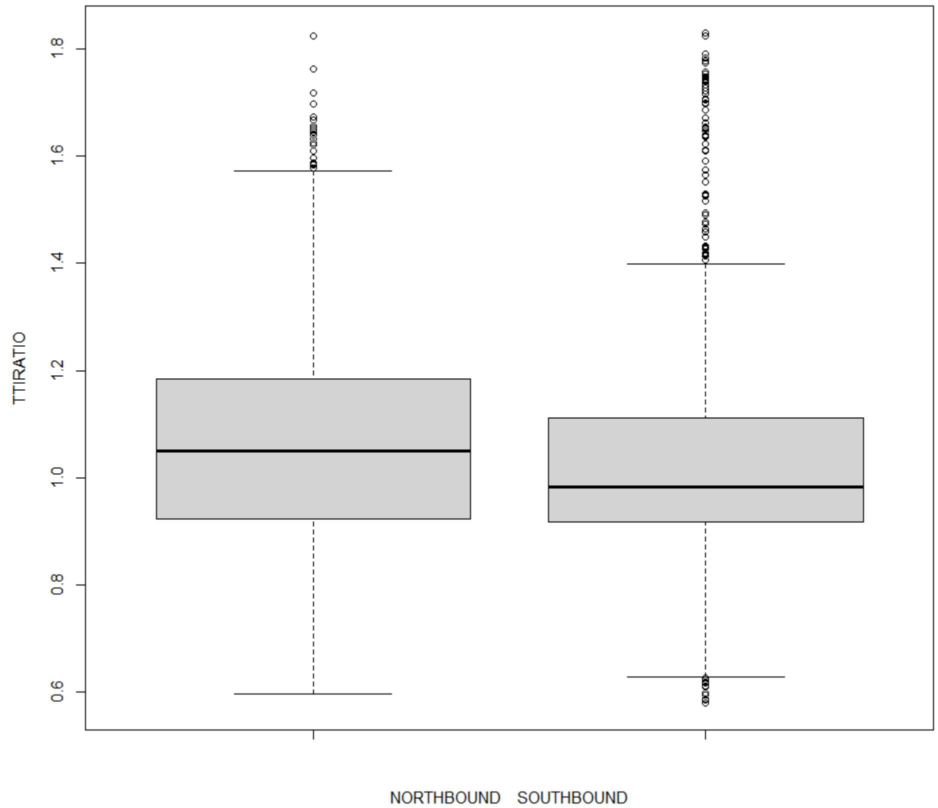

Figure 10 Boxplot of the TTIRATIO in the DFW dataset

In the fleet trip dataset:

The odds ratio $OR(TTIRATIO_S) = \exp(-0.0748) = 0.93(p = 0.019)$

$OR(TTIRATIO_N) = \exp(-0.1281) = 0.88(p = 0.00015^{***})$

The TTIRATIO shows a significant negative impact on the routing selection process. Every 10 percent increase in the TTIRATIO on the north route over the east route causes a 7 to 12 percent decrease in the odds of using the north route.

5.3.2 PTIRATIO

In the fleet trip dataset:

The odds ratio $OR(PTIRATIOS) = \exp(-0.0163) = 0.98 (p = 0.264)$

$OR(PTIRATION) = \exp(-0.012) = 0.99 (p = 0.182)$

Figure 11 shows the same reason for using a 0.1-unit increase in the PTIRATIO to interpret the estimated coefficient. Every 0.1-unit increase in the PTIRATIO causes a 2 percent and 1 percent decrease in the odds of using I-30 in the southbound and northbound dataset, respectively. However, the p values of each coefficient are 0.182 and 0.264 in the southbound and northbound datasets, respectively. The DFW dataset cannot prove the PTIRATIO significantly impacts the fleet drivers' routing decision-making process.

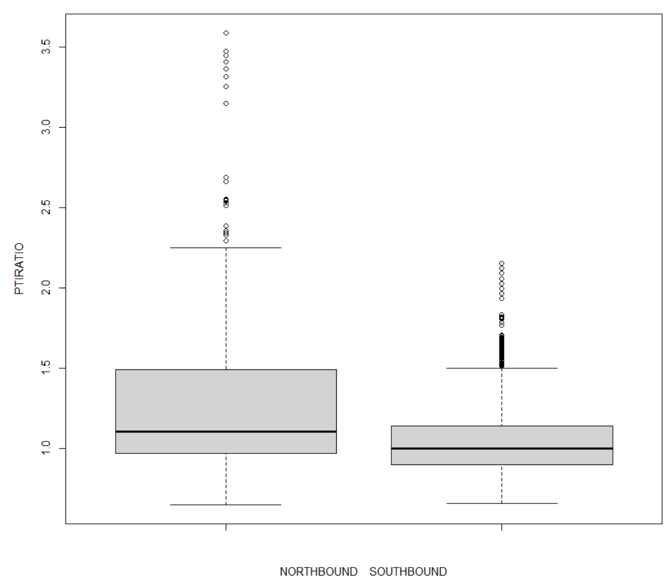

Figure 11 Boxplot of the PTIRATIO in the DFW dataset

5.3.3 Geospatial

In the fleet trip dataset:

The odds ratio $OR(GEOSPATIAL_{IE\_vs\_EE\ \_S}) = exp(1.46) = 4.3(p = 0.000^{***})$

$OR(GEOSPATIAL_{EI\_vs\_EE\ \_S}) = exp(0.397) = 1.49(p = 0.106^{*})$

$OR(GEOSPATIAL_{II\_vs\_EE\ \_S}) = exp(1.522) = 4.58(p = 0.000^{***})$

The odds ratio $OR(GEOSPATIAL_{IE\_vs\_EE\ \_N}) = exp(0.421) = 1.52(p = 0.055^{*})$

$OR(GEOSPATIAL_{EI\_vs\_EE\ \_N}) = exp(0.76) = 2.14(p = 0.0005^{***})$

$OR(GEOSPATIAL_{II\_vs\_EE\ \_N}) = exp(1.43) = 4.18(p = 0.000^{***})$

The geospatial type is a nominal variable with four unordered levels. EE is set as the reference group. The odds ratio represents the odds of the fleet drivers selecting I-30 and I-35E with EI, IE, and II geospatial type trips over the odds of choosing I-30 and I-35E with EE geospatial type trips. Compared to EE trips, other geospatial type trips have a greater preference for using I-30 and I-35E. For example, in the southbound direction, the odds of IE fleet trip drivers picking I-30 and I-35E are 330 percent higher than the odds of EE fleet trip drivers picking I-30 and I-35E. In the northbound direction, the odds of II fleet trip drivers choosing I-30 are 318 percent higher than the odds of EE fleet trip drivers choosing I-30. These results show that those EE trips have a preference for using I-20.

5.3.4 Rush Hour

In the fleet trip dataset:

The odds ratio $OR(\text{RUSH HOUR}_{\text{off-rush hour\_vs\_afternoon rush hour\_S}})$

$= \exp(0.387) = 1.47(p = 0.042^{**})$

$OR(\text{RUSH HOUR}_{\text{off-rush hour\_vs\_afternoon rush hour\_N}})$

$= \exp(0.675) = 1.96(p = 0.00025^{****})$

$OR(\text{RUSH HOUR}_{\text{morning rush hour\_vs\_afternoon rush hour\_S}})$

$= \exp(-0.209) = 0.81(p = 0.205)$

$OR(\text{RUSH HOUR}_{\text{morning rush hour\_vs\_afternoon rush hour\_N}})$

$= \exp(0.483) = 1.62(p = 0.02^{**})$

The odds ratios of off-rush hour over afternoon rush hour are larger than one, which indicates fleet drivers have more confidence in driving on I-30 during off-rush hours. The odds of fleet drivers choosing to use I-30 and I-35E to go through downtown Dallas during off-rush hour are 47 percent greater than the odds during the afternoon rush hours in the southbound dataset, and 96 percent greater in the northbound dataset. The odds of morning rush hour are 20 percent less than that of the afternoon rush hour in the southbound direction. However, the odds of morning rush hour are 62 percent greater than that of afternoon rush hour in the northbound direction.

5.3.5 Weekday

In the fleet trip dataset:

The odds ratio $OR(WEEKDAY_{weekday\_vs\_weekend\_S}) = exp(0.115) = 1.12 (p = 0.293)$

$OR(WEEKDAY_{weekday\_vs\_weekend\_N}) = exp(0.133) = 1.14 (p = 0.241)$

The odds ratio shows fleet drivers have slightly more interest in driving on I-30 and I-35E rather than on I-20 during weekdays.

6. TTIRATIO & PTIRATIO COMPARISON

TTIRATIO and PTIRATIO are two main factors we investigated in this study. The regression results of TTITATIO from two datasets are higly consistant. However, PTIRATIO factor shows certain uncertainty.

6.1 TTIRATIO

The regression results in Table 10 demonstrate that the TTIRATIO has a significant negative impact on the routing decision-making process. This result indicates that with every 0.1-unit increase in the TTIRATIO, the odds of using the north route decrease by 5 to 12 percent.

Table 10 Odds ratio with a 0.1-unit increase in the TTIRATIO

Dataset	Odds Ratio	Change
DFW northbound	0.88	−12%
DFW southbound	0.93	−7%
Maryland northbound	0.95	−5%
Maryland southbound	0.92	−8%

6.2 PTIRATIO

The negative coefficients from both datasets confirm that the PTIRATIO has a negative impact on the routing selection in Table 11. In other words, the usage of that route decreases when the travel time reliability decreases. With a 0.1-unit increase in the PTIRATIO on two similar routes, the odds change from 1 to 3 percent.

Table 11 Odds ratio with a 0.1-unit increase in the PTIRATIO

Dataset	Odds Ratio	Change
DFW northbound	0.99	−1%
DFW southbound	0.98	−2%
Maryland northbound	0.99	−1%
Maryland southbound	0.97	−3%

However, with the fleet data in this research, the southbound dataset in Maryland is the only dataset that shows that the PTIRATIO significantly impacts routing behavior. Others are not statistically significant. The awareness of travel time reliability requires route users to have a sophisticated understanding of those routes' historical traffic conditions, or to have frequently driven on those routes, and previous experiences could help them generate a sense of travel time reliability on both routes. The original data provide a unique device ID for every recorded trip. This unique identification of the geospatial information recording device can categorize those trips by the frequency of the device ID. Trips with high occurrences of the same device ID are frequent users of the selected research routes. It is highly possible that the drivers of these trips could be aware of the travel time reliability of the selected routes. Table 12 and Table 13 verifies the assumption and supports the statement that the PTIRATIO has a significant impact on routing selection. The results from the two datasets are consistent. With a 0.1-unit increase in the PTIRATIO, the odds of drivers using the north route to pass the research area decreases 13 percent in the DFW area and 10 percent in the Maryland area.

Table 12 Logistic regression with frequent fleet drivers

```
==============================================
               Maryland         DFW
----------------------------------------------
  PTIRATIO      -1.089        -1.362
               p = 0.054*   p = 0.004***

  Constant      -0.414      1.155
               p = 0.300    p = 0.025***

----------------------------------------------
Observations     459      213
Log Likelihood   -198.045   -139.217
Akaike Inf. Crit.  400.090    282.434
==============================================
Note:        *p<0.1; **p<0.05; ***p<0.01

Logistic regression with frequent fleet drivers
==============================================
Significant test are one-side tests

----------------------------------------------
```

Table 13 Odds ratio with a 0.1-unit increase in the PTIRATIO of highly frequent trips

Dataset	Odds Ratio	Change
DFW	0.87	−13%
Maryland	0.90	−10%

7. CONCLUSIONS

In this study, the significance of the influence of selected factors on the routing decision-making process was investigated. The results from the binary logistic regression on the Maryland dataset not only indicate whether the factor has a significant effect but also quantify the change in odds (percentage of fleet drivers driving on the north loop over the percentage of fleet drivers driving on the east loop) when the value of each factor changes. The regression results from the DFW dataset validate the results from the Maryland dataset.

TTIRATIO has a significant negative impact on the routing decision-making process. PTIRATIO has a negative impact on the routing decision-making process. However, this factor only shows its significance in the dataset with frequent drivers. This is a critical finding of this research. Travel time reliability is not a concern for those drivers are not familiar with alternative routes. The main reason is that only those drivers who use those routes frequently can generate the sense of travel time reliability. Otherwise, travel time reliability is not a significant factor in routing decision-making process.

Geospatial is a nominal variable with four unordered levels. Table 13 shows in both datasets, the EE spatial type is the reference group. The dummy variables all have a significant impact on routing selection. Fleet trips with a different geospatial type have different routing preference.

Rush hour is defined as a nominal variable with three unordered levels. Afternoon rush hour is set as the reference group. The regression results demonstrate that fleet vehicles have different routing preferences during different rush hour periods. The nominal variable weekday shows interesting results in the two datasets. It is a significant factor in the Maryland dataset, and it is not a significant factor in the DFW dataset. In the DFW area, fleet drivers did not show significant preference on routing selection during weekdays and weekends. In the Maryland area, the preference on routing selection is significantly different during weekdays and weekends. The possible explanation is that traffic on the research routes in DFW did not show much difference during weekdays and weekends. However, the traffic conditions of the research area in Maryland might vary during weekdays and weekends.

REFERENCES

[1] Knorring, J., R. He, and A. Kornhauser. Analysis of route choice decisions by long-haul truck drivers. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1923, 2005, pp. 46-60.

[2] Hunt, J. D., and J. E. Abraham. Stated response examination of factors influencing commercial movement route choice behavior. In *Applications of Advanced Technologies in Transportation Engineering (2004)*, 2004. pp. 72-77.

[3] Liu, H. X., W. Recker, and A. Chen. Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A: Policy and Practice*, Vol. 38, No. 6, 2004, pp. 435-453.

[4] Dia, H. An agent-based approach to modelling driver route choice behaviour under the influence of real-time information. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 5, 2002, pp. 331-349.

[5] Figliozzi, M. A., L. Kingdon, and A. Wilkitzki. Analysis of freight tours in a congested urban area using disaggregated data: characteristics and data collection challenges. Proceedings 2nd Annual National Urban Freight Conference, Long Beach, CA. December, 2007.

[6] Bitzios, D., and L. Ferreira. Factors affecting route choice of commercial vehicle drivers. *Australasian Transport Research Forum*, No. 18, 1993, pp.27-40.

[7] Eisele, W., L. Rilett, K. Mhoon, and C. Spiegelman. Using intelligent transportation systems travel-time data for multimodal analyses and system monitoring. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1768, 2001, pp. 148-156.

[8] Feng, T., T. Arentze, and H. Timmermans. Capturing preference heterogeneity of truck drivers' route choice behavior with context effects using a latent class model. *EJTIR*, Vol. 13, No. 4, 2013, pp. 259-273.

[9] Lu, X., S. Gao, and E. Ben-Elia. Information impacts on route choice and learning behavior in a congested network: Experimental approach. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2243, 2011, pp. 89-98.

[10] Abdel-Aty, M. A., R. Kitamura, and P. P. Jovanis. Investigating effect of travel time variability on route choice using repeated-measurement stated preference data. *Transportation Research Record*, No. 1493, 1995, pp. 39-45.

[11] Kerkman, K., T. Arentze, A. Borgers, and A. Kemperman. Car drivers' compliance with route advice and willingness to choose socially desirable routes. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2322, 2012, pp. 102-109.

[12] Golob, T. F., and A. C. Regan. Trucking industry preferences for traveler information for drivers using wireless Internet-enabled devices. *Transportation Research Part C: Emerging Technologies*, Vol. 13, No. 3, 2005, pp. 235-250.

[13] Pearmain, D., and E. P. Kroes. Stated preference techniques: a guide to practice. Steer Davis Gleave and Hague Consulting Group, Hauge. 1990.

[14] Eisele, W. L., B. Naik, and L. R. Rilett. Estimating route travel time reliability from simultaneously collected link and route vehicle probe data and roadway sensor data. *International Journal of Urban Sciences*, Vol. 19, No. 3, 2015, pp. 286-304.

[15] Nowakowska, M. Logistic models in crash severity classification based on road characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2148, 2010, pp. 16-26.

[16] Dissanayake, S., and J. Lu. Analysis of severity of young driver crashes: sequential binary logistic regression modeling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1784, 2002, pp. 108-114.

[17] Ben-Akiva, M. E., and S. R. Lerman. *Discrete choice analysis: theory and application to travel demand*. MIT Press, 1985.