

Deriving on-Trip route choices of truck drivers by utilizing Bluetooth data, loop detector data and variable message sign data

Sharma, Salil; Snelder, Maaïke; Lint, Hans Van

DOI

[10.1109/MTITS.2019.8883311](https://doi.org/10.1109/MTITS.2019.8883311)

Publication date

2019

Document Version

Final published version

Published in

MT-ITS 2019 - 6th International Conference on Models and Technologies for Intelligent Transportation Systems

Citation (APA)

Sharma, S., Snelder, M., & Lint, H. V. (2019). Deriving on-Trip route choices of truck drivers by utilizing Bluetooth data, loop detector data and variable message sign data. In *MT-ITS 2019 - 6th International Conference on Models and Technologies for Intelligent Transportation Systems* Article 8883311 IEEE. <https://doi.org/10.1109/MTITS.2019.8883311>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' – Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Deriving on-trip route choices of truck drivers by utilizing Bluetooth data, loop detector data and variable message sign data

Salil Sharma

*Civil Engineering and Geosciences
Delft University of Technology
Delft, Netherlands
S.Sharma-4@tudelft.nl*

Maaïke Snelder

*Civil Engineering and Geosciences
Delft University of Technology; TNO
Delft, Netherlands; The Hague, Netherlands
M.Snelder@tudelft.nl*

Hans van Lint

*Civil Engineering and Geosciences
Delft University of Technology
Delft, Netherlands
J.W.C.vanLint@tudelft.nl*

Abstract—On important truck-dominated motorways, a large share of traffic consists of trucks. Our hypothesis is that these trucks do not always make optimal routing decisions which cause inefficiencies in the traffic system. Therefore, route choice of truck drivers is of interest to both transport planners and traffic management authorities. The objectives of this paper are two-fold. First, this paper models on-trip route choices of the truck drivers. Second, we assess the inefficiencies of those routing decisions. This paper utilizes Bluetooth data, loop detector data, and variable message sign data to model the route choices of truck drivers. To the best of our knowledge, this is the first time that Bluetooth data have been used for the estimation of route choice models of truck drivers. The trucks are inferred from Bluetooth data by applying a Gaussian mixture model-based clustering technique. We apply both a binary logit model and a mixed logit model to derive the route choices of truck drivers on a case study between the port of Rotterdam and hinterland in the Netherlands. The predictive performance of the model is tested by conducting out-of-sample validation. The model results indicate truck drivers significantly value travel distance, instantaneous travel time and lane closure information en-route. The estimate of travel distance varies significantly among truck drivers. While 38% of truck drivers do not take the shortest time path, 48% of truck drivers do not choose the system-optimal path. These inefficiencies imply that traffic management solutions have the potential to improve the performance of truck-dominated motorways.

Index Terms—on-trip, en-route, route choice, truck driver, binary logit, mixed logit, bluetooth, loop detector, variable message sign, truck-dominated motorways, freight corridors

I. INTRODUCTION

A large share of traffic consists of trucks on truck-dominated motorways. Truck drivers navigate through the network by making routing decisions. However, the routing behavior of truck drivers may differ from their counterparts since logistics companies may have provided them with a pre-trip route plan to reach their destination. Although significant research efforts are dedicated to understanding the route choices of passenger

car drivers, little attention has been devoted to truck drivers [1]. It is imperative for both transport planners and road management authorities to understand the routing behavior of truck drivers. A better understanding of the on-trip behavior of truck drivers will support the design of appropriate traffic management solutions to ensure consistent performance of the truck-dominated motorways.

The studies on modeling route choice can be categorized into one of two groups based on the availability of the information to the road user and the instant at which the road users make a decision about their routes: pre-trip route choice and on-trip route choice. In pre-trip route choice models, the road users are assumed to choose their routes before starting the trip and they are assumed to have perfect information about the traffic conditions. On the contrary, the on-trip route choice models assume that the road users may deviate from their pre-trip routes based on the current traffic conditions [2]. This paper focuses on on-trip route choices of truck drivers. In this paper, terms such as routes and path are used interchangeably.

On-trip routing decisions reflect the responses of drivers towards current traffic information. This information can be disseminated to the drivers through roadside panels, variable message signs, or the navigation devices. Previous works have either used stated-preference (SP) surveys ([3]–[9]) or revealed-preference (RP) data ([1], [10]–[13]) to derive route choices of truck drivers. The strengths and weaknesses of both methods are widely known. SP data have limitations due to a difference in claimed and observed routing decisions. On the other hand, RP data can reveal the actual choices of the truck drivers, generally contextual information is lacking. It has been suggested by [14] to combine RP and SP data sources to collect freight data. Until now, the impact of current traffic information such as lane closures and driving experience based notion of travel time reliability is only studied in SP studies ([3], [4], [6]). In this paper, we enrich an RP dataset with contextual information by utilizing multiple data sources to overcome the limitations of previous RP/SP studies. In the

The research presented in this paper is a part of the project “ToGRIP-Grip on Freight Trips” supported by the Netherlands Organization for Scientific Research (NWO), TKI Dinalog, Commit2data, Port of Rotterdam, SmartPort, Portbase, TLN, Deltalinqs, Rijkswaterstaat, and TNO.

context of RP studies, GPS dataset is often used to model route choices of truck drivers. With advances in technology, Bluetooth is another cost-effective solution which is used as an RP data source in this paper. Bluetooth data provide origin and destination information of vehicles. Previously it has been used to understand a vehicle's diversion due to bridge closure [15] and pedestrian's route choices at the train station [16]. Besides, we utilize loop-detector data to derive the value of travel time reliability and variable message sign data to retrieve lane closure information on a path.

Consequently, the objectives of this paper are two-fold. First, we utilize Bluetooth data, loop detector data and variable message sign data to model on-trip route choices of truck drivers. Second, we evaluate the inefficiencies in their route choices by presenting a data-driven approach. Since Bluetooth data do not provide mode classification, a Gaussian mixture model-based clustering method is applied to extract trucks. Afterward, we apply both binary logit and mixed logit models to model the on-trip routing behavior of truck drivers on a case study between the port of Rotterdam and hinterland. An out-of-sample validation is conducted to assess the predictive power of the model. Lastly, we present data-driven metrics to analyze the efficiencies of the routing decisions of truck drivers from both user's and system's perspectives.

The contributions of this paper are as follows. First, we present a method to extract trucks from Bluetooth data. Second, we present Bluetooth as an RP data source to model the route choices of truck drivers between the port of Rotterdam and hinterland in the Netherlands. Third, we study the impact of lane closures and travel time reliability on observed routing decisions (i.e., RP setting) of truck drivers in contrast to previous studies in the SP setting. Fourth, we present data-driven metrics to assess the efficiency of truck driver's routing decisions.

This paper is structured as follows. Section II presents a literature review. The study area is described in section III. Section IV presents truck related data extraction method from Bluetooth data. Section V presents the methodology, model specification, and choice data. The results of the estimated model are presented and discussed in section VI. Section VII discusses the efficiency of on-trip routing decisions, and section VIII concludes the paper.

II. LITERATURE REVIEW

This section presents a literature review for the key concepts used in this paper: route choice modeling of truck drivers and mode classification using Bluetooth data.

A. Route choice modeling of truck drivers

SP surveys and GPS data (RP) have been primarily used to collect freight data. In [3], a logit model is developed to analyze the impact of variable message signs on driver's diversion decisions. The data is collected via an on-site survey in the form of a questionnaire. The study further explores how the diversion behavior of truck drivers differs from non-truck drivers. They note that the diversion behavior of truck

drivers may be restricted since they have less feasible routes to follow. The familiarity with a route is found to be a significant variable for truck drivers. In [4], the response of truck drivers is investigated towards variable message signs under incident occurrence. They conduct a stated-preference survey of about 100 truck drivers operating in Athens, Greece. A random-effects ordered probit model has been utilized to model the truck driver's response to diversion. Their results indicate that the long delay, displayed through variable message signs and the availability of an alternative route increase the probability for truck driver diversion.

A stated choice experiment is conducted in [5] to analyze the route choices of truck drivers. The questionnaire is filled by truck drivers and a contact person from the trucking companies located in the Eindhoven region, the Netherlands. Using a mixed logit model, their results indicate that the truck drivers and planners have a preference for the highways, shortest time route and uncongested route. In [6], the route choices of truck drivers are explored under variable risk situations. In their experiments, the truck drivers prefer a risk-averse route while choosing between a short and uncertain route and a longer but more certain route. Their findings suggest that truck drivers value reliability more than the shortest path.

In [7], 252 truck drivers in the United States and Canada are interviewed to understand their decision-making process. [8] utilizes the same data and estimates a random effects logit model. The findings suggest that there exists a wide variation in truck driver's route choices. Other factors such as driver employment terms, bearer of toll costs, driver compensation methods, shipment characteristics and magnitude of delays affect their decisions. For example, the results reflect that the truck drivers avoid the toll roads when they are responsible for the cost but they become indifferent to the toll cost when they are not responsible for the cost.

In [9], a survey is sent to the trucking companies, which include shippers, carriers and receivers, in Washington (WA) state, United States to understand their routing priorities. The data is analyzed for three different latent classes. The results indicate that the long-haul providers value travel distance, refueling locations, parking availability, size and weight limits, and hours of service limit as contributing factors in determining their route choices.

GPS data have also been used by the researchers to understand the diversity of the routes chosen by truck drivers and to model their routing decisions. In [10], the GPS records of the trucks, hosted at the American Transportation Research Institute (ATRI), are analyzed for the routes chosen by truck drivers between an origin-destination (OD) pair. The findings can be valuable for choice set generation schemes. In [11], the diversity of the route choices are analyzed both for short-haul trucks and long-haul truck drivers. Their findings indicate that the long-haul truck drivers, in contrast to short-haul truck drivers, prefer alternative routes if there is high travel time variability on the most used route. For both short and long-haul truck drivers, OD pairs with a higher number of observed trips lead to more number of unique routes. Observing the

route choices of the trucks can be applied to generating the choice set for modeling purposes.

In [12], the route choices of long-haul truck drivers are modeled using the GPS data set collected for the U.S. highway network. In most of the cases considered in this study, the truck drivers have to make a choice between the route going through downtown and a bypass route. They develop a binary logistic model by using travel distance and travel time as the explanatory variables. Their findings suggest that the truck drivers are primarily travel time minimizers since the time parameter is significantly higher than the distance parameter. In [1], an error component logit model is used to analyze the route choices of the long-haul truck drivers in England using GPS data. The results indicate that truck drivers prefer routes with low travel cost and travel time. For urban freight truck drivers, a route choice model has been developed in [13] by utilizing the truck GPS data from the Tokyo metropolitan area. The route choice model is based on the concept of maximum route-overlapping ratio developed by [17]. The explanatory variables are types of roads and their respective load limits. The results indicate that truck drivers prefer the roads which allow more load to be carried.

A summary of studies on route choice modeling of truck drivers is presented in Table I. The RP based studies use travel distance and travel time, type of roads, and load limits as key explanatory variables. In contrast, SP-based studies look further and use variable message signs, roadside amenities, driving experiences, and the like as other explanatory variables. This paper, being an RP based study, tests the effect of lane closures and travel time unreliability on on-trip routing decisions of truck drivers by utilizing loop detector data and variable message sign data.

B. Travel mode classification using Bluetooth data

Bluetooth data do not provide modal information. Previous works have used both supervised and unsupervised learning schemes for mode identification. In [18], a Bluetooth based detection model is proposed to distinguish among three different modes (autos such as passenger cars, motorcycles, and trucks; cyclists; and pedestrians) using a genetic algorithm and neural network in a supervised learning setting. In [19], the class of Bluetooth device and the strength of the signal are utilized to distinguish between motor vehicles and bicycles. This approach, however, requires information about the MAC addresses which may not be available in general due to privacy regulations. Based on the travel time observations between two Bluetooth sensors, [20] utilize clustering methods to classify the road users (car drivers and cyclists) into different user classes. They report that Gaussian mixture models work better than the k-means when tested over different road segments. Above studies have used clustering to distinguish between the modes with different mechanical properties. This paper classifies passenger cars and trucks which share the same road space and have similar mechanical properties.

III. STUDY AREA

The study area is selected based on the locations of Bluetooth sensors and their ability to detect the passing traffic. Moreover, it is located near the port of Rotterdam and witnesses high truck percentages. The study area comprises the motorway ring (A20, A15, A16, and A4) surrounding Rotterdam in the Netherlands (see Figure 1). We consider the traffic going towards the port of Rotterdam (node B) from the hinterland (node A). Two route choices, marked by path 1 and 2, are considered. Each path is divided into two segments. Segment 1 runs from Bluetooth station 1 to 3, segment 2 from Bluetooth station 3 to 2, segment 3 from Bluetooth station 1 to 4 and segment 4 from Bluetooth station 4 to 2.

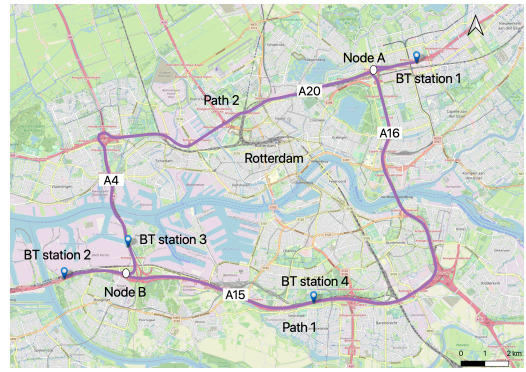


Fig. 1: Map showing the origin, destination and location of four Bluetooth stations

IV. IDENTIFYING TRUCKS FROM BLUETOOTH DATA

This section describes the data sources and the pre-processing steps so that the data can be utilized for route choice modeling. The Bluetooth data is analyzed with python 3.6 and the code is available at <https://github.com/salilsharma/Bluetooth>.

A. Description of Bluetooth data

ToGRIP-Bluetooth service provides the Bluetooth data collected by the port of Rotterdam. When queried, the service returns data in a json format. The real MAC address is converted to an 11 digit vehicle ID using hashing thus the privacy is retained at the user level. The Bluetooth sensor records the time stamp and the strength of the vehicle identification for every MAC address associated with a passing vehicle. The travel time between two Bluetooth sensors can be estimated from the time stamps of the corresponding MAC address. Bluetooth data retrieved from ToGRIP-Bluetooth service are coded with UTC time zone; therefore, it is necessary to convert UTC time to CET/CEST depending on the time of the year.

1) *Preliminary analysis:* The study area comprises 4 Bluetooth stations. Each path is divided into two segments (see Figure 1) marked by the locations of Bluetooth stations. For segments 2 and 4, clusters of travel times can be observed (see

TABLE I: Summary of route choice modeling of truck drivers

Studies	Research Design			
	Data	Study focus	Model	Variables
[1]	GPS data	Motorway, A Road, B Road in England	Error component logit model	Travel cost (function of free flow travel time and type of roadway) and travel time
[12]	GPS data	Highway in USA	Binary logistic model	Travel distance and travel time
[13]	GPS data	Tokyo Metropolitan area, Japan	Maximum route overlapping ratio model	Type of roads and load limit
[3]	Survey	Borman expressway in USA	Binary logit model	Gender, familiarity with the route, driving experience on Borman expressway, content of variable message sign, and trust in that information
[4]	Survey	Athens metropolitan area, Greece	Random-effects ordered probit model	Source of information, driving experience in congested conditions, professional driving experience, provision of alternative route, content of variable message sign, vehicle ownership, and perceived utility from variable message sign
[5]	Survey	Trucking companies located in Eindhoven region, Netherlands	Mixed logit model	Travel time, congestion, road category, road pricing, roadside amenities, area surrounding the roadway
[8]	Survey	Highways in USA and Canada	Random effects logit model	Travel time, toll, delay, route passing through downtown, and entity responsible for paying tolls
[9]	Survey	Washington State (WA) freight companies in USA	Item response theory and latent class analysis	Travel distance, stops, tolls, roadside amenities, hazardous material, load limits, driver availability, and driving hours restrictions.
This paper	Bluetooth data, loop detector data, and variable message sign data	Motorway (A-type) in the Netherlands	Binary logit model, random coefficient mixed logit model with panel effects	Instantaneous travel time, travel distance, travel time unreliability, and lane closures

Figure 2). The clusters could be attributed by the differential speed limits observed in the Netherlands. The truck drivers weighing more than 3.5 metric tonnes have a speed limit of 80 km/h on motorways. However, the clusters are not observed for the segments 1 and 3. It can be inferred that the clusters in travel time plots are observed over short segments of the motorways with restricted route choice possibilities.

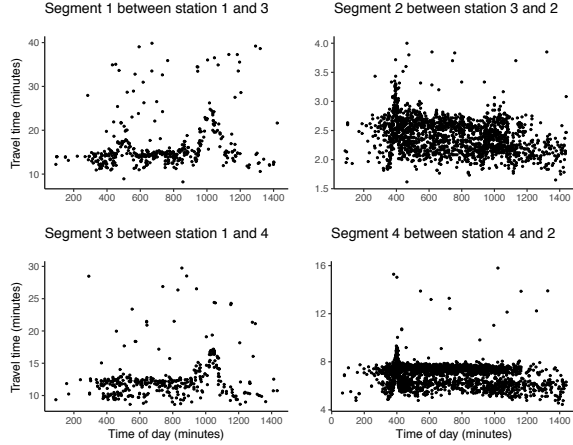


Fig. 2: Bluetooth travel time observations on 24 Nov. 2017

2) Outlier removal from Bluetooth field observations:

Before applying a clustering algorithm, it is necessary to remove the outliers. Tukey's method [21] has been employed to detect and remove the outliers because of its resistance to extreme values as it uses quartiles. Tukey's method utilizes the

inter-quartile range (IQR) to detect outliers. IQR is the distance between the lower (Q1) and upper (Q3) quartiles. The points lying outside the interval $[Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR]$ are marked as possible outliers.

B. Clustering Bluetooth data based on travel time

Based on preliminary analysis, it has been observed that the travel time clusters are formed for certain segments of the motorways. However, the separation may not be clear for all segments. For instance, Figure 2 shows that the clusters are fuzzy for segment 2. Therefore, a Gaussian mixture model-based clustering technique [22] has been employed in this paper. The purpose of clustering the travel time observations is to place them into two groups: fast and slow vehicles. An example of clustering is shown in Figure 3. For congested conditions, we observe more vehicles in the slow vehicles' group. Next, we present a method which uses speed over the full path to extract trucks from the slow vehicles' group.

C. Truck-related data extraction

A general methodology is presented here based on the learnings from the preliminary analysis. The methodology consists of the following steps.

- 1) For the said time period, i.e., four weeks, find all the vehicles that have passed through a path. Remove outliers and store the vehicle Ids in a master list.
- 2) Identify a short stretch of the path where vehicles can be clustered based on travel times. Find the common vehicle Ids that belong to the slow vehicle cluster and master list from step 1.
- 3) From the common vehicle Ids, select the vehicles which have traversed the path with a maximum speed of 80

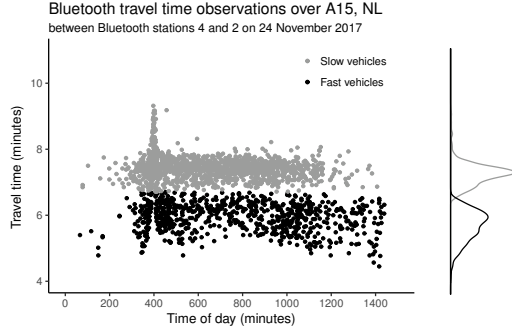


Fig. 3: Example showing clusters of vehicles after applying Gaussian mixture model based clustering

km/h. This speed threshold refers to the speed limit for trucks on motorways in the Netherlands. The vehicle Ids thus extracted can be classified as trucks and used for route choice model estimation. Since outliers are removed in step 1, the trucks extracted from this method will be the ones that have reached their destination from the origin without making a stop in between.

V. ROUTE CHOICE MODEL

This section describes a binary logit and a mixed logit model and their model specifications to derive on-trip route choices of truck drivers. The statistics of the data collected for modeling purposes is also discussed. Table II presents the notations that are used throughout this paper.

TABLE II: Notations used in this paper

Variable	Description
ITT	Instantaneous travel time (min)
TD	Travel distance (km)
LC	Lane closures
$TTUR$	Travel time unreliability
SD	Standard Deviation
Min	Minimum
Max	Maximum
$ITT_{P1}, TD_{P1}, TTUR_{P1}, LC_{P1}$	P1 denotes path 1
$ITT_{P2}, TD_{P2}, TTUR_{P2}, LC_{P2}$	P2 denotes path 2
$\mathcal{L}(\beta_0)$	Initial log likelihood
$\mathcal{L}(\hat{\beta})$	Final log likelihood
$\bar{\rho}^2$	Adjusted McFadden's rho-squared

A. Theoretical background

For route choice modeling, a multinomial logit (MNL) model (used as binary logit in this paper) and a mixed logit model with panel effects are used. The MNL model is a discrete choice model based on random utility theory which assumes that the individual is perfectly rational and selects the alternative with the highest utility. However, the analyst is assumed to have incomplete information; therefore, uncertainty has to be taken into account. Therefore, the utility is modeled as a random variable so as to reflect the uncertainty

[24]. The utility U_{in} that an individual n receives from choosing alternative i from the choice set C_n is described by the following equation.

$$U_{in} = V_{in} + \epsilon_{in} \quad (1)$$

where V_{in} is the deterministic part of the utility and ϵ_{in} is the error term which is independent and identically Gumbel distributed.

The alternative with the highest utility is chosen. The probability that alternative i is chosen by the individual n from choice set C_n is given by:

$$P_{in} = Pr(U_{in} \geq U_{jn}, \forall j \in C_n) \quad (2)$$

V_{in} involves the explanatory variables while distributional assumptions are made on the joint distribution of error terms $\epsilon_{in} = (\epsilon_{1n}, \dots, \epsilon_{jn})$. Therefore, the probability that a given individual n chooses alternative i from the choice set C_n is:

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \quad (3)$$

The MNL model falls short in explaining the taste heterogeneity among the individuals in a population; besides, it cannot capture the panel effects induced by the repetitive decision making of the individuals. The mixed logit (ML) model obviates these limitations of the MNL model. Following [25], the ML model can be expressed as follows.

$$P_{in} = \int L_{in}(\beta) f(\beta) d\beta \quad (4)$$

where $L_{in}(\beta)$ is the logit probability evaluated at parameters β . $f(\beta)$ is a density function which denotes that estimates vary over individuals or decision makers.

To account for repeated choices made by each individual, logit probability can be expressed as follows.

$$L_{in}(\beta) = \prod_{t=1}^T L_{itn} \quad (5)$$

which involves a product of logit probabilities, one for each time period, $i = \{i_1, \dots, i_T\}$.

B. Model specification

Following attributes are considered for the paths and utility is specified as a linear sum of these attributes: total distance of a path, instantaneous travel time of a path, travel time unreliability of a path, and the maximum number of lanes closed along a path. Road pricing is not implemented in the Netherlands thus it is not considered.

It is assumed that truck drivers take into account the current traffic conditions as well as their own experience in the past to make routing decisions on-trip. Current traffic conditions may be disseminated to truck drivers through roadside panels, navigation devices, or logistic companies. For the ML model, we consider that these attributes are normally distributed around a mean and with a standard deviation.

C. Choice data

For data preparation, a period of four consecutive weeks, excluding the weekends, from 30-10-2017 to 24-11-2017 has been selected which contains 20 workdays. Out of 1671 observations, 1293 truck drivers choose path 2 and 378 choose path 1. While travel distance is a fixed variable, other variables are computed at the instant a truck driver reaches the decision node (i.e., node A in Figure 1).

1) *Instantaneous travel time of a path*: Loop detector data, provided by Regiolab-Delft service, is used to compute instantaneous travel times. MATLAB version R2018a is used to call Regiolab-Delft service. Loop detectors are installed at every 500 m of motorways in the Netherlands. It is assumed that the traffic conditions do not change for every section between the detector locations. Aggregated speed values for every detector location are retrieved every minute. Then for every such section, the instantaneous travel times are calculated from the speed data. For every path, the travel times of individual sections are added.

2) *Travel time unreliability of a path*: λ_{skew} is selected to measure the travel time unreliability of a path. This indicator captures the skewness of the day-to-day travel time variabilities. In contrast to other measures of unreliability which are sensitive to extreme events or outliers, λ_{skew} can be interpreted as the likeliness of incurring a very bad travel time (relative to the median) [26]. Mathematically, it is defined as the ratio of the distance between the 90th and 50th percentile and the distance between the 50th and 10th percentile.

λ_{skew} is estimated for different time periods of a day. For each time period of a day, the previous 10 days of travel times are used to arrive at the value of λ_{skew} , which thrusts upon the value individuals associated with their recent driving experiences. Four time periods are considered: morning peak (06:30-09:30), day (09:30-16:00), evening peak (16:00-19:00), and late evening to early morning (19:00-06:30).

3) *Maximum number of lanes closed along a path*: The lane closures are retrieved from VMS data provided by the Regiolab-Delft service. The lane closures denote a reduction in capacity which could be a proxy for the incidents. This variable is added to test if truck drivers change their routes in response to an incident downstream. At the instant truck drivers reaches the decision node, we select a maximum number of lanes closed along a path in our analysis.

4) *Descriptive statistics of choice data*: Descriptive statistics of choice data along with the correlation matrix and variance inflation factor (VIF) are reported in Table III. The cells for variables TD_{P1} and TD_{P2} are empty in the correlation table since these values do not vary across the dataset. Since the value of VIF is less than 5, the effect of multicollinearity on the parameter estimates can be ruled out [23].

VI. MODEL ESTIMATION AND VALIDATION

The route choice models are estimated using PythonBiogeme version 2.6 [27]. We report and compare the results obtained from both binary logit and mixed logit model.

A. Binary logit model

Table IV presents the estimates of the route choice model. The signs for travel time, travel distance, travel time unreliability, and lane closures are all negative as expected. The alternative specific constant is found to be insignificant and its effect is observed to be subsumed by the travel distance parameter.

For out-of-sample validation, the size of the validation set is chosen as 20%. We randomly divide the whole dataset into estimation and validation dataset. For every estimation and validation dataset, we estimate the model parameters and calculate model fit, i.e., average log-likelihood, on the validation dataset. This procedure is repeated 500 times. Figure 4 shows the distribution of the parameter's estimates computed from 500 different estimation dataset. The mean of parameter estimate is represented by the black dashed line. The expected value of the average log-likelihood is -0.520 after analyzing the model fit based on 500 samples (see Figure 4). The model fit is within -0.69 (equal-probability model) and zero which shows the good predictive power of the binary logit model.

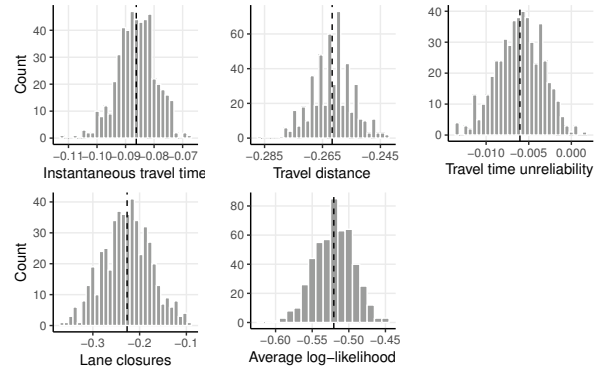


Fig. 4: Distribution of parameter estimates and average log-likelihood for binary logit model

B. Mixed logit model

A simulation was performed using 500 random draws. Table IV reports the results obtained from the mixed logit model. The ML model assumes that truck drivers have a taste sensitivity towards path attributes, and it also accounts for panel effects. The signs for travel time, travel distance, travel time unreliability, and lane closures are negative as expected.

The ML model is compared with the binary logit model. The likelihood ratio test is: $-2(-867.758 + 848.337) = 38.842$. The number of degrees of freedom for this test is equal to the difference in a number of parameters of both binary logit and ML model, i.e., $8-4=4$. The threshold for the test at 5% level is therefore $\chi^2_{4,0.05} = 9.488$. Since $38.842 > 9.488$, we can reject the hypothesis that the two models are equivalent at 5% level. Therefore, the ML model outperforms the binary logit model.

The estimate of travel distance varies significantly in the population. About 81% of the distribution is below zero and

TABLE III: Descriptive statistics of dataset and correlation between variables

Variable	Descriptive statistics				Correlation matrix								VIF
	Mean	SD	Min	Max	ITT _{P1}	TTUR _{P1}	LC _{P1}	TD _{P1}	ITT _{P2}	TTUR _{P2}	LC _{P2}	TD _{P2}	
ITT _{P1}	22.36	0.42	15.00	68.06	1.000								2.245
TTUR _{P1}	3.92	7.47	0.18	40.58	-0.184	1.000							1.151
LC _{P1}	0.26	0.49	0.00	3.00	-0.207	0.175	1.000						1.085
TD _{P1}	27.60	0.00	27.60	27.60	–	–	–	1.000					1.983
ITT _{P2}	21.19	6.19	14.55	57.12	0.559	-0.217	-0.061	–	1.000				1.616
TTUR _{P2}	4.56	7.76	0.37	50.72	-0.168	0.178	0.135	–	-0.221	1.000			1.121
LC _{P2}	0.18	0.49	0.00	3.00	0.483	0.144	-0.056	–	0.066	0.128	1.000		1.558
TD _{P2}	23.10	0.00	23.10	23.10	–	–	–	–	–	–	–	1.000	1.998

19% above. This implies that four out of five truck drivers prefer the shortest distance path. For one out of five truck drivers, utility increases with the longer route.

TABLE IV: On-trip route choice model for truck drivers

Parameters		Binary logit		Mixed logit with panel effects	
		Value	Robust t-test	Value	Robust t-test
ITT	Mean	-0.0866	-6.39	-0.152	-4.89
	SD			0.0197	0.21
TD	Mean	-0.262	-19.54	-0.463	-6.33
	SD			0.512	4.40
TTUR	Mean	-0.00594	-1.10	-0.00899	-0.98
	SD			-0.00125	-0.53
LC	Mean	-0.229	-2.36	-0.414	-2.35
	SD			-0.493	-1.03
Number of observations			1671		1671
Number of individuals			1419		1419
$\mathcal{L}(\beta_0)$			-1158.249		-1158.249
$\mathcal{L}(\beta)$			-867.758		-848.337
$\bar{\rho}^2$			0.247		0.261

The estimate of travel time unreliability is found to be insignificant even after taking into account the panel effects. The choice data reveal that few truck drivers make the journey between the same origin and destination more than once within a month. Since the notion of travel time unreliability builds upon experiences, the dataset does not have that many truck drivers traveling between the same OD pair more often.

VII. COMPUTING EFFICIENCY OF TRUCK DRIVERS' ROUTING DECISIONS

The previous section has reported key attributes which truck drivers consider while making on-trip routing decisions. In this section, we assess the efficiency of those routing decisions from both the user's and system's perspectives.

A. User-centric routing efficiency

In this paper, an user-centric decision can be defined as the one where truck drivers arrive at a decision node and choose a path with least instantaneous travel time. Figure 5 shows that 62% of truck drivers choose the shortest time path and this proportion increases when the time difference between paths increases. The dataset is further divided into regular conditions (no lane closures on any path) and lane closures on any path. It is observed that truck drivers make more conscious choices

when they are informed about the lane closures. Moreover, truck drivers have bounded rationality as they are more likely to choose the shortest time path if the time difference between the two paths is more than 10 minutes.

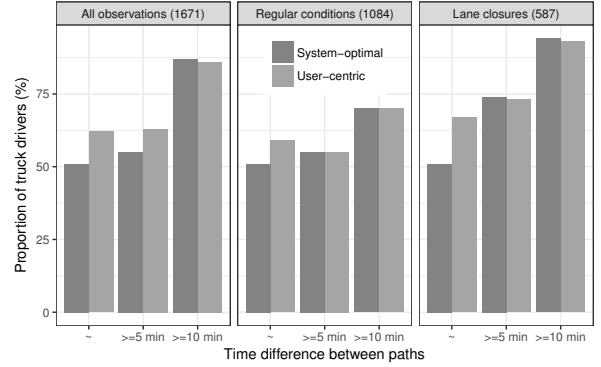


Fig. 5: Efficiency of truck drivers' routing decisions

B. System-optimal routing efficiency

In [28], the vehicles are assigned to the minimum time-dependent marginal travel time paths to achieve a system-optimal state. However, it requires other data such as an OD matrix and link cost functions. In this paper, a system optimal routing decision is approximated as the one where truck drivers arrive at a decision node and choose a path which satisfies two criteria simultaneously: the path should have enough spare capacity and the instantaneous travel time on it should not be worse than that of shortest time path. The least instantaneous travel time condition guarantees that such routing decisions do not increase system-wide travel time. We use density as an indicator to interpret the spare capacity of a path. To compute density, we first divide a path into smaller segments bounded by the loop-detectors. Then we utilize flow and speed data retrieved from Regiolab-Delft service to compute section-specific density values. If the maximum of all such density values is less than a nominal value of critical density, i.e., 25 veh/km/lane, then that path is assumed to have spare capacity. The nominal value of critical density used in this paper aligns with the field-measured value on a real expressway [29]. Figure 5 shows that 52% of truck drivers make system-optimal route choices; this proportion

increases when the time difference between paths increases. If the travel time difference between the two paths is more than 10 minutes, truck drivers are inclined to make system-optimal routing decisions.

VIII. CONCLUSION

This paper presents an on-trip route choice model for truck drivers by utilizing Bluetooth data, loop detector data, and variable message sign data. This analysis is useful for truck-dominated motorways where a large share of traffic consists of trucks. We present a method to extract truck related data from Bluetooth data by applying a Gaussian-mixture model based technique. The case study for truck-dominated motorways between the port of Rotterdam and hinterland in the Netherlands shows that truck drivers value significantly travel time, travel distance and lane closures en-route. The mixed logit model shows that the estimate of travel distance varies significantly in the population. Three out of five truck drivers choose the shortest time path, and this proportion increases if they could distinguish easily the time difference between the alternatives. Only 52% of truck drivers choose a path with enough spare capacity or make system-optimal routing decisions. Routing efficiency can be improved by utilizing traffic management solutions. We can guide truck drivers effectively at the decision node. Moreover, dynamic road pricing schemes can be utilized to alter the route choices of truck drivers.

ACKNOWLEDGMENT

The authors would like to thank the port of Rotterdam for providing us with the Bluetooth data and Peter Knoppers for making the data available through ToGRIP-Bluetooth service. The first author would like to thank Taoufik Bakri and Ali Nadi for their help in processing the Bluetooth data.

REFERENCES

- [1] S. Hess, M. Quddus, N. Rieser-Schüssler, and A. Daly, "Developing advanced route choice models for heavy goods vehicles using GPS data," *Transportation Research Part E: Logistics and Transportation Review*, vol. 77, pp. 29–44, 2015.
- [2] A. J. Pel, M. C. J. Bliemer, and S. P. Hoogendoorn, "Hybrid Route Choice Modeling in Dynamic Traffic Assignment," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2091, pp. 100–107, 2009.
- [3] S. Peeta, J. L. Ramos, and R. Pasupathy, "Content of variable message signs and on-line driver behavior," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1725, pp. 102–108, 2000.
- [4] M. Pouloupoulou, I. Spyropoulou, and C. Antoniou, "Parameters Affecting Professional Driver Response to VMS Messages," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2015-October, 2015, pp. 1563–1568.
- [5] T. Arentze, T. Feng, H. Timmermans, and J. Robbroeks, "Context-dependent influence of road attributes and pricing policies on route choice behavior of truck drivers: Results of a conjoint choice experiment," *Transportation*, vol. 39, no. 6, pp. 1173–1188, 2012.
- [6] E. A. I. Bogers and H. J. Van Zuylen, "The importance of reliability in route choices in freight transport for various actors on various levels," in *Proceedings of the European Transport Conference*, 2004, pp. 149–161.
- [7] Y. Sun, T. Toledo, K. Rosa, M. E. Ben-Akiva, K. Flanagan, R. Sanchez, and E. Spissu, "Route choice characteristics for truckers," *Transportation Research Record*, no. 2354, pp. 115–121, 2013.
- [8] T. Toledo, Y. Sun, K. Rosa, M. Ben-Akiva, K. Flanagan, R. Sanchez, and E. Spissu, *Decision-Making Process and Factors Affecting Truck Routing*, 2013, pp. 233–249.
- [9] M. Rowell, A. Gagliano, and A. Goodchild, "Identifying truck route choice priorities: The implications for travel models," *Transportation Letters*, vol. 6, no. 2, pp. 98–106, 2014.
- [10] M. Kamali, A. Ermagun, K. Viswanathan, and A. R. Pinjari, "Deriving Truck Route Choice from Large GPS Data Streams," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2563, pp. 62–70, 2016.
- [11] T. D. Luong, D. Tahlyan, and A. R. Pinjari, "Comprehensive Exploratory Analysis of Truck Route Choice Diversity in Florida," *Transportation Research Record*, 2018.
- [12] J. H. Knorrning, R. He, and A. L. Kornhauser, "Analysis of route choice decisions by long-haul truck drivers," *Transportation Research Record*, no. 1923, pp. 46–60, 2005.
- [13] H. Oka, Y. Hagino, T. Kenmochi, R. Tani, R. Nishi, K. Endo, and D. Fukuda, "Predicting travel pattern changes of freight trucks in the Tokyo Metropolitan area based on the latest large-scale urban freight survey and route choice modeling," *Transportation Research Part E: Logistics and Transportation Review*, 2018.
- [14] M. E. Ben-Akiva, T. Toledo, J. Santos, N. Cox, F. Zhao, Y. J. Lee, and V. Marzano, "Freight data collection using GPS and web-based surveys: Insights from US truck drivers' survey and perspectives for urban freight," *Case Studies on Transport Policy*, vol. 4, no. 1, pp. 38–44, 2016.
- [15] A. M. Hainen, J. S. Wasson, S. M. L. Hubbard, S. M. Remias, G. D. Farnsworth, and D. M. Bullock, "Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles," *Transportation Research Record*, no. 2256, pp. 43–50, 2011.
- [16] J. van den Heuvel, A. Voskamp, W. Daamen, S. P. Hoogendoorn, M. Chraïbi, M. Boltes, A. Schadschneider, and A. Seyfried, "Using Bluetooth to Estimate the Impact of Congestion on Pedestrian Route Choice at Train Stations," Cham: Springer International Publishing, 2015, pp. 73–82.
- [17] T. Hyodo, N. Suzuki, and K. Takahashi, "Modeling of Bicycle Route and Destination Choice Behavior for Bicycle Road Network Plan," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1705, pp. 70–76, 2000.
- [18] S. Yang and Y.-J. Wu, "Travel mode identification using bluetooth technology," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, pp. 1–15, 2017.
- [19] B. N. Araghi, R. Krishnan, and H. Lahrmann, "Mode-Specific Travel Time Estimation Using Bluetooth Technology," *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, pp. 219–228, 2016.
- [20] S. Liu, J. McGree, G. White, and W. Dale, "Transport mode identification by clustering travel time data," *ANZIAM Journal*, vol. 56, pp. M95–M116, 2014.
- [21] J. W. Tukey, *Exploratory data analysis*. Reading, Mass. : Addison-Wesley, 1993.
- [22] D. Reynolds, "Gaussian Mixture Models," S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, 2009, pp. 659–663.
- [23] R. M. O'Brien, "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality & Quantity*, vol. 41, no. 5, pp. 673–690, 2007.
- [24] M. Ben-Akiva and M. Bierlaire, *Discrete Choice Models with Applications to Departure Time and Route Choice*, pages 7–37. Springer US, Boston, MA, 2003.
- [25] K. E. Train, *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press, 2009.
- [26] J. W. C. van Lint, H. J. van Zuylen, and H. Tu, "Travel time unreliability on freeways: Why measures based on variance tell only half the story," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 258–277, 2008.
- [27] M. Bierlaire, *PythonBiogeme: a short introduction*. Report TRANSP-OR 160706, Series on Biogeme, 2016.
- [28] S. Peeta and H. S. Mahmassani, "System optimal and user equilibrium time-dependent traffic assignment in congested networks," *Annals of Operations Research*, vol. 60, no. 1, pp. 81–113, 1995.
- [29] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S. Tadaki, and S. Yukawa, "Traffic jams without bottlenecks: experimental evidence for the physical mechanism of the formation of a jam," *New J. Phys.*, vol. 10, no. 3, pp. 033001, 2008.